Providing Some Quality of Service for Secondary Users in Cognitive Radios Using Time Slotted Systems

Hudson Okii and Idris A. Rai

Makerere University
P.O. Box 7062
Kampala, Uganda
`Okihudsn@yahoo.com`
`rai@cit.mak.ac.ug`

Abstract. The current research in cognitive radio has been considering absolute guarantee for primary users allowing secondary users access to spectrum only if there is no primary users with data to send. At high arrival rate of primary users this might lead to complete starvation of secondary users and yet it is possible to release some spectrum to secondary users by delaying primary users without affecting their quality. We propose a resource allocation scheme that uses delayed time periods of primary users to transmit secondary user's data packets without jeopardizing the quality of primary users. We analytically modeled the scheme using M/G/1 queue. Our numerical experiments demonstrate that secondary users can be offered some quality of service by delaying primary users in the system to a limit that does not degrade their performance.

Keywords: cognitive radio networks, spectrum assignment, quality of service, queueing theory.

1 Introduction

Recent technological advances have led to growth in use of high data rate wireless application and services. The already crowded and congested radio spectrum has become scarce [2]. *Cognitive radio* (CR) which is a smart programmable radio, capable of sensing interference, learning environment, and dynamically accessing the spectrum is a promising technology to alleviate the increasing stress on the fixed and limited radio spectrum. In cognitive radio networks, the *secondary user* (SU) (unlicensed) user can periodically search and identify available channels in the spectrum to communicate among themselves without disturbing communication of the *primary user* (PU) (licensed) users. *Dynamic spectrum access* (DSA) which is the ability of a secondary user to sense and access spectrum opportunistically can resolve this problem by allowing secondary users to transmit in the assigned but under-utilized frequency bands, provided that the primary user are sufficiently protected.

R. Popescu-Zeletin et al. (Eds.) AFRICOMM 2011, LNICST 92, pp. 24–34, 2012.

Literatures by authors in [5], [6] have addressed the area of dynamic spectrum access in time domain, aiming at exploiting idle periods between bursty transmissions of PU using sense-then-transmit spectrum access strategy.

Delay analysis for a cognitive radio network have been studied by authors in [7], [8], [9]. An approach using a finite state markov chain based queueing model to quantitatively analyse the performance metrics in terms of average packet delay, head of line delay and packet drop rate has been done by authors in [9]. It is worth noting that based on the analytical model, tradeoff among the performance metrics was identified and when and where the cost for favouring the secondary user is worthy was also identified.

Our work is different and distinct from past works in spectrum sharing in cognitive radios using time slotted system where secondary users can only access spectrum when primary users channels are idle as done by authors in [4], [1], [3]. Besides, none of the work in those literatures tries to care for secondary users like our proposed scheme does. To the best of our knowledge, our work is the first of its kind on spectrum sharing which proposes a scheme to provide *quality of service* (QoS) to SU without affecting the quality of PUs at all. Quality of service is defined as the guarantees provided on the ability of network to deliver predictable performance like availability (uptime), bandwidth (throughput), latency (delay), and error rate. Our work is an extension of work done by authors in [4]. In particular, we adopt the analytical methodology and our models are derived from an extension of models in that particular paper.

The rest of the paper is organized as follows: in the next section, we discuss the proposed scheme, and we derive its mathematical models in Section 3. We present and discuss numerical results of the proposed scheme in Section 4, and finally we conclude the paper in Section 5.

2 The Proposed Spectrum Allocation Scheme

In this section, we present the proposed spectrum allocation scheme for cognitive radio network. As pointed out earlier, we propose a scheme that in addition to making use of the unused spectrum, it considers delaying primary user's data in order to provide some service guarantees such as to avoid complete starvation to secondary users.

We consider a time slotted cognitive wireless network where a primary user is the owner of the network. We focus on a cognitive network with one primary and many secondary users uplink. In a conventional cognitive radio network, primary users receive full access to the link without any consideration of secondary users at all because they are the ones who pay for the spectrum (the channels). However, because primary users may not be fully utilizing the provided spectrum, secondary users can opportunistically utilize the channel when primary user is idle and also can share it with primary user when another primary user is present. The cognitive radio access network maintains two queues one for each type of users. In both PU and SU queues, requests are served in a first come first served (FCFS) order of service with SU services only if there is no data in

PU queue. The transmission of the packets is implemented by having the secondary user perform spectrum sensing at the beginning of the slot. If there is no primary users signal at the beginning of the slot, the remainder of the slot can be used for secondary transmission. We assume that there is perfect sensing and time synchronization, and that all packets are one slot in time duration and the system time is slotted with a fixed unit time slot. The network is assumed to operate in noise free and error free channel conditions. Transmission of the packet can only begin at the beginning of the slot so that even if a packet arrives at the middle of the slot, it has to wait a half a slot duration, even if the channel is free. Secondary user is assumed to receive acknowledgment indicating a successful packet transmission, the transmission of acknowledgment is assumed to be error free. If SU transmits a packet and fails to receive the acknowledgment, it retransmits the packet when the channel becomes available again. Each time slot is owned distinctly a primary user. In such a system, when the arrival rate of primary users is very high, secondary users are completely starved. To avoid this in situations where secondary users are to be given some consideration, we propose a scheduling scheme shown in figure 1 that delays primary user when they wish to transmit. Instead, the scheduler gives some priority to secondary users to transmit. Note that we assume that secondary users will always be delayed for a number of time slots before they are given such priority. Otherwise, the scheduler will offer absolute priority to secondary users to the expense of primary users extended delay.

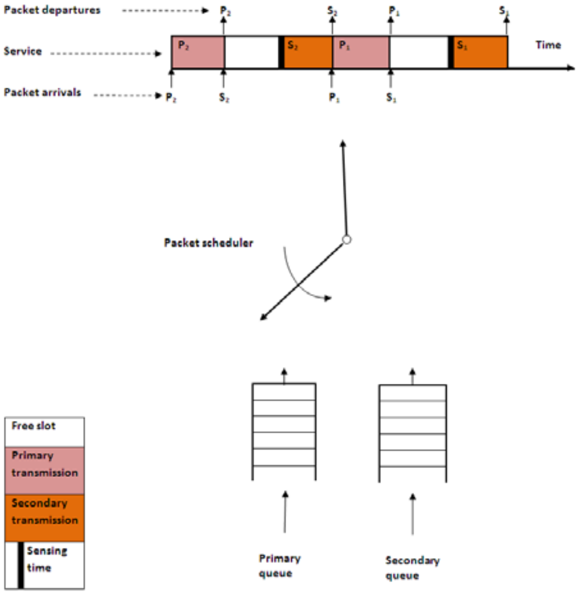
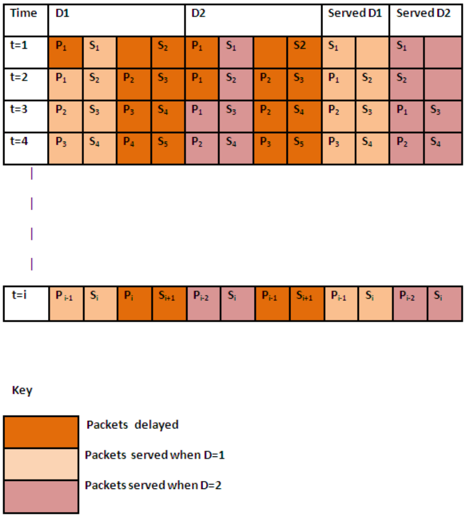


Fig. 1. The model

The number of times primary user's request is delayed in the system before it is allowed to transmit depends on the utilization of the system and the quality requirement of the application. Primary user's requests is not delayed indefinitely, instead primary users application's maximum tolerable delay limit is used to ensure that primary users requests are not starved. The maximum tolerable delay limit depends on the user application. When the number of times primary users request has been delayed approaches maximum tolerable delay limit, the scheme does not delay this requests any longer. It is served immediately so that the quality of its application is not jeopardized. Figure 2 presents the proposed scheduling scheme that schedules primary user's request when $W^P(D = 1)$ and $W^P(D = 2)$ at $t = i$. W^P is the packet delay time for primary users and D is the number of times a single primary users request can be delayed. Let $t = 0, 1, 2...$, index the request service periods, each of 1 time slot. Primary users requests are denoted by P_i and secondary users request are denoted by S_i, let $i = 0, 1, 2, 3...$, index request number. When $D = 1$ at time period $t = i$, primary users packet P_i is delayed by the service of primary users requests P_{i-1} and secondary users request S_i. When $D = 2$ at time period $t = i$, primary users packet P_{i-1} is delayed by the service of primary users request P_{i-2} and secondary users request S_i. When $D = 1$ at time period $t = i$, secondary users packet S_{i+1} is delayed by the service of primary users requests P_{i-1} and secondary users request S_i. When $D = 2$ at time period $t = i$, primary users packet S_{i+1} is delayed by the service of primary users requestP_{i-2} and secondary users request S_i.



Time	D1				D2				Served D1		Served D2	
t=1	P_1	S_1		S_2	P_1	S_1		S2	S_1		S_1	
t=2	P_1	S_2	P_2	S_3	P_1	S_2	P_2	S_3	P_1	S_2	S_2	
t=3	P_2	S_3	P_3	S_4	P_1	S_3	P_2	S_4	P_2	S_3	P_1	S_3
t=4	P_3	S_4	P_4	S_5	P_2	S_4	P_3	S_5	P_3	S_4	P_2	S_4
t=i	P_{i-1}	S_i	P_i	S_{i+1}	P_{i-2}	S_i	P_{i-1}	S_{i+1}	P_{i-1}	S_i	P_{i-2}	S_i

Key

Packets delayed

Packets served when D=1

Packets served when D=2

Fig. 2. Request schedule for $W^P(D = 1)$ and $W^P(D = 2)$ at $t = i$

3 Models of the Proposed Scheme

3.1 Mathematical Background

In this section, we derive mathematical expressions of average time that packets of different user types under the proposed scheme spend in the system (aka *response time*). We first derive the expressions under M/G/1 queue which assumes a general packet size distribution G and then consider M/D/1 to derive the expressions for the case when all packets are of the same size.

Let the subscript or superscript r of the following notations represent either primary (P) or secondary (S) users. We denote the mean arrival rate as λ_r, $\overline{X}_r = \frac{1}{\mu_r}$ denotes the average service time of a user, N_Q^r as the average number of users packets in a queue, and W_Q^r as the average waiting time in the queue. We also denote W^r as the total time spent by a packet of any user in the system, TD as mean waiting time until the beginning of the slot, ρ^r as the utilization factor, $W^r(d = i)$ as the average waiting time of one user type after delaying the other user type i times before receiving service.

In what follows, we derive the expressions of response time of packets of different user types under the proposed scheme. We first derive the expressions under M/G/1 queue which assumes a general packet size distribution G and then consider M/D/1 to derive the expressions for the case when all packets are of the same size.

Average Delay of Primary Users. Response time is defined as the total time a packet spends in the system. We model the scheduler to be able to analytically estimate the mean response of packets that belong to different types of users. In this section, we derive the expressions for mean response time of primary user's packet when it is not delayed. We consider a tagged packet that belongs to a primary user. Its mean response time is delayed by primary user's packet it finds in the scheduler's primary queue upon its arrival. The average delay of the tagged primary user's packet is therefore given as

$$W_Q^P = TD + \frac{N_Q^P}{\mu} \tag{1}$$

By using Little's theorem on N_Q^P in the above equation, we obtain the expression for the mean waiting time of the packet as $W_Q^P = \frac{TD}{(1-\rho^P)}$.

Recalling the fact that the average packet delay (time spent in the system) is given by sum of the mean waiting time in the queue and the average service time of the packet, we can express the packet's mean response time for the primary user as follows:

$$W^P = \overline{X}_P + \frac{TD}{(1-\rho^P)} \tag{2}$$

In the proposed scheme, the priority of the primary users is relaxed by delaying them in the system up to the delay value that does not affect the underlying application's quality. We therefore derive the expressions for primary user's packet mean response time after it has been delayed i times before receiving service. Again, let that packet is tagged. Its average delay is composed of its mean service time and the mean service times of all primary and secondary users it finds in the system upon its arrival. The average waiting time of the tagged primary user's packet is therefore given as

$$W_Q^P(D=i) = TD + \frac{N_Q^P}{\mu} + \frac{i}{\mu_s} \tag{3}$$

By using Little's theorem, we obtain

$$W_Q^P(D=i) = \frac{TD + \frac{i}{\mu_s}}{1-\rho^P} \tag{4}$$

Adding the mean service time of the packet to Equation (4), we obtain packet's average response time for a primary user as

$$W^P(D=i) = \overline{X}_P + \frac{TD + \frac{i}{\mu_s}}{1-\rho^P} \tag{5}$$

Next we derive the expression of the mean response time of a secondary user's packets.

Average Delay of Secondary Users. We start with the expression for secondary user's packet mean response time before primary user's packets are delayed to receive service. We again consider a tagged packet that is transmitted by a secondary user. The tagged packet is delayed by primary users packets found in the queue, secondary user's packet found in the queue, and other primary users that finds it in the system. The average delay of the tagged secondary user's packet is therefore given as

$$W_Q^S = TD + \frac{1}{\mu}N_Q^P + \frac{1}{\mu}N_Q^S + \frac{1}{\mu}\lambda_P W_Q^S \tag{6}$$

By using Little's theorem on N_Q^P and N_Q^S, we obtain the expression for the mean waiting time of the packet as

$$W_Q^S = \frac{TD}{(1-\rho^P)(1-\rho^P-\rho^S)} \tag{7}$$

and finally the average delay of a packet from secondary user is given by

$$W^S = \overline{X}_S + \frac{TD}{(1-\rho^P)(1-\rho^P-\rho^S)} \tag{8}$$

Now let us consider tagged secondary user with data to transmit. We derive the expression for the mean response time of secondary user's packet after primary user's packet has been delayed i times before receiving service which is the essence of the proposed scheme. The average delay of the tagged secondary user is composed of its mean service time and the sum of service times of all primary and secondary users it finds in the system including the packet that is receiving service upon its arrival. The average delay of the tagged secondary user's packet is therefore given as

$$W_Q^S(D=i) = TD + \frac{1}{\mu}N_Q^P + \frac{1}{\mu}N_Q^S \tag{9}$$

By applying Little's theorem on N_Q^P and N_Q^S, we obtain the expression of the mean waiting time of the tagged packet as follows

$$W_Q^S(D=i) = \frac{TD}{(1-\rho^P)(1-\rho^S)} \tag{10}$$

Finally, the average delay of the tagged packet is given as

$$W^S(D=i) = \overline{X}_S + \frac{TD}{(1-\rho^P)(1-\rho^S)} \tag{11}$$

Average Delay of Primary and Secondary Users Using M/D/1 System. In this section, we address the special case where time is slotted with deterministic service time of one slot (M/D/1) and the fact that packet service can only start at the beginning of the slot. We assume that the service time is one slot and newly arrived packet has to wait for 1/2 slot before the beginning of slot. We can substitute these values in $\overline{X}_P$ and TD in equation (2) to obtain

$$W^P = 1 + \frac{\frac{1}{2}}{(1-\rho^P)}. \tag{12}$$

We also substitute the service time of 1 slot and the waiting time of 1/2 of a slot in $\overline{X}_P$ and TD in equation (5) to obtain

$$W^P(D=i) = 1 + \frac{\frac{1}{2} + \frac{i}{\mu_s}}{1-\rho^P}. \tag{13}$$

Substituting the service time of 1 slot and the waiting time of 1/2 of a slot in $\overline{X}_P$ and TD in equation (8) when primary user's packet is delayed before receiving service, we obtain

$$W^S(D=i) = 1 + \frac{\frac{1}{2}}{(1-\rho^P)(1-\rho^P-\rho^S)}. \tag{14}$$

And substituting the service time of 1 slot and the waiting time of 1/2 of a slot in $\overline{X}_P$ and TD in equation (11) when primary user's packet has been delayed i times before receiving service to obtain

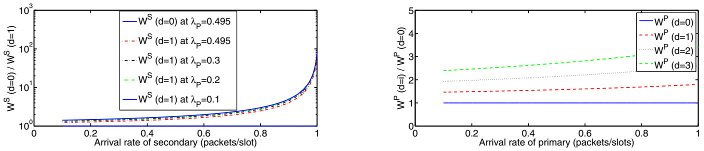
$$W^S(D=i) = 1 + \frac{\frac{1}{2}}{(1-\rho^P)(1-\rho^S)}. \tag{15}$$

4 Performance Evaluation

In this section we discuss numerical results that show the performance of the proposed scheme when secondary users are favoured to access spectrum by delaying primary users, We assume that secondary user's access to spectrun depends on the utilization of the system and the quality requirement of the application, we also evaluate the effect of delaying primary user's packet on varying applications.

4.1 Evaluation of Average Waiting Time for Secondary Users Packet When $W^S(d = 0)$ and When $W^S(d = i)$

In this section we analyze the effect of secondary user's arrival rate on its average waiting time when primary user is delayed several times before receiving service and when its arrival rate is fixed and at a service rate equals that of the secondary user.



(a) Ratio of $\frac{W^S(d=0)}{W^S(d=i)}$ (b) Ratio of $\frac{W^P(d=i)}{W^P(d=0)}$

Fig. 3. The delay performance for secondary and primary users

Fig 3(a) presents the ratio of $\frac{W^S(d=0)}{W^S(d=i)}$ as a function of its arrival rate for $d = i$ (where i=1) when the primary users arrival rate is varied at $\lambda = 0.495$, $\lambda = 0.3$, $\lambda = 0.2$ and $\lambda = 0.1$ and at equal service rate of $\mu = 1.5$. It can be observed that with high primary user arrival rate fixed at $\lambda = 0.495$, the ratio of $W^S(d = 0)$ at secondary users arrival rate of 0.1 is approximately 100/87. This indicates a reduction in average waiting time of secondary users when delayed by 1 time slot at given arrivals rates by approximately 13 percent. With low primary users arrival rate fixed at $\lambda = 0.1$, the ratio of $W^S(d = 0)$ at secondary arrival rate of 0.1 is approximately 100/79, which indicates a reduction in secondary users average waiting time by 21 percent. The result shows that the mean response time of secondary user's packets deteriorates when the primary arrival data rate is high.

Evaluation of packet delay time for primary users when $W^P(d = 0)$ and when $W^P(d = i)$. We now analyze the effect of primary user's arrival rate on its average delay time when the primary data is delayed several times before receiving service and when the secondary user's arrival rate is fixed.

Fig 3(b) shows $\frac{W^P(d=i)}{W^P(d=0)}$ as a function of primary users arrival rates for $d = 0, d = 1, d = 2$, and $d = 3$ when the secondary users arrival rate is fixed at $\lambda = 0.4$

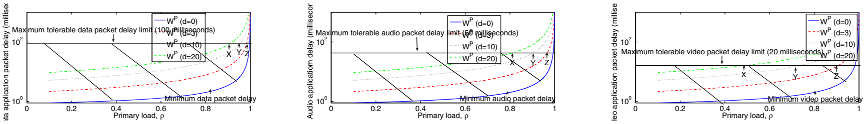
and at equal service rate of $\mu = 1.5$. The figure shows that the ratio of $\frac{W^P(d=1)}{W^P(d=0)}$ is approximately 1.5 to 1 and the ratio of $\frac{W^P(d=2)}{W^P(d=0)}$ is approximately 2.1 to 1. It can be seen that at primary users arrival rate of 0.1, for each $d = 0, d = 1, d = 2$, and $d = 3$, there is an increase in the ratio of $\frac{W^P(d=i)}{W^P(d=0)}$ by a factor of approximately 0.5 slots per request delay.

4.2 Evaluation of Packet Delay Time Limit for Data, Audio and Video Applications

In this section we analyze the effect of primary users packet delay as a function of its load on data, audio and video applications. According to International Telecommunication Union (Y.1541 and Y.1221), the maximum tolerable delay limit of audio, video, and data application requests should not be more than 50, 20, and 100 milliseconds respectively. These values will be used in the evaluation and analysis of the results.

Figure 4(a) presents data packet delay in (millisecond) as a function of primary users load ρ when the secondary users arrival rate is fixed at $\lambda = 0.4$ and at equal service rate of $\mu = 1.5$. It can be seen that the maximum tolerable delay limit for data application is 100 milliseconds. Beyond this limit, the quality of the application is degraded.

Figure 4(b) presents audio packet delay in (millisecond) as a function of primary users load ρ when the secondary users arrival rate is fixed at $\lambda = 0.4$ and at equal service rate of $\mu = 1.5$. It can be seen that the maximum tolerable delay limit for audio application is 50 milliseconds, Beyond this limit, the quality of the application is degraded.



(a) Data packet delay (b) Audio packet delay (c) Video packet delay

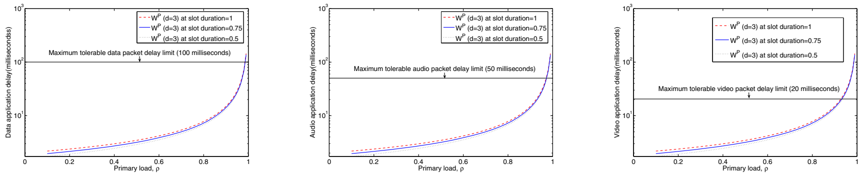
Fig. 4. The packet delay performance for data, audio and video applications

Figure 4(c) presents video packet delay in (millisecond) as a function of primary users load ρ when the secondary user's arrival rate is fixed at $\lambda = 0.4$ and at equal service rate of $\mu = 1.5$. It can be seen that the maximum tolerable delay limit for video application is 20 milliseconds. Beyond this limit, the quality of the application is degraded.

In this section, our observation and analysis of the figures concludes that data application tolerates higher packet delays at lower load than audio and video applications for a fixed secondary user's arrival rate. It can be observed that primary user's requests can be delayed up to a maximum tolerable limit depending on the application's delay tolerance. The maximum tolerable delay limit for video applications is reached at a load of $\rho = 0.57$ when primary users requests are delayed 20 times. For audio applications, the maximum tolerable delay limit is reached at a load of $\rho = 0.82$ when primary users requests are delayed 20 times. In the case of data applications, the maximum tolerable delay limit is reached at a load of $\rho = 0.92$ when primary users requests are delayed 20 times. Therefore the above results show that video application is less delay tolerant than audio application which in turn is less delay tolerant than data application. Hence video application can be delayed fewer number of times before degrading it's quality of service compared to audio application which can be delayed more number of times. Data application can be delayed more times than video and audio.

4.3 Evaluation of Packet Delay Time Limit for Data, Audio and Video Applications under Varying Time Slot Duration

In this section we analyze the effect of primary users packet delay as a function of its load on data, audio and video applications when primary user's packet are delayed three times under varying time slots of 1, 0.75, and 0.5 time slot duration.



(a) Data packet delay (b) Audio packet delay (c) Video packet delay

Fig. 5. The packet delay performance at varying slot durations

Figure 5(a), (b) and (c) present different PU application packet delay in (millisecond) as a function of primary users load ρ when primary user's packet are delayed three times under varying time slot durations of 1, 0.75, and 0.5. It can be observed that the lower the slot duration, the higher the number of times a PU packet can be delayed before it reaches its maximum tolerable limit. The number of times a PU packet is delayed for slot duration 0.5 is higher than for slot duration 0.75 which in turn is higher than for slot duration 1.

5 Conclusion

Unlike conventional spectrum assignment schemes in CR which guarantee absolute priority to PU, we propose an assignment scheme that provides some quality guarantees to SU. Different applications require different quality guarantees, and therefore it is possible to delay some PU data in the system for some time without affecting its quality of service. The novel proposed spectrum assignment approach seems credible for various applications because it directly favors SU so that it is not completely starved in situations when arrival rate of PU is high, unlike other previous schemes [9]and [5].

We extend the models derived from the previous work by authors in [4] to derive analytical expressions of the mean waiting time of SU and PU packets in the proposed systems using queuing theory. We used the models to numerically evaluate the proposed scheme. The results clearly show that it is possible to provide some quality of service to secondary users while preserving the acceptable quality requirements of the PU.

It should be noted that the proposed scheme does not suggest providing absolute guarantees to SU. The models present a general case only. In realistic implementation, PU will be delayed to give room to SU packets only if they SU have been delayed for some time.

References

1. Chen, T., Zhang, H., Katz, M.D., Zhou, Z.: Swarm Intelligence Based Dynamic Control Channel Assignment in CogMesh. In: IEEE International Conference on Communications (2008)
2. FCC, Report of the spectrum efficiency working group, FCC Spectrum Policy Task Force, Tech. Rep. (November 2002)
3. Kang, B., Park, H., Kim, Y., Woo, S., Ban, S.: Out-of-band cooperative spectrum sensing in cognitive radio system of multiple spectrum bands. In: Proceedings of the 3rd WSEAS International Conference on Circuits, Systems, Signal and Telecommunications (2009)
4. Suliman, I., Lehtomäki, J.: Queueing analysis of opportunistic access in cognitive radios. In: First International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace and Electronics Systems Technology, Wireless VITAE 2009 (2009)
5. Xu, Y., Wang, J., Wu, Q.: Interference-Throughput Tradeoff in Dynamic Spectrum Access: Analysis Based on Discrete-Time Queuing Subjected to Bursty Preemption. In: 4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (2009)
6. Zhao, Q.C., Geirhofer, S., Tong, L., Sadler, B.M.: Optimal dynamic spectrum access via periodic channel sensing. In: Wireless Communications and Networking Conference (2007)
7. Wang, S., Zhang, J., Tong, L.: Delay Analysis for Cognitive Radio Networks with Random Access: A Fluid Queue View. In: The Proceedings of INFOCOM 2010 (2010)
8. Zhang, C., Wang, X., Guan, X.: Quality-of-Service in Cognitive Radio Networks with Collaborative Sensing. In: The Proceedings of GLOBECOM 2009 (2009)
9. Su, H., Zhang, X.: Secondary user friendly TDMA Scheduling for Primary Users in Cognitive Radio Networks. In: The Proceedings of GLOBECOM 2009 (2009)