Saturation Throughput in a Heterogeneous Multi-channel Cognitive Radio Network

Vicent Pla*, Stijn De Vuyst†, Koen De Turck†, Elena Bernal-Mor*, Jorge Martinez-Bauset* and Sabine Wittevrongel†

*ITACA research institute, Universitat Politècnica de València (UPV), València, Spain
†SMACS Research Group, Dept. of Telecommunications and Information Processing, Ghent University, Gent, Belgium

Abstract—In this paper we consider a cognitive radio network in a heterogeneous multi-channel scenario where channels are different in terms of achievable bit rates. As a result, secondary users (SUs) will see a set of channels which are different not only in their physical capacities but also in the activity profile of PUs. Consequently, from the point of view of the effective throughput that SUs can obtain in each channel there exists a trade-off between the rate they can obtain in a certain channel and the amount of time they can use it. We develop a model that allows to solve this trade-off by computing the maximum effective throughput that SUs can achieve in each channel.

I. INTRODUCTION

Cognitive Radio (CR) networks are envisaged as the key technology to realize dynamic spectrum access. Such paradigm shift in wireless communications aims at solving the scarcity of radio spectrum [1]. The problem of spectrum scarcity is the result of, or is exacerbated by, the long-running static spectrum allocation policies, which are based on assigning spectrum bands to license holders on a long-term basis for large geographical regions. While there is an increasing demand of spectrum, those spectrum management policies have lead to an important underutilization (both temporally and spatially) of a big part of the assigned bands: conducted spectrum occupancy measurement studies yield average utilization figures as low as 5.2% [2], and below 20% in big cities such as New York or Chicago [3]. The CR concept proposes to boost spectrum utilization by allowing CR users (secondary users, SU) to access the licensed wireless channel in an opportunistic manner so that interference to licensed users (primary users, PU) is kept to a minimum.

The idea of CR is undoubtedly compelling and its realization will induce a huge advance in wireless communications. However, there are many challenges and open questions that have to be addressed before CR networks become practically realizable [4].

Spectrum management is carried out by cognitive users through a series of tasks that form a cognitive cycle. These tasks can also be divided into four major spectrum management functions [4], [5]: spectrum sensing, spectrum de-

Part of this research has been funded by the Euro-NF Specific Joint Research Project NETCO (NETworking over COgnitive radio networks). The Spanish authors have been supported by the Spanish government through projects TIN2008-06739-C04-02 and TIN2010-21378-C02-02.
scenario of interest in this paper since they address either the single channel case or the case of multiple channels but where PUs access them in random order. In [13], [14] large-scale measurement campaigns and subsequent thorough statistical analyses are reported. However, none of these works has specifically addressed the statistical characterization of white spaces duration.

The problem of obtaining the saturation throughput of SUs as addressed in this paper can be seen as that of computing the average service completion time in a queueing system with service interruptions. While the queueing literature on systems with service interruptions is abundant and rather general and complex systems have been analyzed (see for instance the related work sections in [15], [16]), to the best of our knowledge none of the existing models covers the specific case that we tackle here.

The rest of the paper is structured as follows. In Section II we detail the system characteristics and a matrix-analytic model is developed and analyzed. Section III presents a series of numerical experiments which illustrate the capabilities of our analysis and provide useful insights into the phenomena arising in the studied scenario. Finally, Section IV concludes the paper.

II. MODEL DESCRIPTION AND ANALYSIS

We consider a set of $N$ frequency bands (channels) and assume channels are numbered according to their radio characteristics, e.g. SNIR, in decreasing quality, i.e., channel 1 is the best and channel $N$ is the worst. These characteristics are homogeneous for all users in the system, both PUs and SUs, and are considered to be static for the time scale of interest.

PUs access the channels following an order of preference, i.e., an idle channel $n$ will be occupied upon arrival of a PU only if all the better channels, $1, \ldots, n-1$, are busy.

SUs can transmit at rate $R_i$ on the $i$-th channel, $R_1 > R_2 > \cdots > R_N$ and transmit data units, e.g. packets, in a preemptive non-resume manner (as in [6]).

We assume that PUs arrive according to a Poisson process of rate $\lambda$. When a PU arrives it is assigned the idle channel with the lowest index. If all channels are occupied the PU is blocked. The channel holding time is assumed to be exponentially distributed with rate $\mu$. By assuming that the holding time of PUs is the same in all channels we are implicitly considering that PUs generate streaming traffic, i.e., while a higher feasible rate will have an impact on the perceived quality the service duration is kept the same. Due to the lack of space we have only addressed this case here, but the model and analysis can be easily modified to consider PUs generating elastic traffic.

We consider a fixed size of SU’s packets $L = D + H$ in bits, where $D$ is the payload size and $H$ the header size.

Let $x = (x_1, \ldots , x_N)$ represent the system where $x_i = 0$ if the $i$-th channel is idle and $x_i = 1$ if it is busy.

Let $I_0(x) = \{i|x_i = 0\}$ and $I_1(x) = \{i|x_i = 1\}$, i.e., $I_0(x)$ (respectively, $I_1(x)$) is the set of indices corresponding to idle (busy) channels.

The transition rate $q_{x,y}$ from a state $x$ to a state $y$ is given as

$$q_{x,y} = \begin{cases} \mu & \text{if } y = x - e_i \text{ and } i \in I_1(x), \\ \lambda & \text{if } y = x + e_i \text{ and } i = \min I_0(x), \\ 0 & \text{otherwise} \end{cases}$$

where $e_i$ denotes a vector with a 1 on the $i$-th position and 0’s elsewhere.

Let $Q = [q_{x,y}]$ denote the infinitesimal generator of the system obtained by considering states sorted in lexicographical order, and $\pi$ the vector of stationary probabilities.

Consider now an arbitrary channel $i = 1, \ldots, N$. Let us define

$$I(i) = \{x|x_i = 0\} \quad \text{and} \quad B(i) = \{x|x_i = 1\} = \{1, \ldots, N\} \setminus I(i),$$

i.e., $I(i)$ (respectively, $B(i)$) is the set of states where channel $i$ is idle (busy).

Define also

$$Q_{00}^{(i)} = [q_{x,y}]_{x,y \in I(i)} \quad \text{and} \quad Q_{10}^{(i)} = [q_{x,y}]_{x \in B(i), y \in I(i)}.$$

The matrix $Q_{00}^{(i)}$ contains the transition rates among states in $I(i)$ and it is obtained by removing from $Q$ the rows and columns that correspond to states in $B(i)$. The matrix $Q_{10}^{(i)}$ contains the transition rates from states in $B(i)$ to states in $I(i)$, and it is obtained by removing from $Q$ the rows that correspond to states in $I(i)$ and the columns that correspond to states in $B(i)$. Likewise, let $\pi_0^{(i)}$ ($\pi_1^{(i)}$) be a row vector with the stationary probabilities for states in $I(i)$ ($B(i)$), which are obtained by taking the appropriate entries from $\pi$.

We are interested in the duration of the busy period $B_p(i)$ and of the idle period $I_p(i)$ (for an arbitrary channel $i$). The busy period clearly corresponds to the PU holding time of channel $i$, which follows an exponential distribution of rate $\mu$ for all $i$.

The idle period corresponds to the sojourn time in the set of states $I(i)$. Thus, the duration of the idle period can be represented by the phase-type distribution $P_{H}(\alpha(i), T(i))$ [17], where $T(i) = Q_{00}^{(i)}$ and the probability vector $\alpha(i)$, which contains the probabilities of initiating the sojourn at each of the states in $I(i)$, is given by

$$\alpha(i) = \frac{1}{\pi_1^{(i)} Q_{10}^{(i)}} \pi_1^{(i)} Q_{10}^{(i)},$$

where $\mathbf{1}$ is a column vector of 1’s.

The saturation throughput for SUs can be calculated as follows. The probability that at least $k$ packets can be transmitted during an idle period on channel $i$ is given as

$$p_k^{(i)} = P\left[I_p^{(i)} \geq k \cdot \frac{L}{R_i}\right] = 1 - F_{I_p^{(i)}}\left(k \frac{L}{R_i}\right) = \alpha(i) e^{-\frac{kL}{R_i} T(i)} \mathbf{1},$$

where $F_{I_p^{(i)}}(\cdot)$ is the cumulative distribution function of the idle period duration on channel $i$. 
Let $N^{(i)}_s$ be the number of SU packets that are successfully transmitted during an idle period on channel $i$, then

$$E[N_s^{(i)}] = \sum_{k=1}^{\infty} \rho_k^{(i)} = \sum_{k=1}^{\infty} \alpha(i) e^{-\frac{\rho_k^{(i)}}{\mu}} = \alpha(i) e^{-\frac{\rho_1^{(i)}}{\mu}} T(i) \mathbf{1},$$

(1)

where $I$ is the identity matrix.

The saturation throughput $\gamma_s^{(i)}$ for SU on channel $i$ can be now computed as the average number of (payload) bits that can be transmitted during an idle period over the average duration of the sequence of an idle plus a busy period. This leads to

$$\gamma_s^{(i)} = \frac{D \cdot E[N_s^{(i)}]}{E[I_p^{(i)}] + E[I_p^{(i)}]} = \frac{\alpha(i) \left(e^{-\frac{\rho_1^{(i)}}{\mu}} T(i) - I\right)^{-1} \mathbf{1}}{-\alpha(i) T(i)^{-1} \mathbf{1} + 1/\mu} D. \quad (2)$$

Besides the throughput, a useful metric could also be the probability that the transmission of a SU packet is interrupted. Let $\nu_i$ be the interruption probability for SU transmissions on channel $i$, then

$$\nu_i = \frac{1}{1 + E[N_s^{(i)}]},$$

since during an idle time there are $N_s^{(i)}$ successful transmissions and exactly one interrupted transmission. For the same reason, the number of transmissions per time unit on channel $i$ is

$$\frac{E[N_s^{(i)}] + 1}{E[I_p^{(i)}] + E[I_p^{(i)}]}.$$\nThe overall interruption probability $\nu$ for SU transmissions on any of the $N$ channels results from weighting the probabilities $\nu_i$ by the fractions of transmissions per time unit on each of the channels, i.e.

$$\nu = \frac{1}{\sum_{i=1}^{N} \frac{E[N_s^{(i)}] + 1}{E[I_p^{(i)}] + E[I_p^{(i)}]} \sum_{j=1}^{N} E[B_p^{(j)}] + E[I_s^{(j)}]}.$$\n
$$= \frac{1}{\gamma_s/D + \sum_{j=1}^{N} \frac{1}{E[I_p^{(j)}] + E[I_p^{(j)}]} \sum_{j=1}^{N} E[B_p^{(j)}] + E[I_s^{(j)}]}.$$\n
Introducing the harmonic mean $\beta$ of the mean cycle lengths on the channels, i.e. of the mean times between subsequent PU allocations on the channels, we find

$$\nu = \frac{N}{\beta \gamma_s/D + N}, \quad \text{with} \quad \frac{N}{\beta} = \sum_{i=1}^{N} \frac{1}{E[B_p^{(i)}] + E[I_s^{(i)}]}.$$\n
Finally, from a QoS perspective, it is interesting to know how long it takes for a SU packet to be effectively transmitted. Let us assume that if a packet’s transmission on a certain channel is interrupted by a PU, it will be retransmitted on the same channel. We define the transmission delay $d_i$ of a packet on channel $i$ as the time period between the start of this packet’s first transmission and the end of its final (successful) transmission. The average delay on channel $i$ is given by $E[d_i] = D/\gamma_s^{(i)}$, while the delay of an arbitrary SU packet has mean value $E[d_i] = ND/\gamma_s$.

III. NUMERICAL EXAMPLES AND DISCUSSION

To demonstrate the feasibility of our analysis, we now consider some specific scenarios for a system of $N = 10$ channels. Since the ordering of the channels $i = 1, \ldots, 10$ results from a ranking from best to worst, it is reasonable to assume a Zipf law for their bandwidths $R_i$, i.e.

$$R_i = \frac{R_1}{\rho^i}, \quad \text{and} \quad R = \sum_{i=1}^{N} R_i,$$ \quad (3)

for some shape parameter $\theta > 0$ and with $R_1$ the bandwidth of the best channel. We choose $R_1 = 1$ Mbps and $\theta=0.5034$, such that the total available bandwidth of all channels together is $R=5$ Mbps. The average PU holding time on all channels is equal to $1/\mu = 10$ ms. The offered load of PU traffic to the system is defined as

$$\rho_p = \frac{\lambda}{N/\mu}.$$\n
In Fig. 1 we look at the influence of the PU load $\rho_p$ on the saturation throughput. The packet length is chosen to be $L=1000$ bit, half of which is header information $H=500$ bit. The upper dashed curve shows the saturation throughput of the channels if the transmission of SU packets over the channels is not impaired by PU traffic, i.e. if $\lambda=0$. Since the payload consists of half a packet, this curve corresponds to $R_1/2$, as given in (3). The figure clearly demonstrates that for a small PU load $\rho_p$, only the lower channels are affected, as one expects. For a higher PU load, the upper channels carry a significant part of the PU traffic as well, which results in a decreasing SU throughput in those channels.

In Fig. 2 the SU throughput of each channel is shown for some values of the SU packet length $L$, in case $\rho_p=0.2$. One observes that with increasing packet length, the bulk of the throughput shifts more and more towards the upper channels. Since the lower channels carry most of the PU traffic, the ‘gaps’ between the PU transmissions are small there and larger.
SU packets will no longer fit in. For $L = 32$ kb there are almost no gaps on channel 1 that are large enough to contain a complete SU packet, which results in a throughput $\gamma_s^{(1)}$ that is almost zero.

The crucial influence of the SU packet length is further demonstrated in Fig. 3. The total throughput $\gamma_s = \sum_{i=1}^{N} \gamma_s^{(i)}$ on all channels together is shown as a function of the packet length $L$, still assuming a header of 500 bit for each packet. Regardless of the amount of carried PU traffic, the throughput of SU data first increases with $L$, reaches an optimal point and then decreases again to zero. This can be explained as follows. If small SU packets are used, they will consist mostly of header information and have little payload. Therefore, although a lot of packets may be transmitted over the channels, the useful throughput is small. On the other hand, if the packet length $L$ is large, the packets will no longer fit in the gaps between the PU transmissions, resulting in frequent interruption and subsequent retransmission of SU packets. This is especially the case on the lower channels with best quality (as was demonstrated in Fig. 2) because more PU traffic will be allocated there, resulting in smaller gaps. Consequently, we can conclude that there will always be an optimal value of the packet length $L^*$ that maximises the overall throughput $\gamma_s$ and which depends on the PU load of the system.

In Fig. 4, we show $\log_{10} \nu$ as a function of the PU packet length $L$ for different values of the PU load. As expected, the interruption probability $\nu$ increases with both. Note that transmission interruptions do not depend on the header size of the packets, so no specific value for $H$ has been considered here.

In Fig. 5 we show a doubly logarithmic plot of the transmission delay, in case the header size is 500 bit and the packet length is either 1 kb or 5 kb. Obviously, under light PU traffic conditions $\rho_p \ll 1$, the delay is equal to the transmission time $L/R_i$ because SU transmissions are almost never interrupted. For increasing load $\rho_p$, more and more retransmissions are required, resulting in an increasing transmission delay on all channels. In case of Fig. 5 we see that the curves for the different channels overtake each other in the region $\rho_p < 0.5$. As we already discussed, this is due to the PU traffic gradually occupying the system, starting with the lowest channels. A striking observation however, is that the order of the curves switches again for higher values of $\rho_p$. Although not entirely visible on the figure, in extreme PU overload situations $\rho_p \gg 1$, the mean SU transmission delays on the channels are ordered as $E[d_1] < \ldots < E[d_N]$, even though channel $i$ carries more PU traffic than channel $i+1$. Hence we can conclude that for extreme overload, just as for extreme underload, the intrinsic rates $R_i$ determine which channel has the best throughput. In between, which channel has the best throughput is also (and possibly predominantly) influenced by the amount of PU occupation of the channels.

IV. CONCLUSIONS

In this contribution we assess the maximum data throughput for unlicensed users (SUs) in the cognitive radio network.
paradigm. We use a model with \( N \) channels having heterogeneous transmission rates \( R_i \), where incoming requests of licensed users (PUs) are allocated to the highest-rate channel that is unoccupied. SU packets, which are assumed to be of fixed length \( L \), can only use the ‘white spaces’ on a channel between the PU allocations. Any ongoing SU transmission will be forcefully interrupted if a new PU is allocated to the channel, which results in a preemptive repeat discipline for unlicensed packets. The model is tackled analytically, using a Markov chain description of the channels’ occupation. Expressions are obtained for the maximum useful SU throughput as well as related performance criteria such as interruption probability and mean transmission delay.

The results are demonstrated by means of some specific scenarios. We quantitatively show e.g. that careful choice of the packet length \( L \) is extremely important for maximizing the throughput. Additionally, concerning the question which channel achieves the best SU throughput we observed the following. As expected, for very low PU activity this is determined by the intrinsic rates \( R_i \) of the channels. If PU activity becomes significant, the SUs may achieve the best throughput on channels with lower rate because the highest-rate channels are the ones preferably used by the PUs. Surprisingly however, for extremely high PU activity on the system, the best SU throughput is again achieved on the channel with the highest intrinsic rate.

REFERENCES