

Secondary User Scheduling Under Throughput Guarantees for the Primary Network

Dževdan Kapetanović, M. Majid Butt, Symeon Chatzinotas and Björn Ottersten
Interdisciplinary Center for Security, Reliability and Trust (SnT)
University of Luxembourg, Luxembourg
Email: {dzevdan.kapetanovic, majid.butt, symeon.chatzinotas, bjorn.ottersten}@uni.lu

Abstract—This work addresses scheduling in a cognitive radio scenario where a minimum throughput for the downlink primary network (PN) is guaranteed to each user with an associated violation probability (probability of not obtaining the guaranteed throughput). The primary network is surrounded by multiple downlink secondary networks which aiming to maximize its network throughput. Scheduling in PN is performed independent of the secondary networks. Some information about the PN is available at the central scheduler that is responsible for scheduling the secondary networks. The contribution of this work is to apply a novel scheduler to the PN which is more robust to QoS degradations resulting from the secondary networks than other state of the art schedulers. This is validated by numerical simulations of the cognitive radio network.

I. INTRODUCTION

Quality of service (QoS) guarantees in terms of minimum throughput is one of the goals for the network service providers as minimum throughput guarantees determines the worst case performance of the underlying systems. In practice, it is not possible to provide a minimum throughput guarantee with zero violation probability; which is defined as the probability that a users average throughput is below the minimum throughput. Most of the work aims at providing these guarantees on the long term basis. However, the size of the practical window to calculate the average throughput is quite small, and provides new challenges for resource allocation mechanisms.

This scheduling problem has been widely investigated for independent systems [1], but has not been addressed in a cognitive radio (CR) context. As shown through measurement campaigns, a large percentage of the available spectrum is under-utilized either temporally or spatially [2]. In this case, a primary system could share a frequency band with a secondary system which uses the resources opportunistically so that it does not degrade the primary QoS. Based on the side (cognitive) information available at the secondary scheduler, three classes of transmission strategies can be identified, namely interweave, underlay and overlay [3]. We consider interweave and underlay scenarios and describe them in Section II.

In this paper, we consider scheduling users in a primary macrocell and scheduling of a set of secondary coordinated femtocells. This system model has been already considered in [4], [5] focusing on physical layer interference mitigation. More specifically, a cognitive interference alignment has been

proposed for enabling the primary-secondary coexistence by exploiting multiuser MIMO schemes. In this contribution, we investigate cognitive scheduling algorithms which maximizes secondary throughput while protecting the primary QoS using interweave and underlay techniques.

The existing literature addresses network scheduling in different settings, e.g., the authors in [6] discuss a framework for maximizing energy efficiency in terms of bits per joule for a minimum throughput guarantee. Reference [7][8] address specifically violation probability for a given minimum throughput guarantee in the practical window case.

Similar issues have been addressed in CR domain as well. The authors in [9] formulate the throughput maximization problem in a network subject to energy consumption restrictions and an energy consumption minimization problem subject to minimum throughput guarantees. The throughput of the secondary network (SN) in an interweave scenario is characterized in [10] under the constraint that minimum throughput of the PN is guaranteed.

Most of the existing work focuses on maintaining a minimum throughput guarantee for the PN with zero violation probability. In practice, this is hard due to unpredictable fading channels. Our work investigates the problem in the settings where a non-zero violation probability in minimum throughput is allowed for the PN and defined as a QoS parameter. Secondly, the time scale for averaging the throughput for a primary user is of the order of few tens (e.g. 16 for WIMAX [8]) and averaging of throughput over infinite time gives misleading results some times as shown in [1]. We characterize the tradeoff between gain in average throughput of the SN against a (small) loss in terms of violation probability at a guaranteed throughput (per user) for the PN.

The rest of the paper is organized as follows. We discuss the system model in our work in Section II. Section III presents the scheduling scheme and cognitive network scenarios. We evaluate the performance of our schemes through numerical results in Section IV and Section V concludes with the main contributions of the work.

II. SYSTEM MODEL AND PRELIMINARIES

We consider a PN downlink system, operating in time division multiplex (TDM) mode, where the scheduler schedules one user per time slot. The PN is unaware of multiple

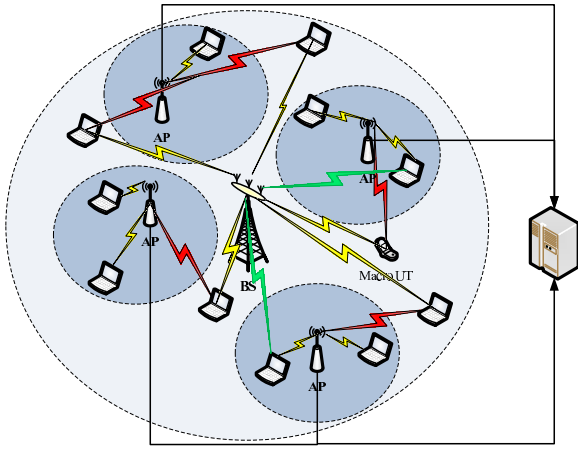


Fig. 1. A primary downlink, TDM system with macro user terminals (UT) operating in the same frequency band as multiple secondary networks/access points (APs). Yellow signals are downlink signals in PN and SN, red is interference from SN base station to primary users and green interference from PN base station to SNs users.

secondary networks (e.g., femtocells) that coexist in the same frequency band as downlink TDM systems. This scenario is depicted in Figure 1. A central scheduler (CS) has perfect knowledge of each SN (obtained from cooperation with the base stations), and is responsible for scheduling one user in one of the secondary networks for transmission in a time slot¹. We assume N primary users in the PN, M secondary networks and K secondary users in each SN (the main outcome of this work holds also when the number of users differ in the secondary networks). h_j , $1 \leq j \leq N$, denotes the channel coefficient of a primary user during a time slot, while \hat{h}_{jk} the channel coefficient of the k^{th} user in the j^{th} SN. A channel coefficient is characterized by a fast fading environment, where the channel outcomes remain constant during the time span of a time slot and are independent and identically distributed (i.i.d.) across the time slots and the users in a network. This model is called a *block fading model*. Hence, the users in all networks are statistically symmetrical with respect to the channel distribution. Let f_{jk} denote the channel from the primary base station to the k^{th} user in the j^{th} SN. Similarly, let \hat{f}_{jk} be the channel from the base station of the j^{th} SN to the k^{th} user in the PN. It is assumed that these cross channels are temporally and spatially i.i.d..

In terms of Channel State Information (CSI), the primary base station has perfect knowledge of h_j , $1 \leq j \leq N$, while the secondary base stations and the CS have perfect knowledge of \hat{h}_{jk} , $1 \leq j \leq M$, $1 \leq k \leq K$. This could be achieved by training in each of the networks. Moreover, the CS has perfect knowledge of \hat{f}_{jk} , but it has no knowledge of f_{jk} , $1 \leq j \leq M$, $1 \leq k \leq K$. The reason for this is that the secondary base stations could learn the channels from each primary user during the uplink training phase in the PN, and transmit this knowledge to the CS via a dedicated link (reciprocity of the channels is assumed to hold). However, for the interference

from the primary base station to each secondary user, imposing a similar dedicated link from the primary base station to the CS is not feasible. In addition, we assume that each SN can detect which primary user is scheduled for each time slot, based on the scheduling information usually contained in the frame headers of the transmitted packets.

Let $R_k[t]$ denote the achievable rate of user k at time slot t in bits/channel use. The average throughput $\bar{T}_k[t]$ of user k up to time slot t is defined as

$$\bar{T}_k[t] \triangleq \frac{\sum_{j=1}^{L_W} R_k[t-j]}{L_W}, \quad (1)$$

where $R_k[t-1], \dots, R_k[t-L_W]$ are the rates allocated to user k during the last L_W time slots. Due to TDM, some of these rates can be zero, which occurs in the time slots where user k was not scheduled. Hence, the average throughput for a user is calculated across a window of L_W time slots.

The violation probability for a user k is defined as

$$\delta_k(T_G) \triangleq \lim_{t \rightarrow \infty} \frac{\sum_{j=1}^t I(\bar{T}_k[j] < T_G)}{t}. \quad (2)$$

where $I(\cdot)$ denotes a standard indicator function that is one if the argument is true, zero otherwise. It is clear from the definition of δ_k that $0 \leq \delta_k \leq 1$. The constant T_G is the minimum guaranteed throughput, i.e., the least rate guaranteed to each user. Hence, δ_k is the probability that the average throughput of the user k falls below the guaranteed throughput.

The PN scheduler aims at minimizing the violation probability given a certain T_G , while the CS aims at maximizing the average sum rate of each SN without severely degrading the primary transmission. To achieve this, two well known CR techniques are investigated next.

A. Interweave Transmission

In this scheme, a SN transmits only when an empty time slot is detected, e.g., through spectrum sensing. It is assumed that no false alarms or missed detection occur in this context. As a result, for each time slot i , either a primary user n receives

$$y_n[i] = h_n[i]x_n[i] + z[i] \quad (3)$$

if the PN transmits, or secondary user k in network j receives

$$\hat{y}_{j,k}[i] = \hat{h}_{j,k}[i]\hat{x}_{j,k}[i] + \hat{z}[i] \quad (4)$$

if the SN j is scheduled for transmission. Here, $x_n[i]$ and $\hat{x}_{j,k}[i]$ are the symbols to user n in the PN and user k in the j^{th} SN at time slot i , with average transmission power of P_p and P_s , respectively. $z[i]$ and $\hat{z}[i]$ are complex AWGN noise samples with noise variance N_0 at time slot i . If a primary user is scheduled at time slot i , its achievable rate is

$$R_n[i] = \log_2(1 + P_p|h_n[i]|^2/N_0), \quad (5)$$

while it is zero for all secondary users. Similarly, if a secondary user is scheduled instead, its achievable rate is

$$R_{j,k}[i] = \log_2(1 + P_s|\hat{h}_{j,k}[i]|^2/N_0), \quad (6)$$

while primary users get zero rate.

¹This corresponds to TDM across the secondary networks, each operating with TDM

B. Underlay Transmission

In this scheme, the secondary system can transmit in the same time slot as the primary system, as long as the induced interference towards the scheduled primary user is below a certain threshold I . Since the CS is aware of the channels $\hat{f}_{j,k}$ from the secondary base stations toward the primary users, it can calculate the induced interference, which equals $P_s |\hat{f}_{j,k}|^2$, and use this information for scheduling. The received signals at the scheduled primary user n and the scheduled secondary user k in the j^{th} network, in time slot i , are

$$y_n[i] = h_n[i]x_n[i] + \hat{f}_{j,n}[i]\hat{x}_{j,k}[i] + z[i] \quad (7)$$

and

$$\hat{y}_{j,k}[i] = \hat{h}_{j,k}[i]\hat{x}_{j,k}[i] + f_{j,k}[i]x_n[i] + \hat{z}[i], \quad (8)$$

respectively. In contrast to interweave transmission, the achievable rate at primary user n in time slot i is given by

$$R_n[i] = \log_2(1 + P_p |h_n[i]|^2 / (N_0 + P_s |\hat{f}_{j,n}|^2)), \quad (9)$$

and for secondary user k in network j in time slot i by

$$R_{j,k}[i] = \log_2(1 + P_s |\hat{h}_{j,k}[i]|^2 / (N_0 + P_p |f_{j,k}|^2)). \quad (10)$$

Hence, due to cross interference between the networks, the instantaneous rate is reduced compared to interweave transmission. However, an advantage is that at each time slot, both the PN and some SN receive a non-zero rate, which is not the case for interweave transmission.

C. Background: Throughput Deadline Scheduling

In [1], a scheduling algorithm, Throughput Deadline Scheduling (TDS), was proposed to maximize the user throughput guarantees for a given violation probability in a TDMA based wireless network. This algorithm is the base for the scheme proposed in this work. Therefore, we briefly review this scheme. We define the term *throughput deadline* $D_k[t]$ for a user k at time slot t as the maximum number of time slots available until his average throughput falls below T_G if he is *not scheduled* continuously².

In a time slot t , a user is scheduled for transmission if he/she maximizes

$$k^* = \arg \max_k g_k[t] R_k[t] \quad (11)$$

where the priority function $g_k[t]$ is given by

$$g_k[t] = \begin{cases} \frac{1}{(D_k[t])^\alpha} \left(\frac{\hat{D}_k[t+1]+1}{D_k[t]} \right)^\beta & \bar{T}_k[t] > T_G \\ (\hat{D}_k[t+1]+1)^\beta & \bar{T}_k[t] \leq T_G. \end{cases} \quad (12)$$

where $\hat{D}_k[t+1]$ denotes the new deadline of the user if scheduled in time slot t .

In the absence of secondary networks, the scheme proposed in [1] provides the currently best known solution to the optimization problem

$$\begin{aligned} \text{Max } T_G^P \\ \text{s.t. } \delta_k(T_G^P) \leq c, \quad 1 \leq k \leq N, \quad 0 \leq c \leq 1 \end{aligned} \quad (13)$$

²For the exact computation, please refer to [1].

where c denotes a constant and T_G^P is the minimum guaranteed throughput to each user. Since each user experiences same channel conditions, any reasonable scheduler produces $\delta_k(T_G^P) = \delta(T_G^P)$, $1 \leq k \leq N$ (i.e., each user experiences the same violation probability). In [1], it has been shown numerically that the scheme outperforms proportional fairness scheduling (PFS) and other state of the art algorithms in the finite window case.

III. PROBLEM FORMULATION AND SCHEDULING

In cognitive settings, the PN is guaranteed T_G^P with a violation probability³ $\delta(T_G^P)$ in the absence of secondary networks. The PN aims to maintain δ for the same T_G^P in the presence of SNs, ideally. However, the *degradation* in performance, in terms of increase in δ , is unavoidable for a given T_G^P due to presence of secondary networks. The resource sharing between the primary and the secondary networks should not degrade the throughput guarantees of the PN severely. In this work, we characterize the degradation in throughput guarantee for the PN as a function of $\delta_d(T_G^P)$ (henceforth denoted as δ_d , implicitly assuming its dependence on T_G^P), which defines incremental violation probability due to presence of SNs. Thus, the effective violation probability of the PN in presence of SNs is $\delta_0 = \delta + \delta_d$. The network designer guarantees a violation probability δ_0 for a throughput T_G^P to the PN, where δ_d is the margin exploited by the SNs in order to increase their throughput.

For a fixed T_G^P, δ, δ_d tuple at the PN, the objective of each SN is to maximize its average sum rate $\bar{T}_s^j(T_G^P, \delta, \delta_d)$, defined as

$$\bar{T}_s^j(T_G^P, \delta, \delta_d) \triangleq \lim_{t \rightarrow \infty} \frac{\sum_{k=1}^t R_{j,n_k}[k]}{t},$$

where n_k is the scheduled user at time slot k in the j^{th} SN. Thus, the optimization problem for the j^{th} SN is

$$\begin{aligned} \text{Max } \quad & \bar{T}_s^j(T_G^P, \delta, \delta_d) \\ \text{s.t. } \quad & \delta_d \leq \epsilon \end{aligned} \quad (14)$$

It should be noted that we do not solve the optimization problems in (13) and (14) jointly but compute the solution sequentially as the PN is supposed to optimize its resources independent of the SN. As the objective is to maximize the average throughput of the SN, the CS employs the Maximum Throughput Scheduler (MTS) for the SNs and schedules the secondary user with the best instantaneous rate $R_k[t]$ [11] out of the joint set of all SNs. An example of such secondary networks can be machine to machine communication networks, where QoS requirements usually are not stringent in terms of latency, and an improvement in average throughput at the cost of small degradation in QoS of the PN is acceptable. Thus, in our problem setting, CS is fixed and the gain is achieved by selecting the appropriate scheduler in the PN that achieves a low violation probability for PN users.

³From here on denoted as δ , implicitly understood that it depends on T_G^P .

A. Scheduling: Interweave Transmission

The interweave approach to achieve a good solution to (14) is for the PN to abandon a time slot in favor of a SN. In this work, this is done by the following rule

$$\max_k g_k^p[t] \leq \gamma(T_G^P, \delta, \delta_d), \quad (15)$$

where $g_k^p[t]$ is here interpreted as a priority function for user k , with the property that if the user has a relatively large average throughput, then the priority function is small, otherwise it is large. The priority function depends on the scheduler that is used. For TDS, $g_k^p[t]$ is given by (12), while it is different for other schedulers; e.g., for PFS, $g_k^p[t] = 1/\bar{T}_k[t]$. The value of the upper bound $\gamma(T_G^P, \delta, \delta_d)$ is such that the rule in (15) gives rise to a degradation δ_d for the specific T_G^P, δ pair in the PN. Hence, δ_d is determined by the upper bound for a fixed T_G^P, δ pair, and vice versa. Increasing the upper bound in (15) increases the degradation δ_d in (14), since time slots will be abandoned more frequently by the PN. The increase in the upper bound will also increase the average sum rate per SN, since the secondary users will be scheduled more often. Note that once a time slot is abandoned by the PN, all primary users get zero rate at that time slot. Instead, if some interference from the secondary networks to the PN is allowed, the primary users will get a non-zero rate at every time slot. This is the rationale of underlay transmission.

B. Scheduling: Underlay Transmission

If underlay transmission is applied to (14), δ_d is determined from the interference generated by a secondary base station j towards a primary user k in each time slot. Thus, the constraint in (14) translates into

$$|\hat{f}_{j,k}|^2 < I_0(T_G^P, \delta, \delta_d) \quad (16)$$

where $I_0(T_G^P, \delta, \delta_d)$ is the maximum interference allowed for a given T_G^P, δ, δ_d tuple, that satisfies the constraint $\delta_d \leq \epsilon$.

In the proposed scheme, we allow scheduling of at most one of the SNs in a given time slot. We select a set of users from each SN which satisfy the constraint in (16), and apply the MTS scheme to schedule one user out of this joint set of users. Hence, at any time slot, one user from one SN is scheduled.

We do not mathematically solve the optimization problems formulated in (13) and (14) due to presence of finite parameters like number of users, window size, T_G^P, δ_d involved. We evaluate the tradeoff between gain in sum rate in SN against loss in terms of violation probability for the PN in Section IV.

IV. PERFORMANCE EVALUATION AND DISCUSSIONS

Monte Carlo simulations are performed across 150 000 time slots in order to evaluate the performance of the cognitive techniques discussed above. Optimal values for the constants in (12), for different values of T_G^P, L_W and N , are taken from [1]. Comparisons of TDS with the PFS scheduler in [12] and the Adaptive Scheduling Algorithm 2 (ASA2) from [8] are made. ASA2 has been proposed specifically for finite window

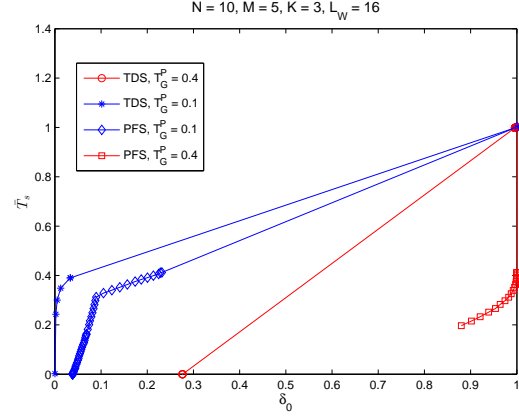


Fig. 2. Interweave with TDS and PFS schedulers at the PN. Clearly, TDS outperforms PFS. Note the discontinuity in outage probability, arising from the finite window size.

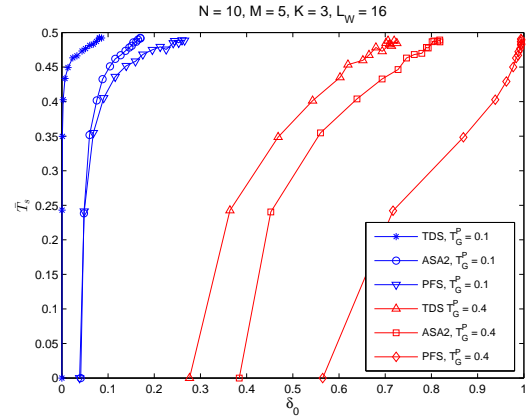


Fig. 3. Underlay with a window size $L_W = 16$. TDS outperforms the competing schedulers.

case where MTS algorithm is applied only to the users with non-zero violation probability. Once a user gets guaranteed throughput after scheduling, he remains out of contention. In all figures, the average transmission power at the PN and SNs is 10 dB. Moreover, a PN with 10 users surrounded by 5 SNs each having 3 users is assumed. This corresponds to a typical femtocell environment, where several femtocells with small amount of users surround a larger PN. Due to the symmetry of the problem, all SNs end up having the same average sum rate, which is denoted as \bar{T}_s in the figures. In each figure, \bar{T}_s versus δ_0 is illustrated, which characterizes the tradeoff between SN gains and the degradation in violation probability for the PN. By continuously increasing the upper bounds in (15) and (16), the degradation δ_d , along with \bar{T}_s , increases and thus also the violation probability δ_0 experienced by the PN. In all figures, for each T_G^P and scheduling method, the initial violation probability δ (attained when SNs are absent) equals the starting value on the x-axis.

Figure 2 shows the performance of the interweave scheme for window size $L_W = 16$ (Mobile WIMAX standard) and different values of T_G^P . For all values, TDS outperforms PFS, while a discontinuous "jump" occurs at some point for both schedulers. This jump results from the finite window size,

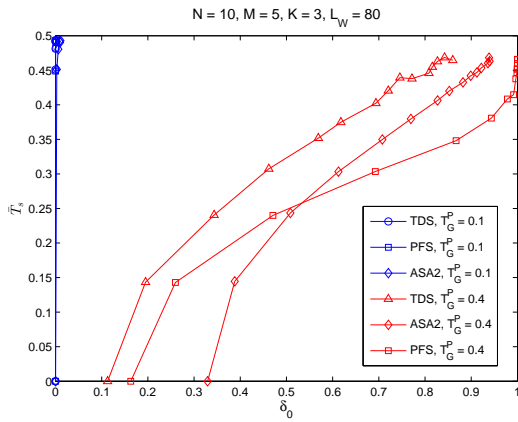


Fig. 4. Underlay with a window size $L_W = 80$. Note that all schedulers provide maximum average throughput at T_G^P with no violation probability degradation.

which makes the functions $g_k^P[t]$, $1 \leq k \leq N$, take on discrete rather than continuous values. It follows from (12) that $g_k[t]$ for TDS always takes on a finite number of values, since the throughput deadline is an integer. For PFS, $g_k[t] = 1/\bar{T}_k[t]$, and since $\bar{T}_k[t]$ is calculated across a finite window, it will have discontinuous jumps (although theoretically the value can be any real number). Once the upper bound in (15) becomes large enough, the functions $g_k[t]$ will rarely be larger than the upper bound, which results in that the schedulers fail and users are always in outage.

In Figure 3, the different schedulers are compared with the underlay technique. Again, TDS has the best performance, followed by ASA2 and PFS. The curves saturate at a violation probability smaller than 1, since the average interference to the primary system from any secondary user is finite, and thus the rates in the PN are large enough for escaping continuous outage. Underlay is also illustrated in Figure 4, but with a window size of $L_W = 80$ instead (European Winner I project). As seen, for $T_G^P = 0.1$, the SN throughput is maximized at virtually zero outage for all schedulers, showing their robustness to interference for small T_G^P . For higher values of T_G^P , TDS still exhibits best performance.

Comparison between the interweave and underlay approaches is presented in Fig. 5. Underlay has a significantly better performance than interweave, with TDS again showing best performance, followed by ASA2 and PFS. The interweave curves terminate just before the discontinuity point (when δ_0 becomes 1). Hence, the interweave scheme is not able to achieve the same maximum \bar{T}_s as underlay for any $\delta_0 < 1$.

V. CONCLUSIONS

This paper investigated a downlink CR scenario, with the aim of maximizing the average sum rate of SNs, under QoS constraints for the PN. The QoS constraint is a violation probability for a guaranteed minimum throughput to each primary user, and depends heavily on the scheduler used in the PN. Both interweave and underlay cognitive techniques are investigated with different schedulers at the PN. Our results demonstrate that the recently proposed TDS scheduler outperforms other state of the art schedulers for both interweave

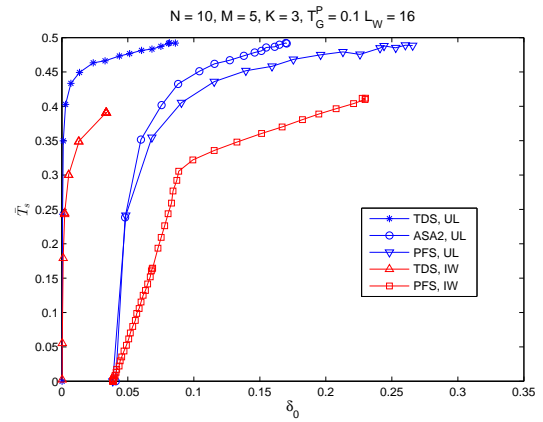


Fig. 5. Comparison between interweave (IW) and underlay (UL). Interweave is unable to provide same network throughput as underlay for $\delta_0 < 1$.

and underlay techniques, providing a larger SN sum rate for a given violation probability and minimum throughput guarantee at the PN. Moreover, underlay outperforms interweave in spite of interference created by the SNs.

ACKNOWLEDGEMENT

This work was supported by the National Research Fund, Luxembourg under the CORE project CO2SAT: Cooperative and Cognitive Architectures for Satellite Networks.

REFERENCES

- [1] M. M. Butt, D. Kapetanovic, and B. Ottersten, "Maximizing minimum throughput guarantees: The small violation probability region," *IEEE Wireless Communication Letter*, vol. PP, no. 99, pp. 1–4, 2013.
- [2] K. Patil, R. Prasad, and K. Skouby, "A survey of worldwide spectrum occupancy measurement campaigns for cognitive radio," in *International Conference on Devices and Communications*, 2011, pp. 1–5.
- [3] A. Goldsmith, S. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894–914, 2009.
- [4] B. O. S.K. Sharma, S. Chatzinotas, "Cognitive interference alignment in heterogeneous networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, 2013.
- [5] S. Chatzinotas and B. Ottersten, "Cognitive interference alignment between small cells and a macrocell," in *19th International Conference on Telecommunications (ICT)*, 2012, pp. 1–6.
- [6] C. Li, S. H. Song, J. Zhang, and K. Letaief, "Maximizing energy efficiency in wireless networks with a minimum average throughput requirement," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Shanghai, China, April 2012.
- [7] V. Hassel, G. E. Øien, and D. Gesbert, "Throughput guarantees for wireless networks with opportunistic scheduling: a comparative study," *IEEE Transactions on Wireless Communications*, vol. 6, no. 12, pp. 4215–4220, 2007.
- [8] J. Rasool, V. Hassel, S. de la Kethulle de Ryhove, and G. Oien, "Opportunistic scheduling policies for improved throughput guarantees in wireless networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, no. 1, p. 43, 2011.
- [9] S. Bayhan and F. Alagoz, "Scheduling in centralized cognitive radio networks for energy efficiency," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 2, pp. 582–595, 2013.
- [10] J.-P. Hong and W. Choi, "Throughput characteristics by multiuser diversity in a cognitive radio system," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3749–3763, 2011.
- [11] R. Knopp and P. Humblet, "Information capacity and power control in single cell multiuser communications," in *IEEE, Int. conf. on communications*, Seattle, WA, June 1995.
- [12] P. Viswanath, D. N. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 46, no. 6, pp. 1277–1294, June 2002.