Wireless Communications and Mobile Computing

Efficient Spectrum Usage for Wireless Communications

Lead Guest Editor: Ivan Marsa-Maestre
Guest Editors: Takayuki Ito, Sofie Pollin, Alessandro Chiumento, and Jose M. Gimenez-Guzman

Efficient Spectrum Usage for Wireless Communications

Wireless Communications and Mobile Computing

Efficient Spectrum Usage for Wireless Communications

Lead Guest Editor: Ivan Marsa-Maestre
Guest Editors: Takayuki Ito, Sofie Pollin, Alessandro Chiumento, and Jose M. Gimenez-Guzman

Editorial Board

Contents

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 8719849, 2 pages
https://doi.org/10.1155/2019/8719849

Editorial

Efficient Spectrum Usage for Wireless Communications

Ivan Marsa-Maestre,[1] **Takayuki Ito**,[2] **Sofie Pollin**,[3] **Alessandro Chiumento**,[4] **and Jose Manuel Gimenez-Guzman**[5]

[1]*Edificio Politécnico, Dcho E-246, Campus Universitario, Alcala de Henares, Madrid 28805, Spain*
[2]*Graduate School of Engineering, Nagoya Institute of Technology, Room 701, Building No. 4, Gokiso, Showa, Nagoya 466-8555, Japan*
[3]*KU Leuven, Kasteelpark Arenberg 10, Box 2444, Room 02.21, 3001 Leuven, Belgium*
[4]*KU Leuven, Kasteelpark Arenberg 10, Box 2444, 3001 Leuven, Belgium*
[5]*Edificio Politécnico, Dcho E-247, Campus Universitario, Alcala de Henares, Madrid 28805, Spain*

Correspondence should be addressed to Ivan Marsa-Maestre; ivan.marsa@uah.es

Received 9 December 2018; Accepted 10 December 2018; Published 3 January 2019

Wireless technologies have reached an impressive popularity in the last years. However, the radio spectrum is very limited, and therefore as wireless communications have become more and more widespread, problems related to spectrum scarcity have arisen. Radio spectrum, as the physical support for wireless communication, both for fixed applications and especially for mobile broadband, is becoming an extremely strategic, valued, and demanded resource. Therefore, technologies and techniques enabling a more flexible access for service providers and clients and a more efficient and effective usage are needed.

Spectrum scarcity can be addressed from many different perspectives. For instance, from the point of view of signal processing, we can look for higher spectral efficiency in modulations. In the paper "Simple Algorithms for Estimating the Symbol Timing Offset in DCT-Based Multicarrier Systems," F. Cruz-Roldán *et al.* enhance multicarrier modulation based on discrete cosine transform (DCT-MCM), proposing two new blind algorithms to perform tight timing offset and coarse frequency synchronization, which addresses the problem of symbol timing offset in these modulations.

Scheduling the usage of radio resources is also an effective strategy, especially in cellular networks. In the paper "Dynamic Tradeoff between Energy and Throughput in Wireless 5G Networks," C. Gueguen and M. Manini consider radio resource allocation for the mobiles within a single access point coverage zone. Their Dynamic Tradeoff scheduler is able to prioritize energy efficiency or spectral efficiency and delay depending on the network traffic load.

Cognitive radio takes a more active approach, making the devices responsible of efficient spectrum utilization by sensing spectrum usage and adjusting transmission parameters to accommodate communications in unused resources. In the paper "Discrete-Time Analysis of Cognitive Radio Networks with Nonsaturated Source of Secondary Users," V. Pla *et al.* address one of the fundamental problems in cognitive radio: sensing for the detection of white spaces when they occur. Authors use Markovian models to analyze and evaluate sensing strategies.

On a higher level, coordination may happen in a centralized or distributed manner, by establishing protocols allowing base stations and clients to increase spectrum utilization while avoiding interferences. Distributed scenarios like multihop wireless networks may be specially challenging to coordinate. In the paper "Optimal Multicommodity Spectrum-Efficient Routing in Multihop Wireless Networks," M. Saad addresses the optimization of end-to-end spectral efficiency in Multihop Wireless Networks where there are multiple source-destination pairs active at the same moment. Authors provide two alternative approaches, using fixed-size and variable-size time slots.

Optimization techniques, artificial intelligence approaches, or economic paradigms may contribute greatly to a more efficient spectrum usage in wireless communications. Regarding optimization techniques, in the paper "On the Goodness of Using Orthogonal Channels in WLAN IEEE 802.11 in Realistic Scenarios," J. M. Gimenez-Guzman *et al.* study the behavior of optimization techniques and heuristics

in Wi-Fi channel assignments, evaluating the gain (or rather, the lack of gain) of having more available channels instead of the usual "orthogonal" set. For a different domain, in the paper "Optimization of Cell Size in Ultra-Dense Networks with Multiattribute User Types and Different Frequency Bands," Y. Wei and S.-H. Hwang propose a multiple-objective optimization model for ultra-dense cellular networks (UDN), showing the potential of new higher frequency bands in these scenarios.

An economic paradigm is explored in the paper "Strategic Interaction between Operators in the Context of Spectrum Sharing for 5G Networks." E. J. Sacoto-Cabrera *et al.* analyze spectrum sharing in 5G networks from an economic perspective, comparing the monopoly situation, the pooling agreement, and the priority sharing agreement. For this last scenario, authors show the conditions in which network sharing is incentive compatible.

Finally, regulatory frameworks like the recent Licensed Shared Access (LSA) may play a crucial role in helping mobile networks operators to make a more efficient usage of the scarce bandwidth resources. In the paper "A Practical Perspective on 5G-Ready Highly Dynamic Spectrum Management with LSA," P. Masek *et al.* explore an experimental extension of LSA-based spectrum management in LTE which is able to operate in a highly dynamic manner.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

The guest editorial team would like to thank the authors and the reviewers for their contributions in bringing this special issue to fruition.

Ivan Marsa-Maestre
Takayuki Ito
Sofie Pollin
Alessandro Chiumento
Jose Manuel Gimenez-Guzman

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 7367028, 12 pages
https://doi.org/10.1155/2019/7367028

Research Article

Discrete-Time Analysis of Cognitive Radio Networks with Nonsaturated Source of Secondary Users

Vicent Pla,[1] **Attahiru S. Alfa**,[2,3]
Jorge Martinez-Bauset,[1] **and Vicente Casares-Giner**[1]

[1] *Universitat Politècnica de València, Spain*
[2] *University of Manitoba, Winnipeg, MB, Canada*
[3] *University of Pretoria, Pretoria, South Africa*

Correspondence should be addressed to Vicent Pla; vpla@upv.es

Received 22 June 2018; Accepted 22 November 2018; Published 2 January 2019

Guest Editor: Takayuki Ito

Sensing is a fundamental aspect in cognitive radio networks and one of the most complex issues. In the design of sensing strategies, a number of tradeoffs arise between throughput, interference to primary users, and energy consumption. This paper provides several Markovian models that enable the analysis and evaluation of sensing strategies under a broad range of conditions. The occupation of a channel by primary users is modeled as alternating idle and busy intervals, which are represented by a Markov phase renewal process. The behavior of secondary users is represented mainly through the duration of transmissions, sensing periods, and idle intervals between consecutive sensing periods. These durations are modeled by *phase-type* distributions, which endow the model with a high degree of generality. Unlike our previous work, here the source of secondary users is nonsaturated, which is a more practical assumption. The arrival of secondary users is modeled by the versatile *Markovian arrival process*, and models for both finite and infinite queues are built. Furthermore, the proposed models also incorporate a quite general representation of the resumption policy of an SU transmission after being interrupted by PUs activity. A comprehensive analysis of the proposed models is carried out to derive several key performance indicators in cognitive radio networks. Finally, some numerical results are presented to show that, despite the generality and versatility of the proposed models, their numerical evaluation is perfectly feasible.

1. Introduction

The evolution and widespread deployment of wireless communications have generated an incessant and increasing demand for radio spectrum. This, combined with the static spectrum allocation policies that have been in place for quite some time, has led to a situation of spectrum scarcity (i.e., of unassigned frequency bands), at the same time, an important underutilization of a substantial part of the assigned bands.

Cognitive Radio (CR) is viewed as the enabling technology for dynamic spectrum access, which would allow solving the seeming paradox between spectrum scarcity and underutilization [1]. The basic idea of CR is to allow the unlicensed users, known as secondary users (SUs), to access licensed channels opportunistically when they are not in use by licensed users, known as primary users (PUs). This way, the interference that SUs produce to PUs should be kept to a minimum.

In this context, *white space* refers to spectrum that is not used by the PUs during a certain time interval at a specific location. The key to success for CR consists of effectively and efficiently sensing radio channels to detect *white space* when it occurs. An ideal, although unrealizable, sensing strategy would detect a portion of white space right after it starts and, likewise, would detect the end of it immediately after the transmission of a PU begins. Furthermore, such an ideal sensing strategy would only consume the minimum amount of energy needed for sensing. There has been a considerable long list of research papers over the years dealing with how to manage and implement CR. For details see [2] and other references therein.

As noted in [3], sensing is a major and challenging issue in CR networks. The choice of detection parameters poses a series of tradeoffs between achievable throughput, energy consumption, and interference caused PUs [4–13]. In general,

if SUs spend more time on channel sensing, they obtain lower throughput, but the interference caused to PUs is also lowered. This is referred to as the *sensing-throughput tradeoff* (STT), which has been studied in a large number of papers (e.g., [4, 5, 9] and references therein). Furthermore, more channel sensing also raises the consumption of energy, which constitutes a critical aspect in certain scenarios (e.g., sensor networks). Consequently, a significant number of studies have considered energy efficiency a crucial part of spectrum sensing [7, 8, 10–13].

Several aspects come into play while considering sensing. Some examples are the duration of sensing periods, the width of the frequency band being scanned to search for unused channels, etc. For details see [2]. The vast majority of the models developed for studying CR networks, and specifically for spectrum sensing, assume that busy and idle times follow an exponential distribution (or geometric in discrete time); for example, see [14, 15] and all the above references to sensing studies. Using measurements, the authors of [16] showed that the channel idle time can be modeled by a lognormal distribution. This result was confirmed in [17], where it is shown that idle times follow a lognormal distribution for long durations, and a geometric distribution for short durations. Correlations between idle and busy times were also overlooked by most the previous papers. However, one could logically expect that some correlation exists between different intervals, as has in fact been noted in [17]. Our previous work in [2] was one of the first few ones to introduce correlation and also allow more general intervals.

The research on sensing strategies for CR networks is not new, and neither is the application of mathematical modeling for analysis and optimization of those strategies. However, the models presented in this paper embody a number of contributions which stem from the generality of the model assumptions. Our purpose in this paper is to propose a number of models that enable the analysis and evaluation of sensing strategies in cognitive radio networks under a broad range of conditions. More specifically, a Markov phase renewal process [18] is used to model the channel availability for SUs. This allows to consider a wide variety of distributions for the duration of idle and busy intervals and also to capture correlations between consecutive intervals. The authors of [19] proposed a similar model for the activity of PUs. However, correlations between different intervals were not considered in their model.

The behavior of SUs is represented mainly through the duration of transmissions, sensing periods, and idle intervals between consecutive sensing periods. These durations are modeled by *phase-type* distributions [18], which endow the model with a high degree of generality. In the literature of mathematical modeling, it is widely acknowledged that phase-type distributions offer an excellent compromise between applicability and tractability.

Our model of SUs also allows sensing errors, which can be misdetections and false alarms. For both of them, two different situations are distinguished, depending on whether the SU is only sensing or sensing and transmitting. This allows us to capture a broad range of SUs sensing capabilities.

However, most important in this paper is that we now have to introduce the arrival process of the SUs since the source is no longer saturated. We used the Markovian arrival process (MAP) to represent the SU arrival process. The MAP is a very versatile arrival process which can capture correlations and still allow modeling and computational tractability.

In our previous paper [2] on this class of problems, we assume that the source of SUs is saturated (i.e., there is always at least an SU waiting and ready to transmit). That assumption is not too realistic, even though it does give us an idea of the best we can achieve if there are always some SUs looking to transmit. In this current paper, we have relaxed that assumption. This makes the model more realistic while making it slightly more challenging as now we need to introduce an arrival process for the SUs. In addition, this new model creates a situation where a channel could be idle simply because there are no PUs or SUs needing to use it. These are the main contributions of this paper.

The rest of this paper is organized as follows. Section 2 introduces the model of the channel from the perspective of the SUs. The model of the secondary network, which includes the behavior of SUs, is described in Section 3. In Section 4 we describe the models for the complete system and their analysis when sensing is assumed to be ideal; this assumption is relaxed in Section 5. Some numerical results are presented in Section 6 to exemplify the capabilities of the proposed models and to show the feasibility of their numerical evaluation. Finally, the paper is concluded in Section 7.

2. Channel Availability

We consider a single channel that can be idle or busy from the perspective of the SUs. The activity of PUs in this channel is described by a discrete-time Markov chain (DTMC) X_k with state space $\{1, 2, \ldots, n_b, n_b+1, \ldots, n_b+n_i\}$. The channel is *busy* (b) if $X_k \in \{1, 2, \ldots, n_b\}$, and *idle* (i) if $X_k \in \{1, 2, \ldots, n_b\}$. We assume that during a time slot the condition of the channel changes at most once.

Let the matrix D_b represent the transitions between busy states and d_{bi} represent the transitions from busy to idle states. Similarly, D_i represents the transitions between idle states, and d_{ib} represents the transitions from idle to busy states. The matrices D_b and D_i are substochastic and of orders n_b and n_i, respectively.

Based on this, the transition matrix of X_k can be written as

$$D = \begin{bmatrix} D_b & d_{bi} \\ d_{ib} & D_i \end{bmatrix}. \tag{1}$$

3. Secondary Network

This section discusses SUs actions and how SUs interact with PUs through the channel status. Throughout this paper, sometimes we use "the SU" to refer to the set of all the SUs that can use the channel or to the SU that is at the head of the waiting line of SUs. We assume there is some coordination mechanism among SUs to perform channel sensing and to

Table 1: State classification in the saturated system.

state #	Channel state	SU state	phases defining internal states	number of internal states
1	busy	sleeping	BPH, LPH, TPH*	$n_b n_\ell (n_t + 1)$
2	busy	sensing	BPH, TPH*	$n_b 1 (n_t + 1)$
3	idle	sleeping	IPH, LPH, TPH*	$n_i n_\ell (n_t + 1)$
4	idle	sensing	IPH, SPH, TPH*	$n_i n_s (n_t + 1)$
5	idle	transmitting	IPH, TPH	$n_i n_t$

BPH: channel busy phase; IPH: channel idle phase; LPH: SU sleeping phase; SPH: SU sensing phase; TPH: SU trans. phase; TPH*: TPH (extended) at interruption.

arrange channel access. Although this coordination mechanism is not trivial, its study is beyond the scope of this paper.

The SU can be in one of the following three modes: *sleeping*, *sensing*, or *transmitting*. The continuous period of time that the SU remains in one of these modes is called a *cycle* or *period*. Thus, we talk about, for example, a *sleeping period* or a *sensing cycle*. Next, we detail the characteristics of each type of cycle and how they alternate between them.

(a) Sleeping. The duration of a sleeping period is modeled by the phase-type distribution $(\boldsymbol{\delta}, L)$ of order n_ℓ. A sleeping period is always followed by a sensing cycle.

(b) Sensing. During a sensing cycle, the SU performs a series of consecutive channel state measurements. If a measurement senses the channel as busy, the sensing period is interrupted and the SU enters the sleeping mode. The maximum number of measurements that would be taken is defined by the PH distribution with representation $(\boldsymbol{\beta}, S)$, of order n_s. If the channel is sensed as idle in all the measurements of the cycle, the SU can initiate a transmitting period.

(c) Transmitting. During a transmitting cycle the SU attempts to transmit a message. The required transmission time (i.e., the number of time slots) to transmit the message is given by the PH distribution with representation $(\boldsymbol{\alpha}, T)$ of order n_t. If the channel becomes busy and the SU is capable (can sense the channel while transmitting) of detecting it, the message transmission is interrupted and the SU switches to sleeping mode. If the transmission of the message is fully completed, the SU goes into sensing mode.

The SUs arrive according to a Markovian arrival process (MAP) represented by two substochastic matrices G_0 and G_1 of order n_a. Then, the mean arrival rate, λ, is given as $\lambda = \boldsymbol{\pi}_G G_1 \mathbf{1}$, where π_G is the probability vector satisfying $\boldsymbol{\pi}_G = \boldsymbol{\pi}_G (G_0 + G_1)$ and $\boldsymbol{\pi}_G \mathbf{1} = 1$ and 1 is a column vector of ones of appropriate dimensions.

An SU that arrives and finds the channel busy waits in a buffer of size $N \leq \infty$.

If an SU is in service when a PU arrives, the SU's service is interrupted since the PU has preemptive priority. Now we specify the resumption policy followed by SUs when a message transmission was interrupted by PUs activity. This policy is described by matrix Q with elements Q_{ij}. Suppose the SU's service was interrupted in phase i, let its service restart in phase j with probability Q_{ij} at resumption. It is clear that if it is a preemptive resume then $Q = I$, whereas if it is a preemptive repeat then $Q = \mathbf{1}\boldsymbol{\alpha}$. Hence, the matrix Q is a general representation.

4. System Model I: Ideal Sensing

For the sake of clarity, we first assume that the SU receives perfect knowledge of the state of the channel. In Section 5 we relax that condition and assume that there could be errors in the sensing carried out by the SU.

Sensing takes place only when there is an SU in the system waiting to get access. Hence, when there is an SU in the system either it is receiving service because there is white space or it is waiting. The waiting could be because the channel is busy with PUs, there are other SUs ahead of it, or it is sensing the system for white space.

Here we assume that the time needed to perform a channel measurement cannot be neglected compared to the transmission time of a data unit. The length of a time slot is set so that a sensing measurement can be performed and know the result by the end of the slot. A conservative approach is applied to establish the outcome of the measurement: a sensing measurement taken during the time slot $[k-1, k)$ does not return *idle* as a result if the channel was busy at time instants $k-1$ or k.

Although, as mentioned above, the duration of a single measurement cannot be neglected, in this section we assume that SUs can sense and transmit simultaneously. This could reflect instances in which SUs are equipped with two radios. The case in which SUs cannot transmit and sense simultaneously can be covered by the model with imperfect sensing described in Section 5.

4.1. Head-of-the-Line (HoL) SU. Even though we are mainly interested in the nonsaturated case, first we consider the saturated case, from which we obtain the DTMC that constitutes the basis for the nonsaturated model.

Let the states of this system be classified as shown in Table 1.

In all states, we keep track of the channel phase (either a busy or an idle phase). When the SU is transmitting, we need to keep track of the transmission phase. Similarly, when the SU is sleeping (sensing) we need to keep track of the sleeping (sensing) phase, and also of how the last transmission epoch ended. A transmission epoch can finish because the channel becomes busy, or because the SU transmission finishes. In the

first case, we record at which of the n_t transmission phases the SU was interrupted, and an additional phase is used to indicate that the SU transmission finished completely. Thus, in total $n_t + 1$ phases are used to represent how the last transmission epoch ended. Note also that when the channel is busy and the SU starts a sensing cycle, only the first sensing phase is required as the SU switches to the sleeping mode after the first sensing measure. Naturally, this situation will change when we consider imperfect sensing later on.

Then we have the following discrete-time Markov chain (DTMC) for the saturated case:

$$P_1^{(\mathrm{sat})} = \begin{bmatrix} D_b \otimes L \otimes I & D_b \otimes \boldsymbol{l} \otimes I & d_{bi} \otimes L \otimes I & d_{bi} \otimes (\boldsymbol{l}\boldsymbol{\beta}) \otimes I & 0 \\ D_b \otimes \boldsymbol{\delta} \otimes I & 0 & d_{bi} \otimes \boldsymbol{\delta} \otimes I & 0 & 0 \\ d_{ib} \otimes L \otimes I & d_{ib} \otimes \boldsymbol{l} \otimes I & D_i \otimes L \otimes I & D_i \otimes (\boldsymbol{l}\boldsymbol{\beta}) \otimes I & 0 \\ d_{ib} \otimes (\mathbf{1}\boldsymbol{\delta}) \otimes I & 0 & 0 & D_i \otimes S \otimes I & D_i \otimes \boldsymbol{s} \otimes Q^* \\ d_{ib} \otimes \boldsymbol{\delta} \otimes \overline{T} & 0 & 0 & D_i \otimes \boldsymbol{\beta} \otimes \overline{\boldsymbol{t}} & D_i \otimes T \end{bmatrix}, \tag{2}$$

where $\overline{T} = [I \ \ \mathbf{0}]$, $\overline{\boldsymbol{t}} = [0 \ \ \boldsymbol{t}]$, and $Q^* = \left[\begin{smallmatrix} Q \\ \boldsymbol{\alpha} \end{smallmatrix}\right]$.

This transition matrix captures the full behavior of this saturated system with general preemptive discipline.

However, in order to study the nonsaturated system, we need to extract, from the above matrix, the following four matrices:

$$H_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & D_i \otimes \boldsymbol{\beta} \otimes \overline{\boldsymbol{t}} & 0 \end{bmatrix}, \tag{3}$$

$$H_0 = P_1^{(\mathrm{sat})} - H_1, \tag{4}$$

$$F_0 = \begin{bmatrix} 0 & D_b \otimes [0_{1\times n_t} \ \ 1] & 0 & d_{bi} \otimes \boldsymbol{\beta} \otimes [0_{1\times n_t} \ \ 1] & 0 \\ 0 & d_{ib} \otimes [0_{1\times n_t} \ \ 1] & 0 & D_i \otimes \boldsymbol{\beta} \otimes [0_{1\times n_t} \ \ 1] & 0 \end{bmatrix}, \tag{5}$$

and

$$F_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & D_i \otimes \boldsymbol{t} \end{bmatrix}. \tag{6}$$

The matrices H_0 and H_1 are used to capture the transitions that affect the state of the channel and the state of the head-of-line SU: H_0 corresponds to the case where the head-of-line SU does not finish its service and H_1 to the one when it does. In H_0 and H_1, it is assumed that there is at least one SU in the queue (at both the initial and the final state). In a similar manner, the transitions from and to an empty queue are captured by F_0 and F_1, respectively. Note that the arrival process has not been taken into account yet.

4.2. Number of SUs in the System: The Nonsaturated Case. In order to study the distribution of the number of SUs in the system, we use a quasi-birth-and-death (QBD) Markov chain structure. The *level* of the QBD represents the number of SUs in the system, and the phases in each level represent the rest of the system state information. Specifically, for level $\ell = 0$ (i.e., when the queue is empty) the phases of the level represent the phase of the arrival process (i.e., the MAP) and the phase of the channel. For the rest of levels ($\ell = 1, 2, \ldots$), the phases represent the phase of the MAP plus the same state information as in the saturated case.

We now consider the cases of finite buffer ($N < \infty$) and infinite buffer ($N = \infty$) separately.

4.2.1. Stationary Behavior for Finite Buffer ($N < \infty$). When the buffer size is finite, we have the discrete-time Markov chain (DTMC) with transition matrix

$$P = \begin{bmatrix} B & C & & & & \\ E & A_1 & A_0 & & & \\ & A_2 & A_1 & A_0 & & \\ & & A_2 & A_1 & A_0 & \\ & & & \ddots & \ddots & \ddots \\ & & & & A_2 & A_1 + A_0 \end{bmatrix}, \tag{7}$$

where

$$B = G_0 \otimes D, \tag{8}$$

$$C = G_1 \otimes F_0, \tag{9}$$

$$E = G_0 \otimes F_1, \tag{10}$$

and

$$A_0 = G_1 \otimes H_0, \tag{11}$$

$$A_1 = G_0 \otimes H_0 + G_1 \otimes H_1, \tag{12}$$

$$A_2 = G_0 \otimes H_1. \tag{13}$$

Note that the number of block rows and block columns of the transition matrix P coincides with the number of levels of the QBD, that is, $N + 1$.

Let the stationary vector associated with this DTMC be $\boldsymbol{x} = [\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]$. Then we have

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{x}P, \\ \boldsymbol{x}\mathbf{1} &= 1. \end{aligned} \tag{14}$$

This vector $\boldsymbol{x}$ exists and is unique, provided the DTMC represented by P is irreducible.

The vector $\boldsymbol{x}$ can be obtained by standard methods used for finite structured DTMCs (e.g., the *block state reduction*, the *block state reduction,* or the *folding algorithm*). Given the vector $\boldsymbol{x}$, we can easily obtain the following performance measures.

(a) Mean Number of SUs in the System. Let X be the number of SUs in the system at steady state, and let

$$p_k = \mathrm{P}(X = k) = \boldsymbol{x}_i\mathbf{1}, \quad k = 0, \ldots, N. \tag{15}$$

Then,

$$\mathrm{E}[X] = \sum_{i=1}^{N} i p_i. \tag{16}$$

(b) Loss Probability. The probability that an SU is lost because there is no buffer space to accommodate it is referred to as the loss probability, and it is given as

$$P_L = \frac{\boldsymbol{x}_N A_0 \mathbf{1}}{\lambda}. \tag{17}$$

(c) Throughput (or Effective Utilization). Let us define the throughput, γ, as the probability that an SU is transmitting during a time slot of white space or, equivalently, the mean number of SU transmitting slots per time unit.

Let us first introduce

$$\boldsymbol{\pi} = \begin{bmatrix} \boldsymbol{\pi}_1 & \boldsymbol{\pi}_2 & \boldsymbol{\pi}_3 & \boldsymbol{\pi}_4 & \boldsymbol{\pi}_5 \end{bmatrix} = \left(\sum_{n=1}^{N} \boldsymbol{x}_n \right) \cdot (\mathbf{1} \otimes I) \tag{18}$$

which contains the marginal probabilities for states of the saturated system; first the marginal distribution of the phases for levels above 0 is computed, and then the phase of the arrival process is dropped by summing along its dimension (right multiplication by $\mathbf{1} \otimes I$).

Now an expression for the throughput can be written as

$$\begin{aligned} \gamma = \boldsymbol{\pi}_5\mathbf{1} &= \sum_{n=1}^{N} \boldsymbol{x} \cdot (\mathbf{1} \otimes I) \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{bmatrix} \\ &= \sum_{n=1}^{N} \boldsymbol{x} \cdot \left(\mathbf{1} \otimes \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{bmatrix} \right). \end{aligned} \tag{19}$$

(d) Goodput. Goodput γ_g is the mean number of useful data units transmitted by an SU per unit of time. Even though we still are assuming perfect sensing, not all transmitted data units can be considered useful. For example, the data unit transmitted during a slot in which the presence of the PU was detected does not represent a useful transmission. Moreover, depending on the service resumption policy, a part or all of the transmitted units up to this point may not be useful transmissions either.

Nonetheless, if the system is stable (and it is clearly the case when $N < \infty$) all SUs accepted to the system are eventually served in full. Therefore, the goodput can be easily obtained as

$$\gamma_g = \lambda(1 - P_\ell)\,\mathrm{E}[L_t], \tag{20}$$

where L_t is the maximum length of an SU transmission phase, and its mean value is given by $\mathrm{E}[L_t] = \boldsymbol{\alpha}(I - T)^{-1}\mathbf{1}$.

(e) Distribution of the Sojourn Time of SUs That Joined the Queue. Here we develop the expressions to obtain the distribution of the time an SU spends in the system (W), which includes the time spent at the head of the line. A similar approach could be followed to derive the distribution for the waiting time until the SU reaches the head of the line.

Let $\boldsymbol{y} = [\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{N-1}]$ be the probability vector as seen by arriving customers that are not blocked. Then

$$\boldsymbol{y}_0 = \left(\lambda(1 - P_\ell)\right)^{-1} \left(\boldsymbol{x}_0 (G_1 \otimes F_0) + \boldsymbol{x}_1 (G_1 \otimes H_1)\right), \tag{21}$$

$$\begin{aligned} &\boldsymbol{y}_n \\ &\quad = \left(\lambda(1 - P_\ell)\right)^{-1} \left(\boldsymbol{x}_n (G_1 \otimes H_0) + \boldsymbol{x}_{n+1} (G_1 \otimes H_1)\right), \\ &\qquad\qquad n = 1, \ldots, N-1. \end{aligned} \tag{22}$$

We now consider the block absorbing DTMC with transition matrix

$$\widetilde{P} = \begin{bmatrix} I & & & & \\ I \otimes F_1 & I \otimes H_0 & & & \\ & I \otimes H_1 & I \otimes H_0 & & \\ & & \ddots & \ddots & \\ & & & I \otimes H_1 & I \otimes H_0 \end{bmatrix}, \tag{23}$$

with the initial probabilities given by $[\mathbf{0}, \boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{N-1}]$.

This DTMC can be further simplified by leaving out the DTMC the phase of the arrival process. By doing this, the transition matrix becomes

$$\widehat{P} = \begin{bmatrix} I & & & \\ F_1 & H_0 & & \\ & H_1 & H_0 & \\ & & \ddots & \ddots \\ & & & H_1 & H_0 \end{bmatrix}, \tag{24}$$

and the initial probabilities are given by

$$\boldsymbol{z}_0(0) = \mathbf{0}, \tag{25}$$

$$\boldsymbol{z}_n(0) = \boldsymbol{y}_{n-1}(\mathbf{1} \otimes I), \quad n = 1, \ldots, N, \tag{26}$$

or equivalently,

$$\boldsymbol{z}_0(0) = \mathbf{0}, \tag{27}$$

$$\boldsymbol{z}_1(0) = \left(\lambda(1-P_\ell)\right)^{-1} \cdot \left(\boldsymbol{x}_0\left((G_1\mathbf{1})\otimes F_0\right) + \boldsymbol{x}_1\left((G_1\mathbf{1})\otimes H_1\right)\right), \tag{28}$$

$$\boldsymbol{z}_n(0) = \left(\lambda(1-P_\ell)\right)^{-1} \cdot \left(\boldsymbol{x}_{n-1}\left((G_1\mathbf{1})\otimes H_0\right) + \boldsymbol{x}_n\left((G_1\mathbf{1})\otimes H_1\right)\right), \quad n = 2,\ldots,N, \tag{29}$$

Now, the probabilities at time k can be obtained recursively as

$$\boldsymbol{z}_0(k) = \boldsymbol{z}_0(k-1)I + \boldsymbol{z}_1(k-1)F_1 = \boldsymbol{z}_1(k-1)F_1, \tag{30}$$

$$\boldsymbol{z}_n(k) = \boldsymbol{z}_n(k-1)H_0 + \boldsymbol{z}_{n+1}(k-1)H_1, \quad n = 1,\ldots,N-1, \tag{31}$$

$$\boldsymbol{z}_N(k) = \boldsymbol{z}_N(k-1)H_0 \tag{32}$$

and, from here, the distribution of the sojourn time follows easily

$$P(W \le k) = \boldsymbol{z}_0(k)\mathbf{1} \quad k = 0,1,\ldots. \tag{33}$$

4.2.2. Stationary Behavior for Infinite Buffer ($N = \infty$). Suppose now the buffer space for the SUs is unlimited. Then we have the associated transition matrix for the DTMC given as

$$P = \begin{bmatrix} B & C & & & \\ E & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & A_2 & A_1 & A_0 \\ & & & \ddots & \ddots & \ddots \end{bmatrix}. \tag{34}$$

The associated stationary vector $\boldsymbol{x} = [\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots]$ can be obtained by using the matrix-geometric method. First, we point out that this stationary vector $\boldsymbol{x}$ exists and is unique, provided that the standard stability conditions are met. These conditions are given as follows. Let $A = A_0 + A_1 + A_2$ with $\boldsymbol{\pi} = \boldsymbol{\pi}A$, $\boldsymbol{\pi}\mathbf{1} = 1$. We know that provided A is irreducible, then $\boldsymbol{\pi}$ exists and it is unique. The conditions for the vector $\boldsymbol{x}$ to exist and be unique are that $\boldsymbol{\pi}A_2\mathbf{1} > \boldsymbol{\pi}A_0\mathbf{1}$. Provided these conditions are met, we have

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i R, \quad i \ge 1, \tag{35}$$

where R is the minimal nonnegative solution to the matrix quadratic equation

$$R = A_0 + RA_1 + R^2A_2. \tag{36}$$

The boundary vectors $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ are obtained from solving

$$[\boldsymbol{x}_0, \boldsymbol{x}_1] = [\boldsymbol{x}_0, \boldsymbol{x}_1]\begin{bmatrix} B & C \\ E & A_1 + RA_2 \end{bmatrix}, \tag{37}$$

normalized by

$$\boldsymbol{x}_0\mathbf{1} + \boldsymbol{x}_1(I-R)^{-1}\mathbf{1} = 1. \tag{38}$$

From the knowledge of the vector $\boldsymbol{x}$ we can easily obtain the performance measures of interest, which are the same as in the finite buffer case in Section 4.2.1. Here, clearly, $P_\ell = 0$. The expressions for the other performance measures are given below; we focus only on the changes with respect to Section 4.2.1.

(a) Mean Number of SUs in the System. Let X be the number of SUs in the system and $p_k = \boldsymbol{x}_i\mathbf{1}$ $k \ge 0$ its distribution, then

$$p_0 = \boldsymbol{x}_0\mathbf{1}, \quad p_k = \boldsymbol{x}_k\mathbf{1} = \boldsymbol{x}_1R^{k-1}\mathbf{1}, \quad k \ge 1, \tag{39}$$

and

$$\mathrm{E}[X] = \sum_{i=1}^{\infty} ip_i = \boldsymbol{x}_1\left(\sum_{i=1}^{\infty} iR^{i-1}\right)\mathbf{1} = \boldsymbol{x}_1(I-R)^{-2}\mathbf{1}. \tag{40}$$

(b) Throughput. Now

$$\begin{bmatrix}\boldsymbol{\pi}_1 & \boldsymbol{\pi}_2 & \boldsymbol{\pi}_3 & \boldsymbol{\pi}_4 & \boldsymbol{\pi}_5\end{bmatrix} = \left(\sum_{n=1}^{\infty}\boldsymbol{x}_n\right)\cdot(\mathbf{1}\otimes I) = \boldsymbol{x}_1(I-R)^{-1}(\mathbf{1}\otimes I), \tag{41}$$

and thus,

$$\gamma = \sum_{n=1}^{\infty}\boldsymbol{\pi}_5\mathbf{1} = \boldsymbol{x}_1(I-R)^{-1}\left(\mathbf{1}\otimes\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\\ \mathbf{1}\end{bmatrix}\right). \tag{42}$$

(c) Goodput

$$\gamma_g = \lambda\mathrm{E}[L_t] = \lambda\boldsymbol{\alpha}(I-T)^{-1}\mathbf{1}. \tag{43}$$

(d) Distribution of the Sojourn Time of SUs That Joined the Queue

$$P(W \le k) = \boldsymbol{z}_0(k)\mathbf{1}. \tag{44}$$

As in Section 4.2.1,

$$\boldsymbol{z}_0(k) = \boldsymbol{z}_0(k-1)I + \boldsymbol{z}_1(k-1)F_1, \tag{45}$$

$$\boldsymbol{z}_n(k) = \boldsymbol{z}_n(k-1)H_0 + \boldsymbol{z}_{n+1}(k-1)H_1, \quad n > 0, \tag{46}$$

and

$$\boldsymbol{z}_0(0) = \mathbf{0}, \tag{47}$$

$$\boldsymbol{z}_1(0) = \lambda^{-1}\left(\boldsymbol{x}_0\left((G_1\mathbf{1})\otimes F_0\right) + \boldsymbol{x}_1\left((G_1\mathbf{1})\otimes H_1\right)\right), \tag{48}$$

$$\boldsymbol{z}_n(0) = \boldsymbol{x}_{n-1}J = \boldsymbol{x}_{n-2}RJ = \cdots = \boldsymbol{x}_1R^{n-2}J, \quad n > 1, \tag{49}$$

where $J = \lambda^{-1}((G_1\mathbf{1})\otimes H_0 + R((G_1\mathbf{1})\otimes H_1))$.

4.3. Comparison of Sensing Strategies. In order to compare any two sensing strategies, we need to determine what performance measure is going to be used for that purpose. In the finite buffer case, we can compare, for the two sensing strategies, the number of SUs in the system, the throughput, the loss probabilities, etc. Each of those comparisons requires that we compute the stationary vector associated with the finite state Markov chain.

However, if we assume that the buffer size is large enough to be approximated by an infinite buffer system, then we can just compare the distribution of the number of SUs in the system. For a service provider, the most relevant performance measure for comparison in that case is the tail behavior of the system. As done in practice, we can assume that when one system is dominant in queue length, then it is also dominant in waiting times. Thus, for a quick comparison of different sensing strategies, we can use the tail behavior and skip the steps that require us to obtain the vector $\boldsymbol{x}$ or even the matrix R.

4.3.1. Tail Probabilities for Comparing Sensing Strategies. Now we assume that our buffer size is infinite and show how to compare the tail behavior of the number of SUs in the system.

For a quasi-birth-and-death (QBD) type of Markov chain, we know from the literature that the probability that the number of customers in the system is at least k, σ_k, can be approximated as

$$p_k = \sigma\eta^k + o\left(\eta^k\right), \quad \text{as } k \longrightarrow \infty, \tag{50}$$

where η is called the decay rate and is equal to Perron Frobenius eigenvalue of the matrix R, and σ is a constant which depends on the boundary behavior of the Markov chain. The decay rate η can be obtained as the unique solution in $(0, 1)$ to the nonlinear equation

$$\eta = \chi(\eta), \tag{51}$$

where $\chi(z)$ is the maximal absolute eigenvalue of the matrix

$$A(z) = A_0 + zA_1 + z^2A_2, \quad |z| \le 1. \tag{52}$$

Generally, η can be computed using the bisection method.

4.3.2. Hazard Rate Order for Comparisons. In this section, we present a simple tool for comparing different sensing strategies. Since the primary interest is in not letting very long queues build up for the secondary users, a good measure is the tail probability of the number of SUs in the system. The hazard rate order is a useful tool for this purpose.

Consider two discrete random variables X and Y. Shaked and Shanthikumar showed in [20, Theorem (1.B.7)] that X is less than Y based on hazard rate, that is, $X \le_{hr} Y$, if

$$\frac{\mathrm{P}\{X \ge k\}}{\sum_{i\ge k}\mathrm{P}\{X \ge i\}} \le \frac{\mathrm{P}\{Y \ge k\}}{\sum_{i\ge k}\mathrm{P}\{Y \ge i\}}, \quad \forall k \ge 0. \tag{53}$$

Similarly, we say X is stochastically less than Y (i.e., $X \le_{st} Y$) if

$$\mathrm{P}\{X \ge k\} \le \mathrm{P}\{Y \ge k\}, \quad \forall k \ge 0. \tag{54}$$

Table 2: Sensing errors and their probabilities.

	Misdetection	False alarm
Type 1	ϕ_1	θ_1
Type 2	ϕ_2	θ_2

In [20, Theorem (1.B.1)] they further showed that

$$\begin{aligned} &\text{if } X \le_{hr} Y, \\ &\text{then } X \le_{st} Y. \end{aligned} \tag{55}$$

Now consider two sensing strategies A and B, with the number of SUs in the system given as X_A and X_B, respectively. Letting $p_k(A)$ an $p_k(B)$ be the associated tail probabilities as mentioned in Section 4.3.1, and also with η_j, and σ_j, $j = A, B$ being the respective decay rates and constants, then we can write

$$p_k(j) = \sigma_j\eta_j^k + o\left(\eta_j^k\right), \quad \text{as } k \longrightarrow \infty, \ j = A, B. \tag{56}$$

Let us now assume that there is a $K < \infty$ such that $o(\eta_j^{K+k}) = 0$, $k \ge 0$, then we can say that

$$p_{K+k}(j) = \sigma_j\eta_j^{K+k}, \quad \forall k \ge 0, \ j = A, B. \tag{57}$$

For the purpose of comparing two sensing strategies, we apply the following theorem.

Theorem 1. *If our threshold of the number in the system is K, then*

$$X_A \le_{st} X_B \quad \text{if } \eta_A \ge \eta_B. \tag{58}$$

Proof. This is based on the fact that

$$\frac{\mathrm{P}\{X \ge k\}}{\sum_{i\ge k}\mathrm{P}\{X \ge i\}} = \frac{\sigma\eta^k}{\sum_{i\ge k}\sigma\eta^i} = \frac{\sigma\eta^k}{\sigma\eta^k(1-\eta)^{-1}} = 1 - \eta, \quad k \ge K. \tag{59}$$

In summary, we have, for a threshold of K,

$$X_A \le_{st} X_B \quad \text{if } \eta_A \ge \eta_B. \tag{60}$$

□

5. System Model II: Imperfect Sensing

In this model we allow sensing errors, which can be misdetections and false alarms. A misdetection (false alarm) occurs when the channel is considered to be idle (busy) when it is actually busy (idle). In addition, for both of them (misdetection and false alarm) we differentiate two different types, depending on whether the SU is only sensing (type 1) or sensing and transmitting (type 2). The probability that a type-i misdetection occurs is ϕ_i, with $\overline{\phi}_i = 1 - \phi_i$ and $i = 1, 2$; and the probability that a type-i false alarm occurs is θ_i, with $\overline{\theta}_i = 1 - \theta_i$ and $i = 1, 2$; Table 2 shows a summary of these probabilities.

Table 3: State classification in the saturated system with sensing errors.

state #	channel state	SU	phases defining internal states	number of internal states
1	busy	sleeping	BPH, LPH, TPH*	$n_b n_\ell (n_t + 1)$
2	busy	sensing	BPH, SPH, TPH*	$n_b n_s (n_t + 1)$
3	busy	transmitting	BPH, TPH	$n_b n_t$
4	idle	sleeping	IPH, LPH, TPH*	$n_i n_\ell (n_t + 1)$
5	idle	sensing	IPH, SPH, TPH*	$n_i n_s (n_t + 1)$
6	idle	transmitting	IPH, TPH	$n_i n_t$

BPH: channel busy phase; IPH: channel idle phase; LPH: SU sleeping phase; SPH: SU sensing phase; TPH: SU trans. phase; TPH*: TPH (extended) at interruption.

We add another state to the set from Model I (channel busy, SU transmitting). Now, the states of this system are classified as shown in Table 3.

Unlike in Model I, in (channel busy, SU sensing) we now have to keep track of the sensing phase. The case in which the system is saturated with SUs is considered first.

The associated transition matrix is

$$P_2^{(\mathrm{sat})} = \begin{bmatrix} D_b \otimes L \otimes I & D_b \otimes (\boldsymbol{l}\boldsymbol{\beta}) \otimes I & 0 & d_{bi} \otimes L \otimes I & d_{bi} \otimes (\boldsymbol{l}\boldsymbol{\beta}) \otimes I & 0 \\ \overline{\phi}_1 D_b \otimes (\mathbf{1}\boldsymbol{\delta}) \otimes I & \phi_1 D_b \otimes S \otimes I & \phi_1 D_b \otimes \boldsymbol{s} \otimes Q^* & \overline{\phi}_1 d_{bi} \otimes (\mathbf{1}\boldsymbol{\delta}) \otimes I & \phi_1 d_{bi} \otimes S \otimes I & \phi_1 d_{bi} \otimes \boldsymbol{s} \otimes Q^* \\ \overline{\phi_2} D_b \otimes \boldsymbol{\delta} \otimes \overline{T} & \phi_2 D_b \otimes \boldsymbol{\beta} \otimes \overline{\boldsymbol{t}} & \phi_2 D_b \otimes T & \overline{\phi}_2 d_{bi} \otimes \boldsymbol{\delta} \otimes \overline{T} & \phi_2 d_{bi} \otimes \boldsymbol{\beta} \otimes \overline{\mathbf{t}} & \phi_2 d_{bi} \otimes T \\ d_{ib} \otimes L \otimes I & d_{ib} \otimes (\boldsymbol{l}\boldsymbol{\beta}) \otimes I & 0 & D_i \otimes L \otimes I & D_i \otimes (\boldsymbol{l}\boldsymbol{\beta}) \otimes I & 0 \\ \overline{\phi}_1 d_{ib} \otimes (\mathbf{1}\boldsymbol{\delta}) \otimes I & \phi_1 d_{ib} \otimes S \otimes I & \phi_1 d_{ib} \otimes \mathbf{s} \otimes Q^* & \theta_1 D_i \otimes (\mathbf{1}\boldsymbol{\delta}) \otimes I & \overline{\theta}_1 D_i \otimes S \otimes I & \overline{\theta}_1 D_i \otimes \boldsymbol{s} \otimes Q^* \\ \overline{\phi}_2 d_{ib} \otimes \boldsymbol{\delta} \otimes \overline{T} & \phi_2 d_{ib} \otimes \boldsymbol{\beta} \otimes \overline{\boldsymbol{t}} & \phi_2 d_{ib} \otimes T & \theta_2 D_i \otimes \boldsymbol{\delta} \otimes \overline{T} & \overline{\theta}_2 D_i \otimes \boldsymbol{\beta} \otimes \overline{\boldsymbol{t}} & \overline{\theta}_2 D_i \otimes T \end{bmatrix}. \quad (61)$$

As we did in Model I, in order to study the case of nonsaturated SUs we need to extract, from the above matrix, the following set of matrices:

$$F_0 = \begin{bmatrix} 0 & D_b \otimes \boldsymbol{\beta} \otimes [0_{1\times n_t} \ 1] & 0 & 0 & d_{bi} \otimes \boldsymbol{\beta} \otimes [0_{1\times n_t} \ 1] & 0 \\ 0 & d_{ib} \otimes \boldsymbol{\beta} \otimes [0_{1\times n_t} \ 1] & 0 & 0 & D_i \otimes \boldsymbol{\beta} \otimes [0_{1\times n_t} \ 1] & 0 \end{bmatrix}, \quad (62)$$

$$F_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \phi_2 D_b \otimes \boldsymbol{t} & \phi_2 d_{bi} \otimes \boldsymbol{t} \\ 0 & 0 \\ 0 & 0 \\ \phi_2 d_{ib} \otimes \boldsymbol{t} & \overline{\theta}_2 D_i \otimes \boldsymbol{t} \end{bmatrix}. \quad (63)$$

$$H_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \phi_2 D_b \otimes \boldsymbol{\beta} \otimes \overline{\boldsymbol{t}} & 0 & 0 & \phi_2 d_{bi} \otimes \boldsymbol{\beta} \otimes \overline{\boldsymbol{t}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \phi_2 d_{ib} \otimes \boldsymbol{\beta} \otimes \overline{\boldsymbol{t}} & 0 & 0 & \overline{\theta}_2 D_i \otimes \boldsymbol{\beta} \otimes \overline{\boldsymbol{t}} & 0 \end{bmatrix}, \quad (64)$$

$$H_0 = P_2^{(\mathrm{sat})} - H_1. \quad (65)$$

Using the same ideas as in the case of Model I, we can easily study the nonsaturated cases, both finite and infinite buffer situations. We skip this for Model II as it is merely a repetition of the procedure used in Model I.

6. Numerical Results

In this section, we present some results to exemplify the capabilities of the proposed models and to show the feasibility of their numerical evaluation.

Since the source of SUs is not saturated, the arrival process of SUs must be taken into account. In our numerical results, the arrival of SUs is modeled by a *platoon arrival process* (PAP) in which the following magnitudes follow geometric distributions: interplatoon times (mean value, 100); intraplatoon interarrival times (mean value, 20); and number of arrivals in a platoon (mean value, N_p); see [18, pp. 49–50] for further details. The mean number of arrivals in a platoon, N_p, is varied from 1 to 40 so that the arrival rate of SUs varies from 0.01 to 0.0455 arrivals per time slot. Then, the matrices of the MAP are

$$G_0 = \begin{bmatrix} 1 - \dfrac{1}{100} & 0 \\ 0 & 1 - \dfrac{1}{20} \end{bmatrix}, \quad (66)$$

$$G_1 = \begin{bmatrix} \frac{1}{N_p}\frac{1}{100} & \left(1-\frac{1}{N_p}\right)\frac{1}{100} \\ \frac{1}{N_p}\frac{1}{20} & \left(1-\frac{1}{N_p}\right)\frac{1}{20} \end{bmatrix}. \quad (67)$$

The number of slots in a sensing cycle has a deterministic value equal to 4, that is,

$$n_s = 4, \quad (68)$$

$$\boldsymbol{\beta} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}, \quad (69)$$

$$S = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (70)$$

The duration of a sleep period is uniformly distributed in $\{1, 2, \ldots, 10\}$:

$$n_\ell = 10, \quad (71)$$

$$\boldsymbol{\delta} = \begin{bmatrix} 0.1 & 0.1 & \cdots & 0.1 \end{bmatrix}, \quad (72)$$

$$L = \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ & 1 & 0 & & \\ & & \ddots & \ddots & \\ & & & 1 & 0 \end{bmatrix}. \quad (73)$$

The duration of an uninterrupted SU transmission follows a negative binomial distribution with mean 10 slots and variance 2.5:

$$n_t = 8, \quad (74)$$

$$\boldsymbol{\alpha} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (75)$$

$$T = \begin{bmatrix} 0.6 & 0.4 & & & \\ & 0.6 & 0.4 & & \\ & & \ddots & \ddots & \\ & & & 0.6 & 0.4 \end{bmatrix}. \quad (76)$$

A preemptive repeat scheme is used when the transmission of an SU is interrupted by the transmission of a PU. Thus,

$$Q = \mathbf{1}\boldsymbol{\alpha} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix}. \quad (77)$$

Finally, we consider ideal sensing, $\phi_1 = \phi_2 = \theta_1 = \theta_2 = 0$, and four different buffer sizes, $N = 1, 10, 100, \infty$.

As in [2], the model proposed in [21] is used for channel occupancy. This model exhibits a self-similar behavior over an adjustable range of time-scales. Then, the transition matrix in (1) becomes

$$D = \left[\begin{array}{c|ccccc} 1-\sum_{k=1}^{n-1}\frac{1}{a^k} & \frac{1}{a} & \frac{1}{a^2} & \cdots & \frac{1}{a^{n-1}} \\ \hline \frac{b}{a} & 1-\frac{b}{a} & & & \\ \left(\frac{b}{a}\right)^2 & & 1-\left(\frac{b}{a}\right)^2 & & \\ \vdots & & & \ddots & \\ \left(\frac{b}{a}\right)^{n-1} & & & & 1-\left(\frac{b}{a}\right)^{n-1} \end{array}\right] \quad (78)$$

where n, a, and b are the parameters of the model. The range of time-scales where the process shows a self-similar behavior can be adjusted the value of parameter n. Once the value of n is set, the values of a and b can be determined by fitting the channel utilization factor $\rho_{\mathrm{PU}} = (1 - 1/b)/(1 - 1/b^n)$, and the average burst length $\mathrm{E}[B] = (\sum_{k=1}^{n-1} a^{-k})^{-1}$. The following values have been considered here: $n = 8$, $\rho_{\mathrm{PU}} = 0.2$, and $\mathrm{E}[B] = 10$.

The chosen channel model, arrival process, and PH distributions do not intend to fit any specific scenario characterized by empirical results. Instead, our purpose is to show how a wide variety of distributions and characterizations can be easily incorporated into our models and to show the feasibility of their numerical evaluation. For this purpose, the channel model and the PH distributions that have been used are of a relatively high number of phases and, as shown in this section, the resulting DTMC is still amenable to numerical evaluation using a conventional desktop computer. Using distributions and models with a similar number of phases would allow obtaining a good fit of the available data in a broad variety of practical cases, while the required computational effort would be comparable to that of the examples shown here.

Figures 1, 2, 3, and 4 show, respectively, the mean number of SUs in the system, loss probability, throughput, and goodput. All of them are represented as a function of the mean number of arrivals in a platoon (N_p) for different sizes of the buffer. When N_p increases so do the rate and burstiness of the arrival process of SUs. Thus, as it may be expected, the curves for in all four graphs exhibit an increasing trend.

In the cases where buffer size is finite, the value of the mean number of SUs in the system (Figure 1) rapidly saturates to about half the size of the buffer. This observation is consistent with the behavior of the loss probability (see Figure 2).

The curves for the throughput and goodput (Figures 3 and 4) show a similar shape, but the values of the goodput are lower than those of the throughput (around 12% lower on average). The latter is due to the fact that unfinished SU transmissions have to be repeated from the beginning the next time the SU gains access to the channel.

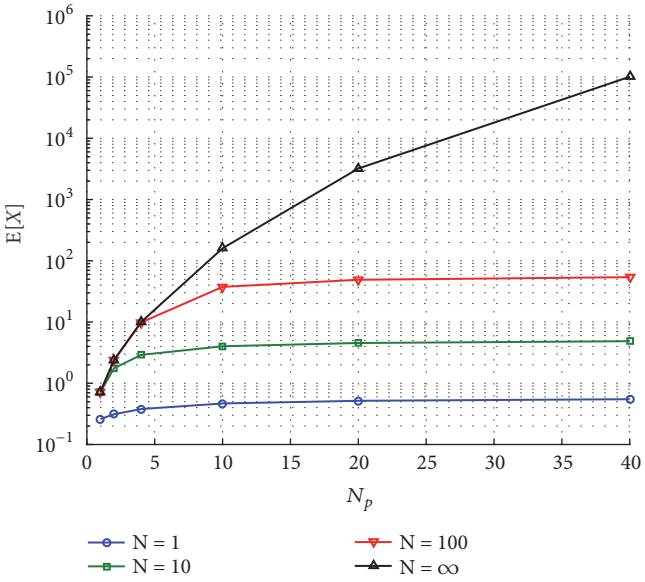


Figure 1: Mean number of SUs in the system, E[X].

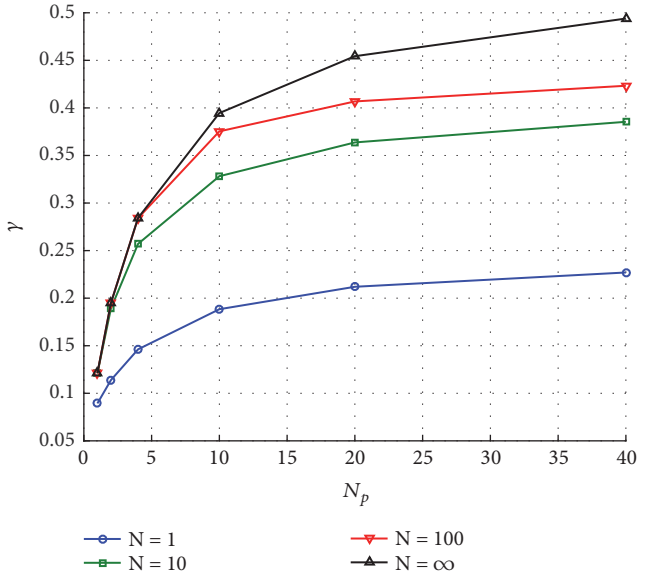


Figure 3: Throughput, γ.

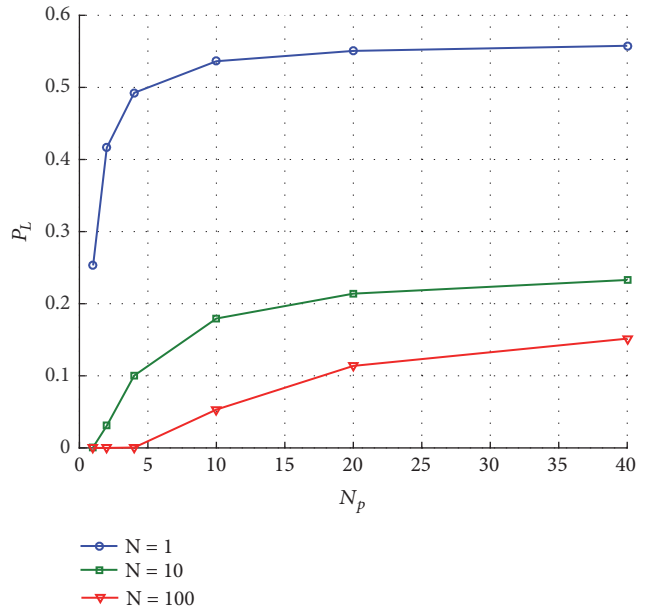


Figure 2: Loss probability, P_L.

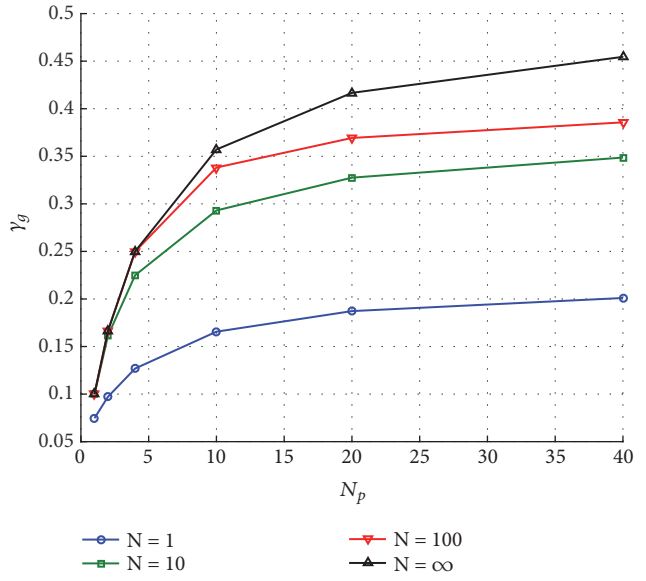


Figure 4: Goodput, γ_g.

In our examples here, more important than the values of the performance parameters themselves is the computational effort required to compute them. In Figure 5 we show the required execution time for our implementation of the analysis in Matlab 2015 that was run in a laptop with an Intel Core i7-4702MQ, 2.2 GHz, and 16 GB RAM.

As observed, the execution time is of the order of 1 minute in the worst case ($N = 100$). It is also observed that, for the finite buffer model, the required execution time increases linearly with the buffer size. Therefore, when the buffer size increases, at some point the computational effort for the finite buffer case exceeds that of the infinite buffer one. Note, however, that if the buffer size is large enough, then it can be approximated by an infinite buffer system.

7. Conclusions

In this paper, we have proposed and analyzed a number of Markovian models that enable the analysis and evaluation of sensing strategies in cognitive radio networks under a broad range of conditions. The proposed models are quite versatile and general. A Markov phase renewal process is used to model the channel availability for secondary users (SUs). This allows considering a wide variety of distributions for the duration of idle and busy intervals and also capturing correlations between consecutive intervals. The behavior of SUs is

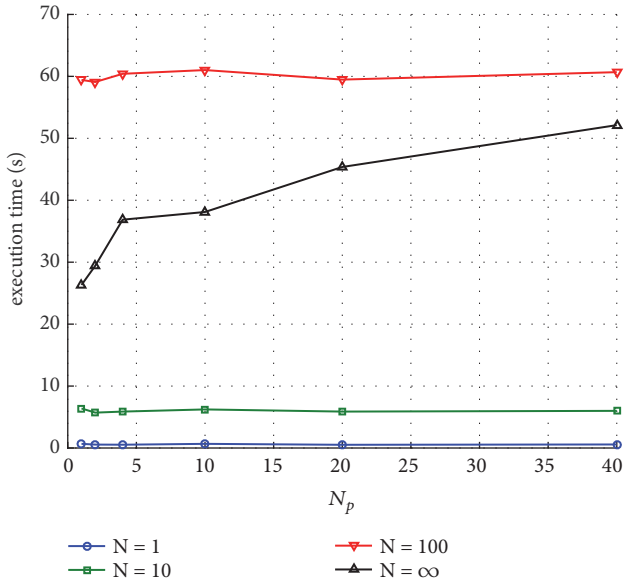


Figure 5: Execution time in seconds.

modeled using general *phase-type* distributions and sensing capabilities. A major contribution with respect our previous work is that the unrealistic assumption that the source of SUs is saturated has been relaxed in the models presented here, and the arrival process of the SUs is introduced. This arrival process is represented by the Markovian arrival process, which is also very general and versatile. Furthermore, we consider a general buffer size, either finite or infinite, for the waiting SUs and a fairly general resumption of an SU transmission after being interrupted by PUs activity. Finally, some numerical examples and results have been presented to show the capabilities of our models and the feasibility of their numerical analysis.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The research of V. Pla was partially supported by the Ministry of Education of Spain under Grant PR2011-0055. The research of A. S. Alfa was partially supported by the NRF SARChI Chair in ASN cohosted by UP and CSIR.

References

[1] M. Conti, S. Chong, S. Fdida et al., "Research challenges towards the Future Internet," *Computer Communications*, vol. 34, no. 18, pp. 2115–2134, 2011.

[2] A. S. Alfa, V. Pla, J. Martinez-Bauset, and V. Casares-Giner, "Discrete time analysis of cognitive radio networks with imperfect sensing and saturated source of secondary users," *Computer Communications*, vol. 79, pp. 53–65, 2016.

[3] T. Yücek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 1, pp. 116–130, 2009.

[4] J. Zhang, F.-C. Zheng, X.-Q. Gao, and H.-B. Zhu, "Which is better for opportunistic spectrum access: The duration-fixed or duration-variable MAC frame?" *IEEE Transactions on Vehicular Technology*, vol. 64, no. 1, pp. 198–208, 2015.

[5] W. S. Jeon, J. Gu, and D. G. Jeong, "Optimal sensing strategy against collision with primary user in cognitive radio systems," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 3, pp. 1230–1235, 2015.

[6] H. Pradhan, S. S. Kalamkar, and A. Banerjee, "Sensing-Throughput Tradeoff in Cognitive Radio with Random Arrivals and Departures of Multiple Primary Users," *IEEE Communications Letters*, vol. 19, no. 3, pp. 415–418, 2015.

[7] L.-F. Huang, S.-L. Zhou, D. Guo, and H.-C. Chao, "MHC-MAC: Cognitive MAC with asynchronous-assembly line mode for improving spectrum utilization and network capacity," *Mathematical and Computer Modelling*, vol. 57, no. 11-12, pp. 2742–2749, 2013.

[8] S. Wang, Y. Wang, J. P. Coon, and A. Doufexi, "Energy-efficient spectrum sensing and access for cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 2, pp. 906–912, 2012.

[9] L. Tang, Y. Chen, E. L. Hines, and M.-S. Alouini, "Effect of primary user traffic on sensing-throughput tradeoff for cognitive radios," *IEEE Transactions on Wireless Communications*, vol. 10, no. 4, pp. 1063–1068, 2011.

[10] Y. Wu and D. H. K. Tsang, "Energy-efficient spectrum sensing and transmission for cognitive radio system," *IEEE Communications Letters*, vol. 15, no. 5, pp. 545–547, 2011.

[11] Y. Wu, V. K. N. Lau, D. H. K. Tsang, and L. P. Qian, "Energy-efficient delay-constrained transmission and sensing for cognitive radio systems," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 7, pp. 3100–3113, 2012.

[12] A. T. Hoang, Y.-C. Liang, D. T. C. Wong, Y. Zeng, and R. Zhang, "Opportunistic spectrum access for energy-constrained cognitive radios," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1206–1211, 2009.

[13] N. B. Chang and M. Liu, "Optimal channel probing and transmission scheduling for opportunistic spectrum access," *IEEE/ACM Transactions on Networking*, vol. 17, no. 6, pp. 1805–1818, 2009.

[14] W.-Y. Lee and I. F. Akyildiz, "Optimal spectrum sensing framework for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, pp. 3845–3857, 2008.

[15] S. Gunawardena and W. Zhuang, "Service response time of elastic data traffic in cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 3, pp. 559–570, 2013.

[16] S. Geirhofer, L. Tong, and B. M. Sadler, "Dynamic spectrum access in the time domain: Modeling and exploiting white space," *IEEE Communications Magazine*, vol. 45, no. 5, pp. 66–72, 2007.

[17] M. Wellens, J. Riihijärvi, and P. Mähönen, "Empirical time and frequency domain models of spectrum use," *Physical Communication*, vol. 2, no. 1-2, pp. 10–32, 2009.

[18] A. S. Alfa, *Applied discrete-time queues*, Springer, New York, NY, USA, 2nd edition, 2015.

[19] Y. H. Bae, A. S. Alfa, and B. D. Choi, "Analysis of a contention-based opportunistic spectrum access under general channel activity model," *Performance Evaluation*, vol. 68, no. 3, pp. 271–289, 2011.

[20] M. Shaked and J. G. Shanthikumar, *Stochastic Orders and Their Applications*, Academic Press, San Diego, Calif, USA, 1994.

[21] S. Robert and J.-Y. Le Boudec, "New models for pseudo self-similar traffic," *Performance Evaluation*, vol. 30, no. 1-2, pp. 57–68, 1997.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 5742712, 11 pages
https://doi.org/10.1155/2018/5742712

Research Article

On the Goodness of Using Orthogonal Channels in WLAN IEEE 802.11 in Realistic Scenarios

Jose Manuel Gimenez-Guzman[1], **Ivan Marsa-Maestre**[1], **David Orden**[2], **Enrique de la Hoz**[1] and **Takayuki Ito**[3]

[1]*Computer Engineering Department, University of Alcala, Spain*
[2]*Department of Physics and Mathematics, University of Alcala, Spain*
[3]*Department of Computer Science, Nagoya Institute of Technology, Japan*

Correspondence should be addressed to Jose Manuel Gimenez-Guzman; josem.gimenez@uah.es

Received 22 June 2018; Revised 29 October 2018; Accepted 12 November 2018; Published 22 November 2018

Academic Editor: Laurie Cuthbert

Due to the high density of Wi-Fi networks, especially in the unlicensed 2.4 GHz frequency band, channel assignment has become a critical duty for achieving a satisfactory user experience. Probably, the main peculiarity of Wi-Fi networks is the partial overlap of the radio channels that can be used by access points. For that reason, a number of works avoid cochannel interferences by using only channels which are far enough from each other to have no interferences, the so-called orthogonal channels. However, there is a range of choices between using the whole spectrum and using only orthogonal channels. In this work we evaluate the influence of the choice of channel set in realistic settings, using both optimization and heuristic approaches. Results show that the optimizer is not able to achieve better results when using the whole spectrum instead of restricting to only the orthogonal channels. In fact, the optimizer uses mainly the orthogonal channels when they are available, while the heuristics considered lose performance when more channels are available. We believe this insight will be useful to design new heuristics for Wi-Fi channel assignment.

1. Introduction and State of the Art

Wireless technologies have reached an impressive popularity in the last years. Probably, the most widespread wireless technology is the one based in the standard family IEEE 802.11, commercially known as Wi-Fi. However, the radio spectrum is very limited, and this scarcity is especially important in unlicensed frequency bands, like the ones where Wi-Fi operates in. Moreover, in unlicensed frequency bands, in addition to the Wi-Fi devices we also find other devices that can cause harmful interferences to Wi-Fi, like microwave ovens, Bluetooth devices, or baby monitors. Due to this high saturation and scarcity of the radio spectrum, it is of paramount importance for Wi-Fi devices to choose the adequate channel to operate in, from the set of available channels.

One of the prominent peculiarities of Wi-Fi networks is that the available frequency channels are partially overlapped. For example, in the most popular 2.4 GHz frequency band, from the 13 available channels (depending on the world region this number can be different) only a subset of them do not overlap (orthogonal channels). As the problem of channel assignment in WLAN environments is a current and increasingly important problem, it has been extensively studied by the scientific community. Doubtlessly, the main work that compiles the most prominent efforts in this area is the survey [1]. In this work, WLAN channel assignment techniques are classified into two categories: (i) centrally managed and (ii) uncoordinated. In addition to the papers cited in [1] related to channel assignment, since its publication to the present, some of the most prominent works published in this category are [2–8].

Although the classification provided in [1] is the most logical when studying the problem of Wi-Fi channel assignment, there is an additional dimension which is especially relevant for this work: whether they consider only the orthogonal channels or the whole spectrum of channels. Note that the

orthogonal channels are channels 1, 6, and 11 when we have 11 available channels at hand (e.g., in the US) and channels 1, 7, and 13 when we have 13 channels available (e.g., in the EU). The main efforts that consider only the three orthogonal channels are [2, 3, 5, 9–12]. On the other hand, there are many approaches that consider the whole spectrum, like, for example, [13–17]. At this point, one could wonder why there are so many papers that consider only three orthogonal channels. Probably, the first idea could be that solving the problem with three nonoverlapping channels is easier to solve that its counterpart that considers 13 overlapped channels. In addition, using three orthogonal channels is easier to evaluate, especially with simulators, as many of them do not consider overlaps between channels.

On the other hand, some of the works [18–20] studying the effect of using partially overlapped channels draw contradictory conclusions. For example, in [18, 19] authors conclude that using partially overlapped channels in Wi-Fi enables a performance improvement. However, other approaches like [20] consider that the gain of using partially overlapped channels is not guaranteed. We believe the respective biases of those studies are due to the fact that the comparison is done using only a heuristic approach, so the results strongly depend on the heuristic considered.

One of the reasons to use heuristics for these studies is the high complexity of optimal channel assignments in Wi-Fi networks, which is already stated in [1, 21]. Some papers [5, 21] formulate channel assignment as an optimization problem, but due to its NP-hardness, they propose a heuristic algorithm to solve it. From the works cited in the survey [1], only two of them [9, 16] make use of optimization techniques (in particular, they both use Integer Linear Programming).

In this paper we thoroughly study the effect of varying the distance between Wi-Fi channels allowed in assignments, ranging from 1 (using the whole spectrum) to 6 (using only orthogonal channels in a 13 channel spectrum), and we perform this study not only analyzing the usual heuristics, but also optimal assignments obtained via nonlinear optimizers. The main contributions of this paper can be summarized as follows:

(i) We model the problem of Wi-Fi channel assignment as a multilayer weighted geometric digraph (Section 2.2).

(ii) We model the goodness of Wi-Fi channel assignments with a nonlinear utility model, which takes into account the saturation effect occurring when signal to interference ratio (SIR) is either very high or very low (Section 2.4).

(iii) We consider channel assignments obtained by a centralized nonlinear optimizer based on simulated annealing, apart from the usual channel assignment heuristics (Section 3).

(iv) We propose a set of realistic Wi-Fi scenarios for performance evaluation, constructed from real Wi-Fi maps from University buildings (Section 4.1).

(v) We study the effect of increasing the number of available channels, either with or without taking into account cochannel interference (Section 4). We also compare the overall gain obtained for the most widespread channel spectrum settings: 11 and 13 channels. We analyze not only the aggregate utility of the solutions, but also the distribution of channels in the assignments.

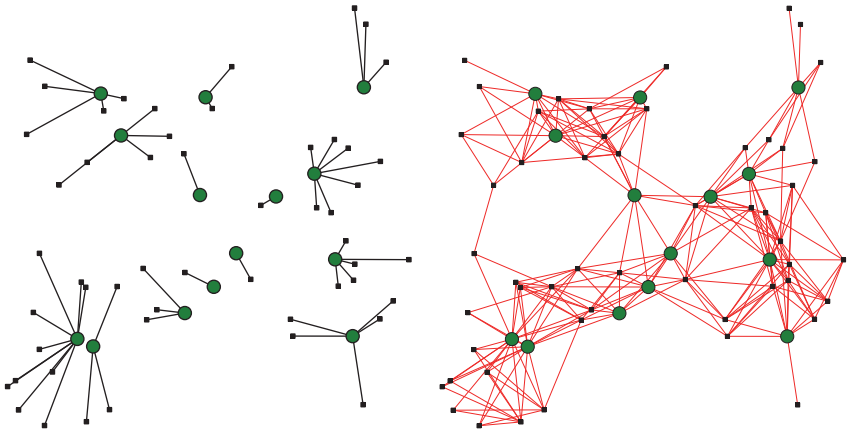
This is, to the best of our knowledge, the first time that the effect of channel distance and number of channels on the goodness of channel assignments is studied, using optimization techniques and in a realistic setting. Since the publication of [1], the only works (apart from our own [22–24]) that use optimization techniques to solve the channel assignment problem in WLANs are [11, 12]. However, neither of these papers studies the effect of using overlapping channels instead of using nonoverlapping channels. As the papers that have study the possibility of using overlapping channels have contradictory conclusions, we think that this paper fills a gap in the literature, as it tackles the channel assignment as an optimization problem. Moreover, another novelty of the paper consists of considering intermediate cases in the set of available channels that can be assigned to access points, i.e., intermediate situations between using only the three nonoverlapping channels and using all the channels of the spectrum, as we can have more than three channels at hand while avoiding the severe interferences that appear between adjacent channels.

The paper can be structured as follows. In Section 2 we describe the Wi-Fi architecture under consideration, later describing accurately the model we propose for such an architecture. Section 3 is devoted to the description of the optimization and heuristic algorithms used for assigning channels in Wi-Fi settings. The scenarios under study and the results of the paper are shown in Section 4. Finally, Section 5 summarizes the paper and the most important conclusions, proposing some future research lines.

2. Network Model

2.1. Wi-Fi Architecture. IEEE 802.11 technology, commercially known as Wi-Fi, is, without any doubt, the most widespread WLAN technology currently deployed. We focus on Wi-Fi networks operating in infrastructure mode, so there are two types of wireless network elements: access points (APs) and wireless devices (WDs). The goal is to assign a channel to each AP, which will be the channel used for the communication between that AP and its associated WDs. Note that the term WD includes a wide range of devices, such as laptops, smartphones, ebooks, TVs, or any other device able to connect to a Wi-Fi network. In infrastructure mode, direct communication occurs only between WDs and APs, as communications between WDs are not permitted without passing through an AP.

Although the frequency band of 5 GHz is also used in Wi-Fi, we focus on the most used and congested frequency band where Wi-Fi can operate, the unlicensed 2.4 GHz frequency band. This band is composed by 11 to 14 frequency channels where devices can operate. The choice between 11, 13, or 14 channels depends on the world region, with 11 channels used



(a) Connectivity layer (b) Signal layer

Figure 1: Example of graph layers of a Wi-Fi scenario.

in USA and 13 channels used in Europe. For that reason, we focus on the scenario where we have 13 channels at hand, although in the results we compare the differences of having 11 instead of 13 available channels.

The main peculiarity of the Wi-Fi technology is that those channels partially overlap, which makes the problem of frequency assignment more complex, as each channel may interfere its neighboring channels. A first decision when assigning channels to APs in a Wi-Fi network will be to decide the number of channels to take into consideration. Many of the algorithms and implementations for Wi-Fi channel assignment focus on the use of orthogonal channels only (i.e., channels that do not interfere with each other). This choice has the advantage that there are less harmful interferences.

2.2. Graph Modeling. In this work, we model the Wi-Fi network by means of a multilayer graph [25]. Multilayer graphs have recently emerged as a powerful tool to model and capture multiple related aspects of complex systems. In our model, we consider two layers: connectivity layer and signal layer. Both layers are geometric graphs with two different kinds of vertices: APs and WDs.

The connectivity layer is a geometric undirected graph, where the edges represent the logical links between APs and WDs; i.e., the connectivity layer links each WD to the AP to which it is attached to. As it is expected that a WD attaches to the AP that offers the most powerful signal and we assume that power losses depend on the distance, we have attached each WD to its closest AP.

On the other hand, the signal layer is a geometric weighted directed graph (or weighted digraph), whose edges represent all the signals that are received in each vertex. Note that each edge will have a weight that represents the power signal received. We use a directed graph to account for the higher impact of an interference signal coming from an AP rather than coming from a WD; i.e., we consider that the interference produced from an AP is more harmful than the one produced by a WD. For that reason, we have two edges between each WD and AP vertices which interfere, with different weights for the edge representing the interference from the AP to the WD and the edge representing the interference from the WD to the AP.

In the signal layer two vertices are connected by an edge provided that the distance between them is lower than the *coverage area* (*CA*), which is defined as the distance at which the power signal has decreased enough to be equal than the sensitivity of the receiver. Section 2.3 defines the propagation and interference models, and describes how to compute *CA*.

Note that in the signal layer we have two types of edges: interferences and the desired signal. Interferences are those undesired received signals that can come from other APs or from WDs that are attached to other APs. Note that two WDs attached to the same AP do not interfere, as these communications are coordinated by that AP. On the other hand, the desired signal is the information signal we want to receive.

A graphical example of the layers of a Wi-Fi scenario with 14 APs and 49 WDs is shown in Figure 1. Figure 1(a) represents the connectivity layer, while Figure 1(b) shows the signal layer, according to the definitions previously provided. For practical purposes, note that in the graphical representation of the undirected graph of the signal layer we join the two edges between vertices (one for each direction) into a single one.

2.3. Propagation and Interference Models. The geometric properties of the multilayer graph allow us consider realistic propagation effects of the radio signals. Our purpose is to assign weights to the edges of the signal layer in such a way that, as mentioned above, each weight represents the power of the signal represented by each edge. First of all, we have to define the propagation model used to represent the power losses that a radio signal suffers from transmission to reception. We have made use of the model proposed in [26], which defines that power loss (in dB) as

$$P_{loss} = 40 \log_{10} d + 20 \log_{10} f - 20 \log_{10} (h_t h_r). \quad (1)$$

Table 1: Spectral overlap between Wi-Fi channels [27].

$\lvert c_i - c_j \rvert$	0	1	2	3	4	5	6	...
Spectral overlap	1	0.8	0.5	0.2	0.1	0.001	0	0

Equation (1) is defined for situations where transmitting and receiving antennas are close to the ground (between 1 m and 2.5 m), being h_t and h_r their height above the floor expressed in meters. Note that d represents the distance between the transmitter and the receiver (in meters) and f is the frequency (in GHz) of the signal. As the frequency for the different channels where Wi-Fi operates in is very similar, we have assumed f to be constant and equal to 2.4 GHz, obtaining

$$P_{loss} = 7.6 + 40 \log_{10} d - 20 \log_{10} (h_t h_r). \quad (2)$$

As we want to assign the weights to each edge of the signal layer, we are going to identify two different cases: (i) the signal is the one we want to receive, i.e., it contains the desired information, and (ii) the signal is undesired, i.e., it is an interference.

As the Wi-Fi network is operating in infrastructure mode, the desired signal is produced always between a WD and the AP to which the WD is attached to and vice versa. The weight of the edges that represent the desired signal between vertices i and j (V_i and V_j) can be computed as

$$S_{V_i \longrightarrow V_j} = P_t + G_t + G_r - L - P_{loss}, \quad (3)$$

where P_t stands for the transmitting power, expressed in dBm, G_t and G_r stand for the transmission and reception antennas gains, respectively, L stands for losses due to obstacles, like walls, windows, and so on, and is expressed in dB, and P_{loss} is computed with (2).

To compute the weight of the edges representing interferences in the signal layer (between V_i and V_j) we use

$$I_{V_i \longrightarrow V_j} = P_t + G_t + G_r - L - P_{loss} + \Psi + \eta \left(\left| c_i - c_j \right| \right), \quad (4)$$

where every value is expressed in logarithmic scale. It is important to define parameters Ψ and, specially, $\eta(|c_i - c_j|)$. First, we can define Ψ as the activity index that represents the aforementioned more harmful effect of the interferences produced by APs ($\Psi = 0.5$) rather than produced by WDs ($\Psi = 0.2$). Second, the η function represents the cochannel interference, so it is a paramount parameter for the purpose of this work. We have considered the effect of partially overlapped channels, represented by this function η, empirically measured in [27], although the results are not very different from those obtained in other works [18, 28, 29]. More specifically, $\eta(|c_i - c_j|)$ represents the cochannel interference between vertices i and j when they are assigned channels c_i and c_j, respectively. Table 1 shows the cochannel interference defined in [27] and used in this work. As we can see, the overlap between channels decreases as they are farther apart in the spectrum.

Finally, once the interferences have been defined, we can define how we have computed the coverage area (CA) defined in Section 2.2. As we stated, CA represents the distance at which the power signal transmitted by a device is equal to the sensitivity of the receivers (σ). From the definitions above this CA can be easily computed as

$$CA = 10^{(P_t + G_t + G_r - L - \sigma - 7.6 + 20 \log_{10}(h_t h_r))/40}. \quad (5)$$

2.4. Utility of the Solutions. To measure the goodness of a particular channel assignment we use the signal to interference ratio (SIR). As the edges of the signal layer include the power of the different signals that appear in the network, SIR can be directly computed from that layer.

The computation of the SIR is slightly different depending on whether the SIR is being computed for a WD or an AP, so we distinguish both cases. For the i-th WD (vertex V_i in the graph), that is attached to j-th AP (vertex V_j in the graph), SIR can be computed as the quotient between the weight of the edge of the desired signal and the sum of the weights of the edges representing interferences (denoted by subindex k); i.e.

$$SIR_i = \frac{S_{V_j \longrightarrow V_i}}{\sum_k I_{V_k \longrightarrow V_i}}. \quad (6)$$

We have distinguished the SIR value for WDs and for APs as, for APs, we have so many desired signals ($S_{V_j \longrightarrow V_i}$) as WDs it has attached that AP. In that case, we have chosen the worst case; i.e., for the desired signal we have chosen the one that is received with less power, resulting in the SIR that is minimum for every WD.

Although the SIR represents a good performance parameter to evaluate channel assignments, it can be improved to represent a more realistic situation. As we know, user experience depends on the SIR, but this experience has upper and lower bounds, as shown in [30]. If SIR value for a certain device is below a certain value (SIR_{min}), the signal quality is very poor and devices cannot keep connected and, thus, throughput (and hence perceived utility) equals to zero. On the other hand, when SIR is above a certain value SIR_{max} the throughput is limited by the technology, so it has no impact to increase the SIR beyond that value. For those reasons we introduce, as a performance parameter, the utility for device i, named U_i. Parameter U_i ranges from 0 to 1, representing 0 the situation where the SIR is below SIR_{min} and 1 when SIR is above SIR_{max}. For the values between SIR_{min} and SIR_{max} we have used a linear function, as shown in Figure 2.

Finally, the utility for a certain channel assignment can be computed as the sum of the utilities for all the vertices in the graph:

$$U = \sum_{\forall i} U_i. \quad (7)$$

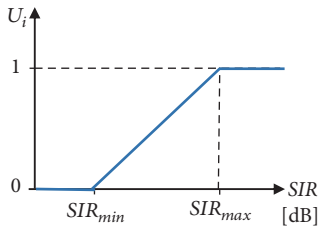


Figure 2: Relation between *SIR* and utility for vertex *i*.

3. Channel Assignment Approaches Used in the Comparison

In our study, according to the classification in the aforementioned survey [1], we have chosen for benchmarking both a centralized and a distributed approach. We have also made experiments with random assignments, to use them as a baseline. In the following we briefly describe the three channel assignment techniques used in the study.

3.1. Baseline: Random Channel Assignment. As a reference, we have used random channel assignments; that is, we have measured the utility values obtained when choosing a random channel for each AP, using a uniform distribution. This gives us a baseline to compare with other approaches, and to have a better view of whether they achieve a reasonably good solution or not. As we will discuss in Section 4, this also gives us an indication of how the search problem becomes harder as the number of available channels change.

3.2. Least Congested Channel Search (LCCS). Distributed heuristics are the *de facto* standard for Wi-Fi-channel assignment. In particular, Least Congested Channel Search (LCCS) [31] works by letting each access point to choose its desired channel autonomously. In order to do so, access points periodically count the number of active wireless devices at each channel, choosing the least used channel for operation. Since LCCS is asynchronous by definition and its convergence is not guaranteed, it is difficult to define a final assignment to use for comparison. To overcome this difficulty, we have implemented a *LCCS hill-climber*, which means that we conduct a number of iterations, allowing the LCCS algorithm to run asynchronously once for each of the APs (with the ordering of the APs randomly changing) at each iteration. If, at the end of the iteration, the overall utility has increased, we keep the changes. Otherwise, we revert the channel assignments before iterating again. We have set the number of iterations to 3000 in our setting.

3.3. Optimization Using Simulated Annealing (SA). Simulated annealing is a centralized nonlinear optimization technique [32], which we have used successfully in the past in frequency assignment problems [22]. The idea of simulated annealing is to search through the solution space at each iteration of the algorithm by evaluating some neighboring solutions to the current state, and deciding whether to move to one of these neighboring solutions with a finite probability, which depends on the utility variation of the movement (the greater the utility loss, the less likely to move) and a parameter called *annealing temperature*, with higher probability of moving when the temperature is high. The temperature is progressively decreasing as the algorithm progresses, guaranteeing its convergence. In this study we used an implementation of SA similar to the one used in [22], with only one neighboring solution evaluated at each iteration (generated by randomly changing the channel assigned to a randomly chosen AP), and by using $P_a = e^{-\Delta u/T}$ as the probability of movement, where Δu is the utility loss for the movement and the temperature T starts at 1 and decreases linearly down to zero as the algorithm iterates. We have used the same number of iterations for SA as the one set for LCCS.

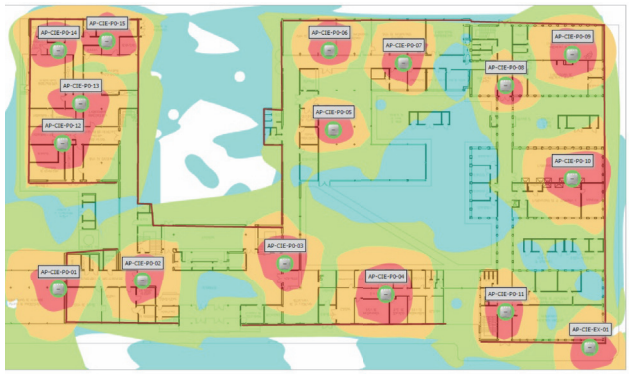
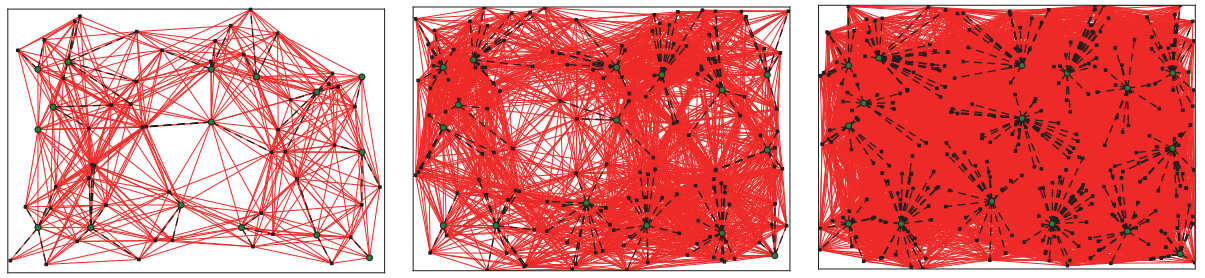


Figure 3: Layout under study.

4. Numerical Study

4.1. Scenarios under Study and Parameters Used. We have made use of realistic settings to evaluate the effect of the use of orthogonal channels in Wi-Fi technology. In particular, we have considered the real layout of the Wi-Fi campus network from the Science Building of the University of Burgos (Spain), which we show as a heat map representation in Figure 3. With this layout, we have considered a wide range of situations increasing the density of WDs. More specifically, we have considered scenarios with 50, 100, 150, 200, 250 and 300 WDs. This choice originates very different situations regarding interferences, as it can be seen in Figure 4, where we show the graph models for some of these scenarios, including the simplest (with 50 WDs) and the most complex (with 300 WDs). In this figure, we have merged the two layers of the graph model, showing APs as green circles and WDs as smaller black squares. Moreover, the edges of the signal layer are depicted with solid red lines and the edges of the connectivity layer with black dashed lines.

As mentioned before, our goal will be to assign the frequency channels to the different devices in the network in order to maximize the experience perceived by network users, i.e., in order to maximize the utility function U defined in Section 2.4. Typically, channel assignment can be recognized as a graph coloring problem [33], so an example of the result of applying any of the channel assignment algorithms can be shown in Figure 5, where we have represented each of the 13 Wi-Fi channels by a color.



(a) 50 WDs (b) 150 WDs (c) 300 WDs

Figure 4: Graph model.

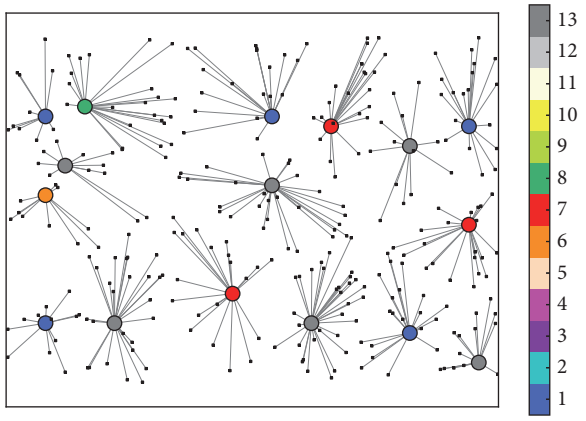


Figure 5: Example of coloring with 300 WDs.

Table 2: Summary of parameters.

Parameter	Value
P_t	30 mW
G_t	0 dB
G_r	0 dB
L	40 dB
σ	−90 dBm
h_t	1.5 m
h_r	1.5 m
Ψ (APs)	0.5
Ψ (WDs)	0.2
SIR_{min}	10 dB
SIR_{max}	40 dB

As there are many parameters that define the models proposed in Section 2, we summarize the parameters used in Table 2, as defined in [22].

4.2. Effect of the Number of Channels. In this section we evaluate the impact of considering the use of only orthogonal channels, in comparison with the use of a wider number of channels up to a maximum of 13 channels. In all cases we have considered evenly spaced possibilities. For that reason, the number of channels used in each case, together with the specific channels considered for each setting, is as shown in Table 3. Note that we have considered all the possibilities that produce evenly spaced channels.

Table 3: Number of channels and channels in use.

Number of channels	Separation	Channels in use
3	6	1, 7, 13
4	4	1, 5, 9, 13
5	3	1, 4, 7, 10, 13
7	2	1, 3, 5, 7, 9, 11, 13
13	1	1-13

4.2.1. Without Cochannel Interference. First of all, we evaluate the effect of the number of channels assuming there are no cochannel interferences, i.e., assuming that all the available 13 channels are orthogonal. Although this situation is not realistic, this experiment is performed to analyze the effect that cochannel interference has on performance, because all the subsequent experiments consider such an interference. In addition, this first experiment lets us validate the model and analyze the effect of the number of channels if cochannel interferences could be avoided.

In this setting, the cochannel interference is 1 for $c_i = c_j$ and 0 otherwise. Results are shown in Figure 6. This figure shows the mean and 95% confidence intervals (showed with the shaded area) obtained as a result of 100 random assignments, and for the different number of WDs under study. Comparing the different curves, we conclude that, as the number of WDs increases, the utility also increases. This is an expected result, as there are more devices to account for the utility when the number of WDs increases. Another identified pattern, also expected, is that the obtained utility increases as we can make use of more channels. This makes sense, since the availability of new channels reduces the amount of "monochromatic edges", that is, adjacent vertices in the signal layer using the same channel, so interference reduces drastically. In the limit, if we had as many available channels as APs, interference would be zero and utility would be maximized.

4.2.2. With Cochannel Interference. In spite of the results shown in the previous section, it is of paramount importance

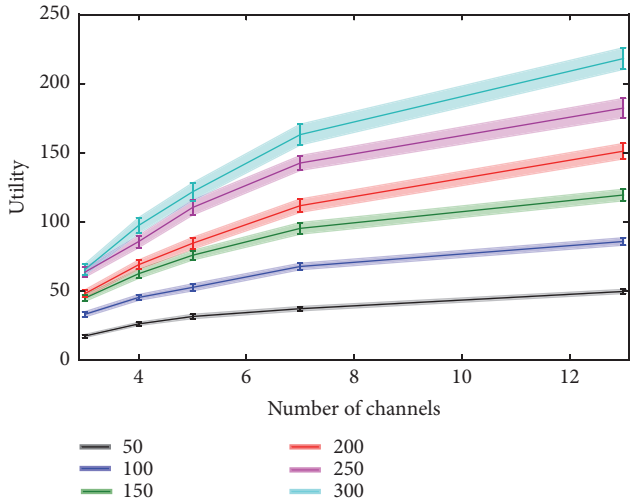


Figure 6: Effect of the number of channels without cochannel interference.

to consider the cochannel interference, as it is probably the most peculiar feature of Wi-Fi networks. As stated above, for the cochannel interference we have used the data obtained in [27].

Now we evaluate the effect that considering orthogonal channels has in performance, in comparison with using an increasing number of evenly spaced channels. This experiment has been driven for different scenarios and using different algorithms to assign channels. The behavior is shown in Figure 7, where we also show the confidence intervals at 95%. To make confidence intervals narrow enough for the the results to be statistically significant, we have run the LCCS and random channel assignment approaches 1000 times, while the SA optimizer has been run for 50 times (the standard deviation for the SA results is much lower). It is very remarkable the local minimum that appears at every curve when we have 4 channels at hand. This effect is due to the fact that the experiments that consider 4 channels are the only ones that cannot use channel 7, and, as we will see later, channel 7 is one of the most widely used in optimal assignments. In particular, having 4 channels available permits us to use channels 5 and 9, which are not orthogonal either to each other or to channels 1 and 13, respectively. However, results show that it is better to use those new channels (and hence enduring the associated interferences) than to avoid them (therefore having only the two channels 1 and 13, which are orthogonal, but being forced to have a number of high-interference monochromatic edges).

However, the main conclusion from Figure 7 is that, when we increase the number of available channels and consider cochannel interference, the utility achieved does not increase. This result is counterintuitive, and it probably justifies the wide number of research studies that only consider three orthogonal channels in Wi-Fi. In fact, following Figure 7 we could recommend not to use more than 3 channels, as the complexity of the problem increases with no gain in performance, probably because the network layout can be efficiently covered with three colors or, for the scenarios under study, because the AP placement has been done considering that only the orthogonal channels were to be used. In particular, we notice that the optimizer SA is not able to obtain higher gains in utility when it can use a higher number of channels. In addition, it must be noted that the problem complexity of using 13 channels is much higher than considering 3, as the state space of possible solutions increases from 3^N to 13^N, where N is the number of APs in the network.

If we compare the performance of the different techniques under evaluation (Figure 7), we notice that random is the worst technique, followed by LCCS. As it could be expected, SA offers the best results. In fact, as SA is an optimizer, we could consider that the results obtained by SA can be considered as upper bounds in performance. As SA requires the complete information about the network, we do not present SA as a realistic technique to be easily used in real networks (it could only be used when we use a central controller). For that reason, comparing the performance of Random and LCCS with SA we can estimate the room for improvement that widely used techniques have. Moreover, it is very important to note that, analyzing the utility achieved by Random and LCCS as the number of available channels increase, this utility decreases. This conclusion is very important, as it can be deduced that when using random or LCCS assignments, it is worth to use 3 nonoverlapping channels, because not choosing the channel correctly can generate more undesired interferences.

In addition to the previous comparison, we now inspect the channels used by the different techniques and for the different deployments. First, and for the sake of space, it is important to note that we focus on the scenario with 200 WDs as a reference, but identical conclusions can be drawn from the rest of scenarios. On one hand, Figure 8 shows the proportion of colors actually used in the channel assignments obtained by the optimizer for the different number of available colors. When we use 3 channels, we note that they are used evenly, as they do not interfere with each other at all. However, when 4 channels are available, we note that channels 5 and 9 are less used than 1 and 13, so the conclusions given before about the utility achieved when having 4 channels are confirmed. Finally, it is worth to note that when we have 5 and 7 channels available, the optimizer only uses the three orthogonal channels (1, 7, and 13), being the use of other channels negligible. A similar conclusion can be given when we have 13 available channels, although the use of other channels different than 1, 7 and 13 is a bit higher.

On the other hand, Figure 9 is the equivalent but for LCCS. In this case, we can notice that LCCS makes use of more channels than the three orthogonal ones, but the most widely used are channels 1 and 13.

4.3. Comparison between the Use of 11 and 13 Channels. As it has been previously shown, increasing the number of channels does not increase the performance, when we consider cochannel interference. For that reason, now we conduct a study comparing the performance that can be achieved when using 11 or 13 channels. Although the choice of having 11 or 13 channels at hand depends on the world region and cannot be chosen individually, following the results of

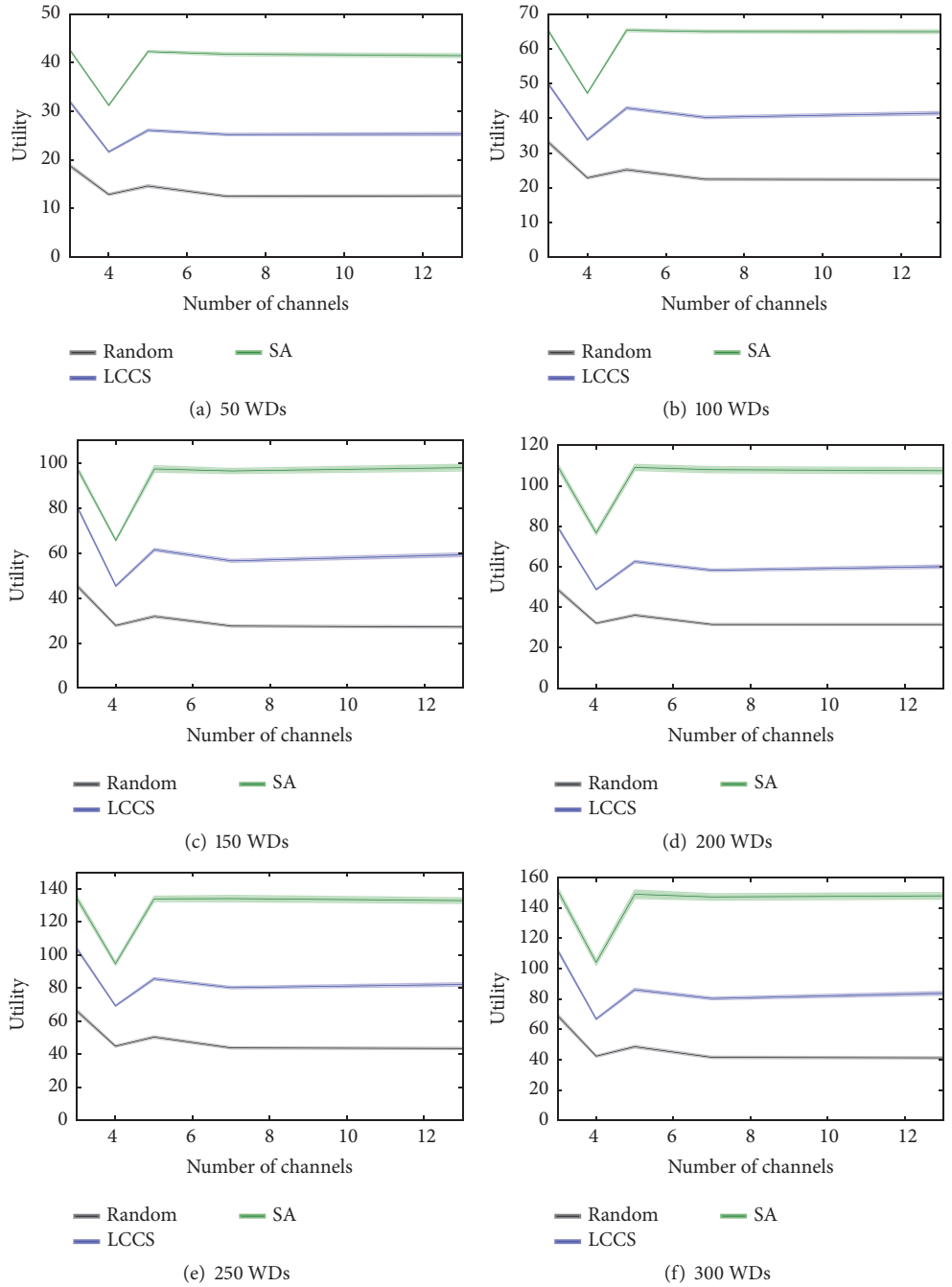


Figure 7: Achieved utility for different number of available channels.

the previous section, it is worth to evaluate the differences that both possibilities have. The main results are summarized in Figure 10, where we show the quotient between the utility achieved when letting the use of 13 channels and 11 channels. Results show that, for all the assignment algorithms under study, the performance of having 13 channels at hand is about 15%-25% better than having only 11 channels. This effect is due to the fact that, although the best choice is to use a reduced number of channels, those channels are more spaced when using 13 channels.

5. Conclusions

Channel assignment is probably the most important configuration issue when deploying Wi-Fi networks. As there is an increasing number of Wi-Fi networks, together with other devices like microwaves, baby monitors, and so on sharing the same spectrum, we must accurately choose in which frequency channel must operate each access point. This frequency planning problem becomes more complex as one of the most peculiarities of Wi-Fi networks is that frequency

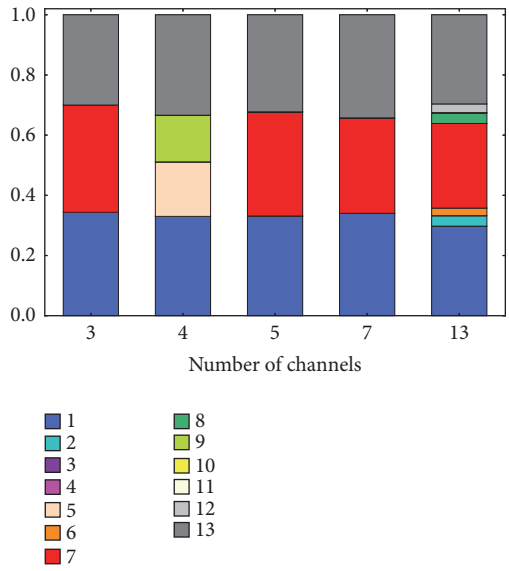


Figure 8: Proportion of colors used in SA with 200 WDs.

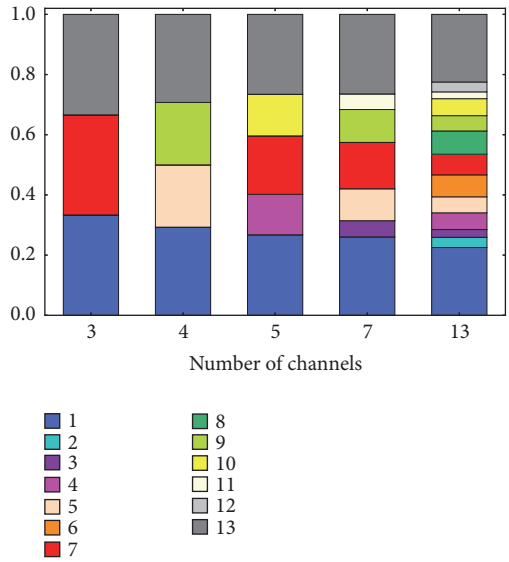


Figure 9: Proportion of colors used in LCCS with 200 WDs.

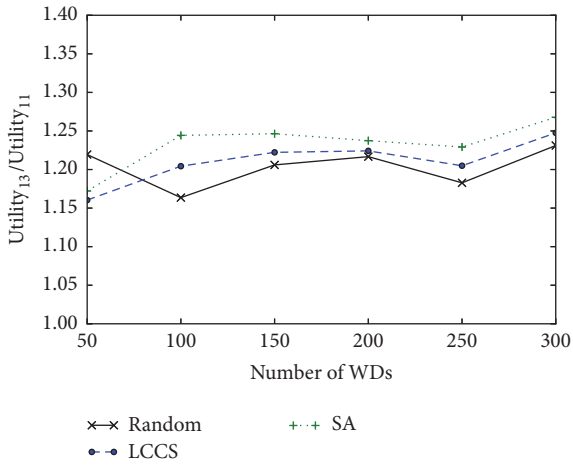


Figure 10: Comparison between 11 and 13 channels.

channels partially overlap between them, raising the problem of cochannel interference. Many works devoted to channel assignment in Wi-Fi networks avoid the use of overlapped channels by using a reduced number of noncolliding channels, called orthogonal channels. In this paper we evaluate the goodness of using only the orthogonal channels instead of a higher number of overlapping channels up to the maximum number of channels available. Moreover, we evaluate our proposal in realistic settings. Although we also consider some usual heuristic channel assignments algorithms, we also use an optimization algorithm based on the simulated annealing technique. Due to its complexity, the Wi-Fi channel assignment problem has been very scarcely addressed as an optimization procedure, being most of the proposals heuristic approaches.

The first study performed in the paper is related to the impact of the number of channels available when we do not consider the cochannel interference. Although this situation is unrealistic, it is included in order to compare its results to the situations where we consider cochannel interference. Results show that, increasing the number of nonoverlapping channels, we improve performance as we have more opportunities to choose a channel that avoids interferences with the channels used in other access points.

Secondly and introducing cochannel interference, we have also studied the effect of the number of channels available. In this comparison we have first determined that the best performance is obtained by the optimizer based on simulated annealing, followed by the heuristic LCCS and, finally, the random assignment, which is useful mainly as a baseline. Furthermore, we conclude that the optimizer is not able to achieve better channel assignments when it is permitted to use more channels than just the orthogonal ones. In fact, it is shown that the algorithm mainly uses the three orthogonal channels even when it is allowed to use the whole spectrum. It is particularly significant the fact that LCCS (the *de facto* standard for Wi-Fi channel assignment) degrades as more channels are available to choose, probably due to the fact that local interference evaluation avoids channel choices which would be necessary for the globally optimal solutions (we could talk of "sacrificing local utility for the global welfare"). This leads us to think that better heuristics could be designed taking into account the aggregation of utilities further than at the access point level (e.g., considering clusters of APs). The design of such heuristics and the study of the conditions in which they succeed will be the scope of our most immediate future work. Finally, we would like to introduce dynamicity in our model, to see how the different approaches react to changes during and after convergence and the utility loss associated with the cost of (re)convergence.

Data Availability

The layout of the scenarios under study can be found on the following public URL: http://www.ubu.es/servicio-ide-informatica-y-comunicaciones/catalogo-de-servicios/red-wifi-vpn/wi-fi/planos-de-cobertura-de-la-red-wifi. Regarding the reproducibility of the results, the authors have included in the paper all the information needed to replicate them.

However, and upon a request, they can send the resulting channel assignments obtained in the paper, which constitute the results of the experiments of the paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work has been supported by the Spanish Ministry of Economy, Industry and Competitiveness Grants TIN2016-80622-P (AEI/FEDER, UE). David Orden has been partially supported by MINECO Project MTM2014-54207, Spanish Ministry of Science Project MTM2017-83750-P (AEI/FEDER, UE), and European Union H2020-MSCA-RISE Project 734922-CONNECT. Authors would like to thank the University of Burgos for granting them access to the wireless infrastructure layout and heat map used in the paper.

References

[1] S. Chieochan, E. Hossain, and J. Diamond, "Channel assignment schemes for infrastructure-based 802.11 WLANs: a survey," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 1, pp. 124–136, 2010.

[2] M. Drieberg, F.-C. Zheng, and R. Ahmad, "MICPA: A client-assisted channel assignment scheme for throughput enhancement in WLANs," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 9, pp. 4497–4508, 2011.

[3] X. Yue, C.-F. Wong, and S.-H. G. Chan, "CACAO: Distributed client-assisted channel assignment optimization for uncoordinated WLANs," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 9, pp. 1433–1440, 2011.

[4] D. J. Leith, P. Clifford, V. Badarla, and D. Malone, "WLAN channel selection without communication," *Computer Networks*, vol. 56, no. 4, pp. 1424–1441, 2012.

[5] D. Gong, M. Zhao, and Y. Yang, "Channel assignment in multi-rate 802.11n WLANs," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '13)*, pp. 392–397, IEEE, Shanghai, China, 2013.

[6] A. Bhartia, D. Chakrabarty, K. Chintalapudi, L. Qiu, B. Radunovic, and R. Ramjee, "IQ-hopping: Distributed oblivious channel selection for Wireless Networks," in *Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 81–90, Germany, 2016.

[7] Y. M. Kwon, K. Choi, M. Kim, and M. Y. Chung, "Distributed channel selection scheme based on the number of interfering stations in WLAN," *Ad Hoc Networks*, vol. 39, pp. 45–55, 2015.

[8] H. Kasasbeh, F. Wang, L. Cao, and R. Viswanathan, "Generous throughput oriented channel assignment for infra-structured WiFi networks," in *Proceedings of the 2017 IEEE Wireless Communications and Networking Conference, WCNC 2017*, USA, March 2017.

[9] Y. Lee, K. Kim, and Y. Choi, "Optimization of AP placement and channel assignment in wireless LANs," in *Proceedings of the 27th Annual IEEE Conference on Local Computer Networks, LCN 2002*, pp. 831–836, USA, November 2002.

[10] J. K. Chen, G. De Veciana, and T. S. Rappaport, "Improved measurement-based frequency allocation algorithms for wireless networks," in *Proceedings of the 50th Annual IEEE Global Telecommunications Conference, GLOBECOM 2007*, pp. 4790–4795, USA, November 2007.

[11] M. Elwekeil, M. Alghoniemy, M. El-Khamy, H. Furukawa, and O. Muta, "Optimal channel assignment for IEEE 802.11 Multi-cell WLANs," in *Proceedings of the 20th European Signal Processing Conference, EUSIPCO 2012*, pp. 694–698, Romania, August 2012.

[12] M. Seyedebrahimi, F. Bouhafs, A. Raschella, M. MacKay, and Q. Shi, "SDN-based channel assignment algorithm for interference management in dense Wi-Fi networks," in *Proceedings of the 2016 European Conference on Networks and Communications, EUCNC 2016*, pp. 128–132, Greece, June 2016.

[13] P. Mahonen, J. Riihijarvi, and M. Petrova, "Automatic channel allocation for small wireless local area networks using graph colouring algorithm approach," in *Proceedings of the 2004 IEEE 15th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 536–539, Barcelona, Spain.

[14] A. Mishra, S. Banerjee, and W. Arbaugh, "Weighted coloring based channel assignment for WLANs," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 9, no. 3, pp. 19–31, 2005.

[15] A. Mishra, V. Brik, S. Banerjee, A. Srinivasan, and W. Arbaugh, "A client-driven approach for channel management in wireless LANs," in *Proceedings of the INFOCOM 2006: 25th IEEE International Conference on Computer Communications*, Spain, April 2006.

[16] M. Haidar, R. Akl, H. Al-Rizzo, and Y. Chan, "Channel assignment and load distribution in a power-managed WLAN," in *Proceedings of the 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC'07*, Greece, September 2007.

[17] R. Akl and A. Arepally, "Dynamic Channel Assignment in IEEE 802.11 Networks," in *Proceedings of the 1st IEEE International Conference on Portable Information Devices, (PIDs '07)*, pp. 1–5, USA, March 2007.

[18] J. Wang, W. Shi, K. Cui, F. Jin, and Y. Li, "Partially overlapped channel assignment for multi-channel multi-radio wireless mesh networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, p. 25, 2015.

[19] M. Abusubaih, "Using Partially Overlapping Channels in Home 802.11g WLANs," *Wireless Personal Communications*, vol. 88, no. 2, pp. 295–303, 2016.

[20] M. Doering, L. Budzisz, D. Willkomm, and A. Wolisz, "About the practicality of using partially overlapping channels in IEEE 802.11 b/g networks," in *Proceedings of the 2013 IEEE International Conference on Communications, ICC 2013*, pp. 5110–5114, Hungary, June 2013.

[21] Y. Cui, W. Li, and X. Cheng, "Partially overlapping channel assignment based on "node orthogonality" for 802.11 wireless networks," in *Proceedings of the IEEE INFOCOM 2011*, pp. 361–365, April 2011.

[22] E. De La Hoz, J. M. Gimenez-Guzman, I. Marsa-Maestre, and D. Orden, "Automated negotiation for resource assignment in wireless surveillance sensor networks," *Sensors*, vol. 15, no. 11, pp. 29547–29568, 2015.

[23] E. De La, I. Hoz, J. M. Gimenez-Guzman, D. Orden, and M. Klein, "Multi-agent nonlinear negotiation for Wi-Fi channel assignment," in *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pp. 1035–1043, 2017.

[24] D. Orden, J. Gimenez-Guzman, I. Marsa-Maestre, and E. de la Hoz, "Spectrum Graph Coloring and Applications to Wi-Fi Channel Assignment," *Symmetry*, vol. 10, no. 3, p. 65, 2018.

[25] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *Journal of Complex Networks*, vol. 2, no. 3, pp. 203–271, 2014.

[26] D. B. Green and M. S. Obaidat, "An accurate line of sight propagation performance model for Ad-Hoc 802.11 wireless LAN (WLAN) devices," *IEEE International Conference on Communications*, vol. 5, pp. 3424–3428, 2002.

[27] K. R. Chowdhury and I. F. Akyildiz, "Cognitive Wireless Mesh Networks with dynamic spectrum access," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 168–181, 2008.

[28] A. Mishra, V. Shrivastava, S. Banerjee, and W. Arbaugh, "Partially overlapped channels not considered harmful," *ACM SIGMETRICS Performance Evaluation Review*, vol. 34, no. 1, pp. 63–74, 2006.

[29] V. Shrivastava, S. Rayanchu, J. Yoon, and S. Banerjee, "802.11n under the microscope," in *Proceedings of the Internet Measurement Conference 2008, IMC'08*, pp. 105–110, Greece, October 2008.

[30] A. Bazzi, "On uncoordinated multi user multi RAT combining," in *Proceedings of the IEEE 74th Vehicular Technology Conference, VTC Fall 2011*, 6 pages, IEEE, USA, September 2011.

[31] M. Achanta, "Method and apparatus for least congested channel scan for wireless access points," *uS Patent App. 10/959*, p. 446, 2006.

[32] T. Ito, M. Klein, and H. Hattori, "A multi-issue negotiation protocol among agents with nonlinear utility functions," *Multiagent and Grid Systems*, vol. 4, no. 1, pp. 67–83, 2008.

[33] J. L. Gross, J. Yellen, and P. Zhang, *Handbook of Graph Theory*, CRC Press, Boca Raton, Fla, USA, 2nd edition, 2013.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 8319749, 10 pages
https://doi.org/10.1155/2018/8319749

Research Article

Optimization of Cell Size in Ultra-Dense Networks with Multiattribute User Types and Different Frequency Bands

Yiqiao Wei and Seung-Hoon Hwang

Division of Electronics and Electrical Engineering, Dongguk University, Seoul, Republic of Korea

Correspondence should be addressed to Seung-Hoon Hwang; shwang@dongguk.edu

Received 25 May 2018; Revised 17 September 2018; Accepted 25 September 2018; Published 18 October 2018

Guest Editor: Ivan Marsa-Maestre

Ultra-dense cellular networks (UDNs) represent the trend for 5G networks in dense urban environments. With the aim of exploring the optimal extent of network densification under different performance requirements and the trade-off between the network capacity and deployment cost in UDNs, a multiple-objective optimization model is proposed. This novel optimization design consists of a multiattribute user type in which users are grouped based on their propagation conditions and an infinitesimal dividing modeling method termed the ring method for network capacity dimensioning. The optimal cell size is estimated to maximize the total network capacity and minimize the deployment cost under different levels of user capacity demand. Additionally, the corresponding total network capacity and the required number of base stations are presented. Furthermore, two conventional frequency bands, 800 MHz and 1.8 GHz, and two new bands, 3.5 GHz and mmWave 28 GHz, are considered to investigate their feasibility and the potential of higher frequency bands in the 5G network.

1. Introduction

Existing cellular networks are being challenged by the explosive growth of traffic demand from a variety of mobile and internet services such as high definition live video streaming and mobile online gaming. This challenge will become more severe with emerging new services such as 3D multimedia, augmented reality, and virtual reality. Because these new services will progressively be delivered over wireless communication, the mobile and wireless traffic volume has been predicted to increase 1000-fold in the near future [1]. More specifically, the area throughput in some densely populated areas may reach tens of Tb/s/km^2 and the data rate experienced by users might exceed 1 Gbps [2]. To improve system capacity and meet the expected demand, better wireless modulation and additional spectrum bandwidth can be employed. However, the potential gain seems to be limited. Therefore, the most efficient way to increase cell density is by reducing coverage [3].

Motivated by the above, the ultra-dense cellular network (UDN) is emerging as a promising technology with new characteristics for fifth-generation (5G) networks [3–5]. The network evolved from the traditional macro-cell-only Homogenous Network (HomNet) to a multiband Heterogeneous Network (HetNet) where macro cells operating at low frequency bands underlie small cells operating at high frequencies. The UDN can be seen as another evolution from HetNet with further densification of small cells [5]. The basic idea of this new paradigm is to shorten the distance between the access node and end user as much as possible and thus densely deploy small cells in crowded districts or hotspots where enormous data traffic is generated. The average intersite distance (ISD) for UDNs is reduced to around or less than 100 m in contrast to the 400 m distance in the traditional 4th-generation (4G) deployment [6]. The UDN can be defined as small cell deployment in dense urban scenarios where the active user density is high, with about 600 active users per km^2 [5–7]. Together with higher frequency bands where much wider bandwidth is available, UDNs are expected to fulfill very high demands on system capacity and achievable end-user data rates [8]. Currently, cellular systems mainly operate at frequency bands below 3 GHz. However,

higher frequencies up to millimeter-wave (mmWave) are considered candidates for 5G deployment [9].

The strategy of operating UDNs on new and higher frequency bands brings new and multiple challenges for network operators in network planning and actual deployment, including network architecture, resource management, difficult handover control, and interference management [3]. Other specific challenges this paper focuses on are as follows.

(i) One challenge is to find the fundamental limits in network densification, i.e., to what extent the cell size can be reduced. Network densification cannot be continued since too close a distance between cells will generate high interference [5]. As discussed in [4], this is considered a key question for future network designs

(ii) Another is to minimize the cost of the network design. There is always an inherent trade-off between the cost and other objectives such as capacity and quality of service (QoS) [10]. This is more significant in network densification since the cost will increase greatly as more cells are needed

(iii) There are various propagation conditions for massive numbers of users in the urban environment. Line-of-sight (LOS) components become more probable when users are closer to the base station (BS). Therefore, both LOS and non-line-of-sight (NLOS) should be considered in the study of dense networks [5]. The target for UDNs is usually dense urban areas where users are often located in buildings and above the ground. It is important to emphasize indoor usage because up to 80% of data traffic is generated in indoor environments [11]. Therefore, it is necessary but difficult to take all these factors into consideration in UDN network research

(iv) Another challenge is to further study the feasibility of high frequency bands in dense urban scenarios. As argued in [12], the value and feasibility of higher frequency bands need to be further considered. This is because, at lower and higher frequency bands, each spectrum has its own pros and cons in urban environments. At lower frequencies, the signal has less path and penetration loss. Meanwhile, the interference from adjacent cells is higher. On the other hand, at higher frequencies, the available system bandwidth is increased, and the interference is decreased. However, more signal loss occurs

With the aims of exploring the proper extent of cell densification and investigating the trade-off between capacity and deployment cost for UDNs, this work proposes an innovative optimization model to determine the cell size in UDNs with multiple objectives including maximizing network capacity, minimizing the deployment cost, and guaranteeing each user's capacity demand. Multiple factors are taken into consideration in the system model such as various user propagation conditions, 3D urban environments, and two urban network deployment scenarios. Furthermore, four frequency bands are employed for the carrier frequency to investigate their pros and cons in UDN deployment. The paper is organized as follows: In Section 2, we review the related literature. The system model is introduced in Section 3 and the optimization formulation is described in Section 4. Based on the simulation, numerical results are presented in Section 4. Finally, our findings and further works are described in Section 5 along with the conclusion.

2. Related Works

Since it evolved from multiband HetNet, 5G UDN jointly consists of small cells and macro cells. However, the macro cells will probably be configured only to transmit control data to solve user handover problems in the small cells while the small cells will be in charge of high speed user data transmission [4]. Many advanced handover technologies have been proposed for UDNs. A software defined network-enabled authentication handover system was proposed in [13] that involves user-specific context information sharing and privacy protection using multiple network paths to transmit data. These two solutions have been proven to simplify the handover by reducing latency and enhancing privacy protection in 5G networks. In [14], measurement-based mmWave dynamic channel models were innovatively employed to study the handover in mmWave systems. With extensive simulation, the proposed dual connectivity framework, which allows fast switching between LTE and mmWave radio access as well as secondary cell handover across mmWave eNodeBs (eNBs), was demonstrated to improve the latency and throughput stability of mmWave systems.

On the other hand, many studies place emphasis on the user throughput and network capacity of UDNs since this is the motivation for network densification. Aiming at bringing detailed analysis of user throughput and network capacity of UDN, a network model with various configurations including different ISDs, user equipment (UE) densities, and frequency bands was used in [15] to reveal the gain of network densification. The simulation results showed that dense cell deployment with an ISD of 35 m can increase the average UE throughput by more than 7.56 times and that using a 10 GHz band with a bandwidth of 500 MHz can further increase the network capacity up to 5-fold.

However, finding the optimal site position or configuration for the purpose of planning objectives like coverage, capacity, and deployment cost requires further research such as network planning or cell optimization. Since these objectives are always interdependent and interrelated [10], joint optimization of multiple objectives is an important research topic in UDN deployment for further investigation [5]. A novel Cognitive Radio Network- (CRN-) applied 5G UDN architecture was designed in [16] and a graph-based algorithm and genetic-based algorithm were used to maximize the user throughput and minimize communication interference. The results showed that the proposed optimization algorithms can effectively improve network performance in terms of throughput and signal-to-interference-plus-noise ratio (SINR). A novel optimization design for UDNs that involves dividing massive numbers of users into groups based on moving speed and selecting suitable subnets was proposed

in [17]. Optimization is then performed to coordinate suitable UDN subnets for each group to meet huge service demands with the minimum quantity of resources. The numerical results showed that the proposed approach can make the best use of bandwidth resources to provide services that meet user demands. With the purpose of balancing the trade-off between energy efficiency (EE) and spectrum efficiency (SE), an improved version of the nondominated sorting genetic algorithm-II (NSGA-II) [18] intelligent approach was proposed in [19] to optimize the performance of EE and SE by jointly allocating transmission power and resource blocks to users in UDNs. This was the first time NSGA-II was applied in a UDN for such optimization and the novel improved version has several advantages including fast nondominated sorting and less complexity. All the aforementioned literature details optimization cell planning for 5G UDN but the objective and methods are distinctive. The distinctive contributions and innovations of the proposed system model in this paper are summarized as follows.

(i) Instead of optimizing either the network capacity or deployment cost, this work is aimed at balancing the trade-off between both parameters for dense small cells

(ii) The optimal solutions, i.e., optimal ISDs within the constraint of different user capacity demands, provide insights to determine the ideal extent of network densification

(iii) A novel method involving multiattribute user types is utilized in the proposed system model. Instead of grouping a large number of users based on single user attributes, like user mobility in [17], this method is designed to cover all possible propagation conditions for users including LOS or NLOS, outdoor or indoor, and UE heights. Other works [15–17, 19] consider only individual or limited combinations of these possible propagation conditions in real urban environments. This method is described in more detail in Section 3

(iv) An innovative infinitesimal dividing modeling method termed the ring method is proposed in this work for network capacity dimensioning and optimization. Compared to the summation of the capacity of all users in the network in [15, 16, 19], the ring method is specially designed to properly derive the number of users in each group since the LOS probability is assumed to be a function of the distance between the user and the BSs. Detailed analysis of the proposed ring method is presented in Section 4

(v) Compared to other works, several frequency bands, specifically two conventional frequency bands (800 MHz and 1800 MHz) and two 5G frequency bands (3.5 GHz and 28 GHz), are investigated in both urban macro (UMa) and urban micro- (UMi-) street canyon deployment scenarios. Although multiple frequency bands can be deployed in the two 5G frequency bands, small cells using high frequency bands will be in charge of high speed user data transmission and macro cells at low frequency bands will probably be configured only to transmit control data regarding handovers [4]. Therefore, the two 5G frequency bands are the main targets and the two conventional bands are also investigated for comparison and to demonstrate they may be not competent in the task of transmitting user data in UDNs

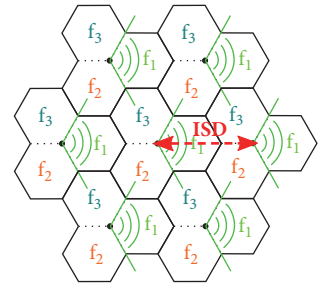


Figure 1: Cell layout of network.

3. System Model

We consider the deployment of a UDN cellular network in a dense urban area of 1 km^2. The goal is to maximize the total network throughput and minimize the deployment cost under certain user capacity constraints. The system model is as follows.

3.1. Cell Layout. Figure 1 details the cell layout in this work. The hexagonal grid three-sector site described in [20] is used. A frequency reuse number of three is assumed where the frequencies used in three sectors of each cell are different. Due to the above, the user only receives interference signals from three BSs when the interference is only considered on the first tier. Generally, there are two kinds of small cell BSs, namely, fully functioning BSs (picocells and femtocells) and macro-extension access points [5]. Here, picocells installed outdoors are considered the BSs in the network.

3.2. User Distribution and Multiattribute User Types. The users are assumed to be uniformly distributed in the target area, which means that the number of active users equals the product of active user density and area. In reality, users in urban areas are always in different propagation conditions such as LOS or NLOS and outdoor or indoor. As shown in Table 1, the users are classified into four types by using different propagation conditions as multiple attributes. The detailed conditions for each user type are presented in Figure 2. The specific number of users of each type is calculated according to the indoor user ratio and LOS probability in [20].

3.3. Carrier Frequency and System Bandwidth. This work investigates four carrier frequencies, namely, 800 MHz, 1800 MHz, 3.5 GHz, and 28 GHz. 800 MHz was popular for cellular network deployment before the 3G. Frequency bands below 3 GHz such as 1.8 GHz were introduced in order to employ 3G and 4G mobile communication services. For the 5G, higher frequency bands up to mmWave bands are expected

Table 1: *Multiattribute user types* in urban scenario.

User type	User condition
Type 1	LOS + outdoor
Type 2	LOS + indoor
Type 3	NLOS + outdoor
Type 4	NLOS + indoor

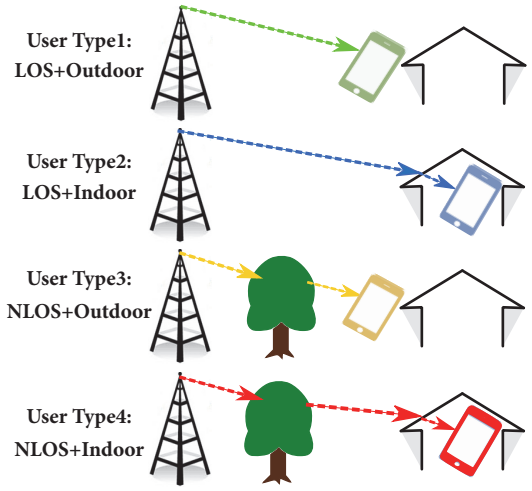


Figure 2: Diagram for four user types.

to provide more bandwidth resources to fulfill the anticipated capacity demand. Among them, 3.5 GHz and 28 GHz are two important candidate bands because a large portion of spectrum resources is available at 3.5 GHz below 6 GHz and at 28 GHz above 6 GHz [9]. Regarding the system bandwidth, 5% of the carrier frequency is assumed. It is reasonable and necessary to consider different bandwidths for different carrier frequencies, since much larger available bandwidth is the main motivation for the consideration of high frequency bands in 5G.

3.4. Radio Propagation Model. Empirical propagation models are broadly used to predict the path loss in feasibility studies and initial deployment in cellular network planning. Conventional propagation models like the Hata and Stanford University Interim (SUI) models used in another of our research works [21] are only applicable to low frequency bands and lack the parameters to simulate an urban environment. Therefore, the 3D propagation model from 3GPP [20] is used. This model supports frequencies ranging from 500 MHz to 100 GHz in urban scenarios and provides diverse parameters such as 3D propagation distance, environment height, standard shadow fading, LOS probability, indoor user rationing, and outdoor-to-indoor penetration loss. These parameters are critical for simulation of the proposed model with multiattribute user types based on user propagation conditions. The urban scenario is further divided into two subscenarios in [20]: UMa and UMi-street canyon. In the UMa urban scenario, the BS antennas are mounted above the rooftop level of surrounding buildings. In contrast, the antennas are below the building rooftops in the UMi-street canyon scenarios. Another difference is that the ISD for UMa is larger than that for the UMi-street canyon. These two scenarios are, respectively, investigated in this work.

4. Problem Formulation and Optimization

4.1. Problem Formulation Using the Ring Method. For UDN deployment, the key challenge is to provide top quality services while minimizing network cost. Focusing on this issue, the goal of this work is to maximize the total network capacity while keeping the deployment cost as low as possible. The total cost is roughly proportional to the number of deployed small cells in the network. Therefore, the total cost of the network is measured as the number of BSs that need to be deployed in the network. The constraint is to guarantee that every user is satisfied with a specific capacity demand, which means the capacity of every user should be higher than or equal to the demand bound constraint. Therefore, it becomes a multiple-objective optimization problem which can be formulated as

$$\begin{aligned} &\text{Maximize } \{C_{network}\}, \\ &\text{Minimize } \{\mathrm{N_{BS}}\}, \end{aligned} \tag{1}$$

within the constraints

$$C_{user} \geq C_0, \tag{2}$$

where $C_{network}$ is the total capacity of the network, C_0 is the capacity demand bound constraint, and N_{BS} is the number of required BSs:

$$N_{BS} = \frac{1000^2}{S_{cell}} = \frac{1000^2}{3 * \left(2.6 * (ISD/3)^2\right)}, \tag{3}$$

where S_{cell} is the area of one cell and is determined by the ISD (m).

Now the main challenge is how to quantify the total network capacity $C_{network}$. An infinitesimal dividing modeling method termed the ring method is proposed in this work. As Figure 3 shows, one cell is divided into many small parts and the width of each part is 1 m. d_i is the distance between the inflection point (marked as a dot in Figure 3) and the BS. Because 1 m is quite small compared to the cell coverage, each part can be seen as a circular arc, which is referred to as a ring here. All users in any single ring $Ring_i$ can be assumed to have the same distance d_i to the BS. Four types of users are distributed in $Ring_i$, according to the LOS probability and indoor user ratio.

Figure 4 is the flowchart for the method used to derive the total network capacity $C_{network}$. The detailed process is explained as follows. As defined in [22], the minimal distance between the UE and picocell BS is 2 m. Therefore, the area of $Ring_i$ can be expressed as

$$S_i = \begin{cases} \sqrt{3}d_i + \dfrac{\sqrt{3}}{2}, & 2 \leq d_i \leq \dfrac{ISD}{3} - 1 \\ \dfrac{\sqrt{3}}{3}ISD, & \dfrac{ISD}{3} - 1 < d_i \leq \dfrac{2}{3}ISD - 1. \end{cases} \tag{4}$$

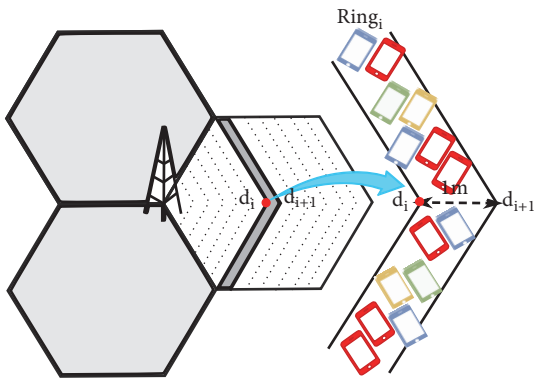


FIGURE 3: An illustrative diagram of the ring method.

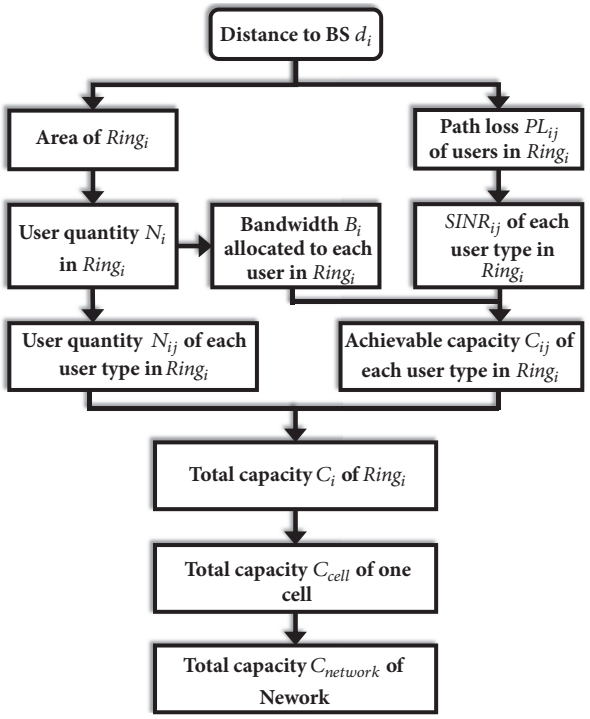


FIGURE 4: Flowchart for the derivation of total network capacity.

Since the users are uniformly distributed in the cell, the number of active users N_i in $Ring_i$ equals the product of active user density ρ_{user} and ring area S_i:

$$N_i = \rho_{user} * \mathrm{S_i}. \tag{5}$$

It is assumed that bandwidth is uniformly allocated to every user in the cell so the bandwidth B_i allocated to each user equals

$$B_i = \frac{0.05 * f_C}{N_{reuse} * N_{cell}} = \frac{0.05 * f_C}{N_{reuse} * (\rho_{user} * S_{cell})}, \tag{6}$$

where f_C is the carrier frequency, N_{reuse} is the frequency reuse number which equals three here, and N_{cell} is the number of users in one cell. Ignoring feeder loss, the received power P_{ij} of UE type j in $Ring_i$ can be expressed as

$$P_{ij} = \begin{cases} P_T + G_T + G_R - PL_{ij}(d_i) - SF, & \text{outdoor UE} \\ P_T + G_T + G_R - PL(d_i) - L_{O2I}(f_c) - SF, & \text{indoor UE,} \end{cases} \tag{7}$$

where P_T is the transmitter power of the base station antenna, G_T is the transmitter antenna gain, G_R is the receiver antenna gain, and SF is shadow fading. The path loss, PL_{ij}, and the outdoor-to-indoor penetration loss, L_{O2I}, are calculated using the 3D path loss model and the low-loss out to indoor penetration model in Technical Report 38.901 by 3GPP, respectively [20]. Note that the indoor user is assumed to be standing by the wall or window where there is no inside loss after signal penetration through the wall or window. In Figure 2, the interference I_{ij} of UE type j in $Ring_i$ is the sum of the power of three received interference signals transmitted by the adjacent three BSs:

$$I_{ij} = \begin{cases} \sum_{k=1}^{3} \left\{P_T + G_T + G_R - PL_{ijn} - SF\right\}, & \text{outdoor UE} \\ \sum_{1}^{3} \left\{P_T + G_T + G_R - PL_{ijn} - SF - L_{O2I}(f_C)\right\}, & \text{indoor UE,} \end{cases} \tag{8}$$

where PL_{ijn} is the path loss of the interference signal transmitted by interference BS n and is also calculated with the 3D path loss model. After deriving the received signal power P_{ij} and the interference I_{ij} power, the $SINR_{ij}$ of UE type j in $Ring_i$ can be achieved by

$$SINR_{ij} = \frac{P_{ij}}{Noise + I_{ij}}. \tag{9}$$

Then C_{ij}, the user maximum achievable capacity of UE type j in $Ring_i$, can be calculated using the Shannon capacity:

$$C_{ij} = B_i * \log_2(1 + SINR). \tag{10}$$

As defined in [20], the indoor user ratio is 80% in urban areas and the LOS probability is a function of the distance d_i. Then, N_{ij}, the user quantity of UE type j in $Ring_i$, can be calculated as

$$N_{ij} = \begin{cases} 0.2N_i * P_{LOS}(d_i), & j = 1 \\ 0.8N_i * P_{LOS}(d_i), & j = 2 \\ 0.2N_i * (1 - P_{LOS}(d_i)), & j = 3 \\ 0.8N_i * (1 - P_{LOS}(d_i)), & j = 4. \end{cases} \tag{11}$$

Note that when N_{ij} is smaller than 1, it represents the probability that the user is in *Ringi* and under type j. Combining (10) and (11), the total capacity of $Ring_i$ equals the sum of the capacities for all users:

$$C_i = \sum_{j=1}^{4} C_{ij} * N_{ij}. \tag{12}$$

The sum of the total capacity of all rings is the total capacity of one cell C_{cell}:

$$C_{cell} = \sum_{i=1}^{(2/3)ISD-2} C_i. \quad (13)$$

Finally, the total capacity of the network is

$$C_{network} = C_{cell} * N_{cell}, \quad (14)$$

where the number of cells in the network N_{cell} is determined by the optimization ISD result. At this point, optimization objective equation (1) and constraint equation (2) are ready for the following optimization procedure.

4.2. Optimization Procedure. The goal of this work is to maximize the total network capacity while keeping the deployment cost as low as possible within the constraint that every user is satisfied with a specific capacity demand. In this optimization problem, ISD is the only variable and the two objectives are maximizing the total network capacity and minimizing the number of deployed small cells. It is difficult to handle small cell planning for UDNs by considering multiple-objective functions at the same time and within the constraints. The genetic optimization algorithm is introduced to solve this multiple-objective optimization problem and derive the optimal ISD. The genetic algorithm is a type of optimization algorithm that imitates the biological processes of reproduction and natural selection to solve for the target function solutions [23]. It is widely used in wireless communication [16, 19], and results in [16] indicate that the genetic algorithm outperforms other network planning algorithms like graph theory. The NSGA-II [18] is exoteric non-domination-based and is arguably the most famous genetic algorithm for optimizing multiobjectives. With the aim of a less bug-prone and convenient implementation, we use the *gamultiobj* function in the MATLAB global optimization toolbox. It uses a controlled, elitist genetic algorithm, which is a variant of the NSGA-II. The *gamultiobj* function can be expressed as

$$X = gamultiobj\left(f_{\text{fitness}}, n_{\text{VARS}}, LB, UB, f_{\text{constraint}}\right), \quad (15)$$

where X is a set of optimization results, i.e., optimal ISDs. Here, the fitness function $f_{fitness}$ is represented by (1) and the constraint function $f_{constraint}$ by (2), n_{VAS} is the dimension of the optimization problem (equal to one here since ISD is the only variable), and LB and UB are the lower and upper bounds of the variable ISD, respectively.

During the optimization procedure, *gamultiobj* first randomly creates the initial population, which is a set of individuals (variable ISDs), with respect to the bounds. Then *gamultiobj* evaluates the objective function $f_{fitness}$ and constraint $f_{constraint}$ for the population and uses those values to create scores for the population. Next, the individuals with the best scores are selected as parent individuals to create children individuals of the next generation by mutation and crossover. This iteration starts over until a termination criterion applies, such as when the average relative change in the best fitness function value is less than or equal to a predefined *FunctionTolerance* or when the generation number exceeds the given *MaxGenerations*. At this point, *gamultiobj* will output the optimized ISD.

Table 2: Simulation parameters [15, 20, 22, 24, 25].

Parameters	UMa	UMi-street canyon
Carrier frequency f_C	800 MHz 1.8 GHz 3.5 GHz 28 GHz	
System bandwidth	5% * f_C	
TX antenna height (m)	25	10
Outdoor user RX antenna height (m)	1.5	1.5
Indoor user RX antenna height (m)	9	9
TX antenna gain, G_T (dBi)	5	5
RX antenna gain, G_R (dBi)	2	2
BS TX power, P_T (dBm)	30	30
Thermal noise density (dBm/Hz)	-174	-174
Effective environment height (m)	1	1
Active user density d_{user} (users/km2)	600	600
Indoor user ratio	0.8	0.8
Minimum distance between UE and BS (m)	2	2
Shadow fading in LOS (dB)	4	4
Shadow fading in NLOS (dB)	6	7.82

5. Simulation Results

Four carrier frequency bands are investigated and the available system bandwidth is set to 5% of the carrier frequency. For each carrier frequency, the maximum achievable capacity guaranteed for every user will be limited. In the simulation, the optimization algorithms cannot provide the right result when the capacity demand bound constraint C_0 reaches its limit. Due to this, the optimal ISD is derived against C_0 varying from 1 Mbps to the maximum value. The initial population in MATLAB *gamultiobj* is randomly created so the optimization result for each run can be slightly different. The results presented in this section are the average values from 10 runs. The simulation parameters including the picocell BS antennas and 3D propagation model are summarized in Table 2. Table 3 lists the *gamultiobj* parameters which are tuned for proper functioning and relative stable output results. The other *gamultiobj* parameters all use the provided default values.

The results for the optimal ISD against the capacity demand bound constraint are presented in Figure 5. The corresponding total network capacity and the number of BSs needed are shown in Figures 6 and 7, respectively. In Figure 5, the optimal ISD decreases as the user capacity

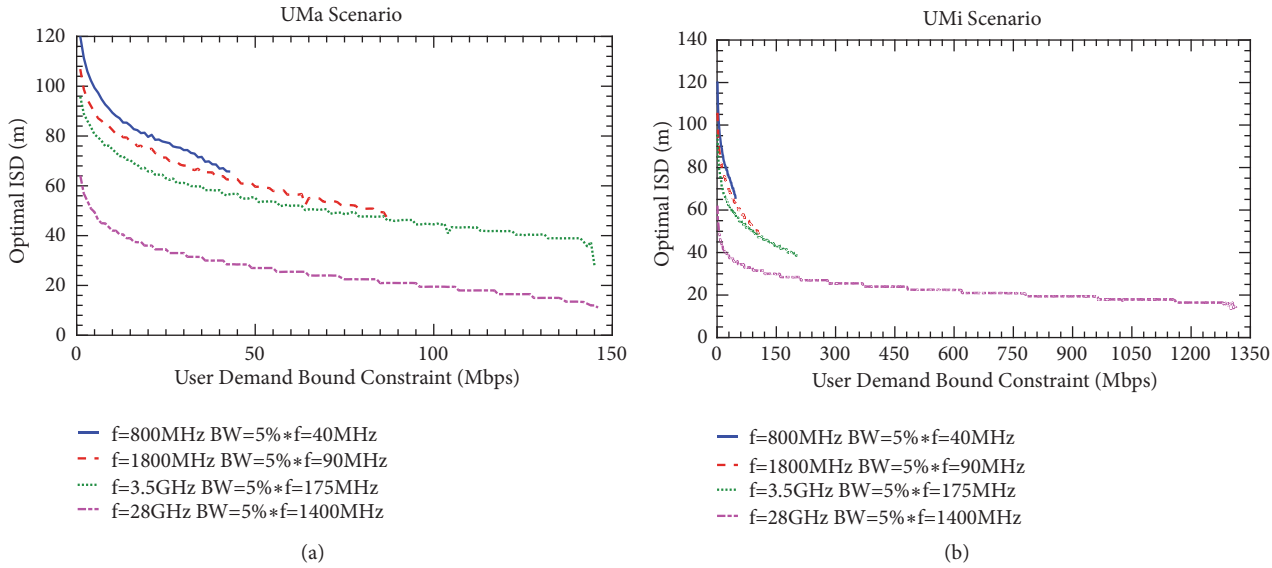


Figure 5: Optimal ISD vs. user demand bound constraint. (a) UMa scenario; (b) UMi-street canyon scenario.

Table 3: Parameters used in MATLAB *gamultiobj* function.

Parameters	Value
LB	5 (m)
UB	500 (m)
FunctionTolerance	1e-6
MaxGenerations	300
PopulationSize	100
ParetoFraction	0.3

demand increases. In smaller cells, the path losses for users are smaller and the bandwidth allocated to each user becomes larger since fewer users exist in the cell. For the same user demand bound, the higher the frequency is, the smaller the optimal cell size becomes. Because higher frequencies suffer severe path and penetration loss, the cell size needs to be reduced to provide sufficient SINR to the cell edge and indoor users. In the UMa scenario, the maximum capacity which every user can be satisfied with is 43 Mbps for a frequency of 800 MHz and ISD of 65 m and 87 Mbps for a frequency of 1.8 GHz and ISD of 47 m. For 3.5 GHz, up to 145 Mbps can be guaranteed when the ISD is reduced to 30 m.

There are two reasons why higher frequencies can support higher user demand. One is that higher frequencies may bring about lower intercell interference. Another reason is that the achievable capacity is determined by bandwidth as well as SINR. For the same SINR, a smaller bandwidth of 800 MHz and 1.8 GHz limits the maximum achievable capacity. Meanwhile, 3.5 GHz can provide much wider bandwidth, which can support much higher user demand. However, note that 28 GHz can guarantee almost the same maximum achievable capacity as 3.5 GHz despite the enormous bandwidth. This is because the propagation loss for 28 GHz is so high in the UMa scenario that even the huge bandwidth cannot compensate for it. On the other hand, the bandwidth of 28 GHz is shown to be advantageous in the UMi-street canyon scenario as every user can be guaranteed up to 1.3 Gbps when ISD is decreased to 10 m. Additionally, 3.5 GHz yields 50 Mbps higher guaranteed user capacity in the UMi-street canyon scenario. The difference between the two scenarios results from the 15 m higher BS antenna height in the UMa scenario compared to the UMi-street canyon scenario. In the 3D propagation model, the propagation distance has three dimensions, which means a higher antenna can provide longer propagation distance for the same horizontal distance between BS and UE. That is, the BS antenna is much closer to users in the UMi-street canyon scenario. The primary disadvantage of high frequency bands, i.e., severe propagation loss, has a lighter effect on user achievable capacity. Therefore, the achievable capacity can be increased with the wider bandwidth of higher frequency bands.

Figure 6 shows the total network capacity versus user demand bound constraints. Generally, the network capacity keeps growing but reaches a limit as the guaranteed user demand increases. In Figure 6(a), the total network capacity for 28 GHz reaches the peak rapidly and then starts to decrease. The network needs to be densely positioned to provide higher capacity to users at the cell edges or indoors. However, the total network capacity cannot increase without limit because the number of active users in the network is limited. Therefore, some cells may not include active users when cells are extremely dense. In addition, when network densification surpasses a certain threshold, some users may receive more severe interference from the interfering BSs. When conventional bands such as 800 MHz and 1.8 GHz are employed, the maximum network capacity is limited below 12 Tbps/km^2 in UMa and 20 Tbps/km^2 in the UMi scenario. New bands at higher frequencies have larger total network capacity. For 3.5 GHz, the total network capacity can reach up to 50 Tbps/km^2 in the UMi scenario. For 28 GHz, the

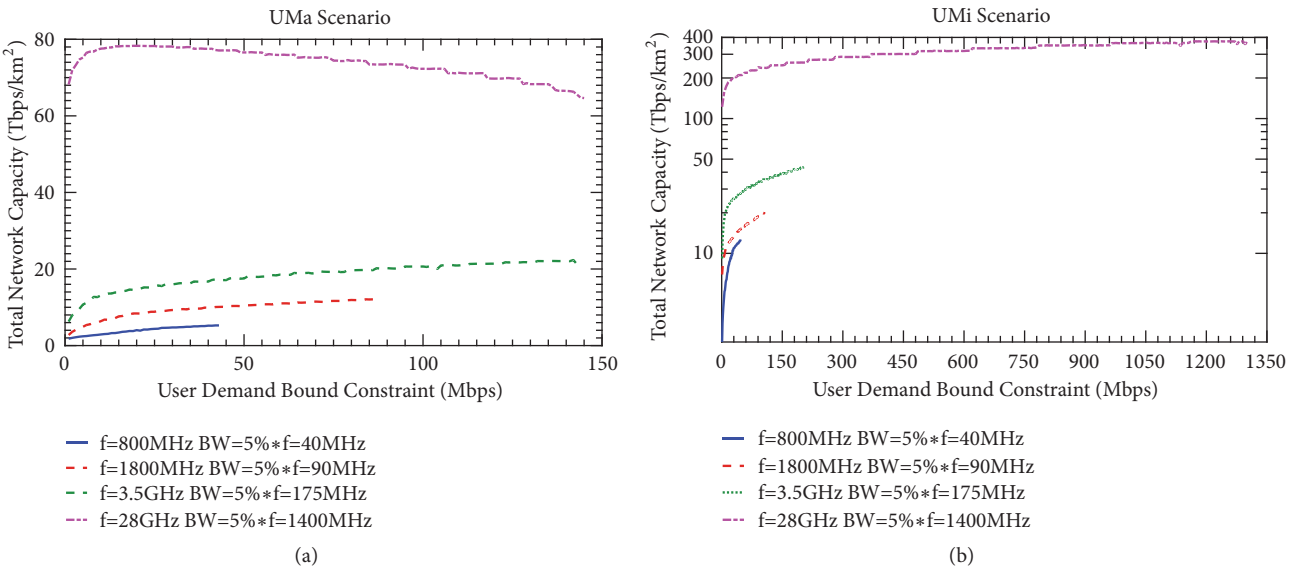


Figure 6: Total network capacity vs. user demand bound constraint. (a) UMa scenario; (b) UMi-street canyon scenario.

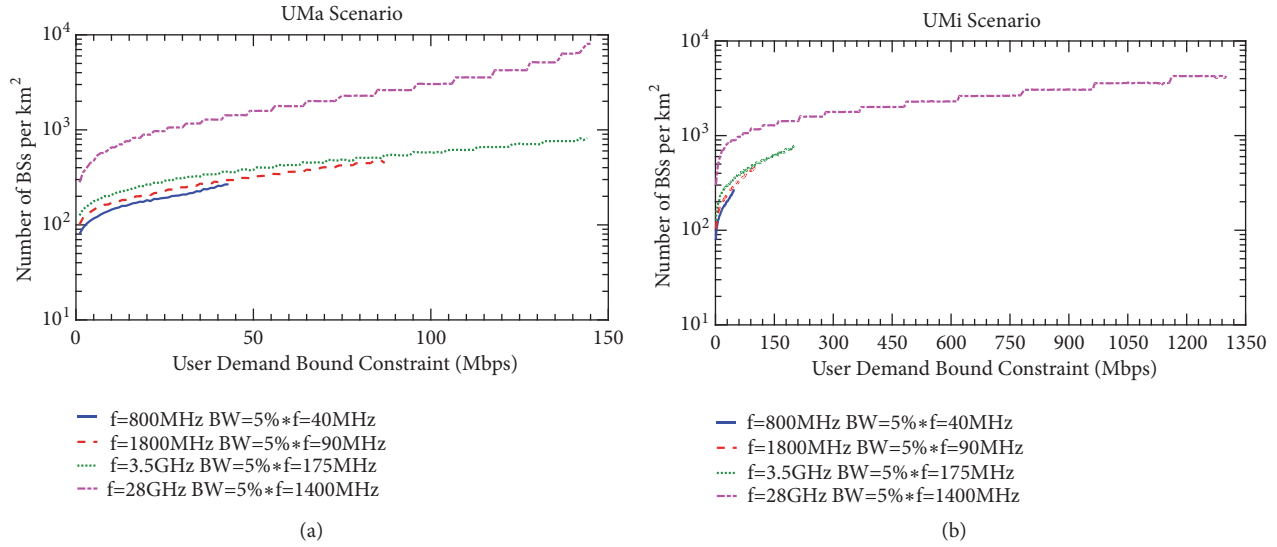


Figure 7: Number of needed BSs vs. user demand bound constraint. (a) UMa scenario; (b) UMi-street canyon scenario.

capacity may be about 80 Tbps/km^2 in the UMa scenario and hundreds of Tbps/km^2 in the UMi scenario. Similar to Figure 5, UMi is advantageous compared to UMa since the BS antennas are closer to users in the UMi-street canyon scenario. UDNs are expected to cover urban areas with small cells, providing a higher data rate than 100 Mbps to every user [6]. Several services in 5G may require a user achievable data rate of at least 1 Gbps and an area capacity of tens of Tbps/km^2 [26]. Figures 5 and 6 show that it is difficult to meet these requirements if UDNs are deployed with conventional bands such as 800 MHz and 1.8 GHz. The new 28 GHz band has been proven to have the potential to meet the requirements for 5G. Furthermore, it is more advantageous in the UMi scenario, i.e., lower antenna heights.

There is a cost when high frequency bands are employed. Figure 7 shows the number of BSs per km^2 when every user is guaranteed a certain capacity. The cost of network deployment is generally proportional to the number of BSs in the area. In both UMa and UMi scenarios, the higher the carrier frequency is, the more the BSs needed for the same user demand bound constraint are. In particular, the number of BSs required for 28 GHz is almost five times that required for 800 MHz or 1.8 GHz. To provide at least 1 Gbps for every user, operators have to deploy around 3500 BSs per km^2. Therefore, the deployment cost with high frequency bands is much higher than that with conventional bands. To meet lower user demand, the conventional lower frequency band can be a more economical choice for the deployment of UDNs

since the number of BSs is lower. However, even though the network capacity for 3.5 GHz is twice that for 1.8 GHz, the number of BSs is slightly different. Therefore, 3.5 GHz may be the more economic choice. If the target area is a hotspot that requires an extremely high user data rate and area capacity, higher frequencies such as 28 GHz, at the expense of great deployment cost, may be the only choice for operators.

6. Conclusions

In this paper, a multiple-objective optimization model for UDNs was proposed for both UMa and UMi-street canyon scenarios. The novel optimization design includes multi-attribute user types in which users are grouped based on their propagation conditions and an infinitesimal dividing modeling method termed the ring method for network capacity dimensioning. The optimal ISD was evaluated to maximize the total network capacity while minimizing the deployment cost under certain user capacity demand bound constraints. Four frequency bands, 800 MHz, 1.8 GHz, 3.5 GHz, and 28 GHz, were investigated for the carrier frequency. The numerical results showed that it would be difficult for conventional bands to meet the requirements of UDN deployment for 5G services. Meanwhile, new higher frequency bands demonstrated the potential to provide 100 Mbps or up to 1 Gbps for every user in the network and tens or even hundreds of Tbps/km^2 in terms of total network capacity. Considering the trade-off between the capacity and deployment cost, 3.5 GHz can be a more economical choice for operators with user demands lower than 200 Mbps. For hotspots with extremely high user-experienced data rates and dense data traffic, 28 GHz in the UMi-street canyon scenario can be employed but the ISD needs to be less than 20 m. This may result in huge deployment costs such as thousands of BSs per km^2.

Data Availability

The simulation data to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the research program of Dongguk University, 2018.

References

[1] A. Osseiran, F. Boccardi, V. Braun et al., "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, 2014.

[2] N. Alliance, "5G white paper, Next generation mobile networks," in *White paper*, 2015.

[3] S. Chen, F. Qin, B. Hu, X. Li, and Z. Chen, "User-centric ultra-dense networks for 5G: Challenges, methodologies, and directions," *IEEE Wireless Communications Magazine*, vol. 23, no. 2, pp. 78–85, 2016.

[4] X. Ge, S. Tu, G. Mao, C.-X. Wang, and T. Han, "5G ultra-dense cellular networks," *IEEE Wireless Communications Magazine*, vol. 23, no. 1, pp. 72–79, 2016.

[5] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-Dense Networks: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2522–2545, 2016.

[6] Nokia, "Ultra dense network (UDN) white paper," in *Nokia Solutions and Neworks Oy*, 2016.

[7] D. González González, E. Mutafungwa, B. Haile, J. Hämäläinen, and H. Poveda, "A Planning and Optimization Framework for Ultra Dense Cellular Deployments," *Mobile Information Systems*, vol. 2017, 2017.

[8] R. Baldemair, T. Irnich, K. Balachandran et al., "Ultra-dense networks in millimeter-wave frequencies," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 202–208, 2015.

[9] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu, "5G network capacity: Key elements and technologies," *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 71–78, 2014.

[10] M. Jaber, Z. Dawy, N. Akl, and E. Yaacoub, "Tutorial on LTE/LTE-A cellular network dimensioning using iterative statistical analysis," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1355–1383, 2016.

[11] P. Sorrells, "Wrestling with the Data Tsunami Indoors," in *White paper*, CommScope, Bellevue, WA, USA, 2014.

[12] J. M. Peha, "Cellular Competition and the Weighted Spectrum Screen," *SSRN Electronic Journal*.

[13] X. Duan and X. Wang, "Authentication handover and privacy protection in 5G hetnets using software-defined networking," *IEEE Communications Magazine*, vol. 53, no. 4, pp. 28–35, 2015.

[14] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Improved Handover Through Dual Connectivity in 5G mmWave Mobile Networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 2069–2084, 2017.

[15] D. López-Pérez, M. Ding, H. Claussen, and A. H. Jafari, "Towards 1 Gbps/UE in cellular systems: understanding ultra-dense small cell deployments," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2078–2101, 2015.

[16] F.-H. Tseng, L.-D. Chou, H.-C. Chao, and J. Wang, "Ultra-dense small cell planning using cognitive radio network toward 5G," *IEEE Wireless Communications Magazine*, vol. 22, no. 6, pp. 76–83, 2015.

[17] J. Zhu, M. Zhao, and S. Zhou, "An Optimization Design of Ultra Dense Networks Balancing Mobility and Densification," *IEEE Access*, vol. 6, pp. 32339–32348, 2018.

[18] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[19] S. Xu, R. Li, and Q. Yang, "Improved genetic algorithm based intelligent resource allocation in 5G Ultra Dense networks," in *Proceedings of the 2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, Barcelona, April 2018.

[20] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz (Release 14)" in *3GPP TR 38.901 V14.0.0*, 2017.

[21] Y. Wei and S. Hwang, "Investigation of spectrum values in rural environments," *ICT Express*, In press.

[22] 3GPP, "Radio frequency (RF) requirements for lte pico node b (release 14)," in *3GPP TR 36.931 V14.0.0*, 2017.

[23] M. Mitchell, *An introduction to genetic algorithms*, MIT press, 1998.

[24] 3GPP, "Evolved universal terrestrial radio access (E-UTRA); radio frequency (RF) system scenarios (Release 14)," *TR 36.942 V14.0.0*, 2017.

[25] H. Claussen, D. Lopez-Perez, L. Ho, S. Kucera, and R. Razavi, *Small Cell Networks: Deployment, Management, And Optimization*, John Wiley & Sons, 2017.

[26] G. Liu and D. Jiang, "5G: Vision and Requirements for Mobile Communication System towards Year 2020," *Chinese Journal of Engineering*, vol. 2016, 2016.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 4308913, 10 pages
https://doi.org/10.1155/2018/4308913

Research Article

Strategic Interaction between Operators in the Context of Spectrum Sharing for 5G Networks

Erwin J. Sacoto-Cabrera,[1,2] **Angel Sanchis-Cano**,[1] **Luis Guijarro**,[1] **José Ramón Vidal**,[1] **and Vicent Pla**[1]

[1]*ITACA, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain*
[2]*GIHP4C, Universidad Politécnica Salesiana-Sede Cuenca, Calle Vieja 12-30 y Elia Liut, 010102 Cuenca, Ecuador*

Correspondence should be addressed to Erwin J. Sacoto-Cabrera; esacoto@ups.edu.ec

Received 30 May 2018; Revised 13 September 2018; Accepted 26 September 2018; Published 11 October 2018

Guest Editor: Ivan Marsa-Maestre

5G networks will make network sharing agreements between mobile operators technically possible. However, depending on the agreed and implemented quality-of-service isolation, the provision of services may lead to unsustainable business cases. In this paper, the economic feasibility of such arrangements is analyzed for the case of two operators. Concretely, while one network operator owns the spectrum, one virtual operator does not, and each one provides service to its subscriber base. Two sharing alternatives, namely, pooling and priority sharing, are studied regarding the profits that each operator gets. We conclude that the network operator is worse off under any circumstances under a pooling agreement, while a lump sum payment may leave the network operator better off under a priority sharing agreement.

1. Introduction

The fifth generation (5G) wireless networks will appear in the market in 2020 and is expected to improve customers' quality of service (QoS) significantly with the increasing of data volume in mobile networks [1]. Furthermore, 5G networks will have the capacity to provide different QoS to users with a wide range of needs, such as new applications, different traffic types, and a wide range of services [2, 3].

The sharing of any sort of resources among network operators has been shown as an efficient way of spreading out the technology at a lower cost. The resources in this paper will be typically spectrum that is owned by the network operator, but similar analysis may be performed when the resources are, e.g., radio-access-network (RAN), network roaming, and core-network [4, 5]. We will refer to the sharing of resources with the general term *network sharing*. Authors of [6] describe a scenario where infrastructure providers enable the leasing of resources, which allow service providers to share the cost of providing coverage with virtual operators, Over-The-Top providers and industry vertical market players. However, network sharing is based on the dynamic deployment and scaling of functions in the 5G networks and requires a careful orchestration of resources to preserve bilateral agreements among operators, as stated by [7].

Regarding the costs involved in moving to 5G, they can be significantly reduced through network densification and network sharing. The business models analyzing these scenarios such as [8] are usually focused on the relation between the service provider and the network operator, where their physical and logical network infrastructures are tightly coupled. There are several ways to implement network sharing, as proposed in [9]. A possible new business model could be based on the sharing of resources, where the owner of the resources and the operator of the resources are different actors.

In the literature, some researches are focused on the study of the economic feasibility of different network sharing scenarios based on elements of queuing theory, such as [10, 11], as well as microeconomics and game theory concepts [12, 13]. For instance, in the context of network sharing, in [14] the effects of infrastructure sharing and competition regulation

on cellular network planning are analyzed. The authors of [15] analyze a network slicing scenario, based on game theory and a framework called "share-constrained proportional allocation"; the results obtained provide an effective and implementable scheme for dynamically sharing resources across slices. In [16], the authors investigate the generic instruments for addressing a spectrum sharing in 4G networks from a cooperative games perspective and conclude that, in most of the considered instances, the mobile network operators (MNOs) are better off building a unique shared RAN than creating subcoalitions or building individual RANs due to the combined gain from spectrum aggregation and cost reduction from sharing the network infrastructure. In [17], the authors investigate a particular class of congestion games and learning mechanisms to design a distributed solution to the wireless network slicing problem. In the same way, there are some studies that analyze the network sharing, based on game theory, such as [18–21]. In the framework of the queue theory, in [22] the effects of queueing delays and users' related costs on the management and control computing resources are analyzed. The authors of [23] propose an analysis from a pricing perspective, based on priority queuing (PQ) and Generalized Processor Sharing, with the issue of maximizing networks operators revenue. In [24], the authors investigate the priority queuing as a way to establish service differentiation; to do that, they consider the Discriminatory Processor Sharing discipline for two models of service with different QoS and determine the prices that maximize the provider's profit. In the same way, [25] studies pricing for heterogeneous services based on a priority queuing as a way to establish service differentiation. Additionally, in [26] the authors analyze the cooperation strategies among mobile network operators competitors, customers, and different types of partners based on network sharing. However, the mentioned studies do not analyze the incentives to mobile operators to engage in spectrum sharing schemes using priority sharing disciplines in a network sharing environment from an economic point of view.

Our aim in this paper is to proceed one step further and model the strategic interaction in network sharing between a network operator and a virtual operator, providing service to their users. We analyze two alternatives: pooling and priority sharing. The analysis is conducted by means of game theory. Our main contribution is to prove that network sharing is economically viable and allows the network and virtual operators to coexist under a priority sharing agreement.

The rest of this paper is organized as follows. In Section 2, we describe in detail the model with the actors, the utility of each actor, and the pricing scheme. In Section 3, we analyze and solve the subscription and pricing strategies of the different models. In Section 4, we show and discuss the results. Finally, in Section 5, we present the conclusions.

2. Model Description

2.1. General Model. This section models the following two sharing agreements:

(i) Pooling: the operators share equally the resource.

(ii) Priority sharing: the operators share the resource under different priorities.

In addition, we analyze in the Appendix the case where there is only one operator providing service; i.e., no priority is implemented and the operator behaves as a monopolist. This case will be used as a baseline for comparison in Section 4. A summary of the notation used in this paper is given in Table 1.

The network operator (aka. Operator 1) is the spectrum owner, while the virtual operator (aka. Operator 2) leases the resource under the specific sharing agreement. As part of the sharing agreement, the operator undertakes to communicate the data of the number of users to the other operator. We assume that each operator has its own subscriber base.

The network, i.e., the spectrum resource that is used for providing service, is modeled as an M/M/1/∞ queue, where packets are generated by the subscribers of the two operators and are served according to the scheduling discipline described below. Operator i has n_i subscribers that independently generate packets following a Poisson process with a rate λ_d, so that for each operator we can define a Poisson packet arrival process with rate $\lambda_i = \lambda_d n_i$, and the sum of the two processes is also Poisson with rate $\lambda = \lambda_1 + \lambda_2$. The service times of all packets are exponentially distributed with mean $1/\mu$. We assume for stability reasons that $\lambda < \mu$.

Under the pooling agreement, the scheduling discipline is First-Come-First-Served and there is no priority. The mean packet system time T_i can be computed as

$$T_1 = T_2 = \left(\frac{1}{\mu - (n_1 + n_2)\lambda_d} \right). \tag{1}$$

Under the priority sharing agreement, we propose that the service provided by the network be modeled by a Discriminatory Processor Sharing (DPS) discipline, where each customer has a relative priority and receives service at an instantaneous rate proportional to the priority [11, 27, 28]. A DPS queue basically works as follows: if there are n customers with priorities $x_1, x_2, \ldots, x_n$ ($x_i \leq 1$ for $i = 1, \ldots, n$), then the customer i is served at a fraction $x_i / \sum_{j=1}^{n} x_j$ of the servers capacity [11]. In our analysis we use a DPS model with two relative priorities: $x_1 = 1 - \gamma$ for the network operator and $x_2 = \gamma$ for the virtual operator, where $x_1 + x_2 = 1$ and $0 \leq \gamma \leq 1$. DPS is a mechanism much more flexible than priority queuing in order to model the sharing of a common resource, thanks to the γ parameter. T_i can be computed [11] as

$$T_1 = \frac{1}{\mu - \lambda_d n_1 - \lambda_d n_2} \left(1 + \frac{\lambda_d n_2 (2\gamma - 1)}{\mu - (1 - \gamma)\lambda_d n_1 - \gamma \lambda_d n_2} \right), \tag{2}$$

$$T_2 = \frac{1}{\mu - \lambda_d n_1 - \lambda_d n_2} \left(1 - \frac{\lambda_d n_1 (2\gamma - 1)}{\mu - (1 - \gamma)\lambda_d n_1 - \gamma \lambda_d n_2} \right). \tag{3}$$

Note that Preemptive Resume PQ, if the users were serviced in a processor sharing manner, is a special case of DPS discipline, since $\gamma = 0$ gives strict priority to operator 1 users while $\gamma = 1$ would give strict priority to operator 2 users [24].

Table 1: Summary of notations.

		Eq.
General Model		
Mean packet system time	T_i	(1)
Mean service rate	μ	(1)
Subscribers mean arrival rate	λ_d	(1)
Operator 1's number of users	n_1	(1)
Operator 2's number of users	n_2	(1)
Virtual Operator's relative priority	γ	(2)
Quality perceived by the users	Q_i	(4)
Users sensitivity to delay	α	(4)
Conversion factor	c	(4)
Users utility	U_i	(5)
Price charged by operator i	p_i	(5)
Operator i's profits	Π_i	(6)
Game Model		
Operator 1's best response	BR_1	(11)
Operator 2's best response	BR_2	(12)
Operator 1's equilibrium price	p_1^*	(13)
Operator 2's equilibrium price	p_2^*	(14)
Operator 1's equilibrium number of users	n_1^*	(16)
Operator 2's equilibrium number of users	n_2^*	(16)
Arbitrarily small positive value	ϵ	
Boundary cases I-II	$\widehat{p}_{1\gamma}$	(34)
Boundary cases I-II	$\widehat{p}_{2\gamma}$	(35)
Boundary cases III-IV	$\widehat{p}_{1U}$	(36)
Boundary cases II-IV	$\widehat{p}_{2U}$	(37)
The Appendix- Monopoly		
Monopolistic operator's profits	Π_0	(A.1)
Monopolistic operator's number of users	n_0	(A.1)
Price charged by monopolistic operator	p_0	(A.1)
Users Utility	U_0	(A.3)
Monopolistic operator's optimal price	p_0^*	(A.9)

We quantify the quality perceived by the users of each operator borrowing a variation of the expression used in [23, 25, 29, 30]. In order to analyze the impact of the priority parameter γ, we assume that the users are homogeneous in their service time requirements for both operators.

$$Q_i \equiv cT_i^{-\alpha}, \quad i = 1, 2, \tag{4}$$

where $c > 0$ is a conversion factor and T_i has been defined and computed above. We define $0 \leq \alpha \leq 1$ as the user sensitivity to delay, since a greater α translates into a worse perceived quality, for a given delay. The user utility is proposed to be given by the difference between the quality perceived by the users in monetary units minus the price charged by the operator. Specifically, the user utility we model has the advantage of being interpretable in monetary-equivalent terms: since cT_i^{α} is the value that the outcome has to user i and p_i (in monetary units per second (m.u./s.)) is the price that the user pays, U_i can be seen as the benefit (m.u./s.) of user i.

$$U_i \equiv cT_i^{-\alpha} - p_i, \quad i = 1, 2. \tag{5}$$

We set the utility when users do not subscribe to the service to zero.

Note that a similar approach for modeling the user utility can be found at, e.g., [22, 31–33]. And this form of utility function can be related to the quasi-linear function widely used in microeconomic and telecommunications networks analysis as described in [34].

The network operator charges a price p_1 to its users, while the virtual operator charges a price p_2. We assume that no costs are incurred by the operators. The operators' profits will then be given by

$$\Pi_i = \lambda_d n_i (p_i) p_i, \quad i = 1, 2. \tag{6}$$

Remark that the profits functions described in (6) should have been decreased by the operator's costs. Important

components of these costs are the operating costs and the investment costs. We do not consider the operating costs, as their inclusion does not provide additional insight since they do not depend on the price of the service, while it makes the expression of profits less explicit. Related to the investment costs, we can consider them constant, given that the time scale on which the provider can adapt his network capacity μ is relatively long compared to the time scale on which prices vary [25, 30].

2.2. Game Model. From the above model description, we can observe the following strategic interactions:

(i) Flow *i's* subscription decision is influenced by operator *i's* pricing decision.

(ii) Operator *i's* profit depends on flow *i's* subscription decision.

(iii) Flow *i's* subscription decision depends on flow *j's* decision with ($j \neq i$), through the Q_i factor.

(iv) Operator *i's* profit is then influenced by operator *j's* pricing decision, indirectly through flow *j's* subscription decision.

We conclude then that the above strategic scenario is amenable to game theory. Specifically, each operator will choose its pricing strategy p_i aiming at maximizing its profit Π_i.

We will use a two-stage sequential game of the multi-leader follower kind, where, in the first stage, each operator chooses a price p_i in order to maximize its profit Π_i. The game is solved using backward induction [35], which means that at Stage I players proceed strategically anticipating the solution of Stage II.

In the second stage, for each operator, the following applies. Each user takes its own subscription decision, trying to maximize the utility it gets from either subscribing to the operator or not. The user will observe the price published by the operator and will take the decision of service subscription. Assuming that number of users is high enough, the individual subscription decision of one user does not affect the utility of the rest of the users. Under these conditions, the equilibrium reached is that postulated by Wardrop [36]. Applying this notion of equilibrium, known as *Wardrop equilibrium*, to the user subscription problem, we postulate that

(i) either $n_i \geq 0$ and $U_i = 0$

(ii) or $n_i = 0$ and $U_i < 0$.

Combining the above alternatives, we can identify four outcomes:

(i) Case I:

$$\begin{aligned} U_1 &= 0, \\ U_2 &= 0 \\ \text{and } n_1 &\geq 0, \\ n_2 &\geq 0 \end{aligned} \tag{7}$$

(ii) Case II:

$$\begin{aligned} U_1 &= 0, \\ U_2 &< 0 \\ \text{and } n_1 &\geq 0, \\ n_2 &= 0 \end{aligned} \tag{8}$$

(iii) Case III:

$$\begin{aligned} U_1 &< 0, \\ U_2 &= 0 \\ \text{and } n_1 &= 0, \\ n_2 &\geq 0 \end{aligned} \tag{9}$$

(iv) Case IV:

$$\begin{aligned} U_1 &< 0, \\ U_2 &< 0 \\ \text{and } n_1 &= 0, \\ n_2 &= 0 \end{aligned} \tag{10}$$

In the first stage, each operator chooses a price in order to maximize its profits in a simultaneous and independent way. Each operator is not only aware of the users subscription decision in the second stage, but also of the rational behavior of the other operator. Under such circumstances, the equilibrium reached is the *Nash equilibrium* [37], where no player has an incentive to change its own strategy unilaterally.

The general method to discover the set of Nash equilibria is to obtain the best-response (BR) function of each operator and identify the crossing points [34]. The BR functions are defined as follows:

$$BR_1(p_2) = \arg\max_{p_1} \Pi_1(p_1, p_2), \tag{11}$$

$$BR_2(p_1) = \arg\max_{p_2} \Pi_2(p_1, p_2). \tag{12}$$

Once we have obtained the BR functions, we can obtain the set of Nash equilibria solving the following system of equations:

$$p_1^* = \arg\max_{p_1} \Pi_1(p_1, p_2^*), \tag{13}$$

$$p_2^* = \arg\max_{p_2} \Pi_2(p_1^*, p_2). \tag{14}$$

3. Models Analysis

In this section, we analyze the models described in Section 2.

3.1. Pooling. In this subsection, operator 1 and operator 2 are sharing the spectrum resource. The utility of the users can be obtained replacing (1) in (5).

$$U_i = c \cdot (\mu - (n_1 + n_2) \cdot \lambda_d)^{\alpha} - p_i \quad i = 1, 2 \tag{15}$$

We obtain the equilibrium values for n_1 and n_2, given p_1 and p_2, as follows.

(i) Case I: from (7) and (15) we obtain

$$n_1^* + n_2^* = \frac{\left[\mu - (p_1/c)^{1/\alpha}\right]}{\lambda_d} = \frac{\left[\mu - (p_2/c)^{1/\alpha}\right]}{\lambda_d}, \tag{16}$$

which is only feasible if $p_1 = p_2$. Without loss of generality, we can assume that $n_1^* = n_2^*$, so that

$$n_1^* = \frac{\left[\mu - (p_1/c)^{1/\alpha}\right]}{2\lambda_d}, \tag{17}$$

$$n_2^* = \frac{\left[\mu - (p_2/c)^{1/\alpha}\right]}{2\lambda_d}. \tag{18}$$

(ii) Case II: from (8) and (15) we obtain

$$n_1^* = \frac{\left[\mu - (p_1/c)^{1/\alpha}\right]}{\lambda_d}, \tag{19}$$

$$n_2^* = 0, \tag{20}$$

$$p_1 \leq c\mu^{\alpha}. \tag{21}$$

$$p_2 > p_1. \tag{22}$$

(iii) Case III: from (9) and (15) we obtain

$$n_1^* = 0, \tag{23}$$

$$n_2^* = \frac{\left[\mu - (p_2/c)^{1/\alpha}\right]}{\lambda_d}, \tag{24}$$

$$p_2 \leq c\mu^{\alpha}, \tag{25}$$

$$p_1 > p_2. \tag{26}$$

(iv) Case IV: finally, from (10) and (15)

$$n_1^* = 0 \tag{27}$$

$$n_2^* = 0, \tag{28}$$

$$p_1 > c\mu^{\alpha}, \tag{29}$$

$$p_2 > c\mu^{\alpha}. \tag{30}$$

As a corollary, price values for Case I comply with $p_1 \leq c\mu^{\alpha}$.

Table 2 summarizes the results of the users' subscription decision under a pooling agreement. Figure 1 shows a graphical representation of the regions for a specific configuration of the parameters.

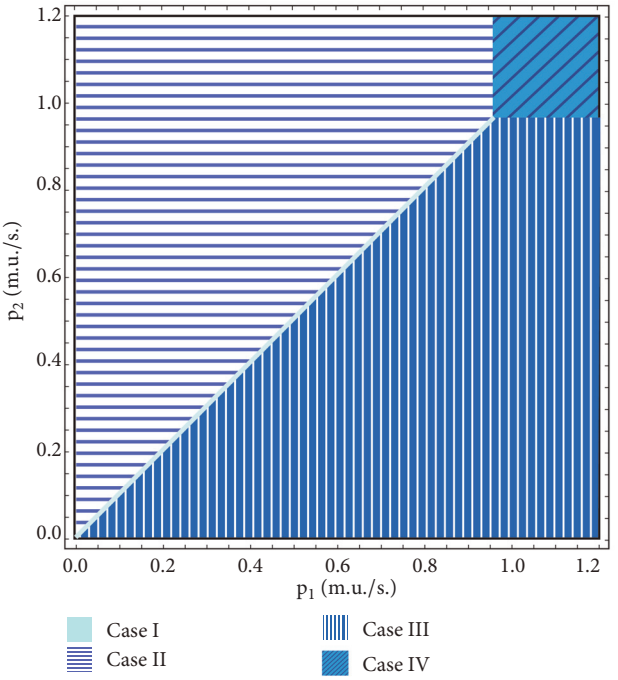


Figure 1: Wardrop equilibrium regions for pooling with $c = 1$, $\alpha = 0.8$, $\mu = 1$ packets per second (p/s), $\lambda_d = 0.5$ p/s.

At this point we proceed to analyze the equilibrium prices p_i^*, given the values for n_1^* and n_2^*. For this purpose we use (6) and we obtain the BRs as defined by (11) and (12):

$$BR_1(p_2) = \begin{cases} p_1 \geq 0 & \text{if } p_2 = 0 \\ p_2 - \epsilon & \text{if } 0 < p_2 \leq p_0^* \\ c\mu^{\alpha}\left(\dfrac{\alpha}{1+\alpha}\right)^{\alpha} & \text{if } p_2 > p_0^* \end{cases}$$

$$BR_2(p_1) = \begin{cases} p_2 \geq 0 & \text{if } p_1 = 0 \\ p_1 - \epsilon & \text{if } 0 < p_1 \leq p_0^* \\ c\mu^{\alpha}\left(\dfrac{\alpha}{1+\alpha}\right)^{\alpha} & \text{if } p_1 > p_0^* \end{cases} \tag{31}$$

where p_0^* is the profit maximizing price for a monopolistic operator, which is given by (A.9), and ϵ is an arbitrarily small positive value. The Nash equilibrium is the intersection of the BRs, which is easily derived as $p_1^* = 0$ and $p_2^* = 0$, with zero profits $\Pi_1^* = 0, \Pi_2^* = 0$. We conclude that a pooling agreement between a network operator and a virtual operator generates strategic interaction that drives prices and profits to zero.

3.2. Priority Sharing. In this section, we analyze the case where operator 1 and operator 2 share the spectrum resource according to a distribution of relative priority. As justified in

Table 2: Users subscription decision-pooling.

Case	n_1^*	n_2^*	p_1	p_2
I	$\frac{[\mu-(p_1/c)^{1/\alpha}]}{2\lambda_d}$	$\frac{[\mu-(p_2/c)^{1/\alpha}]}{2\lambda_d}$	$p_1 \le c\mu^\alpha$	$p_2 = p_1$
II	$\frac{[\mu-(p_1/c)^{1/\alpha}]}{\lambda_d}$	0	$0 \le p_1 \le c\mu^\alpha$	$p_2 > p_1$
III	0	$\frac{[\mu-(p_2/c)^{1/\alpha}]}{\lambda_d}$	$p_1 > p_2$	$0 \le p_2 \le c\mu^\alpha$
IV	0	0	$p_1 > c\mu^\alpha$	$p_2 > c\mu^\alpha$

Section 2.1, this distribution is modeled by means of DPS and T_i are now given by (2) and (3), so that the user utility will be

$$U_1 = c\left(\frac{\mu+(\gamma-1)(\lambda_d n_1+\lambda_d n_2)}{(\mu-\lambda_d n_1-\lambda_d n_2)(\mu-\gamma\lambda_d n_2+(\gamma-1)\lambda_d n_1)}\right)^{-\alpha} - p_1, \quad (32)$$

$$U_2 = c\left(\frac{\mu-\gamma(\lambda_d n_1+\lambda_d n_2)}{(\mu-\lambda_d n_1-\lambda_d n_2)(\mu-\gamma\lambda_d n_2+(\gamma-1)\lambda_d n_1)}\right)^{-\alpha} - p_2. \quad (33)$$

Following a similar reasoning as in Section 3.1, we obtain the equilibrium values for n_1^* and n_2^* and the constraints on p_1 and p_2, which are shown in Table 3. Figure 2 shows a graphical representation of the cases for a specific configuration of the parameters.

The following expressions for the cases boundaries apply:

$$\widehat{p}_{1\gamma} = c\left(\frac{\gamma\mu(p_2/c)^{-1/\alpha}-\gamma+1}{-\gamma\mu+\gamma(p_2/c)^{1/\alpha}+\mu}\right)^{-\alpha}, \quad (34)$$

$$\widehat{p}_{2\gamma} = c\left(\frac{(p_1/c)^{-1/\alpha}\left((\gamma-1)\mu-\gamma(p_1/c)^{1/\alpha}\right)}{(\gamma-1)(p_1/c)^{1/\alpha}-\gamma\mu}\right)^{-\alpha}, \quad (35)$$

$$\widehat{p}_{1U} = c\mu^\alpha, \quad (36)$$

$$\widehat{p}_{2U} = c\mu^\alpha. \quad (37)$$

Note that when $\gamma = 1/2$, the priority sharing agreement reduces to the pooling agreement, since FCFS and DPS result in the same value for T_i.

Unlike Section 3.1, we have proceeded numerically in the analysis of Stage I, as regards the BRs and the Nash equilibrium.

4. Results and Discussion

In this section, numerical values are computed for the users, prices, and profits for the network and virtual operators under a priority sharing agreement. Note that the prices and the profits are zero for pooling, as concluded in Section 3.1.

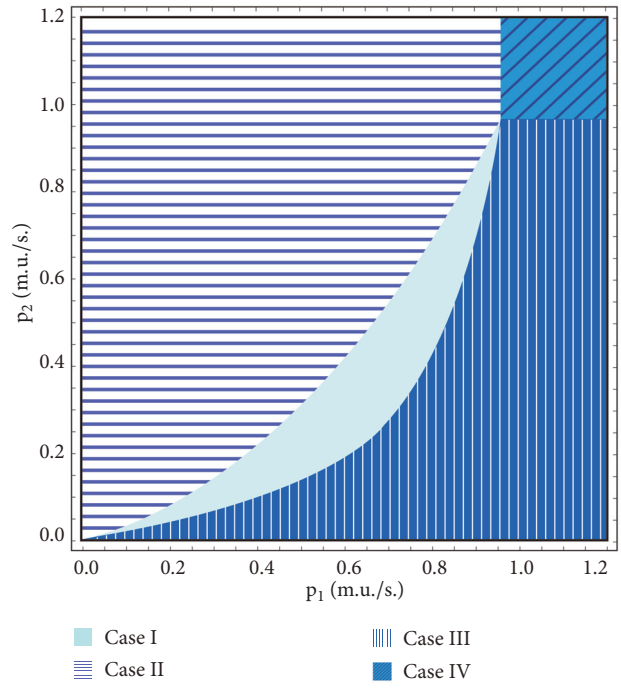


Figure 2: Wardrop equilibrium regions for priority sharing with $\gamma = 1/10$, $c = 1$, $\alpha = 0.8$, $\mu = 1$ packets per second (p/s), $\lambda_d = 0.5$ p/s.

We set the parameters $\alpha = \{0.5, 0.8, 1\}$, $c = 1$, $\mu = 1$ packet/s, $\lambda_d = 0.01$ packet/s and we vary $\gamma \in [0, 1]$. Figures 3, 4, 5, 6, 7, and 8 show the prices, the number of users, and profits, respectively. The values p_0^*, n_0^*, and Π_0^* refer to the monopoly (the Appendix), while the values p_1^*, n_1^*, and Π_1^* refer to the network operator, and the values p_2^*, n_2^*, and Π_2^* refer to the virtual operator.

Figure 3 shows the effect of α and γ (the virtual operator's relative priority) on the equilibrium prices of the monopolistic operator and operator 1. The price p_1 decreases as α and γ increase, since the QoS and priority $(1-\gamma)$ of the service that the network operator can provide to its subscribers are lower. However, when γ approaches 1 the price p_1 increases. In this situation the network operator is giving strict priority to the virtual operator; i.e., the DPS discipline behaves as a PQ. This specific setting is analyzed in [24, 29, 30] and the results match those shown in Figure 3. The monopolistic price p_0^* is a constant value because it does not depend on γ (Table 4). When $\gamma \in]0, 1[$, p_0^* is greater than the equilibrium prices,

Table 3: Users subscription decision-priority sharing.

Case	n_1^*	n_2^*	p_1	p_2
I ($\gamma \neq 1/2$)	$\frac{1}{\lambda_d}\frac{1}{(\gamma-1)(c/p_1)^{1/\alpha}+\gamma(c/p_2)^{1/\alpha}}$ $-\frac{1}{\lambda_d}\frac{\mu}{\gamma(c/p_2)^{-1/\alpha}(c/p_1)^{1/\alpha}+\gamma-1}$	$\frac{1}{\lambda_d}\frac{\mu}{(\gamma-1)(c/p_2)^{1/\alpha}(c/p_1)^{-1/\alpha}+\gamma}$ $-\frac{1}{\lambda_d}\frac{1}{(\gamma-1)(c/p_1)^{1/\alpha}+\gamma(c/p_2)^{1/\alpha}}$	$p_1 \leq \widehat{p}_{1\gamma}$	$p_2 \leq \widehat{p}_{2\gamma}$
I ($\gamma = 1/2$)	$\frac{\left[\mu-(p_1/c)^{1/\alpha}\right]}{2\lambda_d}$	$\frac{\left[\mu-(p_2/c)^{1/\alpha}\right]}{2\lambda_d}$	$p_1 \leq \widehat{p}_{1\gamma}$	$p_2 = p_1$
II	$\left[\mu-\left(\frac{p_1}{c}\right)^{1/\alpha}\right]\left(\frac{1}{\lambda_d}\right)$	0	$0 \leq p_1 \leq \widehat{p}_{1U}$	$p_2 > \widehat{p}_{2\gamma}$
III	0	$\left[\mu-\left(\frac{p_2}{c}\right)^{1/\alpha}\right]\left(\frac{1}{\lambda_d}\right)$	$p_1 > \widehat{p}_{1\gamma}$	$0 \leq p_2 \leq \widehat{p}_{2U}$
IV	0	0	$p_1 > \widehat{p}_{1U}$	$p_2 > \widehat{p}_{2U}$

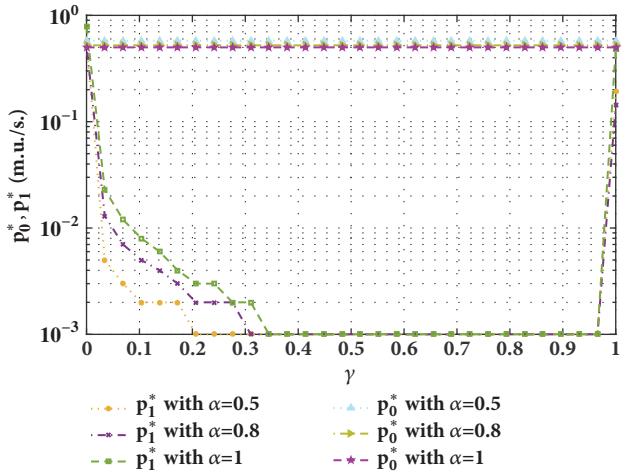


Figure 3: Prices of users for the network operator and the monopolistic operator as a function of γ for different values of α.

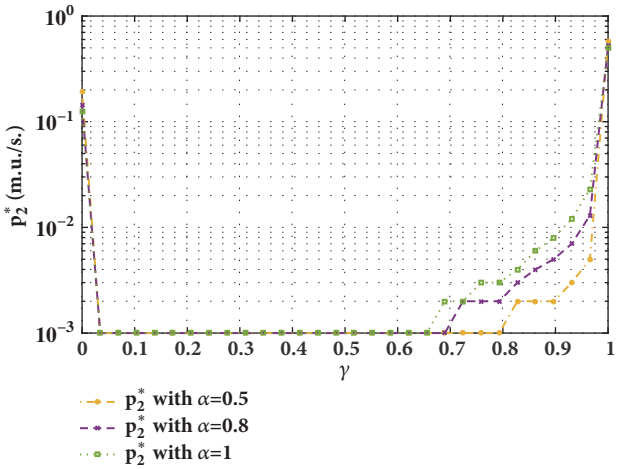


Figure 4: Prices of users of the virtual operator as a function of γ for different values of α.

whereas $p_i^* = p_0^*$ when operator i is given strict priority, since it receives a preemptive priority service; i.e., it is not affected by the nonpriority users.

Figure 4 shows the effect of α and γ on the equilibrium prices of operator 2. The prices (p_1, p_2) behavior is symmetrical; i.e., p_2 exhibits the same behavior as the one described above but when α increases and γ decreases from 1 to 0.

We can conclude that the entry of the virtual operator causes the price charged both by the network operator and by the virtual operator to decrease; therefore, it is beneficial for the users. And we have numerically investigated that PQ is the limit case of DPS always providing the highest prices.

Figure 5 shows the effect of α and γ on the number of users of operator 1. n_1 increases when α increases from 0 to 1 and γ increases from 0 to 0.5, which can be explained by the decrease in price p_1. At the same time, operator 2 loses all its users, and n_2 gets to zero. When $\gamma = 0.5$, the system behaves as in the pooling agreement, i.e., zero prices and equal level of subscription. When γ increases from 0.5 to 1, the number of users switches and operator 1 loses its users. Figure 6 shows the effect of α and γ on the number of users of operator 2. The behavior of n_2 and n_1 is symmetrical; i.e., n_2 exhibits the same behavior as the one described above but when α increases from 0 to 1 and γ decreases from 1 to 0.

The results show that the operator with the higher priority (e.g., operator 1 in the left hand side of Figure 5) is able to provide a satisfactory service to a maximum number of users, although the equilibrium price needs to be very low but nonnull (Figure 3). For γ near the values of 0 and 1, the number of users that the operator with the higher priority can serve is lower than the maximum; however, the equilibrium price is not so low as before, and the profits are greater (Figure 7). This specific setting is analyzed in [24, 29, 30], where the discipline was Preemptive Resume Priority Queueing (which, if users are served in a processor-sharing manner, is equivalent to DPS for $\gamma = 0$ or 1). The explanation for these results is related to the fact that only when γ is near 0 or 1, the QoS received by each user group is sufficiently different from each other so that the competition between the two operators is not too fierce and the prices need not be too low.

Figure 7 shows the equilibrium profit of each operator as α and γ vary. We observe that the network operator suffers a significant reduction in its profit Π_1 when α and γ increase between $0 < \gamma < 1$. The profits only recover when γ approaches

Table 4: Users, prices, and profit of one operator case.

Case	n_0^*	p_0^*	$\Pi_0^*(p_0)$
I	$\left[\mu - \left(\frac{p_o}{c}\right)^{1/\alpha}\right]\left(\frac{1}{\lambda_0}\right)$	$p_0^* = c\mu^\alpha\left(\frac{\alpha}{1+\alpha}\right)^\alpha$	$\Pi_0(p_0^*) = c\mu^{\alpha+1}\left(\frac{\alpha}{1+\alpha}\right)^{\alpha+1}\left(\frac{1}{\alpha}\right)$
II	0	$p_0 > c\mu^\alpha$	$\Pi_0(p_0^*) = 0$

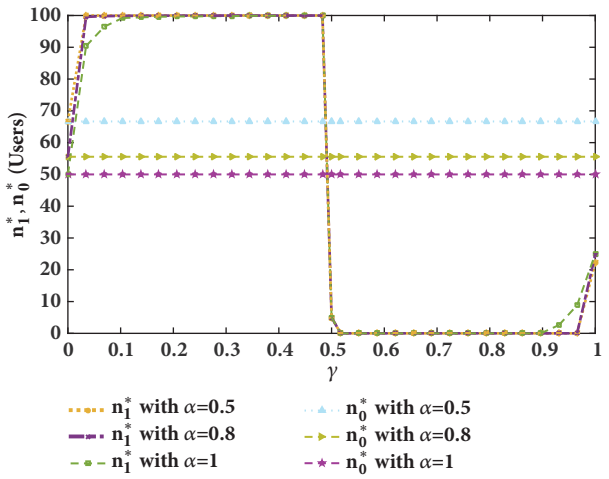


Figure 5: Number of users for the network operator and the monopolistic operator as a function of γ for different values of α.

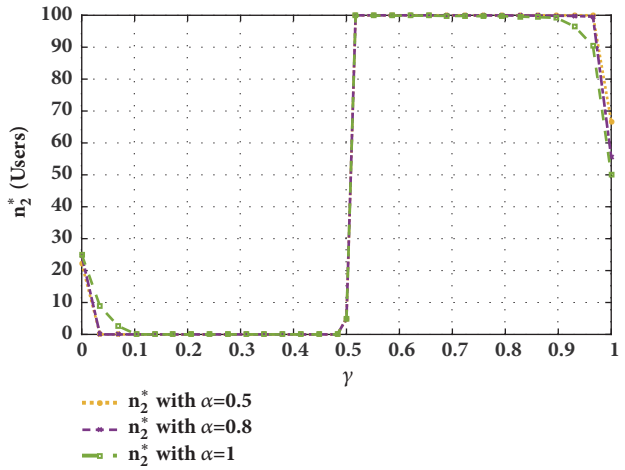


Figure 6: Number of users for the virtual operator as a function of γ for different values of α.

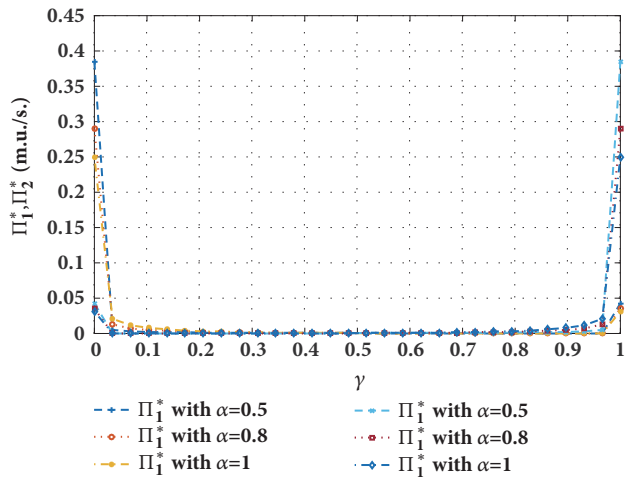


Figure 7: Profits for the network operator and the virtual operator as a function of γ for different values of α.

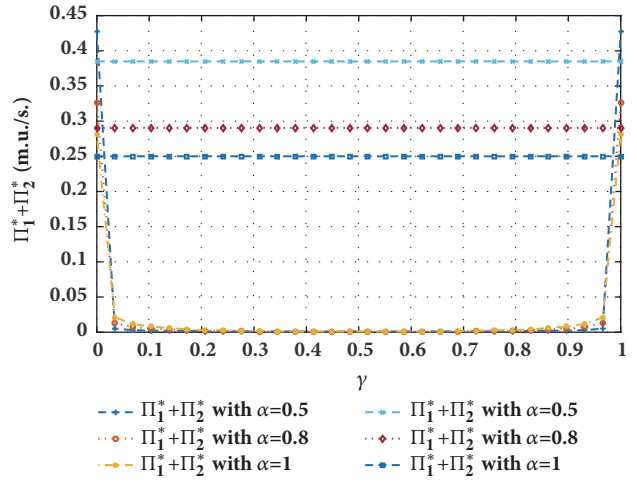


Figure 8: $\Pi_1^* + \Pi_2^*$ and profits for the monopolistic operator as a function of γ for different values of α.

0 or 1; i.e., the DPS discipline behaves as a PQ. The behavior described is consistent with the combined behavior of the prices and the number of users, described above. The virtual operator behavior is symmetrical; i.e., Π_2 exibits the same behavior as the one described above but when α increases and γ decreases from 1 to 0.

When compared to the profit that the network operator would get in a monopoly, Π_0 (the Appendix), we conclude that the network operator is always worse off under a priority sharing agreement. Obviously, it is also worse off under a pooling agreement. However, we can determine a lump sum payment m that the virtual operator would make to the network operator in order to provide an incentive to the latter. Indeed, this payment should not only improve the situation of the network operator, that is, $\Pi_1 + m \geq \Pi_0$, but also allow the virtual operator to obtain nonnegative profits $\Pi_2 - m \geq 0$. Joining the two conditions we obtain $\Pi_0 - \Pi_1 \leq m \leq \Pi_2$. Thus, a necessary condition for the existence of a possible payment is $\Pi_1 + \Pi_2 \geq \Pi_0$. This condition obviously does not hold for the pooling agreement. As regards the priority

sharing agreement, Figure 8 shows that this condition holds only for sufficiently small values of γ or $1-\gamma$.

We can conclude that network sharing is incentive compatible under the following conditions: first, that a priority sharing agreement is reached and second, that either a very low or a very high relative priority is granted to the virtual operator willing to access the network operator resources.

5. Conclusions

In this paper, we have evaluated the issue of network sharing between a network operator and a virtual operator for profit-maximization purposes. We have studied the economic viability of the sharing under pooling and priority sharing agreements, and these are compared with a monopoly. Our main results suggest that the network operator is worse off under any circumstances under a pooling agreement. The entry of a virtual operator is desirable from the point of view of the users' prices. Additionally, for each parameter configuration there exist a range of values of γ, for which a lump sum payment can be designed so that the network operator has an incentive to let the virtual operator enter.

Appendix

Monopoly

In this appendix, we study the case where only the network operator provides the service to users. The first stage described in Section 2.2 is reduced to an optimal decision by the network operator in order to maximize its profits (A.1),whereas the second stage is reduced to two cases.

$$\Pi_0 = \lambda_0 n_0 p_0 \tag{A.1}$$

The utility in this case is

$$U_0 = cT_0^{-\alpha} - p_0 \tag{A.2}$$

where $T_0 = (1/(\mu - n_0\lambda_0))$, and replacing T_0 in (A.2), we obtain

$$U_0 = c\left(\mu - n_0\lambda_0\right)^{\alpha} - p_0 \tag{A.3}$$

Analyzing the users' subscription decision we observe that given a price p_0 announced by the operator, the Wardrop equilibrium will be as follows.

(i) Case I: the number of users subscribing increases until the utility is zero. Therefore, the condition for this case is

$$U_0 = 0 \tag{A.4}$$

Solving (A.3) under condition (A.4), the number of users is then

$$n_0 = \left[\mu - \left(\frac{p_o}{c}\right)^{1/\alpha}\right]\left(\frac{1}{\lambda_0}\right) \tag{A.5}$$

(ii) Case II: the price in (A.3) is so high that the utility is always negative. Therefore the condition for this case is

$$U_0 < 0 \tag{A.6}$$

Under condition (A.6), the users do not subscribe the service. Therefore, the number of users is

$$n_0 = 0 \tag{A.7}$$

Assuming equilibrium is Case I, we can obtain the profits replacing (A.5) in (A.1).

$$\Pi_0 = \left[\mu - \left(\frac{p_0}{c}\right)^{1/\alpha}\right] p_0 \tag{A.8}$$

We can maximize the profit setting its derivative with respect to the price (p_0) equal to zero; the result of (p_0^*) is

$$p_0^* = c\mu^{\alpha}\left(\frac{\alpha}{1+\alpha}\right)^{\alpha} \tag{A.9}$$

Finally we can obtain the maximum profit replacing (A.9) in (A.1).

$$\Pi_0\left(p_0^*\right) = c\mu^{\alpha+1}\left(\frac{\alpha}{1+\alpha}\right)^{\alpha+1}\left(\frac{1}{\alpha}\right) \tag{A.10}$$

Table 4 shows the results of the prices and profit for this case.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was supported by the Spanish Ministry of Economy and Competitiveness through projects TIN2013-47272-C2-1-R (cosupported by the European Social Fund) and BES-2014-068998 and partially supported by the Salesian Polytechnic University of Ecuador through a Ph.D. scholarship granted to the first author.

References

[1] V. Tikhvinskiy and G. Bochechka, *Conceptual aspects of 5G construction*, 2013.

[2] X. Zhang, W. Cheng, and H. Zhang, "Heterogeneous statistical QoS provisioning over 5G mobile wireless networks," *IEEE Network*, vol. 28, no. 6, pp. 46–53, 2014.

[3] W. Cheng, X. Zhang, and H. Zhang, "Heterogeneous statistical QoS provisioning over 5G wireless full-duplex networks," in *Proceedings of the IEEE INFOCOM 2015 - IEEE Conference on Computer Communications*, pp. 55–63, Kowloon, Hong Kong, April 2015.

[4] P. Marsch, I. Da Silva, O. Bulakci et al., "5G Radio Access Network Architecture: Design Guidelines and Key Considerations," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 24–32, 2016.

[5] A. Kliks, B. Musznicki, K. Kowalik, and P. Kryszkiewicz, "Perspectives for resource sharing in 5G networks," *Telecommunication Systems*, vol. 68, no. 4, pp. 605–619, 2018.

[6] K. Samdanis, X. Costa-Perez, and V. Sciancalepore, "From network sharing to multi-tenancy: The 5G network slice broker," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 32–39, 2016.

[7] M. Andrews, M. Bradonjić, and I. Saniee, "Quantifying the benefits of infrastructure sharing," in *Proceedings of the the 12th workshop*, pp. 1–6, Cambridge, Massachusetts, June 2017.

[8] N. Alliance, *"5g white paper," Next generation mobile networks, white paper*, 2015.

[9] J. G. Andrews, S. Buzzi, and W. Choi, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[10] N. Chee-Hock and S. Boon-Hee, "Queueing Modelling Fundamentals: With Applications in Communication Networks: Second Edition," *Queueing Modelling Fundamentals: With Applications in Communication Networks: Second Edition*, pp. 1–271, 2008.

[11] R. Hassin and M. Haviv, *To queue or not to queue: Equilibrium behavior in queueing systems*, vol. 59, Springer Science & Business Media, 2003.

[12] J. C. Harsanyi and R. Selten, *A general theory of equilibrium selection in games*, MIT Press, Cambridge, MA, 1988.

[13] S. Lasaulce, M. Debbah, and E. Altman, "Methodologies for analyzing equilibria in wireless games: A look at pure, mixed, and correlated equilibria," *IEEE Signal Processing Magazine*, vol. 26, no. 5, pp. 41–52, 2009.

[14] P. Di Francesco, F. Malandrino, T. K. Forde, and L. A. DaSilva, "A Sharing- and Competition-Aware Framework for Cellular Network Evolution Planning," *IEEE Transactions on Cognitive Communications and Networking*, vol. 1, no. 2, pp. 230–243, 2015.

[15] P. Caballero, A. Banchs, G. De Veciana, and X. Costa-Perez, "Network slicing games: Enabling customization in multi-tenant networks," in *Proceedings of the 2017 IEEE Conference on Computer Communications, INFOCOM 2017*, USA, May 2017.

[16] L. Cano, A. Capone, G. Carello, M. Cesana, and M. Passacantando, "Cooperative Infrastructure and Spectrum Sharing in Heterogeneous Mobile Networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 10, pp. 2617–2629, 2016.

[17] S. DOro, F. Restuccia, T. Melodia, and S. Palazzo, "Low-Complexity Distributed Radio Access Network Slicing: Algorithms and Experimental Results," https://arxiv.org/abs/1803.07586.

[18] S.-L. Hew and L. B. White, "Cooperative resource allocation games in shared networks: Symmetric and asymmetric fair bargaining models," *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4166–4175, 2008.

[19] F. Teng, D. Guo, and M. L. Honig, "Sharing of Unlicensed Spectrum by Strategic Operators," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 3, pp. 668–679, 2017.

[20] Y.-T. Lin, H. Tembine, and K.-C. Chen, "Inter-operator spectrum sharing in future cellular systems," in *Proceedings of the 2012 IEEE Global Communications Conference, GLOBECOM 2012*, pp. 2597–2602, USA, December 2012.

[21] Y. Zhu, G. Meng, Y. Qiu, Z. Ji, G. Xie, and Y. Song, "Spectrum Sharing for Mobile Virtual Networks: A Game Theoretic Approach," in *Proceedings of the 2018 13th APCA International Conference on Automatic Control and Soft Computing (CONTROLO)*, pp. 179–183, Ponta Delgada, Portugal, June 2018.

[22] H. Mendelson, "Pricing computer services: Queueing effects," *Communications of the ACM*, vol. 28, no. 3, pp. 312–321, 1985.

[23] Y. Hayel, D. Ros, and B. Tuffin, "Less-than-best-effort services: pricing and scheduling," in *Proceedings of the IEEE INFOCOM 2004*, pp. 66–75, Hong Kong, PR China.

[24] Y. Hayel and B. Tuffin, "Pricing for Heterogeneous Services at a Discriminatory Processor Sharing Queue," in *NETWORKING 2005. Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications Systems*, vol. 3462 of *Lecture Notes in Computer Science*, pp. 816–827, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[25] M. Mandjes, "Pricing strategies under heterogeneous service requirements," *Computer Networks*, vol. 42, no. 2, pp. 231–249, 2003.

[26] J. Markendahl, *Mobile network operators and cooperation: A tele-economic study of infrastructure sharing and mobile payment services [Ph.D. dissertation]*, KTH, 2011.

[27] L. Kleinrock, "Time-shared systems: a theoretical treatment," *Journal of the ACM*, vol. 14, no. 2, pp. 242–261, 1967.

[28] E. Altman, K. Avrachenkov, and U. Ayesta, "A survey on discriminatory processor sharing," *Queueing Systems*, vol. 53, no. 1-2, pp. 53–63, 2006.

[29] L. Guijarro, V. Pla, and B. Tuffin, "Entry game under opportunistic access in cognitive radio networks: A priority queue model," in *Proceedings of the 6th IFIP/IEEE Wireless Days Conference, WD 2013*, Spain, November 2013.

[30] A. Sanchis-Cano, L. Guijarro, V. Pla, and J. R. Vidal, "Economic viability of HTC and MTC service provision on a common network infrastructure," in *Proceedings of the 14th IEEE Annual Consumer Communications and Networking Conference, CCNC 2017*, pp. 1044–1050, USA, January 2017.

[31] P. Vamvakas, E. E. Tsiropoulou, and S. Papavassiliou, "Dynamic Provider Selection & Power Resource Management in Competitive Wireless Communication Markets," *Mobile Networks and Applications*, vol. 23, no. 1, pp. 86–99, 2018.

[32] E. Altman, T. Boulogne, R. El-Azouzi, T. Jiménez, and L. Wynter, "A survey on networking games in telecommunications," *Computers & Operations Research*, vol. 33, no. 2, pp. 286–311, 2006.

[33] P. Belleflamme and M. Peitz, *Industrial organization: markets and strategies*, Cambridge University Press, 2015.

[34] P. Maillé and B. Tuffin, *Telecommunication Network Economics: from Theory to Applications*, Cambridge University Press, Cambridge, UK, 2014.

[35] E. N. Barron, *Game Theory: An Introduction*, John Wiley & Sons, 2013.

[36] J. G. Wardrop, "Some theoretical aspects of road traffic research," in *Proceedings of the Inst Civil Engineers*, London, UK.

[37] H. Varian, "Intermediate Microeconomics: A Modern Approach," in *Proceedings of the Ninth International Student Edition*, 2014.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 7484786, 12 pages
https://doi.org/10.1155/2018/7484786

Research Article

Dynamic Tradeoff between Energy and Throughput in Wireless 5G Networks

Cédric Gueguen and Malo Manini

B-Com, University of Rennes 1/IRISA, France

Correspondence should be addressed to Cédric Gueguen; cedric.gueguen@irisa.fr

Received 1 June 2018; Accepted 4 August 2018; Published 4 September 2018

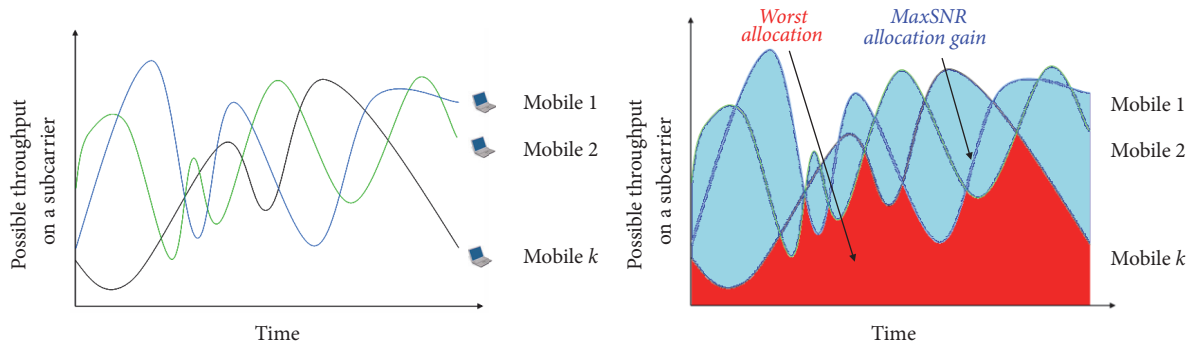
Academic Editor: Jose M. Gimenez-Guzman

Even though system energy and spectral efficiency are major issues in wireless network, reaching these objectives conjointly seems very difficult and requires the usage of tradeoffs. Moreover, depending on the context, the importance of either varies. In underloaded context, guaranteeing high Quality of Service (QoS) is easily achievable due to large surplus of available radio resources and focus should be put on energy rather than system throughput. On the contrary, in an overloaded context, the lack of available radio resources required that resources allocation algorithms focus on system capacity in order to preserve QoS. Since the major issue of the network is to satisfy users, in this specific case, energy consumption must become lesser important. Many specialized solutions have been proposed that focus either on energy saving or on throughput maximization. They provide high performances, respectively, on their specific network traffic load context, previously described, but are not optimized outside. Other solutions that proposed static tradeoffs provide average performances but can not be fully efficient in all scenarios. In this paper, we propose a Dynamic Tradeoff between energy and throughput efficiency that adapts the scheduler priorities to the network context and particularly to the traffic load. Considering the context, the scheduler is able to adjust its behavior in order to maintain high QoS while reducing as much energy as possible. Performance evaluation will show that the proposed solution succeeds to minimize energy consumption better than energy focused scheduler in underloaded context while being able to reach the same spectral efficiency as throughput oriented scheduler in highly loaded context.

1. Introduction

The constant growing number of users which each are more and more demanding in terms of throughput and delay constraints leads us to develop new resource allocation algorithms that increase spectral efficiency while guaranteeing high fairness. In addition, ensuring high Quality of Experience (QoE) can not be reached without offering a good and sustainable mobility that required new resource allocation strategies which provide low energy consumption in order to increase battery lifetime.

Traditional resource allocation strategies used in wireless networks were originally and primarily designed for the wired context. Consequently, these conventional access methods like Round Robin (RR) and Random Access (RA) are not well adapted to the wireless environment and provide very poor throughput. Intensive research efforts have been given in order to propose throughput efficient schedulers and opportunistic approaches have emerged as the best way. The best known is called Maximum Signal to Noise Ratio (MaxSNR) scheduler [1, 2]. It preferably allocates the resources to the user with the most favourable channel conditions at a given time. It takes benefit of multiuser and frequency diversity in order to maximize the system throughput (Figure 1(b)). However users close to the access point have a better average throughput per Resource Unit (RU) than far users. This induces that, with MaxSNR scheduler, close users have statistically more chances to have access to the medium. In consequence, far users will often obtain radio resources after close users making them overpassing their QoS requirement and being unsatisfied. In order to solve this issue, Proportional Fair (PF) and PF-based algorithms have been proposed [3–8]. The basic principle is to allocate resources to a user when its channel conditions are



(a) Throughput/Resource Unit (RU) variation over the time due to multipath fading

(b) Opportunistic system capacity increase

Figure 1: Benefit of opportunistic scheduling strategies on spectral efficiency and system capacity.

the most favourable with respect to its time average. This approach is more fair than MaxSNR since all users have statistically the same probability to access radio resources. Therefore, PF increases the benefits of multiuser diversity which reinforce the opportunistic resource allocation behavior conducting in spectral efficiency increase. However, all these schedulers have a severe lack in terms of energy management.

In order to offer more battery autonomy to users, solutions focusing on energy have been developed. The Power-based Proportional Fairness (PPF) [9] proposes PF-based scheduler that avoids the inefficient allocations (with low SNR) and delays flows that have high average energy consumption. This slightly increases energy efficiency since this gives access to the medium only to users with good SNR and allows always using higher modulation orders that are the most profitable but potentially could segregate users with high traffic load (that will use more radio resources and consequently use more energy). In addition, the best way to minimize energy consumption is not only to optimize the modulation but mainly to maximize the sleep time. The Opportunistic Energy Aware scheduler (OEA) [10] is built on this principle. It exploits active-sleep mode and channel condition together. While other schedulers can potentially activate all users, the OEA limits this number. This allows compressing the transmission time (i.e., active mode), greedy in energy. Considering the channel condition in the allocation process, only allocations with good modulation are also conserved. T-MAC [11] is another strategy that can be considered as an extreme version of OEA. It only schedules a single user by time slots that strongly maximize sleep time but, by losing multiuser diversity benefit, provide lower throughput. All these energy specialized schedulers lack fairness and have limited spectral efficiency. Therefore this limits their scope of usage to underloaded context. Since energy efficiency guarantee must not evade QoS requirement and the system capacity optimization, new approaches must be developed in order to bring together high spectral efficiency, fairness, and energy consumption minimization whatever considered traffic load.

Previously we had proposed a Fairness-Energy-Throughput Optimized Tradeoff Scheduler (FETOT) [12]. This solution tries to provide the best tradeoff between system capacity and energy efficiency while providing fairness. It takes into account the radio condition in order to avoid bad allocation in terms of throughput. A correction factor on the distance is adequately integrated in the algorithm in order to offer the same high fairness considering far and close users like PF. This scheduler is also built to compress the transmission time but, contrary to the OEA, FETOT is able to take a full benefit on the multiuser diversity thanks to a new tradeoff parameter. The result is that FETOT combined the advantages of MaxSNR, PF, and OEA, respectively, on system capacity, fairness, and energy efficiency. However this tradeoff is static and performances can be enhanced making the tradeoff dynamic and always adapted to the context.

In this paper, we propose a new algorithm Dynamic Tradeoff scheduler (DT) that dynamically adjusts its behavior to the traffic load context. In an underloaded system, radio resources are abundant and the system can easily satisfy all users. Consequently, in these contexts, DT detects the surplus of unused radio resources and orients its scheduling strategy to be energy aware. It makes a better usage of multiuser diversity than OEA that allows to preserve more energy than this specialized algorithm even in its scope of usage. In an overloaded system, radio resources are highly valued and system meets high difficulties to satisfy all users. In these contexts, DT detects the lack of available radio resources and orients its scheduling strategy to increase spectral efficiency in order to withstand the load increase. It offers the same system capacity as PF and outperforms MaxSNR. Between these two extreme contexts, DT takes into account the bandwidth usage ratio to smoothly adapt and adjust its energy-throughput tradeoff to the traffic load. Performance evaluation will show that users are always satisfied with fairness as well as PF while always preserving as much energy as possible.

This paper is constructed as follows: Section 2 presents the system description, Section 3 describes the Dynamic Tradeoff

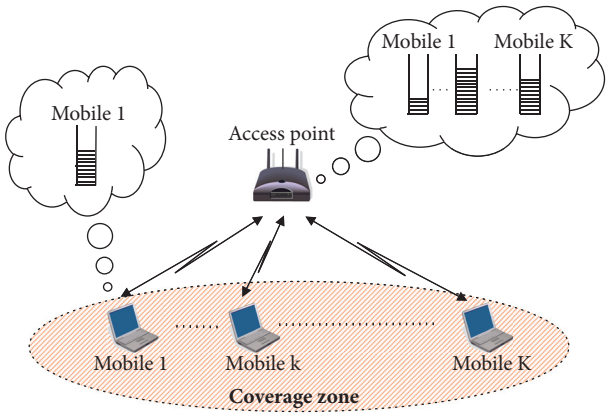


FIGURE 2: Allocation of radio resources among the set of mobiles situated in the coverage zone of an access point.

algorithm, Section 4 shows performances evaluations, and Section 5 is the conclusion.

2. System Description

We focus on the proper allocation of radio resources among the set of mobiles situated in the coverage zone of an access point (Figure 2). We consider a centralized approach. The packets originating from the backhaul network are buffered in the access point which schedules the downlink transmissions. In the uplink, the mobiles signal their traffic backlog to the access point which builds the uplink resource mapping.

We assume that the physical layer is operated using the structure described in Figure 3 which ensures a good compatibility with the OFDM based transmission mode of the IEEE 802.16-2004 [13, 14]. The total available bandwidth is divided into subfrequency bands or subcarriers. The radio resource is further divided into the time domain in frames. Each frame is itself divided into time slots (TS) of constant duration. The time slot duration is an integer multiple of the OFDM symbol duration. The number of subcarriers is chosen so that the width of each subfrequency band is inferior to the coherence bandwidth of the channel. Moreover, the frame duration is fixed to a value much smaller than the coherence time (inverse of the Doppler spread) of the channel. With these assumptions, the transmission on each subcarrier is subject to flat fading with a channel state that can be considered static during each frame.

The elementary Resource Unit (RU) is defined as any (subcarrier, time slot) pair. Each of these RUs may be allocated to any mobile with a specific modulation order. Transmissions performed on different RUs by different mobiles have independent channel state variations [15]. On each RU, the modulation scheme is QAM with a modulation order adapted to the channel state between the access point and the mobile to which it is allocated. This provides the flexible resource allocation framework required for opportunistic scheduling.

The system is operated using time division duplexing with four subframes: the *downlink feedback subframe*, the *downlink data subframe*, the *uplink contention subframe*, and the *uplink data subframe*. The uplink and downlink data subframes are used for transmission of user data. In the downlink feedback subframe, the access point sends control information towards its mobiles. This control information is used for signalling to each mobile to which RU(s) has been allocated in the next uplink and downlink data subframes, the modulation order selected for each of these RUs, and the recommended emission power in the uplink. In the uplink contention subframe, the active mobiles send their current traffic backlog and information elements such as QoS measures and transmit power. The uplink contention subframe is also used by the mobiles for establishing their connections. This frame structure supposes a perfect time and frequency synchronization between the mobiles and the access point as described in [16]. Therefore, each frame starts with a long preamble used for synchronization purposes. Additional preambles may also be used in the frame.

3. Dynamic Tradeoff Scheduler

The DT scheduling algorithm relies on weights that set the dynamic priorities for allocating the radio resources. These weights are built in order to satisfy three major objectives that are explained separately in the following: system capacity maximization, fairness, and energy consumption minimization. Then we will present a calibration of the function that adds an ability for the scheduler to adequately tune the multiuser diversity usage considering the context and relative objectives, merging previous weights in a balanced DT solution.

3.1. System Throughput Maximization. The DT scheduler optimizes the system throughput in a MAC/PHY opportunistic approach. Data integrity requirements of users are enforced considering each user independently, adapting the modulation and the transmit power to the user specific channel state. At each frame allocation, the scheduler computes the maximum number of bits $q_{k,n}$ that can be transmitted in a TS of subcarrier n if assigned to user k while keeping below its Bit Error Rate target ($BER_{target,k}$), for all k and all n:

$$q_{k,n} \leq \left\lfloor \log_2 \left(1 + \frac{3P \times T_s \times (1/d_k)^{\beta} \times \alpha_{k,n}^2}{2N_0 \left[\mathrm{erfc}^{-1} \left(BER_{target,k}/2 \right) \right]^2} \right) \right\rfloor, \quad (1)$$

where P is the transmission power, N_0 is the spectral density of noise, T_s is the OFDM symbol duration, d_k is the distance to the access point of the user k, and $\alpha_{k,n}^2$ represents the flat fading experienced by this user on subcarrier n. In the following, $\alpha_{k,n}$ is Rayleigh distributed with an expectation equal to unity. The exponent β corresponds to the experienced path loss and goes from 2 to 4 considering environment density level. Due to multipath fading, the potential number of bits that a user can transmit on a RU will fluctuate around this value over the time.

We further assume that the supported QAM modulation orders are limited such that q belongs to the set $S = \{0, 2, 4, \ldots, q_{max}\}$. Hence, the maximum number of bits $m_{k,n}$

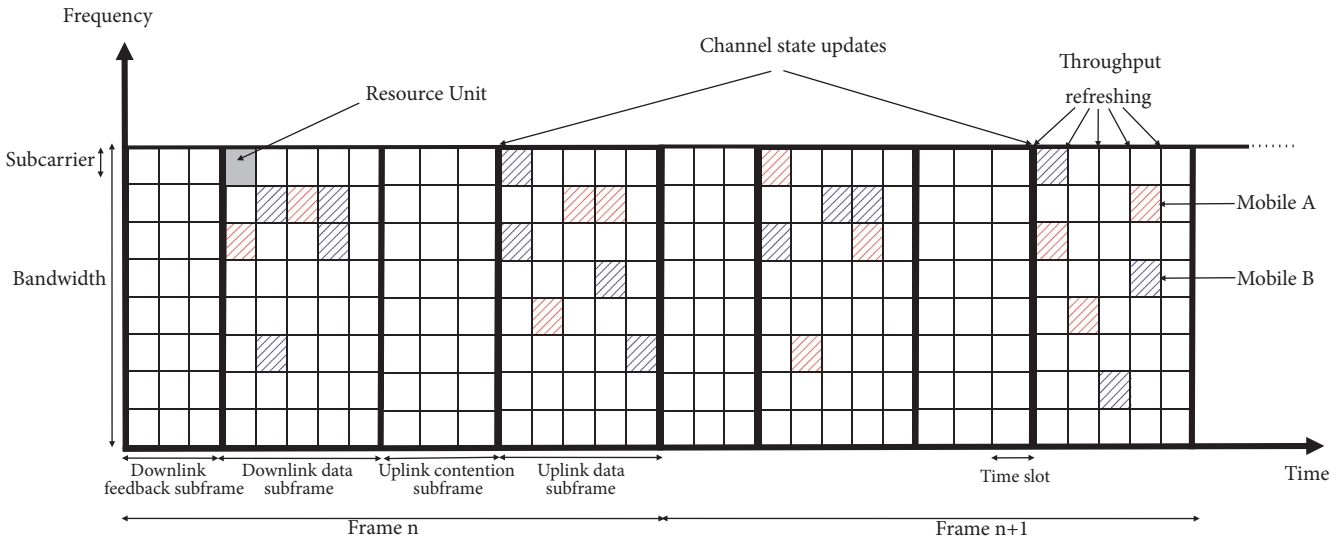


Figure 3: DT frame structure in TDD mode.

that will be transmitted on a TS of subcarrier n if this RU is allocated to the user k is

$$m_{k,n} = \max \{q \in S,\ q \leq q_{k,n}\}. \quad (2)$$

MaxSNR based schemes allocate the RU to the user which has the greatest $m_{k,n}$ values. This strategy maximizes the system capacity at short time scale but is highly unfair considering users far to the access point that are often delaying out of their delay requirement. In order to provide more fairness considering users locations while preserving the system throughput maximization, a fairness parameter is introduced in DT.

3.2. Fairness Guarantee. DT integrates in its scheduling process the fairness parameter proposed in [8]. Called "Compensation Factor" (CF_k), this parameter takes into account the current path loss impact on the average achievable bit rate of mobile k:

$$CF_k = \frac{b_{ref}}{b_k}. \quad (3)$$

b_{ref} is a reference number of bits that may be transmitted on a subcarrier considering a reference free space path loss a_{ref} for a reference distance d_{ref} to the access point and a multipath fading equal to unity:

$$b_{ref} = \log_2 \left(1 + \frac{3P_{max} \times T_s \times a_{ref}}{2N_0 \left[\text{erfc}^{-1}\left(BER_{target}/2\right)\right]^2}\right). \quad (4)$$

b_k represents the same quantity but considering a distance d_k to the access point:

$$b_k = \log_2 \left(1 + \frac{3P_{max} \times T_s \times a_{ref} \times \left(d_{ref}/d_k\right)^{\beta}}{2N_0 \left[\text{erfc}^{-1}\left(BER_{target}/2\right)\right]^2}\right), \quad (5)$$

with β the experienced path loss exponent.

Adequately combining and taking into account both $m_{k,n}$ and CF_k in the allocation process ($m_{k,n} * CF_k$), DT considers all mobiles virtually at the same position in the scheduling decision. CF_k adequately compensates the lower spectral efficiency of far mobiles bringing high fairness in the allocation process. An equal throughput can be provided to each mobile while keeping the MaxSNR opportunistic scheduling advantages thanks to the $m_{k,n}$ parameters which take into account the channel state. Moreover, in contrast with MaxSNR which satisfy much faster the mobiles which are close to the access point, DT keeps more mobiles active but with a relatively low traffic backlog. Satisfaction of delay constraints is more uniform and, by better preserving the multiuser diversity, a more efficient usage of the bandwidth has been highlighted. This jointly ensures fairness and system throughput maximization. If two mobiles have an equal priority for RU, this one is given to the mobile which has the highest buffer occupancy further strengthening fairness. At this step, DT optimizes the throughput and guarantee high fairness but highly suffers of an inefficient energy management as the same level as PF. In order to provide energy consumption minimization while preserving the system throughput maximization and fairness, an energy parameter is introduced.

3.3. Energy Consumption Minimization. The third major objective of the DT is to provide efficient energy management in addition to the system throughput optimization and fairness. Existing opportunistic resource mapping (as MaxSNR or PF for example) basically overexploits multiuser diversity which induces horizontal allocation. Indeed, due to flat fading during a frame, often the same user strictly experienced the greatest channel condition on each TS of a given subcarrier. Consequently, with classical opportunistic schedulers, the same user often receives all the TS of a subcarrier and needs to stay in active mode during a long time. We can potentially have one different selected user on each available subcarrier. Consequently, during all TS, many

selected users can not be set in sleep mode. They consume a lot of power to transmit few bit during a long time (with many allocated TS but on few subcarriers).

The DT scheduler integrates a modified version of the energy efficient OEA solution [10], keeping its energy benefit without its fairness and system capacity failure. Energy consumption is minimized particularly by increasing the sleeping mode duration. In order to achieve this goal, DT extends the classical OEA opportunistic cross-layer design to obtain a new vertical opportunistic resource mapping. When a user is in active mode, DT tries, like OEA, to benefit from its activation in order to compress its time of activity and to transmit more bit per "used" TS. Like this, DT allows to significantly increase sleeping mode duration and energy preservation. Originally, OEA scheduler computed an "energy transmission cost" (ETC_k) parameter (in Watt). It is based on the energy cost of user k to transmit on a RU:

$$ETC_k = A_k * Cn_k + (1 - A_k) * (C_k + Cn_k). \quad (6)$$

When the user k is in active mode, $A_k = 1$ otherwise $A_k = 0$ (i.e., sleep mode). In addition, Cn_k and C_k are two constants (in Watt). C_k represents the energy needed to wake up the user k from the sleep mode to the active mode. Cn_k represents the energy needed to transmit on a n^{th} allocated subcarrier. The energy cost to transmit on the first RU (C_k) is higher than the cost to transmit on n^{th} (Cn_k) since the cost to move to sleep mode to active mode and transmit is greatly higher than just transmit some supplementary bits while user is already active.

ETC_k is used in OEA scheduler but has the negative side effect to highly reduce the usage done of the multiuser diversity. This drastically and negatively impacts the OEA system capacity optimization. In order to keep its energy minimization properties while fixing this throughput issue, DT integrates a modified ETC_k parameter that we called "Throughput-Energy Tradeoff" parameter TET_k:

$$TET_k = A_k * Cn_k + (1 - A_k) * \left(\frac{C_k}{MD} + Cn_k\right), \quad (7)$$

where MD is a multiuser diversity factor. The higher MD is, the more the system increases the number of active users at the same time, intensifying the multiuser usage and consequently the global system throughput at the expense of the energy consumption (infinite MD value makes TET_k constant and induces DT similar to a PF resource allocation). On the contrary, low MD value makes DT decreasing the number of active users at the same time, reducing energy consumption at the expense of the multiuser diversity usage that provides a resource allocation close to OEA scheduling (excepting that this version is strongly more fair due to Section 3.2). After large performance evaluation studies we found that $MD = 10$ provides a very efficient static tradeoff between energy consumption minimization and spectral efficiency. These works had led to a proposition of a new scheduler called FETOT in [12]. It allowed making an adequate usage of the multiuser diversity in order to provide the same system capacity as MaxSNR, same fairness as PF, and an energy minimization very close to the OEA results. However, we are convinced that the usage of a static MD value is not optimal. Even if FETOT provides a very good static overall tradeoff, this can be highly improved with a solution able to adapt and tune the MD (and consequently the tradeoff) to the network traffic load context. Indeed, in very low traffic load context, energy minimization must be the only objective. With the increase of the traffic load, more attention must be done on spectral efficiency in adequate tradeoff. In high traffic load, to improve spectral efficiency becomes the primary goal in order to continue to satisfy users and energy minimization priority must be relegated. The main contribution of this paper is to propose a new scheduler that combined all previously described parameters and used a dynamic MD parameter to adapt priority to the context.

3.4. DT Merging of Priorities. The DT scheduler allocates the radio resource n to the mobile k that has the greatest $DT_{k,n}$ value such as

$$DT_{k,n} = \frac{m_{k,n} * CF_k}{A_k * Cn_k + (1 - A_k) * (Cn_k/MD + Cn_k)} \quad (8)$$

Taking into account $m_{k,n}$ allows optimizing system capacity avoiding unprofitable radio resource allocation, CF_k allows staying fair in the allocation process regarding user location, and the other parameter allows fighting versus energy waste. Particularly, by adjusting the multiuser diversity usage thanks to good function of MD, DT could select the minimum number of users per timeslot to have a good energy efficiency while respecting the QoS requirements. However when the system is more loaded, DT could increase the multiuser diversity thanks to a higher value of MD in order to obtain a better spectral efficiency to support the load.

3.5. Study of the Multiuser Diversity Factor. The multiuser diversity has an important impact on the load resistance and on the energy consumption. Finding an efficient way to adapt its usage to the context thanks to a well-tuned MD factor is challenging:

(i) The first step would be to determine the extremes values of the MD which correspond the best to extreme configurations: when the system is clearly underloaded, the only concern is the energy consumption, and when the system is largely overloaded, the main focus has to be on the QoS requirements.

(ii) The second step is to find a smooth and adequate transition between those two extremes values based on adapted inputs.

3.5.1. Study of Different Static MD Values in order to Detect the More Efficient in Extreme Scenarios. Figure 4 shows the performances of preliminary versions of DT using static value of MD factor. For different traffic loads it shows the energy transmission cost per bit (Figure 4(a)), the spectral efficiency (Figure 4(b)), the bandwidth usage ratio (bandwidth usage ratio of the number of allocated resource units divided by the total number of radio resource units in the system, in average, per frame) (Figure 4(c)), and the packet delay (Figure 4(d)).

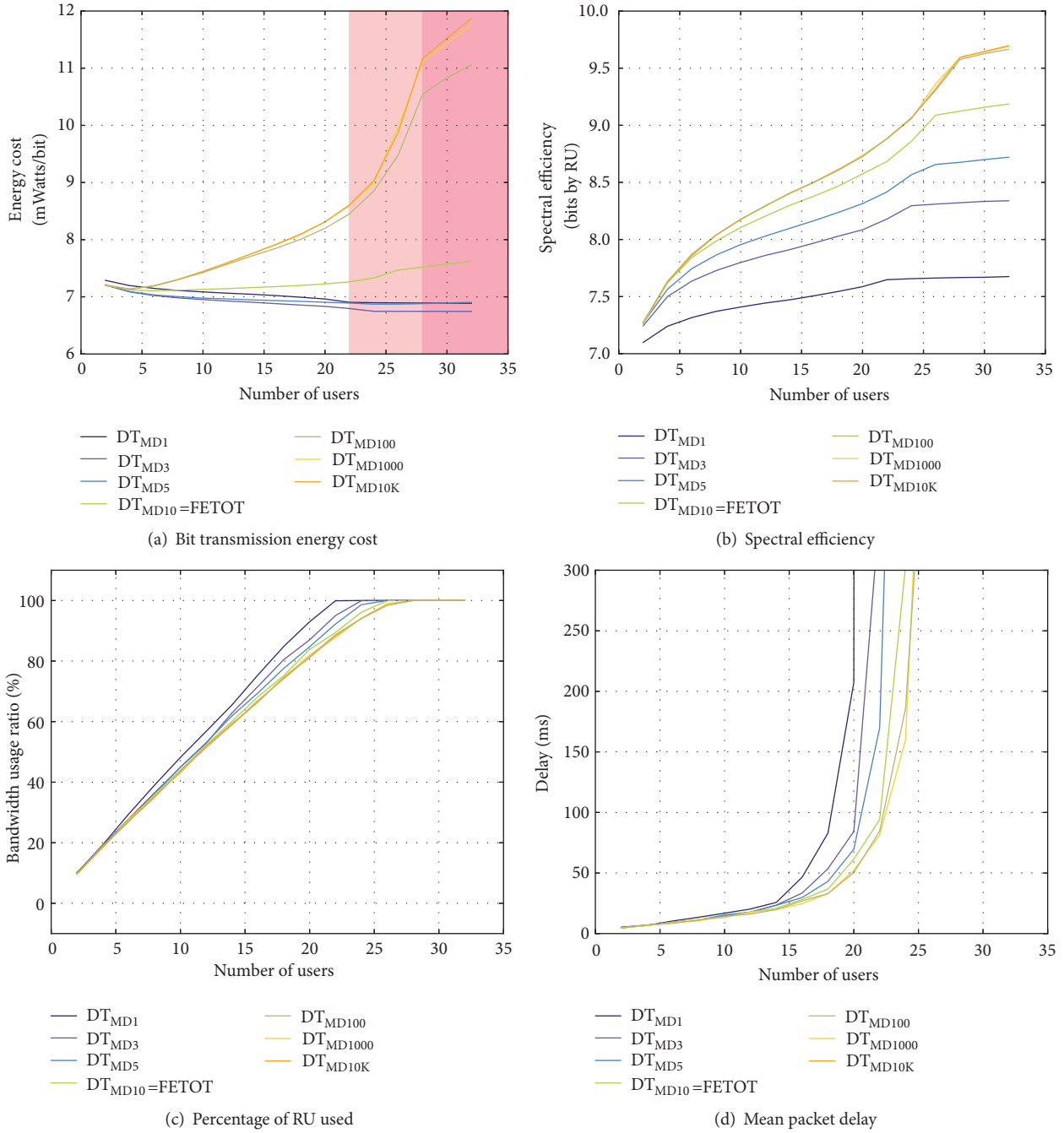


FIGURE 4: System capacity and spectral efficiency study obtained with different static *MD* values.

In a noncongested system (i.e., when delay and bandwidth usage ratio are low, here with a number of users < 15 users), the focus should exclusively be put on the energy efficiency. As we can see in Figure 4(a) a value too big of *MD* (>10) induces excessive consumption due to several users simultaneously active on same time slots. However choosing the smallest value is not a good option either. Indeed if the *MD* is too small, opportunistic behavior is drastically reduced, and the spectral efficiency (Figure 4(b)) is not good enough to evacuate the necessary amount of information in a short time. Even if the scheduler could appear to be more energy efficient due to a drastically limited number of active user at the same time, it is not at long time scale since users will transmit during longer periods due to very low spectral efficiency. Consequently, in extreme and very low loaded context, $MD = 3$ seems to be the most adequate value in order to reach the minimization energy consumption objective (Figure 4(a)).

In a congested system (i.e., when delay is high, bandwidth usage ratio very close or equal to 100 %, here with a number of users > 20 users), the focus should exclusively be put on the spectral efficiency since the priority is to maintain a good level of QoS. Concerning bandwidth usage ratio, Figure 4(c) underlines that all *MD* values superior or equal to 100 allow better withstanding extreme traffic loads providing same spectral efficiency (Figure 4(b)) and best delays (Figure 4(d)). However, having a look at the energy efficiency (Figure 4(a)), we notice a slight advantage to $MD = 100$ over superior value that drive us to consider MD value around 100 as the most adequate values in this extreme highly loaded context.

3.5.2. Dynamic MD Function Calibration. Originally in previous works (FETOT) [12], we show that a fixed *MD* value set at 10 could represent an average good tradeoff. However it is not the best suitable solution for extreme cases as shown above. Adaptive solution can be developed to outperform FETOT in those situations with a dynamic usage of multiuser diversity that can be obtained thanks to a dynamic *MD* according to the context and particularly to the traffic load (that should define the scheduler priorities/goals). We propose in DT to define *MD* as an increasing function of the bandwidth usage ratio. This parameter simply and accurately informs on the state of the system and on the difficulties or not for the scheduler to maintain the QoS to user. Low bandwidth usage ratio values, inducing low MD value (MD=3), underline to DT to focus on energy. High bandwidth usage ratio, which required to focus on spectral efficiency, will induce high MD value (MD around 100) that will improve multiuser diversity usage. In order to link these two extremes, we proposed an heuristic:

$$MD_x = C + \beta x^{\alpha} \tag{9}$$

where x is the bandwidth usage ratio, C is a constant that defined the starting value of the MD function when the system is underloaded, and β corresponds to the other extreme when the system is overloaded. In the following, we set C to 3 and β to 100 according to Section 3.5.1. The parameter α allows setting the reactivity of the function to the traffic load variation. An appropriate calibration of α is highly important.

3.5.3. Studies of α. It is important that the MD function gives low values when bandwidth usage ratio is low. Since QoS is easily guaranteed, DT has to limit the multiuser diversity usage in order to focus on energy consumption minimization. When traffic load increases, MD function must increase its output in adequation with the difficulties met by the scheduler to conserve high QoS. Figure 6 represents *MD* variation depending on traffic load (measured with the bandwidth usage ratio) for different value of α. As we can notice in this figure, the α parameter directly impacts how this *MD* value will increase from the traffic load. If α is set equal to 1, the *MD* function is linear and multiuser diversity usage will be constantly increased with the bandwidth usage ratio. It is not optimal since no QoS difficulties are met with low bandwidth usage ratio values and problems are experienced only when we they come closer to 100%. On the contrary, high value of α (typically $\alpha = 40$) makes MD function growing too late in order to satisfy the QoS. Indeed, in realistic scenario, with the variability of the traffic, even with an average measurable bandwidth usage ratio inferior to 100% but close to this limit, temporary short term congestion can occur decreasing QoS. In these cases multiuser diversity usage must be intensified and this can be done by DT scheduler if MD function is well calibrated. Detecting when MD function must begin to grow is a difficult task and relies on the elasticity of the traffic. In order to define the best value of α we decide to evaluate all possible α values performances in extensive simulations (Figure 5). We used the more realistic traffic models (MPEG-4, Voice, Videoconference, etc.) that highly complicate the task of the schedulers.

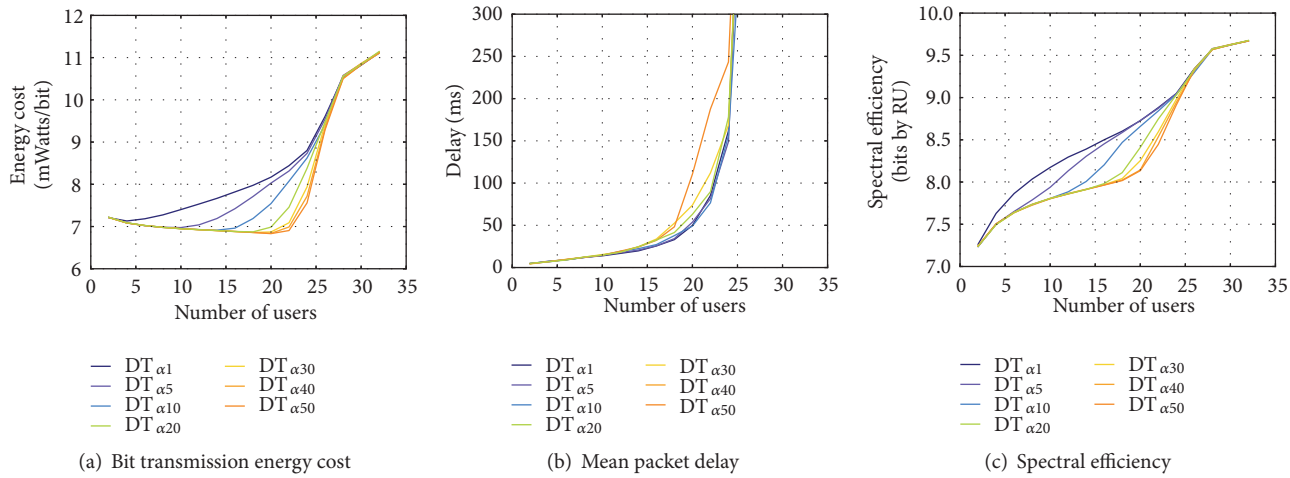
Figure 5(a) shows the impact of α regarding the energy efficiency. It is the most important objective for all the left part of the figure since delays are very low (Figure 5(b)). Choosing a small α value such as 1 has a very bad impact on the energy consumption that increases quickly. This is due to the fact that the algorithm is too much reactive on the traffic load increase, uselessly exploits a supplementary of multiuser diversity, and futilely tries to focus on the QoS. Indeed, the same very good values of delay are obtained for all α values inferior or equal to 20 (Figure 5(b)). Higher values than 20 increase MD too late and provide worst delay, and lower values provide same delay but more energy consumption. If we consider that a final goal is to be able to maintain the best QoS while minimizing energy as much as possible, the most suitable α value for the MD function is 20.

4. Performance Evaluations

4.1. Context and Simulation Setup. Performance evaluation results are obtained using discrete event simulations. In the simulations, we assume C_k and Cn_k are fixed, respectively, equal to 110.2 mW and 46.8 mW, for all k in accordance with measured hardware consumption. The BER target is taken equal to 10^{-3}. We also consider that all users run realistic Variable Bit Rate applications [17] that generate high volume of data with high sporadicity and require tight delay constraints which substantially complicate the task of the scheduler. In order to study the influence of the distance of users on the scheduling performances, a first half of mobiles is situated close to the access point and has a mean $m_{k,n}$ equal to 8 bits. The second half is more far from the access point such as their mean $m_{k,n}$ equal to 6 bits. All performance criteria are done studying the influence of the traffic load. This one varies adding users 2 by 2 (each time, 1 close user and 1 far user).

4.2. Simulation Results

4.2.1. Spectral Efficiency and Throughput. Figure 7(a) shows the spectral efficiency obtained with each scheduler for different traffic load in the system. Since RR does not take into account radio conditions and therefore is not opportunistic, it does not take any advantage of multiuser diversity and its spectral efficiency is constant and low. State-of-the-art energy focused schedulers (T-MAC, OEA) drastically limit the usage



(a) Bit transmission energy cost (b) Mean packet delay (c) Spectral efficiency

Figure 5: Study of α on system capacity and delay.

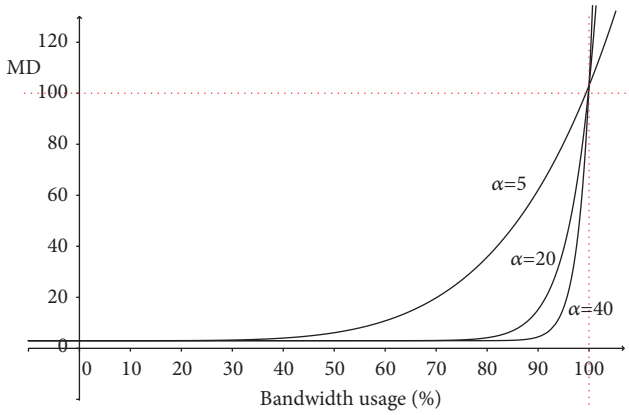


Figure 6: Variation of *MD* according to α value.

of the multiuser diversity in their allocation process offering slightly better results. On the contrary MaxSNR, highly opportunist, provides a large gain. However, as explained in Section 3.1, MaxSNR has a lack on fairness and is not able to take all the benefits of the multiuser diversity and is highly outperformed by PF. FETOT makes a tradeoff between energy and throughput providing spectral efficiency results close to MaxSNR.

Thanks to its dynamic MD parameter based on the bandwidth usage ratio, DT has lesser spectral efficiency in low traffic load context using a moderate usage of the multiuser diversity focusing its efforts on energy. However, when it becomes necessary (i.e., system approach congestion (Figure 7(b))), its MD factor adequately increases and raises the DT usage of the multiuser diversity improving the spectral efficiency at the same level as PF reaching the same overall maximum system capacity (Figure 7(b)).

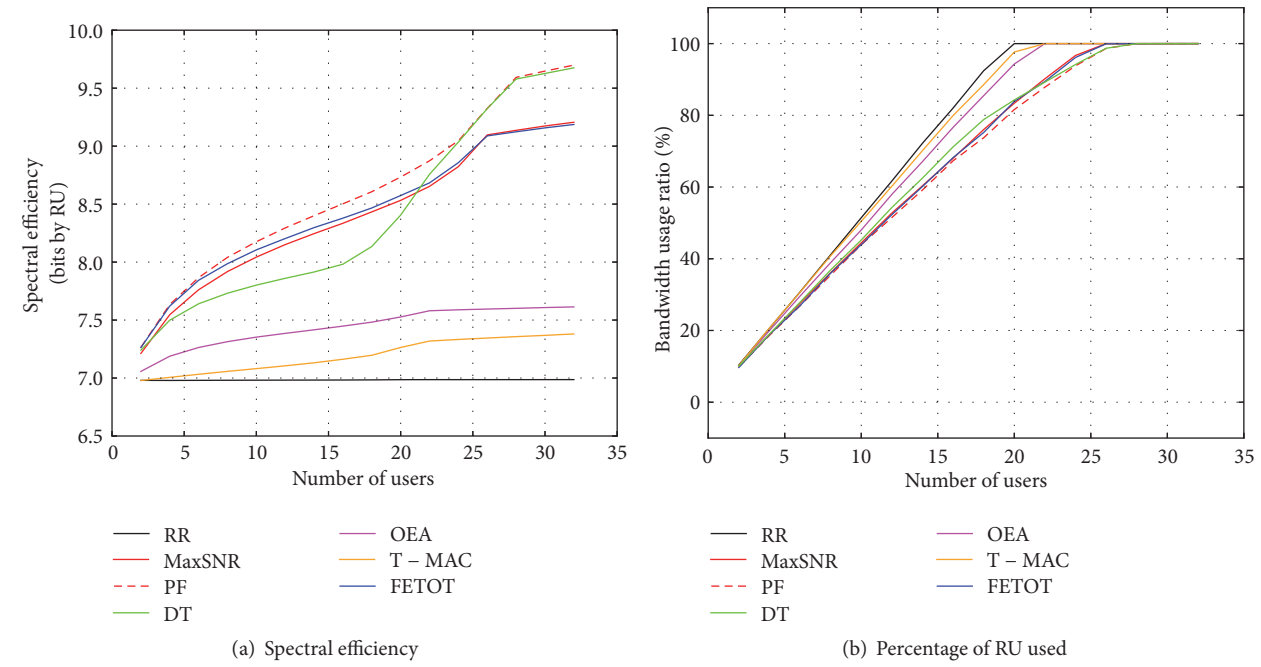
4.2.2. Delay and Fairness. A major QoS key performance indicator is the latency. Figure 8(a) represents the mean packet delay experienced in the system in milliseconds according to the number of users showing the traffic load. We can notice that 2 groups emerged:

(i) First, RR, T-MAC, and OEA have the worst results. Having a low spectral efficiency (Figure 7(a)), they failed to support a large amount of traffic load with good QoS.

(ii) Secondary, MaxSNR, PF, FETOT, and DT are able to better sustain higher load increase with acceptable delay.

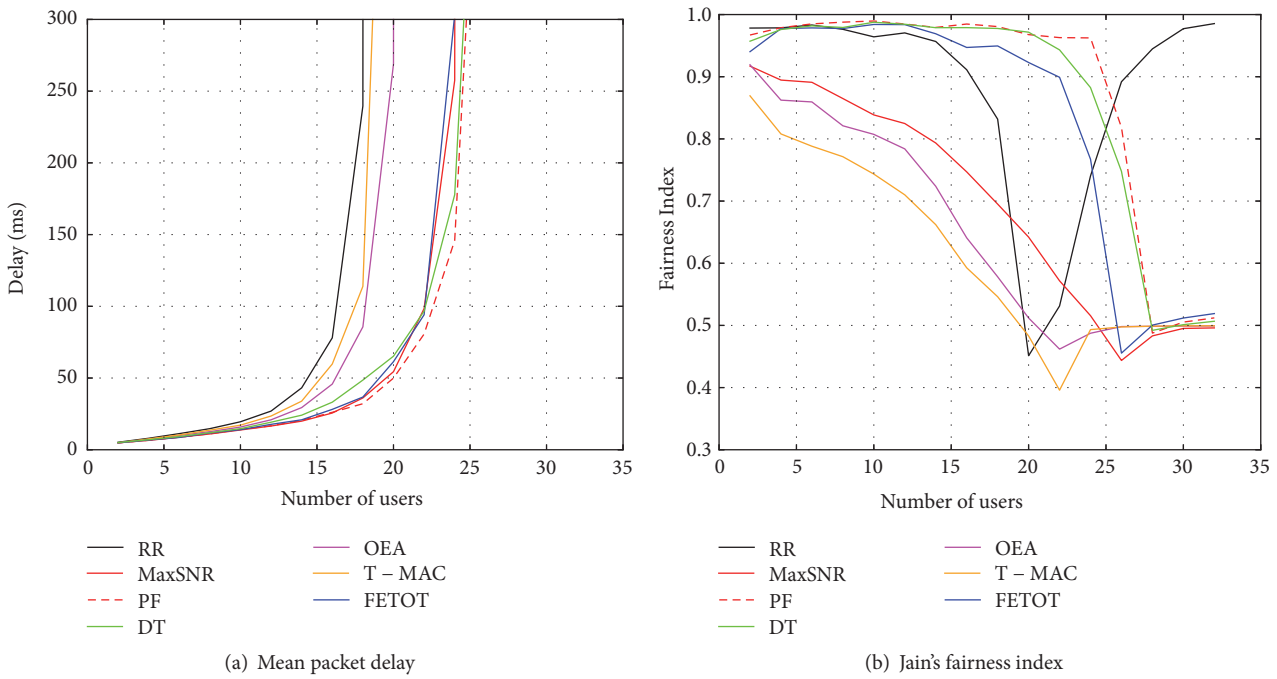
Figure 8(b) focuses on fairness computed thanks to the Jain's fairness index applied on mean packet delay. (With n defined as the current number of users in the system, Jain's fairness index can vary between $1/n$ and 1, respectively, associated with the most unfair scheduling to the most fair.) T-MAC, OEA, and MaxSNR significantly penalize user far from the access point and have decreasingly fairness results with the traffic load increase. On the contrary more fair solutions as RR, PF, FETOT, and DT achieve to reach an high fairness. Note that after congestion fairness cannot be guaranteed since global mean packet delay is infinite. Consequently the capacity of these schedulers to maintain high fairness is directly related to their spectral efficiency and when they have no longer available radio resources, fairness disappeared. (Note that when the system capacity is highly overpassed, Jain's fairness index can increase due to comparison of close but unacceptable huge value of delay.) Respectively, the most fair schedulers are consequently PF, DT, FETOT, RR, MaxSNR, OEA, and T-MAC.

4.2.3. Energy Consumption. Figure 9 shows the abilities of each scheduler to be energy efficient. RR widely provides the worst results. This is due to its nonopportunistic behavior that makes possible highly inefficient resource allocation in terms of bit per RU and corresponding to a significant energy and RU wastes. In addition, due to a cycling user selection, many users can be simultaneously activated (Figure 9(d)) increasing again the energy waste (with more of 20 users, the system is overloaded and RR fails to provide the sufficient amount of RUs required by each user; they are often forced to stay in sleep mode even with data to transmit due to the lack of RUs; more often in forced sleep mode, the users consumed



(a) Spectral efficiency (b) Percentage of RU used

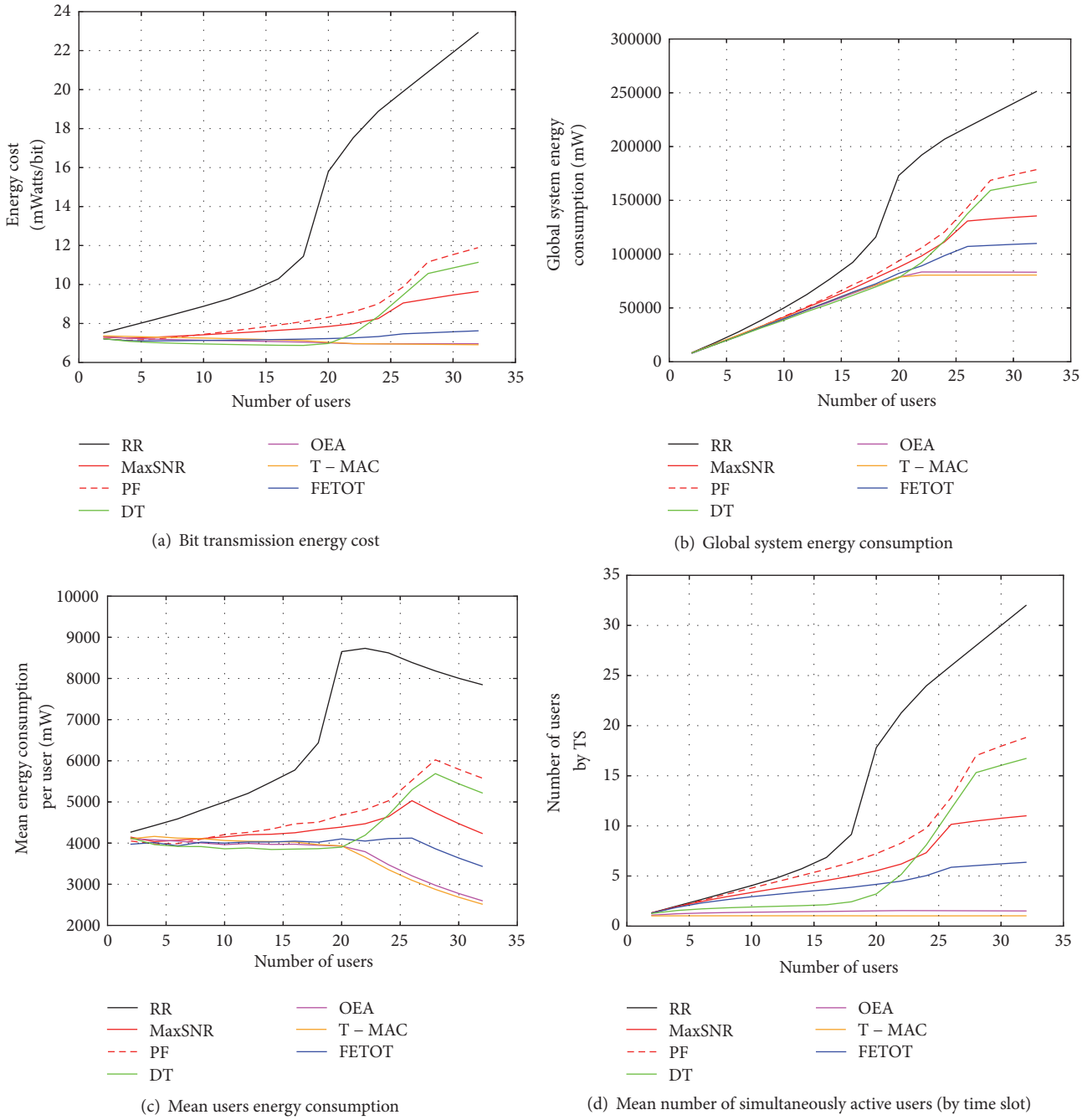
Figure 7: Schedulers system capacity study.



(a) Mean packet delay (b) Jain's fairness index

Figure 8: Schedulers abilities to guarantee high Quality of Service.

less energy over the time; this explains why, with more than 20 users, the RR curve decreases (Figure 9(c))) since more users pay the high transmission activation price C_k. Limiting the usage of the multiuser diversity to a low value whatever the context (Figure 9(d)), T-MAC and OEA provide very good energy consumption (Figure 9(c)). Note that these good results must be put into perspectives. Indeed, those solutions continue to search to minimize energy consumption even when traffic loads increase and this stubborn behavior conducts these schedulers to quickly reach congestion (Figure 7(b)) with high delay (Figure 8(a)). On the contrary, PF, fully exploiting the multiuser diversity (Figure 9(d)), consumes more energy (Figure 9(c)) but less than RR thanks to strongly better spectral efficiency. Focusing on MaxSNR, its energy results are slightly better than PF. Indeed, this scheduler has a tendency to segregate a part of users (far from



(a) Bit transmission energy cost

(b) Global system energy consumption

(c) Mean users energy consumption

(d) Mean number of simultaneously active users (by time slot)

Figure 9: Scheduler energy efficiency.

the access point) and consequently obtains reduced benefits of multiuser diversity usage. This is a weakness in order to improve spectral efficiency but an advantage to increase user sleep duration. FETOT provides better energy efficiency than MaxSNR, very close to OEA, when the traffic load is low (below 20 users). Using an adequate static tradeoff, energy consumption stays reasonable even when traffic reaches higher value but, when necessary and contrary to OEA, this is less done at the expense of spectral efficiency that stays close to MaxSNR (Figure 7(a)).

Considering underloaded contexts (number of users inferior to 20), guaranteeing high Quality of Service (QoS) is easily achievable by DT (Figure 8(a)) due to large surplus of available radio resource units (Figure 7(b)) and focus should be put on energy rather than system throughput. Figures 9(a), 9(b), and 9(c) underline that DT is the scheduler that better optimizes the multiuser diversity usage in this context. Few users are simultaneously activated per time slot (close to T-Mac and OEA (Figure 9(d))) but, contrary to the specialized state-of-the-art energy aware schedulers, DT provides an

adequate spectral efficiency forbidden inefficient resource allocation. This combination allows to better compress the transmission time and therefore better optimize energy consumption. Considering highly loaded context (number of users superior to 20), the lack of available radio resources (Figure 7(b)) required that schedulers focus on system capacity in order to preserve QoS. Energy consumption must become a lesser priority. In this context, DT behavior slightly sacrifices energy in order to sustain the network viability and then favours high spectral efficiency that reaches values close to PF (Figure 7(a)) which provides acceptable delay as long as possible (close to PF).

5. Conclusion

Reaching both low system energy consumption and high spectral efficiency is very difficult tasks in wireless network. Specialized solutions as MaxSNR, PF, or T-MAC have been well designed to well answer one of these criteria failing to the second. Other solutions propose static tradeoffs that provide good average results on these two metrics without success outperforming specialized scheduler in their focused domain. In this paper, we underline that the network objectives must be dependant of the context and particularly to the traffic load. In underloaded context, guaranteeing high Quality of Service (QoS) is easily achievable due to large surplus of available radio resources and the focus must be put on energy rather than system throughput. On the contrary, in a highly traffic loaded context, the lack of available radio resources required that resources allocation algorithms focus on system capacity in order to preserve QoS and satisfy users; thus energy consumption must become lesser important. The main contribution of this paper is to propose a Dynamic Tradeoff (DT) scheduler able to tune its priorities and the multiuser usage benefit according to the network traffic load context. It provides a better energy efficiency than specialized energy aware scheduler when it is feasible while providing the same spectral efficiency and delays as throughput oriented scheduler when it is required. This is achieved with a fairness special attention that is also guarantee. Future works could focus on other metrics like mean packet delay in order to adapt the multiuser usage to different contexts.

Data Availability

The data used to support the findings of this study are included within the article.

Disclosure

Their research did not receive specific funding but was performed as part of the employment of the authors: IRT b-com.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, pp. 1747–1758, 1999.

[2] X. Wang and W. Xiang, "An OFDM-TDMA/SA MAC protocol with QoS constraints for broadband wireless LANs," *Wireless Networks*, vol. 12, no. 2, pp. 159–170, 2006.

[3] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1277–1294, 2002.

[4] H. Kim, K. Kim, Y. Han, and J. Lee, "An efficient scheduling algorithm for QoS in wireless packet data transmission," in *Proceedings of the IEEE Int. Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 2244–2248, Pavilhao Altantico, Lisboa, Portugal.

[5] W. Anchun, X. Liang, Z. Shidong, X. Xibin, and Y. Yan, "Dynamic resource management in the fourth generation wireless systems," in *Proceedings of the ICCT 2003 - International Conference on Communication Technology*, vol. 2, pp. 1095–1098, Beijing, China.

[6] P. Svedman, S. K. Wilson, and B. Ottersten, "A QoS-aware proportional fair scheduler for opportunistic OFDM," in *Proceedings of the 2004 IEEE 60th Vehicular Technology Conference, VTC2004-Fall: Wireless Technologies for Global Security*, pp. 558–562, USA, September 2004.

[7] H. Kim, K. Kim, Y. Han, and S. Yun, "A proportional fair scheduling for multicarrier transmission systems," in *Proceedings of the IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004*, pp. 409–413, Los Angeles, CA, USA.

[8] C. Gueguen and S. Baey, "A fair MaxSNR scheduling scheme for multiuser OFDM wireless systems," in *Proceedings of the 2009 IEEE 20th Personal, Indoor and Mobile Radio Communications Symposium, PIMRC 2009*, Japan, September 2009.

[9] M. R. del Olmo, R. Torrea-Duran, A. G. Orozco-Lugo, and M. Moonen, "Energy-efficient user scheduling algorithm for lte networks," in *Proceedings of the 5th Joint WIC/IEEE SP Symp. Inf. Theory Signal Process*, p. 1, 2015.

[10] C. Gueguen, "Opportunistic Energy Aware Scheduler for Wireless Networks," in *Proceedings of the 2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, Dresden, Germany, June 2013.

[11] T. van Dam and K. Langendoen, "An adaptive energy-efficient MAC protocol for wireless sensor networks," in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems (SenSys '03)*, pp. 171–180, New York, NY, USA, November 2003.

[12] C. Gueguen and M. Manini, "Fairness-Energy-Throughput Optimized Trade-off in Wireless Networks," in *Proceedings of the 2018 International Conference on Computing, Networking and Communications (ICNC)*, pp. 815–821, Maui, HI, March 2018.

[13] IEEEStd802.16-2004, *IEEE standard for local and metropolitan area networks, part 16: Air interface for fixed broadband wireless access systems*, 2004.

[14] C. Hoymann, "Analysis and performance evaluation of the OFDM-based metropolitan area network IEEE 802.16," *Computer Networks*, vol. 49, no. 3, pp. 341–363, 2005.

[15] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared

wireless link," *IEEE Communications Magazine*, vol. 39, no. 2, pp. 150–153, 2001.

[16] J. van de Beek, P. Borjesson, M. Boucheret et al., "A time and frequency synchronization scheme for multiuser OFDM," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 11, pp. 1900–1914.

[17] S. Tanwir and H. Perros, "A survey of VBR video traffic models," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1778–1802, 2013.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 2103868, 10 pages
https://doi.org/10.1155/2018/2103868

Research Article

A Practical Perspective on 5G-Ready Highly Dynamic Spectrum Management with LSA

Pavel Masek,[1] **Evgeny Mokrov**,[2] **Krystof Zeman**,[1] **Aleksey Ponomarenko-Timofeev**,[3] **Alexander Pyattaev**,[3] **Sergey Nesterov**,[4] **Sergey Andreev**,[3] **Jiri Hosek**,[1] **Konstantin Samouylov**,[2] **and Yevgeni Koucheryavy**[3]

[1]*Brno University of Technology, Czech Republic*
[2]*Peoples' Friendship University of Russia (RUDN University), Russia*
[3]*Tampere University of Technology, Finland*
[4]*St. Petersburg Polytechnic University (SPbPU), Russia*

Correspondence should be addressed to Jiri Hosek; hosek@feec.vutbr.cz

Received 13 April 2018; Revised 12 July 2018; Accepted 24 July 2018; Published 2 September 2018

Academic Editor: Ivan Marsa-Maestre

A diversity of wireless technologies will collaborate to support the fifth-generation (5G) communication networks with their demanding applications and services. Despite decisive progress in many enabling solutions, next-generation cellular deployments may still suffer from a glaring lack of bandwidth due to inefficient utilization of radio spectrum, which calls for immediate action. To this end, several capable frameworks have recently emerged to all help the mobile network operators (MNOs) leverage the abundant frequency bands that are utilized lightly by other incumbents. Along these lines, the recent Licensed Shared Access (LSA) regulatory framework allows for controlled sharing of spectrum between an incumbent and a licensee, such as the MNO, which coexist geographically. This powerful concept has been subject to several early technology demonstrations that confirm its implementation feasibility. However, the full potential of LSA-based spectrum management can only become available if it is empowered to operate *dynamically* and at high space-time-frequency granularity. Complementing the prior efforts, we in this work outline the functionality that is required by the LSA system to achieve the much needed flexible operation as well as report on the results of our respective live trial that employs a full-fledged commercial-grade cellular network deployment. Our practical results become instrumental to facilitate more dynamic bandwidth sharing and thus promise to advance on the degrees of spectrum utilization in future 5G systems without compromising the service quality of their users.

1. Introduction

The fifth-generation (5G) wireless systems aim to decisively advance on the levels of spectral and energy efficiency, user-experienced throughput, as well as communication latency and reliability. They prepare to rely on leveraging extremely high frequency (i.e., mmWave) spectrum bands, employing massive Multiple-Input Multiple-Output (MIMO) techniques, as well as deploying increased numbers of small cells with various sizes and across different frequencies. However, the use of mmWave radios is costly and the key enabling technology is still under standardization, whereas massive MIMO requires complex and expensive coordination that is difficult to achieve in practice. Therefore, the main feasible method to offer larger capacity on existing pre-5G deployments is via extreme network densification.

Today, the mobile network operators (MNOs) are however struggling to deploy a higher density of small cells due to the need of extra investment that is not compensated by the actual revenues [1, 2]. On the other hand, multiple field measurement campaigns strongly evidence that the conventional spectrum below 6 GHz may be substantially underutilized across space, time, and frequency [3]. This is a consequence of the legacy "command-and-control" spectrum management approach that used to create static and overprotective allocations [4]. Hence, as a viable alternative to deploying additional small cells, the MNOs may quickly boost the capacity on their deployments with more *dynamic*

and market-friendly spectrum management mechanisms. These should be made available in the emerging 5G systems [5].

With more dynamic spectrum management, the expensive frequency bands could be shared between different stakeholders flexibly, as opposed to exclusive use of licensed spectrum. This may go far beyond opening up unlicensed frequencies for collective uncontrolled use and promises to unlock the much needed additional bandwidth that is currently employed sparsely by its existing incumbents. It can also improve the utilization of presently allocated spectrum across its various dimensions (space, time, frequency), which is essential to support the throughput-hungry 5G applications. To this effect, powerful spectrum sharing technologies emerged recently, such as LTE in unlicensed spectrum (LTE-U), Licensed Assisted Access (LAA), MulteFire, Citizens Broadband Radio Service (CBRS), and Licensed Shared Access (LSA) [6–8].

The latter framework is an evolution of the industry-driven Authorized Shared Access (ASA) technology for controlled spectrum sharing between the incumbent holding the rights to use the frequency bands and the licensee (e.g., the MNO), who is utilizing such spectrum temporarily [9] This concept has been taken forward by the European Commission (EC) to develop a new "individual licensing regime" for authorized spectrum sharing [10]. According to the EC's Radio Spectrum Policy Group (RSPG), the LSA framework enables a limited number of licensees to operate in a frequency band already assigned to one or more incumbents in accordance with well-defined sharing rules. As a result, all of the authorized users, including the incumbents, can maintain their desired Quality of Service (QoS) requirements (see RSPG 13-538).

Ever since its introduction several years ago, the LSA concept has spawned an avalanche of engineering, business, and regulatory work that focused on adapting it promptly for practical applications [11]. This development has been facilitated by the Conference Europeenne des Postes et des Telecommunications (CEPT) as it had established two project teams, PT52 and PT53 (ETSI TS 103 113), to ensure that there are no barriers to the adoption of LSA in 2.3 – 2.4 GHz bands from a regulatory perspective (EC Mandate on MFCN for 2.3–2.4 GHz, 2014) [12]. In parallel, the European Telecommunications Standards Institute (ETSI) has been targeting to outline the LSA system architecture in their respective technical specifications (ETSI TS 103 154 and ETSI TS 103 235). In the US, a Notice of Proposed Rulemaking (NPRM) in 3.5 GHz band was introduced by the FCC [13, 14].

As a result of this concentrated effort, the LSA has soon been ready for practical demonstrations, which took place in Spain (2015), Italy (2016), France (2016), Finland (2016), Czech Republic (2016), and the Netherlands (2017). However, many of these past activities considered near-static LSA operation with longer-term allocations since they primarily addressed the technical feasibility of LSA implementation [15]. Continuing our initial work in [16] and more recent technology groundwork in [17], we here complement these earlier initiatives with a new perspective on highly dynamic spectrum management within the LSA framework. In this paper, we specifically emphasize the QoS aspects and the corresponding service reliability performance as our work unveils the practical limits of dynamic LSA operation based on a real-world trial in a live LTE system.

The rest of this text is organized as follows. In Section 2, an overview of recent activities connected with dynamic spectrum management is offered that refines the requirements for vertical communication use-cases. A discussion on the envisaged dynamic LSA system is also contributed. The principles of our dynamic LSA implementation are discussed in detail in Section 3. We pay particular attention to design targets to ensure that the dynamic LSA framework can achieve the expected degrees of control accuracy. Measurement methodology and results follow in Section 4, while the last Section 5 concludes this work.

2. Dynamic Spectrum Management Overview

Numerous activities and contributions are dedicated to the definition of vertical communication use-cases by refining their requirements and assessing how all of these can be addressed in the 5G specification and standardization process. It is important to note that most of these activities are coming from research, e.g., European FP 7 and H 2020 projects as well as the mobile communications industry community (Next Generation Mobile Consortium (NGMN), 3GPP), while similar initiatives engaging other important vertical industry stakeholders are still rather small-scale [18–20].

With the emerging LSA framework, flexible and more dynamic spectrum sharing may be enabled, which becomes increasingly valuable for demanding 5G applications [21]. The advanced services that can benefit from cross-band spectrum aggregation are those that require massive bandwidths but have difficulty to be supported by the existing MNO deployments (e.g., augmented and virtual reality) [22]. Another category that may take advantage of highly dynamic LSA operation is industrial Internet of Things (IoT) applications across different verticals, especially those requiring reliable operation and dedicated QoS guarantees (e.g., automotive). Finally, LSA can improve a wide range of local broadband services, such as those where the MNOs do not have a possibility of deploying exclusive licensed spectrum (e.g., enterprise) [23, 24].

We thus expect that as LSA technologies mature, an increasing variety of 5G applications and services will be capable of taking advantage of more efficient geographic-temporal spectrum management. In [16], the LSA functionality required to enable truly dynamic spectrum sharing at the timescales of seconds is outlined. Continuing this research, in [25] we addressed a typical cell scenario under the "limit power" policy by capturing the produced interference as a key parameter in the cellular network that employs LSA mechanisms—a summary on our prior system-level evaluations of dynamic LSA operation is available online; see http://winter-group.net/dyn-lsa-sim-res/. In a follow-up research [26], an advanced user satisfaction-aware spectrum management strategy for dynamic LSA management in 5G

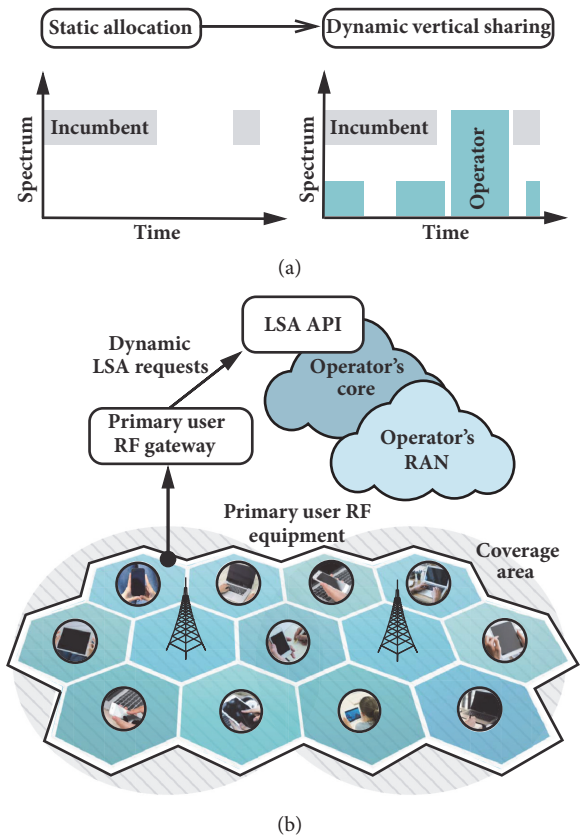


FIGURE 1: (a) Improved spectrum utilization with dynamic sharing. (b) High-level architecture of dynamic LSA operation.

networks was proposed to balance both the user satisfaction and the MNO utilization. All together, the mentioned capabilities are crucial for the incumbent systems with high-speed mobility (e.g., express trains and airplanes) and may offer much improved performance as compared to rigid and near-static LSA implementations. Given that LSA is an example of vertical sharing (see Figure 1(a)), multiple spectrum users across the same geographical area can operate at different priority tiers. For instance, the LSA licensee (e.g., a commercial LTE network) may avoid causing interference to the LSA incumbent (e.g., an air traffic control system).

2.1. Our Contribution. Our envisaged dynamic LSA system is intended to operate according to the high-level architecture captured in Figure 1(b). The primary user of the spectrum (i.e., the incumbent) identifies its target limitations (across space, time, and frequency), where the wireless interference constraints have to be met by the secondary user (i.e., the LSA licensee, such as the MNO). Further, a corresponding LSA request is issued and transferred to the operator's dedicated Application Programming Interface (API), where it is then converted into specific Radio Access Network (RAN) instructions (e.g., interference estimation, transmit power reduction, frequency band change, and LSA spectrum usage policy). Once these commands are received by the operator's cloud, its RAN executes the required actions as instructed.

Considering the key mechanisms and constraints in place for dynamic LSA systems, the proposed functionality as discussed below enables the RAN to respond to the received LSA-specific requests issued by the incumbent in near real time. The primary benefit of our approach is in flexible bandwidth segmentation, which is much more fine-grained and adaptive than in previous LSA implementations. With such dynamic and on-demand network configuration, substantial radio resources can be made available to both the incumbent(s) and the LSA licensee(s), since the proposed logic tightly applies to the time domain. Accordingly, the timescale of radio resource utilization has higher granularity than in static LSA approaches. At the time instants when the incumbent does not utilize its bandwidth resources, they are released automatically, without any additional administrative overheads or delays associated with the LSA database updates.

In addition, our approach efficiently leverages the spatial dimension of the shared spectrum resources. That is, the locations where the bandwidth has to be released by the LSA licensee (the MNO) back to the incumbent can be obtained with higher precision than what is possible with the conventional LSA setups, in which the geographical blocks are typically represented as a coarse grid on the map. In the following sections, we continue by providing a systematic perspective on our development efforts to implement a dynamic LSA system. We also highlight the crucial design choices with regard to the communication chain that facilitates near real time LSA operation within a practical LTE network deployment.

3. Principles of Dynamic LSA Implementation

We expect that upcoming LSA implementations will require software and hardware modifications within the existing cellular network infrastructure as well as, potentially, on the side of the user equipment (UE). Despite substantial ongoing efforts to evolve the LTE system as one of the cornerstone 5G technologies, support for highly dynamic LSA operation in practical MNO deployments calls for a dedicated technology development effort. If not reflected comprehensively in further LTE releases, the LSA spectrum sharing mechanisms may be slow to enter the market, where they are much needed at this time. In this section, we address this important demand by exposing the key system functionality required to support highly dynamic LSA in a 3GPP LTE system.

3.1. Proposed Components and Functionality. We recall that wireless technology standardization conventionally begins by defining the functional elements and interfaces between them. Aiming to lay the groundwork for this, we first identify the main functions and interfaces necessary for implementing dynamic LSA mechanisms (see Figure 2). Given that we have recently outlined the core principles behind the dynamic LSA framework as a proof-of-concept study in [17], we build the present system architecture proposal on our rich hands-on experience acquired then. To this end, we

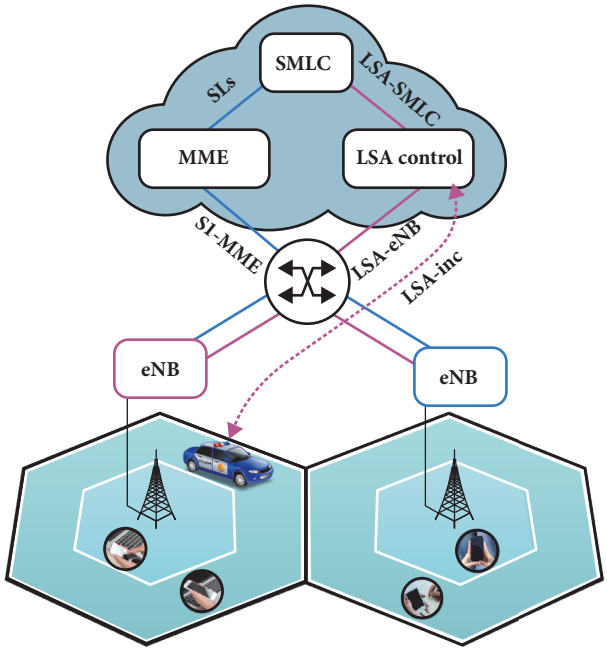


Figure 2: Key elements of proposed dynamic LSA architecture.

rely on a full-fledged cellular system deployment at Brno University of Technology (BUT), Czech Republic, which offers an excellent example of a contemporary 3GPP LTE target system. Therefore, this paper advances on top of our previous works as an enhanced Free Space Path Loss (FSPL) model was constructed to decrease errors of the interference estimation procedure. Also, two representative power allocation policies based on heuristic iterative search were implemented as part of the LSA controller node within the LTE system.

(1) Control is responsible for accepting the incumbent's requests and managing the power allocation across the network. In conventional LSA systems, this function may be performed by the LSA controller, while the LSA repository acts as a proxy. For the intended highly dynamic operation, we utilize a direct interface between the incumbent and the LSA controller to reduce the control-plane latency, since the trial was conducted in a closed environment.

(2) Positioning is used for locating the sources of interference in the network, which is key to efficient power allocation. In cases where the LSA band is utilized for downlink (DL) communication, the positions of all LTE base stations (named eNBs) are typically well known and may thus be programmed into the controller. However, in cases where the LSA is applied to uplink (UL) or TDD bands, it is important to have the position estimates of the UEs as well, since the latter become the main sources of interference. Currently, such information is not available to the LSA controller, but for the current trial we introduced the LSA-SMLC interface, which allows the LSA controller to directly access the LTE positioning data (as calculated based on the cellular signals or reported by the UEs themselves).

(3) Policy assignment offers the capability to issue commands to the individual cells, which is at the very core of the LSA concept. In our dynamic LSA system, this implies setting transmit power constraints for the corresponding radio interfaces—in the limit up to full cell shutdown. For that purpose, we introduce the LSA-eNB interface that aggregates (i) functions normally accessible via OA&M (set the transmit power, shut the cell down, etc.) as well as (ii) certain instructions issued directly to the UE via the RRC interface (initiate handover to a specific cell, switch the band, etc.).

3.2. Envisioned LSA System Operation. Our proposed system design targets to ensure that the dynamic LSA framework can achieve the desired degrees of control accuracy. To this end, we enable the UEs to know their precise location and make them report it to the SMLC. This information can then be extracted from the SMLC via a proprietary system monitoring interface, which acts as LSA-SMLC. The control interface LSA-inc can be implemented as a socket, over which the current coordinates and the threshold power settings can be reliably reported by each of the incumbent's users, thus defining a constraint in the controller's power allocation algorithm. Unlike in static LSA cases, dynamic reports are transient in nature and time out on their own. Hence, the network reverts to its default operation whenever no more reports are being sent. For our below test measurements, such reports were triggered manually.

Most importantly, to address the increased control granularity in the dynamic LSA system, the LSA-eNB interface needs to be implemented as a combination of the OA&M and the direct UE control. One of our core proposals that are instrumental to the dynamic LSA operation is to reduce the transmit power instead of a complete cell shutdown, which considerably improves the system capacity. While it is relatively easy to lower the UL or DL power limit in a cell instead of shutting it down, actually ensuring it in the UEs that are thus forced outside of the cell's coverage area is much more difficult. In our setup, the DL power control has a few possible settings that match the cell coverage area in the DL with its intended service area in the UL. In a commercial-grade test deployment, it may be cumbersome to send the RRC control signals from the core network side that would enforce the UE handover out of the LSA band (as UEs prefer to handover inside their current band when a cell is shut down). If Multi-access Edge Computing (MEC) infrastructure is available, this issue can be resolved by implementing a service that monitors positions of the incumbent entities and UE devices. Furthermore, computational tasks can be offloaded to the MEC layer, if characteristics of the deployment cause an increase in the computational complexity. For example, if we consider a scenario where the number of UEs is significantly high and they move fast, utilizing MEC services may help distribute the computational load. Hence, responsiveness of the system will increase, which in turn will allow to handle high-speed UEs. However, current ETSI documentation does not list this case as a service scenario [27]. To mimic this functionality, the UEs may employ a user-space program that would shut them down instead, since at the time of the trial we did not have access to MEC-enabled hardware. Further, the UEs that fall out of the service area of a particular cell—but

remain capable of receiving its DL signaling—may, according to their protocol, attempt a RACH procedure. These RACH transmissions use power ramping if unsuccessful [28] (and they will be unsuccessful as the cell is instructed not to accept the initiating UEs) and may thus violate the interference constraints. While RACH packets are only short bursts, they may still cause issues in certain cases.

3.3. Important Practical Considerations. Out of the three interfaces identified for the needs of dynamic LSA operation, only two are deployment-ready today. Indeed, reporting the desired interference constraints may be readily achieved with existing IP-based protocols, while connecting the SMLC with the LSA controller is fairly straightforward. On the contrary, ensuring that all of the UEs follow the dynamic power allocation policy in a predictable manner is complicated due to several limitations in the current cellular signaling:

(i) The UEs cannot be forced to handover away from their current serving eNB without a deep integration into the proprietary code inside the MME. While the required functionality may be made available by some of the core network vendors, it is presently not a part of any standard specification. Similarly, UE handovers between the individual cells under the same eNB may not be even reported to the MME, which translates into the need for more proprietary interfaces as of today.

(ii) The service area of a cell with the reduced UL power level may not be easily predicted by the UEs, thus often resulting in futile RACH attempts by the devices that are relatively far away from the cell center (and given the UL power limits). Adding the relevant information elements into the beacon signal could certainly allow to improve on this, but it is neither supported by the current beacon formats nor available in the practical eNBs utilized for testing.

Our LSA controller implementation utilizes all of the needed interfaces together with the relevant power allocation policies that are discussed in the following section. Its current version employs a heuristic iterative search procedure to locate the optimal power assignment across all cells as to match certain performance targets (e.g., maximize the throughput or minimize the number of users that lose service). For the most unsatisfied constraint, a simple algorithm is applied in a loop: a vector of the expected reduction in interference for a 1-dB reduction of power in cell i is computed, and the most impacting cell is chosen. The power in this cell is then reduced by 1 dB and the vector is updated (with its sorted order restored). Once the constraint at hand is no longer the most unsatisfied, another constraint is chosen to proceed further. When all of the constraints are satisfied, the search is complete.

The asymptotic complexity of our proposed algorithm is $O(C \cdot R)$, where C is the number of cells and R is the number of constraints (e.g., incumbent's devices) in the system. The resultant power allocations may then be stored and used to initialize the subsequent runs, thus further reducing the time needed for reaction. This means that our solution can be scaled up to hundreds of cells if desired, without much sacrifice in the response times. To summarize, our proposed system implements all of the dynamic LSA functionality. While some of its components may not be production-ready as of yet (i.e., the LSA-eNB interface) due to the limitations in the underlying LTE subsystems, it clearly confirms that the considered system is not only feasible, but may also be deployed in larger cellular networks with reasonable effort.

Table 1: Main system-level parameters.

Description	Value
3GPP LTE system baseline	Release 10
Division multiplexing	FDD
Number of cells (eNBs)	3
Frequency band	17 (700 MHz)
Bandwidth	5 MHz
Number of resource blocks (RBs)	25
Max. eNB power level	0 dB
Min. eNB power level	-30 dB
Interference threshold	-85 dBm
Noise floor	-100 dBm
Path loss coefficients	5 dBm (concrete) 2 dBm (gypsum)
Path loss model	Enhanced FSPL
Number of terminals (UEs)	4
Transmission data rate	512 kbps
Frequency analyzer	R&S TSMW
Antenna	HL040 log-periodic broadband

4. Our Measurement Methodology and Results

To systematically demonstrate the outlined principles of highly dynamic LSA operation in practice, we conducted a full-scale real-world implementation of our capable LSA-based spectrum sharing system in a commercial-grade 3GPP LTE network deployment. In this unique trial, we focused on the UL LTE channel and evaluated its availability over the LSA frequency bands. The UL system has been preferred due to its significantly higher implementation complexity as compared to the DL LTE channel, which also led to more interesting observations. The primary system-level parameters are summarized in Table 1, while the composition of our trial implementation is detailed in Figure 3.

The UEs under test were continuously communicating with the remote server located on the Internet, which recorded their effective UL and DL bit-rate values over time. Prior to taking measurements, the UEs were configured to target a constant bit-rate (CBR) transmission at 512 kbps, if sufficient UL radio resources were available; otherwise, they utilized all of the remaining UL resources subject to the current transmit power restrictions. The trial focused on analyzing the LSA band and demonstrating its highly dynamic operation. Hence, the UEs were forced to shut down

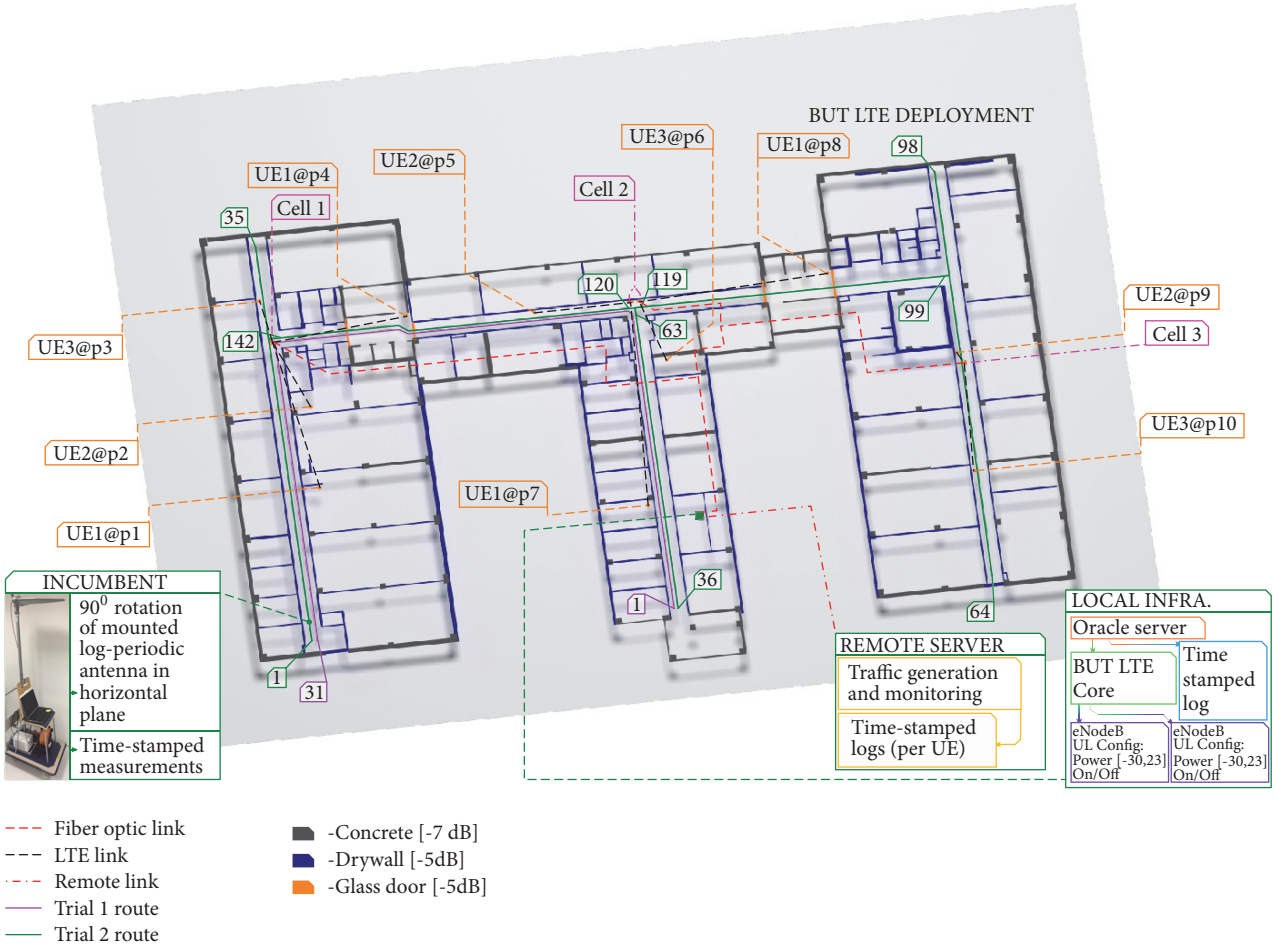


Figure 3: Implemented dynamic LSA setup (architectural components) in our test 3GPP LTE infrastructure.

whenever they were supposed to switch over to the non-LSA frequencies. An enhanced FSPL model was created to decrease errors of the interference estimation procedure. This step was needed, since no precise radio propagation model of the test environment was available. In order to calculate the path loss, we examined which obstacles signal had to penetrate through, while travelling over the direct path from a transmitter to a receiver. Each of the objects was assigned a specific material with its attenuation value that was added to the total path loss. This model helped us avoid situations where interference was estimated incorrectly because of the fact that the pure FSPL model does not take into account walls and other obstacles.

4.1. Power Allocation Policies. Two representative power allocation policies based on heuristic iterative search were implemented as part of the LSA controller: (i) "Derivative" algorithm and (ii) "Shannon" algorithm. Both methods assign the maximum allowed uplink power across the cells such that the total interference from the network towards the moving measurement cart (which represents the incumbent) does not exceed a given threshold. Our trial scenario utilizes 4 UEs, since we were limited in supply. Despite this, we managed to have a system that is well balanced in terms of complexity and the number of participating UEs. They remain stationary and their positions are known, since the dimensions of the trial required precise positioning, thus making the system more complex. This, in turn, makes the measurement setup more prone to behaving incorrectly, should positioning function inaccurately. The interference from a cell (eNB) can typically be approximated by the maximum interference from all the UEs in this cell. Hence, for the 3 utilized cells (see Figure 3), the total interference on the measurement cart is estimated as a sum of the interference levels produced by the UEs closest to the cart, taken across all cells.

(1) Derivative algorithm: this algorithm maximizes the interference reduction against the total power loss. It lowers the power in the closest cell, thus reducing its interference towards the measurement cart, and when the closest cell is shut down, it moves on to the next closest one. The algorithm for power reduction is as in Algorithm 1.

To increase the power in a cell when the cart has moved away, we use Algorithm 2.

Hence, Derivative algorithm provides us with the most interference reduction by lowering the power in the cells that are the closest to the measurement cart.

(2) Shannon algorithm: this algorithm employs the capacity estimate formula $C = B\log_2(1 + S/N)$ to evaluate the

```
while interference I is higher than the threshold I_0 do
    Find the closest active UE;
    Find cell this UE is associated with;
    Lower the power in that cell;
    if power in cell is reduced below its minimum
     feasible level then
        Shut the cell down;
        Remove cell from the list of active cells;
        Remove UEs associated to the cell from the list
         of active UEs;
    end
end
```

Algorithm 1

```
while interference I is lower than the threshold I_0 and
 list of considered cells is not empty do
    Find the farthest considered cell with less than
     maximal feasible power level;
    Find the UEs associated with this cell;
    Calculate the interference gain dI in case of power
     increase;
    if I + dI < I_0 then
        if cell was shut down then
            Turn the cell on;
            Add cell to the list of active cells;
            Add UEs associated with the cell to the list
             of active UEs;
        else
            Increase the power to I_0;
        end
    else
        Remove the cell from the list of considered cells;
    end
end
Reset the list of considered cells;
```

Algorithm 2

changes in the user's effective transmission rate. Importantly, if the user's required bit-rate is below its capacity (subject to the current power restriction), there is no change in the effective transmission rate. The algorithm selects a cell for the power decrease procedure as to maximize the interference reduction subject to the minimal loss in the total effective transmission rate of its users.

Algorithm 3 aims to maintain the UE connectivity to the cells instead of providing them with the highest reliability service. Therefore, it prioritizes keeping the users connected. A user j is considered to be connected to cell i if its UL transmit power level is above P_{ij} and allows it to maintain rate above a certain threshold. The power increasing algorithm works as in Algorithm 4.

Similar to Derivative algorithm, Shannon algorithm first lowers the power in the closest cell, thus reducing its interference towards the measurement cart, but only until the lowest transmission rate in this cell reaches the threshold.

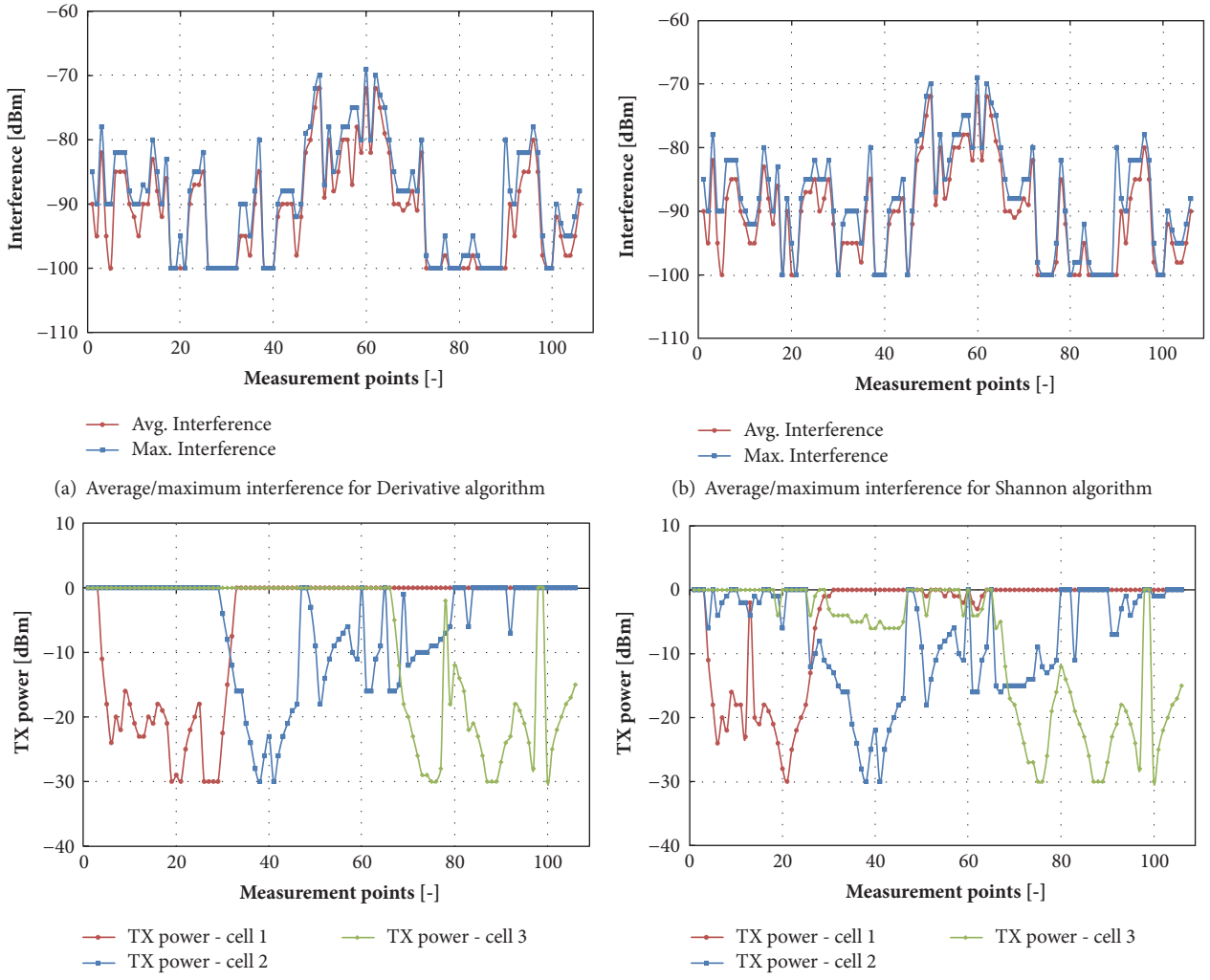
```
for each i from list of considered cells do
    for each j from UEs associated with the cell i do
        calculate the required power level P_ij needed to
         connect to the cell;
    end
end
while interference I is higher than the threshold I_0 do
    if list of considered cells is empty then
        Reset the list of considered cells;
        Reset the list of considered UEs;
        Find the closest UE, which is active;
        Set this UE as inactive;
    end
    Find the closest active UE;
    Find cell this UE is associated with;
    Lower the power in that cell;
    if power in cell i is reduced below its required power
     P_ij needed for the UE j to connect then
        Set power to level P_ij;
        Remove cell from the list of considered cells;
        Remove UEs associated with the cell from the
         list of considered UEs;
    end
end
```

Algorithm 3

```
while interference I is lower than the threshold I_0 and
 list of considered cells is not empty do
    Lower power in all cells to the closest level P_ij;
    Find the farthest inactive UE j;
    Find cell i this UE is associated with;
    Calculate total interference I' if the power in cell i is
     set to level P_ij;
    if I' ≤ I_0 then
        Set power in cell i to P_ij;
        Set UE j as active;
    else
        Use Derivative power increase algorithm;
    end
end
```

Algorithm 4

While the interference threshold is still not reached, the algorithm moves on to the next closest cell. If the power level is such that the bit-rate of the "worst" UE has reached the threshold, but the interference still remains above the target value, the algorithm returns to the first closest cell, drops the "worst" user, and lowers the power in a similar manner again by comparing the second "worst" user's bit-rate against the threshold. Varying the transmission rate threshold P_{ij}, one can control the minimal guaranteed bit-rate. On the other hand, setting the threshold too high can cause more user drops from the network, since a UE is considered dropped when it cannot maintain its threshold transmission



(a) Average/maximum interference for Derivative algorithm

(b) Average/maximum interference for Shannon algorithm

(c) Transmit power allocation for each cell according to Derivative algorithm

(d) Transmit power allocation for each cell according to Shannon algorithm

Figure 4: Measurement-based results produced in the conducted dynamic LSA trial.

rate. Therefore, a sensible solution would be to set the QoS-guaranteed bit-rate equal to the threshold transmission rate.

4.2. Main Hands-On Observations. Our primary objectives in the conducted dynamic LSA trial were to (i) compare the above power allocation policies and (ii) verify whether the corresponding algorithms operate as intended; i.e., they do not breach the interference threshold while meeting their respective optimization targets. The results were collected on a predetermined set of the measurement cart locations placed along the route for the second trial; see green path in Figure 3, where two evaluation scenarios are illustrated: (i) the purple path indicates the measurements for the purposes of QoS assessment (completed in our previous trial [17]), while (ii) the green path stands for the overall testing of the implemented heuristic iterative search logic (conducted in the current trial). As it was mentioned above, measurement points were placed along the route. Their numbering is given in Figure 3 as index of the first and the last measurement point on each segment of the route. Distance between the neighboring points was equal to 1.5 meters. All indexes of the points in this trial refer to points along the green line in Figure 3. Note that the system was notified on the movement of the measurement cart, while at every instant of time the interference was recorded. All of the relevant data were continuously logged for further analysis. Accordingly, Figures 4(a) and 4(b) report on the average (red line) and the maximum (blue line) interference levels received by the measurement cart at each position for both algorithms. It is also important to clarify that since we only had 4 physical UEs, we devised an approach to increase their number in a trial by reusing physical devices at different positions. Hence, the label "UE3@p3" means that the third physical UE was placed at position 3 in our trial. Relocation of the UE was performed only when any interference from it was negligible at the measurement cart, in order to minimize the effect of this procedure.

Comparing Figures 4(c) and 4(d) that illustrate the power allocated according to Derivative and Shannon algorithms, one can observe that when the cart is located in the first cell the Derivative algorithm only alters the power allocation there, while the Shannon algorithm adjusts all three cells in order to raise power and transmission rate in the closest cell. Hence, it can be concluded that Shannon algorithm operates more flexibly over the entire network, while Derivative algorithm mostly concentrates on the nearest cells. From Figures 4(a) and 4(b), where the average and the maximum interference levels received by the measurement cart across the check points are reported, we learn that unlike the Derivative algorithm the Shannon algorithm not only lowers the average interference, but also attempts to maintain it as close to the threshold as possible, so as to increase power and transmission rate in the network. In both plots, however, we can observe an interference threshold breach with the peak value of 10 dBm. This breach spans from point 50 to 70, and it can be seen that at these points the measurement cart was close to the UEs. This can be explained by imperfections in the path loss model, which was used to estimate the interference and calculate the penalties for the UL.

5. Conclusions

This work accentuates the importance of highly dynamic spectrum sharing to leverage additional bandwidth that may be lightly used by its original incumbents. To further improve upon spectrum utilization in demanding 5G systems, we focus on the emerging LSA framework for vertical sharing, where the incumbent(s) and the licensee(s) operate over the same geographical area by utilizing common frequencies in a carefully controlled manner. This concept has been coined in 2013 and since then rapidly took off with many hands-on demonstrations across Europe, primarily in 2016. Supported by visible research initiatives, such as ADEL [3] and CORE++ [6, 11, 13], the LSA functionality has been tested in a number of countries with the emphasis on the feasibility of its early implementation.

Complementing these important efforts, the present study relies on our rigorous past research (please refer to a summary on our prior system-level evaluations of dynamic LSA operation here: http://winter-group.net/dyn-lsa-sim-res/) to advance the state of the art on LSA by outlining its additional functionality required for highly dynamic operation. We therefore elaborate the key principles of system implementation as well as contribute our unique practical methodology based on a live cellular network deployment. The obtained measurements corroborate the rich capabilities of highly dynamic LSA operation in a commercial-grade network as well as report on the crucial performance indicators related to QoS and service-level reliability.

After completing the trial and analyzing the results, we therefore conclude that application of power control policies is viable. Moreover, to a certain extent it is possible to use the considered policies indoors in cases where precise signal propagation model is available. The only locations where the cell was powered off are those near the UE positions. However, since there was no precise model available, and the scale of the deployment was relatively small, some discrepancies were experienced.

As a result, we can state that when applying power control policies in indoor scenarios, there are certain prerequisites. One of them is a precise path loss model. Another one is the capability of pinpointing the locations of the UE devices in case of uplink LSA application. MEC may aid further in this regard in case the UE mobility is present. For example, supplementary sensor information can be collected from the UEs to estimate their movement in a more accurate manner. Moreover, if the UE devices or the incumbents travel at high speeds, we need to update the UL power limits with higher periodicity in order to keep up with the current situation. Pushing the threshold calculation tasks as close as possible and increasing the compute power will help grow the update frequency as well as the maximum number of UE devices in the network.

Going further, we believe that our results will become instrumental to comprehensively reap the benefits of LSA-based highly dynamic spectrum management scenarios, whether in 2.3 – 2.4 GHz frequency bands or at alternative frequencies, such as 3.4 – 3.8 GHz and possibly up to 4.2 GHz in perspective. This will require further demonstration efforts that may rely on our methodology proposed in this work, which could also be useful for other spectrum sharing initiatives across the globe, including CBRS in the US as well as dynamic spectrum utilization at mmWave frequencies.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

The partial presentation of the manuscript was done as part of the successfully defended doctoral thesis of the first author "Heterogeneous Connectivity of Mobile Devices in 5G Wireless Systems" [29].

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The described research was supported by the National Sustainability Program under grant LO1401. For the research, infrastructure of the SIX Center was used. The publication has been prepared with the support of the "RUDN University Program 5-100" and funded by RFBR according to the research projects No. 17-07-00142, No. 17-07-00845. This work has been developed within the framework of the COST Action CA15104, Inclusive Radio Communication Networks for 5G and beyond (IRACON).

References

[1] M. Mustonen, M. Matinmikko, M. Palola, S. Yrjölä, and K. Horneman, "An evolution toward cognitive cellular systems: Licensed shared access for network optimization," *IEEE Communications Magazine*, vol. 53, no. 5, pp. 68–74, 2015.

[2] M. D. P. Guirao, A. Wilzeck, A. Schmidt, K. Septinus, and C. Thein, "Locally and Temporary Shared Spectrum as Opportunity for Vertical Sectors in 5G," *IEEE Network*, vol. 31, no. 6, pp. 24–31, 2017.

[3] A. Morgado, A. Gomes, V. Frascolla et al., "Dynamic LSA for 5G networks the ADEL perspective," in *Proceedings of the European Conference on Networks and Communications, EuCNC 2015*, pp. 190–194, July 2015.

[4] Spectrum Efficiency Working Group et al., "Report of the spectrum efficiency working group," Tech. Rep., Federal Communications Commission, 2002.

[5] M. Hafeez and J. M. H. Elmirghani, "Green Licensed-Shared Access," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 12, pp. 2579–2595, 2015.

[6] S. Yrjölä, V. Hartikainen, L. Tudose, J. Ojaniemi, A. Kivinen, and T. Kippola, "Field Trial of Licensed Shared Access with Enhanced Spectrum Controller Power Control Algorithms and LTE Enablers," *Journal of Signal Processing Systems*, vol. 89, no. 1, pp. 119–132, 2017.

[7] D. Guiducci, C. Carciofi, V. Petrini et al., "Regulatory Pilot on Licensed Shared Access in a Live LTE-TDD Network in IMT Band 40," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 3, pp. 534–549, 2017.

[8] C. Galiotto, G. K. Papageorgiou, K. Voulgaris, M. M. Butt, N. Marchetti, and C. B. Papadias, "Unlocking the Deployment of Spectrum Sharing with a Policy Enforcement Framework," *IEEE Access*, vol. 6, pp. 11793–11803, 2018.

[9] Frequency Management Working Group (WGFM), "Report on ASA Concept," FM(12)084 Annex 47, 2012.

[10] M. Mustonen, T. Chen, H. Saarnisaari, M. Matinmikko, S. Yrjölä, and M. Palola, "Cellular architecture enhancement for supporting the european licensed shared access concept," *IEEE Wireless Communications Magazine*, vol. 21, no. 3, pp. 37–43, 2014.

[11] M. Matinmikko, H. Okkonen, M. Palola, S. Yrjölä, P. Ahokangas, and M. Mustonen, "Spectrum sharing using licensed shared access: the concept and its workflow for LTE-Advanced networks," *IEEE Wireless Communications Magazine*, vol. 21, no. 2, pp. 72–79, 2014.

[12] D. Guiducci, C. Carciofi, V. Petrini et al., "Sharing under licensed shared access in a LTE real test network at 2.3-2.4 GHz," in *Proceedings of the 27th IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications, PIMRC 2016*, Spain, September 2016.

[13] J. Kalliovaara, T. Jokela, R. Ekman et al., "Interference measurements for Licensed Shared Access (LSA) between LTE and wireless cameras in 2.3 GHz band," in *Proceedings of the International Symposium on Dynamic Spectrum Access Networks, (DySPAN'15)*, pp. 123–129, October 2015.

[14] Federal Communications Commission et al., "Enabling innovative small cell use in 3.5 GHz band NPRM & order," FCC-12-148, 2012.

[15] M. Palolo, T. Rautio, M. Matinmikko et al., "Licensed Shared Access (LSA) trial demonstration using real LTE network," in *Proceedings of the 9th International Conference on Cognitive Radio Oriented Wireless Networks (CROWNCOM '14)*, pp. 498–502, June 2014.

[16] A. Ponomarenko-Timofeev, A. Pyattaev, S. Andreev, Y. Koucheryavy, M. Mueck, and I. Karls, "Highly dynamic spectrum management within licensed shared access regulatory framework," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 100–109, 2016.

[17] P. Masek, E. Mokrov, A. Pyattaev et al., "Experimental evaluation of dynamic licensed shared access operation in live 3GPP LTE system," in *Proceedings of the 59th IEEE Global Communications Conference, GLOBECOM 2016*, December 2016.

[18] Siemens Corporate Research, *5G Communication Networks: Vertical Industry Requirements*, white paper.

[19] PMSE and 5G, "white paper, PMSE-XG Project," 2017, http://pmse-xg.research-project.de/.

[20] 3GPP TS 22.261 V15.0.0, Tech. Spec. Group Services and System Aspects, "Service Requirements for the 5G System; Stage 1, "Rel. 15; Mar. 2017.

[21] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 79–89, 2007.

[22] A. A. Daoud, G. Kesidis, and J. Liebeherr, "Zero-determinant strategies: a game-theoretic approach for sharing licensed spectrum bands," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 11, pp. 2297–2308, 2014.

[23] M. M. Buddhikot, "Understanding Dynamic Spectrum Access: Models,Taxonomy and Challenges," in *Proceedings of the 2007 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, pp. 649–663, April 2007.

[24] R. H. Tehrani, S. Vahid, D. Triantafyllopoulou, H. Lee, and K. Moessner, "Licensed spectrum sharing schemes for mobile operators: A survey and outlook," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2591–2623, 2016.

[25] E. Mokrov, A. Ponomarenko-Timofeev, I. Gudkova et al., "Modeling Transmit Power Reduction for a Typical Cell with Licensed Shared Access Capabilities," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, p. 5505, 2018.

[26] Z. Sadreddini, P. Masek, T. Cavdar et al., "Dynamic Resource Sharing in 5G with LSA: Criteria-Based Management Framework," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 7302025, 12 pages, 2018.

[27] ETSI GS MEC-IEG 004: "Mobile-Edge Computing (MEC); Service Scenarios", Ver. 1.1.1, 2015-11.

[28] 3GPP TS 136 321 - V10.5.0 - LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification (3GPP TS 36.321 version 10.5.0 Release 10).

[29] P. Masek, *Heterogeneous Connectivity of Mobile Devices in 5G Wireless Systems [Doctoral thesis]*, Brno University of Technology, Faculty of Electrical Engineering and Communication, Department of Telecommunications, Advised by doc. Ing. Jiri Hosek, Ph.D., 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 3649513, 8 pages
https://doi.org/10.1155/2018/3649513

Research Article

Simple Algorithms for Estimating the Symbol Timing Offset in DCT-Based Multicarrier Systems

Fernando Cruz-Roldán,[1] **José Piñeiro-Ave**,[1] **José L. Rojo-Álvarez**,[2] **and Manuel Blanco-Velasco**[1]

[1]*Department of Signal Theory and Communications (EPS), Universidad de Alcalá, 28805 Alcalá de Henares, Madrid, Spain*
[2]*Department of Signal Theory and Communications, Universidad Rey Juan Carlos, 28943 Fuenlabrada, Madrid, Spain*

Correspondence should be addressed to Fernando Cruz-Roldán; fernando.cruz@uah.es

Received 4 April 2018; Accepted 28 June 2018; Published 15 July 2018

Academic Editor: Takayuki Ito

Multicarrier modulation based on discrete cosine transform (DCT-MCM) is a candidate technique for future wireless communication. DCT-MCM provides a better usage of the spectrum, allowing for much lower out-of-band radiation than conventional OFDM. In this paper, we address the problem of correcting the symbol timing offset in DCT-MCM. Two sliding window-based correlation algorithms are proposed. The new approaches consider a prefix and a suffix that are inserted into each data symbol, which are the replica of part of the useful data. The performance of the proposed approaches is verified and tested, and their usefulness and effectiveness are illustrated by computer simulations.

1. Introduction

The next generation of communication systems (5G and beyond) will employ higher bandwidths and data rates that are able to provide enhanced data services to users, also improving the efficiency of spectrum usage. Consequently, channel partitioning methods will be the dominant medium access techniques due to their appealing features [1], such as higher system performance and spectrum efficiency, typically measured in (bit/s)/Hz per unit area or cell.

Discrete multitone (DMT) modulation and orthogonal frequency division multiplexing (OFDM) are examples of channel partitioning techniques that employ the strategy of replicating the data into each transmitted symbol. This replica of part of the data, usually referred to as redundant samples, leads to a channel partitioning that provides a set of ideally independent subchannels in which channel equalization can be easily performed [1]. DMT and OFDM are efficiently implemented using discrete Fourier transforms (DFTs). These techniques, however, have drawbacks, such as a sensitivity to offsets in time and carrier frequency. Moreover, the temporal pulse shape of DFT is a rectangular window, thus DFT-based multicarrier modulation (DFT-MCM) suffers from high side-lobe radiation.

As alternatives to DFT-based solutions, several authors have proposed the use of discrete cosine transforms (DCTs) [2–6]. DCT-based multicarrier modulation (DCT-MCM) offers benefits such as excellent spectral compaction and energy concentration and less intercarrier interference leakage to adjacent subcarriers or that DCT uses only real arithmetic [2, 3]. Furthermore, DCT-MCM provides a better usage of the spectrum allowing for much lower out-of-band radiation. As has been widely reported (e.g., [3]), the bandwidth of a DCT-MCM system can be only half of the bandwidth required by DFT-MCM with the same number of subcarriers.

In any multicarrier modulation (MCM) systems, symbol timing estimators play an important role in the receiver to find the start of the symbol of the received signal. Numerous methods for DFT-MCM have been proposed since the publication of the maximum likelihood (ML) estimation algorithm in an additive white Gaussian noise (AWGN) channel and the method of Schmidl and Cox [7–9]. We refer the reader to the studies in [10–12] for a

detailed list of recent synchronization approaches and their implementations in some hardware/software environments for different real-world applications. Unfortunately, this field has received less attention for the DCT-MCM, even though synchronization is also crucial and critical. In [13], two joint maximum likelihood frequency offset and phase offset estimators, considering only AWGN channels, were presented. More recently, a new symbol synchronization method for optical fast OFDM is reported in [14]. As a result, the problem of a feasible technique for DCT-MCM still remains open.

The main novelty of this paper is the proposal of simple algorithms to carry out precise symbol timing estimation for DCT-MCM. The new approaches exploit the redundancy inserted as both left and right extensions in DCT-based systems. It is shown by computer simulations that the proposed approaches yield efficient and feasible time estimation solutions, providing satisfactory results in several communication scenarios.

The rest of this paper is organized as follows. In Section 2, the system model for the considered multicarrier transceivers is described. Section 3 presents the proposed algorithms for DCT2e and DCT4e-based systems. Section 4 provides a performance evaluation of the proposed algorithms and its comparison to DFT-based systems, and, finally, concluding remarks are given in Section 5.

2. System Model

Figure 1 is a general block diagram to implement MCM. At the transmitter, the data are processed by an N point inverse transform $\mathbf{T}_a^{-1}$, with N being the number of subchannels or subcarriers. At the receiver, a discrete transform $\mathbf{T}_c$ is performed.

In DFT-MCM (e.g., OFDM/DMT systems), $\mathbf{T}_a^{-1}$ is an IDFT, the redundancy of length N_G, e.g., a cyclic prefix (CP) (see Figure 2(a)), is introduced at the beginning of each N-length data symbol $x[n]$ to be transmitted, $h_{pf}[n]$ is a time-domain equalizer that shortens the effective channel $h_{ch}[n]$ to an appropriate length, $\mathbf{T}_c$ is a DFT, and, finally, the frequency-domain equalizer (FEQ) block corrects the dispersive effect of the transmission channel.

In DCT-MCM, $\mathbf{T}_a^{-1}$ can be any kind of inverse DCT [16], with two different types of redundancy: symmetric extension (SE) or zero-padding (ZP). This work focuses on Type-II even discrete cosine transform (DCT2e) and Type-IV even discrete cosine transform (DCT4e), assuming SE [2, 5]. Unlike DFT-based systems, two symmetrical extensions of N_G samples are appended: a prefix or left extension (LE), and a suffix or right extension (RE). Figures 2(b) and 2(c) show examples of the symmetric extensions for DCT-MCM.

In Figure 1, the received sequence in presence of a temporal delay δ and carrier frequency offset ϵ can be expressed as

$$y[n] = \frac{1}{N}\sum_{k=0}^{N-1} H[k]\cdot X[k]\cdot e^{j(2\pi(k+\varepsilon)(n+\delta)/N)} + z[n], \quad (1)$$

where $H[k] = \mathrm{DFT}\{h[n]\} = \mathrm{DFT}\{h_{ch}[n] * h_{pf}[n]\}$ and $z[n]$ is a term related to the noise. An additional constraint of DCT-MCM is that the channel impulse response $h[n]$ must be symmetric, as shown in Figure 2(d). It is important to highlight that some channels satisfy this condition, such as chromatic dispersion in single-mode fibers [4]. If not, the prefilter $h_{pf}[n]$ of Figure 1 is located at the receiver to enforce the symmetry in $h_{ch}[n]$. There are some appropriate techniques to design this prefilter [2, 6]. For the sake of simplicity, we adopt a model of symmetric discrete-time channel $h[n]$ for DCT-MCM. If this condition is not satisfied, then the training symbol used in [6] can be employed to perform an initial timing symbol synchronization using the algorithms proposed in this paper, and then the channel estimation of $h_{ch}[n]$ can be carried out.

Finally, the channel length of $h_{ch}[n]$ is assumed to be less than the length of the redundancy (SE or CP); i.e., $\nu+1 < N_G$. For a more thorough description of DCT-MCM, we refer the reader to [2, 5, 16].

3. Proposed Metrics

An analysis using the correlation between two signals provides a quantitative measure of the similarity between them. In STO estimation, the idea behind [7–9] and other approaches in [10, 11] is to use correlation functions to find the similarities that are shared by the data part of the symbol and the redundant samples in the prefix/suffix.

In DCT-MCM, the maximum value or peak is also reached (in the absence of noise) when there exists a set of samples that are pairwise correlated. The time position of this maximum value is useful in finding the symbol timing and the phase of the correlation could yield the frequency estimate. Given that in DCT-MCM the redundancy follows a mirror (anti)symmetry and the channel impulse response is a whole-sample symmetry (WS) sequence, there are two blocks of $N_G - \nu$ samples that are pairwise correlated. For example, let us consider the absence of noise, the fact that δ is an integer number and $\epsilon = 0$, and the time instants represented in Figure 3. At the start of the symbol, specifically at instants $n_0 - 1$ and n_0, we have

$$\begin{aligned} y[n_0-1] &= y[n_0] \\ &= \sum_{k=1}^{\nu} h[k]\left(x_e[n_0+\delta-k-1] + x_e[n_0+\delta+k-1]\right) \\ &\quad + h[0]\, x_e[n_0+\delta-1]. \end{aligned} \quad (2)$$

In addition, at the end of the symbol, the receiving signal presents identical (for the HS extension) or opposite (for the HA extension) values in two consecutive samples, $n_0 + N - 1$ and $n_0 + N$. That is, we have

$$\begin{aligned} y[n_0+N-1] &= \gamma y[n_0+N] = \sum_{k=1}^{\nu} h[k] \\ &\quad \cdot \left(x_e[n_0+N+\delta-k-1]\right. \\ &\quad \left. + x_e[n_0+N+\delta+k-1]\right) + h[0]\, x_e[n_0+N+\delta \\ &\quad -1], \end{aligned} \quad (3)$$

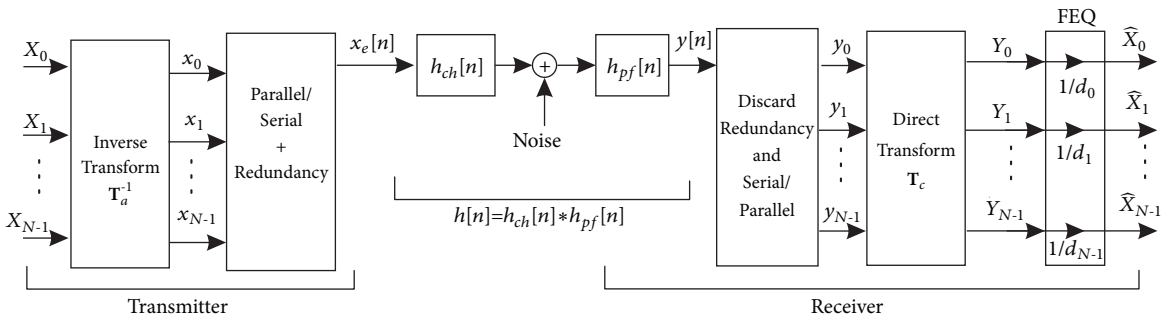


FIGURE 1: Block diagram of a transforms-based multicarrier system over a channel with additive noise.

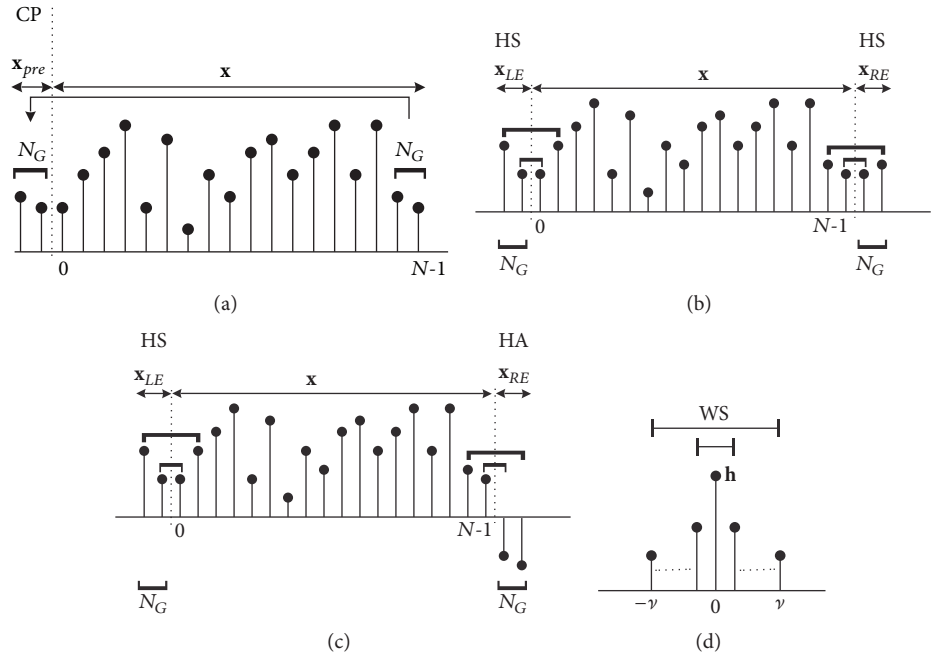


FIGURE 2: (a) Cyclic prefix for DFT-based systems. Symmetrical extensions for (b) DCT2e-based and (c) DCT4e-based systems. HS and HA stand for half-sample symmetry and half-sample antisymmetry, respectively. (d) Symmetry in the channel impulse response **h**. WS stands for whole-sample symmetry ($h[n] = h[-n]$, $1 \le n \le \nu$).

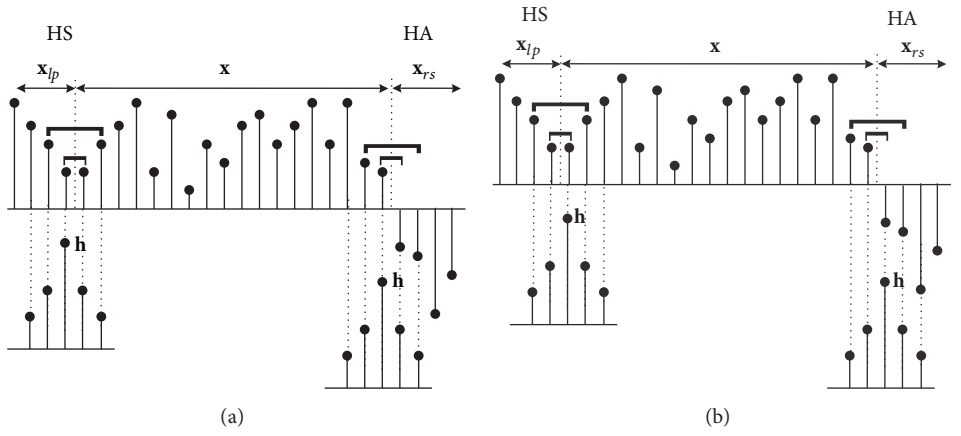


FIGURE 3: Transmitted symbol and channel impulse response at different instants. (a) $n_0 - 1$ and $n_0 + N - 1$. (b) n_0 and $n_0 + N$. ($N_G = 4$, $\nu = 2$).

where $\gamma = 1$ for the HS extension or $\gamma = -1$ for the HA extension (see Figure 2 and [2, 16]). Notice that there are other sets of $N_G - \nu - 1$ identical samples at the start of the received symbol:

$$y\left[n_0 - \ell - 1\right] = y\left[n_0 + \ell\right], \quad \delta \le \ell \le \delta + N_G - \nu - 1, \quad (4)$$

and identical (for the HS extension) or opposite (for the HA extension) samples at the end of the received symbol:

$$y\left[n_0 + N - \ell - 1\right] = \gamma y\left[n_0 + N + \ell\right], \quad \delta \le \ell \le \delta + N_G - \nu - 1. \quad (5)$$

Considering the above, we propose new blind algorithms to perform tight timing offset and coarse frequency synchronization for DCT2e-MCM and DCT4e-MCM. The first proposed approach estimates the timing offset $\widehat{\delta}_{SE}$ using the maximum of the metric:

$$M_{LRE_a}\left[\delta\right] = \left|C_{LE}\left[\delta\right]\right|^2 + \left|C_{RE}\left[\delta\right]\right|^2 - \frac{E_{SE}\left[\delta\right]}{4N_G}, \quad (6)$$

where

$$C_{LE}\left[\delta\right] = \sum_{\ell=\delta}^{\delta+N_G-1} y\left[\ell\right] y^*\left[-\ell - 1\right], \quad C_{RE}\left[\delta\right] = \sum_{\ell=\delta}^{\delta+N_G-1} y\left[N - 1 - \ell\right] \gamma y^*\left[N + \ell\right]. \quad (7)$$

The normalization term $E_{SE}[\delta]$ considers the energy of the windowed signals:

$$E_{SE}\left[\delta\right] = E_{LE}\left[\delta\right] + E_{RE}\left[\delta\right], \quad (8)$$

where

$$E_{LE}\left[\delta\right] = \left(\sum_{\ell=\delta}^{\delta+N_G-1} \left|y\left[-\ell - 1\right]\right|^2 + \sum_{\ell=\delta}^{\delta+N_G-1} \left|y\left[\ell\right]\right|^2\right),$$

$$E_{RE}\left[\delta\right] = \left(\sum_{\ell=\delta}^{\delta+N_G-1} \left|y\left[N - 1 - \ell\right]\right|^2 + \sum_{\ell=\delta}^{\delta+N_G-1} \left|y\left[N + \ell\right]\right|^2\right). \quad (9)$$

The second proposed method is based on Schmidl and Cox's approach. We adapt the metric of [8] to the scheme of prefix and suffix used in DCT-MCM, and, as a result, the timing offset can be estimated using the correlation peak given by

$$M_{LRE_b}\left[\delta\right] = \frac{\left|C_{LE}\left[\delta\right]\right|^2}{E^2_{W_{1p},LE}\left[\delta\right]} + \frac{\left|C_{RE}\left[\delta\right]\right|^2}{E^2_{W_{1s},RE}\left[\delta\right]}, \quad (10)$$

where

$$E_{W_{1p},LE}\left[\delta\right] = \sum_{\ell=\delta}^{\delta+N_G-1} \left|y\left[\ell\right]\right|^2, \quad E_{W_{1s},RE}\left[\delta\right] = \sum_{\ell=\delta}^{\delta+N_G-1} \left|y\left[N - 1 - \ell\right]\right|^2. \quad (11)$$

Table 1: Simulation parameters.

Parameters	Value
System Bandwidth	5 MHz
Sampling Period	200 ns
Carrier Frequency	2 GHz
Modulation and Demodulation	4-QAM
Total Subcarrier Number (N)	512
Length of Redundancy (Examples 2 and 3)	32 (OFDM) 32×2, Left and Right Prefixes (DCT)
Channel Models	ITU Ped A 4 km/h (PED200) [15] ITU Veh A 100 km/h (VEH200) [15]
Channel Equalization	Zero Forcing
Noise Model	iid AWGN
Detection	Hard Decision
Number of Simulations	10000

Finally, assuming perfect symbol synchronization at $\widehat{\delta}_{SE}$ and absence of noise, we have that

$$\left|y\left[\ell + \widehat{\delta}_{SE}\right]\right| = \left|y^*\left[-\ell - 1 + \widehat{\delta}_{SE}\right]\right|, \quad \left|y\left[\ell + N + \widehat{\delta}_{SE}\right]\right| = \gamma\left|y^*\left[-\ell - 1 + N + \widehat{\delta}_{SE}\right]\right|, \quad (12)$$

for those pairwise correlated samples. Considering the above condition, a coarse fractional CFO estimator ($|\varepsilon| \le 0.5$) for DCT-MCM can be carried out by means of the expression

$$\widehat{\varepsilon}_{SE} = \frac{N}{2\pi N_G} \angle\left(C_{LE}\left[\widehat{\delta}_{SE}\right] + \gamma C_{RE}\left[\widehat{\delta}_{SE}\right]\right). \quad (13)$$

4. Experimental Study

In our experiments, we have assumed systems with 512 subcarriers and 4-QAM modulation. Each experiment consists of an average of 10,000 simulation runs with STO independently generated and uniformly distributed over $[-31, 0)$. For the first experiment, the length of the redundant samples is modified to study its effects on different parameters. For the rest, we have considered CPs of 32 samples for the DFT-MCM and the same number of redundant samples for each prefix (LE) and suffix (RE) in DCT-MCM. The setup used in our experiments is summarized in Table 1.

Example 1. The influence of the number of redundant samples (N_G) is studied. The performance of the proposed estimators (6)–(10), for both DCT2e and DCT4e, is assessed and compared with that derived in [10] based on [7], for SNR values of 10 and 16 dB and over discrete memoryless additive white Gaussian noise (AWGN) channels, with no intersymbol interference. The SNR is defined by the ratio of the power of

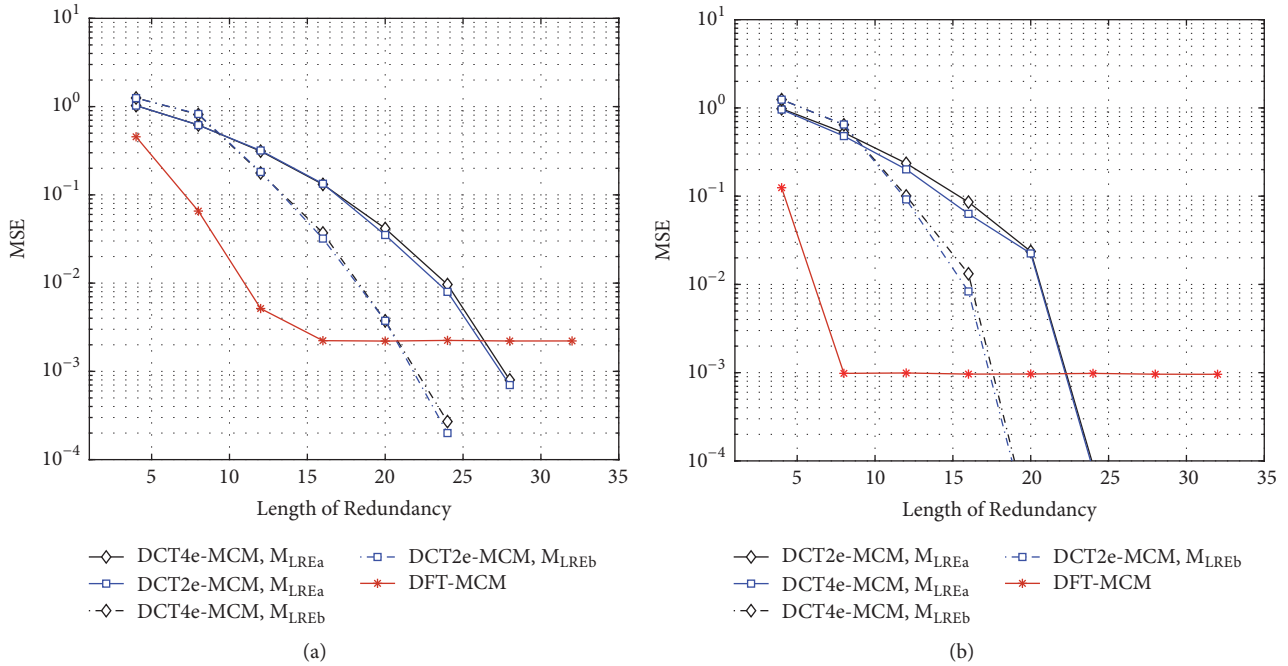


FIGURE 4: Example 1. MSE of STO as an estimator for the AWGN channel and different lengths of the redundancy: (a) SNR = 10 dB, (b) SNR = 16 dB.

the received signal and the variance of the AWGN. Figure 4 shows the resulting mean-square error (MSE) of the STO estimates evaluating the trial and the estimate values of the timing offset. As was documented in [7], it can be seen in Figure 4 that the performance of the DFT-MCM estimator is asymptotically independent of the number of redundant samples, provided that the amount of data for the CP is longer than a certain threshold value, which decreases with the SNR (see Figure 4 and [7, 10]). As the length increases beyond the thresholds, the DFT-MCM time estimator does not improve. On the contrary, the behaviour of the proposed schemes for DCT-MCM is no longer flat. MSE decays monotonically and the improvement is achieved by increasing the SNR and/or the number of redundant samples. Note that MSEs for the DFT-MCM saturate at $2.2 \cdot 10^{-3}$ and $9.9 \cdot 10^{-4}$, and that the performance of the proposed schemes is much better, with MSE$< 10^{-3}$, when SNR = 16 and the number of redundant data is greater than 22 (see (6)) or 17 (see (10)).

We also use our analysis to show how BER changes with the length of the redundant samples. Figure 5 depicts the uncoded BER curves, and, given that the results have been practically indistinguishable for DCT2e and DCT4e, only one curve to represent DCT-MCM is included. Notice that the BER curves become practically saturated for DFT-MCM, and increasing the number of redundant samples does not improve BER. An immediate conclusion that can be drawn from these plots is that the proposed metrics exhibit good performance in the estimation of the STO when the length of the redundant samples is at least one-half of the maximum considered offset. For DCT-MCM, the performance of the proposed approaches for SNR = 10 dB does not improve as the length of the redundant sample increases from 20 to 32, whereas it is possible to obtain BER values lower than 10^{-4} for SNR = 16 dB.

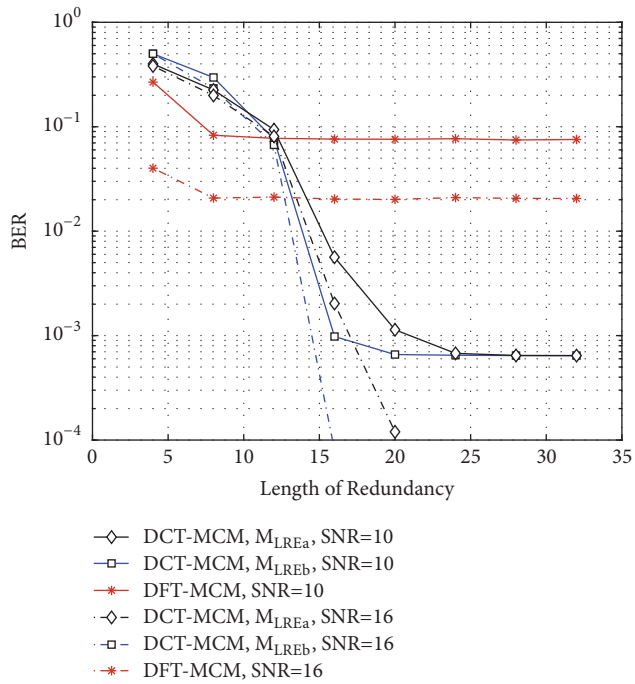


FIGURE 5: Example 1. BER under conditions of STO for different length for the redundancy and SNR values considering AWGN channels.

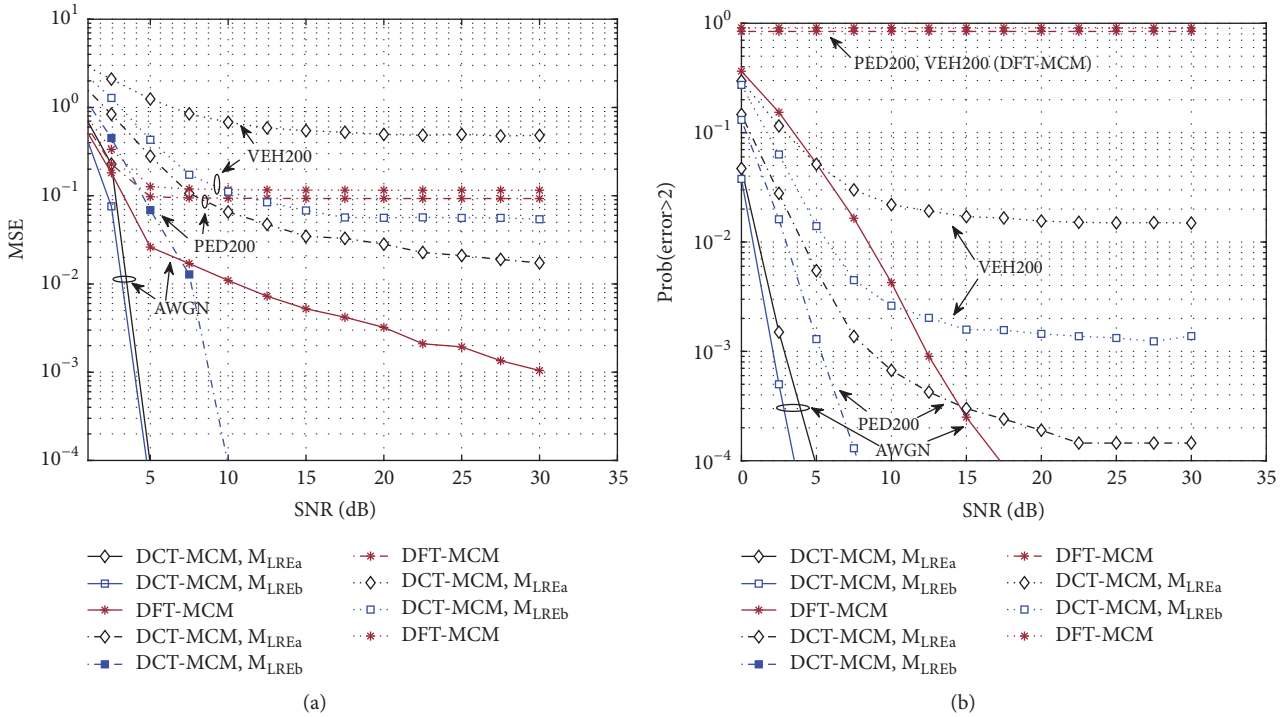


Figure 6: Example 2. Effects of STO for different transceivers and channels. (a) Mean-square error. (b) Probability of failure.

Example 2. In our second set of experiments, all of the results are obtained over three different channels: AWGN and two sets of 100 wireless fading channels each, according to the ITU Pedestrian A and Vehicular A channels [15]. The multipath channels were generated with Matlab's `stdchan` using the channel models `itur3GPAx` and `itur3GVAx` with a carrier frequency $f_c = 2$ GHz and two different sets of parameters: (a) 4 km per hour as pedestrian velocity, $T_s = 200$ ns and length $L = 11$; and (b) 100 km per hour as mobile speed (the moving speed has been chosen considering that the average vehicle speed on European highways lies in the range of 90-100 km per hour under free-flow conditions), $T_s = 200$ ns and length $L = 21$. These channels are referred to as PED200 and VEH200, respectively.

Figure 6(a) shows the resulting mean-square error (MSE) of the STO estimates. As can be seen, the performance of M_{LRE_b} is much better than the others, with the exception of the VEH200 channel, for which [7] provides the best values of MSE.

The estimated temporal offsets have also been obtained. Based on the fact that an accuracy of less than three samples at low SNR makes the approach robust in fading channels [7, 10], we have obtained the probability that the estimated error in the time offset is greater than two samples. The numerical results for the three different transceivers over the channels are displayed in Figure 6(b). These curves show that DCT-based methods significantly improve the performance compared to the DFT-MCM over the entire range of SNR values considered in our experiments.

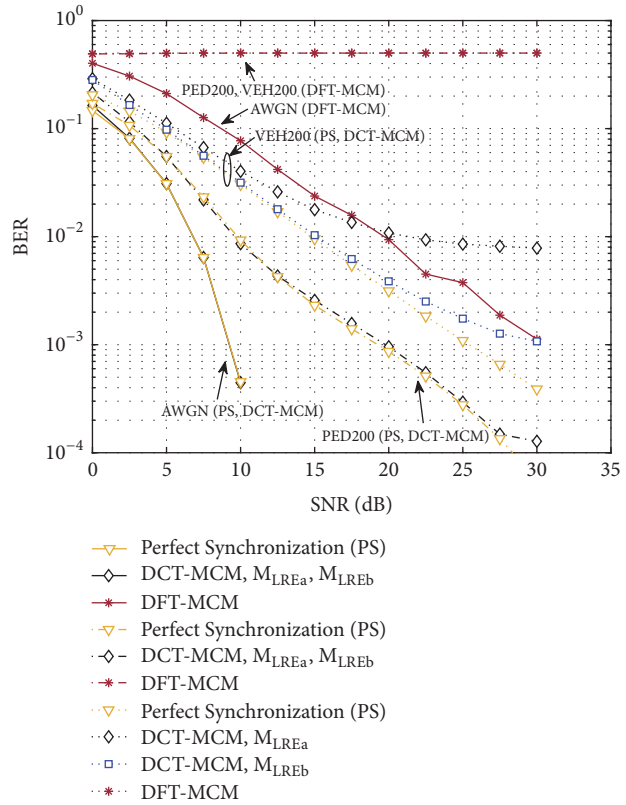


Figure 7: Example 2. BER under symbol timing offset for different transceivers and channels.

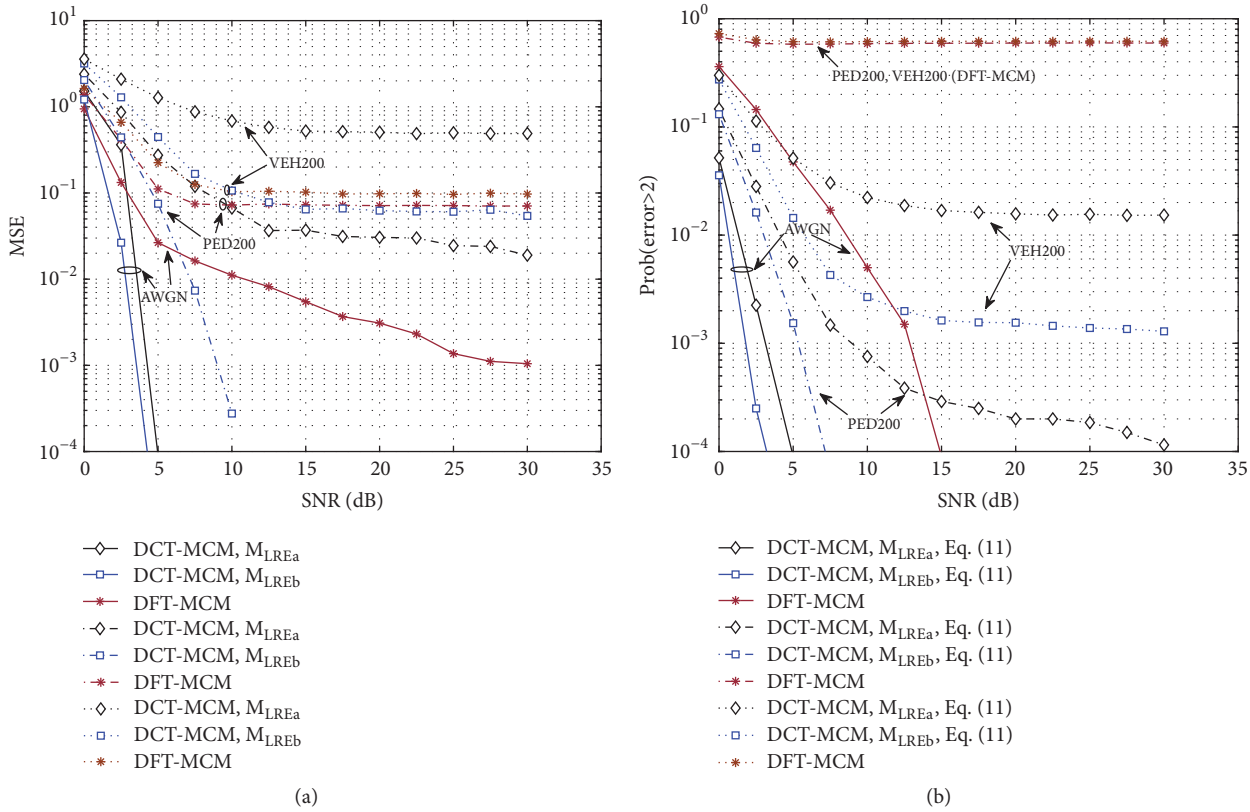


FIGURE 8: Example 3. Effects of STO and CFO for different transceivers. (a) Mean-square error of time offset estimate. (b) Probability of failure for time offset estimate.

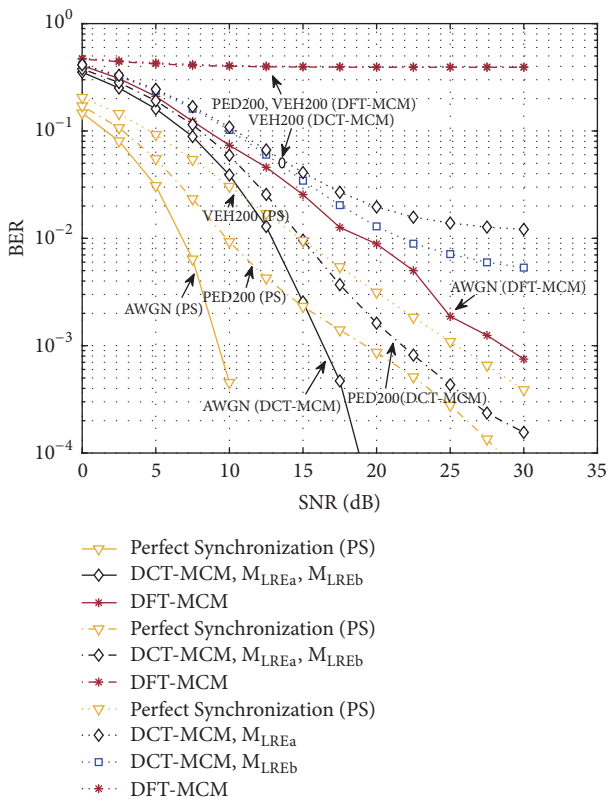


FIGURE 9: Example 3. BER under the effects of STO and CFO for different channels and transceivers.

Next, we compare the BER performance for the three different channels. Figure 7 depicts the obtained results versus the received SNR. In this figure, perfect synchronization (PS) represents the case where STO is perfectly known at the receiver. Interestingly, for a particular channel, the BER performance of the proposed metrics M_{LRE_a} and M_{LRE_b} is practically indistinguishable, and for this reason we only represent one curve. In addition, the proposed methods have the best performance. They are close to the perfect synchronization for the AWGN and PED200 channels, and a performance degradation occurs for the VEH200 channel. On the other hand, the DFT-MCM system is severely affected by a timing offset, especially for the PED200 and VEH200 channels.

Example 3. Lastly, we consider a more complex scenario, where a normalized CFO is added following a uniform distribution in the interval $[-0.04, 0)$, which is estimated by the approach in [7] for DFT-MCM, whereas (13) is employed for DCT-MCM. Figures 8(a) and 8(b) characterize the MSE and $Pr(|\delta - \widehat{\delta}|) > 2$ of the STO estimates, respectively, whereas Figure 9 plots the resulting BER. Comparing the performance of the proposed algorithms with DFT-MCM, it is proved that the new approaches work properly in AWGN and PED200 channels, significantly improving the performance and reducing the BER. The proposed schemes also perform well for the VEH200 channel, although in this case the CFO has an effect on our schemes. On the other

hand, the BER is practically flat for DFT-MCM in the more challenging realistic scenarios of PED200 and VEH200.

5. Conclusions

DCT-MCM offers benefits such as excellent spectral compaction and energy concentration and allows a better usage of the spectrum. In this work, new techniques of symbol timing estimation for DCT-MCM systems are proposed. The new metrics are based on the fact that the correlation between two signals provides a quantitative measure of their similarity, and they take into account the fact that the redundant samples in DCT-MCM are inserted as left (prefix) and right (suffix) extensions satisfying (anti)symmetry properties. A wide set of simulations have been carried out to test the effectiveness of the proposed approaches. It can be established from the results of our experiments that the proposed techniques perform well, especially in Gaussian and low dispersive channels. It is further shown that considering the BER and the probability of the estimate error in the time offset, a performance gain over the conventional DFT-based technique is obtained when the length of the redundancy increases.

Data Availability

The data used to support the findings of this study are not needed actually. The performance study was based on the data randomly generated by the simulation code.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was partially supported by the Spanish Ministry of Economy and Competitiveness through Project TEC2015-64835-C3-1-R.

References

[1] J. Cioffi, "Multichannel modulation," in *Digital Communications*, https://web.stanford.edu/group/cioffi/doc/book/chap4.pdf.

[2] N. Al-Dhahir, H. Minn, and S. Satish, "Optimum DCT-based multicarrier transceivers for frequency-selective channels," *IEEE Transactions on Communications*, vol. 54, no. 5, pp. 911–921, 2006.

[3] P. Tan and N. C. Beaulieu, "A comparison of DCT-based OFDM and DFT-based OFDM in frequency offset and fading channels," *IEEE Transactions on Communications*, vol. 54, no. 11, pp. 2113–2125, 2006.

[4] X. Ouyang and J. Zhao, "Single-tap equalization for fast OFDM signals under generic linear channels," *IEEE Communications Letters*, vol. 18, no. 8, pp. 1319–1322, 2014.

[5] F. Cruz-Roldán, M. E. Domínguez-Jiménez, G. Sansigre-Vidal, J. Pineiro-Ave, and M. Blanco-Velasco, "Single-carrier and multicarrier transceivers based on discrete cosine transform type-IV," *IEEE Transactions on Wireless Communications*, vol. 12, no. 12, pp. 6454–6463, 2013.

[6] F. Cruz-Roldán, M. E. Domínguez-Jiménez, G. Sansigre-Vidal, D. Luengo, and M. Moonen, "DCT-based channel estimation for single- and multicarrier communications," *Signal Processing*, vol. 128, pp. 332–339, 2016.

[7] J.-J. Van De Beek, M. Sandell, and P. O. Börjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Transactions on Signal Processing*, vol. 45, no. 7, pp. 1800–1805, 1997.

[8] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Transactions on Communications*, vol. 45, no. 12, pp. 1613–1621, 1997.

[9] T. Kung and K. K. Parhi, "Optimized joint timing synchronization and channel estimation for OFDM systems," *IEEE Wireless Communications Letters*, vol. 1, no. 3, pp. 149–152, 2012.

[10] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM Wireless Communications with Matlab*, IEEE Press, John Wiley & Sons, 2010.

[11] A. A. Nasir, S. Durrani, H. Mehrpouyan, S. D. Blostein, and R. A. Kennedy, "Timing and carrier synchronization in wireless communication systems: a survey and classification of research in the last 5 years," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, article 180, 2016.

[12] T. H. Pham, V. A. Prasad, and A. S. Madhukumar, "A Hardware-Efficient Synchronization in L-DACS1 for Aeronautical Communications," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 26, no. 5, pp. 924–932, 2018.

[13] F. Gao, T. Cui, A. Nallanathan, and C. Tellambura, "Maximum likelihood based estimation of frequency and phase offset in DCT OFDM systems under non-circular transmissions: algorithms, analysis and comparisons," *IEEE Transactions on Communications*, vol. 56, no. 9, pp. 1425–1429, 2008.

[14] J. Zhao, S. K. Ibrahim, D. Rafique, P. Gunning, and A. D. Ellis, "Symbol synchronization exploiting the symmetric property in optical fast OFDM," *IEEE Photonics Technology Letters*, vol. 23, no. 9, pp. 594–596, 2011.

[15] 3GPP, "Technical specification group radio radio access network. user equipment (UE) radio transmission and reception (FDD) (release 7)," 3GPP TS 3GPP TS 25.101, 3rd Generation Partnership Project, 2007.

[16] F. Cruz-Roldán, M. E. Domínguez-Jiménez, G. Sansigre-Vidal, P. Amo-López, M. Blanco-Velasco, and A. Bravo-Santos, "On the use of discrete cosine transforms for multicarrier communications," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 6085–6090, 2012.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 7985756, 11 pages
https://doi.org/10.1155/2018/7985756

Research Article

Optimal Multicommodity Spectrum-Efficient Routing in Multihop Wireless Networks

Mohamed Saad

Department of Electrical and Computer Engineering, University of Sharjah, Sharjah, UAE

Correspondence should be addressed to Mohamed Saad; msaad@sharjah.ac.ae

Received 16 April 2018; Accepted 13 June 2018; Published 9 July 2018

Academic Editor: Jose M. Gimenez-Guzman

Finding the route with maximum end-to-end spectral efficiency in multihop wireless networks has been subject to interest in the recent literature. All previous studies, however, focused on finding *one* route from a given source to a given destination under the constraint of equal bandwidth sharing. To the best of our knowledge, for the first time, this paper provides extensions to the multicommodity flow case, i.e., the case of multiple simultaneous source-destination (*s-d*) pairs. In particular, given an arbitrary number of *s-d* pairs, we address the problem of finding a route for every *s-d* pair such that the minimum spectral efficiency across all routes is maximized. We provide two alternative approaches, where one is based on fixed-sized time slots and the other is based on variable-sized time slots. For each approach, we derive the *provably* optimal routing algorithm. We also shed the light on the arising tradeoff between the complexity of network-layer route computation and the complexity of medium access control (MAC) layer scheduling of time slots, as well as the amenability to distributed implementation of our proposed algorithms. Our numerical results further illustrate the efficiency of the proposed approaches and their tradeoffs.

1. Introduction

Multihop wireless networks consist of a set of wireless devices that communicate with each other over multiple wireless hops, with participating nodes collaboratively relaying ongoing traffic. Wireless multihop relaying/routing is the foundation for the development and deployment of emerging technologies such as

(i) *client mesh networks*: a set of client devices (tablets, phones, and/or laptops) form a multihop network with *peer-to-peer* relaying;

(ii) *infrastructure wireless mesh networks*: wireless routers/access points are interconnected to provide an infrastructure/backbone for clients;

(iii) *millimeter-wave-based 5G networks*: future 5G networks are envisioned to depend (among others) on ultradense small-cell base stations and the use of millimeter- (mm-) wave spectrum for transmission [1]. The large bandwidth of mm-wave is also accompanied by a high path loss, which necessities the use of multihop relaying across the small-cell base stations [2]. Intelligent routing methods will also be needed for the underlying applications of 5G, e.g., the Internet of Things (IoT) [3].

The end-to-end *spectral efficiency* (in bps/Hz) of a communication route is defined as the rate at which data can be transmitted over the route per unit bandwidth. Therefore, it is an indication of how efficient the channel bandwidth is utilized. Since the bandwidth is a scarce resource in wireless systems, this paper focuses on finding communication routes with maximum spectral efficiencies. In particular, given a multihop wireless network consisting of set of wireless devices and interconnecting wireless links and a set of source-destination (*s-d*) pairs of nodes, this paper addresses the problem of finding a path for each *s-d* pair such that the minimum spectral efficiency of all paths is maximized.

1.1. Related Work. To the best of our knowledge, this is the first systematic, comprehensive study to address wireless spectrum-efficient routing in the case of multiple *s-d* pairs. Related work is presented in two categories:

(1) wireless spectrum-efficient routing,

(2) routing for multiple simultaneous *s-d* pairs.

In what follows we summarize the relevant previous work belonging to both groups.

Spectrum-efficient routing: the recent study in [4] has introduced the following spectrum-efficient routing problem. Given a multihop wireless network that employs time division multiple access (TDMA) and *one s-d* pair, it finds the route with *maximum spectral efficiency* under the constraint of equal bandwidth sharing. On the one hand, the authors of [4] noted that simple shortest path algorithms cannot be used to solve the problem because the resulting routing metric is not isotonic [5]. On the other hand, exhaustive search has an exponential computational complexity because it involves precomputing *all* paths joining a given node pair. Therefore, the study in [4] proposed two efficient, yet *suboptimal* spectrum-efficient routing heuristics. In [6], we have closed the algorithmic gap by introducing the *first* polynomial-time algorithm that solves the spectrum-efficient routing problem (originally introduced in [4]) to exact optimality. In particular, the algorithm presented in [6] has a worst-case computational complexity of $O(N^4)$, where N is the number of nodes. Moreover, we have introduced in [7, 8] an improved algorithm for the same problem that runs in $O(N^3)$-time. Furthermore, we have addressed in [9] the problem of joint routing and power allocation such that a desired spectral efficiency is achieved.

All the studies above focused on finding a route from a given source to a given destination and did not address the interaction between multiple *s-d* pairs; i.e., they did not address the multicommodity flow case. The study in [9] pointed very briefly to the extension to multiple *s-d* pairs for joint routing and power allocation. In contrast, this current paper takes a systematic and comprehensive approach to introduce the multicommodity spectrum-efficient routing problem, which is far from being fully explored.

Multicommodity flow: finding routes for multiple *s-d* pairs simultaneously, such that a network-wide objective is optimized and the link capacities are not exceeded, is known in the computing literature as the multicommodity flow problem. In particular, given a network with capacities on the links, and a set of *s-d* pairs of nodes with associated traffic demands, the problem is to route the demand of each *s-d* pair along *exactly one* route from *s* to *d* without violating/exceeding the link capacities. This problem is known as the integer multicommodity flow problem, or the *unsplittable* flow problem, and is known to be NP-hard. Readers can refer to [10, 11] for more information. If the limitation of routing each demand along exactly one route is relaxed (i.e., each demand can be split across arbitrarily many routes), the resulting *splittable* flow problem can be solved in polynomial-time using linear programming. Readers can refer to [12] for more information. The problem addressed in this paper falls in the category of unsplittable flow, which is in general NP-hard to solve. It is worth noting that the above-mentioned unsplittable/multicommodity flow literature is devised for generic network settings, mostly applicable to wired networks. In this paper, we do not borrow any general-purpose multicommodity flow algorithm from the computing literature. However, we devise new polynomial-time algorithms that harness the special structure of the spectrum-efficient routing problem and provide *provably* optimal solutions.

1.2. Contribution and Paper Outline. In light of the above, the contribution of this paper can be summarized as follows.

(i) To the best of our knowledge, for the first time, we address the problem of spectrum-efficient routing in the case of multiple *s-d* pairs. In particular, given a multihop wireless network and an arbitrary set of *s-d* pairs, we address the problem of finding a route for every *s-d* pair such that the minimum spectral efficiency across all routes is maximized.

(ii) For the problem above, we provide two alternative approaches, where one is based on fixed-sized time slots and the other is based on variable-sized time slots. For each approach, we derive the *provably* optimal routing algorithm. En route, our study sheds the light on the arising tradeoff between the complexity of network-layer route computation versus the complexity of medium access control (MAC) layer scheduling of time slots. Our numerical results further illustrate the efficiency of the proposed approaches, and their tradeoffs.

The remainder of this paper is organized as follows. Section 2 discusses the basics and preliminaries. Section 3 presents the problem formulation and provably optimal algorithm for the fixed time slots approach. The variable-time-slots-based approach, its problem formulation, and provably optimal algorithm are presented in Section 4. Section 5 discusses the tradeoff between the two approaches. Numerical examples and results are presented in Section 6. Section 7 concludes the paper.

2. Preliminaries

To avoid the computational intractability of joint optimal routing and medium access control (MAC) layer scheduling, and following [4, 6–9], it is assumed that a common channel is shared among all nodes using TDMA without spatial reuse, i.e., each node transmits in its own unique time slot. It is demonstrated in [4] that, even though a path is selected assuming no spatial reuse/interference, applying a scheduling technique (separately) that allows some spatial reuse to the selected path can further improve the spectral efficiency. In other words, our framework can still benefit from spatial reuse. It is worth noticing that the MAC layer of the IEEE 802.16 mesh protocol, for example, is based on TDMA (see, e.g., [13]).

A multihop wireless network is modeled as a graph $G = (V, E)$, where V represents the set of nodes (vertices) and E represents the set of links (edges). We let $l \in E$ signify a link in the network. We also let $N = |V|$ and $M = |E|$ denote the number of nodes and links, respectively.

Following [4, 6, 8], we consider the setting in which all transmit devices are constrained by the same symbol-wise

average transmit power P and assume that all devices transmit with power P when transmitting. A possible justification for this assumption is that nodes in *infrastructure* wireless mesh networks are mostly immobile and connected with abundant power supplies. Therefore, for a link $l \in E$, the signal-to-noise ratio (SNR) is given by

$$SNR_l = \frac{PG_l}{N_0 B}, \tag{1}$$

where G_l is the path gain from the sender of link l to the receiver of link l, N_0 is the normalized one-sided power spectral density of the additive white Gaussian noise (at any receiver in the network), and B is the finite bandwidth of the wireless channel.

Now, assume K simultaneous *s-d* pairs are using the network. Each *s-d* pair $i = 1, 2, \ldots, K$ has a source node $s_i \in V$ and a destination node $d_i \in V$. We also let $\mathscr{L}_i$ denote the set of all routes from s_i to d_i. Moreover, we let $L_i \in \mathscr{L}_i$ signify a route from s_i to d_i.

Finally, the spectral efficiency of an arbitrary path L in the network is defined as the bandwidth-normalized end-to-end data rate [4]. In other words,

$$R(L) = \frac{C(L)}{B}, \tag{2}$$

where $R(L)$ is the spectral efficiency (in bps/Hz) of path L, $C(L)$ is the end-to-end achievable rate (in bps) for path L, and B is the channel bandwidth (in Hz).

3. Fixed-Sized Time Slots

The studies in [4, 6–9] focused on a *single s-d* route and assumed the bandwidth is shared equally among its links via TDMA. In other words, each link transmits in its own unique time slot, where the time slots are of fixed size. One way to extend this equal bandwidth sharing to the multicommodity case is to maintain the assumption that the time frame is divided equally among all links of the different *s-d* routes. In particular, if *s-d* pairs $i = 1, 2, \ldots, K$ are served by routes $L_1, L_2, \ldots, L_K$, respectively, then any link on any of the K routes will transmit for a fraction of $1/\sum_{i=1}^{K} |L_i|$ of the time, where $|L_i|$ is the number of hops/links in route L_i. In other words, the time frame will be divided into $\sum_{i=1}^{K} |L_i|$ *fixed-sized* slots, where each slot has a length $1/\sum_{i=1}^{K} |L_i|$ of the frame length.

3.1. Problem Formulation. In light of the above discussion, the TDMA end-to-end achievable data rate on path L_i can be expressed using the well-known Shannon capacity formula as

$$C(L_i) = \frac{B}{\sum_{i=1}^{K} |L_i|} \min_{l \in L_i} \log\left(1 + \frac{PG_l}{N_0 B}\right). \tag{3}$$

Note that the factor $1/\sum_{i=1}^{K} |L_i|$ comes from the sharing of the bandwidth equally among all links (of all routes), i.e., each link on any route is allocated a time fraction of $1/\sum_{i=1}^{K} |L_i|$ for transmission. Note also that the minimum function in (3) results from the fact that the end-to-end data rate of any path L_i is equal to the data rate achieved by its bottleneck link. Using (2), the spectral efficiency of path L_i can, thus, be expressed as follows:

$$R(L_i) = \frac{1}{\sum_{i=1}^{K} |L_i|} \min_{l \in L_i} \log\left(1 + \frac{PG_l}{N_0 B}\right). \tag{4}$$

Consequently, the minimum spectral efficiency, R_{min}, across all active routes $(L_1, L_2, \ldots, L_K)$ can be expressed as

$$R_{min}(L_1, L_2, \ldots, L_K) = \frac{1}{\sum_{i=1}^{K} |L_i|} \min_{l \in \{L_i : i=1,2,\ldots,K\}} \log\left(1 + \frac{PG_l}{N_0 B}\right) \tag{5}$$

Note that $\log(1 + PG_l/N_0 B)$ can be viewed as the width of any link l. Consequently, $\min_{l \in \{L_i : i=1,2,\ldots,K\}} \log(1 + PG_l/N_0 B)$ is the smallest link width used by the set of routes $L_1, L_2, \ldots, L_K$. In other words, the latter represents the width of the narrowest route among $L_1, L_2, \ldots, L_K$. To simplify our notation and algorithm development, we use the following substitution:

$$w(L_1, L_2, \ldots, L_K) = \min_{l \in \{L_i : i=1,2,\ldots,K\}} \log\left(1 + \frac{PG_l}{N_0 B}\right). \tag{6}$$

In other words, $w(L_1, L_2, \ldots, L_K)$ is the smallest link width used by the set of routes $L_1, L_2, \ldots, L_K$; i.e., it represents the width of the narrowest route among $L_1, L_2, \ldots, L_K$ using $\log(1 + PG_l/N_0 B)$ as link widths. Consequently, the minimum spectral efficiency, R_{min}, across all active routes $(L_1, L_2, \ldots, L_K)$ can be rewritten as

$$R_{min}(L_1, L_2, \ldots, L_K) = \frac{w(L_1, L_2, \ldots, L_K)}{\sum_{i=1}^{K} |L_i|}. \tag{7}$$

Therefore, the problem of jointly finding routes for *s-d* pairs $1, 2, \ldots, K$ such that the minimum spectral efficiency across all routes is maximized can be cast as the following optimization problem:

$$\max_{L_i \in \mathscr{L}_i \, \forall 1 \le i \le K} \frac{w(L_1, L_2, \ldots, L_K)}{\sum_{i=1}^{K} |L_i|}. \tag{8}$$

It is worth noting that problem (8) cannot be solved using standard shortest path methods as the resulting routing metric is not isotonic [5]. In particular, even with one *s-d* pair, the routing metric of (8) is not isotonic. See, e.g., [4, 6]. In what follows, we develop a polynomial-time algorithm that provides provably optimal solutions to (8).

3.2. Algorithm. The main idea of the proposed algorithm is an extension of the single *s-d* pair case [6]. In fact, even for multiple paths $L_1, L_2, \ldots, L_K$, the value of $w(L_1, L_2, \ldots, L_K)$ takes one of finite possible values. It is readily seen from (6) that $w(L_1, L_2, \ldots, L_K) \in W$, where W is the set of all link widths in the network; i.e.,

$$W = \left\{\log\left(1 + \frac{PG_l}{N_0 B}\right) : l \in E\right\}. \tag{9}$$

Recall that E is the set of links in the network. Since $|W| = M = O(N^2)$, $w(L_1, L_2, \ldots, L_K)$ can take at most $M = O(N^2)$ values.

The main result follows.

Theorem 1. *Let the set of routes $(L_1^*, L_2^*, \ldots, L_K^*)$ denote the optimal solution to the original multicommodity spectrum-efficient routing problem (8). Let also the set of routes $(L_1^a, L_2^a, \ldots, L_K^a)$ denote the optimal solution to the following modified problem:*

$$\max_{\substack{L_i \in \mathscr{L}_i \, \forall 1 \le i \le K \\ w(L_1, L_2, \ldots, L_K) \ge a}} \frac{a}{\sum_{i=1}^{K} |L_i|}. \tag{10}$$

Then

$$R_{min}\left(L_1^*, L_2^*, \ldots, L_K^*\right) = \max_{a \in W} R_{min}\left(L_1^a, L_2^a, \ldots, L_K^a\right). \tag{11}$$

Proof. First, let the set of routes $(\widehat{L}_1^a, \widehat{L}_2^a, \ldots, \widehat{L}_K^a)$ be an optimal solution to the following subproblem:

$$\max_{\substack{L_i \in \mathscr{L}_i \, \forall 1 \le i \le K \\ w(L_1, L_2, \ldots, L_K) = a}} \frac{w\left(L_1, L_2, \ldots, L_K\right)}{\sum_{i=1}^{K} |L_i|}. \tag{12}$$

Note that (12) is the same as (8) with the additional constraint that $w(L_1, L_2, \ldots, L_K) = a$. By the divide-and-conquer principle [14], and since the union of the sets $\{L_i \in \mathscr{L}_i : 1 \le i \le K, w(L_1, L_2, \ldots, L_K) = a\}$, over all possible values of $a \in W$, covers the route set $\{L_i \in \mathscr{L}_i : 1 \le i \le K\}$, the following is true:

$$R_{min}\left(L_1^*, L_2^*, \ldots, L_K^*\right) = \max_{a \in W} R_{min}\left(\widehat{L}_1^a, \widehat{L}_2^a, \ldots, \widehat{L}_K^a\right). \tag{13}$$

Note that $R_{min}(L_1^*, L_2^*, \ldots, L_K^*)$ and $R_{min}(\widehat{L}_1^a, \widehat{L}_2^a, \ldots, \widehat{L}_K^a)$ represent the optimal objective function values of (8) and (12), respectively.

Moreover, by substituting the equality constraint $w(L_1, L_2, \ldots, L_K) = a$ in its objective function, (12) is equivalent to

$$\max_{\substack{L_i \in \mathscr{L}_i \, \forall 1 \le i \le K \\ w(L_1, L_2, \ldots, L_K) = a}} \frac{a}{\sum_{i=1}^{K} |L_i|}. \tag{14}$$

Now, it is readily seen that (10) is a relaxation of (14). Therefore, if $(L_1^a, L_2^a, \ldots, L_K^a)$ and $(\widehat{L}_1^a, \widehat{L}_2^a, \ldots, \widehat{L}_K^a)$ are the optimal solutions to (10) and (14), respectively, then $a/\sum_{i=1}^{K} |L_i^a| \ge a/\sum_{i=1}^{K} |\widehat{L}_i^a|$. The latter implies that

$$\sum_{i=1}^{K} |L_i^a| \le \sum_{i=1}^{K} \left|\widehat{L}_i^a\right|. \tag{15}$$

Now, the following is true:

$$\begin{aligned} R_{min}\left(L_1^*, \ldots, L_K^*\right) &= \max_{a \in W} \frac{w\left(\widehat{L}_1^a, \widehat{L}_2^a, \ldots, \widehat{L}_K^a\right)}{\sum_{i=1}^{K} \left|\widehat{L}_i^a\right|} \\ &= \max_{a \in W} \frac{a}{\sum_{i=1}^{K} \left|\widehat{L}_i^a\right|} \\ &\le \max_{a \in W} \frac{w\left(L_1^a, L_2^a, \ldots, L_K^a\right)}{\sum_{i=1}^{K} |L_i^a|} \end{aligned} \tag{16}$$

Note that the first equality comes directly from (13). The second equality comes from the fact that the route set $(\widehat{L}_1^a, \widehat{L}_2^a, \ldots, \widehat{L}_K^a)$ is feasible for (12). The inequality comes from (15) and from the fact that the route set $(L_1^a, L_2^a, \ldots, L_K^a)$ is feasible for (10). Consequently, (16) implies that

$$R_{min}\left(L_1^*, L_2^*, \ldots, L_K^*\right) \le \max_{a \in W} R_{min}\left(L_1^a, L_2^a, \ldots, L_K^a\right). \tag{17}$$

Since, among all route sets $\{L_i \in \mathscr{L}_i : 1 \le i \le K\}$, $(L_1^*, L_2^*, \ldots, L_K^*)$ is the route set which maximizes the minimum spectral efficiency, (17) must hold with strict equality. This completes the proof. □

In light of Theorem 1, the multicommodity spectrum-efficient routing problem (8) can be solved using the following procedure:

(i) `For every` $a \in W$, `find the route set` $(L_1^a, L_2^a, \ldots, L_K^a)$ `by solving (10).`

(ii) `Return the route set` $(L_1^*, L_2^*, \ldots, L_K^*) = \text{argmax}_{a \in W} R_{min}(L_1^a, L_2^a, \ldots, L_K^a)$.

Recall that, in the above procedure, W is the set of link widths given by (9), and R_{min} is given by (7). The only remaining issue to show is how to solve (10). Note that, for a given $a \in W$, maximizing $a/\sum_{i=1}^{K} |L_i|$ is equivalent to minimizing $\sum_{i=1}^{K} |L_i|$. Moreover, the latter is minimized if every s-d pair i minimizes $|L_i|$. Consequently, problem (10) is equivalent to finding the minimum-hop path for every s-d pair $1 \le i \le K$, such that the minimum link width across all paths is not less than a. Therefore, for a given value of a, (10) can be solved as follows:

(i) Remove all links $l \in E$ for which $\log(1 + \text{PG}_l/N_0 B) < a$. In the remaining graph, obtain the minimum-hop path for every s-d pair $1 \le i \le K$.

In light of the above discussion, problem (8) can be solved by the following algorithm.

`Algorithm Equal-Time-Slots`

(1) `Let` $W = \{\log(1 + PG_l/N_0 B) : l \in E\}$. `For every` $a \in W$, `do:`

(a) `For every` $s - d$ `pair` $i = 1, 2, \ldots, K$, `do:`

(i) `Remove all links` $l \in E$ `for which` $\log(1 + PG_l/N_0 B) < a$.

(ii) `In the remaining graph, find` L_i^a, `the minimum-hop path from source` s_i `to destination` d_i.

(b) `Let` $R_{min}(L_1^a, L_2^a, \ldots, L_K^a) = w(L_1^a, L_2^a, \ldots, L_K^a)/\sum_{i=1}^{K} |L_i^a|$.

(2) `Return the path set with largest` $R_{min}(L_1^a, L_2^a, \ldots, L_K^a)$.

3.3. Observations. The following observations are in order regarding algorithm *Equal-Time-Slots*.

(i) Step (1a) of algorithm *Equal-Time-Slots* can be implemented by each $s-d$ pair independently, and without any coordination with the other $s-d$ pairs. In particular, $s-d$ pair i (or more precisely source node s_i) obtains its path L_i^a independently.

(ii) Step (1b), however, requires knowledge about all $s-d$ pairs. Therefore, it can be implemented by a centralized entity which knows the hop-count $|L_i^a|$ of every path L_i^a. Alternatively, it requires that all $s-d$ pairs (or source nodes) exchange their information about $|L_i^a|$ with all other nodes using flooding, or any other means of all-to-all communication.

(iii) At the MAC layer, the resulting set of paths will require dividing the time frame into $\sum_{i=1}^{K} |L_i^*|$ *equal-sized* time slots. This simplicity of MAC layer scheduling comes at the expense of the necessity of coordination between $s-d$ pairs during network-layer path computation.

(iv) The computational complexity of algorithm *Equal-Time-Slots* is dominated by the complexity of invoking a shortest path procedure $|W| = M$ times, as in step (1a). Note that step (1a) can be implemented by the different s-d pairs in parallel. Since the number of links M is of $O(N^2)$, where N is the number of nodes, and if the Dijkstra shortest path algorithm is used in every iteration, the overall complexity of algorithm *Equal-Time-Slots* is $O(N^2 \cdot N^2) = O(N^4)$.

4. Variable-Sized Time Slots

An alternative approach to accommodating multiple s-d pairs under the condition of equal bandwidth sharing is to divide the time frame equally among s-d *pairs* (as opposed to dividing the time equally among the *links*). In other words, every s-d pair/path will transmit for a fraction of $1/K$ of the time (assuming K s-d pairs). Consequently, if s-d pair i uses path L_i, then every link along this path will transmit for a fraction of $1/K|L_i|$ of the time. Since different s-d pairs may use paths with different hop-counts, their respective links may use time slots of different sizes.

In this case, the end-to-end spectral efficiency of route L_i serving s-d pair i can be expressed as

$$R(L_i) = \frac{1}{K|L_i|} \min_{l \in L_i} \log\left(1 + \frac{PG_l}{N_0 B}\right). \tag{18}$$

Note that the factor $1/K|L_i|$ comes from the fact that every link $l \in L_i$ transmits for a fraction of $1/K|L_i|$ of the time. Note also that the minimum function in (18) results from the fact that the end-to-end data rate of any path L_i is equal to the data rate achieved by its bottleneck link. It is worth noting that the spectral efficiency for s-d pair i depends on the hop-count $|L_i|$ of its own path L_i only. This is in contrast to the case of equal time slots, where the spectral efficiency for any s-d pair i depends on the hop-counts of all s-d paths $L_1, L_2, \ldots, L_K$.

The problem of maximizing the minimum spectral efficiency across all s-d pairs can, thus, be expressed as

$$\max_{L_i \in \mathscr{L}_i \, \forall 1 \le i \le K} \min_{1 \le i \le K} R(L_i), \tag{19}$$

where $R(L_i)$ is given by (18). It is not hard to see that the minimum spectrum efficiency will be maximized if every s-d pair i maximizes its individual spectral efficiency $R(L_i)$. In other words, every s-d pair i solves the following optimization problem:

$$\max_{L_i \in \mathscr{L}_i} \frac{1}{K|L_i|} \min_{l \in L_i} \log\left(1 + \frac{PG_l}{N_0 B}\right). \tag{20}$$

Moreover, since, for any number of s-d pairs, K is a constant, solving (20) is equivalent to solving the single s-d pair problem. The best known algorithm for solving (20) has been introduced in [8], and can be summarized as follows.

`Algorithm Variable-Time-Slots`
`For every` s-d `pair` $i = 1, 2, \ldots, K$, `do:`

(1) `For` $h = 1, 2, \ldots, N-1$, `do:`

(a) `Find` L_i^h, `the widest path with at most` h `hops connecting` s_i `to` d_i, `using` $\log(1 + PG_l/N_0B)$ `as link labels.`

(b) `Let` $R(L_i^h) = (1/K|L_i^h|) \min_{l \in L_i^h} \log(1 + PG_l/N_0B)$.

(2) `Return the path with largest` $R(L_i^h)$.

The following observations are in order regarding algorithm *Variable-Time-Slots*.

(i) The algorithm can be implemented by each s-d pair in a completely independent manner. In other words, s-d pairs are completely isolated and there is no need for coordination and/or a centralized component.

(ii) The number of s-d pairs K does not affect the computation of the optimal paths. In particular, for any number of s-d pairs, K is a constant, and thus

$$\begin{aligned} &\operatorname*{argmax}_{L_i \in \mathscr{L}_i} \frac{1}{K|L_i|} \min_{l \in L_i} \log\left(1 + \frac{PG_l}{N_0 B}\right) \\ &\quad = \operatorname*{argmax}_{L_i \in \mathscr{L}_i} \frac{1}{|L_i|} \min_{l \in L_i} \log\left(1 + \frac{PG_l}{N_0 B}\right). \end{aligned} \tag{21}$$

In other words, every s-d pair computes its optimal path regardless of how many other s-d pairs exist, and the optimal paths do not change with the change of the number of s-d pairs.

(iii) The number of s-d pairs K, however, is needed for MAC layer scheduling. In particular, every s-d pair transmits for $1/K$ of the time, and every link $l \in L_i$ used by s-d pair i transmits for a fraction of $1/K|L_i|$ of the time.

(iv) Since the paths used by different s-d pairs may have different hop-counts, the resulting time slots may be of different durations.

(v) The algorithm can be implemented by the different *s-d* pairs in parallel. Note also that shortest path algorithms can be modified to compute the widest path. See, e.g., [15]. Moreover, it is worth noticing that the Bellman-Ford shortest path algorithm, in its h^{th} iteration, computes the shortest (or widest) path with at most h hops. Consequently, algorithm *Variable-Time-Slots* can be implemented by invoking the Bellman-Ford procedure only *once*. The overall complexity of algorithm *Variable-Time-Slots* is, thus, $O(N^3)$.

Note that *dynamically* partitioning the TDMA frame into a set of *variable-length* transmission slots is possible. See, e.g., [9, 16]. As an example of the resulting MAC strategy, let the number of *s-d* pairs be 3. Assume also that applying algorithm *Variable-Time-Slots* results in paths L_1, L_2, and L_3 for the 3 *s-d* pairs, where $|L_1| = 3$, $|L_2| = 2$, and $|L_3| = 4$, respectively. The time frame will, thus, be divided into $3 + 2 + 4 = 9$ slots. The normalized slot size is $1/(3 \times 3) = 1/9$ for the first 3 slots (used by path L_1), $1/(3 \times 2) = 1/6$ for the following 2 slots (used by path L_2), and $1/(3 \times 4) = 1/12$ for the last 4 slots (used by path L_3). In other words, the normalized sizes of the 9 slots are 1/9, 1/9, 1/9, 1/6, 1/6, 1/12, 1/12, 1/12, and 1/12, respectively.

In contrast, however, assume that algorithm *Equal-Time-Slots* results in paths L_1, L_2, and L_3 for the 3 *s-d* pairs, where $|L_1| = 3$, $|L_2| = 2$, and $|L_3| = 4$, respectively. In this case, the time frame will be divided into $3 + 2 + 4 = 9$ slots, where the normalized size of each slot is precisely 1/9.

5. Tradeoffs

In this section we summarize the main differences and tradeoffs between algorithms *Equal-Time-Slots* and *Variable-Time-Slots*.

(i) **Distributed implementation**: algorithm *Equal-Time-Slots* has a component that requires centralized knowledge of the paths computed by *all s-d* pairs (in every iteration of the algorithm). Alternatively, coordination and exchange of information is required among *s-d* pairs. Algorithm *Variable-Time-Slots*, however, can be implemented in a fully distributed fashion among *s-d* pairs. No coordination is required among *s-d* pairs in their path computation.

(ii) **Synchronization**: as described above, algorithm *Equal-Time-Slots* is based on the centralized (or coordinated) knowledge of the paths L_i^a for all *s-d* pairs $i = 1, 2, \ldots, K$. This implicitly implies that all existing *s-d* pairs must make their path computation decisions (i.e., invoke their algorithms) in a synchronized way; any *s-d* pair cannot obtain its optimal path without the results of the other *s-d* pairs. Algorithm *Variable-Time-Slots*, however, does not require this synchronization. Any *s-d* pair can compute its optimal path regardless of the other *s-d* pairs. In fact, optimal path computation does not even require the knowledge of the number of existing *s-d* pairs K. The optimal path in case of only *one s-d* pair would remain optimal in the presence of *any* number of *s-d* pairs K. Knowing K is necessary in the transmission (MAC) scheduling phase only.

(iii) **MAC scheduling**: algorithm *Equal-Time-Slots* results in TDMA transmission frames with equal time slots, while algorithm *Variable-Time-Slots* results in TDMA transmission frames with variable time slots.

(iv) **Complexity**: the complexity of *Equal-Time-Slots* is $O(N^4)$, while the complexity of *Variable-Time-Slots* is $O(N^3)$.

It is worth noting that both algorithms (*Equal-Time-Slots* and *Variable-Time-Slots*) require the value of the link SNRs (PG_l/N_0B) to be known at the source nodes computing their respective paths. In practice, the link SNR can be directly measured by received signal strength indicators available on most devices [4], and fed back to the transmitters. Nodes can then exchange their knowledge about the values of PG_l/N_0B for their outgoing links using a *distributed* link-state protocol. Please refer to [9] for further elaboration.

In the following section we provide a numerical study on the performance of both proposed approaches and their tradeoffs.

6. Numerical Results

We consider multihop wireless networks, in which the nodes are located at random positions in a 100×100 two-dimensional area. Without loss of generality, it is assumed that any two nodes can directly communicate; i.e., the network is fully connected. Note that, from an information theoretic point of view, two nodes can always communicate at a sufficiently low rate [4, 17]. The path gain G_l of each link l is assumed to be given by

$$G_l = c \cdot A_l \cdot \left(\max\{d_l, d_0\}\right)^{-4}, \tag{22}$$

where d_l is the length of link l, d_0 is the reference distance for the far-field, A_l is a log-normally distributed random variable (with 0-dB mean and 8-dB log-variance) that reflects shadowing, and c is a constant. Without loss of generality, we set $d_0 = 0.1$ and $c = 0.01$. We test our proposed algorithms on *random* and *independent* network realizations, where in each realization the horizontal and vertical coordinates of each node are chosen randomly (and independently) according to a uniform distribution between 0 and 100, and the path gains are generated randomly (and independently) according to (22). Among the randomly generated nodes, a set of *s-d* pairs is chosen at random. Furthermore, for each tested scenario, we average our results over 10^4 random network realizations. In other words, *every point in each of the following result figures is averaged over* 10^4 *random network realizations*.

The simulation parameters are summarized in Table 1 (note that $E(\cdot)$ and $V(\cdot)$ denote the expected value and variance of a random variable, respectively).

6.1. Effect of the Network Size. First, we vary the number of nodes in the network (N) from 5 to 30. For every value of N,

Table 1: Simulation parameters.

Parameter	Value
Path gain (G_l)	Equation (22)
Path gain constant (c)	0.01
Far-field reference distance (d_0)	0.1
Shadowing (A_l)	log-normal distribution
$E(10 \log A_l)$	0 dB
$V(10 \log A_l)$	8 dB
Node positions	random in 100×100 area
Number of nodes (N)	5 to 30
Number of s-d pairs (K)	5 to 20
Network SNR (P/N_0B)	-20 dB to 80 dB
Simulation software	MATLAB R2017a

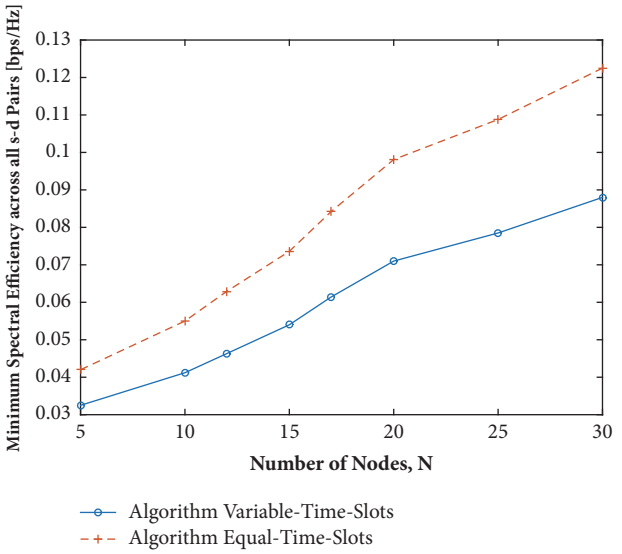


Figure 1: Minimum spectral efficiency among all s-d routes obtained using algorithms *Equal-Time-Slots* and *Variable-Time-Slots*, respectively, versus the number of network nodes.

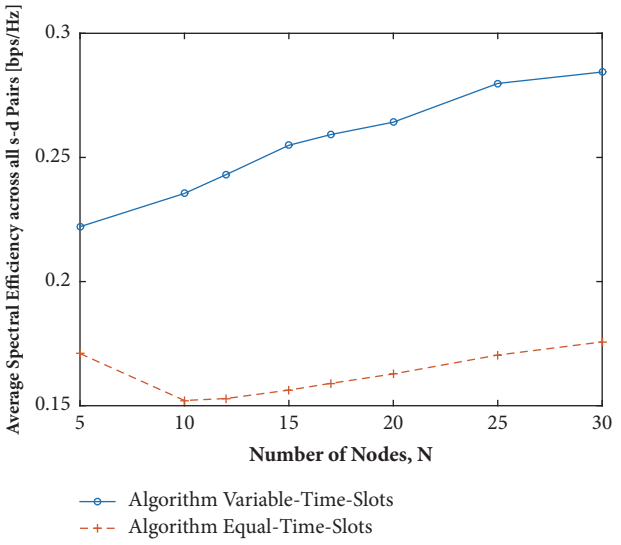


Figure 2: Average spectral efficiency among all s-d routes obtained using algorithms *Equal-Time-Slots* and *Variable-Time-Slots*, respectively, versus the number of network nodes.

we let the number of s-d pairs (K) be 5, and we set the network SNR (P/N_0B) to 80 dB. Figure 1 depicts the *minimum* spectral efficiency among all s-d routes obtained using algorithms *Equal-Time-Slots* and *Variable-Time-Slots*, respectively. It is clearly seen that algorithm *Equal-Time-Slots* results in higher worst-case spectral efficiencies. In particular, the minimum source-destination (s-d) spectral efficiency resulting from algorithm *Equal-Time-Slots* is from 29.23% to 39.2% higher than that of algorithm *Variable-Time-Slots*. Averaged over all experiments for different values of N, the minimum s-d spectral efficiency resulting from algorithm *Equal-Time-Slots* is 36% higher than that of algorithm *Variable-Time-Slots*.

Moreover, Figure 2 depicts the *average* spectral efficiency across all s-d routes obtained using algorithms *Equal-Time-Slots* and *Variable-Time-Slots*, respectively. It is straightforward to see that algorithm *Variable-Time-Slots* results in higher average spectral efficiencies. In particular, the average s-d spectral efficiency resulting from algorithm *Variable-Time-Slots* is from 29.88% to 64.2% higher than that of algorithm *Equal-Time-Slots*. Averaged over all experiments for different values of N, the average s-d spectral efficiency resulting from algorithm *Variable-Time-Slots* is 57.27% higher than that of algorithm *Equal-Time-Slots*. In short, although algorithm *Equal-Time-Slots* has a better worst-case performance (as seen in Figure 1), algorithm *Variable-Time-Slots* has a significantly better average performance (as seen in Figure 2).

Finally, we provide a comparison between the running times of algorithms *Equal-Time-Slots* and *Variable-Time-Slots* per s-d pair. For fairness of comparison, we compare the overall running time of algorithm *Variable-Time-Slots* with the running time of the distributed component of algorithm *Equal-Time-Slots* (i.e., Step (1a), which can be implemented by each s-d pair in isolation). In other words, the running time of the centralized component of algorithm *Equal-Time-Slots* is excluded from the comparison. The running times of both algorithms are depicted in Figure 3. In fact, the (per

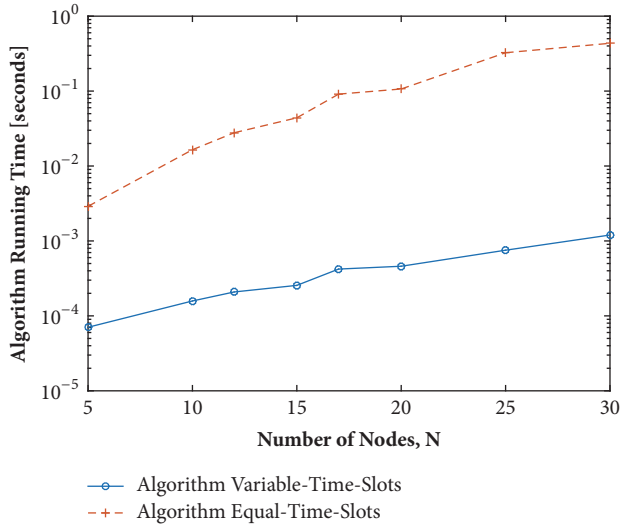


FIGURE 3: Running times of algorithms *Equal-Time-Slots* and *Variable-Time-Slots*.

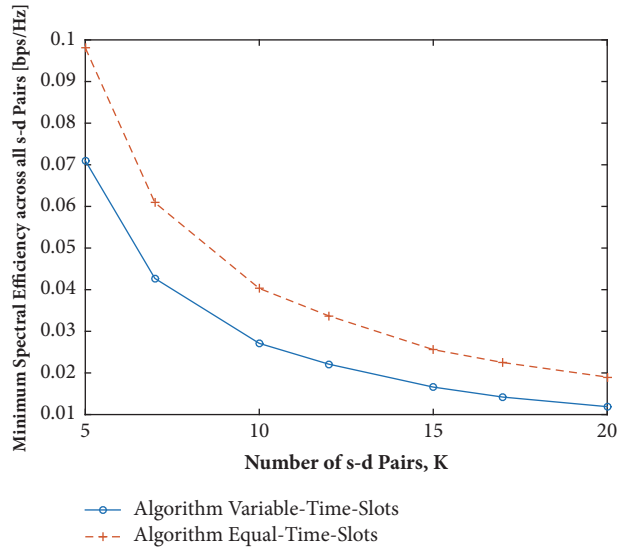


FIGURE 4: Minimum spectral efficiency among all *s-d* routes obtained using algorithms *Equal-Time-Slots* and *Variable-Time-Slots*, respectively, versus the number of *s-d* pairs.

s-d pair) average running time of algorithm *Variable-Time-Slots* is 0.44 milliseconds, while that of algorithm *Equal-Time-Slots* is 0.13 seconds. In other words, although the centralized component of algorithm *Equal-Time-Slots* was not considered in this comparison, the running time of algorithm *Variable-Time-Slots* is on average 99.23% lower than that of algorithm *Equal-Time-Slots*.

6.2. Effect of Number of s-d Pairs. Now, we vary the number of *s-d* pairs (K) from 5 to 20, while the number of nodes is fixed at $N = 20$ and the network SNR is fixed at 80 dB. The results for minimum and average spectral efficiencies across all *s-d* routes are depicted in Figures 4 and 5, respectively. Again, it can be easily seen that algorithm *Equal-Time-Slots* has a better worst-case performance (as seen in Figure 4),

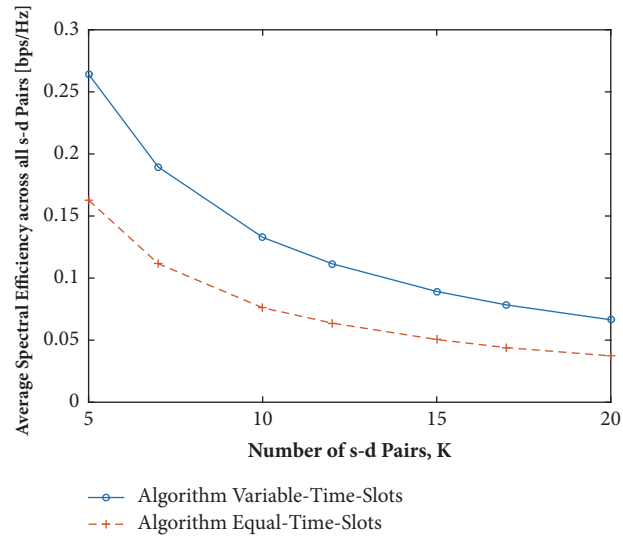


FIGURE 5: Average spectral efficiency among all *s-d* routes obtained using algorithms *Equal-Time-Slots* and *Variable-Time-Slots*, respectively, versus the number of *s-d* pairs.

while algorithm *Variable-Time-Slots* has a significantly better average performance (as seen in Figure 5). In particular, the minimum spectral efficiency across all *s-d* routes obtained by algorithm *Equal-Time-Slots* is from 38.17% to 59.66% higher than that obtained by algorithm *Variable-Time-Slots*. Averaged over all experiments, the minimum spectral efficiency resulting from algorithm *Equal-Time-Slots* is 50.62% higher than that resulting from algorithm *Variable-Time-Slots*. On the other hand, the average spectral efficiency across all *s-d* routes resulting from algorithm *Variable-Time-Slots* is from 62.29% to 78.19% higher than that resulting from algorithm *Variable-Time-Slots*. Averaged over all experiments, algorithm *Variable-Time-Slots* results in 73.50% higher average spectral efficiencies than algorithm *Equal-Time-Slots*.

6.3. Effect of the Network SNR. Now, we vary the network SNR from -20 dB to 80 dB, while the number of nodes is fixed at $N = 20$ and the number of *s-d* pairs is fixed at $K = 5$. The results for minimum and average spectral efficiencies across all *s-d* routes are depicted in Figures 6 and 7, respectively. In consistency with all other results, algorithm *Equal-Time-Slots* consistently shows a better worst-case performance (as seen in Figure 6), while algorithm *Variable-Time-Slots* shows a consistently and significantly better average performance (as seen in Figure 7). In particular, the improvement in *worst-case* spectral efficiencies due to algorithm *Equal-Time-Slots* is between 37.96% and 38.24% (with an average improvement of 37.93% across all experiments). However, the improvement in *average* spectral efficiencies due to algorithm *Variable-Time-Slots* is between 65.18% and 136.37% (with an average improvement of 129.20% across all experiments).

6.4. Comparison against Benchmarks. Finally, we assess the performance of our proposed algorithms in comparison to existing techniques. To this end, we use the following two benchmarks.

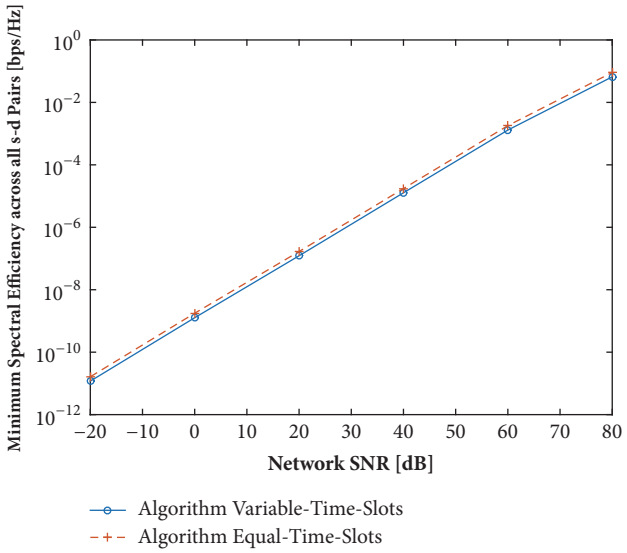


FIGURE 6: Minimum spectral efficiency among all *s-d* routes obtained using algorithms *Equal-Time-Slots* and *Variable-Time-Slots*, respectively, versus the network SNR.

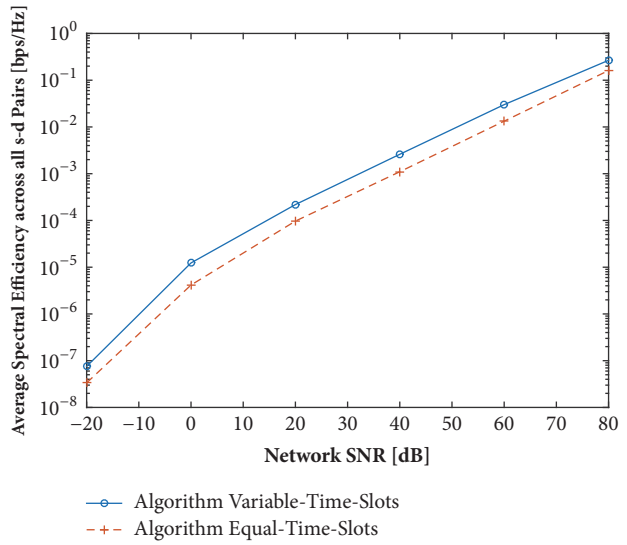


FIGURE 7: Average spectral efficiency among all *s-d* routes obtained using algorithms *Equal-Time-Slots* and *Variable-Time-Slots*, respectively, versus the network SNR.

(i) We compare our algorithms against the spectral efficiency resulting from routing every *s-d* pair along the direct link from s to d. Since all resulting routes are one hop long, dividing the time frame equally between *s-d* pairs or between transmission links is equivalent. Therefore, direct link routing is compared against both proposed algorithms *Equal-Time-Slots* and *Variable-Time-Slots*.

(ii) We also compare our algorithms against the spectral efficiency resulting from routing every *s-d* pair *independently* using the distributed spectrum-efficient routing (DSER) algorithm introduced in [4]. DSER [4] operates by simply finding a shortest path from the source to the destination using $(1 + 2^{\gamma}/SNR_l)$ as the link metric, where γ is the path-loss exponent. Here, $\gamma = 4$. It is assumed that the time frame is divided equally among *s-d* pairs, resulting in variable-length time slots for individual link transmissions. Therefore, DSER is compared against our proposed algorithm *Variable-Time-Slots*.

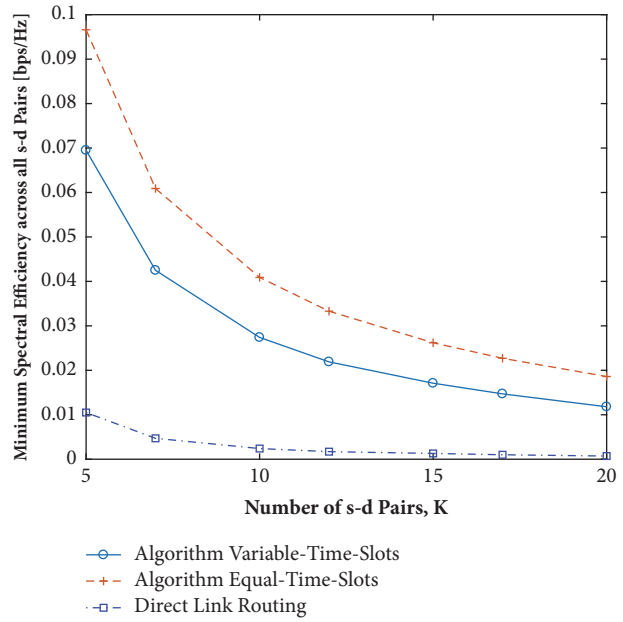


FIGURE 8: Minimum spectral efficiency among all *s-d* routes obtained using algorithms *Equal-Time-Slots* and *Variable-Time-Slots* compared against direct link routing.

To compare them against direct link routing, we vary the number of *s-d* pairs K from 5 to 20, while the number of nodes is fixed at $N = 20$ and the network SNR is fixed at 80 dB. The results for minimum and average spectral efficiencies across all *s-d* routes are depicted in Figures 8 and 9, respectively. Following the same trend, algorithm *Equal-Time-Slots* shows a better worst-case performance, while algorithm *Variable-Time-Slots* shows a better average performance. In particular, the worst-case spectral efficiency resulting from algorithm *Equal-Time-Slots* is 820% to 2500% higher than that resulting from direct routing (with an average improvement of 1724% across all experiments). Moreover, the average spectral efficiency resulting from algorithm *Variable-Time-Slots* is about 30% higher than that resulting from direct routing in all experiments.

To compare against DSER from [4], we vary the number of nodes in the network (N) from 5 to 30, while the number of *s-d* pairs (K) is set to 5 and the network SNR is set to 80 dB. The results for minimum and average spectral efficiencies across all *s-d* routes are depicted in Figures 10 and 11, respectively. The superior performance of our proposed algorithm *Variable-Time-Slots* is clearly seen. In particular, our algorithm results in 93% to 743% higher worst-case spectral efficiencies as compared to direct link routing (with an average improvement of 446% across all experiments) and results in 16% to 68% higher worst-case spectral efficiencies

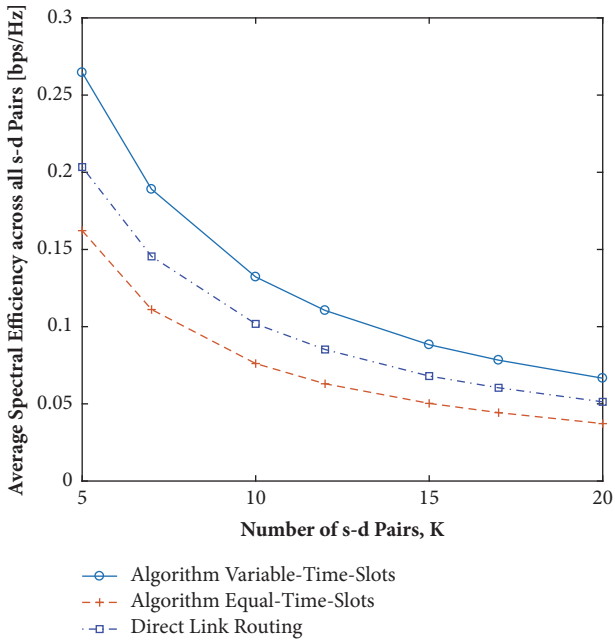


Figure 9: Average spectral efficiency among all *s-d* routes obtained using algorithms *Equal-Time-Slots* and *Variable-Time-Slots* compared against direct link routing.

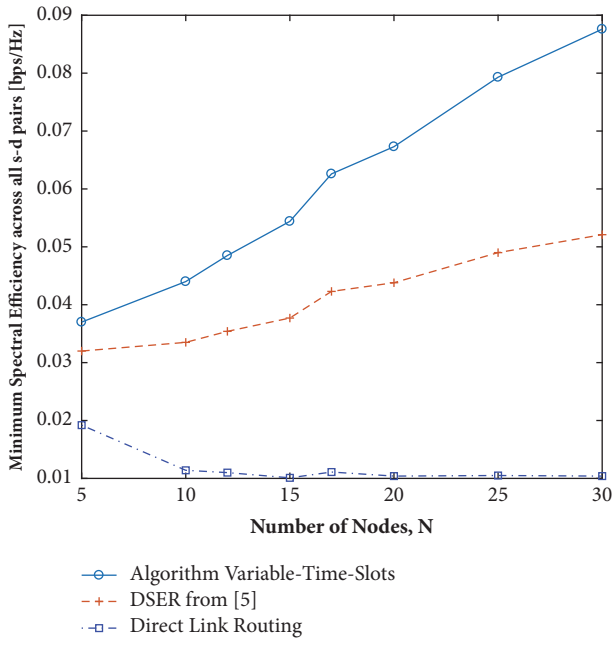


Figure 10: Minimum spectral efficiency among all *s-d* routes obtained using algorithm *Variable-Time-Slots* compared against DSER from [4] and direct link routing.

as compared to DSER (with an average improvement of 45% across all experiments). Moreover, our algorithm results in 6.5% to 39% higher average spectral efficiencies as compared to direct link routing (with an average improvement of 24% across all experiments) and results in 2.9% to 27% higher

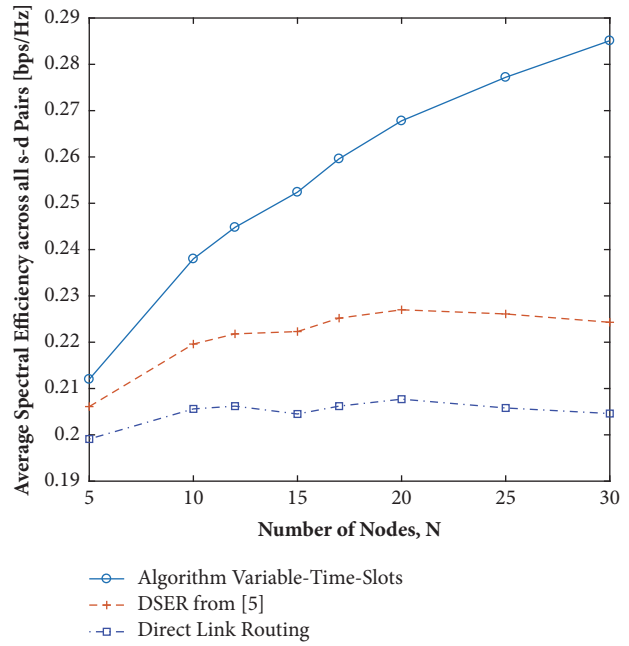


Figure 11: Average spectral efficiency among all *s-d* routes obtained using algorithm *Variable-Time-Slots* compared against DSER from [4] and direct link routing.

average spectral efficiencies as compared to DSER (with an average improvement of 15% across all experiments).

In summary, all our experiments (of varying the network size, number of *s-d* pairs, and network SNR) indicate a similar trend of a better worst-case performance for algorithm *Equal-Time-Slots* versus a better average/typical performance for algorithm *Variable-Time-Slots*. It is worth noting, however, that the improvement in average results due to algorithm *Variable-Time-Slots* is always more significant. Moreover, algorithm *Variable-Time-Slots* enjoys a significantly (more than 99%) lower running time. Moreover, our proposed algorithm *Variable-Time-Slots* has shown a significantly superior worst-case and average performance as compared to DSER from [4], and as compared to direct link routing.

7. Conclusion

To the best of our knowledge, previous work on finding the path with maximum end-to-end spectrum efficiency was restricted to a single *s-d* pair. This paper proposed two alternative approaches for the spectrum-efficient routing problem in the multicommodity flow regime, i.e., in the case of multiple active *s-d* pairs. The routing objective was to maximize the minimum spectrum efficiency achieved across all active *s-d* pairs. The first approach was based on dividing the time frame into equal-sized slots, while the second approach allows dividing the time frame into variable-sized slots. For each approach, we derived the *provably* optimal routing algorithm. We also shed the light on the arising tradeoff between the resulting routing algorithms. In summary, the routing algorithm induced by equal time slots enjoys a better worst-case performance (i.e., higher

worst-case spectrum efficiencies), at the price of a higher computational complexity and the existence of a centralized component requiring coordination and synchronization among all *s-d* pairs in the route computation phase. However, the routing algorithm induced by variable time slots has the advantages of (1) a significantly lower computational complexity (more than 99% reduction in running time), (2) a significantly better average/typical performance (i.e., higher average achieved spectral efficiencies), (3) a significantly better worst-case and average spectral efficiency performance as compared to existing methods, and (4) being entirely distributed with no need for coordination or synchronization among *s-d* pairs. In fact the routing algorithm induced by variable time slots does not even require the knowledge of the number of active *s-d* pairs; the optimal path in case of the existence of a single *s-d* pair remains optimal in case of any number of *s-d* pairs. Knowledge of the number *s-d* pairs is required only in the phase of MAC layer scheduling of time slots. This concludes that algorithm *Variable-Time-Slots* might be preferred for practical implementation.

Data Availability

The details of how the numerical experiment scenarios were generated are clearly explained in Section 6 of the manuscript and can easily be regenerated.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This research was supported in part by the Distributed and Networked Systems Research Group, University of Sharjah, Operating Grant no. 150410, and in part by a University of Sharjah Targeted Project no. 1602040336-P.

References

[1] J. G. Andrews, S. Buzzi, and W. Choi, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[2] W. Feng, Y. Li, D. Jin, L. Su, and S. Chen, "Millimetre-Wave Backhaul for 5G Networks: Challenges and Solutions," *Sensors*, vol. 16, no. 6, p. 892, 2016.

[3] D. Airehrour, J. Gutierrez, and S. K. Ray, "Secure routing for internet of things: A survey," *Journal of Network and Computer Applications*, vol. 66, pp. 198–213, 2016.

[4] D. Chen, M. Haenggi, and J. Laneman, "Distributed spectrum-efficient routing algorithms in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5297–5305, 2008.

[5] J. L. Sobrinho, "Algebra and algorithms for QoS path computation and hop-by-hop routing in the Internet," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp. 541–550, 2002.

[6] M. Saad, "Optimal spectrum-efficient routing in multihop wireless networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 5822–5826, 2009.

[7] M. Saad, "Reduced-complexity spectrum-efficient routing in TDMA multihop wireless networks," in *Proceedings of the 2010 IEEE Symposium on Computers and Communications (ISCC)*, pp. 617–621, Riccione, Italy, June 2010.

[8] M. Saad, "On optimal spectrum-efficient routing in TDMA and FDMA multihop wireless networks," *Computer Communications*, vol. 35, no. 5, pp. 628–636, 2012.

[9] M. Saad, "Joint optimal routing and power allocation for spectral efficiency in multihop wireless networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2530–2539, 2014.

[10] M. Martens and M. Skutella, "Flows on few paths: algorithms and lower bounds," *Networks. An International Journal*, vol. 48, no. 2, pp. 68–76, 2006.

[11] T. H. Szymanski, "Max-flow min-cost routing in a future-internet with improved QoS guarantees," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1485–1497, 2013.

[12] F. Shahrokhi and D. W. Matula, "The maximum concurrent flow problem," *Journal of the ACM*, vol. 37, no. 2, pp. 318–334, 1990.

[13] P. Djukic and S. Valaee, "Delay aware link scheduling for multihop TDMA wireless networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 3, pp. 870–883, 2009.

[14] L. A. Wolsey, *Integer Programming*, John Wiley & Sons, New York, NY. USA, 1998.

[15] R. Guérin and A. Orda, "Computing shortest paths for any number of hops," *IEEE/ACM Transactions on Networking*, vol. 10, no. 5, pp. 613–620, 2002.

[16] H. Zhang and P. Gburzynski, "A variable slot length TDMA protocol for personal communication systems," *Wireless Personal Communications*, vol. 22, no. 3, pp. 409–432, 2002.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1991.