Optimal Power Control and Scheduling under Hard Deadline Constraints for Continuous Fading Channels

Ahmed Ewaisha, Cihan Tepedelenlioğlu

School of Electrical, Computer, and Energy Engineering, Arizona State University, USA.
Email:{ewaisha, cihan}@asu.edu

Abstract—We consider a joint scheduling-and-power-allocation problem of a downlink cellular system. The system consists of two groups of users: real-time (RT) and non-real-time (NRT) users. Given an average power constraint on the base station, the problem is to find an algorithm that satisfies the RT hard deadline constraint and NRT queue stability constraint. We propose a sum-rate-maximizing algorithm that satisfies these constraints. We also show, through simulations, that the proposed algorithm has an average complexity that is close-to-linear in the number of RT users. The power allocation policy in the proposed algorithm has a closed-form expression for the two groups of users. However, interestingly, the power policy of the RT users differ in structure from that of the NRT users. We also show the superiority of the proposed algorithms over existing approaches using extensive simulations.

I. INTRODUCTION

Quality-of-service-based scheduling has received much attention recently. It is shown in [1], [2] and [3] that quality-of-service-aware scheduling results in a better performance compared to best-effort techniques. For example, real-time audio and video applications require algorithms that take hard deadlines into consideration. This is because if a real-time packet is not transmitted on time, the corresponding user might experience intermittent connectivity to its audio or video.

The problem of scheduling for wireless systems under hard-deadline constraints has been widely studied in the literature (see, e.g., [4] and [5] for a survey). In [6] the authors consider binary erasure channels and present a sufficient and necessary condition to determine if a given problem is feasible. The work is extended in [7] to consider general channel fading models. Unlike the time-framed assumption in these works, the authors of [8] assume that arrivals and deadlines do not have to occur at the edges of a time frame. In [9] the authors study the scheduling problem in the presence of real-time and non-real-time data. Unlike real-time (RT) data, non-real-time (NRT) data do not have strict deadlines but have an implicit stability constraint on the queues.

Power allocation has not been considered for RT users in the literature, to the best of our knowledge, except in [10] that considers on-off fading channels. In this paper, we study a throughput maximization problem in a downlink cellular system serving RT and NRT users simultaneously.

We formulate the problem as a joint scheduling-and-power-allocation problem to maximize the sum throughput of the NRT users subject to an average power constraint on the base station (BS), as well as a QoS constraint for each RT user. This QoS constraint requires a minimum ratio of packets to be transmitted by a hard deadline, for each RT user. Perhaps the closest to our work are references [9] and [11]. The former does not consider power allocation, while the latter assumes that only one user can be scheduled per time slot. The contributions in this paper are as follows:

- We present closed-form expressions for the power allocation policy. It is shown that the power allocation expressions for the RT and NRT users have a different structure.
- We present an optimal algorithm satisfying the average power constraint as well as the QoS constraint. We show through simulations that the complexity, in the number of users, of the proposed algorithm is close-to-linear.

More details on the results of this work is presented in [12]. The rest of this paper is organized as follows. In Section II we present the system model and the underlying assumptions. The problem is formulated in Section III and our optimal algorithm is proposed in Section IV. Simulation results are presented in Section V. Finally, the paper is concluded in Section VI.

II. SYSTEM MODEL

We assume a time slotted downlink system with slot duration T seconds. The system has a single base station (BS) having access to a single frequency channel. There are N users in the system indexed by the set $\mathcal{N} \triangleq \{1, \cdots, N\}$. The set of users is divided into the RT users $\mathcal{N}_{\mathrm{R}} \triangleq \{1, \cdots, N_{\mathrm{R}}\}$, and NRT users $\mathcal{N}_{\mathrm{NR}} \triangleq \{N_{\mathrm{R}} + 1, \cdots, N\}$ with N_{R} and $N_{\mathrm{NR}} \triangleq N - N_{\mathrm{R}}$ denoting the number of RT and NRT users, respectively. Following [6], we model the channel between the BS and the ith user as a fading channel with power gain $\gamma_i(k) \in [0, \gamma_{\max}]$ where $\gamma_{\max} < \infty$ is the maximum channel gain that $\gamma_i(k)$ can take during the kth slot. Channel gains are fixed over the whole slot and change independently in subsequent slots and are independent across users. Moreover, the channel state information for all users are known to the BS at the beginning of each slot in a channel estimation technique that is out of the scope of this paper. The reader is referred to, for example, [13] on signal classification techniques that
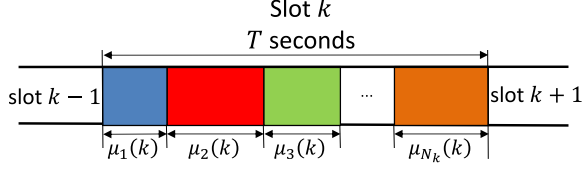
Fig. 1. In the kth slot, the BS chooses $N_k \triangleq |\mathcal{S}_R(k) \cup \mathcal{S}_{NR}(k)|$ users to be scheduled. All time slots have a fixed duration of T seconds.

precede the channel estimation phase if the modulation scheme was unknown.

A. Packet Arrival Model

Let $a_i(k) \in \{0, 1\}$ be the indicator of a packet arrival for user $i \in \mathcal{N}$ at the beginning of the kth slot. $\{a_i(k)\}$ is assumed to be a Bernoulli process with rate λ_i packets per slot and assumed to be independent across all users in the system. Packets arriving at the BS for the RT users are called real-time packets. RT packets have a strict transmission deadline. If an RT packet is not transmitted by this deadline, this packet is dropped out of the system and does not contribute towards the throughput of the user. However, RT user i is satisfied if it receives, on average, more than $q_i\%$ of its total number of packets. We refer to this constraint as the QoS constraint for user i. Here we assume that real-time packets arriving at the beginning of the kth slot have their deadline at the end of this slot.

On the other hand, packets arriving to the BS for the NRT users can be transmitted at any point in time. Thus, packets for NRT user i are stored, at the BS, at user i's (infinite-sized [14]) buffer and served on a first-come-first-serve basis. Since the arrival rate λ_i, for NRT user i, might be higher than what the system can support, we define $r_i(k)$ as an admission controller for user i at slot k. At the beginning of slot k, the BS sets $r_i(k)$ to 1 if the BS decides to admit user i's arrived packet to the buffer, and to 0 otherwise. The time-average number of packets admitted to user i's buffer is $A_i \triangleq \limsup_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}[r_i(k)]$ for all $i \in \mathcal{N}_{NR}$. And the queue associated with NRT user i is given by

$$Q_i(k+1) = (Q_i(k) + Lr_i(k) - \mu_i(k)R_i(k))^+, \quad (1)$$

$i \in \mathcal{N}_{NR}$, where $r_i(k)$ is the admission control decision variable for NRT user i at the beginning of slot k. We note that no admission controller is defined for the RT users since their buffers cannot build up due to the presence of a deadline.

B. Service Model

Following [7] we assume that more than one user can be scheduled in one time slot. However, due to the existence of a single frequency channel in the system, the BS transmits to the scheduled users sequentially as shown in Fig. 1. At the beginning of the kth slot, the BS selects a set of RT users denoted by $\mathcal{S}_R(k) \subseteq \mathcal{N}_R$ and a set of NRT users $\mathcal{S}_{NR}(k) \subseteq \mathcal{N}_{NR}$ to be scheduled during slot k. Moreover, the BS assigns an amount of power $P_i(k)$ for each user $i \in \mathcal{S}_{NR}(k) \cup \mathcal{S}_R(k)$.

This dictates the transmission rate for each user according to the channel capacity given by

$$R_i(k) = \log(1 + P_i(k)\gamma_i(k)). \quad (2)$$

Finally, the BS determines the duration of time, out of the T seconds, that will be allocated for each scheduled user. We define the variable $\mu_i(k)$ to represent the duration of time, in seconds, assigned for user $i \in \mathcal{N}$ during the kth slot (Fig. 1). Hence, $\mu_i(k) \in [0, T]$ for all $i \in \mathcal{N}$. The BS decides the value of $\mu_i(k)$ for each user $i \in \mathcal{N}$ at the beginning of slot k. Since RT users have a strict deadline, then if an RT user is scheduled at slot k, then it should be allocated the channel for a duration of time that allows the transmission of the whole packet. Thus we have

$$\mu_i(k) = \begin{cases} \frac{L}{R_i(k)} & \text{if } i \in \mathcal{S}_R(k) \\ 0 & \text{if } i \in \mathcal{N}_R \setminus \mathcal{S}_R(k) \end{cases}, \quad (3)$$

where L is the number of bits per packet, that is assumed to be fixed for all packets in the system. Equation (3) means that, depending on the transmission power, if RT user i is scheduled at slot k, then it is assigned as much time as required to transmit its L bits. Hence, unlike for the NRT users where $\mu_i(k) \in [0, T]$, $\mu_i(k)$ is further restricted to the set $\{0, L/R_i(k)\}$ for the RT users. For ease of presentation, we denote $\mathbf{Q}(k) \triangleq [Q_1(k), \cdots, Q_{N_{NR}}(k)]^T$. In the next section we present the problem formally.

III. PROBLEM FORMULATION

We are interested in finding the scheduling and power allocation algorithm that maximizes the sum-rate of all NRT users subject to the system constraints. In this paper we restrict our search to slot-based algorithms which, by definition, take the decisions only at the beginning of the time slots.

Now define the average rate of user $i \in \mathcal{N}_{NR}$ to be $\overline{R}_i \triangleq \liminf_{K\to\infty} \sum_{k=1}^{K} \mu_i(k)R_i(k)/(LTK)$ packets per slot. Thus the problem is to find the scheduling, power allocation and packet admission decisions at the beginning of each slot, that solve the following problem

$$\text{maximize} \quad \sum_{i \in \mathcal{N}_{NR}} \overline{R}_i, \quad (4)$$

$$\text{subject to} \quad r_i(k) \leq a_i(k) \forall i \in \mathcal{N}_{NR}, \quad (5)$$

$$\limsup_{k\to\infty} \mathbb{E}[Q_i(k)] < \infty \forall i \in \mathcal{N}_{NR}, \quad (6)$$

$$\overline{R}_i \geq \lambda_i q_i \forall i \in \mathcal{N}_R, \quad (7)$$

$$\limsup_{K\to\infty} \sum_{k\geq 1, i \in \mathcal{N}} \frac{P_i(k)\mu_i(k)}{KT} \leq P_{avg}, \quad (8)$$

$$0 \leq P_i(k) \leq P_{max} \forall i \in \mathcal{N}, \quad (9)$$

$$\sum_{i \in \mathcal{N}} \mu_i(k) = T \forall k \geq 1, \quad (10)$$

where the decision variables are $\boldsymbol{\mu}(k) \triangleq [\mu_i(k)]_{i\in\mathcal{N}}$, $\mathbf{P}(k) \triangleq [P_i(k)]_{i\in\mathcal{N}}$ and $\mathbf{r}(k) \triangleq [r_i(k)]_{i\in\mathcal{N}_{NR}}$, $\forall k \geq 1$. Constraint (5) says that no packets should be admitted to the ith buffer if no packets arrived for user i. Constraint (6) means that the queues of the NRT users have to be stable. Constraint (7) is the RT users' QoS constraint. Constraint (8) is an average power

constraint on the BS transmission power. Finally, constraint (10) guarantees that the sum of durations of transmission of all scheduled users does not exceed the slot duration T. In this paper, we assume that the scheduled NRT user has enough packets, at each slot, to fit the whole slot duration which is a valid assumption in the heavy traffic regime.

IV. PROPOSED ALGORITHM

We use the Lyapunov optimization technique [15] to find an optimal algorithm that solves (4). We do this on three steps: i) We define, in Section (IV-A) a "virtual queue" associated with each average constraint in problem (4). This helps in decoupling the problem across time slots. ii) In Section IV-B, we define a Lyapunov function, its drift and a, per-slot, reward function. iii) Based on the virtual queues and the Lyapunov function, we form and solve an optimization problem, for each slot k, that minimizes the drift-minus-reward expression. The solution of this problem is the proposed power allocation and scheduling algorithm.

A. Problem Decoupling Across Time Slots

We define a virtual queue associated with each RT user as follows

$$Y_i(k+1) = (Y_i(k) + a_i(k)q_i - \mathbb{1}_i(k))^+, \qquad i \in \mathcal{N}_{\mathrm{R}}, \quad (11)$$

where $\mathbb{1}_i(k) \triangleq \mathbb{1}(\mu_i(k))$ with $\mathbb{1}(\cdot) = 1$ if its argument is non-zero and $\mathbb{1}(\cdot) = 0$ otherwise. For notational convenience we denote $\mathbf{Y}(k) \triangleq [Y_1(k), \cdots, Y_{N_{\mathrm{R}}}(k)]^T$. $Y_i(k)$ is a measure of how much constraint (7) is violated for user i. We will later show a sufficient condition on $Y_i(k)$ for constraint (7) to be satisfied. Hence, we say that the virtual queue $Y_i(k)$ is associated with constraint (7). Similarly, we define the virtual queue $X(k)$, associated with constraint (8), as

$$X(k+1) = \left(X(k) + \frac{\sum_{i \in \mathcal{N}} P_i(k)\mu_i(k)}{T} - P_{\mathrm{avg}}\right)^+. \quad (12)$$

To provide a sufficient condition on the virtual queues to satisfy the corresponding constraints, we use the definition of *mean rate stability* of queues [15, Definition 1] to state the following lemma.

Lemma 1. *If, for some $i \in \mathcal{N}_{\mathrm{NR}}$, $\{Y_i(k)\}_{k=0}^{\infty}$ is mean rate stable, then constraint (7) is satisfied for user i.*

Lemma 1 shows that when the virtual queue $Y_i(k)$ is mean rate stable, then constraint (7) is satisfied for user $i \in \mathcal{N}_{\mathrm{NR}}$. Similarly, if $\{X(k)\}_{k=0}^{\infty}$ is mean rate stable, then constraint (8) is satisfied. Thus, our objective would be to devise an algorithm that guarantees the mean rate stability of both $[Y_i(k)]_{i \in \mathcal{N}_{\mathrm{R}}}$ and $X(k)$.

B. Applying the Lyapunov Optimization

The quadratic Lyapunov function is defined as

$$L_{\mathrm{yap}}(U(k)) \triangleq \frac{1}{2}\sum_{i \in \mathcal{N}_{\mathrm{R}}} Y_i^2(k) + \frac{1}{2}\sum_{i \in \mathcal{N}_{\mathrm{NR}}} Q_i^2(k) + \frac{1}{2}X^2(k),$$
$$(13)$$

where $\mathbf{U}(k) \triangleq (\mathbf{Y}(k), \mathbf{Q}(k), X(k))$, and the Lyapunov drift as $\Delta(k) \triangleq \mathbb{E}_{U(k)}[L_{k+1}(\mathbf{U}(k+1)) - L_{\mathrm{yap}}(\mathbf{U}(k))]$ where $\mathbb{E}_{\mathbf{U}(k)}[x] \triangleq \mathbb{E}[x|U(k)]$ is the conditional expectation of the random variable x given $U(k)$. Squaring (1), (11) and (12) taking the conditional expectation then summing over i, the drift becomes bounded by

$$\Delta(k) \leq \frac{C_1}{2} + \Psi(k), \quad (14)$$

where $C_1 \triangleq \sum_{i \in \mathcal{N}_{\mathrm{R}}}\left(q_i^2 + 1\right) + P_{\max}^2 + P_{\mathrm{avg}}^2 + N_{\mathrm{NR}}L^2 + N_{\mathrm{NR}}T^2R_{\max}^2$ and we use $R_{\max} \triangleq \log(1 + P_{\max})$, while

$$\Psi(k) \triangleq \sum_{i \in \mathcal{N}_{\mathrm{R}}} \mathbb{E}_{\mathbf{U}(k)}\left[Y_i(k)\left(\lambda_i q_i - \mathbb{1}_i(k)\right)\right]$$

$$+ X(k)\left(\sum_{i \in \mathcal{N}} \frac{\mathbb{E}_{\mathbf{U}(k)}\left[\mu_i(k)P_i(k)\right]}{T} - P_{\mathrm{avg}}\right)$$

$$+ \sum_{i \in \mathcal{N}_{\mathrm{NR}}} Q_i(k)\left(\mathbb{E}_{\mathbf{U}(k)}\left[Lr_i(k) - \mu_i(k)R_i(k)\right]\right). \quad (15)$$

We define $B_{\max}$ as an arbitrarily chosen positive control parameter that controls the performance of the algorithm. We shall discuss the tradeoff on choosing $B_{\max}$ later on. Since $\mathbb{E}_{\mathbf{U}(k)}[Lr_i(k)]$ represents the average number of bits admitted to NRT user i's buffer at slot k, we refer to $B_{\max}\sum_{i \in \mathcal{N}_{\mathrm{NR}}} \mathbb{E}_{\mathbf{U}(k)}[Lr_i(k)]$ as the "reward term". We subtract this term from both sides of (14), then use (15) and rearrange to bound the drift-minus-reward term as

$$\Delta(k) - B_{\max}\sum_{i \in \mathcal{N}_{\mathrm{NR}}} \mathbb{E}_{\mathbf{U}(k)}[Lr_i(k)] \leq C_1 - X(k)P_{\mathrm{avg}}$$

$$+ \mathbb{E}_{\mathbf{U}(k)}\left[\sum_{i \in \mathcal{N}_{\mathrm{R}}} \Psi_{\mathrm{R}}(i,k)\right] + \mathbb{E}_{\mathbf{U}(k)}\left[\sum_{i \in \mathcal{N}_{\mathrm{NR}}} \Psi_{\mathrm{NR}}(i,k)\mu_i(k)\right]$$

$$+ \mathbb{E}_{\mathbf{U}(k)}\left[\sum_{i \in \mathcal{N}_{\mathrm{NR}}} (Q_i(k) - B_{\max})Lr_i(k)\right] + \sum_{i \in \mathcal{N}_{\mathrm{R}}} Y_i(k)\lambda_i q_i,$$
$$(16)$$

where $\Psi_{\mathrm{R}}(i,k) \triangleq \left(Y_i(k) - \frac{L}{TR_i(k)}X(k)P_i(k)\right)\mathbb{1}_i(k)$ for all $i \in \mathcal{N}_{\mathrm{R}}$ and $\Psi_{\mathrm{NR}}(i,k) \triangleq Q_i(k)R_i(k) - \frac{X(k)P_i(k)}{T}$ for all $i \in \mathcal{N}_{\mathrm{NR}}$ The proposed algorithm schedules the users, allocates their powers and controls the packet admission to minimize the right-hand-side of (16) at each slot. Since the only term in right-hand-side of (16) that is a function in $r_i(k) \ \forall i \in \mathcal{N}_{\mathrm{NR}}$ is the fourth term, we can decouple the admission control problem from the joint scheduling-and-power-allocation problem. Minimizing this term results in the following admission controller: set $r_i(k) = a_i(k)$ if $Q_i(k) < B_{\max}$ and 0 otherwise. Minimizing the remaining terms yields

maximize $\sum_{i \in \mathcal{S}_{\mathrm{R}}(k)} \Psi_{\mathrm{R}}(i,k) + \sum_{i \in \mathcal{N}_{\mathrm{NR}}} \Psi_{\mathrm{NR}}(i,k)\mu_i(k)$
subject to (9) and (10).

$$(17)$$

with decision variables $\mathbf{P}(k)$ and $\boldsymbol{\mu}(k)$. This is a per-slot optimization problem the solution of which is an algorithm that minimizes the upper bound on the drift-minus-reward term defined in (16). Next we show how to solve this problem in

an efficient way.

C. Efficient Solution for the Per-Slot Problem

To solve this problem optimally, we first find the optimal power-allocation-and-scheduling policy for the NRT users through the following lemma.

Lemma 2. *If user $i \in \mathcal{N}_{\mathrm{NR}}$ is scheduled to transmit any of its NRT data during the kth slot, then the optimum power level for this NRT w.r.t. problem* (17) *in the continuous fading case is given by*

$$P_i(k) = \min\left(\left(\frac{Q_i(k)}{X(k)} - \frac{1}{\gamma_i(k)}\right)^+, P_{\max}\right). \quad (18)$$

Moreover, in the heavy traffic regime, the scheduled NRT user, if any, that optimally solves problem (4) *is $i_{\mathrm{NR}}^* = \arg\max_{i \in \mathcal{N}_{\mathrm{NR}}} \Psi_{\mathrm{NR}}^*(i,k)$ with ties broken randomly uniformly, while $\Psi_{\mathrm{NR}}^*(i,k) \triangleq Q_i(k)\log(Q_i(k)) - Q_i(k) + \frac{X(k)}{\gamma_i(k)} - Q_i(k)\log\left(\frac{X(k)}{\gamma_i(k)}\right)$.*

Proof. The proof is omitted for brevity. $\square$

Lemma 2 presents the optimal power and scheduling policy for the NRT users. To solve for the scheduling and power allocation for the RT users, we first solve for $P_i(k)$ assuming a fixed subset $\mathcal{S}_{\mathrm{R}}(k) \subseteq \mathcal{N}_{\mathrm{R}}$, then, the optimum set $\mathcal{S}_{\mathrm{R}}^*(k)$ is the one that maximizes (17). The expression for $P_i(k)$ for the RT users is one of the main contributions of this paper and is presented in the following theorem.

Theorem 1. *In the continuous-fading channel model, given some non-empty set $\mathcal{S}_{\mathrm{R}}(k)$, the power allocation policy*

$$P_i(k) = \min\left(\frac{1}{\gamma_i(k)}\left[\frac{\tilde{\phi}\gamma_i(k) - 1}{W_0\left(\left[\tilde{\phi}\gamma_i(k) - 1\right]e^{-1}\right)} - 1\right], P_{\max}\right), \quad (19)$$

$i \in \mathcal{S}_{\mathrm{R}}(k)$ with $\tilde{\phi} \triangleq (\Psi_{\mathrm{NR}}^*(i_{\mathrm{NR}}^*, k) + \phi)T/X(k)$, is optimal w.r.t. (17) *when ϕ is set to a non-negative value that satisfies* (10) *and $\phi\mu_{i_{\mathrm{NR}}^*}(k) = 0$.*

Proof. See [12] for the complete proof. $\square$

It is clear that the Lambert power policy in (19) has a different structure than the water-filling policy in (18). The reason is because the former is for transmitting packet that have hard deadlines. The following theorem, stated without proof due to lack of space, discusses the monotonicity of the Lambert power policy.

Theorem 2. *Let $\mathcal{S}_{\mathrm{R}}(k)$ be some scheduling RT set at slot k. The power $P_i(k)$ given by* (19) *is monotonically decreasing in $\gamma_i(k)$ $\forall i \in \mathcal{S}_{\mathrm{R}}(k)$.*

In [12], we plot (19) and (18) versus $\gamma_i(k)$ to contrast the fact that, while the water-filling is an increasing function in the channel gain, the Lambert is a decreasing function in the channel gain. This is because the RT user has a single packet of a fixed length to be transmitted. If the channel gain increases, then the power decreases to keep the same transmission rate resulting in the same transmission duration of one slot.

The optimum scheduling algorithm for the RT users is to find, among all subsets of the set $\mathcal{N}_{\mathrm{R}}$, the set that gives the highest objective function of (17).

The following theorem is stated as an effort to achieve an algorithm with a relatively small complexity.

Theorem 3. *At slot k, for any set $\mathcal{S}_{\mathrm{R}}(k)$, if there exists some $i \notin \mathcal{S}_{\mathrm{R}}(k)$ and some $j \in \mathcal{S}_{\mathrm{R}}(k)$ such that $Y_i(k) > Y_j(k)$ and $\gamma_i(k) > \gamma_j(k)$, then $\mathcal{S}_{\mathrm{R}}(k)$ cannot be an optimal RT set, with respect to problem* (17), *for the continuous channel model.*

Proof. See [12] for the complete proof. $\square$

This theorem provides a sufficient condition for non-optimality. In other words, we can make use of this theorem to restrict our search algorithm to the sets that do not satisfy this property. Before presenting the proposed algorithm, we define the set $\mathcal{S}_{\mathrm{subsets}}$ as the set of all possible subsets of the set $\mathcal{N}_{\mathrm{R}}$.

Algorithm 1 Lambert-Strict Algorithm

1: Define the auxiliary functions $\Psi_{\mathrm{X}}(\cdot) : \mathcal{S}_{\mathrm{subsets}} \to \mathbb{R}_+$ and $P_{\mathrm{X}}(\cdot, \cdot) : \mathcal{S}_{\mathrm{subsets}} \times \mathcal{N}_{\mathrm{R}} \to \mathbb{R}_+$.
2: Initialize $P_{\mathrm{X}}(\mathcal{S}, i) = 0$ for all $\mathcal{S} \in \mathcal{S}_{\mathrm{subsets}}$ and all $i \in \mathcal{N}_{\mathrm{R}}$.
3: Find the user i_{NR}^* and its power as given in Lemma 2.
4: **for** $\mathcal{S} \in \mathcal{S}_{\mathrm{subsets}}$ **do**
5:    **if** $\exists$ some $i \notin \mathcal{S}$ and some $j \in \mathcal{S}$ such that $Y_i(k) > Y_j(k)$ and $\gamma_i(k) > \gamma_j(k)$ **then**
6:      Set $\Psi_{\mathrm{X}}(\mathcal{S}) = -\infty$ and go to Step 4 (next set in $\mathcal{S}_{\mathrm{subsets}}$).
7:    **end if**
8:    $\phi \leftarrow \phi_{\max} + \Delta\phi$
9:    **while** $\phi\mu_i(k) \neq 0$ **do**
10:      $\phi \leftarrow \phi - \Delta\phi$. Calculate $P_i(k)$ given by (19) for all $i \in \mathcal{S}$ and set $\mu_{i_{\mathrm{NR}}^*}(k) = T - \sum_{i \in \mathcal{S}} \mu_i(k)$.
11:    **end while**
12:    Set $\Psi_{\mathrm{X}}(\mathcal{S}) = \sum_{i \in \mathcal{S}}(Y_i(k) - X_i(k)\mu_i(k)) + \Psi_{\mathrm{NR}}^*(i_{\mathrm{NR}}^*, k)\mu_{i_{\mathrm{NR}}^*}(k)$ and $P_{\mathrm{X}}(\mathcal{S}, i) = P_i(k) \; \forall i \in \mathcal{S}$.
13: **end for**
14: The scheduling set is $\mathcal{S}_{\mathrm{R}}^*(k) = \arg\max_{\mathcal{S}} \Psi_{\mathrm{X}}(\mathcal{S})$.
15: Set $P_i^*(k) = P_{\mathrm{X}}(\mathcal{S}_{\mathrm{R}}^*(k), i)$ for all $i \in \mathcal{N}_{\mathrm{R}}$, set $\mu_{i_{\mathrm{NR}}^*}(k) = T - \sum_{i \in \mathcal{S}_{\mathrm{R}}^*(k)} \mu_i(k)$ and set $r_i(k) = a_i(k)$ if $Q_i(k) < B_{\max}$ and 0 otherwise $\forall i \in \mathcal{N}_{\mathrm{NR}}$.
16: Update (1), (11) and (12) at the end of the kth slot.

Theorem 4. *For the continuous channel model, if problem 4 is feasible, then for any $B_{\max} > 0$ Algorithm 1 satisfies all constraints in* (4) *and achieves an average sum throughput satisfying*

$$\sum_{i \in \mathcal{N}_{\mathrm{NR}}} \overline{R}_i \geq \sum_{i \in \mathcal{N}_{\mathrm{NR}}} \overline{R}_i^* - \frac{C_1}{LB_{\max}}, \quad (20)$$

where $\overline{R}_i^*$ is the optimal rate for user i w.r.t. (4).

Proof. See [12] for the complete proof. $\square$

Due to the problem being a combinatorial problem with a huge amount of possibilities, we could not reach a closed-form expression for the complexity order of this algorithm.
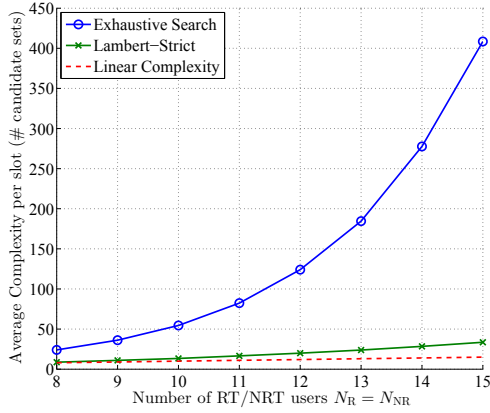
Fig. 2. As the number of NRT users in the system increase the complexity increases exponentially for exhaustive search and nearly linear for the Lambert-Strict algorithm.

Fig. 3. The Lambert-Strict Algorithm yields the same throughput as the exhaustive search algorithm but with a lower average complexity.

However, simulations will show its complexity improvement over the exhaustive search algorithm.

V. SIMULATION RESULTS

We simulate the system given the following parameter values: $\overline{\gamma}_i = 1 \ \forall \ i \in \mathcal{N}$, $L = 1$ bit, $B_{\max} = 100$, $T = 5$, $P_{\text{avg}} = 10$, $q_i = q = 0.9 \ \forall \ i \in \mathcal{N}_{\text{R}}$ and $P_{\max} = 20$. In Fig. 2, we plot the complexity of the Lambert-Strict algorithm as well as the exhaustive search algorithm with exponential complexity versus the number of users N_{R}. The complexity is measured in terms of the average number of iterations, per-slot, which reflects the number of times we have to evaluate the objective function of (17). Since this complexity changes from a slot to the other, we plot the average of this complexity. As the number of users increases, the Lambert-Strict algorithm has an average complexity close to linear. However, there is no sacrifice in the throughput of the NRT users. This is shown in Fig. 3. The reason stems from the optimality of the Lambert-Strict algorithm that does not eliminate any RT users from scheduling unless it is a suboptimal user.

VI. CONCLUSIONS

We discussed the problem of throughput maximization in downlink cellular systems in the presence of RT and NRT users. We formulated the problem as a joint power-allocation-and-scheduling problem. Using the Lyapunov optimization theory, we presented an optimal algorithm that solves the constrained throughput maximization problem. The complexity of the proposed algorithm is shown, through simulations, to have a close-to-linear complexity. Moreover, the power allocations are presented in closed-form expressions for the RT as well as the NRT users. We showed that the NRT power allocation is water-filling-like which is monotonically increasing in the channel gain. On the other hand, the RT power allocation has a totally different structure that we call the "Lambert Power Allocation". It is found that the latter is a decreasing function in the channel gain.
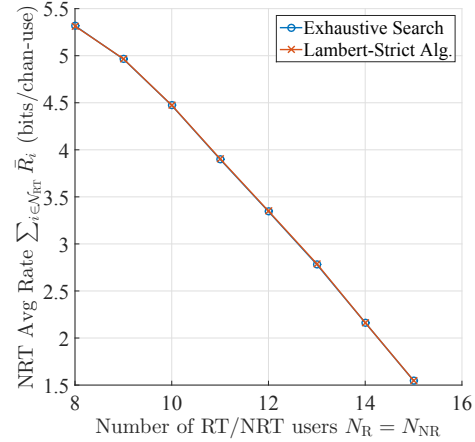
REFERENCES

[1] W. K. Lai and C.-L. Tang, "QoS-aware downlink packet scheduling for LTE networks," *Computer Networks*, vol. 57, no. 7, pp. 1689 – 1698, 2013.

[2] A.E. Ewaisha and C. Tepedelenlioğlu, "Joint scheduling and power control for delay guarantees in heterogeneous cognitive radios," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6298–6309, Sept 2016.

[3] Giuseppe Piro, Luigi Alfredo Grieco, Gennaro Boggia, Rossella Fortuna, and Pietro Camarda, "Two-level downlink scheduling for real-time multimedia services in LTE networks," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 1052–1065, 2011.

[4] I-H. Hou and P.R. Kumar, "A survey of recent results on real-time wireless networking," *Proceedings of the real-time wireless for industrial applications*, 2011.

[5] S. Radhakrishnan, S. Neduncheliyan, and K.K. Thyagharajan, "A review of downlink packet scheduling algorithms for real time traffic in LTE-advanced networks," *Indian Journal of Science and Technology*, vol. 9, no. 4, 2016.

[6] I-H. Hou, V. Borkar, and P.R. Kumar, "A theory of QoS for wireless," in *INFOCOM 2009, IEEE*, April 2009, pp. 486–494.

[7] I-H. Hou and P.R. Kumar, "Scheduling heterogeneous real-time traffic over fading wireless channels," in *IEEE Proceedings of INFOCOM 2010*, March 2010, pp. 1–9.

[8] X. Kang, W. Wang, J.J. Jaramillo, and L. Ying, "On the performance of largest-deficit-first for scheduling real-time traffic in wireless networks," in *Proceedings of the fourteenth ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2013, pp. 99–108.

[9] J.J. Jaramillo and R. Srikant, "Optimal scheduling for fair resource allocation in ad hoc networks with elastic and inelastic traffic," in *Proceedings of INFOCOM 2010*. IEEE, 2010, pp. 1–9.

[10] A.E. Ewaisha and C. Tepedelenlioglu, "Power control and scheduling under hard deadline constraints for On-Off fading channels," in *2017 IEEE Wireless Communications and Networking Conference (WCNC 2017)*, San Francisco, USA, Mar. 2017.

[11] A.E. Ewaisha and C. Tepedelenlioğlu, "Throughput optimization in multichannel cognitive radios with hard-deadline constraints," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2355–2368, April 2016.

[12] A.E. Ewaisha and C. Tepedelenlioğlu, "Optimal power control and scheduling for real-time and non-real-time data," *IEEE Transactions on Vehicular Technology (submitted)*.

[13] Mohammad Bari, Awais Khawar, Miloš Doroslovački, and T Charles Clancy, "Recognizing FM, BPSK and 16-QAM using supervised and unsupervised learning techniques," in *Signals, Systems and Computers, 2015 49th Asilomar Conference on*. IEEE, 2015, pp. 160–163.

[14] D.i Bertsekas and R. Gallager, *Data Networks (2Nd Ed.)*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1992.

[15] C.-P. Li and M.J. Neely, "Delay and Power-Optimal Control in Multi-Class Queueing Systems," *ArXiv e-prints*, Jan. 2011.