Contents lists available at ScienceDirect

Physical Communication

journal homepage: www.elsevier.com/locate/phycom

Full length article

Capacity considerations for uncoordinated communication in geographical spectrum holes☆,☆☆

Erik Axell *, Erik G. Larsson, Danyo Danev

Division of Communication Systems, Department of Electrical Engineering (ISY), Linköping University, S-581 83 Linköping, Sweden

ARTICLE INFO

Keywords:
Cognitive radio
Spectrum hole
Capacity
Achievable rate

ABSTRACT

Cognitive radio is a new concept of reusing a licensed spectrum in an unlicensed manner. The motivation for cognitive radio is various measurements of spectrum utilization, that generally show unused resources in frequency, time and space. These "spectrum holes" could be exploited by cognitive radios. Some studies suggest that the spectrum is extremely underutilized, and that these spectrum holes could provide ten times the capacity of all existing wireless devices together. The spectrum could be reused either during time periods where the primary system is not active, or in geographical positions where the primary system is not operating. In this paper, we deal primarily with the concept of geographical reuse, in a frequency-planned primary network. We perform an analysis of the potential for communication in a geographical spectrum hole, and in particular the achievable sum-rate for a secondary network, to some order of magnitude.

Simulation results show that a substantial sum-rate could be achieved if the secondary users communicate over small distances. For a small number of secondary links, the sum-rate increases linearly with the number of links. However, the spectrum hole gets saturated quite fast, due to interference caused by the secondary users. A spectrum hole may look large, but it disappears as soon as someone starts using it.

1. Introduction

A spectrum is a scarce resource, and operators have made huge financial investments to buy licensed spectrums. The licensed spectrum is intended for specific communication technologies, and no one but the spectrum owner is allowed to use it. Cognitive radio is a new concept of reusing a licensed spectrum in an unlicensed manner [1–3]. The motivation for cognitive radio is various measurements of spectrum utilization, that generally show unused resources in frequency, time and space [4,5]. These "spectrum holes" could be exploited by cognitive radios. The spectrum could be reused either during time periods where the primary system is not active, or in geographical positions where the primary system is not operating. This paper deals primarily with geographical, or spatial, reuse.

The introduction of cognitive radios, sometimes called secondary users, in an existing primary system will create interference and thus a quality degradation of the primary system. In order to reduce the impact on the primary system, cognitive radios have to sense the spectrum and detect whether there are primary users in the vicinity that are currently using the spectrum. The cognitive radios need to be positioned sufficiently far away from the primary users and transmit at very low power levels. This has been analyzed in, e.g. [6] for a single cell, and in [7] for a frequency-planned network. It is unavoidable

☆ Invited Paper.

☆☆ The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 216076. This work was also supported in part by the Swedish Research Council (VR), the Swedish Foundation for Strategic Research (SSF), and the CENIIT foundation. E. Larsson is a Royal Swedish Academy of Sciences (KVA) Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation.

* Corresponding author. Tel.: +46 0 13 281350.
E-mail addresses: axell@isy.liu.se (E. Axell), erik.larsson@isy.liu.se (E.G. Larsson), danyo@isy.liu.se (D. Danev).

1874-4907/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.phycom.2009.03.002

that there has to be some sort of compromise for the primary system to allow secondary users. In [6,7] this compromise is a reduction of the primary cell radius. The cell radius is decreased by a small amount and it is required that the same signal-to-interference-plus-noise ratio (SINR) is experienced by the primary users as without any secondary users. Hence, the aggregated transmit power of the secondary users must be constrained to keep the interference level low. For example, if the cell radius is decreased by 5%, then the transmitter power for the cognitive radios is constrained such that the primary users experience the same SINR as they did before, without any secondary users.

A similar one-cell model for spatial frequency reuse has also been analyzed in [8], together with more general definitions of a spectrum hole and some metrics to quantify the performance of a spectrum sensing algorithm. The focus of this work is more on the uncertainty of detection versus the area that could actually be exploited.

The concept of frequency reuse can be seen as a bucket filled with rocks [8]. Although the bucket is filled there is still plenty of room for sand. However this metaphor requires a large-scale primary system, and a small-scale secondary one. This problem has been stated in [9], and especially the problem of coexistence of a secondary system with primary systems of different scales. The main issue is the coexistence of a secondary system in the presence of a small-scale primary network. The Part 74 wireless microphone users that are a concern to the IEEE 802.22 WRAN is given as an example. Relating to the bucket metaphor, if the bucket is filled with sand we cannot fit any more rocks or sand.

In much of the literature, the main difficulty has been perceived to be the detection of the primary users. Even if that could be solved, we need to know that the spectrum holes can really be exploited and provide some useful data rate. Some studies suggest that the spectrum is extremely underutilized, and that these spectrum holes could provide ten times the capacity of all existing wireless devices together [5]. The aim of this paper is to analyze the potential for communication in a geographical spectrum hole, and in particular the achievable sum-rate for a secondary network, to some order of magnitude. In Section 2 we describe the system model, and Section 3 shows some numerical results. Section 4 proposes some improvements on the individual secondary links by using multiple devices. Section 5 concludes the work.

2. Model

We consider the downlink in a hexagonal frequency-planned network, shown in Fig. 1. We include the main primary base station and the first tier of co-channel interferers. The cell radius is r, and the distance to the first tier of interfering base stations is $D = \sqrt{3n}r$ [10], where $1/n$ is the frequency reuse factor of the primary system (n is the number of frequency groups). The positions of the base stations are denoted by the vectors $\bar{B}_0 = \bar{0}$ and $\bar{B}_m = (D\cos((m-1)\frac{\pi}{3}), D\sin((m-1)\frac{\pi}{3}))$, $m = 1, 2, \ldots, 6$.

Following [6,7] we assume that the cognitive users are permitted to operate only if they are located at a distance at least d from the nearest primary base station. Furthermore, we assume that N cognitive transmitters are spread out uniformly at random in the allowed region, i.e. in the area between the circles of radii $D-d$ and d respectively, as shown in Fig. 1. The positions of the cognitive transmitters are denoted by the vectors $\bar{T}_i = (x_{T,i}, y_{T,i})$, $i = 1, 2, \ldots, N$. For each cognitive transmitter there is an associated cognitive receiver at a distance d_0 from the transmitter and at an angle θ_i. The angle θ_i is uniformly distributed in the interval $[-\pi, \pi]$, but such that the receiver is also in the permitted area. The positions of the receivers are denoted by the vectors $\bar{R}_i = (x_{T,i} + d_0\cos(\theta_i), y_{T,i} + d_0\sin(\theta_i))$, $i = 1, 2, \ldots, N$.

We consider a log-distance path loss model. Thus, we define the channel gain function at distance x,

$$\rho(x) = \left(\frac{x}{x_0}\right)^{-\alpha} 10^{\chi/10}, \quad \chi \sim N(0, \gamma)$$

where α is the path loss exponent, x_0 is a normalization constant and γ is the standard deviation of the lognormal fading in dB. The distance, x, is in general a random variable since we consider random locations for the secondary users. Hence, the received interference is random. However, the distance between a secondary transmitter/receiver pair d_0 is fixed, and the only randomness of the received signal strength is the lognormal fading. The base stations transmit omnidirectionally with power P, and each cognitive user transmits omnidirectionally with power P_c. The transmit power is defined as the power received at the normalization distance x_0.

We assume that the secondary users are uncoordinated and thus transmit simultaneously, using the same channel. Hence, the secondary users will interfere with each other. The interference power experienced by the ith cognitive receiver, and caused by other secondary users can be written as $P_c \sum_{k\neq i} \rho(|\bar{R}_i - \bar{T}_k|)$. The secondary users also receive interfering signals from the primary base stations. The interference power from the primary system, for the ith secondary user, can be written as $P \sum_{n=0}^{6} \rho(|\bar{R}_i - \bar{B}_n|)$. Now the signal-to-interference-plus-noise ratio (SINR) for the ith cognitive receiver becomes

$$\begin{aligned}\mathrm{SINR}_i &= \frac{P_c\rho(d_0)}{\sigma^2 + P_c\sum_{k\neq i}\rho(|\bar{R}_i - \bar{T}_k|) + P\sum_{n=0}^{6}\rho(|\bar{R}_i - \bar{B}_n|)} \\ &= \frac{\rho(d_0)}{\frac{\sigma^2}{P_c} + \sum_{k\neq i}\rho(|\bar{R}_i - \bar{T}_k|) + \frac{P}{P_c}\sum_{n=0}^{6}\rho(|\bar{R}_i - \bar{B}_n|)},\end{aligned}$$

where σ^2 is the receiver noise floor.

The primary system may be either interference limited or noise limited, or somewhere in between. Following [7], we can quantify the operating point of the primary system in terms of the expected interference-to-noise ratio at the cell border without secondary users:

$$\psi \triangleq E\left[\frac{P\sum_{n=1}^{6}\rho(|(r\cos(\phi), r\sin(\phi)) - \bar{B}_n|)}{\sigma^2}\right] = \frac{P\overline{\mu}}{\sigma^2},$$

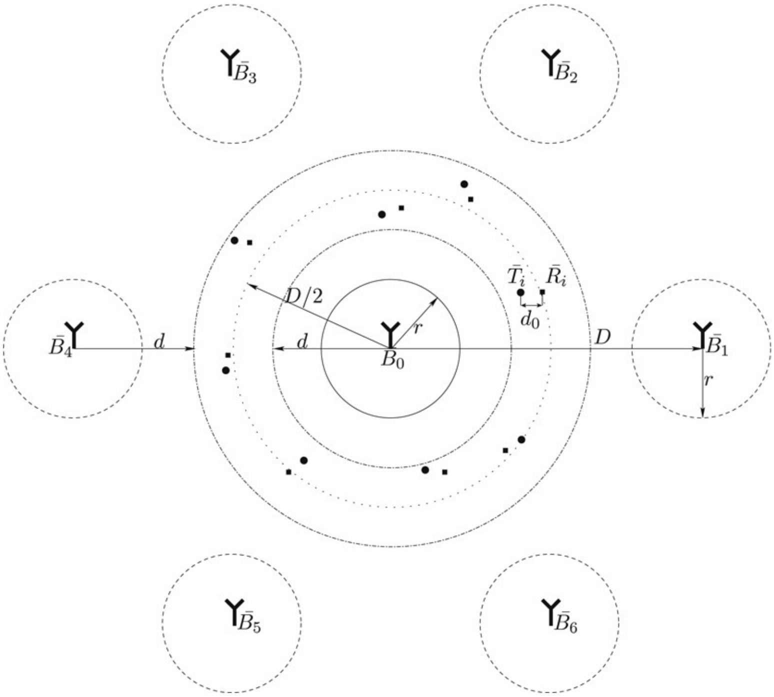

Fig. 1. Model for the cognitive network.

where

$$\overline{\mu} \triangleq E\left[\sum_{n=1}^{6} \rho(|(r\cos(\phi), r\sin(\phi)) - \bar{B}_n|)\right],$$

and the expectation is taken over the lognormal fading and ϕ which is uniformly distributed over $[-\pi, \pi]$. In order to make sure that the cognitive users do not cause too much harmful interference to the primary users, the transmit power P_c must be constrained. We will constrain the aggregate cognitive radio transmit power, such that $NP_c = \epsilon P$, for some $\epsilon > 0$. The choice of ϵ will be discussed later. Thus, we can rewrite the SINR and obtain the following:

$$\mathrm{SINR}_i = \frac{\rho(d_0)}{\frac{N\overline{\mu}}{\epsilon\psi} + \sum_{k\neq i} \rho(|\bar{R}_i - \bar{T}_k|) + \frac{N}{\epsilon}\sum_{n=0}^{6} \rho(|\bar{R}_i - \bar{B}_n|)}.$$

The achievable rate for the ith secondary link is modeled as

$$\begin{aligned} C_i &= \log_2(1 + \mathrm{SINR}_i) \\ &= \log_2\left(1 + \frac{\rho(d_0)}{\frac{N\overline{\mu}}{\epsilon\psi} + \sum_{k\neq i} \rho(|\bar{R}_i - \bar{T}_k|) + \frac{N}{\epsilon}\sum_{n=0}^{6} \rho(|\bar{R}_i - \bar{B}_n|)}\right) \\ &\quad (\text{bits/s/Hz}). \end{aligned}$$

Hence, the sum-rate offered by the network of all secondary users, C, is

$$\begin{aligned} C &= \sum_{i=1}^{N} C_i \\ &= \sum_{i=1}^{N} \log_2\left(1 + \frac{\rho(d_0)}{\frac{N\overline{\mu}}{\epsilon\psi} + \sum_{k\neq i} \rho(|\bar{R}_i - \bar{T}_k|) + \frac{N}{\epsilon}\sum_{n=0}^{6} \rho(|\bar{R}_i - \bar{B}_n|)}\right). \end{aligned}$$

We define the relative area of cognitive operation

$$A = \frac{(D-d)^2 - d^2}{D^2} = 1 - 2\frac{d}{D}.$$

This is the fraction of the total system area in which cognitive operation is permitted. Note that the allowed transmit power P_c depends on the permitted area A and on the primary interference-to-noise operating point ψ. The relationship between these parameters was investigated in [7]. See also [6] for the special case of only a single primary base station. The primary cell radius was decreased, and the secondary transmit power was constrained such that the SINR for the primary users was not decreased compared to the primary system without any secondary users. The allowed aggregated secondary transmit power $NP_c = \epsilon P$ was computed given a relative area A and a primary interference-to-noise operating point ψ. We will use the values of ϵ, ψ and A obtained from this analysis for our simulations.

3. Simulation results

In this section we will show some numerical results from Monte-Carlo simulations. For each number of secondary users N, we generated 5000 realizations of the system model. The achievable sum-rate was then calculated as the mean of the sum-rate over all realizations. For each realization we placed N cognitive transmitters uniformly

Table 1
Parameter values used in the simulations for $n = 7,\ 21$ and $\psi = -10$ dB, obtained from Fig. 3 in [7].

A (%)	1	25	50
ϵ ($n = 21$) (dB)	0	−3	−12
ϵ ($n = 7$) (dB)	0	−5	−20

at random in the allowed area between the circles of radii d and $D - d$ respectively. To obtain a uniform distribution over the circular area we created the transmitter positions $\bar{T}_i = (x_{T,i}, y_{T,i})$, $i = 1, 2, \ldots, N$, in polar coordinates. The angle was uniformly distributed over $[-\pi, \pi]$ and the radius, R, was obtained by

$$R = \sqrt{((D-d)^2 - d^2)X + d^2},$$

where X was uniformly distributed over $[0, 1]$. The receiver positions were then created as

$$\bar{R}_i = (x_{T,i} + d_0 \cos(\theta_i), y_{T,i} + d_0 \sin(\theta_i)) \quad i = 1, 2, \ldots, N,$$

where all θ_i were uniformly distributed over $[-\pi, \pi]$. To make sure that all users were inside the allowed region we simply redrew the angle θ_i whenever a receiver position happened to be outside of the allowed region.

We used $\psi = -10$ dB throughout all simulations. This corresponds to a noise limited primary system. We argue that a practical system where we eventually could make use of this kind of geographical spectrum reuse would typically be noise limited. It could, for example, be a television network with a very sparse frequency reuse and primary transmitters located far away from each other. In addition, simulations have shown that the value of ψ only has a small impact on the results. The values of ϵ and A that were used in the simulations are shown in Table 1, and obtained from Fig. 3 in [7] for $n = 7, 21$ frequency reuses and $\psi = -10$ dB. These values were obtained assuming that the primary cell radius is decreased by 5%, and the primary users experience the same SINR at the 90%-percentile of the distribution, as without any secondary users. Also, in accordance with [7], we use the path loss and shadow fading parameters $\alpha = 4$ and $\gamma = 6$ dB throughout the whole paper. We use frequency reuse $n = 21$, except where otherwise stated.

3.1. Achievable sum-rate

Fig. 2 shows the total system throughput C for $A = 50\%, 25\%, 1\%$ and $\psi = -10$ dB. When increasing N, we observe an increase to some congestion limit. Above this congestion limit, adding more users only causes a throughput degradation due to increased interference. Thus, for a given d, the system throughput is maximized for a certain number of users N. Intuitively we would expect the congestion level to be lower when the allowed region is smaller. The operation region will be saturated for a smaller number of users since the area is smaller. This intuition is confirmed by Fig. 2: the number of users maximizing the throughput is higher when the allowed region is larger. Note also that the total throughput is higher for a 25% cognitive area than for 50% or 1%. Hence, the throughput is neither increasing nor decreasing in d. Rather there is some d that maximizes the total throughput. The interpretation of this is that the allowed transmit power P_c is high and the expected interference from the primary system is low for a small cognitive region, but the interference from the cognitive users increases fast as the number of users N increases. On the other hand, when the allowed region is large, the allowed transmit power P_c is low and the expected interference from the primary system is higher. The optimum seems to be somewhere in between these two extremes. These results are for frequency reuse $n = 21$, but the same behavior is seen also for other frequency reuse factors.

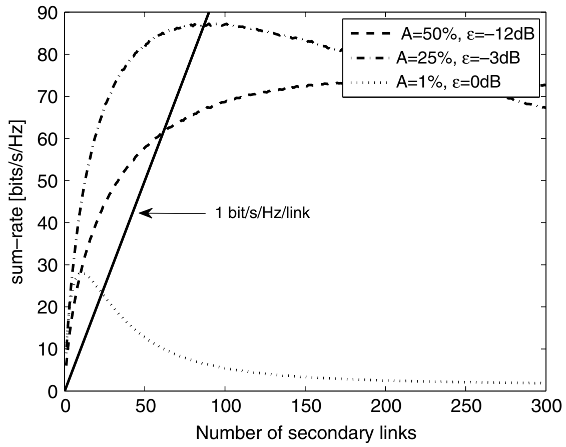

Fig. 2. Total achievable sum-rate for the secondary system for $d_0 = 0.1r$, $\alpha = 4$, $\gamma = 6$ dB, $n = 21$.

As a reference, the solid line shown in Fig. 2 is a straight line with sloping one, and corresponds to the case where each secondary user gets 1 bit/s/Hz. We consider rates above this line as acceptable whereas rates below the line are less acceptable. Arguably links with less spectral efficiency than 1 bit/s/Hz are not very useful. Note that for a large permitted area ($A = 50\%$) many secondary users can coexist and achieve a quite high sum-rate, but the rate per user is not acceptable.

Fig. 3 shows the total achievable rate for different transmitter-receiver distances d_0, both for $n = 7$ and $n = 21$ frequency reuse. Due to the larger operating area for $n = 21$ than $n = 7$ frequency reuse, the secondary users are allowed to use a higher transmit power without causing too much interference for the primary users. Also, since the operating area is larger, the distance between interfering cognitive users is larger on average. As expected, the performance is better for $n = 21$ than for $n = 7$ frequency reuse, both in terms of the total achievable rate and in terms of the number of users that can be allowed before the congestion limit is hit. We also observe a large improvement from decreasing the communication distance. If we compare the maximum total achievable rates, we note that a distance decrease by a factor 10 yields a throughput increase by a factor 10 for $n = 7$ and a factor 20 for $n = 21$. Thus the order of magnitude of the maximum sum-rate seems to stand in inverse proportion to the communication distance between the secondary users. It is also worth noting that the maximum throughput is attained for a larger number

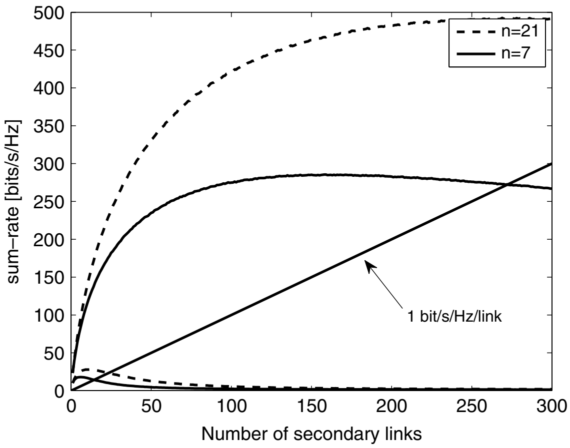

Fig. 3. Comparison of the achievable sum-rates for $d_0 = 0.1r$ and $d_0 = 0.01r$, for $A = 1\%$, $\alpha = 4$, $\gamma = 6$ dB.

of users when the distance is smaller. Hence, both the total throughput and the number of users can be larger for a smaller communication distance.

3.2. Geographic density

We have seen that the total achievable rate attains a maximum for a certain number of secondary users. If there are more users they will get too close to each other and the interference they generate to each other will increase. The question is then how the users should be distributed to achieve a maximum total system throughput. Is the user distribution dense or sparse at the maximum sum-rate operating point?

A reasonable assumption is that the geographic area filled up by each secondary link communicating over a distance d_0 is equal to a circle of radius d_0. Then N secondary links use an area $N\pi d_0^2$. The total area of secondary operation is the area between the circles of radii $D - d$ and d respectively, i.e. $\pi((D-d)^2 - d^2) = \pi AD^2$. We denote by $N_{\max}$ the number of users for which the maximum total achievable rate is attained. We define the effective area, A_E, as the ratio of the area used by $N_{\max}$ secondary users and the total area of secondary operation, i.e.

$$A_E = \frac{N_{\max}\pi d_0^2}{\pi AD^2} = \frac{N_{\max} d_0^2}{3nr^2 A}.$$

Note that this ratio might actually be greater than one, since the secondary links could overlap. Analyzing the simulation results shown in Fig. 2, we obtain the value of $N_{\max}$ in various cases. These values and the effective areas are then shown in Table 2. For all sizes of the allowed secondary operation region we see that the effective area is between 5% and 20% (similar numbers have also been observed for $n = 7$). The conclusion is that the users should be quite sparsely distributed to obtain the maximum system throughput. It is also worth noting that for large operating regions, the rate per link for $N = N_{\max}$ users is "non-acceptable" (in the sense defined in Section 3.1) although the sum-rate is maximized. In this case the users have to be even more sparsely distributed in order to get a satisfactory rate per link.

Table 2
Effective area for $n = 21$, $d_0 = 0.1$ r.

A (%)	1	25	50
$N_{\max}$	12	86	220
A_E (%)	19	5	7

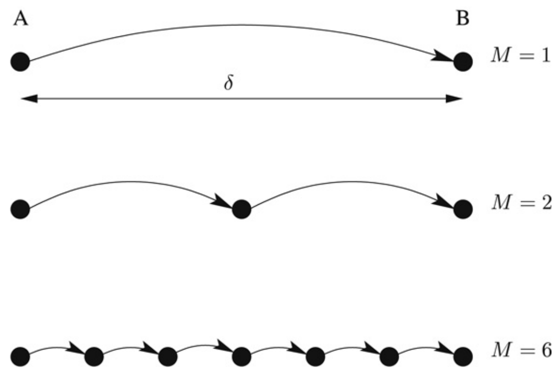

Fig. 4. Point-to-point communication from A to B, either directly or by multihop with time-division multiplexing.

4. Point-to-point improvement

The simulations in Section 3 show that the achievable sum-rate is strongly dependent on the transmission distance d_0. A small distance yields a larger received signal strength for the secondary users and thus a larger achievable rate. This also leads us to another interesting question. Suppose that we want to communicate between a point A and another point B separated by a distance δ, and using power P_{tot}. We neglect shadow fading, i.e. the channel gain at a distance x is simply $\rho(x) = (\frac{x}{x_0})^{-\alpha}$. The achievable rate would then be

$$C = \log_2\left(1 + \frac{P_{\text{tot}}(\delta/x_0)^{-\alpha}}{J}\right), \tag{1}$$

where J is the received noise plus interference power from the primary system.

Assume further that we can alternatively use in total $M + 1$ devices (M links) spread out uniformly on the straight line between A and B, and transmit in a multihop fashion. Then the distance between two neighboring devices is δ/M. We assume that the devices share the channel by time-division multiplexing as shown in Fig. 4, i.e. we let the sub-nodes transmit one at a time, starting at A, to the next sub-node until the message reaches B. Since only one device transmits at a time, each device is allowed to use power P_{tot}. The SINR on each link is in this case

$$\text{SINR} = \frac{P_{\text{tot}}\left(\frac{\delta}{Mx_0}\right)^{-\alpha}}{J}.$$

The received noise plus interference power, J, is assumed to be equal for all devices, since the inter-node distances are