A Hybrid Underlay/Overlay Transmission Mode for Cognitive Radio Networks with Statistical Quality-of-Service Provisioning

Yichen Wang, *Member, IEEE*, Pinyi Ren, *Member, IEEE*, Feifei Gao, *Member, IEEE*, and Zhou Su, *Member, IEEE*

***Abstract*—In order to achieve better statistical Quality-of-Service (QoS) provisioning for cognitive radio networks (CRN), in this paper, we develop a hybrid underlay/overlay transmission mode for CRNs. Specifically, by applying the theory of effective capacity and taking PN's activity statistics into consideration, we first analyze the maximum achievable throughput of the CRN under two dominant transmission modes, namely underlay and overlay, respectively, and provide efficient algorithms to derive optimal transmission strategies for the two modes. Following the analyses, we then propose a hybrid underlay/overlay transmission mode, through which the cognitive users' QoS requirements can be better guaranteed and network throughput can be further improved. Moreover, we analyze the optimal transmission strategies for both underlay and overlay modes under two limiting cases. Analyses indicate that 1) for the loose QoS requirement, optimal transmission strategies for both underlay and overlay modes become the water-filling algorithm; and 2) for the stringent QoS requirement, the cognitive user will transmit with constant rate. Furthermore, the impact of imperfect channel estimations on our proposed transmission mode is discussed. Simulation results are provided to demonstrate the impacts of delay QoS requirements and PN's activity statistics on maximizing the delay-constrained throughput for both underlay and overlay modes and verify the effectiveness of our proposed transmission mode. Moreover, for the overlay mode, we observe that 1) a unique optimal sensing time exists under the given QoS constraint; and 2) the optimal sensing time surprisingly increases as the QoS constraint gets more stringent.**

***Index Terms*—Cognitive radio, underlay, overlay, resource allocation, statistical delay QoS provisioning.**

I. Introduction

WITH the rapid developments of wireless communications technologies, spectrum has become the extremely scarce resource for wireless networks. However, current spectrum utilization efficiency is very low due to the traditional static spectrum allocation policy, where a particular portion of spectrum can be only used by a specific type of wireless communications systems [1]. Cognitive radio (CR), which can effectively realize dynamic spectrum access (DSA), has been regarded as an efficient and promising approach to solve the spectrum shortage and underutilization problems [2]. Accordingly, the cognitive radio networks (CRN) have become one of the most important topics and have been extensively researched in the past ten years from both academia and industry.

In order to perform DSA, two transmission modes for CRNs, which are underlay and overlay, respectively, have been proposed. For the overlay mode [3]–[7], the secondary users (SU) can only access the spectrum licensed to the primary networks (PN) when it is idle, i.e., not currently used by the corresponding PNs. The PNs' spectrum usage status can be obtained by spectrum sensing. For the underlay mode [8]–[15], the SUs do not need to determine the PNs' status and are allowed to use the PNs' spectrum even when PNs are active, but need to confine the interference power to a tolerable level all the time. However, no matter which approach is employed, CRNs must protect PN's QoS from degradation. In [8], authors adopted two methods for PN's QoS provisioning. Specifically, if the SU transmitter can obtain accurate channel state information (CSI) between itself and the primary user (PU) receiver, peak/average interference power constraints are imposed to guarantee that the interference power caused by the SU transmitter is tolerable to the PU receiver. On the contrary, if the accurate inter-system CSI is not available for the SU transmitter, the statistical outage probability that the interference from the SU transmitter to the PU receiver is larger than a specific threshold is used to protect PN's QoS.

Compared with conventional wireless networks, delay Quality-of-Service (QoS) provisioning is a more challenging issue for CRNs due to the following reasons. On the one hand, the performances of CRNs are usually constrained by the maximum transmit power. On the other hand, the CRNs have to meet the interference power constraint to protect the performance of PNs from degradation. Consequently, the CRNs cannot guarantee their QoS requirements through increasing the transmit power. Moreover, the performances of CRNs will be degraded not only by the inter-interferences

Manuscript received May 4, 2013; revised August 31, 2013; accepted November 1, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was E. Jorswieck.

This work was supported in part by the National Natural Science Foundation of China under Grant 61172091, the National Science and Technology Major Project under Grant 2013ZX03003004-003, and the fundamental research funds for the central universities of China.

Y. Wang and P. Ren are with the Department of Information and Communications Engineering, School of Electronic and Information Engineering, Xi'an Jiaotong University, Shaanxi 710049 China (e-mail: {wangyichen0819, pyren}@mail.xjtu.edu.cn). P. Ren is the correspondence author.

F. Gao is with Tsinghua National Laboratory for Information Science and Technology (TNList) State Key Laboratory on Intelligent Technology and Systems, Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: feifeigao@ieee.org).

Z. Su is with the Graduate School of Science and Engineering, Waseda University, Japan (e-mail: zhousu@asagi.waseda.jp).

Digital Object Identifier 10.1109/TWC.2013.010214.130797

1536-1276/14$31.00 © 2014 IEEE

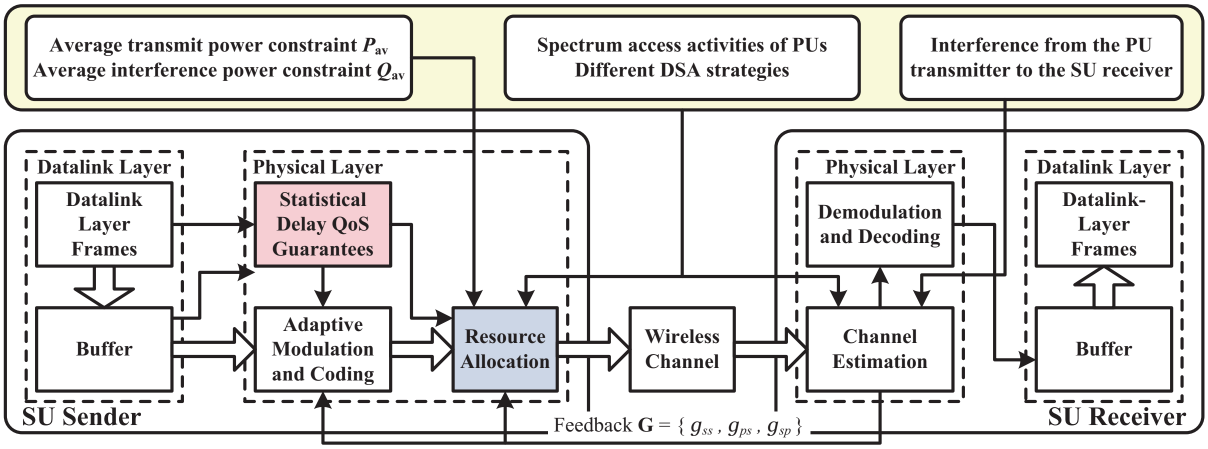


Fig. 1. System framework of the cognitive radio network with statistical delay QoS guarantees.

among SUs, but also by the transmissions of PNs. Although the performances of CRNs can be improved by efficient resource allocation, most resource allocation schemes for CRNs are proposed based on information theory, i.e., optimizing either ergodic capacity or outage capacity [4]–[7], [9]–[15]. However, such an information theoretic framework can only maximize the network throughput either without any delay constraint (ergodic capacity) or with a stringent delay requirement (outage capacity). In realistic wireless services, a wide range of delay constraints may be requested for different applications. Therefore, information theory based resource allocation cannot satisfy the demands for describing a wide range of delay QoS requirements. Moreover, due to the time-varying feature of wireless channels, the deterministic delay QoS provisioning is usually unrealistic for practical wireless networks [16]. Hence, the statistical version should be employed for delay QoS guarantees. Effective capacity [20], which is defined as the maximum constant arrival rate that can be supported by the given time-varying service process, is a powerful approach for statistical delay-QoS provisioning. Tang *et al.* [21] first proposed the statistical QoS driven power and rate adaption strategy over wireless links. This work was extended to the CRNs in [22]–[28]. Specifically, [22]–[24] focus on the underlay mode. The spectrum sensing is considered in [25] and the imperfect channel information is taken into consideration in [26].

Some researches focusing on statistical delay QoS provisioning for CRNs have been investigated, however, there are still some crucial issues need to be solved. First, as PNs' transmissions will significantly affect the performances of CRNs and PNs' activities vary according to time, location, and applications, PNs' activity statistics should be considered, which are ignored in existing investigations for underlay mode. Second, spectrum sensing time is closely related with the performances of CRNs. In order to satisfy the QoS requirements, sensing time should be dynamically adjusted according to different PNs' activity statistics and QoS constraints. Third, spectrum sensing, which is the critically important difference between underlay and overlay, has its own pros and cons. Specifically, on the one hand, determining PNs' status through spectrum sensing allows CRNs to transmit with higher power and results in looser interference constraints, which are beneficial for performance improvement. On the other hand, a certain amount of time will be consumed by spectrum sensing and is unavailable for data transmissions, the performance of CRNs will be correspondingly degraded. It should be noted that there exists some previous works that compare different DSA modes using different tools, such as [29] and [30]. In particular, authors in [29] analyzed three different DSA approaches, which are spreading-based underlay, interference avoidance (IA) based overlay, and spreading-based underlay with IA, respectively. Outage probability was used to evaluate the PN's performance. Analyses showed that the spreading-based underlay with IA can derive lower outage probability than the IA based overlay scheme. In [30], authors investigated an overlay transmission strategy for the multiple-input single-output (MISO) CRN. Corresponding optimization problem was formulated to maximize SU's transmission rate by jointly design the beamformers, phase splitting, and power allocation. Moreover, the performance of the proposed overlay strategy was compared with that of the underlay scheme that combines beamforming, rate splitting, and successive decoding. Numerical results showed that the overlay scheme can provide higher rates in most cases, but it will suffer a significant rate loss due to the learning phase. Consequently, due to distinct features of different DSA modes, they usually have their respective advantages under diverse propagation environments and system parameters.

Motivated by above discussions, in this paper, we develop a novel hybrid underlay/overlay transmission mode for the CRN, where a pair of secondary sender and receiver coexist with a pair of primary sender and receiver by sharing a certain spectrum band. Through using our proposed hybrid transmission mode, we aim at providing better statistical delay QoS provisioning for the CRN. Specifically, by employing the theory of effective capacity, we first analyze the maximum achievable throughput for both underlay and overlay modes as well as the corresponding optimal transmission strategies. Following the analyses, we then design a hybrid underlay/overlay mode for the CRN. In particular, as underlay and overlay modes have their respective advantages under diverse system parameters, our proposed mode divides the two-dimensional plane composed by the QoS constraints and PN's activity statistics into two regions. Then, the SU can dynamically determine whether to use underlay or overlay for DSA under the given QoS constraint and the PN's spectrum-occupancy probability. Moreover, optimal transmission strategies for both underlay and overlay modes under two limiting cases are analyzed. Analyses indicate that 1) optimal transmission strategies under loose QoS requirement become the water-filling algorithm; and 2) for the stringent QoS requirement, the SU will transmit with constant rate. We also discuss the impact of imperfect channel estimations on our proposed transmission mode, which indicates that the underlay mode is more vulnerable to channel estimation errors than the overlay mode. Simulation results are provided to demonstrate the impacts of delay QoS requirements and PN's activity statistics on the maximum delay-constrained throughput for both underlay and overlay modes and verify the effectiveness of our proposed transmission mode. We also observe from the simulation results that the average transmit and interference power constraints play different roles in underlay and overlay modes. Moreover, we observe that the optimal sensing time for the overlay mode increases as the QoS constraint gets more stringent.

The rest of this paper is organized as follows. The system model is described and the background on statistical delay QoS provisioning is provided in Section II. In Section III, we first analyze the optimal transmission strategies for both underlay and overlay modes, then we develop the hybrid underlay/overlay transmission mode. In Section IV, we analyze the optimal transmission strategies for both underlay and overlay modes under two limiting cases and discuss the impact of the imperfect channels estimations on the CRN's performance. Simulation results are provided in Section V. The paper concludes with Section VI.

II. System Model and Background on Statistical Delay QoS Guarantees

A. System Model

We consider a CRN coexisting with a PN by sharing a certain spectrum band with bandwidth B. The PN includes a primary sender (PS) and a primary receiver (PR). The CRN, which is allowed to dynamically access the spectrum bands licensed to the PN, includes a secondary sender (SS) and a secondary receiver (SR). The channel power gains between SS and SR, SS and PR, PS and SR, as well as PS and PR are denoted by g_{ss}, g_{sp}, g_{ps}, and g_{pp}, respectively. All channel power gains are assumed to follow the Nakagami-m fading model. Thus, the probability density function (PDF) of each channel power gain is given by

$$f\left(g\right)=\frac{m^{m}g^{m-1}}{\Gamma\left(m\right)}e^{-mg},\ g\geq 0, \tag{1}$$

where m is the fading parameter. As parameter m varies, where $m\in[1/2,+\infty)$, the Nakagami-m model can span a wide range of fading environments, such as one-sided Gaussian fading channel, Rayleigh fading channel, and additive white Gaussian noise (AWGN) channel. The received noise powers at the SR and all PRs are modeled as independent zero-mean Gaussian random variables with unit variance.

We define the network gain vector (NGV) as $\mathbf{G}\triangleq\{g_{ss},g_{sp},g_{ps}\}$. The PS transmits with constant power $P_{\rm p}$, but the SS transmits with variable power. The statistical delay QoS driven system framework for the CRN is illustrated in Fig. 1. We assume that each frame at the datalink layer of the SS has the same time duration, denoted by T_f. The frames are stored at the transmit buffer and split into bit-streams at the physical layer. The SS employs the adaptive modulation and power control based on the statistical QoS constraint and the channel state information (CSI). We also assume that all channel gains are stationary, ergodic, independent, and block fading processes. The NGV $\mathbf{G}$ can be perfectly estimated by the SR and reliably fed back to the SS[1] (the impact of imperfect channel estimation between SS and PR will be analyzed in Section IV-C and simulation results will be provided in Section V-C). The CRN can dynamically access the spectrum bands licensed to the PN with either underlay mode or overlay mode. We can clearly observe from Fig. 1 that the performance of the SU will be affected by (1) the transmit and interference power constraints, (2) the statistical delay QoS requirement, (3) PN's spectrum access activities, (4) the transmission mode used by SUs, and (5) the interference caused by the PN's transmission.

In this paper, we take the PN's spectrum-occupancy probability into consideration even for the underlay strategy. As the PN's spectrum-occupancy status can be viewed as the two-hypothesis test from the SU's perspective, we denote the probability that the PN does not occupy the channel during the given frame as $P(\mathcal{H}_0)$ and the probability that the selected channel is occupied as $P(\mathcal{H}_1)$.

B. Preliminaries on Statistical QoS Provisioning

In this section, we briefly introduce the concept of statistical delay QoS provisioning. First, we discuss why we adopt such statistical metric for CRN's delay QoS guarantees.

- Currently, the most frequently used method for delay QoS provisioning is to guarantee a minimum transmission rate through resource allocation. Obviously, this approach requires the power control to compensate the deep channel fading. However, it is well known that the transmission rate that can be guaranteed is equal to zero in Rayleigh fading channels with limited transmit power budget. Consequently, such a requirement often cannot be satisfied. The other way for delay QoS provisioning is to strictly control the delay under a specified threshold, but such a hard delay bound is also often unrealistic to guarantee due to the highly time-varying feature of wireless channels.

[1]The CSI between SS and SR, i.e., g_{ss}, can be obtained at SS by the classic channel training, estimation, and feedback mechanisms. The CSI between PS and SR, i.e., g_{ps}, can be derived by SR via estimating the received signal power from PS provided that it knows a priori of the transmit power of PS. The channel power gain g_{sp} between SS and PR can be obtained at SS via the cooperation of the PN. For instance, PR can estimate the channel between itself and SS and then feeds back to SS. An implicit estimation method can also be used as well. If the PN is IEEE 802.16e system and PS is the base station as well as PR is the mobile station, then the PR will transmit channel sounding waveforms to the PS to enable the PS to estimate the channel power gain g_{pp}. Thus, the SS can overhear the uplink channel to obtain g_{sp} [8]. Moreover, The CSI between the PN and CRN can be obtained by using the band manager, which can exchange the channel information between the PN and CN.

- Resource allocation is an efficient way to optimize the performance of wireless networks. Currently, most existing resource allocation schemes are investigated based on information theory, i.e., optimizing the ergodic capacity or outage capacity. However, such an information theoretic framework can only maximize the network capacity either without any delay constraint (ergodic capacity) or with a stringent delay requirement (outage capacity). In realistic wireless services, a wide range of delay constraints may be requested for different applications. In the theory of statistical delay QoS provisioning, a parameter called QoS exponent is defined (it will be introduced later) and the different delay QoS requirements can be reflected by distinct values of QoS exponent, which makes statistical delay QoS provisioning more attractive.
- In the CRN, not only the transmit power constraint needs to be satisfied, but also SU must meet the interference power requirement. Moreover, the performance of CRN will be degraded by PN's transmissions. Consequently, it is more difficult to guarantee the deterministic delay QoS requirement in the CRN.

Based on the above discussions, instead of the deterministic delay QoS requirement, we in this paper employ the statistical version for delay QoS provisioning. The statistical delay QoS guarantee is described by the *delay-bound violation probability*, which can be written as

$$\Pr\{D > D_{\rm th}\} \le P_{\rm th}, \tag{2}$$

where D and $D_{\rm th}$ denote the queueing delay and predefined queueing delay bound, respectively, and $P_{\rm th}$ represents the maximum violation probability. We can find from Eq. (2) that calculating the delay-bound violation probability $\Pr\{D > D_{\rm th}\}$ is critically important for statistical delay QoS provisioning. Fortunately, effective bandwidth [19] and effective capacity [20] are powerful approaches to achieve the above goal. Effective bandwidth is defined as the minimum constant service rate required to support the given data arrival process subject to a QoS constraint specified by the QoS exponent θ. Effective capacity, which is the dual concept of effective bandwidth, determines the maximum constant arrival rate that a given data service process can support in order to guarantee a QoS requirement specified by by the QoS exponent θ.

Consider a stable dynamic discrete-time queueing system with arrival process $\{C[i], i = 1, 2, \cdots\}$ and service process $\{R[i], i = 1, 2, \cdots\}$. Let $\mathcal{C}[t] = \sum_{i=1}^{t} C[i]$ and $\mathcal{R}[t] = \sum_{i=1}^{t} R[i]$ denote the partial sum of the arrival process and service process over time sequence of $i = 1, 2, \cdots, t$, respectively. Then, the effective bandwidth of arrival process and effective capacity of service process, denoted by $\mathrm{E_B}(\theta)$ and $\mathrm{E_C}(\theta)$, respectively, can be written as [19], [20]

$$\left\{ \begin{array}{lcl} \mathrm{E_B}(\theta) &=& \lim\limits_{t\to\infty} \frac{1}{\theta t} \log\left(\mathbb{E}\left\{e^{\theta \mathcal{C}[t]}\right\}\right), \\ \mathrm{E_C}(\theta) &=& -\lim\limits_{t\to\infty} \frac{1}{\theta t} \log\left(\mathbb{E}\left\{e^{-\theta \mathcal{R}[t]}\right\}, \right) \end{array} \right. \tag{3}$$

where θ is called the QoS exponent. If both $\{C[i], i = 1, 2, \cdots\}$ and $\{R[i], i = 1, 2, \cdots\}$ are two uncorrelated processes, the effective bandwidth $\mathrm{E_B}(\theta)$ and effective capacity $\mathrm{E_C}(\theta)$ reduce to

$$\left\{ \begin{array}{lcl} \mathrm{E_B}(\theta) &=& \frac{1}{\theta} \log\left(\mathbb{E}\left\{e^{\theta C[i]}\right\}\right), \\ \mathrm{E_C}(\theta) &=& -\frac{1}{\theta} \log\left(\mathbb{E}\left\{e^{-\theta R[i]}\right\}\right). \end{array} \right. \tag{4}$$

Then, the delay-bound violation probability can be determined by [19]

$$\Pr\{D > D_{\rm th}\} \approx e^{-\theta \mathrm{E_B}(\theta) D_{\rm th}}. \tag{5}$$

We can that if $C[t]$, $D_{\rm th}$, and $P_{\rm th}$ are given, the QoS exponent θ can be determined by

$$\theta = -\frac{1}{\mathrm{E_B}(\theta) D_{\rm th}} \log\left(P_{\rm th}\right). \tag{6}$$

Although Eq. (5) only provides the approximate delay-bound violation probability, this approximate equation has been demonstrated to provide the tight upper-bound of the delay-bound violation probability [20]. Therefore, we can use the QoS exponent θ calculated by Eq. (6) to describe the statistical delay QoS requirement. Specifically, smaller θ leads to looser QoS constraint and larger θ represents more stringent QoS requirement.

Remark: Based on the above discussions, we found that effective bandwidth and effective capacity can efficiently characterize the delay-bound violation probability and thus can employed for statistical delay QoS provisioning. Then, it is natural to raise the following two questions: (1) given the time-varying arrival and service processes, how to determine the QoS exponent θ, such that we can easily analyze the delay-bound violation probability? (2) given the target violation probability, how to find the target θ as a guideline to design the service rate process over the wireless channels?

- For the first question, we can use the theories of effective bandwidth and effective capacity to divide a queuing system with time-varying arrival and service processes into two virtual queuing systems, where the first virtual system has constant service rate described by effective bandwidth $\mathrm{E_B}(\theta)$ and the second virtual system has constant arrival rate described by effective capacity $\mathrm{E_C}(\theta)$. Then, let $\mathrm{E_B}(\theta) = \mathrm{E_C}(\theta)$, we can determine the unique θ such that the delay-bound violation probability can be analyzed.
- For the second question, we can calculate the effective bandwidth and obtain the target θ through Eq. (6). Then, according to the determined QoS exponent θ, we can investigate the corresponding resource allocation scheme to maximize the effective capacity. As effective capacity is the maximum constant arrival rate under the given time-varying service process and QoS requirement, maximizing effective capacity implies that the available delay-guaranteed network throughput is maximized.

As the transceiver cannot determine the properties of data arrival process and wireless channels are usually the bottleneck for data transmissions, in this paper, we mainly focus on investigating the resource allocation strategy subject to the given delay QoS constraint described by the QoS exponent θ, such that the available throughput can be optimized by regulating the data service process.

III. Hybrid Underlay/Overlay Transmission Mode with Statistical QoS Guarantees

Due to the distinct features, different transmission modes usually have their respective advantages under diverse propagation environments and system parameters, which causes that employing a single transmission mode for DSA cannot satisfy the SU's QoS demand. Consequently, we need to develop a new transmission mode to sufficiently guarantee the SU's QoS requirement. In order to achieve this target, we first analyze the optimal transmission strategies for two dominant DSA modes, namely underlay and overlay, respectively. Then, we design a novel statistical delay QoS driven hybrid underlay/overlay transmission mode for the CRN.

A. Optimal Transmission Strategy for Underlay Mode

In this section, we aim at obtaining the optimal transmission strategy for the underlay mode, which can maximize the achievable throughput of the CRN under the given statistical delay QoS requirement determined by the QoS exponent θ. When the underlay mode is applied, the CRN can use the whole frame duration for transmission with adjusted transmit power. Because the PN's spectrum access activities are taken into consideration, the CRN with the underlay mode has two system states during each frame, which are listed as follows:

State 0: The spectrum is not occupied by the PN (idle).

State 1: The spectrum is currently used by the PN (busy).

We denote the service rates of state 0 and 1 as $R_{\rm u}^{(0)}$ and $R_{\rm u}^{(1)}$, respectively, then the maximum achievable service rates for the two states are determined by

$$\begin{cases} R_{\rm u}^{(0)} = T_f B \log_2\left(1+g_{ss}P_{\rm s,u}\right), \\ R_{\rm u}^{(1)} = T_f B \log_2\left(1+\frac{g_{ss}P_{\rm s,u}}{g_{ps}P_{\rm p}+1}\right), \end{cases} \tag{7}$$

where $P_{\rm s,u}$ denotes the transmit power of the SS. Note that as the CRN cannot determine whether the spectrum is currently occupied by the PN or not, the same transmit power is used for the two states. Since the PN's spectrum-occupancy probability is taken into consideration for the underlay mode, the formulation of effective capacity developed in [21]–[24] cannot be directly employed. Therefore, we first provide the effective capacity of the CRN with the underlay mode through the following theorem.

Theorem 1: Suppose the CRN with the underlay mode has two system states, namely state 0 and state 1, respectively. If the probabilities of the two system states are denoted by $P(\mathcal{H}_0)$ and $P(\mathcal{H}_1)$, respectively, and the corresponding service rates are denoted by $R_{\rm u}^{(0)}$ and $R_{\rm u}^{(1)}$, respectively, the effective capacity of the CRN with the underlay mode, denoted by $\mathrm{E}_{\rm C}^{\rm u}(\theta)$, can be written as

$$\mathrm{E}_{\rm C}^{\rm u}(\theta) = -\frac{1}{\theta}\log\left(P(\mathcal{H}_0)\mathbb{E}\left\{e^{-\theta R_{\rm u}^{(0)}}\right\}+P(\mathcal{H}_1)\mathbb{E}\left\{e^{-\theta R_{\rm u}^{(1)}}\right\}\right). \tag{8}$$

Proof: The proof is provided in Appendix A. ■

Because the SS usually has the upper-bounded transmit power, we assume that the SS should satisfy the average transmit power constraint, which can be written as

$$P(\mathcal{H}_0)\mathbb{E}_{\mathbf{G}}\{P_{\rm s,u}\} + P(\mathcal{H}_1)\mathbb{E}_{\mathbf{G}}\{P_{\rm s,u}\} \le P_{\rm av}, \tag{9}$$

where $P_{\rm av}$ represents the maximum average transmit power of the SS and $\mathbb{E}_{\mathbf{G}}\{\cdot\}$ denotes the expectation operation with respective to the NGV $\mathbf{G}$. In order to protect PN's QoS, the CRN should also meet the average interference power constraint imposed by the PN, which can be formulated by

$$\mathbb{E}_{\mathbf{G}}\{g_{sp}P_{\rm s,u}\} \le Q_{\rm av}, \tag{10}$$

where $Q_{\rm av}$ denotes the maximum average interference power that the PR can tolerate.

Our objective is to find the optimal transmit power to maximize the effective capacity of the CRN subject to the upper-bounded average transmit and interference power. Thus, our optimization problem can be mathematically formulated as

$$\begin{aligned}(\mathrm{P1})\ \max_{P_{\rm s,u}}\ & -\frac{1}{\theta}\log\Big(P(\mathcal{H}_0)\mathbb{E}_{\mathbf{G}}\left\{e^{-\theta R_{\rm u}^{(0)}}\right\} \\ & \qquad +P(\mathcal{H}_1)\mathbb{E}_{\mathbf{G}}\left\{e^{-\theta R_{\rm u}^{(1)}}\right\}\Big) \\ \text{s.t.}\ & \text{Eqs. (9) and (10).}\end{aligned} \tag{11}$$

Obviously, the transmit power $P_{\rm s,u}$ is the function of both NGV $\mathbf{G}$ and the QoS exponent θ. Since $\log(\cdot)$ is a monotonically increasing function, plugging Eq. (7) into Eq. (11), the above maximization problem $(\mathrm{P1})$ is equivalent to the following minimization problem:

$$\begin{aligned}(\mathrm{P2})\quad \min_{P_{\rm s,u}}\quad & P(\mathcal{H}_0)\mathbb{E}_{\mathbf{G}}\left\{\left[1+g_{ss}P_{\rm s,u}\right]^{-\beta_{\rm u}}\right\} \\ & +P(\mathcal{H}_1)\mathbb{E}_{\mathbf{G}}\left\{\left[1+\frac{g_{ss}P_{\rm s,u}}{g_{ps}P_{\rm p}+1}\right]^{-\beta_{\rm u}}\right\} \\ \text{s.t.}\quad & \text{Eqs. (9) and (10),}\end{aligned} \tag{12}$$

where $\beta_{\rm u} = \theta T_f B/\log 2$ is the *normalized QoS exponent* and can characterize the statistical delay QoS requirement. We can prove that the problem (P2) is strictly convex over $P_{\rm s,u}$. and thus has a unique optimal solution. Construct the Lagrangian function, denoted by $\mathcal{L}_{\rm u}\left(P_{\rm s,u},\lambda_{\rm u},\mu_{\rm u}\right)$, as follows:

$$\begin{aligned}&\mathcal{L}_{\rm u}\left(P_{\rm s,u},\lambda_{\rm u},\mu_{\rm u}\right) \\ =\ & P(\mathcal{H}_0)\mathbb{E}_{\mathbf{G}}\left\{\left[1+g_{ss}P_{\rm s,u}\right]^{-\beta_{\rm u}}\right\} \\ & +P(\mathcal{H}_1)\mathbb{E}_{\mathbf{G}}\left\{\left[1+\frac{g_{ss}P_{\rm s,u}}{g_{ps}P_{\rm p}+1}\right]^{-\beta_{\rm u}}\right\} \\ & +\lambda_{\rm u}\Big\{\mathbb{E}_{\mathbf{G}}\left[P_{\rm s,u}\right]-P_{\rm av}\Big\}+\mu_{\rm u}\Big\{\mathbb{E}_{\mathbf{G}}\left[g_{sp}P_{\rm s,u}\right]-Q_{\rm av}\Big\},\end{aligned} \tag{13}$$

where $\lambda_{\rm u}$ and $\mu_{\rm u}$ are the Lagrangian multipliers associated with the constraints given by Eqs. (9) and (10), respectively. Then, the Lagrange dual function can be written as $\mathcal{G}_{\rm u}\left(\lambda_{\rm u},\mu_{\rm u}\right) = \min_{P_{\rm s,u}\ge 0}\mathcal{L}_{\rm u}\left(P_{\rm s,u},\lambda_{\rm u},\mu_{\rm u}\right)$. Based on the Lagrange dual decomposition method, it can be decoupled into a series of parallel subproblems with the same structure. For each NGV $\mathbf{G}$, the subproblem can be represented by

$$\begin{aligned}(\mathrm{P3})\ \min_{P_{\rm s,u}\ge 0}\ & P(\mathcal{H}_0)\left[1+g_{ss}P_{\rm s,u}\right]^{-\beta_{\rm u}} \\ & +P(\mathcal{H}_1)\left[1+\frac{g_{ss}P_{\rm s,u}}{g_{ps}P_{\rm p}+1}\right]^{-\beta_{\rm u}}+\lambda_{\rm u}P_{\rm s,u}+\mu_{\rm u}g_{sp}P_{\rm s,u}.\end{aligned} \tag{14}$$

Algorithm 1: Iterative algorithm for obtaining the optimal solution of problem $(\mathbf{P3})$.

1) Initialization: $P_{\mathrm{s,u}}^{(0)}$, $k=0$;
2) Repeat:
   a) Calculate $F'\left(P_{\mathrm{s,u}}^{(k)}\right)$ and $F''\left(P_{\mathrm{s,u}}^{(k)}\right)$ by
   b) Update $P_{\mathrm{s,u}}^{(k+1)}$ by
   $$P_{\mathrm{s,u}}^{(k+1)} = P_{\mathrm{s,u}}^{(k)} - \left[F''\left(P_{\mathrm{s,u}}^{(k)}\right)\right]^{-1} F'\left(P_{\mathrm{s,u}}^{(k)}\right)$$
   c) $k=k+1$;
3) Stop when $\left|F'\left(P_{\mathrm{s,u}}^{(k)}\right)\right| < \varepsilon$, where ε is the termination scalar.
4) The optimal transmit power:
$$P_{\mathrm{s,u}} = \max\left\{P_{\mathrm{s,u}}^{(k)}, 0\right\}$$

Fig. 2. Newton's method for deriving the optimal power allocation of (P3).

Let $F_{\mathrm{u}}(P_{\mathrm{s,u}})$ denotes the objective function in Eq. (14). If $P_{\mathrm{s,u}} > 0$, the following equality must be hold:

$$\begin{aligned}\frac{dF_{\mathrm{u}}(P_{\mathrm{s,u}})}{dP_{\mathrm{s,u}}} &= -P(\mathcal{H}_0)\beta_{\mathrm{u}} g_{ss}\left[1+g_{ss}P_{\mathrm{s,u}}\right]^{-\beta_{\mathrm{u}}-1} \\ &\quad -P(\mathcal{H}_1)\frac{\beta_{\mathrm{u}} g_{ss}}{g_{ps}P_{\mathrm{p}}+1}\left[1+\frac{g_{ss}P_{\mathrm{s,u}}}{g_{ps}P_{\mathrm{p}}+1}\right]^{-\beta_{\mathrm{u}}-1} \\ &\quad +\lambda_{\mathrm{u}} + \mu_{\mathrm{u}} g_{sp} = 0. \end{aligned} \tag{15}$$

Unfortunately, we cannot obtain the closed-form solution from Eq. (15). However, we can easily prove that $F_{\mathrm{u}}(P_{\mathrm{s,u}})$ is twice differentiable over $P_{\mathrm{s,u}}$, thus we can employ Newton's method [31] to iteratively derive the optimal solution and the iterative algorithm is given by Fig. 2.

In order to obtain the optimal solution of problem (P2), we need to determine the optimal Lagrangian multipliers, denoted by λ_{u}^* and μ_{u}^*, respectively. Based on the optimization theory [31], $(\lambda_{\mathrm{u}}^*, \mu_{\mathrm{u}}^*)$ can be derived by solving the following Lagrangian dual problem to problem (P2):

$$\max_{\lambda_{\mathrm{u}}\geq 0,\ \mu_{\mathrm{u}}\geq 0} \quad \mathcal{G}_{\mathrm{u}}\left(\lambda_{\mathrm{u}}, \mu_{\mathrm{u}}\right), \tag{16}$$

where $\mathcal{G}_{\mathrm{u}}\left(\lambda_{\mathrm{u}}, \mu_{\mathrm{u}}\right)$ is the Lagrange dual function. As problem (P2) is convex and there must exists sufficiently small $P_{\mathrm{s,u}} \geq 0$ such that constraints given by Eqs. (9) and (10) hold with strict inequalities, the Slater's condition is satisfied, which implies that strong duality between (P2) its dual problem given by Eq. (16) holds. Consequently, the duality gap between (P2) and its dual problem is zero [31]. In order to maximize dual function $\mathcal{G}_{\mathrm{u}}\left(\lambda_{\mathrm{u}}, \mu_{\mathrm{u}}\right)$, the optimal Lagrangian multipliers $(\lambda_{\mathrm{u}}^*, \mu_{\mathrm{u}}^*)$ can be iteratively obtained by the subgradient method.

Theorem 2: Suppose that the Lagrangian multipliers in the ith iteration are denoted by $\left(\lambda_{\mathrm{u}}^{(i)}, \mu_{\mathrm{u}}^{(i)}\right)$, then the subgradient for $\lambda_{\mathrm{u}}^{(i)}$ and $\mu_{\mathrm{u}}^{(i)}$, denoted by $\Theta_{\mathrm{u}}^{(i)}$ and $\Phi_{\mathrm{u}}^{(i)}$, respectively, are

$$\begin{cases}\Theta_{\mathrm{u}}^{(i)} = \mathbb{E}_{\mathbf{G}}\left\{P_{\mathrm{s,u}}^{(i)}\right\} - P_{\mathrm{av}}, \\ \Phi_{\mathrm{u}}^{(i)} = \mathbb{E}_{\mathbf{G}}\left\{g_{sp}P_{\mathrm{s,u}}^{(i)}\right\} - Q_{\mathrm{av}},\end{cases} \tag{17}$$

where $P_{\mathrm{s,u}}^{(i)}$ is the derived transmit power in the ith iteration.

Based on Theorem 2, the subgradient method for finding the optimal lagrangian multipliers λ_{u}^* and μ_{u}^* is shown in Fig. 3.

Algorithm 2: Subgradient method for determining the optimal Lagrangian multipliers $(\lambda_{\mathrm{u}}^*, \mu_{\mathrm{u}}^*)$ in problem $(\mathbf{P2})$.

1) Initialization: $\lambda_{\mathrm{u}}^{(0)}$, $\mu_{\mathrm{u}}^{(0)}$, $i=0$;
2) Repeat:
   a) Calculate $P_{\mathrm{s,u}}^{(i)}$ by Algorithm 1 under $\lambda_{\mathrm{u}}^{(i)}$ and $\mu_{\mathrm{u}}^{(i)}$;
   b) Calculate $\Theta_{\mathrm{u}}^{(i)}$ and $\Phi_{\mathrm{u}}^{(i)}$ by (17);
   c) Update $\lambda_{\mathrm{u}}^{(i+1)}$ and $\mu_{\mathrm{u}}^{(i+1)}$ by
   $$\begin{aligned}\lambda_{\mathrm{u}}^{(i+1)} &= \lambda_{\mathrm{u}}^{(i)} + \alpha_1\Theta_{\mathrm{u}}^{(i)} \\ \mu_{\mathrm{u}}^{(i+1)} &= \mu_{\mathrm{u}}^{(i)} + \alpha_2\Phi_{\mathrm{u}}^{(i)}\end{aligned}$$
3) Stop when $\lambda_{\mathrm{u}}^{(i)}$ and $\mu_{\mathrm{u}}^{(i)}$ converge, where α_1 and α_2 are the step sizes for finding the optimal λ_{u}^* and μ_{u}^*, respectively.

Fig. 3. Iterative algorithm to determine optimal Lagrangian multipliers $(\lambda_{\mathrm{u}}^*, \mu_{\mathrm{u}}^*)$ for the CRN with underlay mode.

According to Algorithms 1 and 2, we can derive the optimal transmit power for the underlay mode, denoted by $P_{\mathrm{s,u}}^*$.

B. Optimal Transmission Strategy for Overlay Mode

In this section, we aim at deriving the optimal transmission strategy for the overlay mode. As discussed previously, while applying the overlay mode, the SS needs to utilize the spectrum sensing to determine the spectrum occupation status and only the vacant spectrum can be used by the CRN. According to [17], the PN's spectrum occupation status can be determined between the following two hypotheses:

$$\begin{cases}\mathcal{H}_0: & y[i] = n[i] \\ \mathcal{H}_1: & y[i] = \sqrt{g_{ps}[i]}x_{\mathrm{p}}[i] + n[i]\end{cases} \tag{18}$$

where $y[i]$ is the signal received by the SR, $x_{\mathrm{p}}[i]$ is the transmit signal of the PS, and $n[i]$ is the AWGN at the SR. From [17], we have known that the probabilities of false alarm and detection for the energy detector, denoted by P_f and P_d, respectively, can be written as

$$\begin{cases}P_f = \mathcal{Q}\left(\left(\frac{\varepsilon}{\sigma^2}-1\right)\sqrt{\tau f_s}\right) \\ P_d = \mathcal{Q}\left(\left(\frac{\varepsilon}{\sigma^2}-\gamma-1\right)\sqrt{\frac{\tau f_s}{2\gamma+1}}\right)\end{cases} \tag{19}$$

where γ is the received average signal-to-noise ratio, f_s is the sampling frequency, τ is the sensing time, ε denotes the detection threshold, σ^2 is the variance of the AWGN, and $\mathcal{Q}(x) = \left(1/\sqrt{2\pi}\right)\int_x^\infty e^{-(t^2/2)}dt$ represents complementary distribution function of the standard Gaussian. If the target-detection probability, denoted by $\overline{P}_d$, is given, the probability of false alarm can be calculated by [17]

$$P_f = \mathcal{Q}\left(\sqrt{2\gamma+1}\mathcal{Q}^{-1}\left(\overline{P}_d\right) + \sqrt{\tau f_s}\gamma\right). \tag{20}$$

Due to the unavoidable sensing errors, the CRN has four system states, named state 0, 1, 2, and 3, respectively. **State 0** represents the spectrum is not used by the PN and detected as idle. **State 1** represents the spectrum is occupied by the PN but detected as idle. **State 2** denotes the spectrum is not used by PN but detected as busy. **State 3** represents the spectrum is occupied by the PN and detected as busy. As the overlay mode only allows the CRN to access the spectrum only when the spectrum is determined as idle, the service rates of the CRN under the above four states, denoted by $R_{\mathrm{o}}^{(0)}$, $R_{\mathrm{o}}^{(1)}$, $R_{\mathrm{o}}^{(2)}$, and

$R_{\rm o}^{(3)}$, respectively, are determined by

$$\begin{cases} R_{\rm o}^{(0)} &= (T_f-\tau)B\log_2\left(1+g_{ss}P_{\rm s,o}\right), \\ R_{\rm o}^{(1)} &= (T_f-\tau)B\log_2\left(1+\frac{g_{ss}P_{\rm s,o}}{g_{ps}P_{\rm p}+1}\right), \\ R_{\rm o}^{(2)} &= 0, \\ R_{\rm o}^{(3)} &= 0, \end{cases} \tag{21}$$

where $P_{\rm s,o}$ is the transmit power of the SS. Similar with the discussions for the underlay mode, the transition probability matrix for the overlay mode, denoted by $\mathbf{P_o} = \left[p_{\rm o}^{ij}\right]$ $(i,j = 1,\cdots,4)$, can be constructed, where $p_{\rm o}^{ij}$ $(i,j = 0,1,2,3)$ is the transition probability from state i to state j. As the four states are all time-uncorrelated, the transition probabilities $p_{\rm o}^{i0} = P(\mathcal{H}_0)(1-P_f)$, $p_{\rm o}^{i1} = P(\mathcal{H}_1)(1-P_d)$, $p_{\rm o}^{i2} = P(\mathcal{H}_0)P_f$, and $p_{\rm o}^{i3} = P(\mathcal{H}_1)P_d$. Based on the derived transition probability matrix $\mathbf{P_o}$, the following theorem provides the effective capacity of the CRN with the overlay mode.

Theorem 3: Suppose the CRN with the overlay mode has four time-uncorrelated system states, the service rates of the CRN under these four states are denoted by $R_{\rm o}^{(0)}$, $R_{\rm o}^{(1)}$, $R_{\rm o}^{(2)}$, and $R_{\rm o}^{(3)}$, respectively, and the probabilities that spectrum is idle and busy are denoted by $P(\mathcal{H}_0)$ and $P(\mathcal{H}_1)$, respectively. The corresponding effective capacity of the CRN with the overlay mode, denoted by $\mathrm{E}_{\rm C}^{\rm o}(\theta)$, can be written as

$$\begin{aligned} \mathrm{E}_{\rm C}^{\rm o}(\theta) = &-\frac{1}{\theta}\log\bigg(P(\mathcal{H}_0)(1-P_f)\mathbb{E}\left\{e^{-\theta R_{\rm o}^{(0)}}\right\} \\ &+P(\mathcal{H}_1)(1-P_d)\mathbb{E}\left\{e^{-\theta R_{\rm o}^{(1)}}\right\} \\ &+P(\mathcal{H}_0)P_f\mathbb{E}\left\{e^{-\theta R_{\rm o}^{(2)}}\right\}+P(\mathcal{H}_1)P_d\mathbb{E}\left\{e^{-\theta R_{\rm o}^{(3)}}\right\}\bigg), \end{aligned} \tag{22}$$

where P_f and P_d denote the probabilities of false alarm and detection, respectively.

Proof: The proof is similar to theorem 1 and is omitted here due to the lack of space. ■

Similar with the underlay mode, the CRN with the overlay mode also needs to meet the average transmit power constraint, which can be written as

$$\frac{T_f-\tau}{T_f}\Big\{P(\mathcal{H}_0)(1-P_f)+P(\mathcal{H}_1)(1-P_d)\Big\}\mathbb{E}_{\mathbf{G}}\left\{P_{\rm s,o}\right\}\leq P_{\rm av}. \tag{23}$$

Due to sensing errors, the interference from SS to PS is inevitable and thus the CRN should also satisfy the average interference power requirement, which is determined by

$$\frac{T_f-\tau}{T_f}\Big\{P(\mathcal{H}_1)(1-P_d)\mathbb{E}_{\mathbf{G}}\left\{g_{sp}P_{\rm s,o}\right\}\Big\}\leq Q_{\rm av}. \tag{24}$$

Our objective is to derive the optimal transmission strategy in terms of the transmit power $P_{\rm s,o}$ and sensing time τ to maximize the effective capacity of the overlay based CRN under the average transmit and interference power constraints. Consequently, our optimization problem can be formulated as

$$\begin{aligned} \text{(P4)}\ \max_{\tau,P_{\rm s,o}}\ & -\frac{1}{\theta}\log\bigg(P(\mathcal{H}_0)(1-P_f)\mathbb{E}_{\mathbf{G}}\left\{e^{-\theta R_{\rm o}^{(0)}}\right\} \\ & +P(\mathcal{H}_1)(1-P_d)\mathbb{E}_{\mathbf{G}}\left\{e^{-\theta R_{\rm o}^{(1)}}\right\} \\ & +P(\mathcal{H}_0)P_f+P(\mathcal{H}_1)P_d\bigg) \end{aligned} \tag{25}$$

$$\text{s.t.}\quad \text{Eqs. (23) and (24)},\ P_d\geq\overline{P}_d,\ 0<\tau<T_f$$

where $\overline{P}_d$ denotes the target-detection probability. Plugging Eq. (21) into (25), the problem (P4) is equivalent to the following minimization problem:

$$\begin{aligned} \text{(P5)}\ \min_{\tau,P_{\rm s,o}}\ & P(\mathcal{H}_0)(1-P_f)\mathbb{E}_{\mathbf{G}}\Big\{\left[1+g_{ss}P_{\rm s,o}\right]^{-\beta_{\rm o}(\tau)}\Big\} \\ & +P(\mathcal{H}_1)(1-P_d)\mathbb{E}_{\mathbf{G}}\left\{\left[1+\frac{g_{ss}P_{\rm s,o}}{g_{ps}P_{\rm p}+1}\right]^{-\beta_{\rm o}(\tau)}\right\} \\ & +P(\mathcal{H}_0)P_f+P(\mathcal{H}_1)P_d \end{aligned} \tag{26}$$

$$\text{s.t.}\quad \text{Eqs. (23) and (24)},\ P_d\geq\overline{P}_d,\ 0<\tau<T_f.$$

where $\beta_{\rm o}(\tau) = \theta(T_f-\tau)B/\log 2$ is the *normalized QoS exponent* for the overlay mode and is the function of sensing time τ. We can easily prove that the problem (P5) is strictly convex over $P_{\rm s,o}$, thus there is a unique optimal transmit power for the overlay mode under any given P_d and τ. Moreover, in Appendix B, we proved that the optimal solution to problem (P5) can be achieved by $P_d = \overline{P}_d$. Let $\mathbb{E}\left\{G(\tau,P_{\rm s,o})\right\}$ denote the objective function of (P5), the following theorem shows that there exists the optimal sensing time which can minimize $\mathbb{E}\left\{G(\tau,P_{\rm s,o})\right\}$.

Theorem 4: If the probabilities of false alarm and detection are determined by Eq. (19), there exists optimal sensing time τ that can minimize the objective function $\mathbb{E}_{\mathbf{G}}\left\{G(\tau,P_{\rm s,o})\right\}$. Moreover, if $P_f\leq 0.5$, the objective function $\mathbb{E}_{\mathbf{G}}\left\{G(\tau,P_{\rm s,o})\right\}$ is convex with respect to τ.

Proof: The proof is provided in Appendix C. ■

Theorem 4 indicates that 1) the optimal sensing time must exist within the interval $(0,T_f)$; and 2) the optimal sensing time is unique if τ falls into the range in which $P_f\leq 0.5$. Although the objective function $\mathbb{E}\left\{G(\tau,P_{\rm s,o})\right\}$ is convex over τ and $P_{\rm s,o}$, respectively, unfortunately, it is not jointly convex over $(\tau,P_{\rm s,o})$. Consequently, problem (P5) should be solved by two steps. Specifically, in the first step, we calculate the optimal transmit power under the given sensing time. In the second step, the optimal sensing time can be determined by exhausting searching. For any given sensing time τ, construct the Lagrangian function for problem (P5), denoted by $\mathcal{L}_{\rm o}\left(P_{\rm s,o},\lambda_{\rm o},\mu_{\rm o}\right)$, as shown at the top of the next page, where $\lambda_{\rm o}$ and $\mu_{\rm o}$ are Lagrangian multipliers associated with constraints given by Eqs. (23) and (24), respectively. Based on the Lagrange dual decomposition method, under the given τ, the Lagrange dual function $\mathcal{G}_{\rm o}(\lambda_{\rm o},\mu_{\rm o}) = \min_{P_{\rm s,o}\geq 0}\mathcal{L}_{\rm o}\left(P_{\rm s,o},\lambda_{\rm o},\mu_{\rm o}\right)$ can also be decoupled into a series of parallel subproblems with the same structure and each

$$\begin{aligned}\mathcal{L}_{\rm o}\left(P_{\rm s,o},\lambda_{\rm o},\mu_{\rm o}\right) \;=\;& P(\mathcal{H}_0)(1-P_f)\mathbb{E}_{\mathbf{G}}\left\{\left[1+g_{ss}P_{\rm s,o}\right]^{-\beta_{\rm o}(\tau)}\right\}+P(\mathcal{H}_1)(1-\overline{P}_d)\mathbb{E}_{\mathbf{G}}\left\{\left[1+\frac{g_{ss}P_{\rm s,o}}{g_{ps}P_{\rm p}+1}\right]^{-\beta_{\rm o}(\tau)}\right\}\\ &+\lambda_{\rm o}\left\{\frac{T_f-\tau}{T_f}\Big(P(\mathcal{H}_0)(1-P_f)+P(\mathcal{H}_1)(1-\overline{P}_d)\Big)\mathbb{E}_{\mathbf{G}}\left\{P_{\rm s,o}\right\}-P_{\rm av}\right\}\\ &+\mu_{\rm o}\left\{\frac{T_f-\tau}{T_f}P(\mathcal{H}_1)(1-\overline{P}_d)\mathbb{E}_{\mathbf{G}}\left\{g_{sp}P_{\rm s,o}\right\}-Q_{\rm av}\right\}, \end{aligned} \tag{27}$$

subproblem can be represented by

$$\begin{aligned}\text{(P6)}\ \min_{P_{\rm s,o}\geq 0}\ & P(\mathcal{H}_0)(1-P_f)\Big[1+g_{ss}P_{\rm s,o}\Big]^{-\beta_{\rm o}(\tau)}\\ &+P(\mathcal{H}_1)(1-\overline{P}_d)\left[1+\frac{g_{ss}P_{\rm s,o}}{g_{ps}P_{\rm p}+1}\right]^{-\beta_{\rm o}(\tau)}\\ &+\lambda_{\rm o}\frac{T_f-\tau}{T_f}\Big(P(\mathcal{H}_0)(1-P_f)+P(\mathcal{H}_1)(1-\overline{P}_d)\Big)P_{\rm s,o}\\ &+\mu_{\rm o}\frac{T_f-\tau}{T_f}P(\mathcal{H}_1)\big(1-\overline{P}_d\big)g_{sp}P_{\rm s,o}. \end{aligned} \tag{28}$$

Let $F_{\rm o}(P_{\rm s,o})$ denotes the objective function in Eq. (28). If $P_{\rm s,o}>0$, the following equality must hold:

$$\begin{aligned}\frac{dF_{\rm o}(P_{\rm s,o})}{dP_{\rm s,o}} =& -P(\mathcal{H}_0)(1-P_f)g_{ss}\beta_{\rm o}(\tau)\Big[1+g_{ss}P_{\rm s,o}\Big]^{-\beta_{\rm o}(\tau)-1}\\ &-P(\mathcal{H}_1)(1-\overline{P}_d)\frac{g_{ss}\beta_{\rm o}(\tau)}{g_{ps}P_{\rm p}+1}\left[1+\frac{g_{ss}P_{\rm s,o}}{g_{ps}P_{\rm p}+1}\right]^{-\beta_{\rm o}(\tau)-1}\\ &+\lambda_{\rm o}\frac{T_f-\tau}{T_f}\Big(P(\mathcal{H}_0)(1-P_f)+P(\mathcal{H}_1)(1-\overline{P}_d)\Big)\\ &+\mu_{\rm o}g_{sp}\frac{T_f-\tau}{T_f}P(\mathcal{H}_1)(1-\overline{P}_d)=0. \end{aligned} \tag{29}$$

Similar with the analysis for the underlay mode, we cannot derive the closed-form solution from Eq. (29). As $F_{\rm o}(P_{\rm s,o})$ is twice differentiable over $P_{\rm s,o}$, the Newton's method can also be applied to derive the optimal solution for problem (P6). In order to obtain the optimal transmit power under the given sensing time τ, we also employ the subgradient method to iteratively determine the optimal Lagrangian multipliers $(\lambda_{\rm o}^*,\mu_{\rm o}^*)$. If we denote $\left(\lambda_{\rm o}^{(i)},\mu_{\rm o}^{(i)}\right)$ as the Lagrangian multipliers in the ith iteration, we can prove that the subgradient for $\lambda_{\rm o}^{(i)}$ and $\mu_{\rm o}^{(i)}$, denoted by $\Theta_{\rm o}^{(i)}$ and $\Phi_{\rm o}^{(i)}$, respectively, can be written as

$$\begin{cases}\Theta_{\rm o}^{(i)}=\frac{T_f-\tau}{T_f}\Big\{P(\mathcal{H}_0)(1-P_f)\mathbb{E}_{\mathbf{G}}\Big\{P_{\rm s,o}^{(i)}\Big\}\\ \qquad\qquad +P(\mathcal{H}_1)(1-\overline{P}_d)\mathbb{E}_{\mathbf{G}}\Big\{P_{\rm s,o}^{(i)}\Big\}\Big\}-P_{\rm av},\\ \Phi_{\rm o}^{(i)}=\frac{T_f-\tau}{T_f}\Big\{P(\mathcal{H}_1)(1-\overline{P}_d)\mathbb{E}_{\mathbf{G}}\Big\{g_{sp}P_{\rm s,o}^{(i)}\Big\}\Big\}-Q_{\rm av},\end{cases} \tag{30}$$

where $P_{\rm s,o}^{(i)}$ is the obtained transmit power in the ith iteration. Then, we can derive the optimal transmission strategy $\left(\tau^*,P_{\rm s,o}^*\right)$ through the algorithm shown in Fig. 4.

Algorithm 3: Optimal transmission strategy for the cognitive radio network with overlay mode.

00. For $\tau=0:T_f$
01. Initialization: $\lambda_{\rm o}^{(0)}$, $\mu_{\rm o}^{(0)}$, $i=0$;
02. Repeat:
03. Initialization: $P_{\rm s,o}^{(i,0)}$, $k=0$;
04. Repeat:
05. Calculate $F'\left(P_{\rm s,o}^{(i,k)}\right)$ and $F''\left(P_{\rm s,o}^{(i,k)}\right)$;
06. Update $P_{\rm s,o}^{(i,k+1)}$ by
$$P_{\rm s,o}^{(i,k+1)}=P_{\rm s,o}^{(i,k)}-\left[F''\left(P_{\rm s,o}^{(i,k)}\right)\right]^{-1}F'\left(P_{\rm s,o}^{(i,k)}\right)$$
07. Stop when $\left|F'\left(P_{\rm s,o}^{(i,k)}\right)\right|<\varepsilon$, where ε is the prespecified termination scalar;
08. $P_{\rm s,o}^{(i)}=\max\left\{P_{\rm s,o}^{(i,k)},0\right\}$;
09. Calculate the subgradient of $\lambda_{\rm o}^{(i)}$ and $\mu_{\rm o}^{(i)}$ by (30);
10. Update $\lambda_{\rm o}^{(i+1)}$ and $\mu_{\rm o}^{(i+1)}$ by
$$\begin{aligned}\lambda_{\rm o}^{(i+1)} &= \lambda_{\rm o}^{(i)}+\alpha_1\Theta_{\rm o}^{(i)},\\ \mu_{\rm o}^{(i+1)} &= \mu_{\rm o}^{(i)}+\alpha_2\Phi_{\rm o}^{(i)};\end{aligned}$$
11. Stop if $\lambda_{\rm o}^{(i)}$ and $\mu_{\rm o}^{(i)}$ converge, where α_1 and α_2 are the step sizes for finding the optimal $\lambda_{\rm o}^*$ and $\mu_{\rm o}^*$, respectively.
12. End
13. Optimal transmission strategy:
$$\begin{aligned}\tau^* &= \text{argmax}\mathrm{E}_{\rm C}^{\rm o}\left(\theta,\tau,P_{\rm s,o}\right),\\ P_{\rm s,o}^* &= P_{\rm s,o}\big|_{\tau=\tau^*}.\end{aligned}$$

Fig. 4. Algorithm to derive optimal transmission strategy for overlay mode.

C. *Proposed Hybrid Underlay/Overlay Transmission Mode*

In the previous two sections, we have derived optimal transmission strategies for both underlay and overlay modes. However, the two transmission modes usually show different performances due to the following reasons.

1) PN's spectrum access activities: The PN's spectrum access activities will affect the performance of underlay and overlay modes. Specifically, if the PN has heavy traffic load, which means the PN will access the spectrum with large probability, the underlay mode may outperform the overlay mode as the former one allows the CRN use the spectrum all the time, but the latter one will cause the CRN looses lots of transmission opportunities. On the contrary, if the PN accesses the spectrum with low probability, the overlay mode may achieve better performance than the underlay mode.

2) Spectrum sensing: Whether to use spectrum sensing to determine the PN's spectrum occupation status is the critically important difference between the underlay and overlay modes. However, performing spectrum sensing has its own advantages and disadvantages. On the one hand, while applying the over-

$$P(\mathcal{H}_0) \leq \frac{(1-\overline{P}_d)C_{\rm o}^1\left(\theta,\tau^* P_{\rm s,o}^*\right) - C_{\rm u}^1\left(\theta, P_{\rm s,u}^*\right) + \overline{P}_d}{C_{\rm u}^0\left(\theta, P_{\rm s,u}^*\right) - C_{\rm u}^1\left(\theta, P_{\rm s,u}^*\right) - (1-P_f)C_{\rm o}^0\left(\theta,\tau^* P_{\rm s,o}^*\right) + (1-\overline{P}_d)C_{\rm o}^1\left(\theta,\tau^* P_{\rm s,o}^*\right) - P_f + \overline{P}_d}, \tag{31}$$

lay mode, the CRN can use higher power for transmissions and will be imposed looser interference power constraint than the underlay mode. On the other hand, spectrum sensing will consume a certain amount of time, which results that the available transmission time for the underlay mode is longer than that for the overlay mode.

3) CRN's QoS requirements: The QoS requirement will affect the CRN's performance. In particular, when the QoS requirement is loose, the CRN can transmit its data when the channel conditions are good enough. Thus, the CRN may prefer the overlay mode. However, if the QoS requirement is stringent, the CRN may need to guarantee a minimum transmission rate all the time. Consequently, the underlay mode may outperform the overlay mode. Moreover, the CRN's QoS requirement may affect the duration of the spectrum sensing for the overlay mode.

Consequently, we need to develop a new transmission mode for the CRN to more sufficiently guarantee the CRN's QoS demand and improve its performance. As the underlay and overlay modes have their respective advantages under diverse system parameters, the optimal transmission mode for the CRN should adapt the underlay and overlay modes for DSA such that the CRN can achieve the optimal performance and better QoS provisioning. Denote the optimal effective capacity for the underlay and overlay modes by $\mathrm{E}_{\rm C}^{\rm u,opt}(\theta)$ and $\mathrm{E}_{\rm C}^{\rm o,opt}(\theta)$, respectively. Then, our proposed transmission mode is described as follows.

Our Proposed Hybrid Underlay/Overlay Transmission Mode: Consider a two-dimensional plane, where the x-axis represents the QoS exponent θ for the CRN and the y-axis denotes the spectrum idle probability $P\left(\mathcal{H}_0\right)$. Then, we can divide the plane into two regions, denoted by $A_{\rm u}$ and $A_{\rm o}$, respectively. The point $(\theta, P\left(\mathcal{H}_0\right))$ in $A_{\rm u}$ satisfies $\mathrm{E}_{\rm C}^{\rm u,opt}(\theta) \geq \mathrm{E}_{\rm C}^{\rm o,opt}(\theta)$ and the point $A_{\rm o}$ satisfies $\mathrm{E}_{\rm C}^{\rm u,opt}(\theta) < \mathrm{E}_{\rm C}^{\rm o,opt}(\theta)$. Given the delay bound $D_{\rm th}$ and its violation probability $P_{\rm th}$, the corresponding QoS exponent θ can be calculated by Eq. (6). If $(\theta, P\left(\mathcal{H}_0\right))$ falls into $A_{\rm u}$, the CRN should choose the underlay mode. On the contrary, if $(\theta, P\left(\mathcal{H}_0\right))$ falls into $A_{\rm o}$, the CRN should select the overlay mode.

Mathematically, if $\mathrm{E}_{\rm C}^{\rm u,opt}(\theta) \geq \mathrm{E}_{\rm C}^{\rm o,opt}(\theta)$ holds, we must have that Eq. (31) shown at the top of this page holds, where

$$\begin{cases} C_{\rm u}^0\left(\theta, P_{\rm s,u}^*\right) &= \mathbb{E}_{\mathbf{G}}\left\{\left[1+g_s P_{\rm s,u}^*\right]^{-\beta_{\rm u}}\right\}, \\ C_{\rm u}^1\left(\theta, P_{\rm s,u}^*\right) &= \mathbb{E}_{\mathbf{G}}\left\{\left[1+\frac{g_s P_{\rm s,u}^*}{g_{ps}P_{\rm p}+1}\right]^{-\beta_{\rm u}}\right\}, \\ C_{\rm o}^0\left(\theta, \tau^* P_{\rm s,o}^*\right) &= \mathbb{E}_{\mathbf{G}}\left\{\left[1+g_s P_{\rm s,o}^*\right]^{-\beta_{\rm o}(\tau^*)}\right\}, \\ C_{\rm o}^1\left(\theta, \tau^* P_{\rm s,o}^*\right) &= \mathbb{E}_{\mathbf{G}}\left\{\left[1+\frac{g_s P_{\rm s,o}^*}{p_{ps}P_{\rm p}+1}\right]^{-\beta_{\rm o}(\tau^*)}\right\}, \end{cases} \tag{32}$$

$P_{\rm s,u}^*$ is the optimal transmit power for the underlay mode, τ^* and $P_{\rm s,o}^*$ are the optimal sensing time and transmit power for the overlay mode. Clearly, inequality (31) describes the border line for our proposed hybrid underlay/overlay transmission mode. Specifically, for any given spectrum idle probability $P(\mathcal{H}_0)$ and CRN's QoS requirement θ, if inequality (31) is satisfied, the CRN should access the spectrum with the underlay mode; if inequality (31) is not hold, the CRN will use the overlay mode for the DSA.

Remark:

1) Our proposed hybrid underlay/overlay transmission mode combines the advantages of underlay and overlay modes. Specifically, the CRN adapts the underlay and overlay modes according to the PN's spectrum-occupancy probabilities and the CRN's QoS requirements. Consequently, the CRN can access the spectrum with the optimal mode and thus our proposed transmission mode provides better QoS provisioning for the CRN and efficiently improves the CRN's performance. Moreover, our proposed transmission mode is easy to be implemented in realistic systems.

2) The reasons for choosing PN's spectrum access probability and QoS exponent as two metrics for our proposed transmission mode are listed as follows. First, when the CRN and PN are given, the maximum transmit power of the CRN and the maximum interference power that can be tolerated by the PN are determined. Second, the PUs' traffic load, which affects the PN's spectrum access activities and the CRN's performance, is time-varying and can be well reflected by the PN's spectrum-occupancy probabilities. The heavier traffic load will cause higher spectrum-occupancy probability and vice versa. Third, QoS provisioning is critically important to the wireless communications system and especially for the CRN. Moreover, different applications and services usually have diverse QoS requirements. Consequently, our proposed transmission mode simultaneously takes the PN's spectrum-occupancy probabilities and CRN's QoS requirements into consideration.

IV. Performance Analysis

In this section, we first analyze the optimal transmission strategies for the underlay and overlay modes under two limiting cases, i.e., loose delay constraint and stringent delay constraint, respectively. Then we discuss the impact of imperfect channel estimations on the CRN's performance.

A. Analysis for Loose Delay Constraint

In this section, we prove that for loose delay QoS requirement, the optimal transmission strategies for both underlay and overlay modes will become the well-known water-filling algorithm, which means that the ergodic capacity can be achieved.

For the underlay mode, the optimization problem that maximizes the ergodic capacity can be formulated as

$$\begin{aligned}\max_{P_\mathrm{s}\geq 0} \quad & P(\mathcal{H}_0)\mathbb{E}_\mathbf{G}\left\{\log\left(1+g_{ss}P_\mathrm{s}\right)\right\} \\ & +P(\mathcal{H}_1)\mathbb{E}_\mathbf{G}\left\{\log\left(1+\frac{g_{ss}P_\mathrm{s}}{g_{ps}P_\mathrm{p}+1}\right)\right\} \\ \text{s.t.} \quad & \text{Eqs. (9) and (10).}\end{aligned}$$

By applying the Lagrangian dual method, we can easily obtain the optimal transmit power P_s^* for the above optimization problem, which is determined by

$$P_\mathrm{s}^* = \left(\frac{A+\sqrt{B}}{2}\right)^+, \tag{33}$$

where $A = \frac{1}{\lambda+g_{sp}\mu} - \frac{g_{ps}P_\mathrm{p}+2}{g_{ss}}$, $B = A^2 - \frac{4}{g_{ss}}\left\{\frac{g_{ps}P_\mathrm{p}+1}{g_{ss}} - \frac{P(\mathcal{H}_0)(g_{ps}P_\mathrm{p}+1)+P(\mathcal{H}_1)}{\lambda+g_{sp}\mu}\right\}$, λ and μ are Lagrangian multipliers associated with constraints given by Eqs. (9) and (10), respectively. We can observe from Eq. (33) that P_s^* is the water-filling algorithm, where the water-level is $\frac{1}{2(\lambda+g_{sp}\mu)}$. If we can prove that the optimal transmit power $P_\mathrm{s,u}^*$ for the underlay mode equals P_s^* when the delay constraint is loose, our goal is achieved.

When the delay QoS requirement is loose, we have $\theta \to 0$. For the underlay mode, Eq. (15) can be written as

$$\begin{aligned}&\frac{P(\mathcal{H}_0)g_{ss}}{\left[1+g_{ss}P_\mathrm{s,u}\right]^{\beta_\mathrm{u}+1}} + \frac{P(\mathcal{H}_1)g_{ss}}{(g_{ps}P_\mathrm{p}+1)\left[1+\frac{g_{ss}P_\mathrm{s,u}}{g_{ps}P_\mathrm{p}+1}\right]^{\beta_\mathrm{u}+1}} \\ &= \frac{\lambda_\mathrm{u}}{\beta_\mathrm{u}} + g_{sp}\frac{\mu_\mathrm{u}}{\beta_\mathrm{u}}.\end{aligned} \tag{34}$$

Let $\theta \to 0$, we have $\beta_\mathrm{u} \to 0$. We can find that the left-hand side of Eq. (34) is a positive constant under the given $P(\mathcal{H}_0)$, $P(\mathcal{H}_1)$, g_{ss}, g_{ps}, P_p, and $P_\mathrm{s,u}$. Thus, the right-hand side of Eq. (34) must also be a positive constant. As g_{sp} is the determined constant, $\lambda_\mathrm{u}/\beta_\mathrm{u}$ and $\mu_\mathrm{u}/\beta_\mathrm{u}$ should also be constants even if $\beta_\mathrm{u} \to 0$. Consequently, we can replace $\lambda_\mathrm{u}/\beta_\mathrm{u}$ and $\mu_\mathrm{u}/\beta_\mathrm{u}$ by λ_u and μ_u, respectively, and Eq. (34) becomes

$$\begin{aligned}&P(\mathcal{H}_0)g_{ss}\left[1+g_{ss}P_\mathrm{s,u}\right]^{-1} + \frac{P(\mathcal{H}_1)g_{ss}}{g_{ps}P_\mathrm{p}+1}\left[1+\frac{g_{ss}P_\mathrm{s,u}}{g_{ps}P_\mathrm{p}+1}\right]^{-1} \\ &= \lambda_\mathrm{u} + g_{sp}\mu_\mathrm{u}.\end{aligned} \tag{35}$$

Through solving Eq. (35), the optimal transmit power $P_\mathrm{s,u}^*$ of the underlay mode can be written as

$$\lim_{\theta\to 0} P_\mathrm{s,u}^* = \left(\frac{A_\mathrm{u}+\sqrt{B_\mathrm{u}}}{2}\right)^+, \tag{36}$$

where $A_\mathrm{u} = \frac{1}{\lambda_\mathrm{u}+g_{sp}\mu_\mathrm{u}} - \frac{g_{ps}P_\mathrm{p}+2}{g_{ss}}$ and $B_\mathrm{u} = A_\mathrm{u}^2 - \frac{4}{g_{ss}}\left\{\frac{g_{ps}P_\mathrm{p}+1}{g_{ss}} - \frac{P(\mathcal{H}_0)(g_{ps}P_\mathrm{p}+1)+P(\mathcal{H}_1)}{\lambda_\mathrm{u}+g_{sp}\mu_\mathrm{u}}\right\}$. Comparing Eq. (36) with Eq. (33), we can easily find that when the delay QoS constraint is loose, i.e., $\theta \to 0$, the optimal transmit power $P_\mathrm{s,u}^*$ for the underlay mode is the same with the optimal transmit power P_s^* that aims at maximizing the ergodic capacity of the underlay mode. Consequently, we can conclude that **the optimal transmit power for the underlay mode becomes the water-filling algorithm when the delay QoS requirement is loose and the ergodic capacity is achieved**.

For the overlay mode, let $P_\mathrm{idle} = P(\mathcal{H}_0)(1-P_f)$ and $P_\mathrm{miss} = P(\mathcal{H}_1)(1-\overline{P}_d)$. Then, by employing the similar approach, we can derive the optimal transmit power $P_\mathrm{s,o}^*$ of the overlay mode under loose delay QoS requirement, which is given by

$$\lim_{\theta\to 0} P_\mathrm{s,o}^* = \left(\frac{A_\mathrm{o}+\sqrt{B_\mathrm{o}}}{2}\right)^+, \tag{37}$$

where

$$\begin{cases} A_\mathrm{o} = \frac{T_f(P_\mathrm{idle}+P_\mathrm{miss})}{\lambda_\mathrm{o}(T_f-\tau^*)(P_\mathrm{idle}+P_\mathrm{miss})+\mu_\mathrm{o}g_{sp}(T_f-\tau^*)P_\mathrm{miss}} - \frac{g_{ps}P_\mathrm{p}+2}{g_{ss}} \\ B_\mathrm{o} = A_\mathrm{o}^2 - \frac{4}{g_{ss}}\left\{\frac{g_{ps}P_\mathrm{p}+1}{g_{ss}}\right. \\ \qquad \left. - \frac{T_f[P_\mathrm{idle}(g_{ps}P_\mathrm{p}+1)+P_\mathrm{miss}]}{\lambda_\mathrm{o}(T_f-\tau^*)(P_\mathrm{idle}+P_\mathrm{miss})+\mu_\mathrm{o}g_{sp}(T_f-\tau^*)P_\mathrm{miss}}\right\}. \end{cases}$$

We can also easily prove that **the optimal transmit power for the overlay mode reduces to the water-filling algorithm under the loose QoS requirement and is the same with the transmit power that maximizes the ergodic capacity of the overlay mode**.

B. Analysis for Stringent Delay Constraint

In this section, we prove that for stringent delay QoS requirement, which corresponds to $\theta \to \infty$, the optimal transmission strategies for both underlay and overlay modes converge to the channel inversion scheme under which the SU transmits with a constant rate.

For the underlay mode, different from the analyses in the previous section, we cannot obtain the closed-form solution from Eq. (15) when $\theta \to \infty$. Fortunately, as Eq. (15) can be viewed as the linear function of $P(\mathcal{H}_0)$, if we can prove that SU transmits with constant rates under two limiting cases, which are $P(\mathcal{H}_0) \to 0$ and $P(\mathcal{H}_0) \to 1$, respectively, we can determine that SU also operates at a constant rate under any given $P(\mathcal{H}_0) \in (0,1)$. Then, we analyze the power allocation scheme of the underlay mode under $P(\mathcal{H}_0) \to 0$ and $P(\mathcal{H}_0) \to 1$, respectively.

when $P(\mathcal{H}_0) \to 0$, we have $P(\mathcal{H}_1) \to 1$ and the equality given by Eq. (15) becomes

$$\frac{\beta_\mathrm{u}g_{ss}}{g_{ps}P_\mathrm{p}+1}\left[1+\frac{g_{ss}P_\mathrm{s,u}}{g_{ps}P_\mathrm{p}+1}\right]^{-\beta_\mathrm{u}-1} = \lambda_\mathrm{u}+g_{sp}\mu_\mathrm{u}. \tag{38}$$

As $\log(\cdot)$ is a monotonically increasing function, Eq. (38) is equal to the following equality:

$$\begin{aligned}&\log\left(1+\frac{g_{ss}P_\mathrm{s,u}}{g_{ps}P_\mathrm{p}+1}\right) \\ &= \frac{1}{\beta_\mathrm{u}+1}\left\{\log\left(\frac{\beta_\mathrm{u}g_{ss}}{g_{ps}P_\mathrm{p}+1}\right) - \log\left(\lambda_\mathrm{u}+g_{sp}\mu_\mathrm{u}\right)\right\}.\end{aligned} \tag{39}$$

As the left-hand side of Eq. (39) is a positive constant, the right-hand side of Eq. (39) is also a positive constant when $\beta \to \infty$. By performing some mathematical transformations, we can obtain

$$1+\frac{g_{ss}P_\mathrm{s,u}}{g_{ps}P_\mathrm{p}+1} = \Gamma, \tag{40}$$

where $\Gamma = \exp\left\{\frac{1}{\beta_\mathrm{u}+1}\left[\log\left(\frac{\beta_\mathrm{u}g_{ss}}{g_{ps}P_\mathrm{p}+1}\right) - \log\left(\lambda_\mathrm{u}+g_{sp}\mu_\mathrm{u}\right)\right]\right\}$ and needs to satisfy constraints (9) and (10) simultaneously.

Consequently, the optimal transmit power is

$$\lim_{\substack{\theta\to\infty \\ P(\mathcal{H}_0)\to 0}} P^*_{\mathrm{s,u}} = \frac{(m-1)(g_{ps}P_{\mathrm{p}}+1)}{mg_{ss}(P_{\mathrm{p}}+1)}\min\{P_{\mathrm{av}}, Q_{\mathrm{av}}\} \quad (41)$$

if $m \geq 1$. By applying similar approach, the optimal transmit power for the underlay mode when $\theta \to \infty$ and $P(\mathcal{H}_0) \to 1$ can be written as

$$\lim_{\substack{\theta\to\infty \\ P(\mathcal{H}_0)\to 1}} P^*_{\mathrm{s,u}} = \frac{m-1}{mg_{ss}}\min\{P_{\mathrm{av}}, Q_{\mathrm{av}}\}. \quad (42)$$

We can clearly observe that, **when the underlay based CRN needs to meet stringent QoS requirement, the optimal transmit power converges to the channel inversion under two limiting cases**, i.e., $P(\mathcal{H}_0) \to 0$ $P(\mathcal{H}_0) \to 1$, respectively. That means the CRN will transmit with constant rates under these two limiting cases. We denote the constant transmission rates under the two limiting cases as R^0_{u} and R^1_{u}, respectively. As Eq. (15) is the linear function of $P(\mathcal{H}_0)$, we can obtain that, if $0 < P(\mathcal{H}_0) < 1$, the CRN still transmits with a constant rate that belongs to $\left(R^0_{\mathrm{u}}, R^1_{\mathrm{u}}\right)$. In this case, **the optimal transmission strategy for the underlay mode converges to the total channel inversion scheme and the zero-outage capacity is achieved**. Note that, if the Nakagami fading parameter $m < 1$, which implies that the fading is severer than Rayleigh, then total channel inversion scheme does not exist as the transmit power is not enough to totally invert the channel. In this case, **the optimal transmission strategy for the underlay mode becomes the truncated channel inversion scheme and the outage capacity is achieved**.

For the overlay mode, by employing similar method, the optimal transmit power $P^*_{\mathrm{s,o}}$ can be written as

$$\begin{cases} \lim\limits_{\substack{\theta\to\infty \\ P(\mathcal{H}_0)\to 0}} P^*_{\mathrm{s,o}} = \frac{(m-1)(g_{ps}P_{\mathrm{p}}+1)T_f}{mg_{ss}(P_{\mathrm{p}}+1)(T_f-\tau)(1-\overline{P}_d)}\min\{P_{\mathrm{av}}, Q_{\mathrm{av}}\}, \\ \lim\limits_{\substack{\theta\to\infty \\ P(\mathcal{H}_0)\to 1}} P^*_{\mathrm{s,o}} = \frac{(m-1)T_f}{mg_{ss}(T_f-\tau)(1-P_f)}P_{\mathrm{av}}. \end{cases} \quad (43)$$

Consequently, we can also derive the same conclusion for the overlay mode, i.e., **the optimal transmit power for the overlay based CRN converges to the channel inversion and the SU transmits with a constant rate when the QoS requirement is stringent**.

C. Impact of Imperfect Channel Estimations

In this section, we evaluate the impact of imperfect channel estimations on the performance of the CRN. We assume that the channel gain between SS and SR can be perfectly estimated by the SR and reliably fed back to the SS, but the SS can only derive imperfect channel gain between the SS and PR.

Let h_{sp} and $\hat{h}_{sp}$ be the true and estimated channel coefficients for the interference link, i.e., the link between the SS and PR, respectively. We assume that the SS performs the minimum mean square error (MMSE) estimation of h_{sp}. Then, h_{sp} can be written as $h_{sp} = \hat{h}_{sp} + \tilde{h}_{sp}$, where $\tilde{h}_{sp}$ is the channel estimation error of the interference link. We can easily find that $\hat{h}_{sp}$ and $\tilde{h}_{sp}$ are uncorrelated. Moreover, the estimation error $\tilde{h}_{sp}$ is the zero-mean random variable with variance σ_e^2. Then, the true and estimated channel power gains between the SS and PR can be represented by $g_{sp} = |h_{sp}|^2$ and $\hat{g}_{sp} = \left|\hat{h}_{sp}\right|^2$, respectively, and the channel power gain estimation error of the interference link can be written as $\tilde{g}_{sp} = \left|\tilde{h}_{sp}\right|^2$. We modify the network gain vector (NGV) as $\hat{\mathbf{G}} \triangleq \{g_{ss}, g_{ps}, \hat{g}_{sp}\}$. Then, the transmit power of the CRN for the underlay and overlay modes can be written as $P_{\mathrm{s,u}}\left(\theta, \hat{\mathbf{G}}\right)$ and $P_{\mathrm{s,o}}\left(\theta, \hat{\mathbf{G}}\right)$, respectively. For the underlay mode, the average interference power constraint given by Eq. (10) becomes

$$\begin{aligned} Q_{\mathrm{av}} &\geq \mathbb{E}_{\hat{\mathbf{G}},\tilde{g}_{sp}}\left\{g_{sp}P_{\mathrm{s,u}}\left(\theta,\hat{\mathbf{G}}\right)\right\} \\ &= \mathbb{E}_{\hat{\mathbf{G}},\tilde{g}_{sp}}\left\{\left(\hat{g}_{sp}+\tilde{g}_{sp}\right)P_{\mathrm{s,u}}\left(\theta,\hat{\mathbf{G}}\right)\right\} \\ &= \mathbb{E}_{\hat{\mathbf{G}}}\left\{\hat{g}_{sp}P_{\mathrm{s,u}}\left(\theta,\hat{\mathbf{G}}\right)\right\} + \mathbb{E}_{\hat{\mathbf{G}},\tilde{g}_{sp}}\left\{\tilde{g}_{sp}P_{\mathrm{s,u}}\left(\theta,\hat{\mathbf{G}}\right)\right\} \\ &= \mathbb{E}_{\hat{\mathbf{G}}}\left\{\hat{g}_{sp}P_{\mathrm{s,u}}\left(\theta,\hat{\mathbf{G}}\right)\right\} + \sigma_e^2 P_{\mathrm{av}}. \end{aligned} \quad (44)$$

Similarly, the average interference power constraint for the overlay mode given by Eq. (24) can be modified as

$$\begin{aligned} &\frac{T_f-\tau}{T_f}\left\{P(\mathcal{H}_1)(1-P_d)\mathbb{E}_{\hat{\mathbf{G}}}\left[\hat{g}_{sp}P_{\mathrm{s,o}}\left(\theta,\hat{\mathbf{G}}\right)\right]\right\} \\ &\leq Q_{\mathrm{av}} - \frac{T_f-\tau}{T_f}P(\mathcal{H}_1)(1-P_d)\sigma_e^2 P_{\mathrm{av}}. \end{aligned} \quad (45)$$

Based on the above analyses, we find that, 1) in order to protect the PN's QoS, the imperfect channel estimation causes the CRN to satisfy more stringent interference power constraint, which results that the CRN's performance will be correspondingly decreased; and 2) as $\frac{T_f-\tau}{T_f}P(\mathcal{H}_1)(1-P_d) < 1$, the impact of the imperfect channel estimation on the performance of the overlay mode is less than that of the underlay mode. Consequently, the imperfect channel estimation will also degrade the performance of our proposed hybrid underlay/overlay transmission mode, but the CRN has more opportunities to employ the overlay mode for DSA under our proposed transmission mode.

V. Simulation Evaluations and Discussions

In this section, we evaluate the optimal transmission strategies for underlay and overlay modes and illustrate the effectiveness of our proposed hybrid underlay/overlay transmission mode by simulations. In our simulations, we set the frame duration $T_f = 50$ ms, the sampling frequency $f_s = 10^6$ Hz, the target-detection probability $\overline{P}_d = 0.9$, and the fading parameter of the Nakagami-m channel model $m = 2$. The other system parameters are detailed respectively in each of the figures.

A. Optimal Effective Capacity for Underlay, Overlay, and Our Proposed Transmission Modes

Figure 5 presents the normalized effective capacity of the underlay mode as a function of the QoS exponent θ and the spectrum-idle probability $P(\mathcal{H}_0)$. We can observe from Fig. 5 that the QoS exponent θ, which reflects the statistical

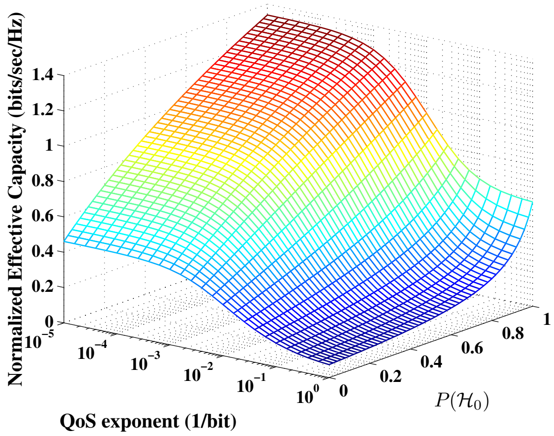


Fig. 5. Normalized effective capacity of the underlay mode under the optimal power allocation scheme. $P_{\rm av} = 10$ dB, $Q_{\rm av} = 0$ dB, and $P_{\rm p} = 10$ dB.

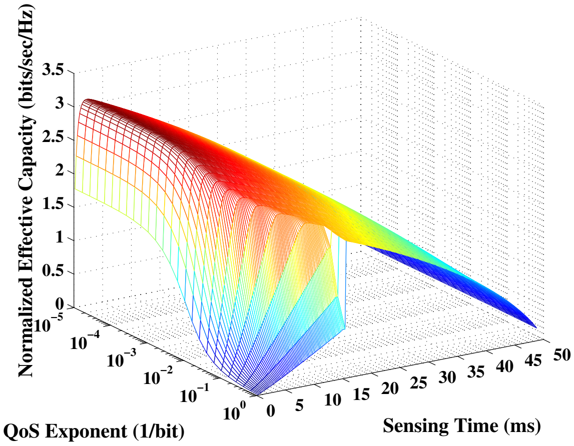
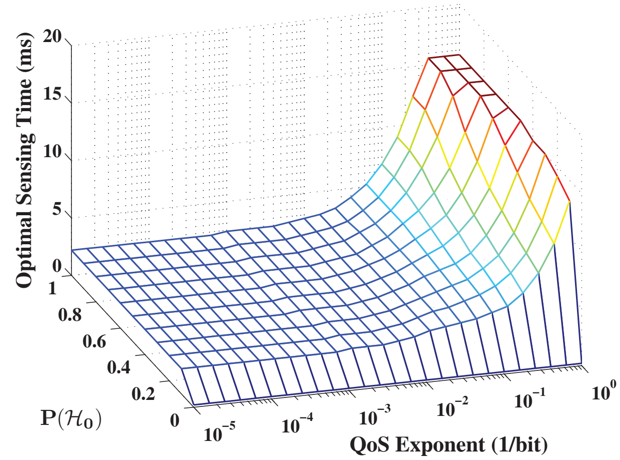
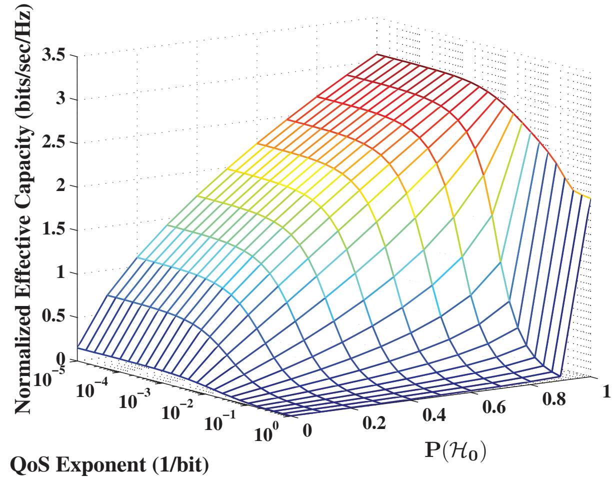


Fig. 6. Normalized effective capacity of the overlay mode as the function of the QoS exponent θ and sensing time τ. $P_{\rm av} = 10$ dB, $Q_{\rm av} = 0$ dB, $P_{\rm p} = 10$ dB, and $P(\mathcal{H}_0) = 1$.



(a) Optimal spectrum sensing time



(b) Normalized effective capacity

Fig. 7. Optimal spectrum sensing time and normalized effective capacity of the overlay mode under the optimal resource allocation scheme. $P_{\rm av} = 10$ dB, $Q_{\rm av} = 0$ dB, and $P_{\rm p} = 10$ dB.

delay QoS requirement, plays a critically important role in the CRN's maximum available throughput. When θ is small, i.e., the QoS constraint is loose, the CRN can realize higher throughput. On the contrary, when θ is large, i.e., the QoS constraint is stringent, the CRN can only support lower arrival rate. This phenomenon can be explained as follows. For small values of θ, the optimal power allocation scheme becomes the water-filling algorithm as discussed in Section IV-A, thus the effective capacity becomes the Shannon capacity and the upper-bounded power resource can be efficiently utilized. For large values of θ, the optimal power allocation scheme converges to the total channel inversion as analyzed in Section IV-B, which results that the effective capacity becomes the zero-outage capacity. Consequently, the power resource is used in a less efficient way since more power will be allocated to the worse channel state. We can also observe from Fig. 5 that the effective capacity of the underlay based CRN is an increasing function of the probability $P(\mathcal{H}_0)$ for any given QoS exponent θ. This is reasonable because that lager $P(\mathcal{H}_0)$ represents higher probability that the spectrum is idle and thus the CRN has more opportunities to transmit with higher rates.

Different from the underlay mode, the CRN with the overlay mode needs to perform spectrum sensing to identify the PN's spectrum-occupancy status and accesses the spectrum only when it is not currently occupied by the PN. Fig. 6 shows the effective capacity of the overlay mode as a function of the QoS exponent θ and sensing time τ. We can observe from Fig. 6 that the optimal sensing time τ^*, which can maximize the CRN's effective capacity, exists under the given QoS requirement. Clearly, different sensing time will result in different performance. Therefore, in order to maximize the effective capacity of overlay based CRN, we need to determine the optimal sensing time. Fig. 6 also shows that as sensing time τ increases, the increase speed of effective capacity in the region of $\tau < \tau^*$ is much faster than the decrease speed of effective capacity in the region of $\tau > \tau^*$. It implies that the impact of underestimation on the performance of the overlay mode is more serious than that of overestimation. The reason for such observation can be explained as follows. For the given target-detection probability $\overline{P}_d$, as sensing time decreases, available transmission time in each frame for the CRN increases, but the probability of false alarm will dramatically increases especially for the short sensing time, which will cause the CRN to loose more transmission opportunities. On the contrary, when the sensing time is long, the probability of false alarm is sufficiently small, Thus, increasing sensing time will only reduce the available transmission time in each frame. Consequently, the underestimation is more fatal than the overestimation.

Fig. 7(a) shows the optimal sensing time of the overlay mode as a function of the QoS exponent θ and the spectrum-idle probability $P(\mathcal{H}_0)$. We can observe from Fig. 7(a) that the optimal sensing time increases as θ increases, i.e., the more stringent QoS constraint results in longer sensing time. Because the power allocation scheme for the overlay mode has the same property as that for the underlay mode, i.e., it will gradually converge to the channel inversion as θ increases, the CRN will transmit with a constant rate. Consequently, the CRN should reduce the probability of false alarm, which causes longer sensing time. Moreover, Fig. 7(a) shows that the optimal sensing time is looked like to be zero when $P(\mathcal{H}_0) = 0$. However, actually, the optimal sensing time for $P(\mathcal{H}_0) = 0$ is very short but is not equal to zero. This is because that when $P(\mathcal{H}_0) = 0$, the effective capacity of the overlay mode will not be affected by the probability of false alarm P_f. Consequently, SS can adjust the detection threshold such that the target-detection probability $\overline{P}_d$ can be achieved by consuming a very short time duration. Based on the derived optimal sensing time, Fig. 7(b) presents the normalized effective capacity of the overlay mode under the optimal transmission strategy. Similar with the underlay mode, we can also observe from Fig. 7(b) that the effective capacity of the overlay mode is a decreasing function of the QoS exponent θ and is an increasing function of the probability $P(\mathcal{H}_0)$.

As shown in Fig. 5 and 7(b), the effective capacity of the underlay and overlay modes have their respective advantages under different QoS requirements and PUs' spectrum-occupancy probabilities, which verifies that the optimal transmission mode for the CRN should dynamically adapt to the underlay and overlay modes according to different system parameters and environments. In our proposed hybrid underlay/overlay transmission mode, the most important issue is to determine the regions A_{u} and A_{o} in the two-dimensional plane, where the x-axis represents the QoS exponent θ and the y-axis denotes the spectrum-idle probability $P(\mathcal{H}_0)$. Based on Fig. 5 and 7(b), Fig. 8 (a) shows the region division. In this figure, our obtained border line divides the plane into two regions. The gray shaded region is A_{u}, where the point $(\theta, P(\mathcal{H}_0))$ satisfies $\mathrm{E}_{\mathrm{C}}^{\mathrm{u,opt}}(\theta) \geq \mathrm{E}_{\mathrm{C}}^{\mathrm{o,opt}}(\theta)$, and the white region is A_{o}, where the point $(\theta, P(\mathcal{H}_0))$ satisfies $\mathrm{E}_{\mathrm{C}}^{\mathrm{u,opt}}(\theta) < \mathrm{E}_{\mathrm{C}}^{\mathrm{o,opt}}(\theta)$. Based on Fig. 8 (a), the hybrid underlay/overlay transmission mode proposed in Section III-C can be performed. If the point $(\theta, P(\mathcal{H}_0))$ falls into the gray shaded region, i.e., A_{u}, the CRN should choose the underlay mode. On the contrary, the CRN should choose the overlay mode if $(\theta, P(\mathcal{H}_0))$ falls into the white region, i.e., A_{o}. Fig. 8 (b) presents the border lines under different average interference power constraints. We can observe from Fig. 8 (b) that the area of the region A_{u} increases when the average interference power constraint Q_{av} becomes larger.

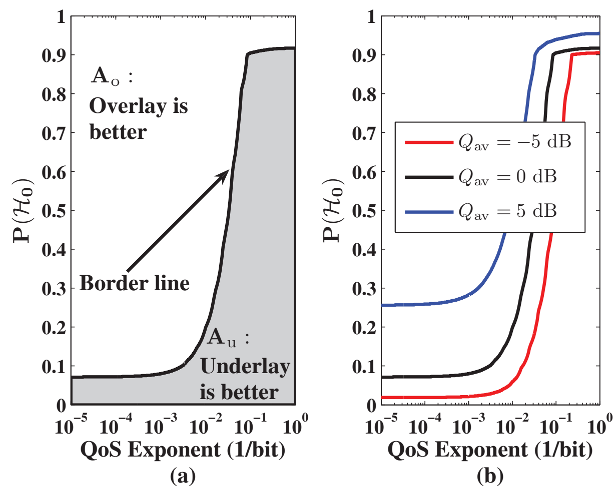
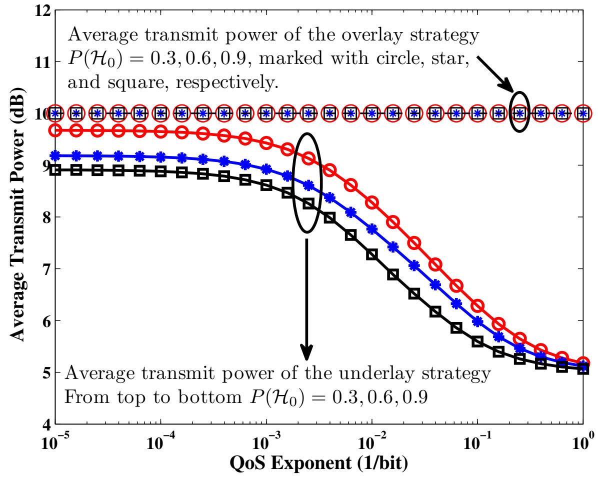


Fig. 8. Border line for our proposed hybrid underlay/overlay transmission mode versus the QoS exponent θ and the spectrum-idle probability $P(\mathcal{H}_0)$. $P_{\mathrm{av}} = 10$ dB, $Q_{\mathrm{av}} = 0$ dB, and $P_{\mathrm{p}} = 10$ dB.

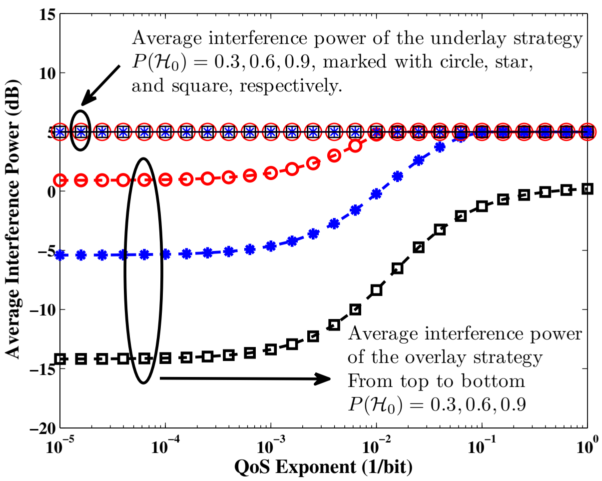
B. Average Transmit and Interference Power for Underlay and Overlay Modes

Due to distinct features of the underlay and overlay modes, the utilized transmit power of the CRN and caused interference power to the PN of these two modes are also different. In this section, we investigate how the average transmit and interference power of underlay and overlay modes vary with the spectrum-idle probability $P(\mathcal{H}_0)$ and QoS exponent θ.

Figure 9(a) shows the utilized average transmit power of the CRN versus the QoS exponent θ under different spectrum-idle probabilities $P(\mathcal{H}_0)$. We can observe from Fig. 9(a) that the utilized average transmit power of the overlay mode remains constant for all θ and do not vary with PN's spectrum-occupancy probability, which means that the average transmit power constraint given by Eq. (23) is always active and the upper-bounded power resource can be fully utilized. On the contrary, the average transmit power of the underlay mode varies with PN's spectrum-occupancy probability. Specifically, the larger the spectrum-idle probability $P(\mathcal{H}_0)$ is, the less the utilized transmit power is. However, the underlay mode can achieve better performance as increasing $P(\mathcal{H}_0)$. Consequently, as the spectrum-idle probability increases, the underlay mode can derive better performance but consume less transmit power, which denotes that the transmit power resource is more efficiently utilized. Moreover, we find that the utilized average transmit power of the underlay mode converges to the average interference power constraint ($Q_{\mathrm{av}} = 5$ dB). The reason of this observation can be explained as follows. When the value of θ is small, the delay QoS constraint is loose, which means that the CRN can tolerate longer delay. In this case, more transmit power will be allocated for transmission when the channel gain g_{ss} between SS and SR is large but the inter-system gains in terms of g_{ps} and g_{sp} are small. Thus, for small θ, the utilized transmit power of the underlay mode will not obviously affected by the predefined average interference power constraint. On the contrary, when the value of θ is large, the CRN needs to satisfy more stringent delay QoS requirement, which causes that a constant



(a) Average transmit power



(b) Average interference power

Fig. 9. Average transmit and interference power for both underlay and overlay modes under the optimal resource allocation schemes. $P_{\mathrm{av}} = 10$ dB, $Q_{\mathrm{av}} = 5$ dB, and $P_{\mathrm{p}} = 10$ dB.

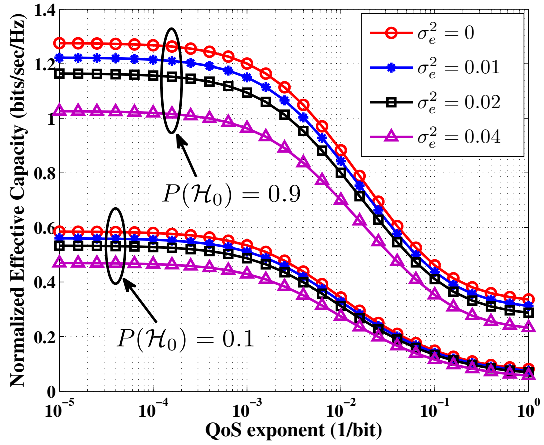
transmission rate needs to be maintained [21]. In this case, in order to meet the interference power requirement, SU has to reduce its transmit power to guarantee that a constant transmission rate can be achieved even though the channel gain g_{sp} of the interference link between SS and PR is large. Consequently, for large θ, the utilized transmit power of the underlay mode will significantly affected by the maximum average interference power Q_{av} and the underlay based CRN will gradually converge to an interference-power constrained system as the QoS constraint becomes more stringent.

Figure 9(b) shows the average interference power of both underlay and overlay modes versus the QoS exponent θ under different spectrum-idle probabilities $P(\mathcal{H}_0)$. We can observe from Fig. 9(b) that the average interference power of both underlay and overlay modes satisfy the predefined maximum interference power, which means that the interference power of both modes are no larger than Q_{av}, but they show distinct features. Specifically, for the underlay mode, the average interference power constraint given by Eq. (10) is always active under different QoS exponents and PN's spectrum-occupancy probabilities. However, for the overlay mode, the average interference power varies with θ and $P(\mathcal{H}_0)$ and our imposed average interference power constraint is not always active. In particular, the average interference power of the overlay mode increases as the QoS exponent θ becomes large and decreases as the spectrum-idle probability $P(\mathcal{H}_0)$ increases. Such phenomenons can be explained as follows. First, when θ becomes larger, the power allocation scheme converts to the channel inversion. In this case, the SS will transmit with constant rate and more transmit power will be allocated to the worse channel conditions. Due to the imperfect spectrum sensing, the SS will use larger power for transmission when the miss-detection happens and thus will cause large interference power. Second, when the spectrum-idle probability $P(\mathcal{H}_0)$ increases, the actual probability of miss-detection $(1 - P(\mathcal{H}_0))\left(1 - \overline{P}_d\right)$ will decrease under the given target-detection probability $\overline{P}_d$ and thus the inference power caused by the overlay mode will be reduced. Moreover, we can find from Fig. 9(a) and 9(b) that, 1) when $P(\mathcal{H}_0)$ is large, the average interference power constraint of the overlay mode always cannot be satisfied and thus the overlay based CRN is a transmit-power constrained system; and 2) when $P(\mathcal{H}_0)$ is small, the characteristic of the overlay based CRN varies with QoS exponent θ. In particular, only the average transmit power constraint is active, i.e., the overlay based CRN is a transmit-power constrained system, for small θ, and both average transmit and interference power constraints are active for large θ.

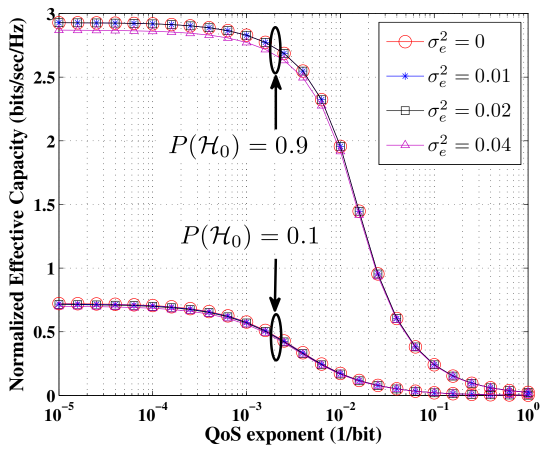
C. Impact of Imperfect Channel Estimations

In this section, we evaluate the impact of channel estimation errors on the performance of underlay and overlay modes as well as our proposed optimal transmission mode. Fig. 10(a) shows the normalized effective capacity of the underlay mode versus the QoS exponent θ under different values of channel estimation error variance σ_e^2. We can observe from Fig. 10(a) that the available maximum throughput of the underlay mode degrades as the channel estimation error variance increases. Fig. 10(b) plots the normalized effective capacity of the overlay mode versus the QoS exponent θ under different values of channel estimation error variance σ_e^2. Similar phenomenon can be observed from Fig. 10(b), i.e., channel estimation error causes the throughput loss and larger σ_e^2 results in more obvious performance degradation. However, under the given channel estimation error variance, the performance loss of the overlay mode is less than that of the underlay mode, which verifies the analysis discussed in Section IV-C.

In order to evaluate the impact of imperfect channel estimations on our proposed hybrid underlay/overlay transmission mode, Fig. 11 shows the border line of our proposed transmission mode under different values of channel estimation error variance σ_e^2. We can observe from Fig. 11 that the region division varies with different values of σ_e^2. Specifically, as the impact of imperfect channel estimations on the underlay mode is more serious that that on the overlay mode, the border line moves down as σ_e^2 increases. Such observation means that the CRN has more opportunities to employ the overlay mode for transmission as the overlay mode is more robust than the underlay mode if the channel estimation error exists.



(a) Underlay mode



(b) Overlay mode

Fig. 10. Normalized effective capacity of both underlay and overlay modes under the corresponding optimal transmission strategies with imperfect channel estimations. $P_{\rm av} = 10$ dB, $Q_{\rm av} = 0$ dB, and $P_{\rm p} = 10$ dB.

VI. Conclusions

In this paper, we developed a statistical delay QoS guarantees based hybrid underlay/overlay transmission mode for the CRN. Specifically, by applying the theory of effective capacity and taking the PNs' activity statistics into consideration, we first analyzed the maximum available throughput of the CRN with underlay and overlay modes and provided efficient algorithms to derive the corresponding optimal transmission strategies. Then, we proposed a novel transmission mode name hybrid underlay/overlay mode for the statistical QoS guaranteed CRN, through which the CRN's QoS requirement can be more sufficiently guaranteed and its throughput can be further improved. Moreover, we analyzed the optimal transmission strategies for both underlay and overlay modes under two limiting cases, which are that the QoS requirement is loose and stringent, respectively. The impact of imperfect channel estimations on the CRN's performance was also discussed. Simulation results were provided to demonstrate the impacts of delay QoS requirements and PN's activity statics on maximizing the delay-constrained throughput for both underlay and overlay modes and verify the effectiveness of our proposed optimal transmission mode.

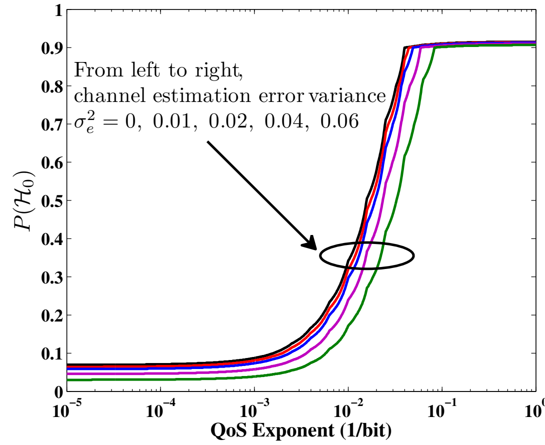


Fig. 11. Border line for our proposed transmission mode with imperfect channel estimations versus the QoS exponent θ and the spectrum-idle probability $P(\mathcal{H}_0)$. $P_{\rm av} = 10$ dB, $Q_{\rm av} = 0$ dB, and $P_{\rm p} = 10$ dB.

Appendix A
Proof of Theorem 1

Proof: Define the diagonal matrix $\mathbf{\Phi}(\theta) \triangleq \mathrm{diag}\Big\{\mathbb{E}[e^{-\theta R_{\rm u}^{(0)}}], \mathbb{E}[e^{-\theta R_{\rm u}^{(1)}}]\Big\}$, then the effective capacity $\mathrm{E}_{\rm C}^{\rm u}(\theta)$ can be written as [19]

$$\mathrm{E}_{\rm C}^{\rm u}(\theta) = -\frac{1}{\theta}\log\Big(\rho\big\{\mathbf{P_u}\mathbf{\Phi}(\theta)\big\}\Big), \tag{46}$$

where $\mathbf{P_u}$ denotes the transition probability matrix of the underlay based CRN and $\rho\{\cdot\}$ is the spectral radius of the matrix. As the CRN has two system states, the transition probability matrix $\mathbf{P_u} = \big[p_{\rm u}^{ij}\big]$ $(i,j = 0,1)$ can be constructed, where $p_{\rm u}^{ij}$ $(i = 0,1; j = 0,1)$ is the transition probability from state i to state j. As the PN independently selects whether to access the spectrum or not at the beginning of each frame, the two system states are time-uncorrelated and thus we have $p_{\rm u}^{00} = p_{\rm u}^{10} = P(\mathcal{H}_0)$ and $p_{\rm u}^{01} = p_{\rm u}^{11} = P(\mathcal{H}_1)$. Then, we can clearly find that $\mathbf{P_u}$ is a matrix with unit rank. Consequently, the spectral radius of matrix $\mathbf{P_u}\mathbf{\Phi}(\theta)$ can be easily calculated by

$$\begin{aligned}\rho\big\{\mathbf{P_u}\mathbf{\Phi}(\theta)\big\} &= \mathrm{trace}\left(\mathbf{P_u}\mathbf{\Phi}\right)\\ &= P(\mathcal{H}_0)\mathbb{E}\left\{e^{-\theta R_{\rm u}^{(0)}}\right\} + P(\mathcal{H}_1)\mathbb{E}\left\{e^{-\theta R_{\rm u}^{(1)}}\right\}.\end{aligned} \tag{47}$$

Substituting the above equation into Eq. (46), we can obtain the effective capacity of the CRN with the underlay mode as shown in Eq. (8). ■

Appendix B
Proof of Optimal Detection Probability

Proof: We can clearly find from Eq. (19) that the probabilities of false alarm and detection are both functions of detection threshold ε and sensing time τ. Let

$$\begin{aligned}G(\varepsilon,\tau,P_{\rm s,o}) = {}& P(\mathcal{H}_0)(1-P_f)C_0(\tau,P_{\rm s,o}) + P(\mathcal{H}_0)P_f\\ &+ P(\mathcal{H}_1)(1-P_d)C_1(\tau,P_{\rm s,o}) + P(\mathcal{H}_1)P_d,\end{aligned} \tag{48}$$

where $C_0(\tau,P_{\rm s,o}) = [1+g_{ss}P_{\rm s,o}]^{-\beta_{\rm o}(\tau)}$ and $C_1(\tau,P_{\rm s,o}) = \Big[1+\frac{g_{ss}P_{\rm s,o}}{p_{ps}P_{\rm p}+1}\Big]^{-\beta_{\rm o}(\tau)}$. Then, $\mathbb{E}_{\mathbf{G}}\left\{G(\varepsilon,\tau,P_{\rm s,o})\right\}$ denotes the objective function of problem (P5) given by Eq. (26). According to [17], for any given sensing time τ, we can choose a detection threshold, denoted by ε_0, such that $P_d|_{\varepsilon=\varepsilon_0} = \overline{P}_d$. For any other ε_1 that satisfies $P_d|_{\varepsilon=\varepsilon_1} > \overline{P}_d$, we must have $\varepsilon_1 < \varepsilon_0$. Then, $P_f|_{\varepsilon=\varepsilon_1} > P_f|_{\varepsilon=\varepsilon_0}$ must hold since P_f is the decreasing function of ε. Consequently, the following inequality must hold:

$$\begin{aligned} & G(\varepsilon_0,\tau,P_{\mathrm{s,o}}) - G(\varepsilon_1,\tau,P_{\mathrm{s,o}}) \\ = \; & P(\mathcal{H}_0)(P_f|_{\varepsilon=\varepsilon_1} - P_f|_{\varepsilon=\varepsilon_0})(C_0(\tau,P_{\mathrm{s,o}})-1) \\ & + P(\mathcal{H}_1)(P_d|_{\varepsilon=\varepsilon_1} - \overline{P}_d)(C_1(\tau,P_{\mathrm{s,o}})-1) \le 0 \end{aligned} \tag{49}$$

due to the fact that $C_0(\tau,P_{\mathrm{s,o}}) \le 1$ and $C_1(\tau,P_{\mathrm{s,o}}) \le 1$ if $P_{\mathrm{s,o}} \ge 0$ and $0 \le \tau \le T_f$. Then, we have $\mathbb{E}_{\mathbf{G}}\left\{G(\varepsilon_0,\tau,P_{\mathrm{s,o}})\right\} \le \mathbb{E}_{\mathbf{G}}\left\{G(\varepsilon_1,\tau,P_{\mathrm{s,o}})\right\}$. Therefore, the optimal solution to problem (P5) can be achieved by $P_d = \overline{P}_d$. We rewrite $G(\varepsilon,\tau,P_{\mathrm{s,o}})$ as $G(\tau,P_{\mathrm{s,o}})$. ■

Appendix C
Proof of Theorem 4

Proof: Let $\alpha = \sqrt{2\gamma+1}\mathcal{Q}^{-1}\left(\overline{P}_d\right)$, then we have $P_f = \mathcal{Q}\left(\alpha + \sqrt{\tau f_s}\gamma\right)$. We can easily obtain that

$$\begin{aligned} & \frac{\partial G(\tau,P_{\mathrm{s,o}})}{\partial \tau} \\ & = P(\mathcal{H}_0)\Big(1-\mathcal{Q}\Big(\alpha+\sqrt{\tau f_s}\gamma\Big)\Big)\frac{\theta B \log\left(1+g_{ss}P_{\mathrm{s,o}}\right)}{\log 2}C_0(\tau,P_{\mathrm{s,o}}) \\ & + P(\mathcal{H}_1)(1-\overline{P}_d)\frac{\theta B}{\log 2}\log\left(1+\frac{g_{ss}P_{\mathrm{s,o}}}{g_{ps}P_{\mathrm{p}}+1}\right)C_1(\tau,P_{\mathrm{s,o}}) \\ & + P(\mathcal{H}_0)\frac{\gamma\sqrt{f_s}}{2\sqrt{2\pi\tau}}\exp\left\{-\frac{\left(\alpha+\sqrt{\tau f_s}\gamma\right)^2}{2}\right\}\Big(C_0(\tau,P_{\mathrm{s,o}})-1\Big). \end{aligned}$$

On the one hand, when $\tau \to 0$, we have

$$\lim_{\tau\to 0}\frac{\partial G(\tau,P_{\mathrm{s,o}})}{\partial \tau} = -\infty. \tag{50}$$

On the other hand, when $\tau \to T_f$, the following inequality must hold:

$$\begin{aligned} & \lim_{\tau\to T_f}\frac{\partial G(\tau,P_{\mathrm{s,o}})}{\partial \tau} \\ & = P(\mathcal{H}_0)\Big(1-\mathcal{Q}\left(\alpha+\sqrt{T_f f_s}\gamma\right)\Big)\frac{\theta B}{\log 2}\log\big(1+g_{ss}P_{\mathrm{s,o}}\big) \\ & + P(\mathcal{H}_1)(1-\overline{P}_d)\frac{\theta B}{\log 2}\log\left(1+\frac{g_{ss}P_{\mathrm{s,o}}}{g_{ps}P_{\mathrm{p}}+1}\right) > 0. \end{aligned} \tag{51}$$

Eqs. (50) and (51) indicate that $G(\tau,P_{\mathrm{s,o}})$ decreases when τ is small and increases when τ goes to T. Therefore, there exists the optimal sensing time that can minimize the objective function $\mathbb{E}\big\{G(\tau,P_{\mathrm{s,o}})\big\}$ within the interval $(0,T)$.

According to [17], when $P_f \le 0.5$, we have $\alpha+\sqrt{\tau f_s}\gamma \ge 0$. Moreover, we can prove that $\frac{\partial P_f}{\partial \tau} < 0$ and $\frac{\partial^2 P_f}{\partial \tau^2} > 0$ must be hold. Then, we have

$$\begin{aligned} & \frac{\partial^2 G(\tau,P_{\mathrm{s,o}})}{\partial \tau^2} \\ & = P(\mathcal{H}_0)(1-P_f)\left[\frac{\theta B}{\log 2}\log\left(1+g_{ss}P_{\mathrm{s,o}}\right)\right]^2 C_0(\tau,P_{\mathrm{s,o}}) \\ & -2P(\mathcal{H}_0)\frac{\partial P_f}{\partial \tau}\frac{\theta B}{\log 2}\log\left(1+g_{ss}P_{\mathrm{s,o}}\right)C_0(\tau,P_{\mathrm{s,o}}) \\ & + P(\mathcal{H}_1)(1-\overline{P}_d)\left[\frac{\theta B}{\log 2}\log\left(1+\frac{g_{ss}P_{\mathrm{s,o}}}{g_{ps}P_{\mathrm{p}}+1}\right)\right]^2 C_1(\tau,P_{\mathrm{s,o}}) \\ & - P(\mathcal{H}_0)\frac{\partial^2 P_f}{\partial \tau^2}\Big(C_0(\tau,P_{\mathrm{s,o}})-1\Big) > 0, \end{aligned}$$

which means that $G(\tau,P_{\mathrm{s,o}})$ is convex over τ. As the expectation operation will not affect the convexity, the objective function $\mathbb{E}\big\{G(\tau,P_{\mathrm{s,o}})\big\}$ is also convex over τ. ■

References

[1] FCC, "Spectrum policy task force report," Nov. 2002. Available: http://www.fcc.gov/sptf/files/SEWGFinalReport-1.pdf

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[3] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 589–600, Mar. 2007.

[4] M. G. Khoshkholgh, K. Navaie, and H. Yanikomeroglu, "Access strategies for spectrum sharing fading environment: overlay, underlay, and mixed," *IEEE Trans. Mobile Comput.*, vol. 9, no. 12, pp. 1780–1793, Dec. 2010.

[5] D. Xu, E. Jung, and X. Liu, "Efficient and fair bandwidth allocation in multichannel cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 11, no. 8, pp. 1372–1385, Aug. 2012.

[6] H. J. Lim, D. Y. Seol, and G. H. Im, "Joint sensing adaption and resource allocation for cognitive radio with imperfect sensing," *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 1091–1100, Apr. 2012.

[7] T. Nadkar, V. Thumar, G. P. S. Tej, S. N Merchant, and U. B. Desai, "Distributed power allocation for secondary users in a cognitive radio scenario," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1576–1586, Apr. 2012.

[8] K. Son, B. C. Jung, and S. Chong, "Power allocation policies with full and partial inter-system channel state information for cognitive radio networks," *Springer Wireless Netw.*, vol. 19, no. 1, pp. 99–113, Jan. 2013.

[9] L. Musavian and S. Aissa, "Capacity and power allocation for spectrum-sharing communications in fading channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 148–156, Jan. 2009.

[10] X. Kang, Y.-C. Liang, A. Nallanathan, H. K. Garg, and R. Zhang, "Optimal power allocation for fading channels in cognitive radio networks: ergodic capacity and outage capacity," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 940–950, Feb. 2009.

[11] X. Kang, R. Zhang, Y.-C. Liang, and H. K. Garg, "Optimal power allocation strategies for fading cognitive radio channels with primary user outage constraint," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 374–383, Feb. 2011.

[12] L. Musavian and S. Aissa, "Fundamental capacity limits of cognitive radio in fading environments with imperfect channel information," *IEEE Trans Wireless Commun.*, vol. 57, no. 11, pp. 3472–3480, Nov. 2009.

[13] H. A. Suraweera, and P. J. Simth, and M. Shafi, "Capacity limits and performance analysis of cognitive radio with imperfect channel knowledge," *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 1811–1822, May 2010.

[14] X. Gong, S. A. Vorobyov, and C. Tellambura, "Optimal bandwidth and power allocation for sum ergodic capacity under fading channels in cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1814–1826, Apr. 2011.

[15] H. Guo and V. C. M. Leung, "Orthogonal transmissions for spectrum underlay MISO cognitive radio," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1266–1270, Apr. 2012.

[16] A. J. Goldsmith and S. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[17] Y.-C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hong, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1326–1337, Apr. 2008.

[18] W. Zhang, R. K. Mallik, and K. B. Letaief, "Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5761–5766, Dec. 2009.

[19] C. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing neyworks," *IEEE Trans. Automatic Control*, vol. 39, no. 5, pp. 913–931, May 1994.

[20] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of Quality of Service," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630–643, July 2003.

[21] J. Tang and X. Zhang, "Quality-of-service driven power and rate adaptation over wireless links," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 3058–3068, Aug. 2007.

[22] L. Musavian and S. Aissa, "Effective capacity of delay-constrained cognitive radio in Nakagami fading channels," *IEEE Trans. Wireless Commun.*, vol. 9, no. 3, pp. 1054–1062, Mar. 2010.

[23] Y. Ma, H. Zhang, D. Yuan, and H. Chen, "Adaptive power allocation with quality-of-service guarantee in cognitive radio networks," *Comput. Commun.*, vol. 32, no. 18, pp. 1975–1982, Dec. 2009.

[24] Y. Wang, P. Ren, and F. Gao, "Power allocation for statistical QoS provisioning in opportunistic multi-relay DF cognitive networks," *IEEE Signal Process. Lett.*, vol. 20, no. 1, pp. 43–46, Jan. 2013.

[25] S. Akin and M. C. Gursoy,"Effective capacity analysis of cognitive radio channels for quality of service provisioning," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3354–3364, Nov. 2010.

[26] S. Akin and M. C. Gursoy,"Performance analysis of cognitive radio systems under QoS constraints and channel uncertainty," *IEEE Trans. Wireless Commun.*, vol. 10, no. 9, pp. 2883–2895, Sept. 2011.

[27] Y. Wang, P. Ren, Q. Du, and Z. Su,"Resource allocation and access strategy selection for QoS provisioning in cognitive networks," in *Proc. 2012 IEEE Int. Conf. Commun.*, pp. 4632–4636.

[28] S. Akin and M. C. Gursoy, "Cognitive radio transmission under QoS constraints and interference limitations," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 301, pp. 1–15, Sept. 2012.

[29] R. Menon, R. M. Buehrer, and J. H. Reed,"Outage probability based comparison of underlay and overlay spectrum sharing techniques," in *Proc. 2005 IEEE International Symp. New Frontiers Dynamic Spectrum Access Netw.*, pp. 101–109.

[30] R. Blasco-Serrano, J. Lv, R. Thobaben, E. Jorswieck, A. Kliks, and M. Skoglund, "Comparison of underlay and overlay spectrum sharing strategies in MISO cognitive channels," in *Proc. 2012 International Conf. Cognitive Radio Oriented Wireless Netw.*, pp. 224–229.

[31] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

Yichen Wang (S'13-M'14) received his B.S. and Ph.D. degrees both from Xi'an Jiaotong University. He is currently an Assistant Professor of Information and Communications Engineering Department, Xi'an Jiaotong University, China. His research interests include mobile wireless communications and networks with emphasis on ad hoc networks, MAC protocol design, cognitive radio techniques, statistical QoS provisioning, and resource allocation over wireless networks. He has published nearly 20 technical papers on international journals and conferences. His work co-authored with his Ph.D. advisor Prof. Pinyi Ren received the Best Letter Award in the IEICE Communications Society for the paper "A throughput-aimed MAC protocol with QoS provision for cognitive Ad Hoc networks" in 2010. He is the frequent reviewers for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE COMMUNICATIONS LETTERS, *EURASIP Journal on Wireless Communications and Networking*, IEEE GLOBECOM, IEEE ICC, etc.

Pinyi Ren (M'10) received his B.S., M.S. and Ph.D. degrees all from Xi'an Jiaotong University, China. He is currently a Professor and the Department Head of Information and Communications Engineering Department, Xi'an Jiaotong University. His current research interests include cognitive radio networks, MIMO systems, game theory in wireless communications, wireless relay, routing, signal detection, etc. He has published more than 80 technical papers in international Journals and conferences. He received the Best Letter Award of IEICE Communications Society in 2010. He has eight authorized Chinese Patents and the Copyrights for three Software Programs.

Prof. Ren serves as an Editor for the Journal of Xi'an Jiaotong University, the Lead Guest Editor for the Special Issue of Mobile Networks and Applications on "Distributed Wireless Networks and Services," and the Lead Guest Editor for the Special Issues of *Journal of Electronics* on "Cognitive Radio." He serves as the General Chair of ICST WICON 2011, the Technical Program Committee members of IEEE GLOBECOM, IEEE ICC, IEEE ICCC, etc. Prof. Ren is a Senior Member of Chinese Institute of Communications, a Member of IEICE and IEICE Communications Society, and a Member of IEEE.

Feifei Gao (S'05-M'09) received the Ph.D. degree from National University of Singapore, Singapore, in 2007. He was a Research Fellow with the Institute for Infocomm Research, A*STAR, Singapore, in 2008 and was an Assistant Professor with the School of Engineering and Science, Jacobs University, Bremen, Germany, from 2009 to 2010. In 2011, he joined the Department of Automation, Tsinghua University, Beijing, China, where he is currently an Associate Professor.

Prof. Gao has authored and coauthored more than 50 refereed IEEE journals and more than 70 IEEE conference proceeding papers. He has served as an editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS LETTERS, and *China Communications* and as a Technical Program Committee member for the IEEE International Conference on Communications, the IEEE Global Communications Conference, the IEEE Vehicular Technology Conference, and the IEEE Personal, Indoor, and Mobile Radio Communications Conference.

Zhou Su (S'03) received the B.E and M.E degrees from Xian Jiaotong University, Xi'an, China, in 1997, 2000, and Ph.D degree from Waseda University, Tokyo, Japan, in 2003, respectively. He was an exchange student between Waseda and Xi'an Jiaotong University from 1999 to 2000. From 2001 to 2004 he had been a research associate at Waseda University and he is currently an associate professor at the same university. His research interests include multimedia communication, web performance and network traffic. He received the best paper award of International Conference CHINACOM2008, and Funai Information Technology Award for Young Researchers 2009. He is a member of the IEEE and IEICE.