A Comparative Study of Scheduling Schemes for Cognitive Radio Networks: A Quality of Service Perspective

V. Mehar Swarup
Cisco Systems Inc.
Cessna Business Park
Bangalore-560103
Email: mehar1988@gmail.com

Vinay J. Ribeiro
Deptt. Of CSE
Indian Institute of Technology
Hauz Khas, New Delhi-110016
Email: vinay@cse.iitd.ernet.in

Amol Gupta
Cosmic Circuits Pvt. Ltd.
AECS Layout- A Block
Bangalore-560037
Email: amolgupta87@gmail.com

***Abstract*—Quality of Service (QoS) based scheduling in Cognitive radio networks (CRN) is a pressing research problem. The objective of this work is to study the feasibility of using CRN for delay sensitive applications in the presence of primary users with random ON/OFF periods.**

The first contribution of this work is the modeling of a cognitive node as a Queuing system and characterization of the average delay encountered by a packet on a cognitive node for various scheduling disciplines. This delay characterization gives the application/user the decision power to choose whether to transmit using a particular node depending upon the delay and rate requirements. It also gives the designer of higher layer the ability to make more informed decisions for routing in a multiple node scenario.

The second contribution of this work is a comparative study of various scheduling disciplines in a cognitive radio environment. We have considered First Come First Serve (FCFS), strict priority and a proposed Queue Proportional Batch Scheduling (QPBS). QPBS maintains a balance between priority and fairness as opposed to the strict priority scheduling where the fairness is sacrificed at the cost of priority and FCFS where there is no notion of priority. This comparative study clearly brings out the relative pros and cons of aforementioned scheduling schemes in a cognitive radio environment. A discrete event based simulation study has been performed to verify the efficacy of mathematical modeling and validate the results.

***Index Terms* CRN, QoS, FCFS scheduling, Strict Priority Scheduling, QPBS**

I. Introduction

Cognitive radio Networks are basically no guarantee networks. An important issue in this context is feasibility of usage of cognitive radio networks for applications which require a guarantee on the Quality of Service (e.g. delay and throughput). Now considering the random nature of the CRNs, the exact delay or throughput guarantees are almost impossible to achieve.

In this work, our focus is on the delay analysis of single hop static wireless cognitive radio network. Optimal channel selection and contention and resolution schemes are not part of this work. We have used queuing theory to analyze the scenarios and, using analogies with priority assignment problems, given an approximation of the average delay encountered by packets of a particular application while using a cognitive node for its transmission. A cognitive node in our scenario is basically a channel on which secondary users can transmit in the absence of primary users.

978-1-4673-5494-3/13/$31.00 © 2013 IEEE

We have answered following questions in the present work:-

1) What is the average delay that a packet has to bear in order to be serviced by a cognitive node?
2) How to schedule packets in order to maintain a fair trade-off between priority and fairness, while meeting QoS constraints? A novel scheduling scheme has also been proposed.
3) A comparison between various scheduling disciplines in cognitive radio environment.

Previously there have been many studies on resource management in cognitive radio networks. People have proposed algorithms for resource management in CRNs [2], [4], [9]. The focus of these is on throughput studies and optimal channel selection strategies, but there is not much work done towards packet scheduling based on application specific tags. In [8] the authors have worked in the direction of application specific scheduling but their focus is on resource management and load balancing among secondary users.

The structure of rest of this paper is as follows. In section II, we have classified the problem into sub parts. In section III, we will build the theoretical foundation for different packet forwarding strategies for cognitive radio scenario. Section IV discusses the implementation details and simulation results and Section V presents a brief comparative study among the scheduling disciplines, finally Section VI concludes the paper.

II. System Modeling

A. The Primary User

Primary users own the spectrum and they deserve interference free transmission and reception. There are various primary user models described in the literature. We have used two different models in our simulations and termed them as type I and type II primary models respectively.

Type I Primary Model: In this model, we have taken an exponential on and exponential off time period, where on and off states alternate [5].

Type II Primary Model: In [8], the authors have taken a poisson arrival of packets and each packet is served according to exponential service time distribution.

We model the primary user as the highest priority user which preempts the transmission of any other secondary user [8]. This is an interesting way of looking at the problem. By viewing the primary user as the highest priority user, we can exploit the research already done for priority queuing scenarios.

B. The Secondary User

Secondary users use the spectrum opportunistically. There are two modes of operation of a secondary user. One is sensing mode in which the secondary user senses the channel and detects the presence or absence of primary user and the SINR of the channel. The second mode is the transmission mode. The secondary packet arrival and service time are taken to be Poisson and exponential respectively.

C. Problem Classification

- *Single Channel (Server) and Single Queue*: In Single queue case, there is only one queue through which all the entities are processed on a first come first serve basis. Delay analysis and admission control for this is discussed in section III-A and simulation results are presented in section IV-A.
- *Single channel (server) and multiple applications (multiple queues)*: In the multiple queue single server scenarios, there are different queues for different application/users (or it can be class of applications). The concept of service differentiation based on one or more attributes of the applications is applicable in this scenario. Strict priority and the proposed **Q**ueue **P**roportional **B**atch **S**cheduling (**QPBS**) scheme come in this sub part.

III. Theoretical Framework

The average delay encountered by a specific application in cognitive environment is a very tricky problem because of the variable nature of input traffic, primary activity and the available bandwidth. The scope of this work is limited to stationary primary activity pattern and a single channel which is interrupted by primary users. The results obtained in this work can be used as a guideline for the scenarios where the primary activity pattern is non-stationary, but the validity of this claim is not verified in this work.

A. Single channel (Server) and Single Queue

This is the case when there is no priority of one application over the other. The packets are served in the order of their arrival. Here we have taken the first model of the primary generator (Type I) as described in section II-A. Since the server is modulated by the primary, the effective capacity (C_{eff}) for secondary is given by:

$$C_{eff} = \gamma \, B \, log_2(1 + SINR) \tag{1}$$

where γ is the duty cycle of the secondary user and is defined as:

$$\gamma = \frac{\mathbb{E}\left[t_{ON}\right]}{\mathbb{E}\left[t_{ON}\right] + \mathbb{E}\left[t_{OFF}\right]} \tag{2}$$

t_{on} and t_{off} are primary off (secondary **on**) and primary on (secondary **off**) times respectively. The on and off times are random variables with probability density function given by:

$$f_{T_{ON}}(t_{on}){=}\lambda \exp(-\lambda t_{on}) \tag{3}$$

$$f_{T_{OFF}}(t_{off}){=}\mu \exp(-\mu t_{off}) \tag{4}$$

If the average packet size is P, then time to serve one packet is given by:

$$AverageServiceTime(E[C]){=}\frac{P}{C_e f f} \tag{5}$$

For the simplicity in the analysis we assume:

$$\frac{P}{B \log_2(1 + SINR)}{=}1 \tag{6}$$

Hence the average service time is given by

$$E[C]{=}1 + \frac{E[t_{off}]}{E[t_{on}]} \tag{7}$$

This problem resembles a single queue, single server first come first server (FCFS) queuing system where the state of the server alternates between available (on) and unavailable (off) with random on and off periods. We find that primary user behavior is analogous to an ON-OFF process with random ON and OFF periods and secondary applications are forming the queue. Hence we can exploit the results achieved for such queuing systems.

Awi Federgrue et al. [3] have modeled the queuing systems with Markov modulated service time and they prove this result Eq. (7) in a more generalized manner, valid for any general on and off time distributions. They have also derived the total waiting time of a packet in the system given by Eq. (8) [3] :

$$E[W]{=}E[C]{+}\frac{\lambda_{Secondary} E[C^2]}{2(1 - \lambda_{Secondary} E[C])}{+}\frac{E[t_{off}^2]}{2(E[t_{on}] + E[t_{off}])} \tag{8}$$

Where $\lambda_{Secondary}$ is the secondary traffic average arrival rate. The queue would become unstable when

$$(1 - \lambda_{Secondary} * E(C)) < 0 \tag{9}$$

The admission policy of the single node is thus formed. On the basis of the primary activity pattern and the secondary rate requirement, we can say following things beforehand:

1) Average waiting time: Knowing this information prior to using a node for communication will be extremely useful. An application/user can decide beforehand whether

to use that particular node or not based on QoS criterion. As shown in the simulation section that this average converges to a stable value over a short span of time, a moving average of the average delay will be a good estimate of the same in a real time scenario.
2) Rate support: Another important feature which lets a user/application know whether their rate requirement will be met.

If any one of the two criteria is not met, the application/user might not use that particular node for communication. This feature gives the designer of upper layers an ability to make more informed routing decisions.

B. Single Channel (Server) and Multiple Queues

We have considered strict priority and queue proportional batch scheduling disciplines in this section

1) Strict priority system: In this scheme, the higher priority entities get absolute priority over lower priority entities. The priorities are assigned on the basis of one or more functionalities of the entities e.g. in our case the priorities may be assigned on the basis of delay.

Under the cognitive environment, in addition to the secondary application to be served, we have primary users which have inherently highest priority. We have used this fact to model the primary. This makes the whole system a strict priority system. This allows us to exploit the results already obtained for the strict priority scheduling. We have used the second primary model (Type II) here as described in section II-A , i.e., the primary packet arrival process is Poisson and the service time is exponential. During the service time of the primary, the secondary has to remain silent.

The average arrival rate of the primary traffic is denoted by λ_1 and secondary average arrival rates are denoted by $\lambda_2 \dots \lambda_r$. If server follows exponential service time distribution for each of the queues, the average waiting time of a packet is given by Eq. (10) [1] :

$$W_p = \frac{\sum_{k=1}^{p} \frac{\lambda_k}{\mu_k^2}}{\left(1 - \sum_{k=1}^{p-1} \frac{\lambda_k}{\mu_k}\right)\left(1 - \sum_{k=1}^{p-1} \frac{\lambda_k}{\mu_k}\right)} \tag{10}$$

Where μ is the parameter for service time distribution.

2) Queue Proportional Batch Scheduling (QPBS) : The strict priority scheme lacks fairness. When there are multiple applications or users wanting to transmit their data through a specific node, the fairness issue comes into the picture. We cannot always give strict priority to one of the applications/users over the other. In this regard, we propose a scheduling scheme which maintains a trade-off between priority and fairness. This idea was motivated by a memory scheduling algorithm called PARBS [6]. The important feature of PARBS is ensuring fairness among different applications by forming a batch.

The other important thing is to avoid the buffer overflow. Hence the number of packets from each queue is decided based on individual queue lengths in a batch. Finally to prefer a delay sensitive application over less delay sensitive application, transmission of packets takes place in sorted order of their delay sensitivities. The algorithm can be briefly summarized in the following steps.

1) Different FCFS queues are maintained for different services. Incoming traffic from different services enters the respective queues.
2) Based on queue length of different queues, decide the proportion allocated to the packets of each service in a batch. Suppose the queue lengths of individual applications are denoted by $Q_1, Q_2 \dots Q_n$,and the batch size is $L + n$, (where L is a constant positive integer) then the number of packets of each application in the batch is given by:

$$L_i = \left\lceil \frac{Q_i}{\sum_{i=1}^{n} Q_i} min\left(\sum_{i=1}^{n} Q_i, L\right) \right\rceil \tag{11}$$

3) Sort the batch based on the delay sensitivity of each application with the most delay sensitive application at the top of the batch. Transmit the packets in the sorted order.

TABLE I
ANALOGY BETWEEN QPBS AND PWRR

Traffic Model(PWRR)	Model	Analogy(QPBS)
Expedited Forwarding(EF)	Poisson arrival, Exponential Service time	Primary Traffic
Assured Forwarding(AF)	Poisson arrival, Exponential Service time	Secondary traffic 1
Best effort forwarding (BE)	Poisson arrival, Exponential Service time	Secondary traffic 2 (AF and BE are similar in our case)

Our proposed scheme has some similarities with the priority weighted round robin scheme (PWRR) [10] with an analogy as presented in Table I. In the DiffServ scheme [7], they have divided the traffic into three classes, namely, Expedited Forwarding (EF), Assured Forwarding (AF) and the Best Effort (BE). We have found an analogy with these traffic classes and our model resembles this model. The difference between the DiffServ scheme and proposed scheme is that the weighting factor (K=ratio of the packets of AF and BE traffic served in one round) in the weighted round robin is not constant in our model but it changes adaptively as the queue lengths change.

Because of this similarity, we have exploited the results obtained by [10] and used them as an approximation for the delay calculations.

IV. SIMULATION RESULTS

We carried out our simulation in SimEvents, a MATLAB based discrete event simulator. The various blocks used in this

TABLE II
SIMULATION PARAMETER VALUES

Parameter	Value
$E[t_{on}]$	1
$E[t_{off}]$	0.1
Secondary Rate	0.5
Average Server Delay	1

simulation setup are described below

- **Primary generators** We have done our simulations with two prevalent models described in the literature. In the first one, we have taken exponential on and off times which alternate. The on and off process are independent of each other. This type of model is widely used in literature including [5] . The motivation of using this model stems from the fact that we can exploit the results obtained by Awi Federgruen et.al [3] .
  In the second model, Primary generates entities which follow poisson arrival and get exponential service time. In other words, the inter-arrival time between the entities is exponential and the service time of the entities is also exponential [8], [10].
  Whenever the primary is off, the secondary can transmit its data; when primary is on, the secondary stops its transmission. We have assumed perfect synchronization among the primary and secondary for the simulation purpose. In a realistic scenario, the secondary has to undergo the sense, contention and transmit cycles.
- **Secondary generators** Secondary activity generators generate entities which follow Poisson arrival. The arrival rate and the distribution can be user defined/application specific.
- **Servers** We have modeled the channels as the servers. Whenever the primary is not using the channel, the secondary can use it. The data rate provided by a wireless channel depends on bandwidth and SINR conditions according to Shannon's formula. We have assumed a very slow fading channel and hence a constant rate providing channel. This means that whenever primary is on it provides a constant rate to primary traffic and when it is off it provides the same rate to the secondary traffic. So a server is a kind of delay block which provides a delay proportional to the size of the entity (packet). The server is modulated by the primary activity generator
- **Queue** When the server is processing primary traffic or in queue-server terminology when it is in off state, the arriving secondary entities are stored in the queue. The average waiting time spent in the queue before getting served by the server is known as queue waiting time

Now we are in a position to discuss the simulation results of all the three scheduling schemes discussed in the earlier sections.

A. Single Server and Single Queue

The single queue case is equivalent to no priority. Users/data are served in order of their arrival. We have taken exponential on-off model for the primary (Type 1). The exact expression for total system delay is given by Eq. (8). To validate the theoretical framework as discussed above, we performed a simulation with parameters as shown in Table II. The parameters taken for simulation are used to validate the theoretical framework, but they can be scaled to the typical values of various applications.

TABLE III
SINGLE QUEUE SINGLE SERVER SERVICE AND WAITING TIMES

Parameter	Theoretical	Simulation
Average service time	1.1	1.074
Average waiting time	1.72	1.74

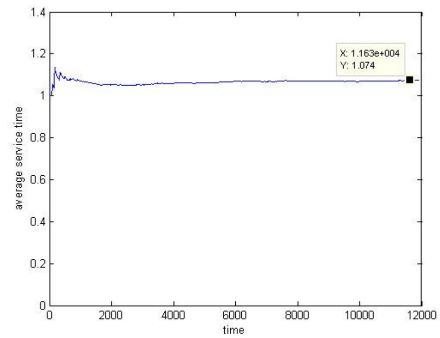

Fig. 1. Average service time

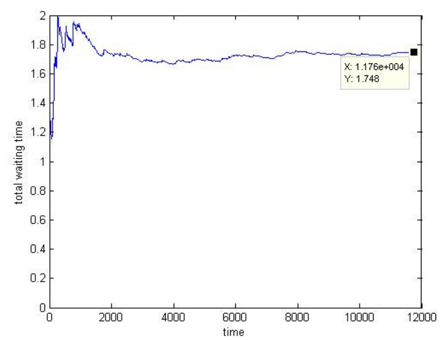

Fig. 2. Average waiting time

The average service and waiting are shown in Fig. 1 and Fig. 2. The comparison between the theoretical Eq. (7), Eq. (8) and simulation values is shown in the Table II. This confirms our modeling and mathematical framework. We can see that averages stabilize over a short period of time provided that the primary activity pattern doesnt change drastically over a short span of time. This is a reasonable assumption to make especially in the case of DTV bands.

B. Single Server and Multiple Queues

In this section, we will present the simulation study done for the second scenario. We have performed simulations for both strict priority and QPBS scheme. A brief comparative study is also presented in Section V.

TABLE IV
A COMPARISON OF THEORETICAL AND EXPERIMENTAL SYSTEM WAITING TIMES IN A STRICT PRIORITY QUEUING SCHEME

IAT1, IAT2	Theoretical delay-high priority	Experimantal delay- high priority	Theoretical delay-low priority	Experimental delay-low priority
1.5, 4.5	3.38	3.4	47.41	46.10
1.5, 6.0	3.18	3.12	24.52	23.10
1.5, 7.5	3.06	2.91	18.56	18.22
1.5, 9.0	2.98	2.96	15.82	15.10
1.5, 10.5	2.92	2.86	14.25	12.20
1.5, 12.0	2.88	2.88	13.24	14.40

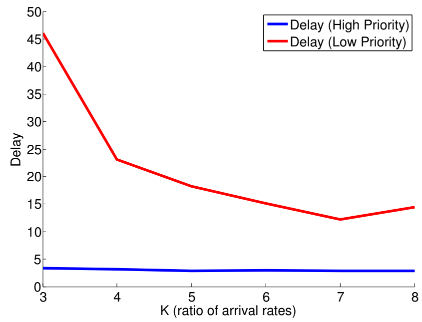

Fig. 3. Delay comparison between high and low priority applications for strict prority scheme

1) *Strict Priority* : The simulation parameters are noted below:
Primary Inter arrival time=20 time units
Primary Service time=1 time unit
Simulation time=40,000 time units

The units of time and rates are time unit and per time unit respectively. The simulations are done to validate the theoretical framework. The values can be scaled to the typical values of rate and inter arrival time of various applications. Simulation is performed with two applications with rates λ_1 and λ_2. Inter-Arrival time1 (**IAT1**) and **I**nter-**A**rrival **T**ime2 (**IAT2**) are the inverse of $\lambda 1$ and $\lambda 2$ respectively. The waiting times are plotted versus the ratio $K = \frac{IAT2}{IAT1} = \frac{\lambda_2}{\lambda_1}$ of the arrival rates. As we expect, in the strict priority scheme, the higher priority traffic gets very less waiting time at the cost of the lower priority traffic. The theoretical value of the system waiting time can be found out if we view the primary as the highest priority user in the result obtained in Eq. (10). Table IV lists the theoretical and simulation values of the system delays(queue waiting time + service time) for the strict priority scheme. Fig. 4 is the graphical illustration of the table IV .

2) *Queue Proportional Batch Scheduling*: In this setup, there are two secondary applications interrupted by a primary application. The packets from the individual queues are taken in proportion to their queue lengths and are sorted within the batch according to their delay priority. The primary average inter arrival time is 20 time units and the average service time is 1 time unit. Both follow exponential distribution. The batch size is taken to be 25. Simulation is performed for 40,000 time units. We observe that the delay trend of the higher and lower priority waveforms is similar in QPBS (Fig.5 QPBS Delays), but there is definite differentiation between the two applications based on their delay priority. This depicts a decent trade-off between priority and fairness, as against the strict priority wherein the high priority packets observe very less delay at the cost of lower priority packets being starved. The distance between the two waveforms is a controllable parameter which is a function of the batch size.
The Fig. 6 compares the simulation results with the theoretical values as obtained by PWRR delay equation for the higher priority secondary application (priority in terms of delay), where the definition of K remains the same.
As anticipated, the theoretically approximated value of waiting time and simulation values do not match exactly but they follow a similar trend. The difference is attributed to the dissimilarities between QPBS and PWRR.

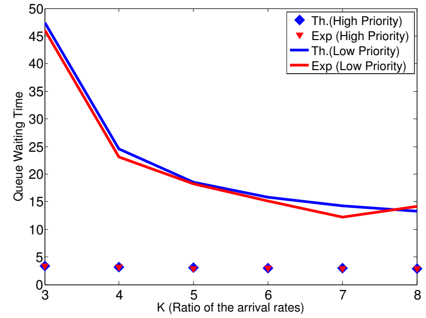

Fig. 4. Comparison of theoretical and experimental waiting times for strict priority scheme

TABLE V
THE AVERAGE DELAYS FOR QPBS

IAT1	IAT2	Delay1	Delay2
1.5	4.5	18.5	20
1.5	6	9.7	13
1.5	7.5	8.1	11.5
1.5	9	7	11.2
1.5	10.5	6.5	11
1.5	12	6.4	10

V. COMPARISON BETWEEN SCHEDULING SCHEMES

The purpose of this comparative study is to demonstrate the trade-off between the priority and fairness as achieved by QPBS. A comparative study encompassing all the existing scheduling disciplines is out of the scope of this work.

As the last step of the multiple queues scenario, we have plotted the delay trade-off curves of the QPBS and strict priority schemes as shown in Fig 7. The red line represents the strict priority scheme and the blue line represents the QPBS

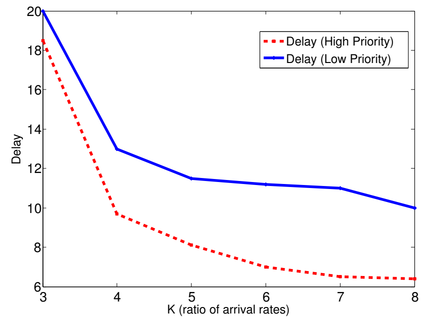

Fig. 5. QPBS low priority and high priority delays

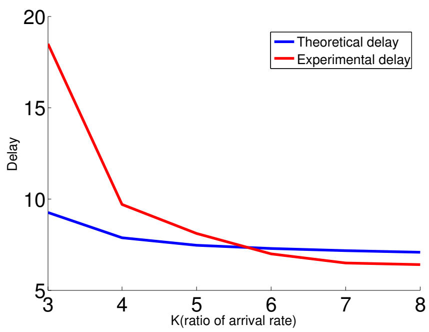

Fig. 6. QPBS experimental value comparison with PWRR theoretical value

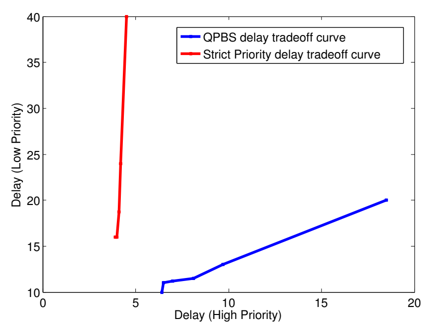

Fig. 7. Comparison of low priority and high priority delays for QPBS and strict priority scheme

scheme. This diagram brings out very important facts to us. First of all, as can be seen from the strict priority waveform that the slope of this waveform is very large. This implies very little trade-off between the high and low priority applications' delay. In other words the delay of the high priority application remains more or less a constant. On the other hand QPBS waveform has a lesser slope, which implies smooth trade-off between the high and low priority applications. In fact slope of QPBS waveform is close to one which implies that, the priority is implicit, i.e., priority within a batch which is visible from the positive intercept of QPBS waveform on y-axis.

VI. Conclusion

In this work our focus was on the analysis of the delay experienced by a packet in a cognitive radio environment following various scheduling schemes and the comparison among these schemes. The problem was simplified to a single channel problem interrupted by primary users. Hence we avoided the complexities of multiple channel scenarios, which can be interesting direction to work in the future. Since we have calculated the average delays, which can make sense only in the case of fixed or very slow variation in primary activity pattern, the effect of variation of the same is also under study.

All the simulations were carried out with artificial data. The real world scenario would incorporate practical considerations such as sense and contention delays, bursty losses and random outages; simulations with the trace of real time applications are another work of future.

Acknowledgment

This work was supported by the research projects titled "Design and development of a rapidly deployable Wimax based mesh network" (RP02177) and "SPARC: Spectrum Aware Rural Connectivity" (RP02565) funded by the Department of Information Technology, Government of India.

References

[1] Alan Cobham. Priority assignment in waiting line problems. *Operations Research*, 2(1):70–76, 1954.

[2] C. Cordeiro, K. Challapali, D. Birru, S. Shankar, et al. Ieee 802.22: an introduction to the first wireless standard based on cognitive radios. *Journal of communications*, 1(1):38–47, 2006.

[3] Awi Federgruen and Linda Green. Queueing systems with service interruptions. *Operations Research*, 34(5):752–768, 1986.

[4] L. Le and E. Hossain. A mac protocol for opportunistic spectrum access in cognitive radio networks. In *Wireless Communications and Networking Conference, 2008. WCNC 2008. IEEE*, pages 1426–1430. IEEE, 2008.

[5] Miguel Lopez-Benitez and Fernando Casadevall. Spectrum occupancy in realistic scenarios and duty cycle model for cognitive radio. *Advances in electronics and telecommunications*, 1(1):26–34, 2010.

[6] O. Mutlu and T. Moscibroda. Parallelism-aware batch scheduling: Enhancing both performance and fairness of shared dram systems. In *ACM SIGARCH Computer Architecture News*, volume 36, pages 63–74. IEEE Computer Society, 2008.

[7] K. Nichols, S. Blake, F. Baker, and D. Black. Rfc 2474: Definition of the differentiated services field (ds field) in the ipv4 and ipv6 headers. *Internet Engineering Task Force*, 1998.

[8] Hsien-Po Shiang and Mihaela van der Schaar. Queuing-based dynamic channel selection for heterogeneous multimedia applications over cognitive radio networks. *IEEE Transaction on Multimedia*, 10(5):896–909, 2008.

[9] H. Su and X. Zhang. Cream-mac: an efficient cognitive radio-enabled multi-channel mac protocol for wireless networks. In *World of Wireless, Mobile and Multimedia Networks, 2008. WoWMoM 2008. 2008 International Symposium on a*, pages 1–8. IEEE, 2008.

[10] Y. Zhang and PO Harrison. Performance of a priority-weighted round robin mechanism for differentiated service networks. In *Computer Communications and Networks, 2007. ICCCN 2007. Proceedings of 16th International Conference on*, pages 1198–1203. IEEE, 2007.