Optimal Spectrum Sharing with ARQ based Legacy Users via Chain Decoding

Nicolò Michelusi†, Member, IEEE
†School of Electrical and Computer Engineering, Purdue University, USA, michelus@purdue.edu

Abstract

This paper investigates the design of access policies in spectrum sharing networks by exploiting the retransmission protocol of legacy primary users (PUs) to improve the spectral efficiency via opportunistic retransmissions at secondary users (SUs) and chain decoding [2]. The optimal access policy which maximizes the SU throughput under a maximum interference constraint to the PU and its performance are found in closed form. It is shown that the optimal policy randomizes among three modes: 1) Idle: the SU remains idle over the retransmission window of the PU, to avoid interfering; 2) Interference cancellation: the SU transmits only after decoding the PU packet, to improve its own throughput via interference cancellation; 3) Always transmit: the SU transmits over the retransmission window of the PU to maximize the future potential of interference cancellation via chain decoding. This structure is exploited to design a stochastic gradient descent algorithm to facilitate learning and adaptation in settings where the model parameters are unknown or vary over time, based on ARQ feedback from the PU and CSI measurements at the SU receiver. It is shown numerically that, for a 10% interference constraint, the optimal access policy yields 15% improvement over a state-of-the-art scheme without selective SU retransmissions, and up to $2\times$ gain over a scheme using a non-adaptive access policy instead of the optimal one.

I. INTRODUCTION

The recent proliferation of mobile devices has been exponential in number as well as heterogeneity, leading to spectrum crunch. The tremendous increase in demand of wireless services requires a shift in network design from exclusive spectrum reservation to spectrum sharing

This research has been funded in part by the grants NSF CNS-1642982 and DARPA #108818. Part of this work appeared at ISIT’17, see [1].
to improve spectrum utilization [3]. Cognitive radios [4] enable the coexistence of incumbent legacy users (primary users, PUs) and opportunistic users (secondary users, SUs) capable of autonomous reconfiguration by learning and adapting to the communication environment [5].

A central question is: how can opportunistic users leverage side information about nearby legacy users (e.g., activity, channel conditions, protocols employed, packets exchanged [6]) to opportunistically access the spectrum and improve their own performance, with minimal or no degradation to existing legacy users [7]? In this paper, we address this question in the context of the retransmission protocol employed by PUs. We consider a wireless network composed of a pair of PUs and a pair of SUs. The PU employs Type-I HARQ [8] to improve reliability, which results in replicas of the PU packet (re)transmitted over subsequent slots, henceforth referred to as ARQ window. With the scheme developed in [9], the SU receiver attempts to decode the PU packet independently in each slot, and replicas of the PU packet are not exploited; thus, in the example of Fig. 1 no SU packets can be decoded with the scheme [9]. However, the SU may leverage these replicas to improve its own throughput via interference cancellation. In [10], we have investigated a scheme, termed backward interference cancellation (BIC), where the SU receiver decodes the PU packet and removes its interference to achieve interference-free transmissions over the entire ARQ window of that PU packet. In the example of Fig. 1, this scheme allows the SU receiver to decode packet S4, after removing the interference of P2, decoded in slot 5, thus outperforming [9].

In [2], we have advanced this concept by allowing the SU to opportunistically retransmit SU packets and buffer the corrupted signals at the SU receiver. In fact, if a previously transmitted and failed SU packet is decoded at the SU receiver, its interference can be removed from previous retransmission attempts of the same, thus facilitating the decoding of the concurrent PU packets; in turn, the interference of these PU packets can be removed to facilitate the decoding of SU packets over their respective ARQ windows. This scheme continues in chain, until no more packets can be decoded, hence the name chain decoding (CD) [2]. In the example of Fig. 1, the retransmission of S3 in slot 4 allows the SU receiver to chain the ARQ windows of P1 and P2, so that all 3 SU packets S1-S3 can be decoded, versus only one decoded with BIC, and none decoded with the scheme in [9]. However, note that the decoding of S1 is delayed by 4 slots. Therefore, the throughput improvement of chain decoding comes at a latency cost in the delivery of SU packets, hence it is suitable for latency-tolerant applications, such as monitoring sensor networks as in [11] and video streaming, see [12, Fig. 1] for a list of potential use cases.
Fig. 1: Example of chain decoding and comparison with BIC [10] and the scheme in [9]. The SU fails in slots 1-4. SUrx decodes P2 in slot 5. With the scheme in [9], SUrx does not remove the interference of P2 in slot 4, no SU packets are decoded. With BIC, SUrx removes the interference of P2 from slot 4 to decode S4, thus decoding 1 SU packet. With chain decoding, SUrx retransmits S3 in slot 4; after decoding P2 in slot 5, chain decoding is initiated: SUrx removes the interference of P2 from slot 4 to decode S3; hence, it removes the interference of S3 from slot 3 to decode P1; finally, it removes the interference of P1 from slots 1-2 to decode S1-S2; overall, SUrx decodes 3 SU packets.

Additionally, as explained in Sec. II-C, chain decoding requires a buffering mechanism at the SU receiver, whereas no buffering is required in [9].

While our previous work [2] proves the optimality of a CD protocol, which dictates the retransmission process at the SU to maximize the potential of interference cancellation at its receiver, it does not investigate the design of an optimal SU access scheme (i.e., whether the SU should transmit or remain idle). Such design question, not addressed in [2] but investigated in this paper, is of great practical interest. In fact, as we will show in Sec. V, information on the structure of the optimal SU access scheme may be exploited to significantly reduce the policy search space and the optimization complexity, thus facilitating learning and adaptation in scenarios where the statistics are unknown or vary over time. Based on the underlay paradigm [13], in this paper we investigate the optimal SU access policy so as to maximize the SU throughput via CD, under an interference constraint to the PU.

We derive the optimal policy and its performance in closed form for the case where the PU enforces reliability, and show that the optimal SU access policy reflects a randomization among
three modes of operation: 1) The SU remains idle over the entire ARQ window; 2) The SU transmits only after its receiver decodes the PU packet; 3) The SU always transmits over the ARQ window. With mode 1), the SU does not interfere with the PU; with mode 2), the SU leverages knowledge of the PU packet to perform interference cancellation and create an interference-free channel for its own data transmission; with mode 3), the SU leverages the full potential of successive interference cancellation via CD over the ARQ window. The optimal randomization among these three modes reflects a strike between maximizing the SU throughput and minimizing the interference caused to the PU. We show numerically that, for a 10% interference constraint, the optimal access policy under CD attains a throughput gain of 15% with respect to BIC, and up to $2\times$ improvement over a CD scheme using a non-adaptive access policy. Importantly, the optimal policy does not require knowledge of the statistics of the model, but only an estimate of the interference level perceived at the PU receiver (for instance, estimated by monitoring the ACK/NACK feedback signal [14]). This feature facilitates learning and adaptation when the statistics of the system are unknown or vary over time. For these scenarios, we present a stochastic optimization framework, where the SU learns the optimal randomization and its transmit rate based solely on ARQ feedback from the PU and CSI measurements at the SU receiver. We prove the effectiveness of this strategy numerically.

Other previous work leverage the retransmission protocol of the PU [9], [14]–[18]. In [15], [16], the primary ARQ process is limited to one retransmission, with incremental redundancy and packet combining, respectively, assuming a slow-fading scenario. In [9], it is shown that the SU throughput is maximized by concentrating the interference to the PU in the first transmissions of the PU packet, but the temporal redundancy of ARQ is not exploited to cancel interference. In [18], the SUs cooperate with the PU by assisting retransmissions of failed packets using distributed orthogonal space-time block code; however, knowledge of the PU packet is not exploited at the SU receiver to perform interference cancellation. Differently from [18], we assume no cooperation with the PU at the SU transmitter, but only interference cancellation at the SU receiver. Differently from these works, in this paper we consider multiple retransmissions (in contrast to [15], [16]), and we exploit the redundancy of the ARQ process (in contrast to [9], [18]). While in this paper, like in [19], we assume that the ARQ feedback is overheard by the SU without errors, practical aspects related to imperfect sensing are considered in [18], [20]. While non-causal knowledge of the PU packet is assumed in [17], in our work we model the dynamic acquisition of the PU packet at the SU receiver, which is of more practical interest.
In [14], the SU exploits ARQ feedback to estimate the throughput loss of the PU and tunes its transmission policy accordingly, based on information theoretic results. In [19], the PU adapts its transmitted power in response to interference; the SU uses the feedback control channel from the PU to control its interference. In this paper, instead, we leverage the structure of the optimal policy to design a simple but effective learning algorithm based on *stochastic gradient descent* [21, Chapter 14], as opposed to approaches based on reinforcement learning [22], which suffer from slow convergence rate due to the need to explore the action and state spaces.

This paper is organized as follows. In Sec. II, we describe the system model; in Sec. III, we introduce the performance metrics and optimization problem. In Sec. IV, we provide the analytical results. In Sec. V, we present the stochastic optimization framework, and in Sec. VI we present numerical results. In Sec. VII we provide concluding remarks.

II. SYSTEM MODEL

The main parameters of the model are given in Table I. We consider a two-user interference network, depicted in Fig. 2, where a primary transmitter PUtx and a secondary transmitter SUtx transmit to their respective receivers, PUrx and SUrx, and generate mutual interference.

Time is divided into slots of fixed duration $\Delta$, corresponding to the transmission of one data packet and the feedback signal from the receiver, see Fig. 1. We assume a block-fading channel.
model, \textit{i.e.}, the channel gains are constant within each slot, i.i.d. over time and independent across links. SUtx and PUtx transmit with constant powers $P_s$ and $P_p$, respectively. $P_s$ may be based on an interference temperature threshold experienced at the PU receiver \cite{19}, and can be estimated using techniques developed in \cite{11}. Assuming AWGN noise at the receivers, we define the SNR of the links SUtx→SUrx, PUtx→PUrx, SUtx→PUrx and PUtx→SUrx at time $t$ as $\gamma_{s,t}, \gamma_{p,t}, \gamma_{sp,t}, \gamma_{ps,t}$, i.i.d. over time and with mean $\bar{\gamma}_s, \bar{\gamma}_p, \bar{\gamma}_{sp}, \bar{\gamma}_{ps}$, respectively.

No channel state information is available at the transmitters. Thus, PUtx transmits with fixed rate $R_p$ [bits/s/Hz], and is data backlogged. SUtx transmits with fixed rate $R_s$ [bits/s/Hz], or remains idle to reduce the interference caused to the PU. We denote the SU access decision in slot $t$ as $a_{S,t} \in \{0, 1\}$, selected according to access policy $\mu$, introduced in Sec. \textbf{II-D}. Thus, $a_{S,t}=1$ if SUtx transmits, and $a_{S,t}=0$ if it remains idle.

We assume that the SU knows the signal characteristics of the PU and is accurately synchronized with the PU system. This is a common assumption in the literature \cite{6}, \cite{9}, \cite{15}, \cite{16}, \cite{18}. The modulation type can be inferred using signal processing techniques such as cyclostationary feature detection \cite{20} or deep neural networks \cite{23}. The codebook information may be obtained if the PUs follow a uniform standard for communication based on a publicized codebook, or periodically broadcast it \cite{13}. Moreover, the SUs perform timing, carrier synchronization and channel equalization by leveraging pilots, preambles, synchronization words or spreading codes used by PUs for coherent detection \cite{20}. The SU pair can use this information to synchronize with the PU system to detect ARQ feedback messages, decode the PU packet, and then reconstruct the PU transmit signal to perform interference cancellation, \textit{e.g.}, using techniques developed in \cite{24}.

Let $\rho_{a_{S,t}}$ be the failure probability for the PU as a function of $a_{S,t} \in \{0, 1\}$. We assume that $\rho_0 < \rho_1$, since transmissions of the PU are more likely to fail under interference from the SU. The PU employs retransmissions (ARQ) in case of transmission failure \cite{8} and enforces perfect reliability, so that a packet is retransmitted until, eventually, it is received successfully. We will evaluate the effect of a finite ARQ deadline numerically in Sec. \textbf{VI}. At the end of slot $t$, PUrx sends a feedback message $y_{P,t} \in \{\text{ACK, NACK}\}$ to PUtx over a dedicated control channel, to notify it about the transmission outcome and, possibly, request a retransmission (NACK). This message is received without error by PUtx and overheard by the SU pair.

We assume all codewords are drawn from a Gaussian codebook, and are sufficiently long to allow reliable decoding whenever the attempted rate is within the mutual information of the
channel. Let $C(SNR) \triangleq \log_2(1 + SNR)$ be the capacity of the Gaussian channel as a function of the SNR at the receiver [25]. We denote the packets being transmitted by the SU and PU with their labels $l_S$ and $l_P$, respectively ($l_S = 0$ if the SU remains idle). We now describe the SU system.

### A. Decoding outcomes at SUrx

The decoding performance at SUrx depends on whether $l_P$ is known or not at SUrx to perform interference cancellation, as a result of a previous successful decoding operation, and on the access decision $a_{S,t} \in \{0, 1\}$, as detailed below.

1) **Case $a_{S,t} = 1$, $l_P$ unknown:** SUrx attempts to decode both $l_S$ and $l_P$. Since SUtx, PUtx and SUrx form a multiple access channel [25], the outcomes at the receiver for a given rate pair $(R_s, R_p)$, as a function of the SNRs $(\gamma_s, \gamma_{ps})$, are as depicted in Fig. 3. We denote their probabilities as

\[
\begin{align*}
\{1\} & : \delta_{sp} \triangleq \mathbb{P}(l_P \& l_S \text{ decoded}), & \{4\} & : v_s \triangleq \mathbb{P}(l_P \rightarrow l_S), \\
\{2\} & : \delta_s \triangleq \mathbb{P}(\text{only } l_S \text{ decoded}), & \{5\} & : v_p \triangleq \mathbb{P}(l_S \rightarrow l_P), \\
\{3\} & : \delta_p \triangleq \mathbb{P}(\text{only } l_P \text{ decoded}), & \{6\} & : v_{sp} \triangleq \mathbb{P}(l_S \leftrightarrow l_P), \\
\{7\} & : v_\emptyset \triangleq \mathbb{P}(\text{failure}),
\end{align*}
\]

Let $C(SNR) = \log_2(1 + SNR)$ be the capacity of the Gaussian channel as a function of the SNR at the receiver [25]. We denote the packets being transmitted by the SU and PU with their labels $l_S$ and $l_P$, respectively ($l_S = 0$ if the SU remains idle). We now describe the SU system.

### A. Decoding outcomes at SUrx

The decoding performance at SUrx depends on whether $l_P$ is known or not at SUrx to perform interference cancellation, as a result of a previous successful decoding operation, and on the access decision $a_{S,t} \in \{0, 1\}$, as detailed below.

1) **Case $a_{S,t} = 1$, $l_P$ unknown:** SUrx attempts to decode both $l_S$ and $l_P$. Since SUtx, PUtx and SUrx form a multiple access channel [25], the outcomes at the receiver for a given rate pair $(R_s, R_p)$, as a function of the SNRs $(\gamma_s, \gamma_{ps})$, are as depicted in Fig. 3. We denote their probabilities as

\[
\begin{align*}
\{1\} & : \delta_{sp} \triangleq \mathbb{P}(l_P \& l_S \text{ decoded}), & \{4\} & : v_s \triangleq \mathbb{P}(l_P \rightarrow l_S), \\
\{2\} & : \delta_s \triangleq \mathbb{P}(\text{only } l_S \text{ decoded}), & \{5\} & : v_p \triangleq \mathbb{P}(l_S \rightarrow l_P), \\
\{3\} & : \delta_p \triangleq \mathbb{P}(\text{only } l_P \text{ decoded}), & \{6\} & : v_{sp} \triangleq \mathbb{P}(l_S \leftrightarrow l_P), \\
\{7\} & : v_\emptyset \triangleq \mathbb{P}(\text{failure}),
\end{align*}
\]
Fig. 3: Decoding regions at SUrx as a function of \((\gamma_s, \gamma_{ps})\). The boundaries correspond to the decoding thresholds of the multiple access channel \([25]\).

computed as the marginals with respect to the distribution of \((\gamma_s, \gamma_{ps})\). In \{1\}, \(l_s\) and \(l_P\) are jointly decoded. In \{2\} (respectively, \{3\}), only \(l_s\) (\(l_P\)) is decoded, by treating the interfering \(l_P\) (\(l_s\)) as noise. In \{4\}, \{5\} or \{6\}, neither \(l_s\) nor \(l_P\) can be currently decoded by SUrx; however, one packet can be decoded after removing the interference from the other. The arrow \(l_X \rightarrow l_Y\) indicates the decoding dependence between \(l_X\) and \(l_Y\), so that \(l_Y\) can be decoded after removing the interference from \(l_X\), but not vice-versa (unless \(l_X \leftrightarrow l_Y\)). In these three cases, the received signal is buffered at SUrx for future recovery via chain decoding, see Sec. [II-C] Finally, in \{7\}, the channel quality is poor, so that neither \(l_s\) nor \(l_P\) can be decoded by SUrx (even after removing their mutual interference) and the signal is discarded.

2) Case \(a_{S,t} = 1\), \(l_P\) known: \(l_P\) is known at SUrx as a result of a previous decoding operation at SUrx, its interference can be removed from the received signal, thus creating an interference-free channel to decode \(l_s\). Therefore, the SU transmission succeeds if \(R_s < C(\gamma_{s,t})\). Since this
event is the union of the four disjoint events \( \{1\}, \{2\}, \{4\} \) and \( \{6\} \) (right side of the decoding threshold \( R_s = C(\gamma_s) \) in Fig. 3), its probability is obtained via (1) as
\[
D_s \triangleq \mathbb{P}(R_s < C(\gamma_{s,t})) = \delta_s + \delta_{sp} + \nu_s + \nu_{sp}.
\] (2)

3) Case \( a_{S,t} = 0, l_P \) unknown: SUtx remains idle and SUrx attempts to decode \( l_P \). Thus, SUrx decodes \( l_P \) successfully if \( R_p < C(\gamma_{ps,t}) \). Since this event is the union of the four disjoint events \( \{1\}, \{3\}, \{5\} \) and \( \{6\} \) (region above the decoding threshold \( R_p = C(\gamma_{ps}) \) in Fig. 3), its probability is obtained via (1) as
\[
D_p \triangleq \mathbb{P}(R_p < C(\gamma_{ps,t})) = \delta_p + \delta_{sp} + \nu_p + \nu_{sp}.
\] (3)

4) Case \( a_{S,t} = 0, l_P \) known: no decoding activity at SUrx.

B. Decoding feedback from SUrx

At the end of each slot, SUrx feeds back \( y_{S,t} \in \{1, \ldots, 7\} \) to SUtx over a dedicated error-free control channel, indicating one of the regions of Fig. 3 (the numbering is given in (1) as \( \{j\} \)). This feedback signal, together with the ARQ feedback signal received from PUrx, allows SUtx to keep track of the chain decoding state, the buffering of corrupted signals and the knowledge of the current PU packet \( l_P \) at SUrx.

C. SU retransmissions, buffering and chain decoding

The SU performs retransmissions and buffering at SUrx to improve the potential of interference cancellation at SUrx. For instance, if \( l_S \rightarrow l_P \) or \( l_P \leftrightarrow l_S \) in a previous slot, the SU may retransmit \( l_S \). If \( l_S \) is decoded by SUrx, its interference can be removed from the previously buffered received signal to recover \( l_P \). In turn, the recovered \( l_P \) may be exploited to recover other SU packets from previously buffered signals received within the ARQ window associated to \( l_P \), via interference cancellation, see example in Fig. 1. The iterative application of interference cancellation on signals buffered at SUrx, is denoted as chain decoding (CD). Thus, when \( l_P \rightarrow l_S, l_S \rightarrow l_P \) or \( l_P \leftrightarrow l_S \), with probability \( \nu_s, \nu_p \) and \( \nu_{sp} \), respectively, SUrx buffers the corresponding received signals. For analytical tractability, we assume an infinite buffer at SUrx. We will evaluate the effect of a finite buffer size numerically in Sec. VI.
Definition 1. The decoding relationship among the SU and PU packets buffered at SUrx is represented by the CD graph, with vertices the set of undeveloped packets, and edges the decoding relationship between them. For instance, if \( l_s \rightarrow l_p \), then \( l_s \) and \( l_p \) are vertices in the CD graph, connected by a directed edge from \( l_s \) to \( l_p \). We define the CD root as the SU packet which, once decoded, triggers the recovery of the largest number of SU packets via CD (see Fig. 4).

The retransmission process at the SU is governed by the packet selection policy: if \( a_{s,t} = 1 \), it selects which SU packet to (re)transmit based on the structure of the CD graph at SUrx. In [2], we have shown that the optimal packet selection policy follows a chain decoding protocol, assumed in the rest of the paper. We refer to [2] for details and proof of optimality. Herein, we describe the CD protocol with the help of Fig. 4. At the start of a new ARQ window (the PU transmits a new packet), the PU packet is unknown at SUrx and the configuration is depicted in Fig. 4.a. The CD graph evolves over the ARQ window, leading to one of the configurations in Fig. 4.a-g. In the configuration of Fig. 4.a, the PU packet and the CD root are not connected:
the CD protocol dictates to retransmit the CD root, so as to maximize the chances of either decoding it, or connecting it to the PU packet (when \( l_P \rightarrow l_S, l_S \rightarrow l_P \) or \( l_P \leftrightarrow l_S \)), leading to one of the configurations in Fig. 4.b,d,f. If the CD root is decoded, then the CD graph is decoded via chain decoding. If \( l_P \rightarrow l_S, l_S \rightarrow l_P \) or \( l_P \leftrightarrow l_S \), respectively, the new configuration becomes the one depicted in Fig. 4.b with \( b = 1 \), Fig. 4.f with \( b = 0 \) or Fig. 4.d with \( b = 0 \), respectively. Finally, if the PU packet is decoded, the new configuration becomes the one depicted in Fig. 4.g. Once the CD root and the PU packet are connected (Figs. 4.b-f), it is optimal to transmit a new packet: this choice maximizes the chance of connecting it to the CD graph and leverage interference cancellation in the future; retransmitting the CD root would be redundant, since it is already connected to the CD graph. In the configuration of Fig. 4.g, the PU packet is known at SUrx, and thus its interference is cancelled. In this case, it is optimal to retransmit the CD root to maximize the chances of decoding the CD graph, by taking advantage of the interference-free channel.

D. SU access policy

At the beginning of slot \( t \), given the history up to slot \( t \), \( H^t = (y_{P,0}^{t-1}, y_{S,0}^{t-1}, a_{S,0}^{t-1}) \), SUtx selects \( a_{S,t} = 1 \) in slot \( t \) with probability \( \mu_t(H^t) \), and \( a_{S,t} = 0 \) otherwise, where \( \mu_t \) denotes the access policy. If \( a_{S,t} = 1 \), the CD protocol described in Sec. II-C dictates which packet to transmit.

III. OPTIMIZATION PROBLEM

We define the average long-term PU throughput as

\[
\bar{T}_P(\mu) \triangleq \lim_{D \to \infty} R_p E \left[ \frac{1}{D} \sum_{t=0}^{D-1} (1 - \rho_{a_{S,t}}) \right],
\]

where \( D \to \infty \) is the horizon length, the expectation is taken with respect to the sequence \( \{a_{S,t}, t \geq 0\} \) generated by the access policy \( \mu \), and the decoding outcomes at SUrx and PUrx. This metric is equivalent to the "stable throughput", which guarantees stability in a systems where packets generated in upper layers are stored in queues before transmission [18]. Since \( \rho_{a_{S,t}} = \rho_0 + a_{S,t}(\rho_1 - \rho_0) \), one can rewrite

\[
\bar{T}_P(\mu) = \bar{T}_{P,\text{max}} \left[ 1 - \nabla(\mu) \right],
\]
where $T_{P,\text{max}} \triangleq R_p (1 - \rho_0)$ is the maximum PU throughput, achievable when the SU remains always idle, and we have defined the PU throughput degradation, relative to the maximum throughput $T_{P,\text{max}}$, as

$$\nabla(\mu) \triangleq \frac{\rho_1 - \rho_0}{1 - \rho_0} \lim_{D \to \infty} \mathbb{E} \left[ \frac{1}{D} \sum_{t=0}^{D-1} a_{S,t} \right].$$

(6)

We can interpret $\nabla(\mu)$ as the throughput loss experienced by the PU as a result of the activity of the SU, which should be limited to reflect higher layer QoS constraints [19]. Similarly, we define the average long-term SU throughput as

$$\bar{T}_S(\mu) \triangleq \lim_{D \to \infty} \mathbb{E} \left[ \frac{1}{D} \sum_{t=0}^{D-1} r_{S,t} \right],$$

(7)

where $r_{S,t}$ is the instantaneous throughput accrued via CD.

The goal is to design the SU access policy $\mu$ so as to maximize $\bar{T}_S(\mu)$, subject to a maximum PU throughput degradation constraint $\nabla \in (0, 1)$ (alternatively, subject to a minimum PU throughput $T_{P,\text{min}} = T_{P,\text{max}} (1 - \nabla)$ via (5)),

$$\text{OP} : \mu^* = \arg \max_{\mu} \bar{T}_S(\mu), \text{ s.t. } \nabla(\mu) \leq \nabla.$$

(8)

Note from (6) that $\nabla(\mu)$ is maximum when $a_{S,t} = 1, \forall t$, yielding $\nabla(\mu) \leq \nabla_{\text{max}} \triangleq \frac{\rho_1 - \rho_0}{1 - \rho_0}$; then, if $\nabla \geq \nabla_{\text{max}}$, the constraint in OP becomes inactive.

The SU throughput and interference are difficult to compute in this form, since the outcome of CD depends on the specific instance of the CD graph. As shown in [2], a simplification can be obtained using the concept of virtual decodability.

**Definition 2.** A packet $l$ in the CD graph is virtually decodable if it becomes decodable by initiating CD at the CD root, following the directed edges in the CD graph (CD root excluded). Otherwise, we say it is virtually undecodable.

Based on this definition, if $l$ is virtually decodable and the CD root is decoded, then $l$ is also decoded via CD. Therefore, if one guarantees to decode the CD root with probability one, eventually any virtually decodable $l$ will also be decoded. Indeed, this is the case: according to the optimal CD rules [2], as explained in Sec. II-C, the CD root is retransmitted (at least) at the beginning of each ARQ window (Fig. 4.a). Thus, eventually, it will be decoded, triggering the
decoding of the entire CD graph via chain decoding; thus, \( l \) can be considered virtually decoded, even if it has not been currently decoded. As a result, there is no loss of generality, in terms of average throughput, if one counts the virtually decodable packets at the present time, rather than at the future time when they are actually decoded via CD. Based on this intuition, in [2] we have shown that \( \bar{T}_S(\mu) \) can be expressed as

\[
\bar{T}_S(\mu) = \lim_{D \to \infty} \inf \mathbb{E} \left[ \frac{1}{D} \sum_{t=0}^{D-1} v_S(a_{S,t}, s_t) \right],
\]

where \( v_S(a_{S,t}, s_t) \) is the expected virtual instantaneous throughput (which counts the virtually decoded SU packets in addition to the currently decoded ones), whose analytical expression is provided in Sec. [IV-A] and \( s_t = (\Phi, b) \) is the state of the CD protocol:

- \( \Phi \) denotes the virtual knowledge of the current PU packet \( l_P \) at SUrx, and takes values in the set \{\( \rightarrow K \), \( \leftarrow K \), \( U \)\}. "K" denotes that the current PU packet \( l_P \) is virtually decodable at SUrx (i.e., either it has been decoded in a previous slot by SUrx as in Fig. 4g, or it is virtually decodable as in Fig. 4c-f); in contrast, "U" denotes the complementary event that \( l_P \) is virtually undecodable (Fig. 4a-b). The unidirectional or bidirectional arrow above "K" indicates the type of edge connecting \( l_P \) to the CD root. In particular, \( \Phi = \rightarrow K \) (Fig. 4c-d) indicates that \( l_P \) and the CD root are mutually decodable after removing their respective interference, i.e., \( l_P \leftrightarrow [\text{CD root}] \); \( \Phi = \leftarrow K \) indicates that either \( l_P \) is known (Fig. 4g), or it can be decoded via CD after decoding the CD root, but not vice versa ([CD root] \( \rightarrow l_P \) but not \( l_P \rightarrow [\text{CD root}] \), Fig. 4e-f).

- if \( \Phi = U \) (\( l_P \) is virtually undecodable), \( b \) denotes the number of virtually undecoded SU packets \( l^{[1]}_S, l^{[2]}_S, \ldots, l^{[b]}_S \) transmitted within the current ARQ window of \( l_P \), such that \( l_P \rightarrow l^{[i]}_S \), as in Fig. 4b. If \( l_P \) becomes virtually decodable in state \( U \), then its interference can be virtually removed, and thus \( l^{[i]}_S, i = 1, 2, \ldots, b \) become virtually decodable as well; in contrast, if \( l_P \) is not virtually decoded within the end of its ARQ window, then \( l^{[1]}_S, \ldots, l^{[b]}_S \) remain undecoded and are discarded, since \( l_P \) will not be transmitted again. Reliability of these SU packets may be enforced via retransmissions requested by higher layer protocols. We set \( b = 0 \) when \( \Phi \in \{ \rightarrow K, \leftarrow K \} \), since \( l_P \) is virtually decodable in these cases and the SU channel is, virtually, interference free.

When a new ARQ window begins, the new PU packet is virtually undecodable and \( b = 0 \) (Fig. 4a), hence the new state becomes \((U, 0)\).
IV. Analysis

Under this equivalent formulation, the operation of the SU is a Markov decision process (MDP) [26], with state \( s_t \in S \) infinite (but countable) state space

\[
S = \{ b | b \geq 0 \} \cup \{ \leftrightarrow K, \rightarrow K \}, \quad (10)
\]

action \( a_{S,t} \in \{ 0, 1 \} \), reward \( v_S(a_{S,t}, s_t) \) (to compute the SU throughput (9)) and cost \( \frac{\rho_1 - \rho_0}{1 - \rho_0} a_{S,t} \) (to compute the PU throughput degradation (6)). Thus, the optimal solution of \( \text{OP} \) is a stationary and state-dependent policy [27], \( \mu_t(\mathcal{H}_t) = \mu(s_t), \forall t \). We let \( \mathcal{U} \) be the set of such policies,

\[
\mathcal{U} \equiv \{ \mu : S \mapsto [0, 1] \}. \quad (11)
\]

The transition probabilities and reward \( v_S(a_{S,t}, s_t) \) of the MDP are characterized in Sec. IV-A. Then, in Sec. IV-B, we investigate the optimal SU access policy.

A. Virtual throughput and transition probabilities

In state \( b \) (Fig. 4.a-b), \( v_S(a_S, b) \) is given by\(^2\)

\[
v_S(a_S, b) = R_s [a_S (\delta_{sp} + \delta_s) + D_p b]. \quad (12)
\]

In fact, with probability \( D_p \), \( l_P \) becomes virtually decodable, along with the \( b \) buffered SU packets connected to it; if \( a_S = 1 \), \( l_S \) is decoded successfully with probability \( \delta_{sp} + \delta_s \), due to the interference from the PU signal.

Remark 1. Note that the probability of virtually decoding \( l_P \) is \( D_p \), irrespective of whether SUtx transmits or remains idle; in fact, when SUtx transmits \( l_S, l_P \) is decoded with probability \( \delta_p + \delta_{sp} \) (see Fig. 3), and it is virtually decoded if \( l_S \rightarrow l_P \) (with probability \( v_p \)) or \( l_P \leftrightarrow l_S \) (with probability \( v_{sp} \)), yielding \( D_p = \delta_p + \delta_{sp} + v_p + v_{sp} \) as the overall probability of (possibly, only virtually) decoding \( l_P \). Then, the virtual decodability of the PU packet is not hampered by the interference caused by SU’s own signal.

\(^1\)For a more compact notation, we write state \((U, b)\) as \( b \), \((\leftrightarrow K, 0)\) as \( \leftrightarrow \) and \((\rightarrow K, 0)\) as \( \rightarrow \).

\(^2\)Note that the SU packets in the “Old CD graph” do not appear in the expression of the virtual throughput, since they have already been virtually decoded in previous ARQ windows.
In state $\leftarrow K$ (Fig. 4.c-d),

$$v_S(a_S, \leftarrow K) = R_s[a_S(D_s - \nu_{sp}) + D_p].$$

(13)

In fact, since $l_P$ is virtually decodable, $l_S$ can be decoded successfully with probability $D_s$ (since the channel is, virtually, interference-free), thus accruing the term $a_SD_s$. With probability $[a_S(\delta_p + \delta_{sp}) + (1-a_S)D_p]$, $l_P$ is decoded; it follows that the CD root is decoded (since $[\text{CD root}]\leftrightarrow l_P$ in state $\Phi = \leftarrow K$), thus accruing one unit of throughput. Finally, with probability $a_S\nu_p$, the transmission outcome is such that $l_S \rightarrow l_P$; it follows that $l_S$ becomes the new CD root leading to the new configuration of Fig. 4.e, and the previous CD root is virtually decoded (since $[\text{previous CD root}]\leftrightarrow l_P$ in state $\Phi = \leftarrow K$), thus accruing one unit of throughput (see [2]).

We obtain (13) by adding up all these terms. Finally, in state $\rightarrow K$,

$$v_S(a_S, \rightarrow K) = a_SR_sD_s,$$

(14)

since $l_P$ is virtually decodable and the channel is (virtually) interference-free. We now derive the transition probabilities $P(s_{t+1} = x|s_t = s, a_{S,t} = a_S) \triangleq P(x|s, a_S)$, by adapting those in [2] to the model of this paper, with backlogged PU and infinite ARQ deadline. From state $b \geq 0$,

$$P(x|b, a_S) = \begin{cases} 
1 - \rho_{aS}(D_p + a_S\nu_s), & x = 0, \text{ if } b = 0, \\
1 - \rho_{aS}, & x = 0, \text{ if } b > 0, \\
\rho_{aS}(1 - D_p - a_S\nu_s), & x = b, \text{ if } b > 0, \\
\rho_1a_S\nu_s, & x = b + 1, \\
\rho_1a_S\nu_{sp}, & x = \leftarrow K, \\
\rho_{aS}(D_p - a_S\nu_{sp}), & x = \rightarrow K. 
\end{cases}$$

(15)

In fact, the PU transmission succeeds with probability $1 - \rho_{aS}$; in this case, a new ARQ window begins with a new PU packet, which is virtually undecodable to the SU, so that the new state becomes $x = 0$. If the transmission outcome is such that $l_P \rightarrow l_S$ and the PU fails, then the signal is buffered and $b$ increases by one unit, thus the state becomes $x = b + 1$. If the PU fails and $l_P$ is virtually decoded (with probability $\rho_{aS}D_p$), then the new state becomes $x = \leftarrow K$ or $x = \rightarrow K$, depending on whether $l_S \rightarrow l_P$ (with probability $D_p - a_S\nu_{sp}$) or $l_S \leftrightarrow l_P$ (with probability

Note that in the configurations of Fig. 4.c-f, the $b$ SU packets such that $l_P \rightarrow l_S^{[1]}, l_S^{[2]}, \ldots, l_S^{[b]}$ are not counted in the virtual throughput, since they have been already virtually decoded in the transitions leading to these configurations (e.g., from Fig. 4.b to Fig. 4.c).

January 25, 2018
\(a_Sv_{sp}\), respectively. Otherwise, the state remains \(x = b\). From state \(\text{\textarrow{K}}\),

\[
P(x|\text{\textarrow{K}},a_S) = \begin{cases} 
1 - \rho_{a_S}, & x = 0, \\
\rho_{a_S} \left(1 - D_p + a_Sv_{sp}\right), & x = \text{\textarrow{K}}, \\
\rho_{a_S} \left(D_p - a_Sv_{sp}\right), & x = \text{\textarrow{K}}.
\end{cases}
\]  

(16)

In fact, the PU transmission succeeds with probability \(1 - \rho_{a_S}\) and the new state becomes \(x = 0\). If the PU fails and the decoding outcome is such that \(l_S \rightarrow l_P\), \(l_S\) becomes the new CD root, and the new state becomes \(x = \text{\textarrow{K}}\). Otherwise, the state does not change. From state \(\text{\textarrow{K}}\),

\[
P(x|\text{\textarrow{K}},a_S) = \begin{cases} 
1 - \rho_{a_S}, & x = 0, \\
\rho_{a_S}, & x = \text{\textarrow{K}}.
\end{cases}
\]  

(17)

In fact, with probability \(1 - \rho_{a_S}\) the PU succeeds, and a new ARQ window begins (state 0). Otherwise, the state does not change.

**B. Optimal SU access policy**

In this section, we derive the optimal SU access policy \(\mu^*\) and its performance in closed form. The main result is given in Theorems 1 and 2 whose proof is provided in Sec. IV-C. We let \(\bar{T}_S^{(GA)}(\epsilon) \triangleq \epsilon R_s D_s\) be the genie-aided SU throughput when SUrx has non-causal knowledge of the PU packets and can remove their interference (hence the success probability is \(D_s\) in each slot), and SUtx transmits with probability \(\epsilon = \min \left\{ \frac{\nabla_{th}}{\nabla_{\max}}, 1 \right\}\) to attain the constraint \(\nabla_{th}\). We let \(\pi_{\mu}\) be the steady-state probability of the MDP under policy \(\mu\).

Being genie-aided, \(\bar{T}_S^{(GA)}(\epsilon)\) is an upper bound to the SU throughput. A simple scheme to attain it is as follows: SUtx remains idle until SUrx decodes the PU packet (state \(\text{\textarrow{K}}\)); hence, it transmits with probability \(\mu(\text{\textarrow{K}})\) until the end of the ARQ window. By transmitting only in state \(\text{\textarrow{K}}\) when \(l_P\) is known at SUrx, the genie-aided throughput \(\bar{T}_S^{(GA)}\) is attained since SUrx can remove the interference of \(l_P\) from the received signal, as in the genie-aided case. If the PU throughput degradation constraint \(\nabla_{th}\) augments, the access probability in state \(\text{\textarrow{K}}\) can be increased accordingly, so as to accrue larger SU throughput, until it becomes \(\mu(\text{\textarrow{K}}) = 1\). At this point, transitions from state 0 (where the SU remains idle) to state \(\text{\textarrow{K}}\) occur with probability \(\rho_0 D_p\) (SUrx decodes the PU packet and the PU requests a retransmission); transitions from state \(\text{\textarrow{K}}\) (where the SU transmits) to state 0 occur with probability \((1 - \rho_1)\) (the PU succeeds and a new
ARQ window begins), hence \( \pi_\mu(\vec{K}) = \rho_0 D_p / [1 - \rho_1 + \rho_0 D_p] \) at steady-state and the SU transmits over a fraction \( \pi_\mu(\vec{K}) \) of the slots, yielding

\[
\nabla_{GA} \triangleq \pi_\mu(\vec{K}) \nabla_{max} = \frac{\rho_0 D_p}{1 - \rho_1 + \rho_0 D_p}
\]  

(18)
as the PU throughput degradation. This fact yields the following theorem.

**Theorem 1.** If \( \nabla_{th} \leq \nabla_{GA} \), then

\[
\begin{align*}
\mu^*(0) &= 0, \\
\mu^*(\vec{K}) &= \mu^*(b) = 1, \quad \forall b > 0, \\
\mu^* (\vec{K}) &= \frac{[1 - \rho_0 (1 - D_p)] \nabla_{th}}{[1 - \rho_0 (1 - D_p)] \nabla_{GA} + (\rho_1 - \rho_0) (\nabla_{th} - \nabla_{GA})};
\end{align*}
\]  

(19)

under such policy,

\[
\bar{T}_S(\mu^*) = \bar{T}^{(GA)}_S \left( \frac{\nabla_{th}}{\nabla_{max}} \right), \quad \nabla(\mu^*) = \nabla_{th}.
\]  

(20)

**Proof.** (20) shows that the genie-aided throughput \( \bar{T}_S^{(GA)}(\epsilon) \) is achievable under policy (19) when \( \nabla_{th} \leq \nabla_{GA} \). Indeed, since \( \mu^*(0) = 0 \), from (15) with \( b = 0 \) it follows that the transition probability to states \( \vec{K} \) and \( b > 0 \) is zero, yielding \( \pi_{\mu^*}(\vec{K}) = 0 \), \( \pi_{\mu^*}(1) = 0 \) and, by induction, \( \pi_{\mu^*}(b) = 0, \forall b \geq 0 \).

Therefore, SUtx never accesses states \( \vec{K} \) and \( b > 0 \); in other words, it remains silent until the PU packet \( l_P \) is decoded at SUrx (state \( \vec{K} \)), as in the genie-aided case.

Policy (19) is randomized only in state \( \vec{K} \). By the property of MDPs [26], the same performance is achieved by a policy that selects probabilistically (or time-shares between) one of the following two modes of operation at the beginning of each ARQ window (in the recurrent state 0): **Idle:** The SU remains idle over the entire ARQ window; **IC** (interference cancellation): The SU transmits (with probability one) only after the current PU packet is decoded at SUrx. With Idle mode, the SU does not interfere at all with the PU; with IC mode, it leverages knowledge of the PU packet to perform interference cancellation, in the event that the SU packet is decoded at SUrx. With Idle mode,

\[
\nabla_{th} \rightarrow 0,
\]  

the SU selects Idle mode with probability \( \xi_1 = 1 \). When \( \nabla_{th} = \nabla_{GA} \), the SU selects IC mode with probability \( \xi_2 = 1 \). When \( 0 < \nabla_{th} < \nabla_{GA} \), the probabilities \( \xi_1 \) and \( \xi_2 = 1 - \xi_1 \) are chosen so as to attain the PU throughput degradation constraint with equality.

When \( \nabla_{th} > \nabla_{GA} \), the SU access probability in state \( \vec{K} \) can no longer be increased; therefore, higher SU throughput can only be achieved by transmitting in state 0 as well. The optimal policy for this case is determined in the following theorem.
Theorem 2. If $\nabla_{GA} < \nabla_{th} < \nabla_{max}$, then

\[
\begin{align*}
\mu^*(0) &= \frac{\nabla_{th} - \nabla_{GA}}{\nabla_{max} - \nabla_{GA} + (\nabla_{max} - \nabla_{th}) (\frac{\rho_1 - \rho_0}{1 - \rho_1 + \rho_0})}, \\
\mu^*(\rightarrow K) &= \mu^*(\leftrightarrow K) = \mu^*(b) = 1, \forall b > 0;
\end{align*}
\]

under such policy,

\[
\begin{align*}
\bar{T}_S(\mu^*) &= \bar{T}_S^{(GA)} \left( \frac{\nabla_{th}}{\nabla_{max}} \right) \\
&\quad + \frac{(1 - \rho_1 + \rho_0 D_p)(1 - \rho_1)}{1 - \rho_1(1 - D_p)} \nabla_{th} - \nabla_{GA} \nabla_{max} \zeta_R, \\
\nabla(\mu^*) &= \nabla_{th},
\end{align*}
\]

where we have defined

\[
\zeta \triangleq \frac{v_{sp}}{1 - \rho_1(1 - D_p + v_{sp})} + \frac{v_s}{1 - \rho_1(1 - D_p)}.
\]

Finally, if $\nabla_{th} \geq \nabla_{max}$, the "always transmit" policy $\mu^*(s) = 1, \forall s \in S$ is optimal, and

\[
\begin{align*}
\bar{T}_S(\mu^*) &= \bar{T}_S^{(GA)} (1) - \frac{(1 - \rho_1)^2}{1 - \rho_1(1 - D_p)} \zeta R_s, \\
\nabla(\mu^*) &= \nabla_{max}.
\end{align*}
\]

If $\nabla_{GA} < \nabla_{th} < \nabla_{max}$, SUTx transmits with non-zero probability in state 0 until it reaches one of the states 1, $\rightarrow K$ or $\leftrightarrow K$, see (21). From this point on, it transmits with probability one until the end of the ARQ window. This policy is randomized only in state 0. By the property of MDPs [26], the same performance is achieved by a policy that selects probabilistically (or time-shares between) one of the following two modes of operation at the beginning of each ARQ window: IC mode as before; Always-TX: The SU always transmits over the entire ARQ window. With Always-TX mode, the SU maximizes the number of SU packets transmitted over the ARQ window and builds up the CD graph; these packets may become decodable via CD, hence this strategy maximizes the aggregate throughput accrued via CD. In the limit $\nabla_{th} \rightarrow \nabla_{GA}$, $\mu^*(0) \rightarrow 0$ and the SU selects IC mode with probability $\xi_2 = 1$. In the limit $\nabla_{th} \rightarrow \nabla_{max}$, $\mu^*(0) \rightarrow 1$ and the SU selects Always-TX mode with probability $\xi_3 = 1$. When $\nabla_{GA} < \nabla_{th} < \nabla_{max}$, $\xi_2$ and $\xi_3 = 1 - \xi_2$ are chosen so as to attain the PU throughput degradation constraint with equality. Finally, if $\nabla_{th} \geq \nabla_{max}$, the constraint becomes inactive and the SU selects Always-TX mode deterministically so as to
maximize the benefits of CD.

Overall, the optimal policy $\mu^*$ reflects a randomization among Idle, IC and Always-TX modes, with probabilities $\xi_1$, $\xi_2$ and $\xi_3 = 1 - \xi_1 - \xi_2$, respectively. If $\nabla_{th} \leq \nabla_{GA}$, $\xi_3 = 0$ so that only Idle and IC modes are used; if $\nabla_{GA} < \nabla_{th} < \nabla_{max}$, $\xi_1 = 0$ so that only IC and Always-TX modes are used; finally, if $\nabla_{th} \geq \nabla_{max}$, $\xi_3 = 1$ so that only Always-TX mode is used.

C. Proof of Theorem 2

We use a geometric approach based on the properties of constrained MDPs [27], [28] to determine, in closed form, the optimal policy and its performance when $\nabla_{th} > \nabla_{GA}$. We make the following definition.

**Definition 3** (Deterministic/randomized policy). A policy $\mu \in \mathcal{U}$ is deterministic if $\mu(s) \in \{0, 1\}, \forall s \in S$; otherwise, $\mu$ is randomized. We let $\mathcal{D} \subset \mathcal{U}$ be the set of deterministic policies.

In other words, $\mu \in \mathcal{D}$ takes a deterministic action in each state; however, the state sequence is random and governed by the transition probabilities under $\mu$, see (15)-(17). Each deterministic policy $\mu \in \mathcal{D}$ attains a black or white circle in Fig. 5 located at coordinates $(\nabla(\mu), \bar{T}_S(\mu))$. The set of randomized policies, instead, attain the convex hull of the points with coordinates given by all the deterministic policies, denoted as $\mathcal{C}(\mathcal{D})$, see Fig. 5. Thus, any point in $\mathcal{C}(\mathcal{D})$ with coordinates $(\nabla(\mu), \bar{T}_S(\mu))$, can be achieved by a (possibly) randomized policy. Any point outside of this region is not attainable.

According to OP, the goal is, for a given PU throughput degradation constraint $\nabla_{th}$, to determine the optimal point $(\nabla(\mu^*), \bar{T}_S(\mu^*))$ within the convex hull $\mathcal{C}(\mathcal{D})$, and the corresponding optimal policy $\mu^*$ maximizing the SU throughput $\bar{T}_S(\mu)$, see Fig. 5. We define the Pareto optimal envelope of the convex hull $\mathcal{C}(\mathcal{D})$, denoted as $\text{PO}(\mathcal{D})$, as the set of points such as no improvement in the SU throughput $\bar{T}_S(\mu)$ is possible without causing additional degradation $\nabla(\mu)$ to the PU. This is indicated by the sequence of solid lines connecting the black circles in the figure. Mathematically,

$$\text{PO}(\mathcal{D}) = \{ (\nabla(\mu), \bar{T}_S(\mu)) : \mu \in \mathcal{U}, \forall \tilde{\mu} \in \mathcal{U}$$

$$\text{s.t. } \tilde{T}_S(\tilde{\mu}) > \bar{T}_S(\mu) \Rightarrow \nabla(\tilde{\mu}) > \nabla(\mu) \}.$$  \hspace{1cm} (27)

\footnote{In fact, a randomized policy can be expressed equivalently as a time-sharing among deterministic policies.}
Accordingly, we define the set of Pareto optimal policies as

\[ U_{PO} \triangleq \{ \mu \in U : (\nabla(\mu), \bar{T}_S(\mu)) \in PO(D) \}. \]  

(28)

Note that any non Pareto optimal policy \( \mu \notin U_{PO} \) is suboptimal since there exists \( \tilde{\mu} \in U \) such that \( \bar{T}_S(\tilde{\mu}) \geq \bar{T}_S(\mu) \) and \( \nabla(\tilde{\mu}) < \nabla(\mu) \). Thus, we have the following result.

**Lemma 1.** The optimal policy is such that \( \mu^* \in U_{PO} \).

Hence, we can limit the search of the optimal policy within the set \( U_{PO} \), which we aim to characterize. Note that \( PO(D) \) is defined by a sequence of segments, each with endpoints defined by a pair of deterministic policies \( \mu^i \) and \( \mu^{i+1} \), \( i \geq 1 \). Without loss of generality, the sequence \( \{ \mu^i, i \geq 1 \} \) is characterized by strictly decreasing values of the interference, \( \nabla(\mu^i) > \nabla(\mu^{i+1}), \forall i \geq 1 \), and \( \mu^{[1]} \) is the deterministic policy which maximizes the SU throughput (unconstrained), \( \mu^{[1]} = \arg \max_{\mu \in D} \bar{T}_S(\mu) \), derived in Lemma 3. Such sequence exhibits the following property.
Lemma 2. \( \{\mu[i], i \geq 1\} \) defines strictly decreasing values of \( \nabla_{th}(\mu) \) and \( \bar{T}_S(\mu) \), i.e.,

\[
\nabla_{th}(\mu[i]) > \nabla_{th}(\mu[i+1]), \quad \bar{T}_S(\mu[i]) > \bar{T}_S(\mu[i+1]).
\]

(29)

Proof. If the above condition is not satisfied, i.e., \( \nabla(\mu[i]) > \nabla(\mu[i+1]) \) and \( \bar{T}_S(\mu[i]) \leq \bar{T}_S(\mu[i+1]) \), then we achieve a contradiction on the Pareto optimality of \( \mu[i] \).

It follows that, given \( \nabla_{th} > 0 \), one can determine \( \mu^* \) and its performance as follows:

- If \( \nabla_{th} \geq \nabla(\mu[1]) \), then \( \mu^* = \mu[1] \); in fact, \( \mu[1] \) achieves the maximum unconstrained throughput and is feasible for the given value of \( \nabla_{th} \);

- Otherwise, let \( i^* \geq 1 \) be the unique index such that \( \nabla(\mu[i^*]) \geq \nabla_{th} > \nabla(\mu[i^*+1]) \); then, the optimal policy is given by a proper randomization (or time-sharing) between \( \mu[i^*] \) and \( \mu[i^*+1] \); we will characterize the form of this randomization throughout the proof.

To characterize \( \PO(D) \), we are left with the problem of finding the sequence \( \{\mu[i], i \geq 1\} \subseteq D \). To this end, we let \( D[i] \subseteq D \) be the set of deterministic policies that interfere strictly less than \( \nabla(\mu[i]) \). Mathematically,

\[
D[i] = \{\mu \in D : \nabla(\mu) < \nabla(\mu[i])\}.
\]

(30)

Then, by construction, \( \mu[i+1] \) is the deterministic policy which minimizes the slope of the segment connecting \( (\nabla(\mu[i]), \bar{T}_S(\mu[i])) \) to \( (\nabla(\mu), \bar{T}_S(\mu)) \) over \( \mu \in D[i] \), i.e.,

\[
\mu[i+1] = \arg \min_{\mu \in \mathcal{D}[i]} \frac{\bar{T}_S(\mu) - \bar{T}_S(\mu[i])}{\nabla(\mu) - \nabla(\mu[i])}, \quad \forall i \geq 1.
\]

(31)

In other words, \( \mu[i+1] \) is the deterministic policy that yields the minimum decrease in SU throughput, relative to the decrease in PU throughput degradation.

Using this algorithm, we now determine \( \mu[1] \) and \( \mu[2] \). Lemma 3 states that \( \mu[1] \) is the Always-TX mode discussed in Sec. IV-B and that it uniquely maximizes the interference \( \nabla(\mu) \). That the Always-TX policy maximizes \( \bar{T}_S(\mu) \) is an intuitive, but non trivial result; indeed on a setting without CD, it was proved that Always-TX is not the throughput maximizing policy, see [29]. Then, Lemma 4 states that \( \mu[2] \) is the IC policy discussed in Sec. IV-B. It follows that, when \( \nabla(\mu[1]) \geq \nabla_{th} > \nabla(\mu[2]) \), the optimal policy is obtained by time-sharing between the Always-TX policy \( \mu[1] \) and the IC policy \( \mu[2] \); alternatively, since the Always-TX and IC policies only differ in state 0, the same result is obtained by randomizing in state 0, yielding (21).
Lemma 3. \( \mu^{[1]} \) is uniquely given by the Always-TX policy

\[
\mu^{[1]}(s) = 1, \quad \forall s \in \mathcal{S}.
\]  

(32)

Moreover, \( \mathcal{D}^{[1]} \equiv \mathcal{D} \setminus \{\mu^{[1]}\} \).

Proof. See Appendix B. \( \square \)

Given \( \mu^{[1]} \) we now determine \( \mu^{[2]} \) as the solution of the optimization problem (31). However, there is no need to minimize over the entire set \( \mathcal{D}^{[1]} \equiv \mathcal{D} \setminus \{\mu^{[1]}\} \). In fact, since \( \text{OP} \) has one constraint, the optimal policy is randomized in at most one state [28]. In particular, any point in the segment connecting \( (\nabla (\mu^{[1]}), \bar{T}_S(\mu^{[1]})) \) to \( (\nabla (\mu^{[2]}), \bar{T}_S(\mu^{[2]})) \) is achievable by a policy that is randomized in at most one state, hence \( \mu^{[1]} \) and \( \mu^{[2]} \) differ in only one state. Letting \( s^{[1]} \) be such state, and

\[
\Delta_s(s) = \chi(s = \hat{s}) \quad \forall s \in \mathcal{S},
\]  

(33)

where \( \chi(\cdot) \) is the indicator function, we can express \( \mu^{[2]} \) as

\[
\mu^{[2]} = \mu^{[1]} - \Delta_s(s),
\]  

(34)

so that \( \mu^{[2]}(s) = \mu^{[1]}(s) = 1, \forall s \neq s^{[1]} \) and \( \mu^{[2]}(s^{[1]}) = 0 \), hence \( \mu^{[2]} \) differs from \( \mu^{[1]} \) only in state \( s^{[1]} \). By leveraging these structural properties into (31), we conclude that

\[
s^{[1]} = \arg \min_{s \in \mathcal{S}} \eta(s),
\]  

(35)

where we have defined the SU access efficiency in state \( s \) as

\[
\eta(s) \triangleq \frac{\bar{T}_S(\mu^{[1]} - \Delta_s) - \bar{T}_S(\mu^{[1]})}{\nabla (\mu^{[1]} - \Delta_s) - \nabla (\mu^{[1]})}.
\]  

(36)

In other words, \( \eta(s) \) amounts to the decrease in SU throughput \( (\bar{T}_S(\mu^{[1]} - \Delta_s) - \bar{T}_S(\mu^{[1]})) \), per unit decrease in PU throughput degradation \( (\nabla (\mu^{[1]} - \Delta_s) - \nabla (\mu^{[1]})) \), as a result of remaining idle in state \( s \). Since the SU aims at maximizing its own throughput, under a PU throughput degradation constraint, \( s^{[1]} \) is chosen as the state \( s \) in (36) that minimizes the loss in SU throughput, per unit decrease of the PU throughput degradation, as captured in (35). By solving (35), we obtain the following result.
Lemma 4. $\mu^{[2]}$ is uniquely given by the IC policy

$$\mu^{[2]}(0) = 0, \quad \mu^{[2]}(s) = 1, \quad \forall s \in S \setminus \{0\}. \quad (37)$$

Proof. In this proof, we evaluate $\eta(s)$ in all states $s \in S$, and show that it is minimized by $s = 0$. We will make use of Appendix A to compute the performance of $\mu^{[1]}$ and $\mu^{[1]} - \Delta_s, \forall s \in S$ in closed form, used to compute $\eta(s)$ in (36). We obtain

$$\eta(m) = \frac{R_s}{\nabla_{\text{max}}} \frac{1 - \rho_1}{1 - \rho_1} \left( 1 - \rho_0 D_p \right) \nabla_{\text{max}} \left[ \left( 1 - \rho_1 \right) \left( 1 - \rho_0 D_p \right) \left( 1 - \rho_1 \right) \nabla_{\text{max}} \right],$$

$$\eta(\vec{K}) = \eta(0) + \frac{R_s}{\nabla_{\text{max}}} \frac{1 - \rho_1}{1 - \rho_1} \left( 1 - \rho_0 D_p \right) \nabla_{\text{max}} \left[ \left( 1 - \rho_1 \right) \left( 1 - \rho_0 D_p \right) \left( 1 - \rho_1 \right) \nabla_{\text{max}} \right],$$

$$\eta(\vec{K}) = \eta(0) + \frac{R_s}{\nabla_{\text{max}}} \frac{1 - \rho_1}{1 - \rho_1} \left( 1 - \rho_0 D_p \right) \nabla_{\text{max}} \left[ \left( 1 - \rho_1 \right) \left( 1 - \rho_0 D_p \right) \left( 1 - \rho_1 \right) \nabla_{\text{max}} \right],$$

where $\eta(0)$ is given by (38) with $m=0$, and $\zeta$ is given by (24).

To conclude, by comparing the SU access efficiencies, it is clear that $\eta(m) > \eta(0), \forall m > 0, \eta(\vec{K}) > \eta(0)$ and $\eta(\vec{K}) > \eta(0)$. Thus, the solution of (31) yields $s^{[1]} = 0$ and $\mu^{[2]} \equiv \mu^{[1]} - \Delta_0$, proving the lemma.

Under policy $\mu = \mu^{[2]}$, we have that $\mu(0) = 0$ and $\mu(\vec{K}) = 1$ hence, following the discussion in Sec. IV-B, the SU accesses only states 0 (where it remains idle) and $\vec{K}$ (where it transmits). Therefore, it transmits with probability one in state $\vec{K}$ only, i.e., after the PU packet becomes known at SUrx; when this happens, the SU packet is decoded via interference cancellation. For this reason, $\mu^{[2]}$ is termed "IC policy". As discussed in Sec. IV-B under such policy we obtain $\nabla(\mu^{[2]}) = \nabla_{\text{GA}}$ as in (18). Thus, if $\nabla_{th} \in [\nabla(\mu^{[2]}), \nabla_{\text{GA}}, \nabla(\mu^{[1]}), \nabla_{\text{max}}]$, the optimal policy is obtained by randomizing between the IC policy $\mu^{[2]}$ and the Always-TX policy $\mu^{[1]}$, or equivalently, by the policy (21) randomized in state 0. The optimal policy given in (21) and its performance in (22) are obtained by enforcing $\nabla(\mu^*) = \nabla_{th}$ to determine the optimal value of $\mu^*(0)$. On the other hand, if $\nabla_{th} \geq \nabla_{\text{max}}$, then the optimal policy is Always-TX ($\mu^{[1]}$), which maximizes the SU throughput and satisfies the constraint $\nabla(\mu^{[1]}), \nabla_{\text{max}} \leq \nabla_{th}$. Thus, we have proved the structure and performance of the optimal policy for the case $\nabla_{th} \geq \nabla_{\text{GA}}$.
V. Online Learning and Adaptation

The Always-TX, IC and Idle modes do not require any knowledge of the statistics of the model, such as the decoding probabilities \( p_0 \) or the PU outage probabilities \( \rho_0 \) and \( \rho_1 \). Thus, the SU needs only to learn the optimal randomization among these three modes of operation. This can be inferred from the throughput degradation experienced by the PU, estimated by monitoring the ACK/NACK feedback: when this estimate is below the PU throughput degradation constraint, the SU may transmit more often by favoring the Always-TX or IC mode (depending on the time-sharing currently in use); when this estimate is above the constraint, the SU may reduce its transmissions by favoring the IC or Idle mode. This feature of the optimal policy facilitates learning and adaptation in practical settings where the statistics of the system are unknown, or vary over time.

In this section, we propose an algorithm based on stochastic gradient descent (SGD) [21, Chapter 14] for the online optimization of the SU access policy and transmit rate \( R_s \), by leveraging the structure of the optimal access policy. Note that the optimal SU access policy \( \mu^* \) of Theorems 1 and 2 is uniquely characterized by a parameter \( \nu \equiv \mu^*(0) + \mu^*(\rightarrow) \in [0, 2] \), related to the access level of the SU: given \( \nu \), we get \( \mu^*(0) = \max\{\nu - 1, 0\} \) and \( \mu^*(\rightarrow) = \min\{\nu, 1\} \), and the degradation to the PU, \( \nabla_{th} \), is related to \( \nu \) via (19) for \( 0 \leq \nu \leq 1 \) and (21) for \( 1 < \nu \leq 2 \). Let \( \bar{T}_S(\nu) \) and \( \bar{T}_P(\nu) \) be the corresponding SU and PU throughputs, respectively, which are increasing and decreasing functions of \( \nu \), respectively. Let \( \bar{T}_{P,\text{min}} \) be the minimum throughput requirement for the PU. This information may be broadcast by the PU system to regulate the access of SUs. The PU throughput degradation constraint \( \nabla_{th} \) in (8) is related to \( \bar{T}_{P,\text{min}} \) via \( \nabla_{th} = 1 - \bar{T}_{P,\text{min}}/\bar{T}_{P,\text{max}} \). The rate \( R_s \) is chosen so as to maximize the SU throughput under no interference from the PU signal, i.e., \( R_s = \arg\max_{r_s} r_s \Pr(r_s < C(\gamma_s)) \). Under Rayleigh fading, we obtain \( \Pr(r_s < C(\gamma_s)) = \exp\left\{ -\frac{2r_s - 1}{\bar{\gamma}_s} \right\} \). Thus, the optimal \( \nu^* \in [0, 2] \) (or equivalently, the optimal policy \( \mu^* \)) and \( R_s^* \geq 0 \) can be expressed as the minimizers of

\[
\min_{\nu, r_s} \frac{1}{2} \left( \bar{T}_P(\nu) - \bar{T}_{P,\text{min}} \right)^2 - r_s \exp\left\{ -\frac{2r_s - 1}{\bar{\gamma}_s} \right\}.
\]

We denote the objective function as \( G(\nu, r_s) \). Consider the optimization with respect to \( \nu \). Since \( \bar{T}_P(\nu) \) is a decreasing function of \( \nu \), if \( \bar{T}_{P,\text{min}} > \bar{T}_{P,\text{max}} \) (hence, \( \bar{T}_P(\nu) < \bar{T}_{P,\text{min}} \)), then the solution is \( \nu = 0 \) (the SU remains idle, and the optimization of \( R_s \) is irrelevant); indeed, in this case, the PU has set an unrealistic demand, hence the SU should remain idle to at least partially
satisfy it. If \( R_p(1-\rho_1) \leq \bar{T}_{P,\min} \leq \bar{T}_{P,\max} \), where \( R_p(1-\rho_1) \) is the PU throughput achieved when the SU always transmits, then the solution is the unique \( \nu^* \) such that \( \bar{T}_P(\nu^*) = \bar{T}_{P,\min} \), i.e., the PU throughput constraint is attained with equality. Finally, if \( \bar{T}_{P,\min} < R_p(1-\rho_1) \), then \( \bar{T}_{P,\min} < \bar{T}_P(\nu), \forall \nu \), hence the solution is \( \nu = 2 \) (the SU always transmits); indeed, in this case, the PU demand can be met even if the SU always transmits.

Problem (41) can be solved using the gradient projection algorithm \([30, \text{Chapter 3}]\). The gradient of the objective function \( G(\nu, r_s) \) with respect to \( \nu \) and \( r_s \) is given by

\[
\begin{align*}
\frac{dG(\nu, r_s)}{d\nu} &= \frac{d\bar{T}_P(\nu)}{d\nu} (\bar{T}_P(\nu) - \bar{T}_{P,\min}) \\
&\propto \bar{T}_{P,\min} - \bar{T}_P(\nu) \triangleq g_1(\nu), \\
\frac{dG(\nu, r_s)}{dr_s} &= \exp \left\{ -2r_s - 1 \right\} \left[ \frac{1}{\bar{\gamma}_s} \ln(2)r_s 2^{r_s} - 1 \right] \\
&\propto \mathbb{E}[a_{s,t}] \ln(2)r_s 2^{r_s} - \bar{\gamma}_s \triangleq g_2(r_s),
\end{align*}
\]

where \( \propto \) denotes proportionality up to a positive multiplicative factor, since \( d\bar{T}_P(\nu)/d\nu < 0 \). Thus, (41) can be solved as

\[
\begin{align*}
\nu_{t+1} &= \left[ \nu_t - \beta_t g_1(\nu_t) \right]_0^2, \\
R_{s,t+1} &= \left[ R_{s,t} - \beta_t g_2(R_{s,t}) \right]^+_0,
\end{align*}
\]

where \( \beta_t > 0 \) is the step-size, \([\cdot]_0^2 = \min\{\max\{\cdot, 0\}, 2\} \) and \([\cdot]^+_0 = \max\{\cdot, 0\} \) are projection operations onto the feasible sets. The policy used at time \( t \) is then given by

\[
\begin{align*}
\mu_t(0) &= \max\{\nu_t - 1, 0\}, \\
\mu_t(\overline{K}) &= \min\{\nu_t, 1\}, \\
\mu_t(\overline{K}) &= \mu_t(b) = 1, \quad \forall b > 0.
\end{align*}
\]

However, typically \( T_P(\nu) \) may not be available to the SU to compute the gradient \( g_1(\nu) \), but only observations of the ACK/NACK feedback sequence \( \{y_{P,t}, t \geq 0\} \); similarly, only realizations of the channel fading \( \gamma_{s,t} \) may be available via channel estimation, instead of the expected channel gain \( \bar{\gamma}_s = \mathbb{E}[\gamma_{s,t}] \) required to compute \( g_2(r_s) \). Thus, we use the SGD algorithm, which replaces \( g_1(\nu_t) \) and \( g_2(r_{s,t}) \) with estimates \( \hat{g}_{1,t} \) and \( \hat{g}_{2,t} \) such that \( \mathbb{E}[\hat{g}_{1,t} | \nu_t] = g_1(\nu_t) \) and \( \mathbb{E}[\hat{g}_{2,t} | \nu_t] = g_2(r_{s,t}) \).
In particular, we choose
\[ \hat{g}_{1,t} = \bar{T}_{P,\text{min}} - R_p \chi(y_{P,t} = \text{ACK}), \]
\[ \hat{g}_{2,t} = a_{S,t} \left[ \ln(2) R_{s,t} 2^{R_{s,t}} - \gamma_{s,t} \right]. \]

We finally obtain
\[ \nu_{t+1} = \left[ \nu_t + \beta_t (R_p \chi(y_{P,t} = \text{ACK}) - \bar{T}_{P,\text{min}}) \right]^2, \]
\[ R_{s,t+1} = \left[ R_{s,t} + \beta_t a_{S,t} (\gamma_{s,t} - \ln(2) R_{s,t} 2^{R_{s,t}}) \right]^+, \]

where \( \nu_0 = 0 \) and \( R_{s,0} = 0 \) (the SU is idle at initialization). Thus, \( \nu \) tends to augment if an ACK is received, so that the SU may transmit more often, and to diminish otherwise; \( R_s \) tends to augment if the channel is good \( (\gamma_{s,t} > \ln(2) R_{s,t} 2^{R_{s,t}}) \), and diminish otherwise. In static scenarios where the parameters of the model do not change, a decreasing step-size is commonly used in stochastic optimization, such as \( \beta_t = \beta_0 / (t + 1) \); in time-varying scenarios, a fixed but small step-size may be used, in order to accommodate adaptation. Note that, if \( a_{S,t} = 0 \), the SU remains idle and the channel fading may not be estimated, yielding \( R_{s,t+1} = R_{s,t} \) as in (50).

**VI. Numerical Results**

In this section, we present numerical results. PUtx is located at position \((0,0)\), PUrx at \((0,d_0)\), at reference distance \(d_0\) from PUtx. SUtx and SUrx are located at positions \((d_{SP},0)\) and \((d_{SP},d_0)\), respectively, where \(d_{SP}\) is the distance between the SU and PU pairs. We assume Rayleigh fading channels. The expected SNR of the link PUtx-PUrx is \( \bar{\gamma}_p = 20 \). For any other link, the expected SNR is given by \( \bar{\gamma}_TR = \bar{\gamma}_p (d_{TR}/d_0)^{-\alpha} \), where \(d_{TR}\) is the distance between the corresponding transmitter (T) and receiver (R), and \( \alpha=2 \) is the pathloss exponent. \( R_p \) and \( R_s \) are chosen so as to maximize the respective PU and SU throughputs under no interference, *i.e.*, \( R_x = \arg \max_{r_x} r_x \mathbb{P}(r_x < C(\gamma_x)), x \in \{p,s\} \). The outage probabilities for the PU are computed as \( \rho_0 = \mathbb{P}(R_s < \gamma_p) \) and \( \rho_1 = \mathbb{P}(R_s < \gamma_p/(1 + \gamma_{sp})) \).

We consider the following schemes: 1) The optimal CD policy (OPCD) given by Theorems 1 and 2. 2) The BIC scheme developed in [10], where, unlike CD, the SU does not perform retransmissions; hence, after decoding the PU packet, it uses interference cancellation only within the current ARQ window only (see example in Fig. 1). 3) The non-adaptive CD scheme (NACD), where the SU adopts the optimal packet selection policy without any access policy.
i.e., it transmits with constant probability $\min\{\nabla_{th}/\nabla_{\text{max}}, 1\}$ in all slots, independently of the state. 4) The "ARQ-oblivious" (AO) scheme, originally proposed in [9], where the SU attempts to jointly decode the SU and PU packets and remove the interference of the latter, by leveraging the PU codebook structure [31]; however, it does not exploit the redundancy of the ARQ mechanism to perform interference cancellation over the ARQ window; its performance is given by $\bar{T}_{\text{AO}} = \min\{\nabla_{th}/\nabla_{\text{max}}, 1\} R_s (\delta_s + \delta_{sp})$. Additionally, we plot the genie-aided throughput $\bar{T}_{\text{S}}^{(GA)} \left(\min\{\nabla_{th}/\nabla_{\text{max}}, 1\}\right)$, which assumes a-priori knowledge of the PU packet.

In Fig. 6, we plot the SU throughput as a function of $d_{SP}/d_0$. The interference constraint to the PU is set to $\nabla_{th} = 0.1$, so that the SU is allowed to degrade the PU throughput by at most 10%. The solid lines refer to the analytical expressions, whereas the markers refer to the SGD algorithm developed in Sec. V evaluated via Monte Carlo simulation over $10^5$ slots. The SGD algorithm is initialized with $\nu_0 = 0$ and $R_{s,0} = 0$, so that the SU is initially idle. This is a conservative behavior, which minimizes the risk of generating harmful interference to the PU in the initial phase when the SU is uninformed, and allows the latter to collect observations before undertaking a more active behavior. We note that SGD closely approaches the analytical curves.

Note that AO lower bounds the performance of CD and BIC, since it does not leverage the interference of the ARQ protocol. In contrast, by assuming non-causal knowledge of the PU packet, "genie-aided" upper bounds the performance. Both AO and "genie-aided" exhibit
a monotonically increasing trend as a function of $d_{SP}$: as the SU and PU pairs move farther away from each other, the interference decreases, hence the SU can afford to transmit more frequently. Similarly, the SU throughputs of CD and BIC monotonically increase and attain the genie-aided throughput for $d_{SP} \lesssim 2d_0$. This is a consequence of the fact that, in this regime, SUtx is close to PUrx, hence its transmissions interfere strongly with PUrx; in this case, the SU may only transmit sparingly by randomizing between Idle and IC modes (policy (19)), which attains the genie-aided upper bound, see (20). As $d_{SP}$ increases beyond $2d_0$, the interference to the PU becomes weaker, the SU may transmit more frequently, hence it starts transmitting in state 0 as well (policy (21)), creating a gap with respect to the genie-aided throughput (case $\nabla_{GA} < \nabla_{th} < \nabla_{max}$ in Theorem 2).

However, surprisingly, CD and BIC do not follow the same monotonic trend as AO and "genie-aided" for $d_{SP} \gtrsim 2d_0$: the performance of CD decreases for $2d_0 \lesssim d_{SP} \lesssim 2.7d_0$, and approaches that of AO for $d_{SP} \rightarrow 3d_0$ (a similar consideration holds for BIC in the range $2d_0 \lesssim d_{SP} \lesssim 2.5d_0$). This counterintuitive result can be explained as follows:

1) As SUrx moves farther away from PUtx, SUrx receives a weaker PU signal, hence it becomes more difficult to decode the PU packet and remove its interference; in turn, it becomes more difficult to decode the CD root and initiate CD, due to the interference from
the PU;

2) As SUtx moves farther away from PUrx, the interference generated by SUtx to PUrx becomes weaker, hence the outage probability $\rho_1$ tends to decrease; therefore, when using the Always-TX mode, the ARQ window tends to shorten (its average duration is $1/(1-\rho_1)$) in the Always-TX mode) since the PU is more likely to succeed, resulting in fewer opportunities to leverage the redundancy of the ARQ protocol.

Overall, we note that OPCD outperforms BIC by up to $\sim 15\%$ and achieves up to $2\times$ throughput improvement over NACD and up to $3\times$ over AO. NACD performs poorly compared to both OPCD and BIC, revealing the importance of using an optimized SU access policy over a non-adaptive one which does not fully leverage the structure of the problem.

In Fig. 7, we plot the trade-off between the SU throughput $\bar{T}_S(\mu)$ and the PU throughput degradation $\nabla(\mu)$ as $\nabla_{th}$ is varied. For $\nabla_{th} \leq \nabla_{GA} \simeq 0.1$, OPCD randomizes between the Idle and IC modes; for $\nabla_{GA} \leq \nabla_{th} \leq \nabla_{max} \simeq 0.6$, it randomizes between the IC and Always-TX modes; for $\nabla_{th} \geq \nabla_{max} \simeq 0.6$ the interference constraint becomes inactive and Always-TX is the only mode of operation. In all cases, $\hat{T}_S(\mu)$ monotonically increases with $\nabla_{th}$, since more opportunities become available to the SU to use the channel. OPCD achieves up to $40\%-60\%$ improvement over BIC, for $\nabla_{th} \gtrapprox 0.3$, up to $2\times$ improvement over NACD for $\nabla_{th} \lesssim 0.1$, and
\[ \text{Value} \]

\[ \begin{align*}
\text{Slot index} & \quad 0 & 1 & 2 & 3 & 4 & 5 & 6 \\
\text{Idle & IC modes} & \quad \text{IC & Always-TX modes} & \quad \text{IC & Always-TX modes} & \quad \text{IC & Always-TX modes} & \quad \text{IC & Always-TX modes} & \quad \text{IC & Always-TX modes} & \quad \text{IC & Always-TX modes} & \quad \text{IC & Always-TX modes} \\
\text{Always-TX mode} & \quad \text{Always-TX mode} & \quad \text{Always-TX mode} & \quad \text{Always-TX mode} & \quad \text{Always-TX mode} & \quad \text{Always-TX mode} & \quad \text{Always-TX mode} & \quad \text{Always-TX mode} \\
\end{align*} \]

Fig. 9: Convergence of the SGD algorithm; \( \nabla_{th} = 0.1 \); the distance varies linearly from \( d_{SP} = 10 \) at time 0 to \( d_{SP} = 0.5 \) at time \( t = 5000 \), and then linearly to \( d_{SP} = 10 \) at time \( t = 10^4 \).

\( \geq 70\% \) improvement over AO, for all range of values.

In the analysis, we assumed that the PU packet is transmitted until successfully decoded at its intended receiver and an infinite buffer to store the received signals at SUrx. In practical systems, these are finite quantities. In Fig. 8 we evaluate the effect of a finite ARQ deadline \( T_{ARQ} \) and a finite buffer size \( B_{max} \). Each unit corresponds to the buffer space required to store one received signal. When the buffer at SUrx is full, the received signal is discarded after the decoding attempt, resulting in missed opportunities to build up the CD graph. Note that, as \( T_{ARQ} \) increases, the performance improves. In fact, the longer the ARQ window, the more opportunities available at the SU to leverage the redundancy of the ARQ process of the PU. Surprisingly, most of the benefits of CD are reaped with a buffer size of only \( B_{max} \approx 4 \) units, with larger buffer sizes yielding only marginal improvements. Generally, CD achieves a non trivial improvement over BIC, demonstrating a better use of the buffer space available at SUrx.

In Fig. 9 we investigate the performance of the SGD algorithm developed in Sec. 5 in a time-varying scenario with constant step-size \( \beta_t \). The distance between SUtx and SUrx is kept fixed, whereas that between the SU and PU pair, \( d_{SP} \), varies as described in the caption. Accordingly, the optimal value of the SU rate \( R_s \) is constant, whereas the optimal value of the parameter \( \nu \) depends on \( d_{SP} \): when \( d_{SP} \) is large (around \( t \approx 0 \) and \( t \approx 10000 \), the SU and PU pair are far away from each other), \( \nu^* \) is large since the SU generates little interference to the
PU and thus may transmit more frequently; in contrast, when $d_{SP}$ is large (around $t \simeq 5000$, the SU and PU pair are far away from each other), $\nu^*$ is small since the SU generates strong interference to the PU, hence it may only transmit sparingly. We note that the SGD algorithm, after an initial convergence phase, closely tracks the optimal values of $R^*_s$ and $\nu^*$. Surprisingly, this is accomplished by only observing the ARQ feedback from PUrx and the channel fading realization, see (49).

VII. CONCLUSIONS

In this paper, we have investigated the design of optimal SU access policies that maximize the SU throughput via chain decoding, subject to an interference constraint to the PU. We have found a closed form expression of the optimal policy and of its performance, and shown that it can be expressed as a randomization among three modes of operation. We have designed an algorithm based on stochastic gradient descent to determine the optimal randomization in practical settings where the statistics of the system are unknown or vary over time. We have shown numerically that, for a 10% interference constraint, the optimal access policy with chain decoding outperforms by 15% a state-of-the-art scheme that does not exploit opportunistic secondary retransmissions, and achieves up to $2\times$ improvement over a chain decoding scheme using a non-adaptive access policy instead of the optimal one.

REFERENCES

[1] N. Michelusi, “Optimal secondary access in retransmission based primary networks via chain decoding,” in IEEE International Symposium on Information Theory (ISIT), June 2017, pp. 1187–1191.
[2] N. Michelusi, P. Popovski, and M. Zorzi, “Optimal cognitive access and packet selection under a primary arq process via chain decoding,” IEEE Transactions on Information Theory, vol. 62, no. 12, pp. 7324–7357, Dec 2016.
[3] “Realizing the Full Potential of Government-Held Spectrum to Spur Economic Growth,” Tech. Rep., July 2012, report to the president. [Online]. Available: http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast_spectrum_report_final_july_20_2012.pdf
[4] Q. Zhao and B. Sadler, “A Survey of Dynamic Spectrum Access,” IEEE Signal Processing Magazine, vol. 24, no. 3, pp. 79–89, May 2007.
[5] N. Michelusi and U. Mitra, “Cross-layer estimation and control for cognitive radio: Exploiting sparse network dynamics,” IEEE Transactions on Cognitive Communications and Networking, vol. 1, no. 1, pp. 128–145, March 2015.
[6] S. Geirhofer, L. Tong, and B. M. Sadler, “Dynamic spectrum access in the time domain: Modeling and exploiting white space,” IEEE Communications Magazine, vol. 45, no. 5, pp. 66–72, May 2007.
[7] J. Peha, “Sharing Spectrum Through Spectrum Policy Reform and Cognitive Radio,” Proceedings of the IEEE, vol. 97, no. 4, pp. 708–719, Apr. 2009.
[8] R. Comroe and D. Costello, “ARQ Schemes for Data Transmission in Mobile Radio Systems,” IEEE Journal on Selected Areas in Communications, vol. 2, no. 4, pp. 472–481, July 1984.

[9] M. Levorato, U. Mitra, and M. Zorzi, “Cognitive Interference Management in Retransmission-Based Wireless Networks,” IEEE Transactions on Information Theory, vol. 58, no. 5, pp. 3023–3046, May 2012.

[10] N. Michelusi, P. Popovski, O. Simeone, M. Levorato, and M. Zorzi, “Cognitive Access Policies under a Primary ARQ Process via Forward-Backward Interference Cancellation,” IEEE Journal on Selected Areas in Communications, vol. 31, no. 11, pp. 2374–2386, November 2013.

[11] M. Sharma, A. Sahoo, and K. D. Nayak, “Channel modeling based on interference temperature in underlay cognitive wireless networks,” in 2008 IEEE International Symposium on Wireless Communication Systems, Oct 2008, pp. 224–228.

[12] GSMA Intelligence, “Understanding 5G: Perspectives on future technological advancements in mobile – GSMAi report,” Tech. Rep., December 2014.

[13] A. Goldsmith, S. Jafar, I. Maric, and S. Srinivasa, “Breaking Spectrum Gridlock With Cognitive Radios: An Information Theoretic Perspective,” Proceedings of the IEEE, vol. 97, no. 5, pp. 894–914, May 2009.

[14] K. Eswaran, M. Gastpar, and K. Ramchandran, “Bits through arqs: Spectrum sharing with a primary packet system,” in 2007 IEEE International Symposium on Information Theory, June 2007, pp. 2171–2175.

[15] R. Tannious and A. Nosratinia, “Cognitive Radio Protocols Based on Exploiting Hybrid ARQ Retransmissions,” IEEE Transactions on Wireless Communications, vol. 9, no. 9, pp. 2833–2841, Sep. 2010.

[16] J. C. F. Li, W. Zhang, A. Nosratinia, and J. Yuan, “SHARP: Spectrum Harvesting with ARQ Retransmission and Probing in Cognitive Radio,” IEEE Transactions on Communications, vol. 61, no. 3, pp. 951–960, March 2013.

[17] A. Jovicic and P. Viswanath, “Cognitive Radio: An Information-Theoretic Perspective,” IEEE Transactions on Information Theory, vol. 55, no. 9, pp. 3945–3958, Sep. 2009.

[18] K. Kulkarni and A. Banerjee, “On Stable Throughput of Cognitive Radio Networks With Cooperating Secondary Users,” IEEE Transactions on Communications, vol. 64, no. 10, pp. 4097–4110, Oct 2016.

[19] R. Zhang, “On active learning and supervised transmission of spectrum sharing based cognitive radios by exploiting hidden primary radio feedback,” IEEE Transactions on Communications, vol. 58, no. 10, pp. 2960–2970, October 2010.

[20] D. Cabric, S. M. Mishra, and R. W. Brodersen, “Implementation issues in spectrum sensing for cognitive radios,” in Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, 2004., vol. 1, Nov 2004, pp. 772–776 Vol.1.

[21] S. Shalev-Shwartz and S. Ben-David, Understanding Machine Learning: From Theory to Algorithms. New York, NY, USA: Cambridge University Press, 2014.

[22] R. S. Sutton and A. G. Barto, Reinforcement Learning : An Introduction. MIT Press, 1998.

[23] X. Liu, D. Yang, and A. El Gamal, “Deep neural network architectures for modulation classification,” in 51st Asilomar Conference on Signals, Systems and Computers, Nov. 2017.

[24] S. Zhang, S. C. Liew, and H. Wang, “Blind known interference cancellation,” IEEE Journal on Selected Areas in Communications, vol. 31, no. 8, pp. 1572–1582, August 2013.

[25] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2nd ed. John Wiley & Sons, Inc., New York, 2006.

[26] D. Bertsekas, Dynamic Programming and Optimal Control. Athena Scientific, 1995.

[27] K. W. Ross, “Randomized and Past-Dependent Policies for Markov Decision Processes with Multiple Constraints,” Operations Research, vol. 37, no. 3, pp. 474–477, 1989.

[28] E. Altman, Constrained Markov Decision Processes. Chapman and Hall, 1999.

[29] N. Michelusi, O. Simeone, M. Levorato, P. Popovski, and M. Zorzi, “Optimal cognitive transmission exploiting redundancy in the primary arq process,” in 2011 Information Theory and Applications Workshop, Feb 2011, pp. 1–10.
APPENDIX A: PERFORMANCE EVALUATION

In the following, by averaging over the random action $a_S \sim \text{Bernoulli}(\mu(s))$ selected under policy $\mu$, we let

$$
\begin{align*}
    v_S(\mu(s), s) &= \mu(s)v_S(1, s) + (1 - \mu(s))v_S(0, s), \\
    P(x|s, \mu(s)) &= \mu(s)P(x|s, 1) + (1 - \mu(s))P(x|s, 0).
\end{align*}
$$
(51)

**Definition 4.** Let $Q_\mu(0) = 1$, and recursively, for $b > 0$,

$$
Q_\mu(b) = \frac{\rho_1v_s\mu(b - 1)}{1 - \mu(b)\rho_1(1 - D_p - v_s) - (1 - \mu(b))\rho_0(1 - D_p)}Q_\mu(b - 1).
$$
(52)

Let

$$
\Psi_\mu(x, y, z) \triangleq \sum_{b=0}^{\infty} Q_\mu(b)[x + \mu(b)y + bz]
$$
(53)

for constants $x, y, z$. Let

$$
W_\mu(\hat{K}) = \frac{\Psi_\mu(0, \rho_1v_sp, 0)}{1 - \rho_{\mu(\hat{K})}(1 - D_p) - \mu(\hat{K})\rho_1v_sp},
$$
(54)

$$
W_\mu(\vec{K}) = \frac{\Psi_\mu(\rho_0D_p(1 - \rho_0)D_p - \rho_1v_sp, 0)}{1 - \rho_{\mu(\vec{K})}} + \frac{[\rho_{\mu(\hat{K})}D_p - \mu(\hat{K})\rho_1v_sp]\Psi_\mu(0, \rho_1v_sp, 0)}{[1 - \rho_{\mu(\hat{K})}(1 - D_p) - \mu(\hat{K})\rho_1v_sp][1 - \rho_{\mu(\vec{K})}]}.
$$
(55)

**Theorem 3.** Under $\mu \in \mathcal{U}$, the steady-state distribution of $s \in \mathcal{S}$ is given by

$$
\pi_\mu(0) = \frac{1}{\Psi_\mu(1, 0, 0) + W_\mu(\hat{K}) + W_\mu(\vec{K})},
$$
(56)

$$
\pi_\mu(b) = Q_\mu(b)\pi_\mu(0), \quad \forall b > 0, \quad \pi_\mu(\hat{K}) = W_\mu(\hat{K})\pi_\mu(0), \quad \pi_\mu(\vec{K}) = W_\mu(\vec{K})\pi_\mu(0);
$$
(57)

the SU throughput and interference are given by

$$
\bar{T}_S(\mu) = \frac{\nabla(\mu)}{\nabla_{\max}} R_sD_s - \pi_\mu(0)R_s\Psi_\mu(0, v_sp + v_s, D_p) + \pi_\mu(0)R_sW_\mu(\hat{K})\left(D_p - \mu(\hat{K})v_sp\right),
$$
(58)

$$
\nabla(\mu) = \nabla_{\max}\pi_\mu(0) \left[ \Psi_\mu(0, 1, 0) + W_\mu(\hat{K})\mu(\hat{K}) + W_\mu(\vec{K})\mu(\vec{K}) \right].
$$
(59)
Proof. Let $\mu \in \mathcal{U}$. Using (15)-(17), we obtain the steady state equations in state $b > 0$ as

$$\pi_\mu(b) = \pi_\mu(b)P(b|b, \mu(b)) + \pi_\mu(b-1)P(b|b-1, \mu(b-1)).$$

Thus, we obtain

$$\pi_\mu(b) = \frac{P(b|b-1, \mu(b-1))}{1 - P(b|b, \mu(b))}\pi_\mu(b-1), \quad (60)$$

and, by using induction on $b$ and (13), we obtain (57), with $Q_\mu(b)$ given by (52).

In state $\mu \rightarrow K$, using (15)-(17) we obtain

$$\pi_\mu(\mu \rightarrow K) = \pi_\mu(K)P(K | \mu \rightarrow K, \mu(\mu)) + \sum_{b=0}^{\infty} \pi_\mu(b)P(K | b, \mu(b)). \quad (61)$$

Using (15)-(16), $\pi_\mu(b) = Q_\mu(b)\pi_\mu(0)$ and the definition of $\Psi_\mu(x,y,z)$ given by (53), we thus obtain $\pi_\mu(\mu \rightarrow K) = W_\mu(K)\pi_\mu(0)$ as in (57), with $W_\mu(K)$ given by (54).

Finally, in state $\mu \rightarrow K$ we have

$$\pi_\mu(\mu \rightarrow K) = \sum_{b=0}^{\infty} \pi_\mu(b)P(K | b, \mu(b)) + \pi_\mu(K)P(K | \mu \rightarrow K, \mu(\mu)) + \pi_\mu(\mu \rightarrow K)P(K | \mu \rightarrow K, \mu(\mu)). \quad (62)$$

Using (15)-(17), $\pi_\mu(b) = Q_\mu(b)\pi_\mu(0)$, $\pi_\mu(\mu \rightarrow K) = W_\mu(K)\pi_\mu(0)$ and the definition of $\Psi_\mu(x,y,z)$ given by (53), we thus obtain $\pi_\mu(\mu \rightarrow K) = W_\mu(K)\pi_\mu(0)$ as in (57), with $W_\mu(K)$ given by (54).

Finally, by normalizing the steady state distribution to unity we obtain (56) from

$$1 = \sum_{b=0}^{\infty} \pi_\mu(b) + \pi_\mu(\mu \rightarrow K) + \pi_\mu(\mu \rightarrow K) = \pi_\mu(0) \left[ \sum_{b=0}^{\infty} Q_\mu(b) + W_\mu(\mu \rightarrow K) + W_\mu(\mu \rightarrow K) \right]. \quad (63)$$

The SU throughput $\bar{T}_S(\mu)$ and interference $\nabla(\mu)$ are obtained by averaging the instantaneous virtual throughput $v_S(a_S, s)$ and interference $\frac{\mu - B_0}{1 - \rho_0} a_S$ over the steady state distribution and access probability and by using the definition of $\Psi_\mu(x,y,z)$ in (53).

\section*{APPENDIX B: PROOF OF LEMMA 3}

Proof. Let $\mu \in \mathcal{D}$ such that $\exists s \in S$ with $\mu(s) = 0$ (i.e., $\mu$ is not the Always-TX policy). Let $m \geq 0$ be the first index such that $\mu(b) = 1, \forall b < m$ and $\mu(m) = 0$ (possibly, $m = \infty$ if $\mu(b) = 1, \forall b \geq 0$). We show that by setting $m = \infty$, $\mu(\mu \rightarrow K) = \mu(\mu \rightarrow K) = 1$, corresponding to the Always-TX policy, the SU throughput and SU interference are simultaneously and uniquely
maximized. This implies that, for any other policy \( \mu \neq \mu[1], \nabla(\mu) < \nabla(\mu[1]), \) hence \( \mathcal{D}[1] \equiv \mathcal{D} \setminus \{\mu[1]\} \) follows by the definition of \( \mathcal{D}[i] \) in (50).

From (52), using the fact that \( \mu(b) = 1, \forall b < m \) and \( \mu(m) = 0 \), we obtain

\[
Q_\mu(b) = \phi_1^b \chi(0 \leq b < m) + \chi(b = m) \phi_1^{b-1} (\chi(m = 0) \phi_1 + \chi(m > 0) \phi_0),
\]

where we have defined

\[
\phi_i \triangleq \frac{\rho_1 v_s}{1 - \rho_i (1 - D_p - i v_s)}, \quad i \in \{0, 1\}.
\]

Therefore, using this expression into (53) and using the fact that \( Q_\mu(b) = 0, \forall b > m \), we obtain

\[
\Psi_\mu(x, y, z) = \sum_{b=0}^{m-1} Q_\mu(b)[x + y + bz] + Q_\mu(m)(x + mz) = \chi(m = 0)x
\]

\[
+ (x + y + mz) \frac{1 - \phi_1^m}{1 - \phi_1} - z \frac{m - (m + 1) \phi_1 + \phi_1^{m+1}}{(1 - \phi_1)^2} + \chi(m > 0)(x + mz) \phi_1^{m-1} \phi_0.
\]

Then, \( \bar{T}_s(\mu) \) and \( \nabla(\mu) \) can be found via (58) and (59) in Appendix A.

We first show that, if \( m < \infty \), \( \mu(\tilde{K}) = \mu(\bar{K}) = 1 \), then both \( \bar{T}_s(\mu) \) and \( \nabla(\mu) \) are strictly smaller than under the Always-TX policy obtained by setting \( m \to \infty \) (Appendix B.A); then, we show that \( \bar{T}_s(\mu) \) and \( \nabla(\mu) \) are strictly increasing functions of \( \mu(\tilde{K}) \) for all \( m \geq 0 \), hence they are strictly maximized by setting \( \mu(\tilde{K}) = 1 \) (Appendix B.B); finally, we show that \( \bar{T}_s(\mu) \) and \( \nabla(\mu) \) are strictly increasing functions of \( \mu(\bar{K}) \) when \( \mu(\bar{K}) = 1 \) for all \( m > 0 \), hence they are strictly maximized by setting \( \mu(\bar{K}) = 1 \) (Appendix B.C). These results together imply that \( \bar{T}_s(\mu) \) and \( \nabla(\mu) \) are uniquely maximized by the Always-TX policy.

A. \( m < \infty \), \( \mu(\tilde{K}) = \mu(\bar{K}) = 1 \)

By evaluating (58) and (59) with \( \mu(\tilde{K}) = \mu(\bar{K}) = 1 \) and using (66), we obtain\(^5\)

\[
\bar{T}_s(\mu) = \frac{R_s D_s}{\nabla_{\text{max}}} \nabla(\mu) - \frac{R_s (1 - \rho_1)}{1 - \rho_1 (1 - D_p)} \left[ 1 + \frac{(\rho_1 - \rho_0) \phi_1^m}{1 - \rho_0 (1 - D_p) - (\rho_1 - \rho_0) \phi_1^m} \right]
\]

\[
\times \left[ (1 - \phi_1^m)(1 - \rho_1) \zeta + m D_p \phi_1^m (\rho_1 - \rho_0) (1 - D_p) \right],
\]

\[
\nabla(\mu) = \nabla_{\text{max}} \left[ 1 - \frac{(1 - \rho_1) \phi_1^m}{1 - \rho_0 (1 - D_p) - (\rho_1 - \rho_0) \phi_1^m} \right],
\]

\(^5\)Due to space constraints, the details of the algebraic derivations are omitted.
where \( \zeta \) is defined in [24]. By taking the limit \( m \to \infty \), we obtain the performance of the Always-TX policy, yielding (25). By comparing it with policy \( \mu \) under consideration, we obtain

\[
\begin{align*}
\bar{T}_S(\text{Always-TX}) - \bar{T}_S(\mu) &= \frac{(1 - \rho_1)\phi_1^m}{1 - \rho_0(1 - D_p) - (\rho_1 - \rho_0)\phi_1^m} R_s \left[ (D_s - \upsilon_s - \upsilon_{sp}) \right] \quad (69) \\
&+ \zeta \rho_1(D_p - \upsilon_{sp}) + \frac{\rho_1\upsilon_s\upsilon_{sp} + (\rho_1 - \rho_0)D_p[m(1 - D_\mu) + \zeta(1 - \rho_1)]}{1 - \rho_1(1 - D_p)} > 0,
\end{align*}
\]

\[
\nabla(\text{Always-TX}) - \nabla(\mu) = \nabla_{\max} \frac{(1 - \rho_1)\phi_1^m}{1 - \rho_0(1 - D_p) - (\rho_1 - \rho_0)\phi_1^m} > 0. \quad (70)
\]

Therefore, when \( \mu(\bar{K}) = \mu(\bar{K}) = 1 \) and \( m < \infty \), \( \bar{T}_s(\mu) \) and \( \nabla(\mu) \) strictly increase under \( \mu^{[1]} \).

B. \( m \leq \infty \), \( \mu(\bar{K}) = 0 \)

We now show that, if \( \mu(\bar{K}) = 0 \), then both \( \bar{T}_S(\mu) \) and \( \nabla(\mu) \) strictly augment by setting \( \mu(\bar{K}) = 1 \); equivalently, \( \bar{T}_S(\mu) \) and \( \nabla(\mu) \) are strictly increasing functions of \( \mu(\bar{K}) \). By computing the derivative with respect to \( \mu(\bar{K}) \), we obtain

\[
\frac{d\bar{T}_S(\mu)}{d\mu(\bar{K})} = \pi_\mu(\bar{K}) \frac{(1 - \rho_0)R_s D_s - (\rho_1 - \rho_0)\bar{T}_S(\mu)}{1 - \rho_1^{\hat{\mu}(\bar{K})}} \geq \pi_\mu(\bar{K}) \frac{1 - \rho_1}{1 - \rho_1^{\hat{\mu}(\bar{K})}} R_s D_s > 0, \quad (71)
\]

where we have used the fact that \( \bar{T}_S(\mu) \leq R_s D_s \) (see Theorem 4 in Appendix C). Similarly,

\[
\frac{d\nabla(\mu)}{d\mu(\bar{K})} = \pi_\mu(\bar{K}) \frac{(1 - \rho_0)\nabla_{\max} - (\rho_1 - \rho_0)\nabla(\mu)}{1 - \rho_1^{\hat{\mu}(\bar{K})}} \geq \pi_\mu(\bar{K}) \frac{1 - \rho_1}{1 - \rho_1^{\hat{\mu}(\bar{K})}} \nabla_{\max} > 0, \quad (72)
\]

where we have used the fact that \( \nabla(\mu) \leq \nabla_{\max} \). The strict inequality is a consequence of the fact that \( \pi_\mu(\bar{K}) > 0 \), i.e., state \( \bar{K} \) is reachable. Thus, \( \bar{T}_S(\mu) \) and \( \nabla(\mu) \) strictly increase by setting \( \mu(\bar{K}) = 1 \) as in the Always-TX policy.

C. \( m \leq \infty \), \( \mu(\bar{K}) = 1 \), \( \mu(\bar{K}) = 0 \)

Finally, we show that, if \( \mu(\bar{K}) = 0 \) and \( \mu(\bar{K}) = 1 \) (since \( \mu(\bar{K}) = 0 \) was already proved in the previous section to achieve strictly smaller SU throughput and interference than \( \mu(\bar{K}) = 1 \)), \( \bar{T}_S(\mu) \) and \( \nabla(\mu) \) are strictly upper bounded by setting \( \mu(\bar{K}) = 1 \).

We have two cases: if \( m = 0 \), then it follows that \( \pi_\mu(\bar{K}) = 0 \) (i.e., state \( \bar{K} \) cannot be accessed under such policy). In this case, \( \bar{T}_S(\mu) \) and \( \nabla(\mu) \) are unchanged if we let \( \mu(\bar{K}) = 1 \). At this point, since \( \mu(\bar{K}) = 1 \) and \( \mu(\bar{K}) = 1 \), by applying the result of Appendix B.A, we see that \( \bar{T}_S(\mu) < \bar{T}_S(\text{Always-TX}) \) and \( \nabla(\mu) < \nabla(\text{Always-TX}) \).
On the other hand, if \( m > 0 \), then it follows that \( \pi_\mu(\bar{K}) > 0 \). The derivatives of \( \bar{T}_S(\mu) \) and \( \nabla(\mu) \) with respect to \( \mu(\bar{K}) \) are given by

\[
\frac{d\bar{T}_S(\mu)}{d\mu(\bar{K})} = \pi_\mu(\bar{K}) \frac{1 - \rho_1(1 - D_p)\mu(\bar{K}) - \mu(\bar{K})\rho_1 \nu}{1 - \rho_\mu(\bar{K})(1 - D_p) - \mu(\bar{K})\rho_1 \nu} \left[D_s - \nu + \rho_1 - \rho_0 \left(D_s - \bar{T}_S(\mu)\right)\right] \tag{73}
\]

\[
+ \left(1 - \rho_\mu(\bar{K})(1 - D_p) - \mu(\bar{K})\rho_1 \nu\right) \left(D_p - \nu\right) > 0,
\]

\[
\frac{d\nabla(\mu)}{d\mu(\bar{K})} = \pi_\mu(\bar{K}) \frac{1 - \rho_1(1 - D_p) - \rho_1 \nu}{1 - \rho_\mu(\bar{K})(1 - D_p) - \mu(\bar{K})\rho_1 \nu} \left[\nu_{\text{max}} + \frac{\rho_1 - \rho_0}{1 - \rho_1} (\nabla_{\text{max}} - \nabla(\mu))\right] > 0. \tag{74}
\]

where we have used the fact that \( \bar{T}_S(\mu) \leq R_s D_s \) (Theorem 4 in Appendix C) and \( \nabla(\mu) \leq \nabla_{\text{max}} \).

Thus, \( \bar{T}_S(\mu) \) and \( \nabla(\mu) \) strictly increase by setting \( \mu(\bar{K}) = 1 \) as in the Always-TX policy.

The lemma is thus proved.

\[\square\]

**APPENDIX C: GENIE-AIDED UPPER BOUND**

**Theorem 4.** For any \( \mu \in U \), we have

\[
\bar{T}_S(\mu) \leq \frac{\nabla(\mu)}{\nabla_{\text{max}}} R_s D_s = \bar{T}_S^{(GA)} \left(\frac{\nabla(\mu)}{\nabla_{\text{max}}}\right) \leq R_s D_s. \tag{75}
\]

**Proof.** Note from (12) and (13) that

\[
v_S(a_s, b) \leq R_s \left[a_s(\delta_s + \delta_s) + D_p b + \{1 - \rho a_s\}(1 - D_p) b + (1 - \rho_1)a_s(\nu + \nu_s)\right] \triangleq v_S^{(up)}(a_s, b),
\]

\[
v_S(a_s, \bar{K}) \leq R_s \left[a_s(D_s - \nu) + D_p + \{1 - \rho a_s\}(1 - D_p) + a_s(1 - \rho_1) \nu_s\right] \triangleq v_S^{(up)}(a_s, \bar{K}).
\]

These upper bounds are obtained by adding the SU packets that would remain undecoded at the end of the ARQ window, but are decodable in the genie-aided case with a-priori knowledge of the PU packet. Hence, we can upper bound \( \bar{T}_S(\mu) \) as

\[
\bar{T}_S(\mu) \leq \sum_{b=0}^{\infty} \pi_\mu(b) v_S^{(up)}(\mu(b), b) + \pi_\mu(\bar{K}) v_S^{(up)}(\mu(\bar{K}), \bar{K}) + \pi_\mu(\bar{K}) v_S(\mu(\bar{K}), \bar{K})
\]

\[
= \sum_{s \in \mathcal{S}} \pi_\mu(s) R_s D_s + \sum_{b=0}^{\infty} \pi_\mu(b) \{A(b) + \mu(b) [B(b) + C(b)]\}, \tag{76}
\]

January 25, 2018
hence
\[ A(b) = bR_s[1 - \rho_0(1 - D_p)], \]
\[ B(b) = -bR_s[\mu_1(1 - D_p - v_s) - \rho_0(1 - D_p)], \]
\[ C(b) = -(b + 1)R_s\rho_1 v_s. \]

We then obtain
\[ T_S(\mu) - T_S^{(GA)} \left( \frac{\nabla(\mu)}{\nabla_{\text{max}}} \right) + \sum_{b=0}^{\infty} \pi_\mu(b) [A(b) + \mu(b)B(b)] + \sum_{b=1}^{\infty} \pi_\mu(0)Q_\mu(b-1)C(b-1) \]
\[ = T_S^{(GA)} \left( \frac{\nabla(\mu)}{\nabla_{\text{max}}} \right) + \sum_{b=0}^{\infty} \pi_\mu(b) [A(b) + \mu(b)B(b)] + \sum_{b=1}^{\infty} \pi_\mu(0)Q_\mu(b) \left( \frac{\mu(b)}{\phi_1} + \frac{1 - \mu(b)}{\phi_0} \right) \]
\[ = T_S^{(GA)} \left( \frac{\nabla(\mu)}{\nabla_{\text{max}}} \right) + \sum_{b=1}^{\infty} \pi_\mu(b) \left[ A(b) + \mu(b)B(b) + \left( \frac{\mu(b)}{\phi_1} + \frac{1 - \mu(b)}{\phi_0} \right) \right] \]
where in the first equality we used the recursion (52) to express \( Q_\mu(b-1) \) in terms of \( Q_\mu(b) \) and (65); in the second we used \( A(0) = B(0) = 0 \). Finally, by inspection we find that
\[ A(b) + \mu(b)B(b) + \left( \frac{\mu(b)}{\phi_1} + \frac{1 - \mu(b)}{\phi_0} \right) C(b-1) = 0, \]
\[ \text{(78)} \]
hence \( T_S(\mu) \leq T_S^{(GA)} \left( \frac{\nabla(\mu)}{\nabla_{\text{max}}} \right) \). The second inequality in (75) follows from \( \nabla(\mu) \leq \nabla_{\text{max}} \). 

\[ \blacksquare \]

**APPENDIX D: EVALUATION OF \( \eta(s) \)**

We refer to Appendix A for the definition of the terms used in this proof.

\( a) \) **Evaluation of \( \eta(m), m \geq 0 \):** In the proof of Lemma 3 in Appendix B, we have shown that \( T_S(\mu^{[1]} - \Delta_m) \) and \( \nabla(\mu^{[1]} - \Delta_m) \) are given by (69) and (70), respectively. Thus, we obtain (38) after computing their ratio and reordering the terms.

\( b) \) **Evaluation of \( \eta(\bar{K}) \):** Since under policy \( \mu \equiv \mu^{[1]} - \Delta_{\bar{K}} \) we have \( \mu(b) = 1, \forall b \), we find that \( Q_\mu(b) = Q_\mu(1)^b \), hence solving (53) we obtain
\[ \Psi_\mu(x, y, z) = \frac{1 - \rho_1(1 - D_p - v_s)}{1 - \rho_1(1 - D_p)} \left( x + y + \frac{\rho_1 v_s}{1 - \rho_1(1 - D_p)} \right). \]

\[ \blacksquare \]
Then, $\bar{T}_S(\mu)$ and $\nabla(\mu)$ can be found by letting $\mu(b) = 1, \forall b$, $\mu(\leftrightarrow) = 0$ and $\mu(\rightarrow) = 1$ in (58) and (59), yielding

$$\bar{T}_S(\mu) = D_s R_s - \frac{(1 - \rho_1)R_s}{[1 - \rho_1(1 - D_p)][1 - \rho_0(1 - D_p)] - (\rho_1 - \rho_0)\rho_1 v_{sp}} \times \rho_1 v_{sp}[D_s - D_p] + v_{sp}[1 - \rho_0(1 - D_p)] + v_s(1 - \rho_1)\frac{1 - \rho_0(1 - D_p)}{1 - \rho_1(1 - D_p)},$$

$$\nabla(\mu) = \nabla_{\max} \left[ 1 - \frac{(1 - \rho_1)\rho_1 v_{sp}}{[1 - \rho_1(1 - D_p)][1 - \rho_0(1 - D_p)] - (\rho_1 - \rho_0)\rho_1 v_{sp}} \right].$$

We obtain (39) after subtracting the performance of the Always-TX policy, given by (25), and then dividing the terms.

c) Evaluation of $\eta(\rightarrow)$: Similarly to the previous case, since under policy $\mu \equiv \mu^{[1]} - \Delta$ we have $\mu(b) = 1, \forall b$, we find that $Q_\mu(b) = Q_\mu(1)^b$, so that $\Psi_\mu(x, y, z)$ is again given by (79).

Then, $\bar{T}_S(\mu)$ and $\nabla(\mu)$ can be found by letting $\mu(b) = 1, \forall b$, $\mu(\leftrightarrow) = 1$ and $\mu(\rightarrow) = 0$ in (58) and (59), yielding

$$\bar{T}_S(\mu) = \frac{1 - \rho_0}{1 - \rho_0 + \rho_1(D_p - v_{sp})} R_s \left[ D_s - \zeta (1 - \rho_1)\frac{1 - \rho_1(1 - D_p + v_{sp})}{1 - \rho_1(1 - D_p)} \right],$$

$$\nabla(\mu) = \nabla_{\max} \frac{1 - \rho_0}{1 - \rho_0 + \rho_1(D_p - v_{sp})}.$$