Primary-Prioritized Markov Approach for Dynamic Spectrum Allocation

Beibei Wang, *Student Member, IEEE,* Zhu Ji, K. J. Ray Liu, *Fellow, IEEE,* and T. Charles Clancy, *Member, IEEE*

***Abstract*—Dynamic spectrum access has become a promising approach to fully utilize the scarce spectrum resources. In a dynamically changing spectrum environment, it is very important to consider the statistics of different users' spectrum access so as to achieve more efficient spectrum allocation. In this paper, we propose a primary-prioritized Markov approach for dynamic spectrum access through modeling the interactions between the primary and the secondary users as continuous-time Markov chains (CTMC). Based on the CTMC models, to compensate the throughput degradation due to the interference among secondary users, we derive the optimal access probabilities for the secondary users, by which the spectrum access of the secondary users is optimally coordinated, and the spectrum dynamics are clearly captured. Therefore, a good tradeoff can be achieved between the spectrum efficiency and fairness. The simulation results show that the proposed primary-prioritized dynamic spectrum access approach under proportional fairness criterion achieves much higher throughput than the CSMA-based random access approaches and the approach achieving max-min fairness. Moreover, it provides fair spectrum sharing among secondary users with only small performance degradation compared to the approach maximizing the overall average throughput.**

***Index Terms*—Cognitive radio networks, dynamic spectrum access, interference management, Markov chain.**

I. INTRODUCTION

THE usage of radio spectrum resources and the regulation of radio emissions are coordinated by national regulatory bodies like the Federal Communications Commission (FCC). The FCC assigns spectrum to license holders or services on a long-term basis for large geographical regions; however, a large portion of the assigned spectrum remains unutilized. The inefficient usage of the limited spectrum necessitates the development of dynamic spectrum access techniques. Recently, the FCC began considering more flexible and comprehensive uses of available spectrum [1], [2], through the use of *cognitive radio* technology [3]. By exploiting the spectrum in an opportunistic fashion, dynamic spectrum access enables secondary users to sense which portions of the spectrum are available, select the best channel, coordinate access to spectrum channels with other users, and vacate the channel when a licensed user appears.

In order to fully utilize the limited spectrum resources, efficiently and fairly sharing the spectrum among secondary users becomes an important issue, especially when multiple dissimilar secondary users coexist in the same portion of the spectrum band. There have been several previous efforts addressing this issue on a negotiated/pricing basis [4]-[9] or an opportunistic basis [10], [11]. A local bargaining mechanism was proposed in [4] to distributively optimize the efficiency of spectrum allocation and maintain bargaining fairness among secondary users. Rule-based approaches were proposed in [5] that regulate users' spectrum access in order to trade-off fairness and utilization with communication costs and algorithmic complexity. In [6], the authors proposed a repeated game approach, in which the spectrum sharing strategy could be enforced using the Nash Equilibrium of dynamic games. In [7] [8], belief-assisted dynamic pricing was used to optimize the overall spectrum efficiency while basing the participating incentives of the selfish users on double auction rules. A centralized spectrum server was considered in [9] to coordinate the transmissions of a group of wireless links sharing a common spectrum. Recently, attention is being drawn to opportunistic spectrum sharing. In [10], a distributed random access protocol was proposed to achieve airtime fairness between dissimilar secondary users in open spectrum wireless networks without considering primary users' activities. The work in [11] examined the impact of secondary user access patterns on blocking probability and achievable improvement in spectrum utilization with statistical multiplexing, and proposed a feasible spectrum sharing scheme.

Although existing dynamic spectrum access schemes have successfully enhanced spectrum efficiency, most of them focus on spectrum allocation among secondary users in a static spectrum environment. Therefore, several fundamental challenges still remain unanswered. First, the radio spectrum environment is constantly changing. In conventional power control to manage mutual interference for a fixed number of secondary users, after each change of the number of contending secondary users, the network needs to re-optimize the power allocation for all users completely. This results in high complexity and much overhead. Second, if a primary user appears in some specific portion of the spectrum, secondary users in that band need to adapt their transmission parameters to avoid interfering with the primary user. Furthermore, in addition to maximizing the overall spectrum utilization, a good spectrum sharing scheme should also achieve fairness among

Manuscript received January 14, 2008; revised May 3, 2008 and June 23, 2008; accepted August 4, 2008. The associate editor coordinating the review of this paper and approving it for publication was H.-H. Chen.

This paper was presented in part at the IEEE Symposia on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), Dublin, Ireland, April 2007.

B. Wang and K. J. R. Liu are with the Department of Electrical and Computer Engineering, and Institute for Systems Research, University of Maryland, College Park, MD 20742, USA (e-mail: {bebewang, kjrliu}@umd.edu).

Z. Ji is with Qualcomm, San Diego, CA 92121, USA (e-mail: zji@qualcomm.com).

T. C. Clancy is with the Laboratory for Telecommunications Sciences, US Department of Defense, College Park, MD 20740, USA (e-mail: clancy@LTSnet.net).

Digital Object Identifier 10.1109/T-WC.2008.080031

dissimilar users. If multiple secondary users are allowed to access the licensed spectrum, dynamically coordinating their access to alleviate mutual interference and avoid conflict with primary users should be carefully considered.

Motivated by the preceding, in this paper we propose a primary-prioritized Markov approach for dynamic spectrum access. Specifically, we propose to model the interactions between the primary users (legacy spectrum holders) and the secondary users (unlicensed users) as continuous-time Markov chains (CTMC), by which we can capture the system evolution dynamics, especially the effect of the primary user's activities on the secondary users. It has been shown in [14], [15] that when unlicensed devices coexist with licensed devices in the same frequency and time simultaneously, the capacity achieved by unlicensed devices with reduced power is very low, while they still cause harmful interference to the licensed users. Therefore, in this paper, we assume that when primary users exist in some spectrum band, secondary users cannot operate in the same band simultaneously.

Further, in order to coordinate secondary spectrum access in a fair and efficient manner, dynamic spectrum access under different criteria is proposed based on the CTMC models. In the proposed approach, the spectrum access of different users is optimally coordinated through the modeling of secondary spectrum access statistics to alleviate mutual interference.

The contributions of the proposed primary-prioritized Markov approach for dynamic spectrum access are multifold. First, the radio system's evolutionary behavior, including the primary user's activities, is thoroughly captured through CTMC modeling. Second, we consider various policies of spectrum access by employing different optimality criteria, among which we focus on the proportional-fair (PF) spectrum access approach to achieve the optimal tradeoff between spectrum utilization efficiency and fairness. Third, the proposed PF spectrum access approach can achieve better performance than the CSMA-based scheme, and can be generalized to spectrum sharing among multiple secondary users.

The remainder of this paper is organized as follows: Dynamic spectrum access system model is described in Section II. The primary-prioritized Markov models are derived in Section III, and dynamic spectrum access approaches based on these models are developed in Section IV. The simulation studies are provided in Section V. Finally, Section VI concludes.

II. System Model

We consider dynamic spectrum access networks where multiple secondary users are allowed to access the temporarily-unused licensed spectrum bands on an opportunistic basis, without conflicting or interfering the primary spectrum holders' usage. Such scenarios can be envisioned in many applications. Considering the fact [1] that heavy spectrum utilization often takes place in unlicensed bands while licensed bands often experience low (e.g., TV bands) or medium (e.g., some cellular bands) utilization, IEEE 802.22 [16] proposes to reuse the fallow TV spectrum without causing any harmful interference to incumbents (e.g., the TV receivers). Moreover, with regard to more efficient utilization of some cellular bands, [17] proposes to share the spectrum between a cellular communication system and wireless local area network (WLAN) systems. In rural areas where there is little demand on the cellular communication system, the WLAN users can efficiently increase their data rates by sharing the spectrum.

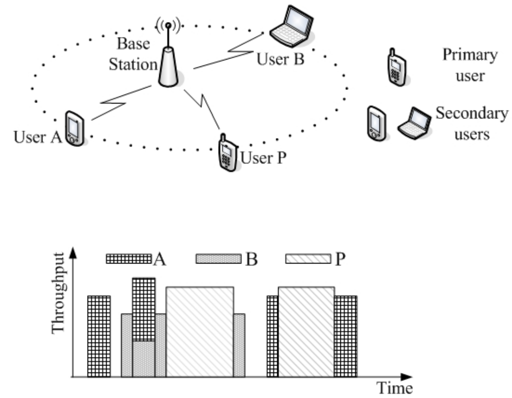


Fig. 1. System model (upper: system diagram; lower: throughput vs. time).

In order to take advantage of the temporally unused spectrum holes in the licensed band, without loss of generality we consider a snapshot of the above spectrum access networks shown in Fig. 1, where two secondary users and one primary user coexist, and the secondary users opportunistically utilize the spectrum holes in the licensed band. Note that the system diagram shown here serves only as an example model to gain more insight and the scenario with multiple secondary users will be studied in details in the following section.

The primary user is denoted by P, which has a license to operate in the spectrum band. The offered traffic for primary user P is modeled with two random processes[1]. The service request is modeled as a Poisson process with rate λ_P s^{-1}. The service duration (holding time) is negative-exponentially distributed with mean time $1/\mu_P$ s, so the departure of user P's traffic is another Poisson process with rate μ_P s^{-1}.

The secondary users are denoted by A and B, and set S is defined as $\mathrm{S} = \{A, B\}$. For each secondary user γ, where $\gamma \in \mathrm{S}$, its service session is similarly characterized by two independent Poisson processes, with arrival rate λ_γ s^{-1} and departure rate μ_γ s^{-1}. They contend to access the spectrum when primary user P is not using the spectrum band.

Since the primary user has a license to operate in the spectrum band, its access should not be affected by the operation of any other secondary user, and priority to access the spectrum is given to primary user P. We assume that the secondary users equipped with cognitive radios are capable of detecting the primary user's activities, i.e., the appearance of the primary user in the spectrum band and its departure from the spectrum. Furthermore, the secondary users' access is assumed to be controlled by a secondary management point so that they can distinguish whether the spectrum is occupied by the primary user or secondary users. Therefore, when primary user P appears, the secondary users should adjust their transmission parameters, for instance reduce the

[1]Identical assumptions that the service requests and departures are Poisson processes can be found in [10], [26] and references therein.

transmit power or vacate the channels and try to transfer their communications to other available bands. The *interference temperature* model [13] is proposed by FCC that allows secondary users to transmit in licensed bands with carefully adjusted power, provided that secondary users' transmission does not raise the interference temperature for that frequency band over the interference temperature limit. Although it can provide better service continuity for the secondary users to remain operating in the band with reduced power, the capacity they can achieve is very low [14], [15]. Therefore, in this paper, we assume that when primary user P appears, any secondary user should vacate and the traffic currently being served is cut off. In the duration of primary user P being served, any entry of the secondary user's traffic into the spectrum is denied until primary user P finishes its service.

In the bottom of Fig. 1, we show an example of the system throughput versus time for the dynamic spectrum access. First, user A accesses the spectrum band, followed by user B. During B's service, user A accesses the band again and shares the spectrum band with user B, which may result in less throughput to both user A and B due to their mutual interference. After user A has finished its service for a while, primary user P accesses the band, and user B's service is interrupted. After user P vacates the band, user B continues its service until its service duration ends. Afterwards, user A accesses the band, and its service is ceased when primary user P appears and resumed when P finishes its service in the way as user B.

For any secondary user γ that operates in the spectrum band alone, its maximal data rate [18] can be represented by

$$r_1^\gamma = W \log_2(1 + \frac{p_\gamma G_{\gamma\gamma}}{n_0}), \tag{1}$$

where W is the communication bandwidth, n_0 is power of the additive white Gaussian noise (AWGN), p_γ is the transmission power for user γ, and $G_{\gamma\gamma}$ is the channel gain for user γ. The secondary users A and B are allowed to share the spectrum band. We assume that the transmitter of a secondary user can vary its data rate through a combination of adaptive modulation and coding, so the transmitter and receiver can employ the highest rate that permits reliable communication, given the signal-to-interference-plus-noise ratio (SINR). We assume that the secondary users use random Gaussian codebooks, so their transmitted signals can be treated as white Gaussian processes and the transmission of other secondary users are treated as Gaussian noise. Then, the maximal rate of user γ when secondary users share the spectrum can be represented by

$$r_2^\gamma = W \log_2(1 + \frac{p_\gamma G_{\gamma\gamma}}{n_0 + \sum_{\alpha \neq \gamma} p_\alpha G_{\alpha\gamma}}), \tag{2}$$

where $\alpha \neq \gamma$, $\alpha \in$ S, and $G_{\alpha\gamma}$ is the channel gain from user α's transmitter to user γ's receiver.

III. Primary-Prioritized Markov Models

In this section, we derive primary-prioritized Markov models to capture the dynamics of spectrum access statistics for the primary user and the secondary users.

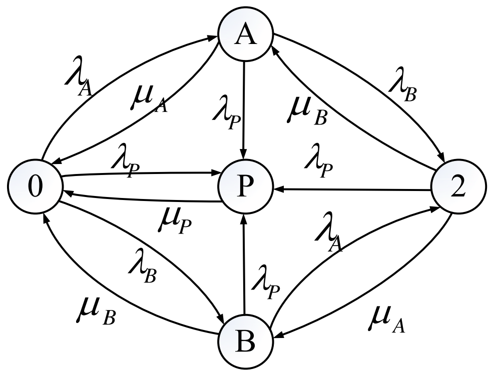


Fig. 2. The rate diagram of CTMC with no queuing.

A. Primary-Prioritized CTMC without Queuing

1) CTMC without Queuing: In dynamic spectrum access, where the secondary users opportunistically access the unused licensed spectrum, priority should be given to the primary user. That is, secondary users cannot operate in the same spectrum band with the primary user at the same time; when the primary user appears in the spectrum band, all secondary users in the same band should stop operating in the spectrum. Moreover, the arrival and departure of different users' traffic are assumed to be independent Poisson processes. Therefore, we model the interactions between the secondary users and the primary user as a primary-prioritized CTMC.

In the CTMC, when the secondary users contend to access the idle spectrum using CSMA, collisions only occur when their service requests arrive exactly at the same time. This case rarely happens for independent Poisson processes. Therefore, in the CTMC model we omit the collision state of the secondary users, and assume their service durations always start from different time instances.

If we assume that when the primary user appears, there is no queuing of the interrupted service for the secondary users, then we can model the spectrum access process as a five-state CTMC shown in Fig. 2. We denote this five-state Markov chain by "CTMC-5" for short, where state 0 means no user operates in the spectrum, state γ means user γ operates in the spectrum with $\gamma \in \{A, B, P\}$, and state 2 means both user A and user B operate in the spectrum.

Assume at first the spectrum band is idle, i.e., CTMC-5 is in state 0. Secondary users contend to operate in the spectrum. Upon the first access attempt of some user, say user A, CTMC-5 enters state A with transition rate λ_A s^{-1}. If user A's service completes before any other user requests spectrum access, CTMC-5 then transits to state 0 with rate μ_A s^{-1}. If user B's service request arrives before A completes its service, CTMC-5 transits to state 2 with rate λ_B s^{-1}, where both secondary users share the spectrum. Once user B (or A)'s service is completed, CTMC-5 transits from state 2 to state A (or B), with rate μ_B (or μ_A) s^{-1}. However, primary user P may, once in a while, appear during the service duration of the secondary users, i.e., when CTMC-5 is in state A, B or 2. At that time, the secondary user's traffic is dropped to avoid conflict with the primary user, and CTMC-5 transits to state P with rate λ_P s^{-1}. During the primary user operating in the spectrum band, no secondary user is given access to the spectrum. CTMC-5 transits to state 0 with rate μ_P s^{-1} only if P completes its service.

The "flow-balance" (the rate at which transitions take place out of state s_i equals to the rate at which transitions take place into state s_i) and the normalization [19] equations governing the above system are given by

$$\mu_A\Pi_A + \mu_P\Pi_P + \mu_B\Pi_B = (\lambda_A + \lambda_B + \lambda_P)\Pi_0, \quad (3)$$
$$\lambda_A\Pi_0 + \mu_B\Pi_2 = (\mu_A + \lambda_P + \lambda_B)\Pi_A, \quad (4)$$
$$\lambda_P(\Pi_0 + \Pi_A + \Pi_2 + \Pi_B) = \mu_P\Pi_P, \quad (5)$$
$$\lambda_B\Pi_0 + \mu_A\Pi_2 = (\mu_B + \lambda_P + \lambda_A)\Pi_B, \quad (6)$$
$$\lambda_B\Pi_A + \lambda_A\Pi_B = (\mu_B + \lambda_P + \mu_A)\Pi_2, \quad (7)$$
$$\Pi_0 + \Pi_A + \Pi_B + \Pi_P + \Pi_2 = 1, \quad (8)$$

where Π_{s_i} represents the stationary probability of being in state s_i, $s_i \in \mathcal{S} \triangleq \{0, A, B, P, 2\}$.

The solutions to the above equations, i.e., the probabilities when the spectrum is occupied by either primary user P or the secondary users, are given by

$$\Pi_P = \lambda_P/(\lambda_P + \mu_P), \quad (9)$$
$$\Pi_A = C_1\lambda_A[\lambda_B\mu_B + (\lambda_P + \mu_B)(\lambda_A + \lambda_P + \mu_A + \mu_B)] \quad (10)$$
$$\Pi_B = C_1\lambda_B[\lambda_A\mu_A + (\lambda_P + \mu_A)(\lambda_B + \lambda_P + \mu_A + \mu_B)] \quad (11)$$
$$\Pi_2 = C_1\lambda_A\lambda_B(\lambda_A + \lambda_B + 2\lambda_P + \mu_A + \mu_B), \quad (12)$$

where, for simplicity, the coefficient C_1 is defined as

$$\begin{aligned} C_1 =& (1 - \Pi_P)[(\lambda_A + \mu_A + \lambda_P)(\lambda_B + \mu_B + \lambda_P) \\ & (\lambda_A + \mu_A + \lambda_B + \mu_B + \lambda_P)]^{-1}. \end{aligned} \quad (13)$$

One of the most important goals in spectrum sharing is efficient spectrum utilization, i.e., high throughput achieved by each secondary user through successful acquisition of a spectrum band. From a statistical point of view, the secondary users want to maximize their average throughput. Given the solutions of the steady state probabilities, we know that Π_{s_i} is the stationary probability that the system is in state s_i, so it can be thought of as the expected long-run fraction of the time that the CTMC spends in state s_i [19]:

$$\Pi_{s_i} = \lim_{T\to\infty} \frac{1}{T}\int_0^T \Pr\{\mathcal{S}(t) = s_i\}dt, \quad (14)$$

where $\mathcal{S}(t)$ is the state of the CTMC at time t. If we define

$$U_\gamma = \lim_{T\to\infty} \frac{1}{T} E\left(\int_0^T R_\gamma(\mathcal{S}(t))dt\right) \quad (15)$$

as the long-run expected average throughput for user γ, where $R_\gamma(\mathcal{S}(t))$ is the throughput of user γ achieved in state $\mathcal{S}(t)$, we have

$$\begin{aligned} U_\gamma &= \lim_{T\to\infty} \frac{1}{T}\int_0^T E(R_\gamma(\mathcal{S}(t)))dt \\ &= \lim_{T\to\infty} \frac{1}{T}\int_0^T \sum_{s_i\in\mathcal{S}} R_\gamma(s_i)\Pr\{\mathcal{S}(t) = s_i\}dt \\ &= \sum_{s_i\in\mathcal{S}} R_\gamma(s_i) \lim_{T\to\infty} \frac{1}{T}\int_0^T \Pr\{\mathcal{S}(t) = s_i\}dt \\ &= \sum_{s_i\in\mathcal{S}} R_\gamma(s_i)\Pi_{s_i}. \end{aligned} \quad (16)$$

The interchanges of limits, integrals, sums, etc. are permitted as long as $\sum_{s_i\in\mathcal{S}} |R|_\gamma(s_i)\Pi_{s_i} < \infty$. Thus, from CTMC-5, we can express the total average throughput for user γ as follows,

$$U_\gamma = \Pi_\gamma r_1^\gamma + \Pi_2 r_2^\gamma, \quad (17)$$

where Π_γ and Π_2 are as solved in (10)-(12), and r_1^γ and r_2^γ are defined in (1) and (2), respectively. The first term on the right-hand side of (17) represents the throughput when user γ occupies the spectrum alone, and the second term represents the throughput when two secondary users share the spectrum.

Therefore, by using CTMC-5, we not only can capture the dynamic utilization of the unused licensed spectrum for secondary users without conflicting with the primary user, but also can study their stationary behaviors and quantify their spectrum utilization from a statistical point of view.

2) Multi-User CTMC without Queuing: The CTMC previously introduced can also be generalized to model the scenario with more than two secondary users. Suppose the set of N secondary users is denoted by $\mathbf{S} = \{1, \cdots, N\}$, then the state space $\mathcal{A}$ consists of $2^N + 1$ combinations of the status of primary user P and the secondary users:

$$\begin{aligned} (\Phi_P, \Phi_{\mathbf{S}}) \in \mathcal{A} \triangleq \{(1, [0, \cdots, 0])\} \bigcup \\ \{(0, \phi_{\mathbf{S}}) : \phi_{\mathbf{S}} \triangleq [n_N, \cdots, n_1] \in \{0, 1\}^N\}, \end{aligned} \quad (18)$$

where state $(1, [0, \cdots, 0])$ represents the case where the primary user is in service in the spectrum band alone, and $\{(0, \phi_{\mathbf{S}})\}$ represents all 2^N states where primary user P is not in service and zero up to N secondary users are in service.

For this generalized Markov model, the rate diagram can be drawn as an N-dimensional hypercube. Each vertex of the hypercube represents a state in $\{(0, \phi_{\mathbf{S}})\}$; each edge connecting two vertices is bi-directional, and it represents the transition that some secondary user begins or completes its service. The center of the hypercube represents state $(1, [0, \cdots, 0])$; a straight line from each vertex to the center represents the transition when primary user P begins its service, and another line from the center to state $(0, [0, \cdots, 0])$ represents the transition when user P completes its service. The stationary probabilities can be obtained by solving the corresponding linear equations in Appendix VII-A.

For each secondary user γ, $\gamma \in \mathbf{S} = \{1, \cdots, N\}$, its average throughput consists of 2^{N-1} components, each of which represents the average throughput when user γ, together with zero up to all the other $N - 1$ secondary users, are in service. Since more secondary users contend the spectrum access, the contention in the generalized Markov model becomes heavier than CTMC-5. As a result, each secondary user shares less spectrum access on average. Moreover, the interference also increases by introducing more secondary users. Therefore, as the number of secondary users increases, the average throughput for each of them is reduced.

B. Primary-Prioritized CTMC with Queuing

1) CTMC with Queuing: In CTMC-5 presented in Section III-A1, the service of the secondary users is forced to stop and be dropped when primary user P appears in the spectrum band. After primary user P completes its service, CTMC-5

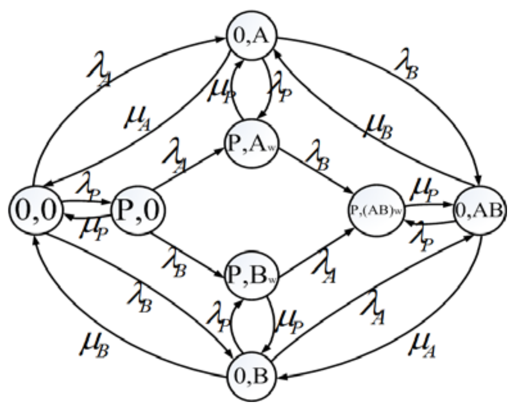


Fig. 3. The rate diagram of CTMC with queuing.

will transit to the idle state. However, there may be some time interval wasted between when the system is in the idle state and the next secondary user accesses the spectrum. In order to further increase the spectrum utilization, queuing of the secondary users' service requests due to the primary user's presence is considered. More specifically, when the spectrum is being occupied by secondary users, upon the appearance of primary user, the secondary users should stop transmission, buffer their interrupted service session, and continue scanning the licensed band until the licensed band becomes available again. Also, if the primary user begins to operate in the previously idle spectrum, new service requests of secondary users are also queued. In this paper, we assume that there is one waiting room for the secondary user, i.e., each user can only buffer a single service request; and if a service request already exists in the queue, the secondary user will direct the following service requests to other available licensed bands to avoid potential delay, and that scenario is beyond the scope of this paper.

By considering the above factors, we model the spectrum access with queuing as an eight-state CTMC, denoted by "CTMC-8". The rate diagram of CTMC-8 is shown in Fig. 3. Compared to CTMC-5 and its dynamics, in CTMC-8 three additional states are introduced: (P, A_w), (P, B_w) and $(P, (AB)_w)$. State (P, γ_w) means primary user P is in service and secondary user γ is waiting, and state $(P, (AB)_w)$ means P is in service and both secondary users are waiting. The transitions in CTMC-8 are briefed as follows. When the spectrum band is occupied by secondary user A, if A detects that primary user P needs to acquire the spectrum band, it buffers the unfinished service session, sensing the licensed band until the end of the primary user's service session, and CTMC-8 transits from state $(0, A)$ to state (P, A_w) with rate λ_P s^{-1}. If primary user P finishes its service before B's access, CTMC-8 transits from state (P, A_w) to $(0, A)$ with rate μ_P s^{-1}. In contrast, if secondary user B requests access to the licensed spectrum before primary user P completes its service duration, B also buffers its service session, and CTMC-8 transits to state $(P, (AB)_w)$ with rate λ_B s^{-1}. In state $(P, (AB)_w)$, both A and B keep sensing the spectrum. Once P vacates, CTMC-8 transits to state $(0, AB)$ with rate μ_P s^{-1}, where A and B share the spectrum band. Also, when CTMC-8 is in state $(P, 0)$, if secondary users attempt to access the spectrum, they will keep sensing the licensed band until the primary user vacates, and CTMC-8 transits to state (P, A_w) or state (P, B_w), with rate λ_A s^{-1} or λ_B s^{-1}, respectively.

The equations governing the above system and the corresponding solutions can be obtained in a similar way as in Section III-A1.

2) Multi-User CTMC with Queuing: CTMC with queuing can also be generalized to model the scenario with more than two secondary users. For the Markov chain with a set $\mathbf{S} = \{1, \cdots, N\}$ of secondary users, the state space $\mathcal{B}$ consists of all possible 2^{N+1} combinations of the status for primary user P and the secondary users:

$$(\Psi_P, \Psi_{\mathbf{S}}) \in \mathcal{B} \triangleq \{(1, \psi_{\mathbf{S}}^w) \bigcup (0, \psi_{\mathbf{S}}) : \psi_{\mathbf{S}} \triangleq [n_N, \cdots, n_1] \in \{0, 1\}^N\}, \tag{19}$$

where $\{(1, \psi_{\mathbf{S}}^w)\}$ represents all 2^N states in which the primary user is in service and zero up to N secondary users are waiting, and $\{(0, \psi_{\mathbf{S}})\}$ represents all 2^N states where primary user P is not in service and a subset of the N secondary users are in service. The rate diagram for this model can be similarly drawn as in Section III-A2, and the stationary probabilities can be obtained as shown in Appendix VII-B. As more secondary users contend the spectrum, in addition to increased interference, more waiting time is also introduced; therefore, the average throughput for each secondary user will be reduced.

IV. Primary-Prioritized Dynamic Spectrum Access

In this section, we will first analyze the effect of secondary users' behavior on the system performance. Then, we propose primary-prioritized dynamic spectrum access with different optimality criteria and compare them to CSMA-based random access approaches.

In order to develop primary-prioritized dynamic spectrum access, it is important to first analyze the behavior of the secondary users. Since the secondary users contend for the spectrum, if they access the spectrum in a greedy manner such that all of their injected traffic is admitted, then the Markov chain is more likely to be in the state where more than one user shares the spectrum. Hence, the secondary users may suffer a throughput degradation due to interference, if there is no control on very high arrival rates. On the other hand, if the secondary users reduce their arrival rates too much so as to avoid interference, the average throughput may be unnecessarily low. Therefore, secondary user spectrum access should be carefully controlled.

In the proposed dynamic spectrum access scheme, we introduce the state-dependent spectrum access probabilities for user A and user B, and the resulting random access process can be approximated by slightly modifying the original CTMCs. Without loss of generality, we take CTMC-5 as an example, and the modified Markov chain is shown in Fig. 4. It is seen from the figure that when one secondary user, e.g. user B, already occupies the spectrum and the system is in state B, user A's spectrum access requests are admitted with probability $a_{A,1}$, where $0 \le a_{A,1} \le 1$. Since on average one

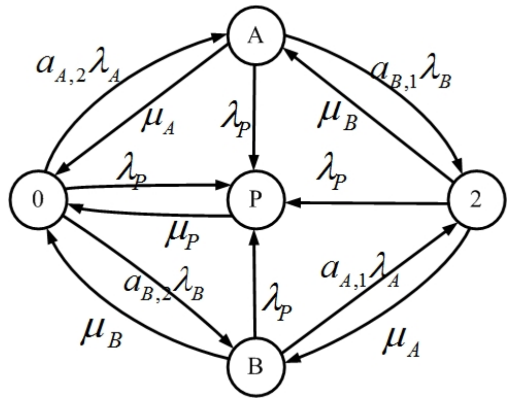


Fig. 4. Modified CTMC with access control (no queuing).

out of $\frac{1}{a_{A,1}}$ user A's access requests are allowed when user B is in service, the chance of coexistence of the secondary users and mutual interference can be reduced. Due to the *decomposition property* of Poisson random process [19], if each access request of user A has a probability $a_{A,1}$ of being admitted, then the number of actual admitted access requests is also a Poisson process with parameter $a_{A,1}\lambda_A$ s^{-1}. Hence, the transition rate from state B to state 2 now becomes $a_{A,1}\lambda_A$ s^{-1}. It is also seen that user A's access requests are admitted with probability $a_{A,2}$ when the spectrum is idle (i.e., the transition from state 0 to state A). However, there is no interference in state A. In order to obtain a high throughput, we assume that when the spectrum is sensed idle, user A is allowed to access the spectrum with probability one, i.e., $a_{A,2} = 1$. In addition, it is expected that if the mutual interference between the secondary users is high, $a_{A,1}$ should be close to 0; if there is little mutual interference, $a_{A,1}$ should be close to 1. User B's spectrum access is controlled in a similar way as user A, because the CTMC is symmetric.

Denote the access probability for user A and user B as vectors $\mathbf{a}_A = [a_{A,1}, a_{A,2}]$, and $\mathbf{a}_B = [a_{B,1}, a_{B,2}]$, respectively. Then, the optimization goal is to determine $\mathbf{a}_A$ and $\mathbf{a}_B$, such that the system performance can be maximized, i.e.,

$$\{\mathbf{a}_\gamma\} = \underset{0\le \mathbf{a}_\gamma \le 1}{\operatorname{argmax}} \quad U(\{\mathbf{a}_\gamma\}), \tag{20}$$

where $\forall\gamma \in \{A, B\}$.

Since a good spectrum sharing scheme not only can efficiently utilize the spectrum resources, but also can provide fairness among different users, we first propose to maximize the average throughput based on PF criterion [20] [21]. Thus, in (20), $U(\mathbf{a}_A, \mathbf{a}_B)$ can be written as

$$U_{PF}(\mathbf{a}_A, \mathbf{a}_B) = \prod_{\gamma\in \mathrm{S}} U_\gamma(\mathbf{a}_A, \mathbf{a}_B). \tag{21}$$

We also consider other criteria to compare with PF, expressed by the following maximal-throughput criterion

$$U(\mathbf{a}_A, \mathbf{a}_B) = \sum_{\gamma\in \mathrm{S}} U_\gamma(\mathbf{a}_A, \mathbf{a}_B), \tag{22}$$

and max-min fairness criterion

$$U(\mathbf{a}_A, \mathbf{a}_B) = \min_{\gamma\in \mathrm{S}} U_\gamma(\mathbf{a}_A, \mathbf{a}_B). \tag{23}$$

For the maximal-throughput optimization, the overall system throughput is maximized, but the users with the worse channel conditions may starve. For the max-min fairness optimization, the performance of the secondary user with the worst channel condition is optimized, while resulting in inferior overall system performance. In this paper, we will demonstrate that the PF dynamic spectrum access is preferred because it can ensure more fairness than the maximal-throughput optimization, while achieve better performance than the max-min fairness optimization. Specifically, the definition of PF is expressed as follows.

Definition: The throughput distribution is proportionally fair if any change in the distribution of throughput pairs results in the sum of the proportional changes of the throughput being non-positive [20], i.e.,

$$\sum_{\gamma\in \mathrm{S}} \frac{U_\gamma(\mathbf{a}_A, \mathbf{a}_B) - U_\gamma^*(\mathbf{a}_A, \mathbf{a}_B)}{U_\gamma^*(\mathbf{a}_A, \mathbf{a}_B)} \le 0, \tag{24}$$

where $U_\gamma^*(\mathbf{a}_A, \mathbf{a}_B)$ is the proportionally fair throughput distribution, and $U_\gamma(\mathbf{a}_A, \mathbf{a}_B)$ is any other feasible throughput distribution for user γ.

As proved in Appendix VII-C, the optimal solution $U_\gamma^*(\mathbf{a}_A, \mathbf{a}_B)$ defined in (24) can be obtained by solving (20), with $U(\mathbf{a}_A, \mathbf{a}_B)$ defined in (21).

As mentioned earlier in this section, we assume $a_{A,2} = a_{B,2} = 1$, then the two access probabilities to be optimized are $a_{A,1}$ and $a_{B,1}$. We denote them by a_A and a_B for simplicity, and can write U_γ as

$$U_\gamma(a_A, a_B) = \Pi_\gamma(a_A, a_B) r_1^\gamma + \Pi_2(a_A, a_B) r_2^\gamma \quad , \tag{25}$$

where

$$\begin{aligned}
\Pi_A(a_A, a_B) &= C_1\lambda_A[(\lambda_P + \mu_B)(a_A\lambda_A \\
&\qquad + \lambda_P + \mu_A + \mu_B) + a_A\lambda_B\mu_B] \\
\Pi_B(a_A, a_B) &= C_1\lambda_B[(\lambda_P + \mu_A)(a_B\lambda_B \\
&\qquad + \lambda_P + \mu_A + \mu_B) + a_B\lambda_A\mu_A] \\
\Pi_2(a_A, a_B) &= C_1\lambda_A\lambda_B[a_A(a_B(\lambda_A + \lambda_B) + \lambda_P \\
&\qquad + \mu_A) + a_B(\lambda_P + \mu_B)]
\end{aligned} \quad , \tag{26}$$

with

$$\begin{aligned}
C_1 = (1 - \Pi_P)\Big\{& a_A\lambda_A[a_B\lambda_B(\lambda_A + \lambda_B + \lambda_P) \\
&+ (\lambda_B + \lambda_P)(\lambda_P + \mu_A) + (\lambda_B + \lambda_P + \mu_A)\mu_B \\
&+ \lambda_A(\lambda_P + \mu_B)] + (\lambda_P + \mu_A + \mu_B)[\lambda_A(\lambda_P + \mu_B) \\
&+ (\lambda_P + \mu_A)(\lambda_B + \lambda_P + \mu_B)] + a_B\lambda_B[(\lambda_P + \mu_A) \\
&(\lambda_B + \lambda_P + \mu_B) + \lambda_A(\lambda_P + \mu_A + \mu_B)]\Big\}^{-1},
\end{aligned} \tag{27}$$

When $0 \le a_\gamma \le 1$, we have $\Pi_A(a_A, a_B) \ge 0$, $\Pi_B(a_A, a_B) \ge 0$, $\Pi_2(a_A, a_B) \ge 0$, and $U_\gamma(a_A, a_B) \ge 0$. Taking derivative of $U_\gamma(a_A, a_B)$ with respect to a_A, we can show that

$$\frac{\partial U_A(a_A, a_B)}{\partial a_A} > 0, \quad \frac{\partial U_B(a_A, a_B)}{\partial a_A} < 0. \tag{28}$$

So when secondary user A is given more chance to access the frequency band, i.e., when a_A increases, $U_A(a_A, a_B)$ becomes larger while $U_B(a_A, a_B)$ shrinks, indicating that there is a possible tradeoff to choose the optimal a_A that maximizes

TABLE I
PRIMARY-PRIORITIZED DYNAMIC SPECTRUM ACCESS

1. Initially primary user P is operating in the spectrum band;
2. Secondary access point obtains optimal access probabilities defined in (29) for secondary users (Other optimality criteria can also be implemented);
3. Once primary user P is sensed to have completed its service, secondary users start to access the spectrum band with the probabilities solved in Step 2 depending on various states;
4. When primary user P re-appears in the band, secondary users currently operating in the band vacate;
5. If secondary users still have service not completed, go back to Step 3; If the statistics of secondary users' services or their locations change, go to Step 2.

$U_{PF}(a_A, a_B) = U_A(a_A, a_B)U_B(a_A, a_B)$. However, it can be seen that there are a lot of variables in $U_\gamma(a_A, a_B)$ and hence the objective function $U_{PF}(a_A, a_B)$. In addition, the utility of each secondary user U_γ is a complicated function of the $\{\lambda_\gamma, \mu_\gamma, a_\gamma\}$'s and the data rates $\{r_1^\gamma, r_2^\gamma\}$'s. Therefore, it is analytically difficult to justify the concavity for arbitrary parameters. Nevertheless, given a specific set of parameters $\{\lambda_\gamma, \mu_\gamma\}$'s and $\{r_1^\gamma, r_2^\gamma\}$'s, we can substitute their values in (26) and determine the concavity of $U_{PF}(a_A, a_B)$ by observing the Hessian matrix $\nabla^2 U_{PF}(a_A, a_B)$, for $0 \leq a_A, a_B \leq 1$. When the two eigenvalues of $\nabla^2 U_{PF}(a_A, a_B)$ are not greater than zero, i.e., the Hessian matrix is negative semidefinite, we can determine that $U_{PF}(a_A, a_B)$ is concave with respect to a_A and a_B, and the optimal access probabilities can be expressed as

$$a_{\gamma,i}^{opt} = \min\{\max(a_{\gamma,i}^*, 0), 1\}, \tag{29}$$

where $a_{\gamma,i}^*$ is the solution to the following equations

$$\frac{\partial U_{PF}(a_A, a_B)}{\partial a_{\gamma,i}} = 0, \quad \forall \gamma, i \in \mathrm{S}. \tag{30}$$

If $U_{PF}(a_A, a_B)$ is not concave, we can check the points on the boundary of the feasible region of (a_A, a_B). For example, for some value of $\lambda_\gamma, \mu_\gamma$, if $r_1^\gamma \gg r_2^\gamma$, indicating heavy mutual interference, function $U_{PF}(a_A, a_B)$ may not be concave. However, the optimal solution of a_γ is 0 to avoid interference. Another instance where $U_{PF}(a_A, a_B)$ is not concave happens when $\lambda_\gamma \ll \mu_\gamma$, and the optimal solution is $a_\gamma = 1$.

We assume that there exists a secondary base station (BS) that can control the medium access for all the secondary users. The secondary users send periodic reports to the BS informing it about their service statistics and date rates. Using the information gathered from all secondary users, the BS evaluates the spectrum utilization, computes the optimal access probability in different states (i.e., when different set of secondary users are in service), and sends the access probability to the secondary users. Based on the above discussions, we illustrate our primary-prioritized Markov approach for dynamic spectrum access in Table I.

The proposed primary-prioritized dynamic spectrum access shares some characteristic with conventional medium access control (MAC) protocols, since they all target appropriate coordination of different users' access to the medium. For instance, in IEEE 802.11 [23], a CSMA/CA mechanism is employed. If the medium is sensed idle, a user transmits its packet; if the medium is sensed busy, then the user may reschedule the retransmission of the packet according to some random back-off time distribution. These kinds of protocols are effective when the medium is not heavily loaded, since they allow users to transmit with minimum delay. However, under heavy traffic load, there is always a chance that users' attempts conflict with each other. If the conflicted users are kept waiting for an idle medium, their packets suffer significant delay and may expire.

In the proposed primary-prioritized dynamic spectrum access, different secondary users are allowed to share the spectrum band simultaneously. This will increase the spectrum utilization because of the following reasons. First, for independent Poisson processes, the service durations of different secondary users are generally not the same. For instance, in CTMC-5, even though user B begins operating in the spectrum band right after user A, it is possible that user A completes its service much earlier than user B. After user B is admitted to occupy the spectrum band, the two secondary users share the spectrum only for a very short time. Once A finishes its service, the Markov chain transits to the state where B operates in the spectrum alone and no interference exists. Using CSMA protocols, however, user B is forced to retransmit its packet after a random back-off time, which may not be short. Therefore, using the proposed approaches, the spectrum can be more efficiently utilized. Furthermore, in the proposed schemes, optimal access probabilities are employed to carefully control the coexistence of the secondary users. In this way, the interference is maintained at a low level.

Also, in a mobile network, the radio spectrum environment is dynamic. When using global optimization approaches specific to a fixed environment, for instance conventional power control to manage mutual interference between a fixed number of secondary users, after each change in the number of contending secondary users, the network needs to re-optimize the power allocation for all users completely. This results in high complexity and much overhead, especially when there are frequent service requests and the service duration is short. In the proposed approach, by controlling the access probabilities for secondary users, there is no need to perform delicate power control to manage the interference, and computational complexity is reduced while the average throughput is maximized in the long-run.

In order to achieve optimal dynamic spectrum access, a certain overhead is needed. More specifically, the overhead mainly comes from access controlling and sensing primary users. To optimally coordinate the access of the secondary users, necessary measurements needs to be taken, such as the throughput and arrival/departure rates for different secondary users. On the other hand, detecting a primary user's presence

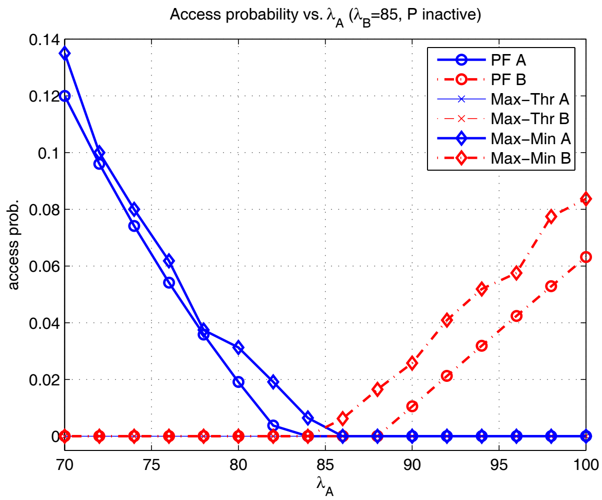


Fig. 5. Access probability vs. λ_A (symmetric-interference, $\lambda_B = 85\ \mathrm{s}^{-1}$).

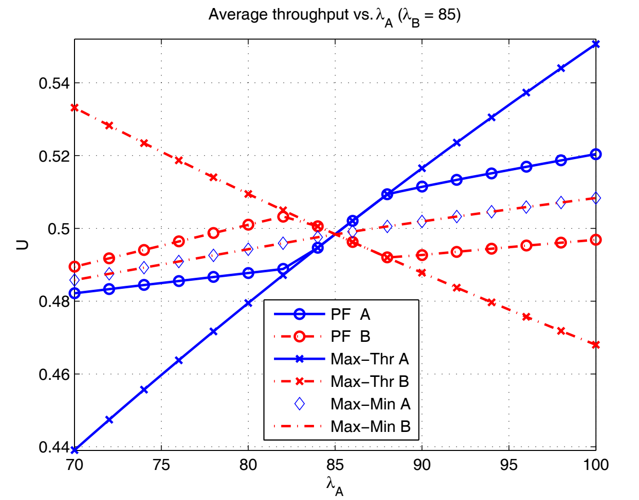


Fig. 6. Average throughput vs. λ_A (symmetric-interference, $\lambda_B = 85\ \mathrm{s}^{-1}$).

relies mainly on the observations from the secondary users and the necessary spectral analysis.

V. Simulation Results and Analysis

In this section, we first compare the performance of CTMC-8 with different optimization goals (maximal-throughput, max-min, and PF). Then we compare the performance of CTMC-8, CTMC-5, and the nonpersistent CSMA-based random access. Finally we show the throughput gain of spectrum sharing among more than two secondary users against the case without access control.

The parameters in the simulations are chosen as follows. We set the bandwidth of the licensed spectrum as $W = 200$ KHz, the transmission power of each secondary user $p_\gamma = 2$ mW, the noise power $n_0 = 10^{-15}$ W, and the propagation loss exponent factor as 3.6. The departure rates μ_A, μ_B, μ_P are set to be $100\ \mathrm{s}^{-1}$. According to [25], in the spectrum band allocated to cellular phone and Specialized Mobile Radio (SMR), the fraction of time that the spectrum is being used by primary users in an urban environment is measured as approximately 45%. Thus, when μ_P is $100\ \mathrm{s}^{-1}$, we set the arrival rate of the primary user $\lambda_P = 85\ \mathrm{s}^{-1}$. The arrival rate of secondary user B is $\lambda_B = 85\ \mathrm{s}^{-1}$, and we vary λ_A from 70 to $100\ \mathrm{s}^{-1}$. In the simulation results, we use "Max-Thr" to denote the maximal-throughput criterion, "Max-Min" to denote the max-min fairness criterion, and "A" and "B" to denote secondary users A and B, respectively.

A. CTMC-8 for the Symmetric-Interference Case

In the first set of simulations, we test the case where two secondary users experience symmetric interference. The transmitter of user A is at $(0\mathrm{m}, 0\mathrm{m})$, and its receiver is at $(200\mathrm{m}, 0\mathrm{m})$. The transmitter of user B is at $(200\mathrm{m}, 460\mathrm{m})$, and its receiver is at $(0\mathrm{m}, 460\mathrm{m})$. According to their symmetric locations, we know that $r_1^B = r_1^A > r_2^B = r_2^A$ from (1) and (2). In Fig. 5, we show the optimal access probability versus λ_A for each secondary user when the other secondary user is transmitting, i.e., the access probability associated with the transition from state $(0, \gamma)$ to $(0, AB)$ in CTMC-8 (see Fig. 3).

Since CTMC-8 is symmetric for the two users, when $\lambda_A < \lambda_B = 85\ \mathrm{s}^{-1}$, user A will have a smaller time share than user B if there is no access probability control. Further because we have $r_1^B = r_1^A > r_2^B = r_2^A$, from the definition of the average throughput in (17), user A will experience a lower average throughput than B. In order to provide more fairness, PF and max-min optimization assigns user B a zero access probability and assigns user A a higher access probability than user B when $\lambda_A < \lambda_B = 85\ \mathrm{s}^{-1}$. With the increase of λ_A, the difference between the two users' time share becomes smaller, so the access probability of user A decreases and is equal to B's access probability when $\lambda_A = \lambda_B$. When $\lambda_A > \lambda_B$, user B is assigned a higher access probability due to a smaller time share, while user A's access requests are denied. However, when $\lambda_A > \lambda_B = 85\ \mathrm{s}^{-1}$, λ_A is much higher and the probability of State A is also higher, in order to reduce the mutual interference, the growth of B's access probability is not symmetric to the decrease of A's. Due to the mutual interference, the maximal-throughput optimization assigns zero access probability to both users when the other user is in service.

In Fig. 6, we show the throughput U_γ for each user. Max-min fairness optimization provides absolute fairness to both users: the two U_γ's are identical and increase as λ_A goes up. In the PF optimization, when $\lambda_A < \lambda_B$, we have $U_A < U_B$. As λ_A becomes higher, U_A increases; however, as shown in Fig. 5, user A's access probability decreases as λ_A increases until $\lambda_A = \lambda_B = 85\ \mathrm{s}^{-1}$, so the mutual interference is managed and U_B also increases. When $\lambda_A = \lambda_B = 85\ \mathrm{s}^{-1}$, $U_A = U_B$, since the secondary users are identical in terms of both channel conditions and service requests. As λ_A further increases, $U_A > U_B$ and U_A keeps increasing; since user B's access probability increases as $\lambda_A > \lambda_B = 85\ \mathrm{s}^{-1}$ (see Fig. 5), U_B also increases. For the maximal-throughput optimization, as seen from Fig. 5, the access probabilities of the two users are both zero, indicating that they are not allowed to transmit simultaneously, so U_A keeps increasing

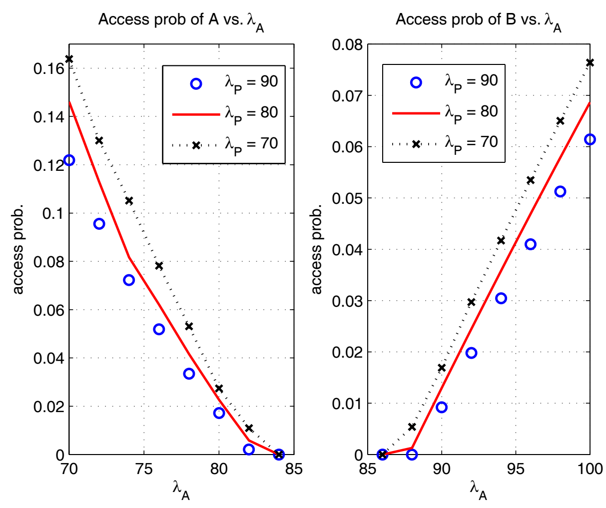


Fig. 7. Access probability for different λ_P ($\lambda_B = 85$ s^{-1}).

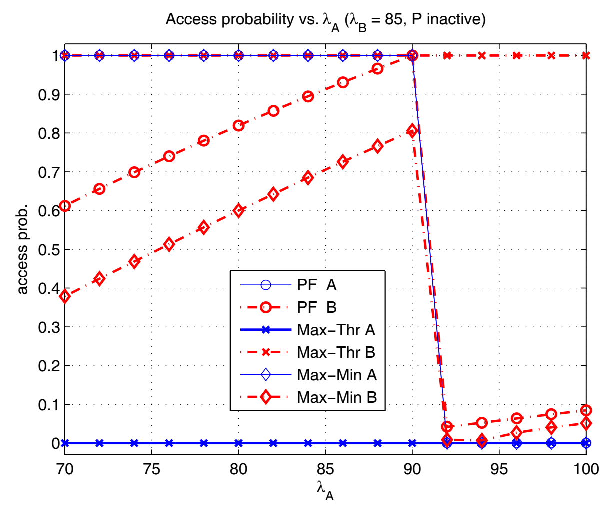


Fig. 8. Access probability vs. λ_A (asymmetric-interference, $\lambda_B = 85$ s^{-1}).

as λ_A increases, while U_B drops quickly, which is unfair.

In Fig. 7, we show the effect of λ_P on the average access probability. In this set of simulations, λ_B is still set as 85 s^{-1} and we vary λ_A from 70 to 100 s^{-1}. We know from Fig. 5 that the user with the higher access rate has a zero access probability when the other user is in service. Therefore, we only demonstrate the nonzero access probability of the user with a lower access rate, i.e., we show user A's access probability when $\lambda_A < \lambda_B = 85$ s^{-1} and user B's access probability when $\lambda_A > \lambda_B = 85$ s^{-1}. In Fig. 7, we compare the average access probability when λ_P is chosen from $\{90, 80, 70\}$ s^{-1}. We know that as λ_P increases, the competition between the secondary users becomes more severe. In order to reduce mutual interference, when λ_A is a fixed value, both users' access probabilities decrease as λ_P becomes larger.

B. CTMC-8 for the Asymmetric-Interference Case

In the second set of simulations, the transmitter of user A is at $(0\text{m}, 0\text{m})$, and its receiver is at $(200\text{m}, 0\text{m})$. The transmitter of user B is at $(185\text{m}, 460\text{m})$, and its receiver is at $(15\text{m}, 460\text{m})$. Under these settings, we have $r_1^B > r_1^A > r_2^B > r_2^A$ from (1) and (2), so the interference is asymmetric. In Fig. 8, we show the optimal access probabilities versus λ_A for each secondary user when the other is transmitting. Since user A has a worse channel condition than user B, for the maximal-throughput optimization, user A's access probability is 0 (e.g., user A's requests are always rejected) when user B is in service, which is unfair. For the PF or max-min fairness optimization, when $\lambda_A < \lambda_B = 85$ s^{-1}, user A's access probability is 1 (e.g., user A's requests are always admitted), while only a part of B's requests are admitted, due to fairness concerns. When λ_A is a little greater than λ_B, unlike the symmetric-interference case, user A's probability is still 1 and higher than B's access probability, because user A has a worse channel condition than B. When λ_A exceeds 90 s^{-1}, the chance of co-existence is so high that the access probabilities for both users drop to avoid interference.

In Fig. 9, we show the average throughput for each secondary user. We know from Fig. 8 that in the maximal-throughput optimization, user A's access probability is 0 and user B's access probability is 1; therefore, U_B is much greater than U_A. The PF optimization greatly reduces the throughput difference between the two users, with only a small loss of total throughput.

C. Comparison with a CSMA-based Scheme

In Fig. 10, we show the overall throughput of the PF dynamic spectrum access for CTMC-8, CTMC-5, and the overall throughput for a CSMA-based scheme [24]. The transmitters for both secondary users are uniformly located in a 200m $\times$ 200m square area, the distance between each transmitter-receiver pair is uniformly distributed in $[100\text{m}, 200\text{m}]$, and the other parameters are the same as in the previous setting. We choose the slotted version of the nonpersistent CSMA to avoid frequent collisions assuming the secondary users experience severe contention for the licensed spectrum, and the slot size is 0.005. So when primary user P is absent and one secondary user γ is transmitting, the later-coming secondary user senses the spectrum in every $0.005/\mu_\gamma$ s until the licensed spectrum is available again.

We can see that the PF access for both CTMCs have better performance than CSMA-based scheme as λ_A increases. This is because in CSMA, the secondary users cannot utilize the spectrum at the same time. Thus, even though interference exists when secondary users share the spectrum, by allowing spectrum sharing between them and optimally controlling their access probabilities, performance gain can still be achieved.

As λ_A increases, the overall throughput of the PF access for both CTMCs increases, while the throughput of CSMA-based scheme decreases. When $\lambda_A = 100$ s^{-1}, CTMC-5 can achieve about 50% throughput gain over CSMA, and CTMC-8 can achieve more than 95% throughput gain. This shows that the proposed PF access approach has a larger capability than CSMA to accommodate more traffic. Moreover, the spectrum efficiency of CTMC-8 is higher than that of CTMC-5, due to queuing of the interrupted service.

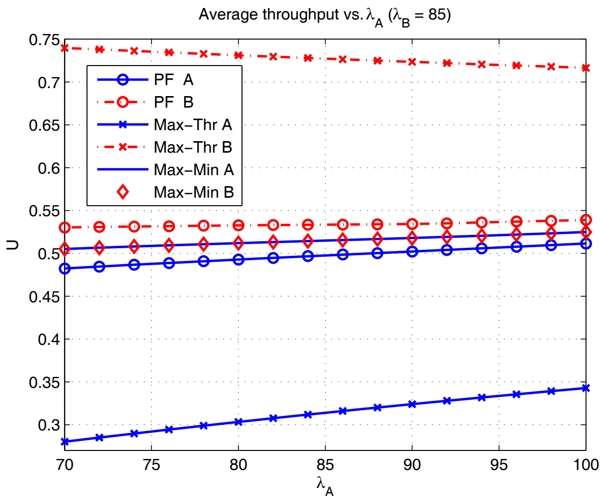


Fig. 9. Average throughput vs. λ_A (asymmetric-interference, $\lambda_B = 85$ s^{-1}).

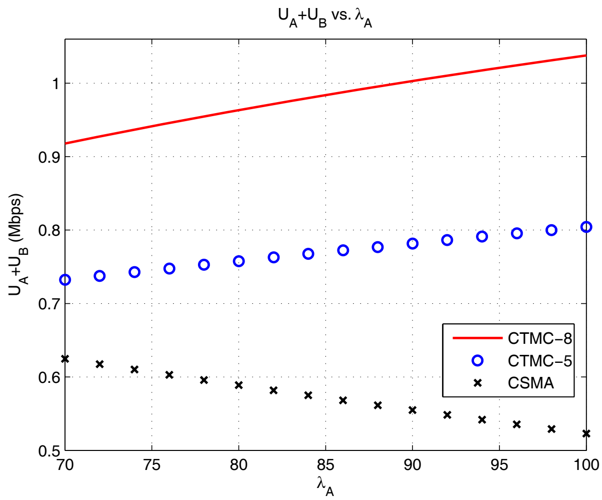


Fig. 10. Overall throughput for CTMC-5, CTMC-8 and CSMA.

D. Comparison with a Uniform-Access-Probability Scheme

In [12], we have proposed a uniform access probability for each secondary user no matter what state the CTMC is in. However, when the licensed spectrum is idle, the access probability may restrain full spectrum utilization. Moreover, the interference condition for one secondary user is varying when different subsets of secondary users share the spectrum. Only optimizing one single access probability may result in a sub-optimal solution. In this subsection, we conduct simulations to compare the scheme proposed in this paper with the one in [12]. In the comparison, we adopt the PF method, while the transmission power, request/service rates, and the locations of the secondary users are all uniformly distributed in a proper interval, and we test 1000 independent experiments to get the average. The histogram of the performance gain (U_{PF}) is shown in Fig. 11. We see that the proposed scheme in this paper with state-dependent access probability achieves on average a 24% higher system throughput than the scheme using a uniform access probability in [12].

E. Spectrum Sharing Among Multiple Secondary Users

Spectrum access with multiple secondary users can also be optimally controlled using a method where the access probabilities are obtained with numerical search algorithms. The transmitter-receiver pair of each user is randomly distributed in a 200m $\times$ 200m square area, and the transmission power is randomly chosen between 1mW and 3mW. In Fig. 12, we compare the total throughput of the proposed PF spectrum access to that without access control (i.e., all service requests are admitted with probability one). By optimizing the access probabilities, the proposed scheme achieves 17% higher throughput on average, since the interference is successfully alleviated. We also see that as the number of competing secondary user increases, the average throughput for each user is greatly reduced, since the spectrum competition becomes much heavier and each user has a smaller spectrum share.

VI. Conclusion

In this paper, we propose a primary-prioritized Markov approach for dynamic spectrum access. We model the interactions between the primary users and the secondary users as continuous-time Markov chains, and optimize the state-dependent access probabilities for secondary users so that the spectrum resources can be efficiently and fairly shared by the secondary users in an opportunistic way without interrupting the primary usage. The simulation results show that the proposed spectrum access with PF criterion can achieve up to 95% performance gain over a CSMA-based random access approach, and also achieves the optimal tradeoff between efficient spectrum utilization and fairness.

In the current work, the spectrum access is coordinated by a secondary management point. A distributed algorithm to find the best access pattern with less measurement overhead and signaling would be interesting future work to address. Secondary users can distributively adapt their access probabilities according to observation about their own data throughput, and the approach proposed in this paper would then serve as an upper bound to the performance of such distributed algorithms.

VII. Appendix

A. Stationary Probabilities for CTMC without Queuing in Multi-User Case

- Notation: Let S_i denote state $(0, [n_N, \cdots, n_1])$, where $n_k \in \{0, 1\}, k = 1, \cdots, N$, and $i = \sum_{j=1}^{N} 2^{j-1} n_j$, S_{2^N} denote state $(1, [0, \cdots, 0])$, and $q_{ij} \triangleq q\{S_i \to S_j\}$ denote the transition rate from state S_i to S_j;
- Construct the generator matrix $\mathbf{Q} = [q_{ij}]$:
  1) for $S_i = (0, [n_N, \cdots, n_j, \cdots, n_1])$, where $i = 0, \cdots, 2^N - 1$, and $j = 1, \cdots, N$, $q\{(0, [n_N, \cdots, n_j, \cdots, n_1]) \to (0, [n_N, \cdots, 1 - n_j, \cdots, n_1])\} = \mu_j(n_j = 1)$, or $\lambda_j(n_j = 0)$; $q\{S_i \to S_{2^N}\} = \lambda_P$; $q_{ii} = -\sum_{j \neq i} q_{ij}$;
  2) $q\{S_{2^N} \to S_0\} = \mu_P$, $q\{S_{2^N} \to S_{2^N}\} = -\mu_P$;

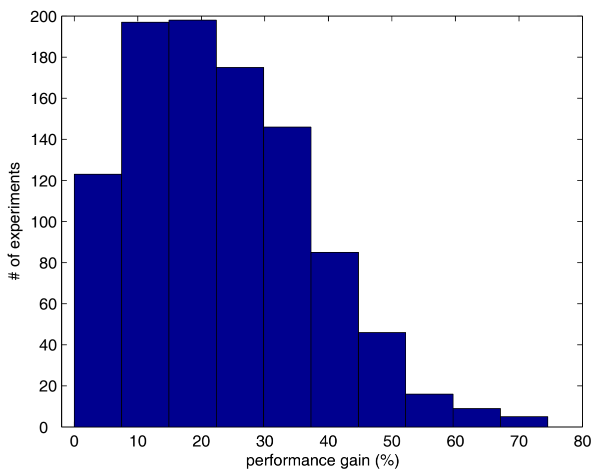


Fig. 11. The histogram of throughput improvement (U_{PF})

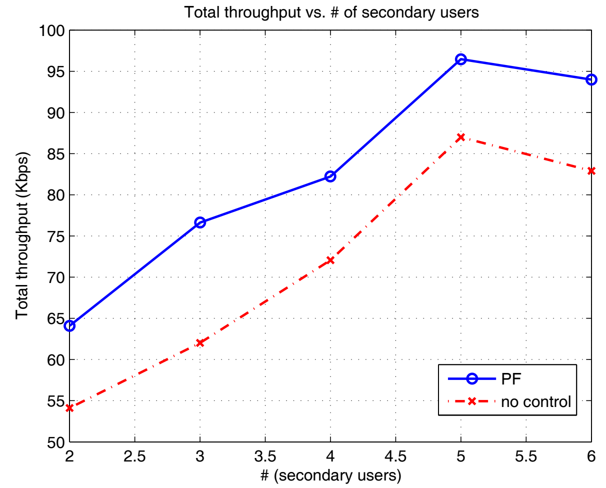


Fig. 12. Comparison of overall throughput for multiple secondary users.

- Solve the stationary probability $\mathbf{\Pi} = [\Pi_{S_0}, \cdots, \Pi_{S_{2N-1}}, \Pi_{S_{2N}}]$ from

$$\mathbf{Q}_{aug}\mathbf{\Pi}^T = \mathbf{b}, \tag{31}$$

where $\mathbf{Q}_{aug} = \begin{bmatrix} \mathbf{Q}^T \\ \mathbf{1}_{1\times(2^N+1)} \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} \mathbf{0}_{(2^N+1)\times 1} \\ 1 \end{bmatrix}$.

B. Stationary Probabilities for CTMC with Queuing in Multi-User Case

- Notation: Let S_i denote state $(0, [n_N, \cdots, n_1])$, and S_i^w denote state $(1, [n_N, \cdots, n_1]^w)$.
- Construct the generator matrix $\mathbf{Q} = [q_{ij}]$:
  1) for $S_i = (0, [n_N, \cdots, n_j, \cdots, n_1])$, where $i = 0, \cdots, 2^N - 1$, and $j = 1, \cdots, N$, $q\{(0, [n_N, \cdots, n_j, \cdots, n_1]) \rightarrow (0, [n_N, \cdots, 1 - n_j, \cdots, n_1])\} = \mu_j(n_j = 1)$, or $\lambda_j(n_j = 0)$; $q\{(1, [n_N, \cdots, n_j, \cdots, n_1]^w) \rightarrow (1, [n_N, \cdots, 1 - n_j, \cdots, n_1]^w)\} = \lambda_j(n_j = 0)$;
  2) $q\{S_i \rightarrow S_i^w\} = \lambda_P$; $q\{S_i^w \rightarrow S_i\} = \mu_P$; $q_{ii} = -\sum_{j\neq i} q_{ij}$;
- Solve the equation array similar to (31).

C. Proof of Equivalence of Eqns. (21) and (24)

Since the function of $\ln$ is monotonic, the PF-based utility defined in (21) is equivalent to

$$\sum_{\gamma\in S} \ln U_\gamma(\mathbf{a}_A, \mathbf{a}_B). \tag{32}$$

Define $\tilde{U}_\gamma = \ln U_\gamma$, then the gradient of $\tilde{U}_\gamma$ at the PF utility U_γ^* is $\frac{\partial \tilde{U}_\gamma}{\partial U_\gamma}\Big|_{U_\gamma^*} = \frac{1}{U_\gamma^*}$.

Since the PF utility U_γ^* optimizes (32), for a small feasible perturbation from the PF utility, we can omit the high-order polynomials in the Taylor series, apply first-order Taylor approximation, and obtain the following condition:

$$\sum_\gamma \frac{\partial \tilde{U}_\gamma}{\partial U_\gamma}\Big|_{U_\gamma^*}(U_\gamma - U_\gamma^*) = \sum_\gamma \frac{U_\gamma - U_\gamma^*}{U_\gamma^*} \leq 0, \tag{33}$$

Since the feasible region for U_γ is a convex set and the logarithm function (32) is strictly concave, (33) holds for any point deviating from the PF utility. Therefore, the definition of the PF criterion in (21) and (24) is equivalent.

VIII. Acknowledgments

The authors would like to thank anonymous reviewers for their valuable comments and feedback. Meanwhile, the first author would like to thank Mahmoud Abdulrehem for helpful discussion.

References

[1] FCC, "Spectrum policy task force report," FCC Document ET Docket No. 02-135, Nov. 2002.

[2] FCC, "Facilitating opportunities for flexible, efficient and reliable spectrum use employing cognitive radio technologies: notice of proposed rule making and order," FCC Document ET Docket No. 03-108, Dec. 2003.

[3] J. Mitola III, "Cognitive radio: an integrated agent architecture for software defined radio," Ph.D. thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2000.

[4] L. Cao and H. Zheng, 'Distributed spectrum allocation via local bargaining," in *Proc. IEEE SECON 2005*, pp. 475-486, Santa Clara, CA, Sept. 2005.

[5] H. Zheng and L. Cao, "Device-centric spectrum management," in *Proc. IEEE DySPAN 2005*, pp. 56-65, Baltimore, MD, Nov. 2005.

[6] R. Etkin, A. Parekh, and D. Tse, "Spectrum sharing for unlicensed bands," *IEEE J. Select. Areas Commun.*, vol. 25, no. 3, pp. 517-528, Apr. 2007.

[7] Z. Ji and K. J. R. Liu, "Dynamic spectrum sharing: a game theoretical overview," *IEEE Commun. Mag.*, pp. 88-94, May 2007.

[8] Z. Ji and K. J. R. Liu, "Belief-assisted pricing for dynamic spectrum allocation in wireless networks with selfish users," in *Proc. IEEE SECON 2006*, pp. 119-127, Reston, VA, Sept. 2006.

[9] C. Raman, R. D. Yates, and N. B. Mandayam, "Scheduling variable rate links via a spectrum server," in *Proc. IEEE DySPAN 2005*, pp. 110-118, Baltimore, MD, Nov. 2005.

[10] Y. Xing, R. Chandramouli, S. Mangold, and S. N. Shankar, "Dynamic spectrum access in open spectrum wireless networks," *IEEE J. Select. Areas Commun.*, vol. 24, no. 3, pp. 626-637, Mar. 2006.

[11] S. Keshavamurthy and K. Chandra, "Multiplexing analysis for spectrum sharing," in *Proc. IEEE MILCOMM'06*, pp. 1-7, Washington, DC, Oct. 2006.

[12] B. Wang, Z. Ji, and K. J. R. Liu, “Primary-prioritized Markov approach for dynamic spectrum access," in *Proc. IEEE DySPAN 2007*, pp. 507-515, Dublin, Ireland, Apr. 2007.

[13] FCC, “Establishement of interference temperature metric to quantify and manage interference and to expand available unlicensed operation in certain fixed mobile and satellite frequency bands," ET Docket 03-289, Notice of Inquiry and Proposed Rulemaking, 2003.

[14] T. Clancy, “Achievable capacity under the interference temperature model," in *Proc. IEEE INFOCOM 2007*, pp. 794-802, Anchorage, AK, May 2007.

[15] T. Clancy, “Formalizing the interference temperature model," *Wiley J. Wireless Commun. and Mobile Computing*, vol. 7, no. 9, pp. 1077-1086, Nov. 2007.

[16] IEEE 802.22 working group on wireless regional area networks [Online]. Available: http://www.ieee802.org/22/.

[17] —, “Improving the sharing of the radio spectrum: final report" [Online]. Available: http://www.ofcom.org.uk/research/technology/overview/ese/share/.

[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1990.

[19] V. G. Kulkarni, *Modeling and Analysis of Stochastic Systems*. CRC Press, 1995.

[20] F. Kelly, “Charging and rate control for elastic traffic," *European Trans. Telecommun.*, vol. 8, no. 1, pp. 33-37, Jan. 1997.

[21] Z. Han, Z. Ji, and K. J. R. Liu, “Fair multiuser channel allocation for OFDMA networks using Nash bargaining solutions and coalitions," *IEEE Trans. Commun.*, vol. 53, no. 8, pp. 1366-1376, Aug. 2005.

[22] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Prentic-Hall, 1992.

[23] —, IEEE Std 802.11a-1999, part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications.

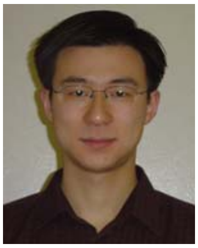
[24] L. Kleinrock and F. Tobagi, “Packet switching in radio channels: part I-carrier sense multiple-access modes and their throughput-delay characteristics," *IEEE Trans. Commun.*, vol. com-23, no. 12, pp. 1400-1416, Dec. 1975.

[25] M. A. McHenry, “NSF spectrum occupancy measurements project summary," Aug. 2005.

[26] Z. Haas, J. Halpern, L. Li, and S. B. Wicker, “A decision-theoretic approach to resource allocation in wireless multimedia networks," in *Proc. Dial M for Mobility*, pp. 86-95, 2000.

Beibei Wang (S’07) received her B.S. degree in electrical engineering (with highest honor) from the University of Science and Technology of China, Hefei, China, in July 2004. She is currently a PhD student in the Department of Electrical and Computer Engineering, University of Maryland, College Park.

Her research interests include resource allocation and management in cognitive radio systems, wireless communications and networking, game theory, wireless multimedia, and security. Ms. Wang received the Graduate School Fellowship from the University of Maryland in 2004–2006.



Zhu Ji received a Ph.D. degree in Electrical and Computer Engineering from University of Maryland, College Park in May 2007. He received B.S. and M.S. degrees in Electronic Engineering from Tsinghua University, Beijing, China, in 2000 and 2003, respectively.

He joined Qualcomm, San Diego, CA in 2007. From 2003 to 2007, he was a graduate research assistant in the Communication and Signal Processing Laboratory, University of Maryland, College Park. From 2000 to 2002, he was a visiting student (research intern) in the Wireless and Networking Group at Microsoft Research Asia, Beijing, China. His research interests are in wireless communications and networking.

K. J. Ray Liu (F’03) is Distinguished Scholar-Teacher of University of Maryland, College Park. He leads the Maryland Signals and Information Group conducting research encompassing broad aspects of information technology including communications and networking, information forensics and security, multimedia signal processing, biomedical imaging, and bioinformatics.

Dr. Liu is the recipient of numerous honors and awards including best paper awards from IEEE Signal Processing Society (twice), IEEE Vehicular Technology Society, and EURASIP; IEEE Signal Processing Society Distinguished Lecturer, EURASIP Meritorious Service Award, and National Science Foundation Young Investigator Award. He also received various teaching and research recognitions from University of Maryland including university-level Invention of the Year Award, and both the Outstanding Research Award and Poole and Kent Senior Faculty Teaching Award from A. James Clark School of Engineering.

Dr. Liu is Vice President – Publications and on the Board of Governor of IEEE Signal Processing Society. He was the Editor-in-Chief of IEEE SIGNAL PROCESSING MAGAZINE and the founding Editor-in-Chief of EURASIP JOURNAL ON APPLIED SIGNAL PROCESSING.

His recent books include *Cooperative Communications and Networking*, Cambridge University Press, 2008; *Resource Allocation for Wireless Networks: Basics, Techniques, and Applications*, Cambridge University Press, 2008; *Ultra-Wideband Communication Systems: The Multiband OFDM Approach*, IEEE-Wiley, 2007; *Network-Aware Security for Group Communications*, Springer, 2007; *Multimedia Fingerprinting Forensics for Traitor Tracing*, Hindawi, 2005.

T. Charles Clancy is a senior researcher with the Laboratory for Telecommunications Sciences, and an adjunct professor of Electrical Engineering at the University of Maryland. He received his M.S. in Electrical Engineering from the University of Illinois and Ph.D. in Computer Science from the University of Maryland. He chairs the HOKEY working group and is the security advisor to the CAPWAP and EAP working groups of the IETF. His research interests include next-generation wireless networks and security.