Primary-Prioritized Markov Approach for Dynamic Spectrum Access

Beibei Wang, Zhu Ji, and K. J. Ray Liu
Department of Electrical and Computer Engineering and Institute for Systems Research
University of Maryland, College Park, MD 20742, USA

***Abstract*— In order to fully utilize the scarce spectrum resources, with the development of cognitive radio technologies, dynamic spectrum access becomes a promising approach to increase the efficiency of spectrum usage. The spectrum access can be designed in an opportunistic way to efficiently and fairly share the spectrum resources among multiple unlicensed users, while not disturbing primary users' spectrum usage. In this paper, we propose a primary-prioritized Markov approach for dynamic spectrum access through modeling the interactions between the primary users and the unlicensed users as continuous-time Markov chains (CTMC). By designing appropriate access probabilities for the unlicensed users, the spectrum dynamics can be captured using CTMC models to optimally coordinate the spectrum access of the unlicensed users so that 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 not only provides fair spectrum sharing among unlicensed users with only small performance degradation compared to the approach maximizing the overall average throughput, but also achieves much higher throughput than CSMA-based random access approaches and the approach achieving max-min fairness.**

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 policy by FCC assigns the 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 recent development of dynamic spectrum access. By exploiting the spectrum in an opportunistic fashion, dynamic spectrum access enables the unlicensed 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, how to efficiently and fairly share the spectrum among the unlicensed users becomes an important issue, especially when multiple and dissimilar unlicensed users coexist in the same portion of spectrum band. There are several previous works addressing this issue, on a negotiated/pricing basis [1]-[6] or an opportunistic basis [7], [8]. Local bargaining mechanism is proposed in [1] to distributively optimize the efficiency of spectrum allocation and maintain bargaining fairness among unlicensed users. In [2], auction mechanisms are proposed for sharing spectrum among multiple users such that the interference is below certain level. Rule-based approaches are proposed in [3] that regulate users' spectrum access to tradeoff fairness and utilization with communication costs and algorithm complexity. In [4], the authors propose a repeated game approach to enlarge the set of achievable rates, in which the spectrum sharing strategy can be enforced by the Nash Equilibrium of dynamic games. In [5] and [6], the authors propose a belief-based dynamic pricing approach to optimize overall spectrum efficiency while keeping the participating incentives of the selfish users based on double auction rules. Recently, attentions are being drawn to opportunistic spectrum sharing. In [7], a distributed random access protocol is proposed to achieve airtime fairness between dissimilar unlicensed users. In [8], based on a multi-server queueing model, feasible operating regimes are rendered for the coexistence of primary and secondary users in the shared spectrum.

Although the existing dynamic spectrum access schemes have achieved some success on enhancing spectrum efficiency, most of them focus on spectrum allocation among unlicensed users without taking the primary users' activities into consideration, or specific for static topologies. However, the radio spectrum environment is constantly changing. For example, if a primary user appears in some specific portion of the spectrum bands, the unlicensed users existing in that spectrum band need to vacate the channels and try to transfer their communications to other available bands. Using the global optimization approach, after each change of the spectrum environment, the network needs to re-optimize the spectrum allocation for all users completely, resulting in high computational complexity. Furthermore, besides maximizing the overall spectrum utilization, a good spectrum sharing scheme should also achieve fairness among dissimilar users. In addition, if multiple unlicensed users are allowed to access the licensed spectrum resources, how to coordinate their access to alleviate interference with each other unlicensed user and avoid conflict with the primary users should be carefully considered.

Motivated by the above reasons, in this paper, we propose a primary-prioritized Markov approach for dynamic spectrum access. Specifically, we propose to model the interactions between the legacy spectrum holders (primary users) and the unlicensed users as continuous-time Markov chains (CTMC), by which we can capture the dynamics of system's evolvements, especially the effect of the primary user's activities on the unlicensed users. Further, in order to coordinate the spectrum access of unlicensed users in a more 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 the unlicensed users is

1-4244-0663-3/07/$20.00 ©2007 IEEE

optimally coordinated by the secondary base station through designing optimal spectrum access probabilities. The contributions of the proposed primary-prioritized Markov approach for dynamic spectrum access are multi-fold. First, the dynamics and the limiting behaviors of the radio system's evolvements including the primary user's activities are thoroughly studied through CTMC models. Second, from a statistical point of view, the proportional fairness (PF) based access approach optimally balances the tradeoff between the efficiency and fairness of dynamic spectrum access compared with the approach maximizing the total throughput or achieving the max-min fairness. Third, the proposed approach can reduce the computational complexity in a dynamically changing radio spectrum environment.

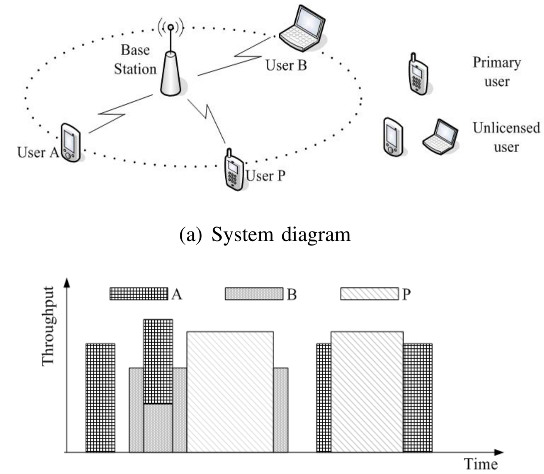
The remainder of this paper is organized as follows: The system model of the dynamic spectrum access is described in Section II. The primary-prioritized Markov approach for dynamic spectrum access is proposed in Section III. The simulation studies are provided in Section IV. Finally, Section V concludes this paper.

II. System Model

We consider the dynamic spectrum access networks where multiple unlicensed 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 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 [12] proposes to reuse the fallow TV spectrum without causing any harmful interference to incumbents (i.e., the TV receivers). With the development of cognitive radios, ancillary services carried by the Ultra-High Frequency (UHF) television transmission, e.g., reallocation of the spectrum to mobile applications are technically feasible. Moreover, with regard to more efficient utilization of some cellular bands, in [13], it is proposed 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.

In order to take advantage of the temporally unused spectrum holes (no device in service) in the licensed band, without loss of generality, we consider a snapshot of the above spectrum access networks shown in Fig. 1(a), where two unlicensed users and one primary user coexist and the unlicensed users opportunistically utilize the spectrum holes in the licensed band. Note that the system diagram shown here serves only as an example model and the scenario with multiple unlicensed 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. The arrival traffic is modeled as a Poisson process with rate λ_P ms^{-1}. The



(a) System diagram

(b) Throughput vs. time

Fig. 1: System model.

spectrum access duration is negative-exponentially distributed with mean time $1/\mu_P$ ms, so the departure of user P's traffic is another Poisson process with rate μ_P ms^{-1}.

The unlicensed users are denoted by A and B, respectively, and set S is defined as S $= \{A, B\}$. For each unlicensed user γ, where $\gamma \in$ S, its traffic pattern is similarly characterized by two independent Poisson processes, with arrival rate λ_γ ms^{-1} and departure rate μ_γ ms^{-1}, respectively. 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 operations of any other unlicensed user. In order not to disturb the primary user's spectrum usage, priority to access the spectrum is given to primary user P. We assume that the unlicensed users equipped with cognitive radios are capable to detect the primary user's activities, i.e., the appearance of the primary user in the spectrum band and its vacation from the spectrum. When primary user P appears, the unlicensed users should adjust their transmission parameters, for instance, reduce the transmit power or vacate the channels and try to transfer their communications to other available bands. The *interference temperature* model [9] is proposed by FCC that allows unlicensed users to transmit in licensed bands with carefully adjusted power, provided that unlicensed 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 unlicensed users to remain operating in the band with reduced power, the capacity they can achieve is very low [10], [11]. Therefore, in this paper, we assume that when primary user P appears, any unlicensed user should vacate and the traffic currently being served is cut off, e.g., dropped or stored in a buffer. In the duration of primary user P being served, any entry of the unlicensed user's traffic into the spectrum is denied until primary user P finishes its service.

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

For any unlicensed user γ that operates in the spectrum band alone, its channel capacity [14] can be represented by

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

where W is the communication bandwidth, n_0 is thermal noise power, p_{γ} is the transmission power for user γ, and $G_{\gamma\gamma}$ is the channel gain for user γ. The unlicensed users A and B are allowed to coexist in the spectrum band, however, there is interference at the receiver side of each user. In this scenario, the capacity of unlicensed user γ can be represented by

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

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 Approach for Dynamic Spectrum Access

In this section, we first derive a primary-prioritized Markov approach to study the dynamic spectrum access process for the primary user and the unlicensed users. Then, we propose dynamic spectrum access among the unlicensed users under different criteria, including the PF, maximal-throughput, and max-min fairness criteria.

A. Primary-Prioritized CTMC without Buffering

1) CTMC without Buffering: In dynamic spectrum access, where the unlicensed users opportunistically access the unused licensed spectrum, priority should be given to the primary user. That is, unlicensed users cannot operate in the spectrum when there exists the primary user; when the primary user appears in the spectrum band, all unlicensed 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 unlicensed users and the primary user as a primary-prioritized CTMC [15].

In the CTMC, when the unlicensed users contend to access the idle spectrum, collisions only occur when their packets arrive exactly at the same time. This case rarely happens for independent Poisson processes. Another type of collisions happens when some unlicensed user's packet arrives while the spectrum band is occupied by the primary user. In this case, collision avoidance techniques are assumed, such as random back-off, or CSMA/CA [20]. Therefore, in the CTMC model we omit the collision state.

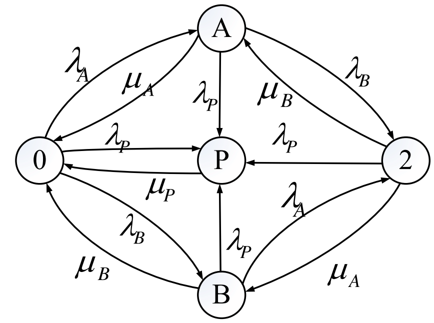


Fig. 2: The rate diagram of CTMC-5.

TABLE I: The Five States of CTMC-5

State	Description
0	No user operates in the spectrum
A	Unlicensed user A operates in the spectrum
B	Unlicensed user B operates in the spectrum
P	Primary user P operates in the spectrum
2	Both A and B operate in the spectrum

If we assume that when the primary user appears, unlicensed users' traffic is dropped, i.e., there is no buffering of the interrupted traffic, 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, and the states of CTMC-5 are described in Table I.

Assume at first the spectrum band is idle, i.e., CTMC-5 is in state 0. Unlicensed 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. If user A's service completes before any other user requests spectrum access, CTMC-5 then transits to state 0 with rate μ_A. If user B's traffic arrives before A completes its service, CTMC-5 transits to state 2 with rate λ_B, where both unlicensed users share the spectrum usage. 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). However, primary user P may, once in a while, appear during the service duration of the unlicensed users, i.e., when CTMC-5 is in state A, B or 2. At that time, the unlicensed user's traffic is dropped to avoid conflict with the primary user, and CTMC-5 transits to state P with rate λ_P. During the primary user operating in the spectrum band, no unlicensed user is given access to the spectrum. CTMC-5 transits to state 0 with rate μ_P 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 [15] equation array governing the above system is given by

$$\mu_A \Pi_A + \mu_P \Pi_P + \mu_B \Pi_B = (\lambda_A + \lambda_B + \lambda_P)\Pi_0, \quad (3)$$

$$\lambda_A \Pi_0 + \mu_B \Pi_2 = (\mu_A + \lambda_P + \lambda_B)\Pi_A, \quad (4)$$

$$\lambda_P(\Pi_0 + \Pi_A + \Pi_2 + \Pi_B) = \mu_P \Pi_P, \quad (5)$$

$$\lambda_B \Pi_0 + \mu_A \Pi_2 = (\mu_B + \lambda_P + \lambda_A)\Pi_B, \quad (6)$$

$$\lambda_B\Pi_A + \lambda_A\Pi_B = (\mu_B + \lambda_P + \mu_A)\Pi_2, \tag{7}$$

$$\Pi_0 + \Pi_A + \Pi_B + \Pi_P + \Pi_2 = 1, \tag{8}$$

where Π_{s_i} represents the stationary probability of being in state s_i, $s_i \in \Omega_S$, and $\Omega_S = \{0, A, B, P, 2\}$.

The solutions of our interest to the above equation array, i.e., the probabilities when the spectrum is occupied by either primary user P or the unlicensed users, are given by

$$\begin{aligned} \Pi_P &= \lambda_P(\lambda_P + \mu_P)^{-1}, \\ \Pi_A &= C_1\lambda_A[\lambda_B\mu_B + (\lambda_P + \mu_B) \\ &\quad (\lambda_A + \lambda_P + \mu_A + \mu_B)], \\ \Pi_B &= C_1\lambda_B[\lambda_A\mu_A + (\lambda_P + \mu_A) \\ &\quad (\lambda_B + \lambda_P + \mu_A + \mu_B)], \\ \Pi_2 &= C_1\lambda_A\lambda_B[\lambda_A + \lambda_B + 2\lambda_P + \mu_A + \mu_B], \end{aligned} \tag{9}$$

where, for simplicity, the coefficient C_1 is defined as

$$\begin{aligned} C_1 = (1 - \Pi_P)[(\lambda_A + \mu_A + \lambda_P)(\lambda_B + \mu_B + \lambda_P) \\ (\lambda_A + \mu_A + \lambda_B + \mu_B + \lambda_P)]^{-1}. \end{aligned} \tag{10}$$

One of the most important goals in spectrum sharing is efficient spectrum utilization, i.e., high throughput each unlicensed user can achieve by successfully acquiring a spectrum band. From a statistic point of view, the unlicensed users want to maximize their average throughputs. 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 equivalently viewed as the ratio of allocation time to state s_i to the entire reference time [7]. Thus, the product of Π_{s_i} and the capacity that unlicensed user γ achieves when operating in state s_i represents one average throughput component acquired by user γ in state s_i. Therefore, from CTMC-5, we can express the total average throughput for user γ as follows,

$$U_\gamma = \Pi_\gamma r_1^\gamma + \Pi_2 r_2^\gamma, \tag{11}$$

where Π_γ and Π_2 are as solved in (9), and r_1^γ and r_2^γ are defined in (1) and (2), respectively. The first term on the right-hand side of (11) represents the throughput when user γ occupies the spectrum alone, and the second term represents the throughput when two unlicensed users share the spectrum.

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

2) CTMC-5 in Multi-User Case: CTMC-5 can also be generalized to model the scenario with more than two unlicensed users. Suppose the set of N unlicensed 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 unlicensed users:

$$\begin{aligned} (\Phi_P, \Phi_{\mathbf{S}}) \in \mathcal{A} \triangleq \{(1, [0, \cdots, 0])\} \\ \bigcup\{(0, \phi_{\mathbf{S}}) : \phi_{\mathbf{S}} \triangleq [n_N, \cdots, n_1] \in \{0, 1\}^N\}, \end{aligned} \tag{12}$$

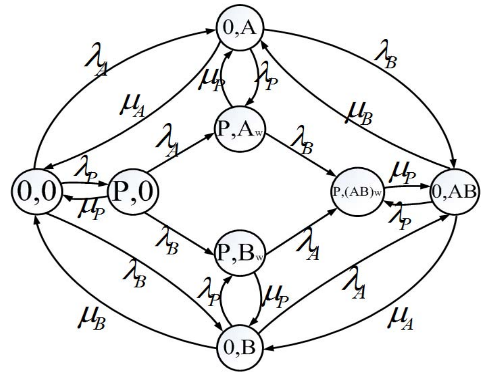


Fig. 3: The rate diagram of CTMC-8.

TABLE II: The Eight States of CTMC-8

Index	State	Description
0	$(0, 0)$	Spectrum is idle
1	$(0, B)$	Unlicensed user B is in service
2	$(0, A)$	Unlicensed user A is in service
3	$(0, AB)$	Both A and B are in service
4	$(P, 0)$	Primary user P is in service
5	(P, B_w)	P is in service; B is waiting
6	(P, A_w)	P is in service; A is waiting
7	$(P, (AB)_w)$	P is in service; A and B are waiting

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 unlicensed users are in service.

For this generalized Markov model, the state transition 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 unlicensed 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 equation array in Appendix VI-A.

For each unlicensed 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 $N - 1$ unlicensed users are in service. Since more unlicensed users contend the spectrum access, the contention in the generalized Markov model becomes heavier than CTMC-5. As a result, each unlicensed user shares less spectrum access in average. Moreover, the interference also increases by introducing more unlicensed users. Therefore, as the number of unlicensed users increases, the average throughput for each of them is reduced.

B. Primary-Prioritized CTMC with Buffering

1) CTMC with Buffering: In CTMC-5 presented in Section III-A, the traffic of the unlicensed users is dropped when primary user P appears in the spectrum band, or when primary user P is operating in the spectrum band. After primary P completes its service, CTMC-5 will transit to the idle state. However, there may be some time interval wasted when the system is in the idle state until the next unlicensed user accesses the spectrum. In order to further increase the spectrum utilization and smooth service for unlicensed users, buffering of the unlicensed user' traffic due to primary user's presence is considered in the subsection. More specifically, when the spectrum is being occupied by some unlicensed users, upon the appearance of primary user P, the unlicensed users store their interrupted traffic in some buffer, continue scanning the spectrum band and immediately resume transmission once the spectrum band becomes idle again. Also, if the primary user begins to operate in the previously idle spectrum, newly coming traffic of the unlicensed user is also stored in some buffer.

Further considering the above factors, we model the spectrum access with buffering of the interrupted traffic for unlicensed users as an eight-state CTMC, denoted by "CTMC-8" for short. The transition diagram of CTMC-8 is shown in Fig. 3, and the eight states of CTMC-8 are described in Table II. Compared to CTMC-5 and its dynamics, in CTMC-8, three additional states are introduced, i.e., states (P, A_w), (P, B_w) and $(P,(AB)_w)$. Their transitions are described as follows. When the spectrum band is occupied by some unlicensed user, e.g., user A, if A detects that primary user P needs to acquire the spectrum band, it stores the unfinished traffic in some buffer, and CTMC-8 transits from state $(0, A)$ to state (P, A_w) with rate λ_P. 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. If, in contrast, the traffic of B arrives before primary user P completes its service duration, B also stores its traffic in some buffer, and CTMC-8 transits to state $(P,(AB)_w)$ with rate λ_B. 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, where A and B share the spectrum band. Also, when CTMC-8 is in state $(P, 0)$, if unlicensed users attempt to access the spectrum, they also store their traffic in some buffer, and CTMC-8 transits to state (P, A_w) or state (P, B_w), with rate λ_A or λ_B, respectively.

The equation array governing the above system is given by

$$(\lambda_A + \lambda_B + \lambda_P)\Pi_0 = \mu_B\Pi_1 + \mu_A\Pi_2 + \mu_P\Pi_4, \tag{13}$$

$$(\mu_B + \lambda_P + \lambda_A)\Pi_1 = \lambda_B\Pi_0 + \mu_A\Pi_3 + \mu_P\Pi_5, \tag{14}$$

$$(\mu_A + \lambda_P + \lambda_B)\Pi_2 = \lambda_A\Pi_0 + \mu_B\Pi_3 + \mu_P\Pi_6, \tag{15}$$

$$(\mu_A + \mu_B + \lambda_P)\Pi_3 = \lambda_A\Pi_1 + \lambda_B\Pi_2 + \mu_P\Pi_7, \tag{16}$$

$$(\mu_P + \lambda_A + \lambda_B)\Pi_4 = \lambda_P\Pi_0, \tag{17}$$

$$(\mu_P + \lambda_A)\Pi_5 = \lambda_P\Pi_1 + \lambda_B\Pi_4, \tag{18}$$

$$(\mu_P + \lambda_B)\Pi_6 = \lambda_P\Pi_2 + \lambda_A\Pi_4, \tag{19}$$

$$\mu_P\Pi_7 = \lambda_B\Pi_6 + \lambda_P\Pi_3 + \lambda_A\Pi_5, \tag{20}$$

$$\Pi_0 + \Pi_1 + \Pi_2 + \Pi_3 + \Pi_4 + \Pi_5 + \Pi_6 + \Pi_7 = 1. \tag{21}$$

The solutions can be obtained in a similar way as in Section III-A. Therefore, by using CTMC-8, which introduced buffering for unlicensed users' traffic, spectrum utilization is further increased compared to CTMC-5, meanwhile, unlicensed users' service is smoothed.

2) CTMC-8 in Multi-User Case: CTMC-8 can also be generalized to model the scenario with more than two unlicensed users. For the Markov chain with a set $\mathbf{S} = \{1, \cdots, N\}$ of unlicensed users, the state space $\mathcal{B}$ consists of all possible 2^{N+1} combinations of the status for primary user P and the unlicensed users:

$$\begin{aligned}(\Psi_P, \Psi_\mathbf{S}) \in \mathcal{B} &\triangleq \{(1, \psi_\mathbf{S}^w) \\ &\bigcup (0, \psi_\mathbf{S}) : \psi_\mathbf{S} \triangleq [n_N, \cdots, n_1] \in \{0,1\}^N\},\end{aligned} \tag{22}$$

where $\{(1, \psi_\mathbf{S}^w)\}$ represents all 2^N states where the primary user is in service and zero up to N unlicensed users are waiting, and $\{(0, \psi_\mathbf{S})\}$ represents all 2^N states where primary user P is not in service and zero up to N unlicensed users are in service. The transition diagram for this model can be similarly drawn as the generalized Markov model for CTMC-5 in Section III-A, while more states representing $(1, \psi_\mathbf{S}^w)$ and corresponding transitions representing their dynamics need to be added.

As more unlicensed users contend the spectrum, besides increased interference, more waiting time is also introduced to every unlicensed user in average. As a result, each unlicensed user achieves a smaller average throughput.

C. Primary-Prioritized Dynamic Spectrum Access

In order to develop the primary-prioritized dynamic spectrum access, it is important to first analyze the behaviors of the unlicensed users. Since the unlicensed users contend for the spectrum, if they access the spectrum in such a greedy manner that all of their injected traffic is admitted, then the Markov chain is more likely to be in the state where more than one users share the spectrum. Hence, the unlicensed 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 unlicensed users reduce their arrival rates too much so as to avoid interference, the average throughput may be unnecessarily low. Therefore, the access to the spectrum of the unlicensed users should be carefully controlled.

In the proposed dynamic spectrum access, we introduce the spectrum access probability for user A and user B, denoted by a_A and a_B, respectively. Then, 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 is Fig. 4. Because each unlicensed user γ's traffic is admitted with probability a_γ, the actual arrival rate is approximated by $a_\gamma\lambda_\gamma$. Correspondingly, the equation array governing the CTMC and the stationary probabilities in (9) are modified by replacing all λ_γ terms with $a_\gamma\lambda_\gamma$.

Then, the optimization goal is to determine a_A and a_B, such that the utility function can be maximized, i.e.,

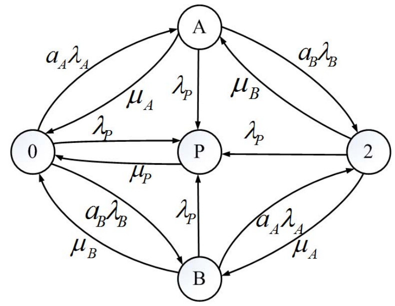


Fig. 4: Modified CTMC-5 with access control.

$$\max_{a_A,\ a_B} \ U(a_A, a_B) \quad \text{s.t. } U_\gamma(a_A, a_B) \geq 0, 0 \leq a_\gamma \leq 1, \tag{23}$$

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

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 [17]. Thus, in (23), $U(a_A, a_B)$ can be written as

$$U(a_A, a_B) = \prod_{\gamma \in S} U_\gamma(a_A, a_B). \tag{24}$$

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

$$U(a_A, a_B) = \sum_{\gamma \in S} U_\gamma(a_A, a_B), \tag{25}$$

and max-min fairness criterion

$$U(a_A, a_B) = \min_{\gamma \in S} U_\gamma(a_A, a_B). \tag{26}$$

For the maximal-throughput optimization, the overall system throughput is maximized. However, since the channel condition for one unlicensed user may be better than that of the other, the user with worse channel condition may starve. For the max-min fairness optimization, the performance of the unlicensed user with worse channel condition is maximized, and strict fairness is guaranteed. However, this criterion penalizes the user with better channel conditions, and thus results in inferior overall system performance. In this paper, we prefer the PF dynamic spectrum access because it can ensure more fairness than the maximal-throughput optimization, while achieve much 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 [17], i.e.,

$$\sum_{\gamma \in S} \frac{U_\gamma(a_A, a_B) - U_\gamma^*(a_A, a_B)}{U_\gamma^*(a_A, a_B)} \leq 0, \tag{27}$$

where $U_\gamma^*(a_A, a_B)$ is the proportionally fair throughput distribution, and $U_\gamma(a_A, a_B)$ is any other feasible throughput distribution for user γ.

Proved in a similar way as in [18], the optimal solution $U_\gamma^*(a_A, a_B)$ defined in (27) can be obtained by solving (23), with $U(a_A, a_B)$ being defined in (24). With the following lemma, we can employ the first-order optimality condition to solve the optimal spectrum access probabilities for unlicensed users in the PF optimization.

Lemma: The $U_{PF}(a_A, a_B)$ defined in (24) for the dynamic spectrum access using PF criterion is concave in a_γ, with $0 \leq a_\gamma \leq 1$, $\forall \gamma \in$ S, and S $= \{A, B\}$.

Proof: Please refer to Appendix VI-B. ■

If $\forall a_\gamma = 0$, neither user A nor user B can access the spectrum band, then we have $U_{PF}(a_A, a_B) = 0$; while $a_\gamma \neq 0$ results in $U_{PF}(a_A, a_B) > 0$. Thus, $a_\gamma = 0$ is not the optimal solution to maximize $U_{PF}(a_A, a_B)$. Therefore, the optimal access probabilities can be expressed as

$$a_\gamma^{opt} = \min\{a_\gamma^*, 1\}, \tag{28}$$

where a_γ^* is the solution to the following equations

$$\frac{\partial U_{PF}(a_A, a_B)}{\partial a_\gamma} = 0, \quad \forall \gamma \in \text{S}. \tag{29}$$

To implement the primary-prioritized dynamic spectrum access, without loss of generality, we assume that there exists a secondary base station that acts as a centralized management point to perform the optimization over a_A, a_B, and manage the spectrum resources.

The proposed primary-prioritized Markov approach for dynamic spectrum access shares some similarity with the conventional medium access control (MAC) protocols, since they all target for appropriate coordination of different users' access to the medium. For instance, in 802.11 standard [20], 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 re-schedule 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 it allows 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 idle medium, their packets suffer significant delay and may expire.

In the proposed primary-prioritized dynamic spectrum access, different unlicensed users are allowed to share the spectrum band simultaneously using CDMA techniques. This will increase the spectrum utilization because of the following reasons. First, for independent Poisson processes, the service durations of different unlicensed 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 unlicensed 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 force to re-transmit its packet after a random

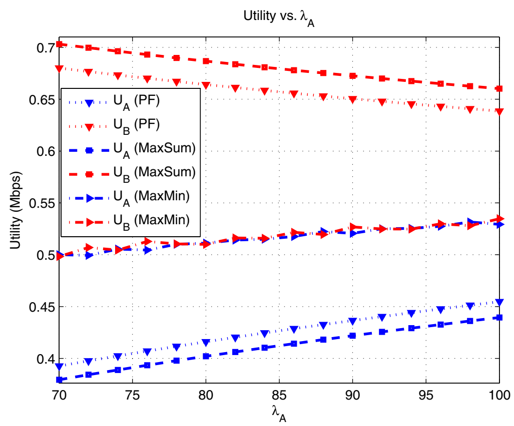


Fig. 5: Each unlicensed user's throughput (Mbps) vs. λ_A.

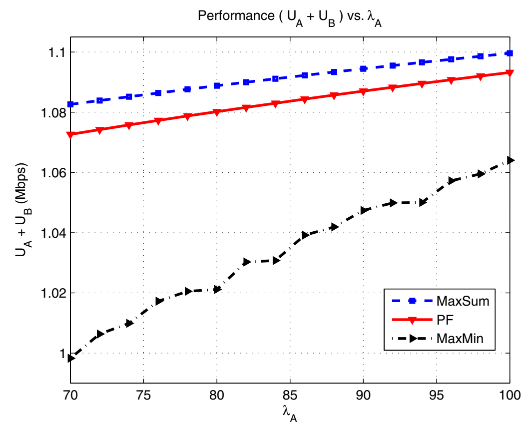


Fig. 6: Overall throughput (Mbps) vs. λ_A.

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 unlicensed users. By doing this, the interference is maintained at a low level.

In addition, from the solutions in (9), the stationary probability of the primary user operating in the spectrum band equals to the probability where the offered traffic for the primary user is simply modeled as an M/M/1 queue [19] and no unlicensed user is allowed to access the unused spectrum. This reflects that in the proposed primary-prioritized CTMC, the primary user is entitled to absolute priority to access the spectrum and its activities are not interrupted.

Also, in a mobile network, the radio spectrum environment is dynamically varying. Using global optimization approaches specific for fixed topologies, after each change of the spectrum environment, the network needs to re-optimize the spectrum allocation for all users completely, resulting high computational complexity. In the proposed approach, by controlling the access probabilities dynamically, computational complexity is reduced while the average throughput is maximized.

IV. Simulation Results and Analysis

In order to evaluate the performance of the proposed primary-prioritized dynamic spectrum access scheme, we first compare the performances of three different optimization goals (maximal-throughput, max-min, and PF) for CTMC-8. Then we compare the performances of CTMC-8, CTMC-5, and the nonpersistent CSMA based random access.

In the simulations, the communication bandwidth is 200 KHz, the transmission power is $p_\gamma = 2$ mW, and the thermal noise power is $n_0 = 10^{-15}$ W. The propagation loss factor is 3.6. The transmitter of user A is at $(0\text{m}, 0\text{m})$, and its receiver is at $(150\text{m}, 0\text{m})$. The transmitter of user B is at $(300\text{m}, 0\text{m})$, and its receiver is at $(400\text{m}, 0\text{m})$. The service rates of P, A and B are all set to be 100 ms^{-1}. The arrival rates of P and B are both chosen as 85 ms^{-1}, while the arrival rate of A varies from 70 to 100 ms^{-1}.

In Fig. 5, we show the utilities of both user A and B versus λ_A of the PF, maximal-throughput, and max-min optimizations for CTMC-8. Since $r_1^B > r_1^A > r_2^B > r_2^A$, for the maximal-throughput optimization, user B has a higher throughput. When $\lambda_A < \lambda_B$, the difference in the throughput is very large. As λ_A increases, A has more chance to access the spectrum, so user A's throughput gradually increases. However, the interference from A to B also increases, so user B suffers a throughput degradation. For the max-min optimization, both users almost have the same throughput, which increases as λ_A increases. This is because the system has to accommodate the user with worse channel condition, and also user A has an increasing spectrum allocation. For the PF optimization, the difference between A and B's throughput is smaller than that of the maximal-throughput optimization. Also, the increment of A's throughput is larger than the decrement of B's throughput as λ_A increases. This shows that the PF optimization is more fair than the maximal-throughput method.

In Fig. 6, we show the overall throughput of two unlicensed users versus λ_A with different optimization goals for CTMC-8. Because the max-min method compensates the user with worse channel condition, its has the worst performance, especially when user A further suffers a lower access when λ_A is very small. The PF method has the performance between the maximal-throughput method and max-min method, while the maximal-throughput method is less fair. In addition, the performance loss of the PF method to that of the maximal-throughput method is small. Therefore, the primary-prioritized PF dynamic spectrum access is a good tradeoff between the fairness and efficiency.

In Fig. 7, we show the overall throughput of the PF dynamic spectrum access for CTMC-8, CTMC-5 and the throughput of the nonpersistent CSMA random access [21] versus λ_A. In CSMA, the propagation delay is set to be 0.005. We can see that PF access approaches for both the two CTMCs have better performances than the CSMA as λ_A increases. This is because in CSMA, two unlicensed users cannot utilize the spectrum at the same time, otherwise information symbols of neither user will be correctly decoded. Thus, even though interference exists when both users share the spectrum, by

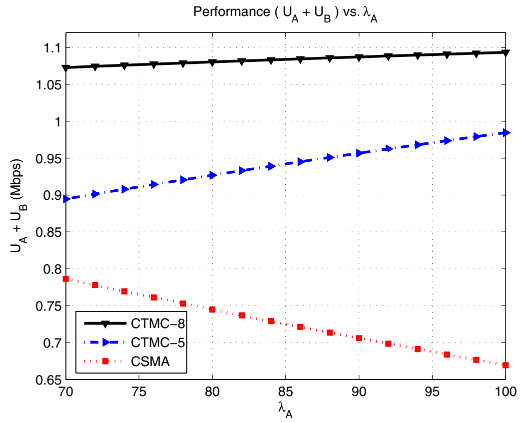


Fig. 7: Comparison of the overall throughput vs. λ_A for CTMC-5, CTMC-8 and CSMA.

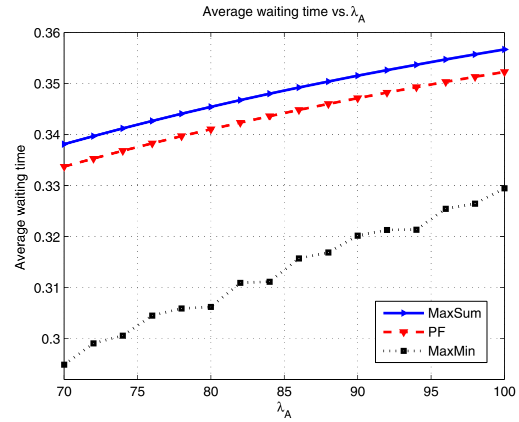


Fig. 8: Average waiting time ratio per unlicensed user vs. λ_A for CTMC-8.

optimally controlling the access probabilities of the unlicensed users, performance gain can still be achieved by the proposed approach. As λ_A increases, the overall throughput of the PF access approach for both the CTMCs increases, while the throughput of CSMA decreases. This shows that the proposed PF access approach has a larger capability than the CSMA approach to accommodate more traffic. As we mentioned in Section III-B, CTMC-8 stores the interrupted traffic in a buffer and immediately resumes it when primary user P completes its spectrum usage, while CTMC-5 simply drops the interrupted traffic and needs to wait for new incoming traffic. We can see that the spectrum utilization of CTMC-8 is higher than that of CTMC-5, since the arrival rates are identical for the two cases while CTMC-8 has a larger overall throughput. However, as λ_A increases, the performance increment of CTMC-8 compared to CTMC-5 gradually decreases. This is because, as shown in Fig. 8, more traffic injection for CTMC-8 results in larger average waiting time allocation for each interrupted unlicensed user and the actual spectrum allocation ratio satiates gradually. From Fig. 8, we also see that the maximal-throughput optimization has the highest average waiting time ratio, since in this spectrum access, the user with better channel condition, i.e., user B, is allowed more frequent spectrum access than that of the other two optimizations and experiences longer waiting. The max-min fairness optimization has the lowest waiting time ratio, because the access of user B are greatly limited to accommodate user A and experiences shorter waiting. The PF dynamic spectrum access has the waiting time ratio between the maximal-throughput and max-min spectrum access optimizations.

In order to evaluate the proposed dynamic spectrum access approach in a more general scenario, we also performed simulations where the transmitters and receivers of the unlicensed users are randomly located in a 400m $\times$ 400m area with the distance between each transmitter-receiver pair being kept 200m. In Fig. 9, we show the overall throughput of the PF dynamic spectrum access for CTMC-8, CTMC-5 and that of the CSMA-based random access protocol, averaged over random locations. We can observe that even under the randomized settings, the proposed dynamic spectrum access approaches still show better performances and larger capabilities to accommodate more traffic than the CSMA-based random access protocol. Also, CTMC-8 achieves more efficient spectrum utilization than CTMC-5 as in the case with fixed transmitter-receiver positions.

V. Conclusions and Future Work

In this paper, we propose a primary-prioritized Markov approach for dynamic spectrum access. In order to study the spectrum dynamics, the interactions between the primary users and the unlicensed users are modeled using continuous-time Markov chains. By studying the optimal spectrum access probabilities of unlicensed users, the spectrum resources can be efficiently and fairly shared by the unlicensed users in an opportunistic way without interrupting the spectrum usage of the primary users. The simulation results show that the proposed primary-prioritized dynamic spectrum access with PF criterion achieves efficient and fair spectrum access with low computational complexity.

There are several avenues for future research. We intend to further extend the current optimization of the access probabilities performed by a secondary base-station to a distributed implementation using a noncooperative game theoretical framework with limited local information observed by each unlicensed user. Also, we would like to consider a more general scenario where multiple primary users coexist with multiple unlicensed users and there are several disjoint spectrum bands to be opportunistically shared with all of the users.

VI. Appendix

A. Stationary Probabilities for CTMC-5 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 $j = 1, \cdots, N$,

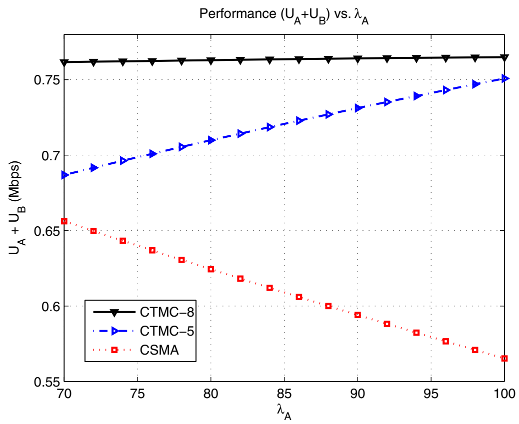


Fig. 9: Comparison of the overall throughput (Mbps) vs. λ_A for CTMC-5, CTMC-8 and CSMA in random locations.

$q\{(0,[n_N,\cdots,n_j,\cdots,n_1]) \rightarrow (0,[n_N,\cdots,1-n_j,\cdots,n_1])\} = \mu_j(n_j=1), \text{or } \lambda_j(n_j=0)$; $q\{S_i \rightarrow S_{2^N}\} = \lambda_P$; $q_{ii} = -\sum_{j\neq i} q_{ij}$;

2) $q\{S_{2^N} \rightarrow S_0\} = \mu_P$, $q\{S_{2^N} \rightarrow S_{2^N}\} = -\mu_P$;

- Solving the equation array

$$\mathbf{\Pi Q} = \mathbf{0}, \sum_{i=0}^{2^N} \Pi_{S_i} = 1, \tag{30}$$

we can obtain the stationary probability vector $\mathbf{\Pi} = [\Pi_{S_0},\cdots,\Pi_{S_{2^N-1}},\Pi_{S_{2^N}}]$.

B. Proof of lemma

Substituting (9) with all λ_γ terms replaced by $a_\gamma\lambda_\gamma$ into (11), we can write U_γ as

$$U_\gamma(a_A,a_B) = \Pi_\gamma(a_A,a_B)r_1^\gamma + \Pi_2(a_A,a_B)r_2^\gamma \quad , \tag{31}$$

where

$$\begin{aligned} \Pi_A(a_A,a_B) &= C_1 a_A\lambda_A[a_B\lambda_B\mu_B \\ &\quad + (\lambda_P+\mu_B)(a_A\lambda_A+\lambda_P+\mu_A+\mu_B)] \\ \Pi_B(a_A,a_B) &= C_1 a_B\lambda_B[a_A\lambda_A\mu_A \\ &\quad + (\lambda_P+\mu_A)(a_B\lambda_B+\lambda_P+\mu_A+\mu_B)], \\ \Pi_2(a_A,a_B) &= C_1 a_A\lambda_A a_B\lambda_B \\ &\quad \times (a_A\lambda_A+a_B\lambda_B+2\lambda_P+\mu_A+\mu_B) \end{aligned} \tag{32}$$

with

$$\begin{aligned} C_1 &= (1-\Pi_P)[(a_A\lambda_A+\mu_A+\lambda_P)(a_B\lambda_B+\mu_B+\lambda_P) \\ &\quad \times (a_A\lambda_A+\mu_A+a_B\lambda_B+\mu_B+\lambda_P)]^{-1}, \end{aligned} \tag{33}$$

When $0 \le a_\gamma \le 1$, we have $\Pi_A(a_A,a_B), \Pi_B(a_A,a_B), \Pi_2(a_A,a_B) \ge 0$. Since the capacity $r_1^\gamma, r_2^\gamma > 0$, we have $U_\gamma(a_A,a_B) \ge 0$. Taking the first order derivative of $U_\gamma(a_A,a_B)$ with respect to a_A, we can show that

$$\frac{\partial U_A(a_A,a_B)}{\partial a_A} > 0, \quad \frac{\partial U_B(a_A,a_B)}{\partial a_A} < 0. \tag{34}$$

So when unlicensed 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 tradeoff to choose the optimal a_A that maximizes $U_{PF}(a_A,a_B) = U_A(a_A,a_B)U_B(a_A,a_B)$. Taking the second order derivative of $U_\gamma(a_A,a_B)$ to a_A, we can verify that $\frac{\partial^2 U_A(a_A,a_B)}{\partial a_A^2}$ consists of four terms, which are all negative, thus, $\frac{\partial^2 U_A(a_A,a_B)}{\partial a_A^2} < 0$, and similarly we can verify that $\frac{\partial^2 U_B(a_A,a_B)}{\partial a_A^2} > 0$. Expanding $\frac{\partial^2 U_{PF}(a_A,a_B)}{\partial a_A^2}$, we have

$$\frac{\partial^2 U_{PF}(a_A,a_B)}{\partial a_A^2} = U_A\frac{\partial^2 U_B}{\partial a_A^2} + 2\frac{\partial U_A}{\partial a_A}\frac{\partial U_B}{\partial a_A} + U_B\frac{\partial^2 U_A}{\partial a_A^2}. \tag{35}$$

After some mathematical manipulations, we can prove that $\frac{\partial^2 U_{PF}(a_A,a_B)}{\partial a_A^2} < 0$. Similarly, we can prove that $\frac{\partial^2 U_{PF}(a_A,a_B)}{\partial a_B^2} < 0$. Therefore, $U_{PF}(a_A,a_B)$ defined in (24) is concave in a_γ.

References

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

[2] J. Huang, R. A. Berry, and M. L. Honig, "Auction-based Spectrum Sharing," in *ACM/Springer Mobile Networks and Applications Journal (MONET)*, vol. 11, no. 3, pp. 405-418, Jun. 2006.

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

[4] R. Etkin, A. Parekh, and D. Tse, "Spectrum sharing for unlicensed bands," in *Proc. of IEEE DySPAN 2005*, pp. 251-258, Baltimore, MD, Nov. 2005.

[5] Z. Ji and K. J. R. Liu, "Dynamic pricing approach for spectrum allocation in wireless networks with selfish users," in *Proc. of IEEE Globecom'06*, San Francisco, CA, Nov. 2006.

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

[7] Y. Xing, R. Chandramouli, S. Mangold, and S. Shankar N, "Dynamic spectrum access in open spectrum wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 626-637, Mar. 2006.

[8] S. Keshavamurthy and K. Chandra, "Multiplexing analysis for spectrum sharing," in *Proc. of IEEE MILCOMM'06*, Washington, D.C., Oct. 2006.

[9] 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.

[10] T. Clancy, "Achievable capacity under the interference temperature model," to appear, in *Proc. of IEEE INFOCOM*, Anchorage, AK, May 2007.

[11] T. Clancy, "Formalizing the interference temperature model," to appear, *Wireless Communications and Mobile Computing*.

[12] IEEE 802.22 working group on wireless regional area networks, http://www.ieee802.org/22/.

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

[14] T. M. Cover and J. A. Thomas, *Elements of information theory*, New York: Wiley-Interscience, 1990.

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

[16] G. Owen, *Game Theory*, 3rd ed. New York: Academic, 2001.

[17] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, no. 1, pp. 33-37, Jan. 1997.

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

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

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

[21] 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.