Approximate Analysis of Cognitive Radio Systems Using Time-Scale Separation and Its Accuracy

Jorge Martinez-Bauset, Vicent Pla, Jose R. Vidal and Luis Guijarro,

Abstract—We model a cognitive radio system as a quasi-birth-death (QBD) process and determine its performance parameters. We also model the system at the quasi-stationary limiting regime. We show that this regime defines the asymptotic system behavior. The performance parameters of interest at this regime are independent of the service time distributions and can be determined by simple recursions. We propose and evaluate a new methodology to determine when the quasi-stationary approximation can be considered a good approximation of the actual system behavior. It requires low computational cost and does not require to solve the exact system.

Index Terms—Cognitive radio networks, traffic analysis, quasi-stationary approximation.

I. Introduction

ANALYZING systems with different user types can be problematic as the higher dimensionality may render the analysis problem computationally intractable. However, when the dynamics of different user types operate at sufficiently separated time-scales one can resort to highly efficient approximations based on *time-scale decomposition*, which can greatly simplify the computations. The interested reader is referred to the seminal work of [1] or to a more recent application example in [2]. The referred technique is especially suited to cognitive radio networks (CRNs), as primary user (PU) transmissions being relatively static to secondary user (SU) ones has been identified as one of the conditions that may lead to a successful deployment of CRNs technology [3].

Our work has been partly motivated by the study presented in [4]. There, a continuous-time Markov chain (CTMC) model is proposed to evaluate the performance of a CRN with dedicated secondary channels. Their model is approximate and is implicitly based on the idea of time-scale decomposition. The same approach is used in [5], but there a delay model is used to study the packet level time scale, whereas we use a loss model and focus at the session level.

The main objective of our study is to evaluate the system performance from the traffic perspective. We propose and analyze two different CTMC models for the system under consideration. The first model is the exact one and corresponds to the case where no time-scale separation is assumed. The second model corresponds to the limiting regime, one in which the PU events occur at a much lower time scale than SU

The authors are with with the Department of Communications, Universidad Politécnica de Valencia (UPV), ETSIT Camí de Vera s/n, 46022 Valencia, Spain. Email: {jmartinez,vpla,jrvidal,lguijar}@dcom.upv.es

events, i.e. the *quasi-stationary* (QS) *regime* [2]. We show that the stationary distribution of the number of SUs in the system at this limiting regime is insensitive to the service time distribution beyond the mean. Therefore, all performance parameters derived from it are also insensitive. In addition, the stationary distribution at the limiting regime can be obtained by simple recursions.

As the QS regime is particularly suited for CRNs, we study the accuracy of approximating the stationary distribution of the system at its actual regime by the one at the QS regime. We refer to it as the *QS approximation*. We propose a new methodology to determine when the QS approximation can be considered a good approximation of the actual system behavior. It requires low computational cost and does not require to solve the exact system. In [4], [5] the QS approximation is used without justifying its range of applicability, i.e., for which range of the system parameters the QS approximation yields accurate results. We provide an interesting numerical study on the impact that varying the ratio of time-scales between PUs and SUs has on the accuracy of the QS approximation and on the system performance. In our study, we keep all other system parameters constant, i.e., traffic offered by PUs and SUs, number of channels, spectrum access scheme, spectrum handover capabilities, etc. This study allows us to relate the real system operating point with the QS limiting regime, which was not done before.

The analytical results are validated by comparison against the results of a simulation model that mimics the physical behavior of the system and therefore it is completely independent from the CTMC models. Moreover, we compare our results with those in [4] and show that the QS approximation in our study yields significantly more accurate results than the one developed there. As in [4], we consider a scenario where there is either a central or distributed control entity to perform channel allocations to SUs. This is one of the potential deployment scenarios for CRNs [6]. The existence of a control entity has been assumed before, see for example [5], [4], [7], [8]. We approach the problem from the traffic perspective, as we believe that the traffic management techniques complement those defined at the physical layer [9], [10].

II. Model Description

We model the PU and SU traffic at the session (connection) level and ignore interactions at the packet level (scheduling, buffer management, etc.). We assume an ideal MAC layer for SUs, which allows a perfect sharing of the allocated channels among the active SUs (all active SUs get the same bandwidth
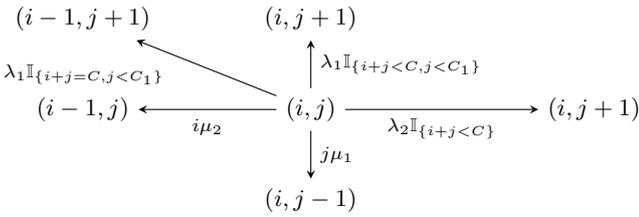
Fig. 1. State and transition diagram.

portion), introduce zero delay and whose control mechanisms consume zero resources. In addition, we also assume that an active SU can sense the arrival of a PU in the same channel instantaneously and reliably. In this sense, the performance parameters obtained can be considered as an upper bound.

The system has C_1 *primary channels* (PCs) that can be shared by PUs and SUs, and C_2 *secondary channels* (SCs) only for SUs. Let $C = C_1 + C_2$ be the total number of channels in the system. Note that the SCs can be obtained from unlicensed bands, e.g. as proposed in [4]. This assumption is applicable to the *coexistence* deployment scenario for CRNs [6]. Alternatively, as it might be of commercial interest for the primary and secondary networks to *cooperate*, the secondary channels may be obtained based on an agreement with the primary network [6].

A SU in the PCs might be forced to vacate its channel if a PU claims it to initiate a new session. As SUs support *spectrum handover*, a vacated SU can continue with its ongoing communication if a free channel is available. Otherwise, it is *forced to terminate*. At a SU arrival, the SU selects one free channel in either set with equal probability, provided that there are free channels in both channel sets [4]. The algorithm used by the SUs to select channels is irrelevant for the performance parameters of interest, as spectrum handover is supported. On the other hand, using a selection scheme that chooses a free SC as first option, and resorts to occupy a PC only when all SCs are occupied, will reduce the interference caused to PUs and the rate of spectrum handovers.

Poisson arrivals and exponentially distributed service times are assumed, as in [4]. In addition, we also study the impact of service time distributions other than the exponential. The arrival rate for PU (SU) sessions is λ_1 (λ_2), their service rate is μ_1 (μ_2), and requests consume 1 (1) channel when accepted.

We denote by (i, j) the system state, when there are i ongoing SU sessions and j PU sessions. The set of feasible states is $\mathcal{S} := \{(i, j) : 0 \leq i + j \leq C, 0 \leq j \leq C_1\}$. The state and transition diagram of the system is depicted in Fig. 1. We denote by $\mathbb{I}_{\{\}}$ an indicator function that is 1 when the condition in braces is satisfied, and 0 otherwise.

If we partition the state space into *levels* and the set of states in *level* i is $\mathcal{L}(i) := \{(i, j) : 0 \leq j \leq \min(C_1, C - i)\}$ the CTMC becomes a quasi-birth-death (QBD) process [11], where transitions can only occur between adjacent levels. The transition rate matrix $\mathbf{Q}$ can be written in block-tridiagonal form. Let $\mathbf{A}_1^{(i)}$ (respectively, $\mathbf{A}_2^{(i)}$ and $\mathbf{A}_0^{(i)}$) the blocks in the main (respectively, lower and upper) diagonal, where $i = 0, \dots, C$ denotes the row block of $\mathbf{Q}$ starting at 0. More

specifically, $\left[\boldsymbol{A}_2^{(i)}\right]_{(k,l)}$ is the transition rate from (i, k) to $(i-1, l)$ $\left[\boldsymbol{A}_0^{(i)}\right]_{(k,l)}$ is the transition rate from (i, k) to $(i+1, l)$ and $\left[\boldsymbol{A}_1^{(i)}\right]_{(k,l)}$ is the transition rate from (i, k) to (i, l) if $k \neq l$, or, if $k = l$, $-\left[\boldsymbol{A}_1^{(i)}\right]_{(k,k)}$ is the total outgoing rate from state (i, k). The matrices $\boldsymbol{A}_0^{(i)}, \boldsymbol{A}_1^{(i)}, \boldsymbol{A}_2^{(i)}$, which are not shown here due to the lack of space, can be easily constructed from the state transition diagram in 1.

By applying the *Linear Level Reduction* algorithm [12], which can solve level-dependent finite QBDs in an efficient manner, the stationary distribution is obtained from $[\boldsymbol{\pi}^{(0)} \cdots \boldsymbol{\pi}^{(Q)}]\mathbf{Q} = \mathbf{0}$ and $\sum_{i=0}^{C_1+C_2} \boldsymbol{\pi}^{(i)}\mathbf{e} = 1$, where $\boldsymbol{\pi}^{(i)} = [\pi(i,0) \dots \pi(i,\alpha)]$, $\pi(i,j)$ is the stationary probability of state (i, j), $\alpha = \min(C_1, C - i)$ and $\mathbf{e}$ is a column vector of 1's of the appropriate size.

The system performance parameters are determined as follows,

$$P_1 = \sum_{i=0}^{C_2} \pi(i, C_1), \quad P_2 = \sum_{i=C_2}^{C_2+C_1} \pi(i, C_1 + C_2 - i), \quad (1)$$

$$P_{ft} = \sum_{i=C_2+1}^{C_2+C_1} \frac{\lambda_1 \pi(i, C_1 + C_2 - i)}{\lambda_2(1 - P_2)} = \frac{\lambda_1(P_2 - \pi(C_2, C_1))}{\lambda_2(1 - P_2)}, \quad (2)$$

$$Th_2 = \sum_{i=1}^{C_1+C_2} i\mu_2 \cdot \boldsymbol{\pi}^{(i)}\mathbf{e}, \quad (3)$$

where P_1 is the PUs blocking probability, which clearly coincides with the one obtained in an Erlang-B loss model with C_1 servers; P_2 is the SUs blocking probability, i.e. the fraction of SU sessions that upon arrival find the system full and are rejected; P_{ft} is the forced termination probability of the SUs, i.e. the rate of SU sessions forced to terminate divided by the rate of accepted SU sessions; and Th_2 is the SUs throughput, i.e the rate of SU sessions successfully completed.

A. QS Limiting Regime

In the *QS regime* it can be assumed that the distribution of the number of SUs in the system reaches equilibrium between consecutive PU events. As PU events are very slow with respect to the SU events, the fraction of SU preemptions is negligible ($P_{ft}^{qs} \approx 0$). Then we can write,

$$\pi(i, j) = \pi_1(j) \cdot \pi_2(i|j). \quad (4)$$

As the PUs have priority over the SUs, $\pi_1(j)$ is the stationary probability of finding j ongoing sessions in an $M/M/C_1/C_1$ system with only PUs. Also, $\pi_2(i|j)$ is the stationary probability of finding i ongoing sessions in an $M/M/C-j/C-j$ system with only SUs. Both $\pi_1(j)$ and $\pi_2(i|j)$ can be determined independently using simple recursions, since their corresponding CTMC are one-dimensional birth-and-death processes.

At the QS regime, the PUs blocking probability is $P_1^{qs} = \pi_1(C_1)$, the SUs blocking probability P_2^{qs} can be determined by expression (1) and the SUs throughput Th_2^{qs} by (3), but using distribution (4) in both cases. Note that $P_{ft}^{qs} \approx 0$. In the scenarios of interest a pure QS limiting regime is not achieved.
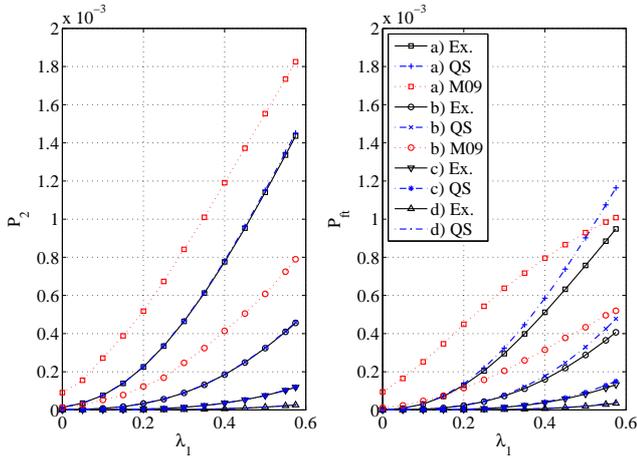
Fig. 2. Blocking and forced termination probabilities for SUs: a) $C_1 = 3$, $C_2 = 3$; b) $C_1 = 4$, $C_2 = 3$; c) $C_1 = 5$, $C_2 = 3$; d) $C_1 = 6$, $C_2 = 3$.

Fig. 3. SUs blocking and forced terminations probabilities as a function of the PUs event rate accelerating factor.

However, in some cases sufficient accuracy can be obtained by using the QS approximation. In those cases, we approximate P_2 by P_2^{qs} and Th_2 by Th_2^{qs}. To approximate P_{ft} we propose to use expression (2) but employing distribution (4). We denote it by $\hat{P}_{ft}$ and has a value bigger than zero in the scenarios of interest.

III. NUMERICAL EVALUATION

In [4] and [5] scalability issues have been used as an argument to deploy the QS approximate solution, as well as to consider systems with a small number of channels. However, the computation time to solve the proposed exact model (i.e., the QBD process) is below 1 s using a conventional laptop, even for systems several times larger than those in [4]. Nevertheless, in order to compare the results with the ones presented in [4], we use the same configurations used there, where $\lambda_2 = 0.2$, $\mu_1 = 0.5$ and $\mu_2 = 0.4$ users/s. Note also that, the computational complexity of the proposed exact model might make its solution unfeasible in different scenarios, like when very large number of channels or heterogeneous users are studied. In those cases, resorting to the QS approximation might be the only way to approach the system solution.

Figure 2 shows the SUs blocking and forced termination probabilities as a function of λ_1. For the scenarios shown, the QS approximation for the SUs blocking probability P_2^{qs} (lines marked as "QS") is excellent, as it practically overlaps the exact value P_2 (lines marked as "Ex."). On the other hand, the difference between the QS approximation for the SUs forced termination probability $\hat{P}_{ft}$ (lines marked as "QS") and P_{ft} (lines marked as "Ex.") becomes smaller as λ_1 decreases, i.e. when the rate of PU events is smaller than the rate of SU events. In other words, the closer the actual system regime is to the QS limiting regime, the more accurate $\hat{P}_{ft}$ is. Finally, note that the numerical results in [4] (lines marked as "M09") are not an accurate approximation to the actual system behavior.

In Fig. 3 we show P_2 and P_{ft} in a scenario where $C_1 = 3$, $C_2 = 3$, $\lambda_2 = 0.625$, $\mu_2 = 0.4$. These performance parameters are plotted as a function of an accelerating factor f, which is used to accelerate the arrival and departure events of the PUs,
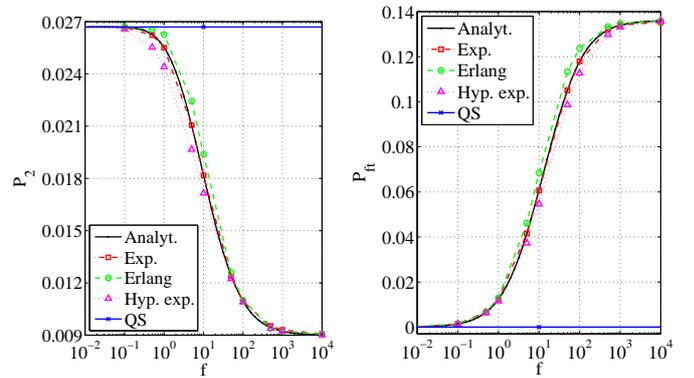
while keeping the offered traffic constant. For each value of f, the PU arrival and service rates are obtained as $\lambda_1 = f \cdot 0.5$ and $\mu_1 = f \cdot 0.5$.

In each figure we show the exact values (continuous line) for exponentially distributed service time distributions and simulation values for three different combinations of the service time distributions. Let s_1 (s_2) be the PUs (SUs) service time random variable. Then, the curves marked as "Expon", "Erlang" and "Hyperexp" have been obtained when s_1 and s_2 follow simultaneously exponential, Erlang or hyperexponential distributions, respectively. The last two with coefficients of variation 0.5 and 2.0, respectively. In all scenarios $E[s_1]$ and $E[s_2]$ are kept constant. In each figure we also plot one horizontal line that corresponds to the QS regime. As expected, when the accelerating factor f decreases the curves tend to the QS regime. It is interesting to note that the limiting regimes define asymptotes for the system behavior that are independent of the service time distributions.

One interesting observation is that in Fig. 3 P_2 increases and P_{ft} decreases as f decreases, i.e., as we approach the QS regime. Although not shown due to the lack of space, the shape of the curve describing the evolution of Th_2 with f is similar to that for P_2, i.e., it increases from 0.535 to 0.608 as f decreases. Then, the higher Th_2, the higher the effective Erlang capacity exploited by the SUs, which is 13.68% higher in the scenario of Fig. 3. We define the effective Erlang capacity as the mean number of channels used by SU communications that complete successfully. Note also that a reduction on P_{ft} (it tends to zero) is perceived by the SUs as an improvement on the QoS of the system, as, in general, it is more annoying for them that an ongoing communication is aborted that a set up request is blocked.

A. Accuracy of the QS Approximation

An intuitive condition to assume that the actual system regime is close to the QS limiting regime is when the rates λ_1 and μ_1 are much smaller than the rates λ_2 and μ_2 [2].

By inspecting the value of these rates in our scenario, and according to [2], we would conclude that the QS approximation would not lead to accurate results. However, our results show that it is quite accurate.

TABLE I
ACCURACY OF THE NEW METHODOLOGY

	P_{ft}	$\hat{P}_{ft}$	error	P_2	P_2^{qs}	error	Th_2/λ_2	Th_2^{qs}/λ_2	error
LL	$1.7 \cdot 10^{-3}$	$2.0 \cdot 10^{-3}$	$3.4 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$8.3 \cdot 10^{-5}$	0.9973	0.9989	0.0016
LH	$4.8 \cdot 10^{-2}$	$6.6 \cdot 10^{-2}$	$1.8 \cdot 10^{-2}$	$5.0 \cdot 10^{-2}$	$6.7 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$	0.9041	0.9334	0.0294
HL	$1.5 \cdot 10^{-3}$	$2.7 \cdot 10^{-3}$	$1.2 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$8.5 \cdot 10^{-5}$	0.9975	0.9989	0.0014
HH	$7.2 \cdot 10^{-2}$	$1.4 \cdot 10^{-1}$	$6.4 \cdot 10^{-2}$	$5.0 \cdot 10^{-2}$	$7.0 \cdot 10^{-2}$	$2.0 \cdot 10^{-2}$	0.8811	0.9295	0.0484
			adjusting f to achieve $\hat{P}_{ft} = 1 \cdot 10^{-3}$						
LL	$8.9 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$	$1.1 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$4.5 \cdot 10^{-5}$	0.9981	0.9989	0.0008
LH	$9.9 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$	$6.3 \cdot 10^{-6}$	$6.6 \cdot 10^{-2}$	$6.7 \cdot 10^{-2}$	$3.8 \cdot 10^{-4}$	0.9329	0.9334	0.0005
HL	$7.1 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$	$2.9 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$3.6 \cdot 10^{-5}$	0.9982	0.9989	0.0007
HH	$9.9 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$	$1.0 \cdot 10^{-5}$	$7.0 \cdot 10^{-2}$	$7.0 \cdot 10^{-2}$	$3.1 \cdot 10^{-4}$	0.9289	0.9295	0.0006

We propose an alternative methodology that determines with much better precision when the QS approximation leads to accurate results. It is based on checking how close is $\hat{P}_{ft}$ to zero. To determine $\hat{P}_{ft}$, we need to obtain the distribution (4), which is computed using simple recursions. Note that the exact distribution $\pi(i, j)$ *is not required*.

To evaluate the goodness of the new methodology, we studied the system with different number of channels and different loads. We used the following system sizes $\{(C_1, C_2)\} = \{(10, 5), (10, 10), (20, 10), (30, 10), (50, 50), (100, 50), (100, 100), (50, 100)\}$. Setting $\mu_1 = \mu_2 = 1$ we adjusted λ_1 and λ_2 to obtain two load conditions, low (L) and high (H), which correspond to blocking probabilities $1 \cdot 10^{-3}$ and $5 \cdot 10^{-2}$, respectively. As an example, we represent the results of the system $(30, 10)$ in Table I, where the characters AB in each row denote the load of PUs (A='L','H') and SUs (B='L','H'). We show two groups of results, one is for the actual system regime and the other is obtained when the acceleration factor f is adjusted to achieve $\hat{P}_{ft} = 1 \cdot 10^{-3}$. Note that we show Th_2/λ_2 and Th_2^{qs}/λ_2, which are the normalized throughputs for the actual and the QS regimes, respectively.

As observed in Table I, when $\hat{P}_{ft}$ is around $1 \cdot 10^{-3}$ or smaller for the actual regime, the accuracy of the QS approximation is very good. For those scenarios where the QS approximation leads to less accurate results, we show that it can be accurate again by decelerating the PU events until we achieve that $\hat{P}_{ft} = 1 \cdot 10^{-3}$. In other words, the proposed restriction on $\hat{P}_{ft}$ is a necessary and sufficient condition for the QS approximation to be accurate. We obtained similar results for the other system sizes, which can be summarized as follows. When $\hat{P}_{ft} = 1 \cdot 10^{-3}$, then the maximum errors obtained by using the QS approximations instead of the exact values are: $|P_2^{qs} - P_2| \leq 5.5 \cdot 10^{-4}$, $|\hat{P}_{ft} - P_{ft}| \leq 3.1 \cdot 10^{-4}$ and $|Th_2^{qs} - Th_2| / \lambda_2 \leq 0.0009$.

IV. CONCLUSIONS

We studied a cognitive radio network with two channel sets, one shared by primary and secondary users and the other dedicated to the secondary users. We modeled the system as a quasi-birth-death process, determined the common performance parameters and validated them by simulation. We also obtained the stationary distribution for the quasi-stationary limiting regime. We showed that the quasi-stationary regime defines the asymptotic system behavior when the rate at with events occur for PUs is much lower than rate at with events occur for SUs. At this limiting regime, the performance parameters are obtained with low computational cost and they are independent of the service time distribution. Finally, we proposed a new methodology to determine when the quasi-stationary approximation can be considered a good approximation of the actual system behavior. We provide intuitive and experimental evidences that show that a small estimated forced termination probability is a necessary and sufficient condition for the quasi-stationary approximation to be accurate.

REFERENCES

[1] P. Courtois, "Decomposability, instabilities, and saturation in multiprogramming systems," *Communications of the ACM*, vol. 18, no. 7, pp. 371–377, 1975.

[2] H.-P. Tan, R. Núñez-Queija, A. F. Gabor, and O. J. Boxma, "Admission control for differentiated services in future generation CDMA networks," *Performance Evaluation*, vol. 66, no. 9-10, pp. 488 – 504, 2009.

[3] C. Politis, "Managing the radio spectrum," *Vehicular Technology Magazine*, vol. 4, no. 1, pp. 20–26, 2009.

[4] H. Al-Mahdi, M. A. Kalil, F. Liers, and A. Mitschele-Thiel, "Increasing spectrum capacity for ad hoc networks using cognitive radios: an analytical model," *IEEE Communications Letters*, vol. 13, no. 9, pp. 676–678, Oct. 2009.

[5] E. Wong and C. Foh, "Analysis of cognitive radio spectrum access with finite user population," *IEEE Communications Letters*, vol. 13, no. 5, pp. 294–296, 2009.

[6] J. Peha, "Sharing spectrum through spectrum policy reform and cognitive radio," *Proceedings of the IEEE*, vol. 97, no. 4, pp. 708–719, 2009.

[7] A. T. Hoang, Y.-C. Liang, and M. H. Islam, "Power control and channel allocation in cognitive radio networks with primary users' cooperation," *IEEE Transactions on Mobile Computing*, vol. 9, no. 3, pp. 348–360, Mar. 2010.

[8] J. Martinez-Bauset, V. Pla, M. Domenech-Benlloch, and D. Pacheco-Paramo, "Admission control and interference management in dynamic spectrum access networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, Article ID 708029, 2010.

[9] A. Rabbachin, T. Quek, H. Shin, and M. Win, "Cognitive network interference," *IEEE J. on Selected Areas in Comms.*, vol. 29, no. 2, pp. 480–493, 2011.

[10] A. Conti, D. Dardari, G. Pasolini, and O. Andrisano, "Bluetooth and IEEE 802.11b coexistence: analytical performance evaluation in fading channels," *IEEE J. on Selected Areas in Comms.*, vol. 21, no. 2, pp. 259–269, 2003.

[11] M. Neuts, *Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach*. The Johns Hopkins University Press, 1981.

[12] D. Gaver, P. Jacobs, and G. Latouche, "Finite birth-and-death models in randomly changing environments," *Advances in Applied Probability*, vol. 16, pp. 715–731, 1984.