Performance Modeling of Finite-Source Cognitive Radio Networks

Béla Almási[*], Tamás Bérczes[*], Attila Kuki[*], János Sztrik[*], and Jinting Wang[†]

Abstract

This paper deals with performance modeling aspects of radio frequency licensing. The utilization of mobile cellular networks can be increased by the idea of the cognitive radio. Licensed users (Primary Users - PUs) and normal users (Secondary Users - SUs) are considered. The main idea is, that the SUs are able to access to the available non-licensed radio frequencies.

A finite-source retrial queueing model with two non independent frequency bands (considered as service units) is proposed for the performance evaluation of the system. A service unit with a priority queue and another service unit with an orbit are assigned to the PUs and SUs, respectively. The users are classified into two classes: the PUs have got a licensed frequency, while the SUs have got a frequency band, too but it suffers from the overloading. We assume that during the service of the non-overloaded band the PUs have preemptive priority over SUs. The involved inter-event times are supposed to be independent and exponentially distributed random variables.

The novelty of this work lies in the fact that we consider the effect of retrial phenomenon of SUs in performance modeling of radio frequency licensing by using a finite-source queueing model which takes the unreliability of radio transmission into account for the first time. In the literature, most work studied the performance of cognitive radio networks under a mixed spectrum environment of licensed and unlicensed bands where the blocked SUs and the preempted SUs are forced to leave the system forever when there are no idle channels in the system. But in practical situation, the blocked SUs and the preempted SUs may do not leave the system forever and try to continue their services after random amount of time.

By the help of an appropriate continuous time Markov chain using MOSEL (MOdeling Specification and Evaluation Language) tool several numerical examples are provided showing the effects of different input parameters on the main performance measures of the cognitive radio networks.

Our primary focus is to determine an optimal number of SUs, where at the

[*]University of Debrecen, Hungary, E-mail: `almasi.bela,berczes.tamas, kuki.attila, sztrik.janos@inf.unideb.hu`

[†]Beijing Jiaotong University, China E-mail: `jtwang@bjtu.cn`

secondary band the gained utilization, that is when switching to the cognitive radio, has a maximum value.

Keywords: cognitive radio networks, performance evaluation, finite-source retrial queueing systems, modeling tools

1 Introduction

The increasing demand for mobile communication has produced a bottleneck point in the radio frequency licensing. In order to increase the performance of mobile cellular networks the idea of the cognitive radio was introduced (see [15]). The users having "cognitive" wireless devices (called Secondary Users, SUs or Normal Users) are able to access to the available non-licensed radio frequencies. This access may not affect or disturb the communication of the licensed users (called Primary Users, PUs), i.e. the "cognitive" radio devices must intelligently release the unlicensed spectrum if a licensed user appears.

The teletraffic theory and modeling the attempt of repeated calls have been investigated widely in the past decades. Interesting result can be found on models, measurements, blocking probabilities, and so on, e.g. in [10], [13], [14], and [21]. The cognitive radio networks (CRN) have been a hot research area recently. As we could realize queueing theory can be successfully applied to establish mathematical models for variety of telecommunication systems (see e.g. [1], [2], [3], [6], [7], [8], [11], [12], [16], [17], [18],). These models can be used to calculate performance measures like mean delay, mean waiting time, utilization of the frequency bands etc. In many cases infinite preemptive priority queueing models are applied (see for example [5],[18]) for modeling and analysis. As the cognitive radio environments are used also in small sized radio network cells (picocells or femtocells) finite- source queueing models can be also appropriate for modeling. Recently in [19] Wong and Foh has introduced a finite-source queueing model to study the performance of a CRN with one frequency band (i.e. one queueing service station). This paper gave us an inspiration to develop a more realistic model.

In the present paper we introduce a finite-source queueing model with two (non independent) frequency bands (channel, server). As it is widely used in CRN modeling (see e.g. [8]) the users are classified into two classes: the Primary Users (PUs) have got a licensed frequency, which does not suffer from overloading feature. The Secondary Users (SUs, or Normal Users) have got a frequency band too, but it suffers from the overloading. If the band of the SUs is engaged then a newly arriving SU request may use the band of the PUs (which is not licensed for the SUs) in a cognitive way: the non-licensed frequency must be released by the SU if a PU request appears. In our environment the band of the PUs is modeled by a queue, where the requests from the PUs' class has preemptive priority over the SUs' request: the SU's request is subject to an immediate "handover" from the PU's band to the SU's band if a PU's request appears. The band of the SUs is described by a retrial queue: if the band is free when the request arrives then it is

transmitted. Otherwise the request goes to the Orbit if both bands are busy. The transmission of the requests from the orbit will be retried after a random holding time.We assume that the radio transmission is not reliable, that is the transmission will fail with a non-zero probability p for both channels. If this happens then the request retransmission process will start immediately.

Hence it should be noted that the novelty of our work lies in the fact that we consider the effect of retrial phenomenon of SUs in performance modeling of radio frequency licensing by using a finite-source queueing model which takes the unreliability of radio transmission into account for the first time. In the literature, most work studied the performance of cognitive radio networks under a mixed spectrum environment of licensed and unlicensed bands where the blocked SUs and the preempted SUs are forced to leave the system forever when there are no idle channels in the system. But in practical situation, the blocked SUs and the preempted SUs may do not leave the system forever and try to continue their services after random amount of time. This is a motivation of our research and model improvement.

The rest of the paper is organized as follows. Section 2 describes the precise mathematical model when a multi-dimensional continuous time Markov-chain is defined for describing the system's dynamics. Formulas of the most important performance measures are also discussed here. For presentation of numerical results the MOdeling Specification and Evaluation Language (MOSEL see [4]) tool is used. In Section 3 the most important system characteristics (utilizations, response times) are determined in order to verify the model. Series of experience are carried out to study how the number of the SUs' influences the average gain of the cognitive radio related to the non-cognitive one. At the end of this section some empirical optimization results are shown. Finally, the paper ends with conclusions.

2   System Model

A finite source queueing system, as illustrated in Fig.1 is used to model the considered cognitive radio network. The queueing system contains two interconnected, not independent sub-systems. The first part is for the requests of the PUs. The number of sources is denoted by N_1. These sources generate high priority requests with exponentially distributed inter-request times with parameter λ_1. The generated requests (jobs, packets) are sent to a single server unit (Primary Channel Service - PCS) with a preemptive priority queue. The service times are exponentially distributed with parameter μ_1. The second part is a finite-source retrial queueing system. The requests from the SUs are generated here. There are N_2 sources, the inter-request times are assumed to be exponentially distributed random variables with parameter λ_2. The single server unit (Secondary Chanel Service - SCS) works according to exponentially distributed service times with parameter μ_2.

A generated high priority packet goes to the primary service unit (i.e. to the radio band licensed by the PUs). If the unit is idle, the service of the packet begins immediately. If the server (i.e. radio channel) is engaged with a high priority
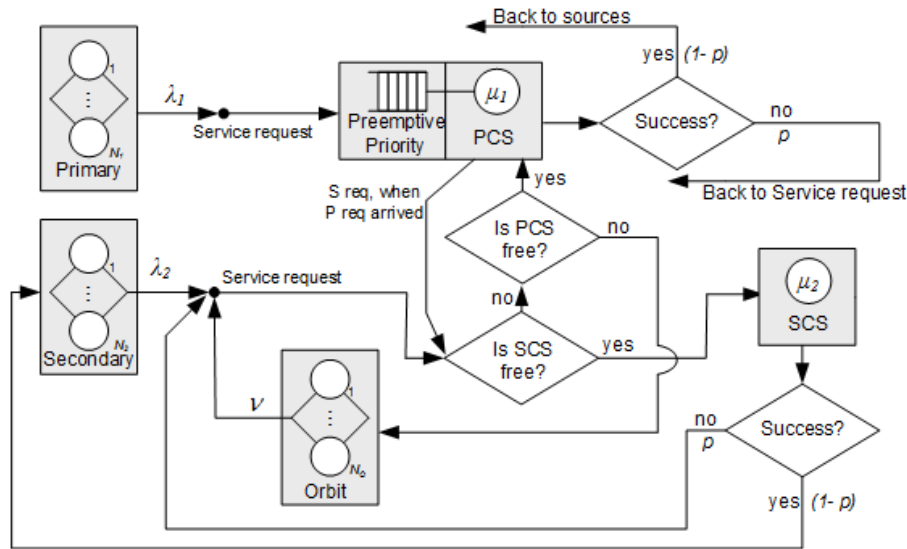
Figure 1: A priority and a retrial queue with components

request, the packet joins to the preemptive priority queue. When the unit is engaged with a request from SUs, the service is interrupted and the interrupted low priority task is sent back to the SCS. Depending on state of the secondary channel the interrupted job is directed to either the server or the orbit. At the PCS the service of the incoming request (which has interrupted the service of the task from SUs) begins. The transmission through the radio channel may produce errors, which can be discovered after the transmission (i.e. after the service). In the model this case has a probability p, and the failed packet is sent back to the appropriate service unit (high priority request to the PCS, low priority request to the SCS). When the submission is successful (with probability $1-p$), the request goes back to the source (to the PUs or SUs, depending on the packet, respectively).

A generated request from SUs looks for the secondary service unit. If the SCS is idle, the service of the task begins immediately. If the SCS is busy, the packet checks the primary service unit. In case of an idle PCS, the service of the low priority packet begins at the high priority channel (PCS). If the PCS is busy, the packet goes to the orbit. From the orbit it retries to be served after an exponentially distributed time with parameter ν. After service, the same transmission failure with the same probability can occur as in the PCS segment. Here the failed request is sent back to the SCS, the token of the successfully submitted packet goes back to the secondary source.

This functionality of the system can be seen on Fig.1.

To create a stochastic process describing the behavior of the system the following notations are introduced

- $k_1(t)$ is the number of high priority sources at time t,

- $k_2(t)$ is the number of low priority (normal) sources at time t,

- $q(t)$ denotes the number of high priority requests in the priority queue at time t,

- $o(t)$ is the number of requests in the orbit at time t.

- $y(t) = 0$ if there is no job in the PCS unit, $y(t) = 1$ if the PCS unit is busy with a job coming from the high priority class, $y(t) = 2$ when the PCS unit is servicing a job coming from the secondary class at time t

- $c(t) = 0$ when the SCS unit is idle and $c(t) = 1$, when the SCS is busy at time t.

It is easy to see that

$$k_1(t) = \begin{cases} N_1 - q(t), & y(t) = 0, 2 \\ N_1 - q(t) - 1, & y(t) = 1 \end{cases}$$

and

$$k_2(t) = \begin{cases} N_2 - o(t) - c(t), & y(t) = 0, 1 \\ N_2 - o(t) - c(t) - 1, & y(t) = 2. \end{cases}$$

The network input parameters are collected in Table 1.

Table 1: List of network parameters

Parameter	Maximum	Value at t
Active primary sources	N_1	$k_1(t)$
Active secondary sources	N_2	$k_2(t)$
Primary generation rate		λ_1
Secondary generation rate		λ_2
Total gen. rate	$\lambda_1 N_1 + \lambda_2 N_2$	$\lambda_1 k_1(t) + \lambda_2 k_2(t)$
Requests in priority queue	$N_1 - 1$	$q(t)$
Requests in orbit	$N_2 - 1$ (orbit size)	$o(t)$
Primary service rate		μ_1
Secondary service rate		μ_2
Retrial rate		ν
Error rate		p

To obtain the steady-state probabilities and performance measures within the Markovian framework, the mathematical tractability of the proposed model should be preserved. Therefore, we follow the classical approach frequently applied in the theory of retrial queues for the performance evaluation of wireless cellular networks, namely, the distributions of inter-event times (i.e., request generation times for low and high priority jobs, service time, retrial time) presented in the system are assumed to be exponentially distributed and totally independent of each other.

Hence the state of the network at a time t can be described by a Continuous Time Markov Chain (CTMC) with 4 dimensions:

$$X(t) = (y(t), q(t); c(t), o(t))$$

The steady-state distributions are denoted by

$$P(y, q, c, o) = \lim_{t \to \infty} P(y(t) = y, q(t) = q, c(t) = c, o(t) = o)$$

Note that in the present case, the unique stationary distribution always exists, because the underlying CTMC is irreducible and the state space of the CTMC is finite. The computation of this distribution is described e. g. in [9]. For computing the steady-state probabilities and the system characteristics, the MOSEL-2 software tool is used. These computations are similar to the ones described in, for example [20].

As soon as we have calculated the distributions defined above, the most important steady-state system performance measures can be obtained in the following way

- *Utilization of the primary server with respect to primary users*

$$U_{11} = \sum_{q=0}^{N_1-1} \sum_{c=0}^{1} \sum_{o=0}^{N_2-c} P(1, q, c, o)$$

- *Utilization of the primary server with respect to secondary users*

$$U_{12} = \sum_{q=0}^{N_1-1} \sum_{c=0}^{1} \sum_{o=0}^{N_2-c} P(2, q, c, o)$$

- *Utilization of the primary server*

$$U_1 = U_{11} + U_{12}$$

- *Utilization of the secondary server*

$$U_2 = \sum_{y=0}^{2} \sum_{q=0}^{N_1-1} \sum_{o=0}^{N_2-1} P(y, q, 1, o)$$

- *Average number of jobs in queue*

$$\overline{Q} = \sum_{y=0}^{2} \sum_{q=0}^{N_2-1} \sum_{c=0}^{1} \sum_{o=0}^{N_2-c} qP(y, q, c, o)$$

- *Average number of jobs in the orbit*

$$\overline{O} = \sum_{y=0}^{2} \sum_{q=0}^{N_2-1} \sum_{c=0}^{1} \sum_{o=0}^{N_2-c} oP(y,q,c,o)$$

- *Average number of jobs of primary users in the network*

$$\overline{M_1} = \overline{Q} + U_{11}$$

- *Average number of jobs of secondary users in the network*

$$\overline{M_2} = \overline{O} + U_{12} + U_2$$

- *Average number of jobs in the network*

$$\overline{M} = \overline{M_1} + \overline{M_2}$$

- *Average number of active primary users*

$$\overline{\Lambda_1} = N_1 - \overline{M_1}$$

- *Average number of active secondary users*

$$\overline{\Lambda_2} = N_2 - \overline{M_2}$$

- *Average generation rate of primary users*

$$\overline{\lambda_1} = \lambda_1 \overline{\Lambda_1}$$

- *Average generation rate of secondary users*

$$\overline{\lambda_2} = \lambda_2 \overline{\Lambda_2}$$

- *Mean response time of primary user's jobs*

$$\overline{T_1} = \frac{\overline{M_1}}{\overline{\lambda_1}}$$

- *Mean response time of secondary user's jobs*

$$\overline{T_2} = \frac{\overline{M_2}}{\overline{\lambda_2}}$$

Table 2: Numerical values of model parameters

Case studies								
No.	N_1	N_2	λ_1	λ_2	μ_1	μ_2	ν	p
Fig. 2	10, 20	50	$x-axis$	0.03	1	1	20	0.1
Fig. 3	10	$x-axis$	0.02	0.03	1	1	20	0.1
Fig. 4	10	10	0.02, 0.08	0.03, 0.08	1	1	$x-axis$	0.1
Fig. 5	10, 20	50	$x-axis$	0.03	1	1	20	0.1
Fig. 6	10	$x-axis$	0.02	0.03	1	1	20	0.1
Fig. 7	10	$x-axis$	0.02	0.03	1	1	20	0.1
Fig. 8	10	$x-axis$	0.02	0.03	1	1	20	0.1
Fig. 9	10	$x-axis$	0.02	0.03	1	1	20	0.1
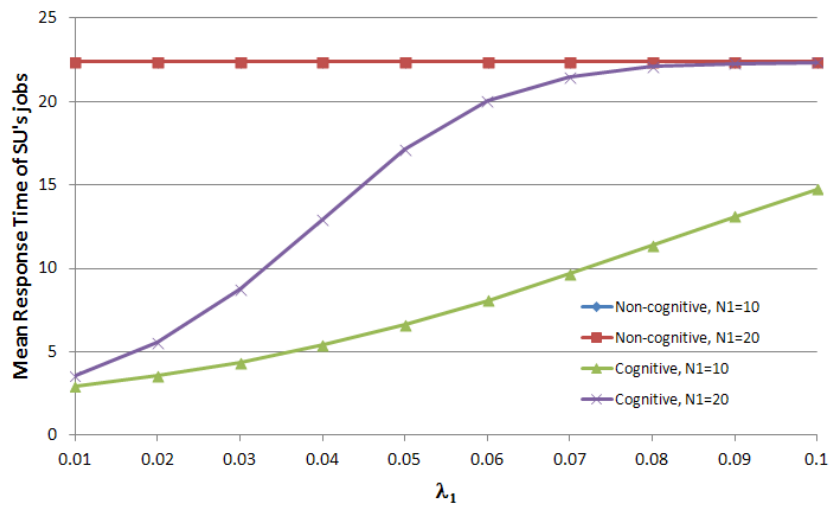
Figure 2: Mean Response Time of SU's jobs vs. λ_1
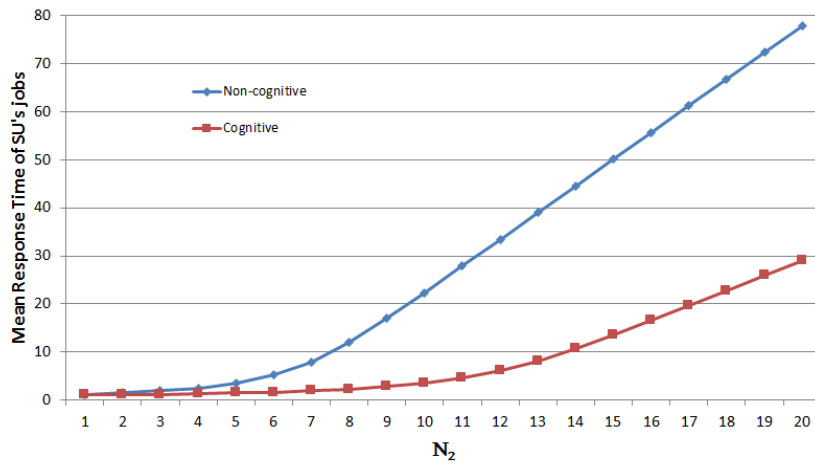
Figure 3: Mean Response Time of SU's jobs vs. N_2
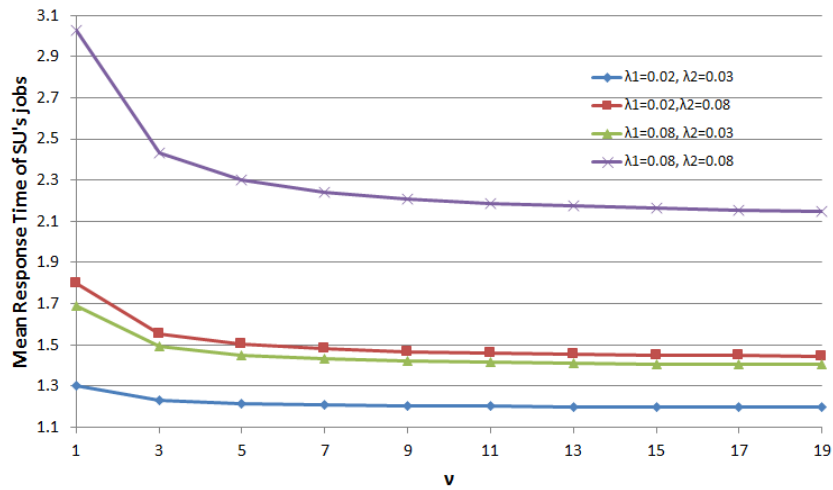
Figure 4: Mean Response Time of SU's jobs vs. ν
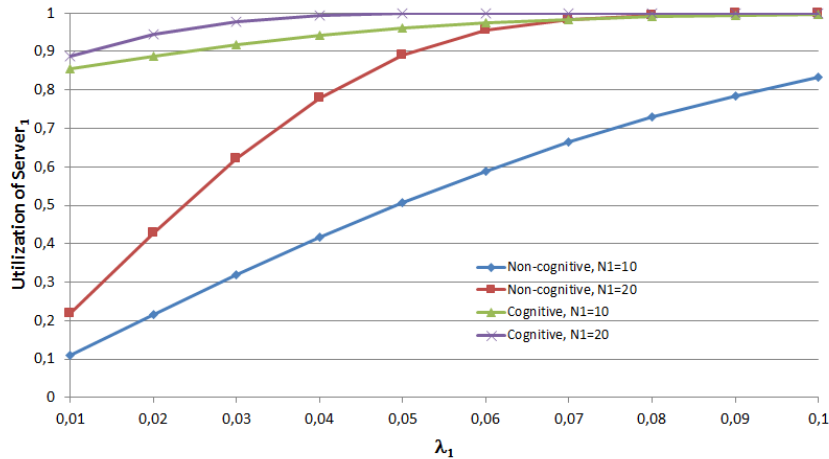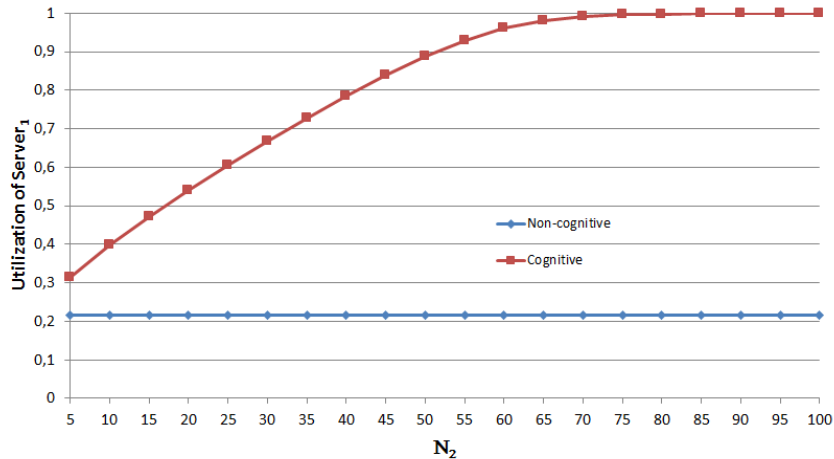
Figure 5: Utilization of PCS vs. λ_1

3 Numerical results

Investigating the functionality and the behavior of the system several numerical calculations were performed. From the steady-state probabilities computed by MOSEL-2 tool the most interesting performance characteristics were obtained, which are graphically presented in this section. The numerical values of model parameters are collected in Table 2.

In the calculations two operational modes were applied. The Non-Cognitive Radio Mode (Non-cognitive on figures), where the secondary users are not enabled to use the high priority primary channel, and the Cognitive Radio Mode (Cognitive on figures), where the secondary users can use the idle primary channel.

On Figure 2 the mean response times of the secondary jobs are displayed as a function of the primary generation rate. Beside the cognitive and non-cognitive radio feature, two different values of the number of primary sources are introduced. In the non-cognitive cases the two subsystems (the flow of primary and secondary jobs) are independent, thus the value of λ_1 and the number of primary source have no effect for this system characteristic. The two non-cognitive lines are identical. In the cognitive case it can be seen, that increasing the value of λ_1, the secondary jobs have less opportunities to use the primary channel, the response times significantly increase.

Figure 3 shows the mean response times of the secondary jobs in cases of different numbers of the secondary source. Natural way, the more secondary users try to use the system, the greater the response time is. Comparing the cognitive and the non-cognitive results, the benefit (gain) of the cognitive radio network is displayed. For low numbers of N_2 the response times are almost the same, but increasing the number of the secondary source, the difference between the response times

Figure 6: Utilization of PCS vs. N_2

increases, as well.

On Figure 4 the effect of the time spent in orbit (it was modeled by a variable retrial rate) to the response time of SCS is displayed.

On Figure 5 the utilization of the PCS is displayed as a function of λ_1. For small values of the primary generation rate the two cognitive lines are highly above the non-cognitive ones, because in cognitive case the idle periods of the PCS are filled up with jobs from the secondary source.

The next figure (Figure 6 the utilization of PCS is shown as a function of the numbers of secondary source. In non-cognitive case the number of N_2 has no effect for this characteristic. The jobs from primary and secondary sources work independently. The utilization is constant. In cognitive case the increasing number of the secondary source increases the utilization of PCS, more and more jobs from the secondary source try to use the PCS.

On Figure 7 the same investigation is performed for the SCS. In non-cognitive case the server load reaches the maximum utilization soon, because the jobs are not able to use the PCS. In cognitive case the PCS also can serve secondary jobs, thus the utilization of SCS will reach the value of 1 only at higher numbers of secondary source.

Figure 8 shows the relative difference (the difference is divided by N_2) of mean orbit size in non-cognitive and cognitive cases. This is the increment of orbit size for one user in function of number of secondary users. When the size of the orbit fills up, this increment has a local maximum value. After this point, the increment will be decreasing, because the orbit still remains full and the number of users are increasing.

On the last figure (Figure 9) the difference of utilizations of SCS (in non-cognitive and cognitive cases) is displayed. With these parameter setup the optimal
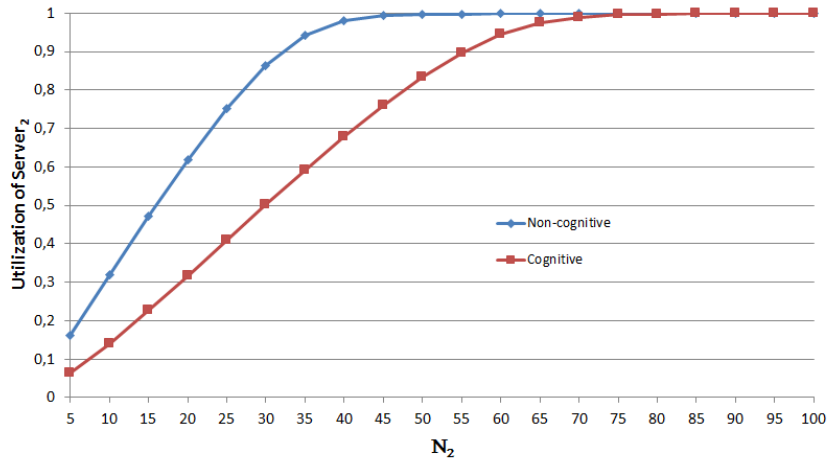
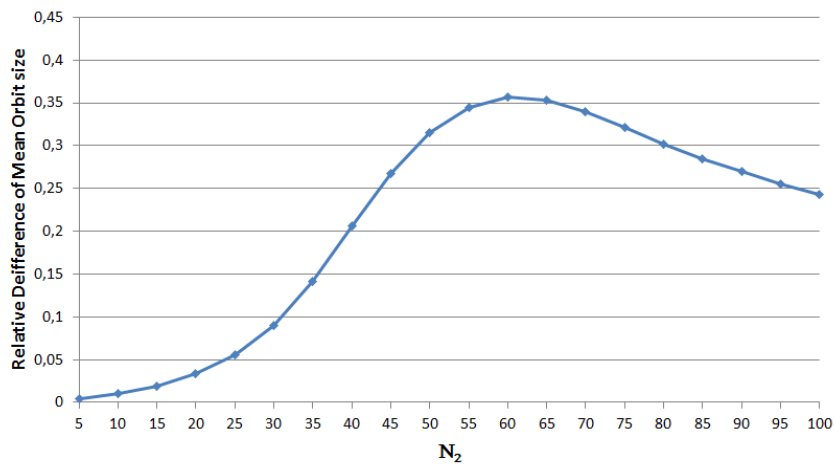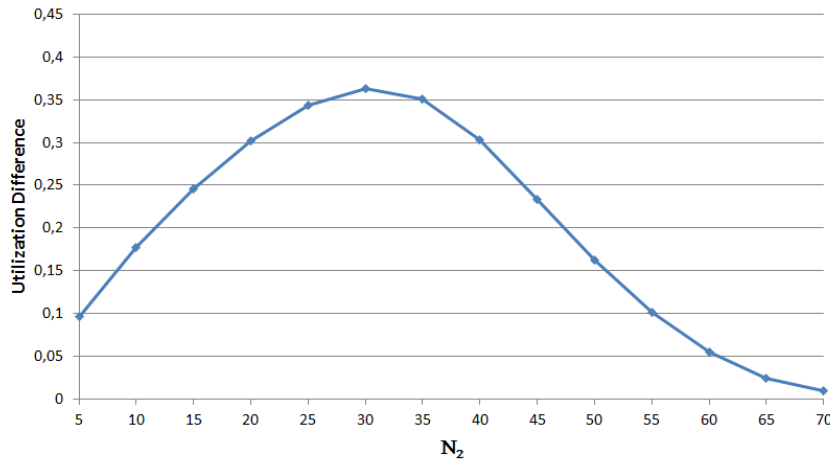Figure 7: Utilization of SCS vs. N_2

Figure 8: Relative Difference of Mean Orbit Size vs. N_2

Figure 9: Difference of utilization of SCS vs. N_2

point of utilization gains can be found at 30, i.e. this is the number of secondary users, when it is worth to change to the cognitive radio devices.

4   Conclusions and further questions

In this paper a novel finite-source retrial queueing model was proposed with two bands servicing primary and secondary users in a cognitive radio network. Primary users have preemptive priority over the secondary ones in servicing at primary channel. At the secondary channel an orbit is installed for the secondary jobs finding the server busy upon arrival. The MOSEL tool was used to carry numerical calculations illustrating the effect of some input parameters on the main performance measures. The primary focus was to determine an optimal number of secondary users, where at the secondary band the gained utilization, that is when switching to the cognitive radio, has a maximum value.

Interesting further works could be to investigate the case of more than one frequency chanels. Other interesting problem is, when the users are able to switch their classes, i.e. a secondary user can advance into the class of primary users. Finally, other distributions, e.g. hipo-exponential distribution can be investigated, as well.

5   Acknowledgments

References

[1] Artalejo, J.R. Accessible bibliography on retrial queues. *Mathematical and Computer Modelling*, 30:1–6, 1999.

[2] Artalejo, J.R. and Gomez-Corral, A. *Retrial Queueing Systems*. Springer, Berlin, 2008.

[3] Basharin, G. P., Samouylov, K. E., Yarkina, N. V., and Gudkova, I. A. A new stage in mathematical teletraffic theory. *Automation Remote Control*, 70(12):1954–1964, 2009.

[4] Begain, K., Bolch, G., and Herold, H. *Practical Performance Modeling, Application of the MOSEL Language,*. Kluwer Academic Publisher, Boston, 2001.

[5] Devroye, N., Vu, M., and Tarock, V. Cognitive radio networks. *IEEE Signal Processing Magazine*, 25:12–23, 2008.

[6] Do, T. Van, Chakka, R., and Sztrik, J. Spectral expansion solution methodology for QBD-M processes and applications in Future Internet engineering. *Advanced Computational Methods for Knowledge Engineering, Studies in Computational Intelligence*, 479:131–142, 2013.

[7] Feng, W. and Kowada, M. Performance analysis of wireless mobile networks with queueing priority and guard channels. *International Transactions on Operational Research*, 15:481–508, 2008.

[8] Gao, S. and Wang, J. Performance analysis of a cognitive radio network based on a preemptive priority retrial queue. *International Journal of Computer Mathematics*, GCOM-2013-0621-A, 2014.

[9] Gross, D.; Shortie, J. F.; Thompson J. M. & Harris C. M. *Fundamentals of Queueing Theory, Fourth Edition*. John Wiley & Sons, Inc., 2008.

[10] Jonin, G. L. The systems with repeated calls: Models, measurements, results. *Fundamentals of Teletraffic Theory. Proceedings of the Third International Seminar on Teletraffic Theory*, pages 197–208, 1980.

[11] Lakatos, L., Szeidl, L., and Telek, M. *Introduction to queuing systems with telecommunication applications*. Springer, Heidelgerg, 2013.

[12] Lee, Y., Park, C. G., and Sim, D. B. Cognitive radio spectrum access with prioritized secondary users. *Applied Mathematics & Information Sciences*, 6:595–601, 2012.

[13] Marsan, M. A.; De Carolis, G.; Leonardi E.; Lo Cigno R. & Meo M. An approximate model for computation of blocking probabilities in cellular networks with repeated calls. *Telecommunication Systems*, 15:53–62, 2000.

[14] Marsan, M. A.; De Carolis, G.; Leonardi E.; Lo Cigno R. & Meo M. Efficient estimation of call blocking probabilities in cellular mobile telephony networks with customer retrials. *IEEE Journal on Selected Areas in Communications*, 19:332–346, 2001.

[15] Mitola, J. *Cognitive radio: an integrated agent architecture for software defined radio.* PhD thesis, KTH, 2000.

[16] Ponomarenko, L., Kim, C. S., and Melikov, A. *Performance analysis and optimization of multi-traffic on communication networks.* Springer, Heidelberg, 2010.

[17] Stasiak, M., Glabowski, M., Wishniewski, A., and Zwierzykowski, P. *Modeling and dimensioning of mobile networks.* John Wiley, Chichester, 2011.

[18] Wang, L.C., Wang, C.W., and Adachi, F. Load-balancing spectrum decision for cognitive radio networks. *IEEE Journal on Selected Areas of Communications*, 29:757–769, 2011.

[19] Wong, E. W. M. and Foh, C. H. Analysis of cognitive radio spectrum access with finite user population. *IEEE Communications Letters*, 13, 2009.

[20] Wüchner, P., Sztrik, J., and de Meer, H. Modeling wireless sensor networks using finite-source retrial queues with unreliable orbit. *Springer Lecture Notes in Computer Science*, 6821:275–285, 2011.

[21] Zukerman, M. & Chia Min Lee. Performance bounds for cellular mobile communications networks with repeated attempts. *Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS 53rd*, 2:996–1000, 2001.