Dynamic Spectrum Allocation in Cognitive Radio: Throughput Calculations

Ronald Nissel[†, ‡] and Markus Rupp[†]
[†] Institute of Telecommunications, Technische Universität Wien, Vienna, Austria
[‡] Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion, ITC, TU Wien
Email: {rnissel, mrupp}@nt.tuwien.ac.at

***Abstract*—In this paper, we use queuing theory (egalitarian process sharing) to derive throughput expressions for dynamic spectrum allocation in cognitive radio systems. In particular, we consider different priority groups, for example, secondary users who can access the channel only if no primary user is active. The aggregated throughput, a measure for social optimality, is also investigated. Moreover, we calculate the user throughput for the case that two network operators share their spectrum by cognitive radio technologies. Primary users of operator 1 are then secondary users of operator 2 and vice versa. The optimal load distribution between primary channel and secondary channel is given by the Nash equilibrium.**

I. Introduction

The number of mobile cellular subscriptions in the European Union exceeds the number of its citizen, proving the great success of mobile communications. However, an increasing number of network applications require even higher data rates, so that each generation of wireless systems has to catch up. While the main objective of 2G was to build a robust and relatively simple voice centric system, 3G and 4G [1] focused on increasing the data rate, employing a large range of possible methods. One promising method which has not been included in the standard so far (due to technical challenges) is cognitive radio [2]. Although cognitive radio consists of different techniques, among them, for example, spectrum sensing, we will keep our analyses simple by focusing on dynamic spectrum allocation. Currently, dynamic spectrum allocation only happens at the user level within a single network operator. Depending on the instantaneous throughput demand of subscribers, the available time-frequency resources are scheduled to specific users in a dynamic way. In contrast to that, the bandwidth of different network operators and services is fixed and assigned by a regulatory authority. However, such fixed spectrum allocation is highly inefficient because the actual utilization is relatively low. Cognitive radio allows a better utilization by allowing secondary users to access the bandwidth in an opportunistic way, that is, if no primary users are active. Many authors, such as [3], expect that cognitive radio will be employed in the next generation of wireless systems (5G), at least to some extent. However, in order to fully exploit all advantages of cognitive radio, some technical challenges need to be solved and maybe new modulation techniques, such as filter bank multicarrier modulation [4], [5], might be necessary.

To evaluate the performance of dynamic spectrum allocation we consider the throughput, defined as average data packet size over average time needed to transmit packets. However, we have to be careful using such a throughput definition in the context of priority groups. For example, secondary users cannot transmit data once primary users require access, leading to an instantaneous throughout of zero. If such event happens often, secondary users might be dissatisfied, irrespectively of an otherwise high instantaneous throughput. In terms of scheduling, we assume an egalitarian processor sharing, that is, all active users share the available bandwidth equally. Clearly, a higher number of active users reduces the bandwidth share of each user, which leads to a smaller individual throughput.

One of the most remarkable results in information theory is the channel capacity, that is, the rate at which error free transmission is possible [6], [7]. For an Additive White Gaussian Noise (AWGN) channel, the capacity C depends linearly on the bandwidth and logarithmically on the Signal-to-Noise Ratio (SNR). Thus, bandwidth and capacity have the same meaning (besides a scaling factor). The maximal data rate is also limited by the technical standard which can be interpreted as capacity as well, although not in an information theoretic sense. However, because spectrum allocation is the main focus of this paper, we keep the analysis simple by assuming a fixed and known capacity C, irrespectively of different definitions and, in particular, independent of the SNR.

Our most important assumptions are:

- The throughput is defined as average data packet size over average time needed to transmit data packets.
- The starting time for data transmissions is modeled by a Poisson process whereas each user has the same arrival rate λ. Thus, n users have the arrival rate $n\lambda$.
- The packet size is exponentially distributed with mean σ.
- All active users share the bandwidth equally (no signaling overhead, no delay) which can be modeled by process sharing in the context of queuing theory.

Novel contribution:

Firstly, we extend the throughput considerations of [8] so that they also account for an opportunistic access of different priority groups, for example, secondary users.

Secondly, we determine the througput for the special case that two network operators jointly share their spectrum. In particular, we investigate the optimal load distribution between primary and secondary channel.

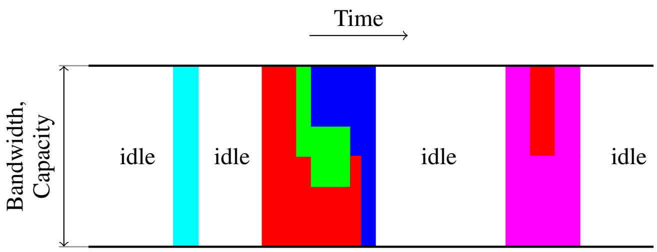

Fig. 1. Dynamic spectrum allocation (statistical multiplexing) means that the capacity is shared (equally) among active users. The instantaneous user throughput then depends on how many users are active and is thus a statistical variable. Note that it can also happen that there are no active users, that is, idle times.

II. Dynamic Spectrum Allocation

Dynamic spectrum allocation (statistical multiplexing) is based on the fact that not all users are active at the same time. There are even times when all users are inactive. Consider for example n users who all have to share the same bandwidth. Usually, only a small subset of these users are active users (they transmit data). If static spectrum allocation were used, every user would only get a throughput of C/n, even if only one user is active. In dynamic spectrum allocation on the other hand, such user would get the whole bandwidth. Only when all users are active at the same time, they will have to share the available bandwidth according to C/n. Thus, statistical multiplexing can never be worse than static frequency multiplexing (ignoring practical issues such as signaling overheads). Figure 1 illustrates the basic concept of dynamic spectrum allocation. Different colors correspond to different active users and the area of the color determines the transmitted data size. Depending on how many users are active, the available capacity is shared (equally) among them so that the user throughput becomes a statistical variable. We model the throughput as continuous flows which are limited by the data size and the available bandwidth. However, in reality such flows are not continuous but discrete. Let us for example consider Long-Term Evolution (LTE) which employs Orthogonal Frequency Division Multiplexing (OFDM) as modulation technique [9]. OFDM can be thought of transmitting symbols over a rectangular time-frequency grid whereas each of such resource elements has a size of approximately 15 kHz $\times$ 71μs. For a 10 MHz LTE signal, only 9 MHz can be used, so that 9 MHz/15kHz=600 subcarriers are available. In theory, each of this subcarriers could be assigned to a specific user at each time interval of 71 μs. However, in order to keep the signaling overhead small, the base station assigns only resource blocks consisting of 12 subcarriers $\times$ 7 OFDM symbols, corresponding to 180 kHz $\times$ 0.5 ms. There exist different strategies to assign the resource blocks to specific users [10], whereas round robin scheduling is the most practical one due to its simplicity. Here, all active users receive the same share of the time-frequency resources without taking the channel quality into account. Our dynamic spectrum allocation model that shares the bandwidth equally reflects exactly such round robin scheduling.

III. Throughput in Cognitive Radio

As already mentioned in Section II, the user throughput is a statistical variable which depends on the number of active users. As performance measure we define the user throughput as average packet size over average time needed to transmit packets. Waiting time and capacity are interchangeable. For example, one can transmit from $0\ldots T/2$ at rate zero and from $T/2\ldots T$ at a rate C, or one can transmit at rate $C/2$ over the whole time interval T. In both cases, the same amount of data is transmitted in the intervall $0\ldots T$. The mean response time of our process sharing system can therefore be modeled by a queuing system (Erlang C) with one service, that is, an M/M/1 queue. The analyses of such queuing system dates back to the beginning of the 20th century and describes the waiting time in telephone networks: the probability that in the next time interval $\varDelta t$, the number of current calls is increased by one, is given by $\lambda\,\varDelta t$, with λ being the arrival rate. Similar, the probability that one call is cleared in the same time interval is given by $\mu\,\varDelta t$, where μ represents the service rate. Interpreting this as a birth-death process for the limit case of $\varDelta t \longrightarrow 0$ allows us to write such system as a Markov chain with the corresponding steady state equilibrium probability that k users are active:

$$P(k \text{ users active}) = \rho^k(1-\rho), \tag{1}$$

where ρ is the traffic load. If there are n users who all have the same arrival rate λ, then the overall arrival rate is given by $n\lambda$ and therefore the load ρ by:

$$\rho = \frac{n\lambda}{\mu} = \frac{n\lambda\sigma}{C}, \tag{2}$$

whereas for the service time μ^{-1} we use the relation $\mu^{-1} = \frac{\sigma}{C}$, with σ being the average data packet size. Note that a steady state equilibrium exists only if $0 \leq \rho < 1$. The M/M/1 queuing system delivers a certain average waiting time. Because our throughput measure is defined as expected packet size over expected response time, the user throughput γ for our dynamic spectrum allocation system becomes, [8]:

$$\gamma = C(1-\rho). \tag{3}$$

By comparing (3) with (1) we recognize that the throughput is given by the capacity times the probability that no user is active ($k=0$). An important property of γ is the following: Suppose m operators of equal size merge (they have the same number of subscribers and the same bandwidth). According to (3), the throughput then increases by a factor of m. As comparison, for static frequency multiplexing the throughput would stay constant. Thus, from this point of view, a single operator provides a higher throughput than many operators because a single operator can better exploit the statistical multiplexing gain.

So far, we recapped well known results from queuing theory. Let us now consider the case of cognitive radio, where we assume $i = 1\ldots I$ priority groups, each having n_i users. The lower the group number i, the higher the priority. For example, the primary users $i = 1$ have priority over all other users. On the other hand, secondary users $i = 2$ can transmit data only if no primary user is active, but have priority over tertiary users. Because primary users are not effected by lower priority users, the primary user throughput γ_1 is the same as in (3), with $\rho = \frac{n_1 \lambda \sigma}{C}$. For secondary user, the available maximum throughput (what is left from the primary users) is given by γ_1. Similar as in (3), the secondary user throughput is then given by this maximum throughput times the probability that neither primary users nor secondary users are active, that is, $(1 - [\rho_1 + \rho_2])$, leading to:

$$\gamma_2 = C(1 - \rho_1)(1 - [\rho_1 + \rho_2]). \tag{4}$$

Besides our considerations above, an alternative way of deriving (4) can be obtained by using the equations in [11] for the special case of $g_1 = 1$, $g_2 = 0$ and by considering the throughput definition given in [8], which relates the average waiting time to the throughput. Similar to (3), we assume that the traffic load in (4) is smaller than one, that is, $[\rho_1 + \rho_2] < 1$. If this is not true, the underlying Markov chain has no steady state equilibrium. This corresponds to the case when the rate of connection buildups is so high that the channel is completely congested (the inflow is higher than the outflow), leading to a throughput of zero.

Similar as for the secondary user throughput, we find the throughput for the i-th priority group by:

$$\gamma_i = \begin{cases} C \prod_{k=1}^{i} \left(1 - \sum_{l=1}^{k} \rho_l\right) & \text{if } 0 \le \sum_{l=1}^{i} \rho_l < 1 \\ 0 & \text{otherwise} \end{cases}, \tag{5}$$

where $\rho_i = \frac{n_i \lambda \sigma}{C}$ represents the traffic load for priority group i. Figure 2 plots the user throughput γ_i for $I = 5$ priority groups. To keep the illustration simple, we assume that all user groups have the same traffic load (number of users). For a very small traffic load, all user achieve approximately the same maximum throughput, given by C. If the traffic load is higher than a certain value, the througput becomes zero due to the strong congestion of the system. Let us for example consider the case of secondary users. Then, a traffic load of $\frac{1}{2}$ leads to a secondary user throughput of zero, although half of the time the spectrum is idle. This is because we consider a steady state equilibrium and the inflow of new data requests from secondary users is too high, leading to the congestion. Note that this is only true if all priority groups have the same traffic load. Indeed, if we allow a smaller number of secondary users, it is possible to utilize the unused spectrum. Similar considerations are true for higher priority groups.

Besides the individual user throughput shown in Figure 2, a good measure for social optimality is the aggregated throughput, defined as:

$$\Gamma_I = \sum_{i=1}^{I} n_i \gamma_i. \tag{6}$$

Here, n_i represents the number of users in priority group i and γ_i the throughput, given in (5). Note that the aggregated

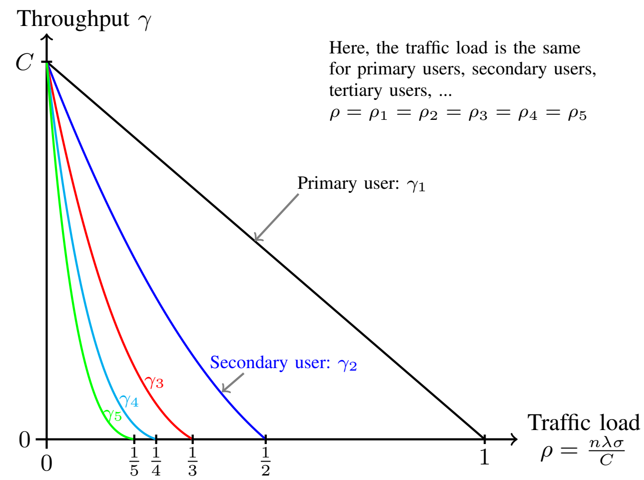

Fig. 2. User throughout for different priority groups. For example, secondary users can access the channel only when no primary user is active, resulting in a lower throughput. If the traffic load is too high, congestion causes the throughput to become zero.

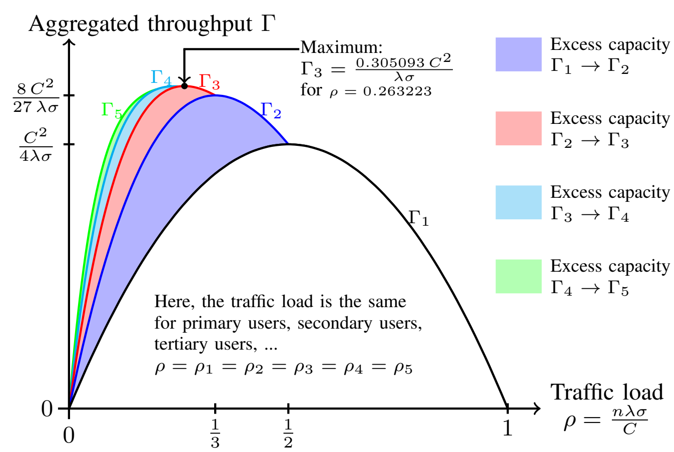

Fig. 3. The aggregated throughput, a measure for social optimality, shows a maximum for a specific traffic load. If there exist only primary users, the maximum aggregated throughput is achieved for a traffic load of $\frac{1}{2}$. However, the overall maximum requires at least three priority groups.

throughput in (6) does not reflect the sum throughput that users can achieve at the same time, which would be the capacity C. The aggregated throughput is more a measure for social optimality and includes the fact that not all users are active at the same time. Figure 3 shows the aggregated throughput for the special case that all priority groups have the same traffic load (number of users). In general, for an increasing traffic load, the aggregated throughput increases, reaches a maximum and then decreases again due to congestion of the system. Because priority groups of higher order (they have a lower priority) utilize the spectrum in an opportunistic way, the aggregated throughput Γ_I can never we worse than Γ_J with $J < I$. The shaded areas describe the excess capacity which can be utilized in an opportunistic way by lower priority users (higher i). Note that the maximum aggregated throughput is given by $\Gamma_3 = \frac{0.31\, C^2}{\lambda\sigma}$ and achieved for a traffic load of

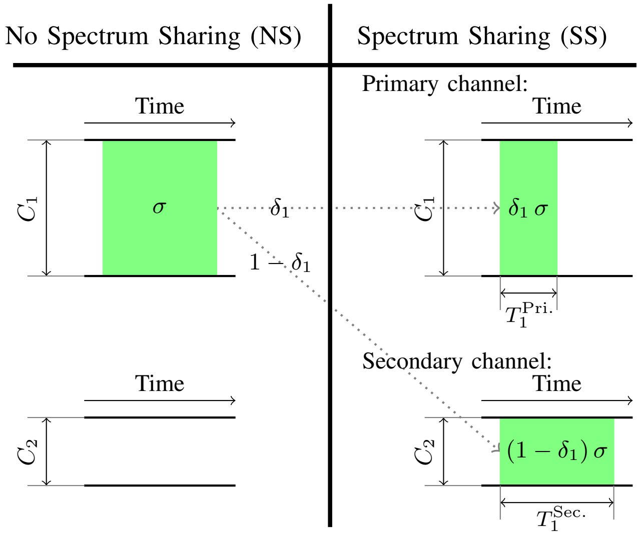

Fig. 4. A packet of size σ is split between primary channel and secondary channel. Note that throughput maximization requires $T_1^{\text{Pri.}} = T_1^{\text{Sec.}}$. Because the required time in the secondary channel depends also on the choices of the other operator, we analyze the outcome using non-cooperative game theory. The Nash equilibrium then corresponds to the optimal load distribution between primary channel an secondary channel.

$\rho = 0.26$ for $I = 3$ priority groups (additional priority groups have no effect because their throughput is zero). Of course, this is only true if all priority groups have the same traffic load.

IV. Joint Spectrum Sharing

In this section we derive the user throughput for the case when two network operators, with capacity C_1, respectively C_2, jointly share their spectrum using cognitive radio. We assume that users of operator 1 can access the capacity of operator 2 in an opportunistic way, that is, only if no user of operator 2 is active. Thus, users of operator 1 are secondary users of the capacity of operator 2, as described in Section III. Conversely, the same concept applies for operator 2, that is, users of operator 2 are secondary users of the capacity of operator 1. Unfortunately, the throughput equations of Section III cannot be straightforwardly applied to describe such spectrum sharing. To understand why this is the case, let us for example consider the throughput in Figure 2. A straightforward approach would be to simply add both throughput measures. For a traffic load larger than $\frac{1}{2}$, we immediately recognize the problem of such a simple throughput definition: it coincides with the primary users throughput due to congestion. However, because "secondary users" have an own primary channel, they could distribute parts of their load to the primary channel, so that congestion is avoided, leading to a higher throughput. We therefore need a more evolved method to describe spectrum sharing, whereas such load distribution lies in the heart of it. Figure 4 shows the basic concept for operator 1: a data packet of size σ is split using the packet size distribution factor $0 \leq \delta_1 \leq 1$, so that the primary channel has to process a packet of size $\delta_1\sigma$ while the secondary channel has to deal with the remaining packet of size $(1-\delta_1)\sigma$. Because the throughput is defined as the average data size over the average time needed to transmit the packet, we can use (3) to determine the time $T_1^{\text{Pri.}}$ a packet of size $\delta_1\sigma$ requires to be transmitted over the primary channel. In the same way, (4) can be used to determine the required time $T_1^{\text{Sec.}}$ for the secondary channel which has to process a packet of size $(1-\delta_1)\sigma$. Doing so allows us to write the required times as:

$$T_1^{\text{Pri.}} = \frac{\delta_1\sigma}{C_1 - \delta_1\sigma\lambda n_1} \tag{7}$$

$$T_1^{\text{Sec.}} = \frac{(1-\delta_1)\sigma}{(C_2 - \delta_2\sigma\lambda n_2 - (1-\delta_1)\sigma\lambda n_1)\left(1 - \frac{\delta_2\sigma\lambda n_2}{C_2}\right)}. \tag{8}$$

Different average times $T_1^{\text{Pri.}} \neq T_1^{\text{Sec.}}$ implies that the load distribution is not optimal because changing it could decrease the overall required time which is given by $T_1^{\text{Pri.}}$ if $T_1^{\text{Pri.}} > T_1^{\text{Sec.}}$ or $T_1^{\text{Sec.}}$ if $T_1^{\text{Sec.}} > T_1^{\text{Pri.}}$. Thus, the throughput is maximized by choosing the packet size distribution δ_1 so that $T_1^{\text{Pri.}} = T_1^{\text{Sec.}}$. A further challenge is the fact that the required time $T_1^{\text{Sec.}}$ also depends on the load distribution δ_2 of operator 2. Such interdependency can be best analyzed in the framework of non-cooperative game theory [12]. By setting $T_1^{\text{Pri.}} = T_1^{\text{Sec.}}$ and rewriting it with respect to δ_j, we obtain the best response functions for $i, j = 1, 2$ with $i \neq j$, by:

$$\begin{aligned}\delta_i(\delta_j) = {} & \frac{1}{2} + \frac{\delta_j n_j}{2n_i} + \frac{C_iC_j + C_j^2}{2\delta_j\sigma^2\lambda^2 n_i n_j} - \frac{C_j}{\sigma\lambda n_i} - \\ & - \Big(\left[C_j(C_i + C_j) - 2C_j\delta_j\sigma\lambda n_j + \delta_j\sigma^2\lambda^2 n_j(n_i + \delta_j n_j)\right]^2 \\ & \qquad - 4C_iC_j\delta_j\sigma^2\lambda^2 n_i n_j\Big)^{1/2} \frac{1}{2\delta_j\sigma^2\lambda^2 n_i n_j}.\end{aligned} \tag{9}$$

The point where the best response functions intersect represents the Nash equilibrium, that is, no operator has an incentive to change his decision single-handedly. In our case there always exists a Nash equilibrium in pure strategies because the best response functions in (9) are continuous and the packet distribution factor is limited between zero and one. Unfortunately, we cannot find a general closed-form solution for the Nash equilibrium. For the special case of $C = C_1 = C_2$ and $n = n_1 = n_2$, however, we find a Nash equilibrium at $\delta_1^* = \delta_2^* = \frac{C}{(2C - \sigma\lambda n)}$. For more general cases we have to rely on numerical methods. Figure 5 shows an example of the best response functions given in (9) for the case of $C = 1$ and $n = 0.5$. The point where the two functions intersect is the Nash equilibrium. Figure 5 also shows other Nash equilibriums for different capacities and different numbers of subscribers whereas for the sake of clarity the best response functions are not shown. Such Nash equilibrium tells the operator how it should split the data packets between primary channel and secondary channel so that the throughput is maximized. As illustrated in Figure 4, the throughput γ_1^{SS} for users of

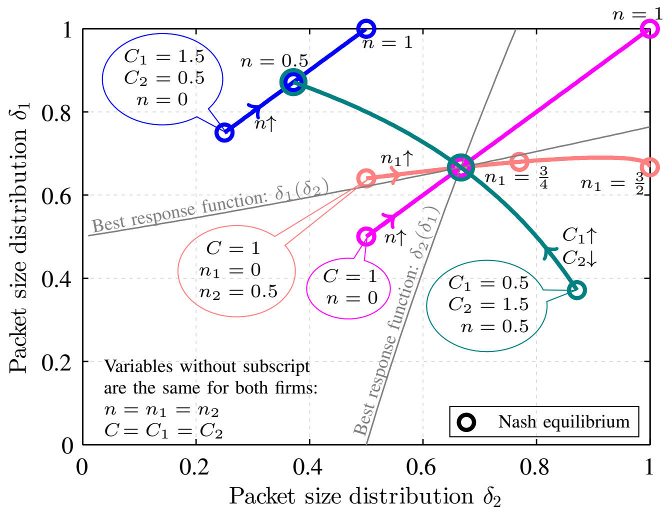

Fig. 5. Best response functions and Nash equilibriums for $\sigma\lambda = 1$ and a few selected values of C and n. The Nash equilibirum delivers the throughput maximizing packet size distribution between primary and secondary channel.

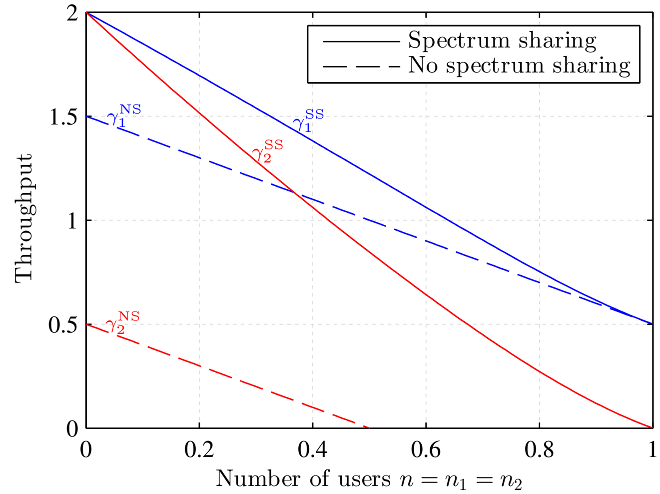

Fig. 6. User throughput in case of no spectrum sharing (NS) and discriminatory spectrum sharing (SS), for the special case of $C_1 = 1.5$, $C_2 = 0.5$ and $\lambda\sigma = 1$. The packet size distribution which maximizes the throughput is shown in Figure 5 (blue curve).

operator 1, respectively γ_2^{SS} for operator 2, can then be found by:

$$\gamma_1^{\text{SS}} = \frac{\sigma}{T_1^{\text{Pri.}}} = \frac{1}{\delta_1^*}\left(C_1 - \delta_1^* \sigma\lambda n_1\right) \tag{10}$$

$$\gamma_2^{\text{SS}} = \frac{\sigma}{T_2^{\text{Pri.}}} = \frac{1}{\delta_2^*}\left(C_2 - \delta_2^* \sigma\lambda n_2\right). \tag{11}$$

Again, only for the special case of $C = C_1 = C_2$ and $n = n_1 = n_2$ we can express (10) and (11) in closed-form, which turns out to be the same as in (3) for one operator who has twice the capacity and twice the traffic load, $\gamma_1^{\text{SS}} = \gamma_2^{\text{SS}} = 2(C - n\lambda\sigma)$. For the special case of $C_1 = 1.5$, $C_2 = 0.5$ and $\sigma\lambda = 1$, Figure 6 shows how the user throughput depends on the number of users n. The corresponding packet size distribution (Nash equilibrium) is given by the blue curve in Figure 5. By employing spectrum sharing, operator 2 gains relatively more throughput than operator 1 while the absolute throughput of operator 1 is still higher, as illustrated in Figure 6.

V. Conclusion

We showed how queuing theory can be used to determine the throughput in cognitive radio systems. Employing such cognitive radio systems allows to increase the spectrum utilization and consequently the aggregated throughput whereas different priority groups guarantee that legacy users are not worse off. Our considerations provide basic elements which can be used in more evolved setups, as shown in Section IV, where we derived the user throughput for the case that two network operators jointly share their spectrum. In particular, if both firms have the same capacity and the same traffic load, the user throughput doubles and coincides with the solution of a single network operator who has twice the capacity and twice the traffic load.

Acknowledgment

The financial support by the Austrian Federal Ministry of Science, Research and Economy, the National Foundation for Research, Technology and Development, and the TU Wien is gratefully acknowledged.

References

[1] M. Rupp, S. Schwarz, and M. Taranetz, *The Vienna LTE-Advanced Simulators: Up and Downlink, Link and System Level Simulation*. Springer, 2016.

[2] ITU Report ITU-R M2225, *Introduction to cognitive radio systems in the land mobile service*, 2011.

[3] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. ten Brink, I. Gaspar, N. Michailow, A. Festag *et al.*, "5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications." *IEEE Communications Magazine*, vol. 52, no. 2, pp. 97–105, 2014.

[4] R. Nissel and M. Rupp, "On pilot-symbol aided channel estimation in FBMC-OQAM," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, March 2016.

[5] ——, "Bit error probability for pilot-symbol aided channel estimation in FBMC-OQAM," in *IEEE International Conference on Communications (ICC)*, Kuala Lumpur, Malaysia, May 2016.

[6] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[7] R. Nissel, S. Caban, and M. Rupp, "Closed-form capacity expression for low complexity BICM with uniform inputs," in *Proceeding of PIMRC*, Hong Kong, Aug 2015.

[8] S. B. Fred, T. Bonald, A. Proutiere, G. Régnié, and J. W. Roberts, "Statistical bandwidth sharing: a study of congestion at flow level," in *ACM SIGCOMM Computer Communication Review*, vol. 31, no. 4. ACM, 2001, pp. 111–122.

[9] R. Nissel and M. Rupp, "Doubly-selective MMSE channel estimation and ICI mitigation for OFDM systems," in *IEEE International Conference on Communications (ICC)*, London, UK, June 2015.

[10] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in LTE cellular networks: Key design issues and a survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 2, pp. 678–700, 2013.

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

[12] A. Mas-Colell, M. D. Whinston, J. R. Green *et al.*, *Microeconomic theory*. Oxford university press New York, 1995, vol. 1.