

On optimization formulations for radio resource allocation subject to common transmission rate

Yi Zhao^{*}, Di Yuan

Department of Information Technology, Uppsala University, 751 05, Uppsala, Sweden

ARTICLE INFO

Keywords:

Integer programming
Modeling
Radio resource allocation

ABSTRACT

We study a radio resource allocation problem in mobile communication systems. As the distinct characteristic of this problem, a common data transmission rate is used on all channels allocated to a user. Because the channels differ in their quality, for each user the achievable rate varies by channel. Thus allocating more channels does not necessarily increase the total rate, as the common rate is constrained to be the lowest one supported by the allocated channels. Radio resource allocation subject to the common-rate constraint is of practical relevance, though little attention has been paid to modeling and solving the problem. We take a mathematical optimization perspective with focus on modeling. We first provide a complexity analysis. Next, several integer linear programming (ILP) formulations for the problem, including compact as well as non-compact models, are derived. The bulk of our analysis consists in a rigorous comparative study of their linear programming (LP) relaxations, to reveal the relationship between the formulations in terms of bounding. Computational experiments are presented to illustrate the numerical performance in bounding and LP-assisted problem solving. Our theoretical analysis and numerical results together serve the aim of setting a ground for the next step of developing model-based and tailored optimization methods.

1. Introduction

Mobile communication systems have been evolving rapidly in the past decades. From an optimization standpoint, a key research topic for mobile networks is radio resource management for utilizing the radio resource efficiently, particularly because the amount of data traffic is constantly growing in the current 4th generation (4G) systems, and the trend will remain or even accelerate in the 5th generation (5G) networks. Along with this growth, 5G networks target a plethora of new services (e.g., vehicular communications), making resource management more challenging than before, as pointed out in Calabrese et al. (2018).

4G/5G networks deploy orthogonal frequency division multiple access (OFDMA). In essence, data transmission takes place in two dimensions, namely time and frequency that are divided into time slots and channels, respectively.¹ Resource management for channel allocation in a time slot, as well as that across multiple time slots, is commonly called scheduling. In any time slot, a user may be scheduled to receive transmission from its base station (BS) on multiple channels, and a channel may be scheduled for at most one user of this BS. The latter is what the term “orthogonal” refers to in OFDMA. Among the BSs,

the channels are reused, i.e., there is generally no dedicated (sub)set of channels for a BS. As a result, a channel is exposed to interference, if the channel is also scheduled in some nearby BSs in the same time slot. Note that even if interference is not present, the channels are still frequency-selective (Tse and Viswanath, 2005), meaning that they vary in quality and hence the achievable data rate from a user’s perspective.

In this paper, we consider the scheduling problem for a BS and one time slot, in the direction of downlink (i.e., data transmission by the BS to its users). For a user, a lower or upper bound on the total rate may be present. The former is for time-critical services (e.g., vehicular applications) where a minimum amount of data must be delivered within the time slot. The presence of upper bound originates from that the amount of data currently buffered at the BS (and hence ready for transmission) is typically limited. As an additional constraint, by mobile communication standards for 4G and 5G (3GPP, 2018a,b), the system has to select a common data rate on all channels allocated to a user. This has a couple of implications. First, allocating more channels to a user does not necessarily increase the total rate, since if the rate supported by the additional channels is low, the common rate will have to be brought down. Second, in effect, the total rate equals the common rate multiplied by the number of allocated channels (that

^{*} Corresponding author.

E-mail addresses: yi.zhao@it.uu.se (Y. Zhao), di.yuan@it.uu.se (D. Yuan).

¹ In communication technology, these entities are more formally known as transmission time interval (TTI) and subcarrier. We use a plain style in terminology throughout this paper for better readability.

must all support the rate). In the literature, little attention has been paid to channel allocation subject to the common-rate constraint. To our knowledge, the current work is the first study of integer linear programming (ILP) formulations for the problem.

It should be remarked that, from a conventional optimization viewpoint, solving our channel allocation and rate selection problem to optimality is “fast”. For representative problem sizes, it is a matter of seconds for a standard integer programming solver to do the job. However, in a practical system, the allocation decision need to be taken at a time scale of milliseconds. In order to approach highly efficient solution methods based on optimization models, deriving and studying mathematical formulations for this problem is crucial. To this end, we propose compact as well as non-compact ILP formulations. We then provide a rigorous comparative study of their linear programming (LP) relaxations, revealing the relationship between these formulations in LP bound. To numerically examine the performance of the models, we conduct experiments of LP-assisted problem solving to obtain integer solutions. Together, the theoretical analysis and the numerical experiments of our work contribute to the forthcoming research of developing model-based and highly efficient methods targeting practical use.

2. Literature review

There is a vast amount of optimization problems in mobile networks, in particular for 4G/5G systems that have a broad range of applications and networking scenarios. Here, we narrow down the review to network planning and resource allocation, with scheduling being part of the latter.

Network planning is a long-term optimization task. Examples of decisions include BS location and antenna configuration, and typical performance metrics are coverage, capacity, throughput, and energy consumption. Network planning has been studied by both the engineering and the operations research (OR) communities. Examples of the former category for 4G/5G are Bioardi et al. (2013), Al-Kanj et al. (2015), and Ali et al. (2020). In Bioardi et al. (2013), the authors propose an ILP model for BS location and configuration, addressing the trade-off between capital cost and energy cost. The work in Al-Kanj et al. (2015) minimizes the number of BSs to be deployed as well as their activation over time for energy efficiency, subject to coverage and interference considerations. The problem is solved using integer programming and heuristics. In Ali et al. (2020), meta-heuristics have been proposed for optimal location of both BS and relay stations, where signal quality is part of the objective function in addition to the deployment cost. In the OR literature, the research orientation has been more on efficient modeling. Mathematical formulations for coverage planning using minimum power, while explicitly accounting for overlap to assist hand-over, have been examined in Chen and Yuan (2010). The work in D’Andreagiovanni et al. (2012) optimizes BS location and transmission power, to maximize the system capacity subject to interference constraints. Modeling is also the focus in Ecker et al. (2019), where the authors propose ILP formulations based on enumeration of interference scenarios for a BS location problem. A recent trend for network planning is the use of machine learning (Dreifuerst et al., 2021).

Resource allocation and scheduling for 4G/5G have been studied mainly by the engineering research community. The optimization may involve one BS or multiple BSs. The performance objective usually targets spectrum efficiency, throughput (sometimes in extended forms that are often referred to as system utility), and power or energy.

For multi-BS resource allocation, interference is one of the most crucial issues to be addressed. In Ng et al. (2012), interference is accounted for by considering how the channels are reused among multiple BSs. The study proposes an iterative optimization algorithm using fractional programming for energy efficiency. The work in Lopez-Perez et al. (2014) proposes a scheme for the purpose of managing and coordinating the amount of interference. Then, for each BS, the

scheduling problem with the objective of power minimization, is formulated as an integer program and solved using smart search. To address interference, Kuang et al. (2016) consider subsets of BSs, each of which is a candidate group of reusing a channel, resembling the column-oriented formulation for graph coloring (Mehrotra and Trick, 1995). This leads to an ILP formulation that is also able to address user association (i.e., the selection of the serving BS of each user), and the Frank–Wolfe method is used to solve the LP relaxation. The work in Wang et al. (2017) considers power allocation in addition to channel allocation and user association, and solves the optimization problem using a method that alternately solves the subproblems obtained by fixing subsets of variables.

The relevance of resource allocation within one BS originates from that it is generally difficult to do multi-BS resource optimization in real time. A more practical approach is some form of decomposition, separating the optimization tasks for inter-BS and single-BS resource allocation. This is indeed the general idea in, for example, Lopez-Perez et al. (2014) and Kuang et al. (2016). For single BS channel allocation, Lin et al. (2009) propose a tailored branch-and-bound algorithm for power minimization. The work in Yuan et al. (2012) provides problem complexity analysis for power minimization along with identifying cases admitting polynomial-time algorithms. Problem complexity analysis for system utility maximization subject to a power budget is reported in Liu (2014) and Liu and Dai (2014), and that for max–min capacity is given in Fallgren (2010). A heuristic algorithm for power minimization of channel allocation, based on repeatedly solving network flow problems, is proposed in Joung et al. (2012). Cohen and Katzie (2010) consider a related though different problem, where resource is allocated in the two-dimensional space of channel and time, with a structure similar to a packing problem. For utility maximization, a fast heuristic is proposed in Ng and Sung (2008), and the authors show the solution is Pareto optimal (i.e., the utility cannot be improved for all users) within some neighborhood of the solution. In Letchford et al. (2017), the authors address rate maximization with channel and power allocation. That the power is assumed to be continuous and the rate function is a logarithmic one leads to a nonlinear integer programming model. The authors propose outer approximation, cuts, and other mechanisms for solving the problem to optimality.

The studies discussed above all assume that the rates on the allocated channels of a user are decoupled. To enhance the practical relevance, extending the optimization problem with the common-rate constraint in channel allocation is of significance. Very recently, this constraint is taken into account in Huang et al. (2021) and Chen et al. (2021). The scope of Huang et al. (2021) and Chen et al. (2021) is however very different from ours, as these studies focus on the use of massive graphic processing units (GPUs) to implement parallel heuristic search for problem solving, whereas our interest lies in optimization modeling.

3. Problem definition

Our optimization problem is defined for a BS with user set $\mathcal{I} = \{1, \dots, I\}$ and channel set $\mathcal{L} = \{1, \dots, L\}$. By mobile system standards (3GPP, 2018b), there is a predefined set of candidate data rates, of which the index set is denoted by $\mathcal{R} = \{1, \dots, R\}$. We use v_r to denote the data rate of $r \in \mathcal{R}$, and assume $v_1 < v_2 < \dots < v_R$. The channel condition differs by channel and user. For user i and channel ℓ , the index of the maximally achievable rate is denoted by $r_{i\ell}$. A user may have a lower or upper bound on the total rate.² For user i , these are denoted by \underline{d}_i and \bar{d}_i , respectively. Moreover, the importance of user i is represented by a positive weight w_i .

² Strictly speaking, the data demand is in bits rather than in rate that is defined in bits per time unit. However, since the problem concerns one time slot, the demand can be equivalently treated in rate.

By the common-rate constraint, one rate in \mathcal{R} is to be selected for data transmission on all channels allocated to a user. Thus, if rate r is selected for a user, a channel ℓ is useful only if $r_{i\ell} \geq r$. In effect, if n channels are allocated to user i with rate r , the total rate for user i is nv_r , and this has to be within the interval $[\underline{d}_i, \bar{d}_i]$. Note that, due to the common-rate constraint, if channel ℓ is allocated to user i , the realized rate on the channel may be lower than $r_{i\ell}$. In the rest of the paper, we say channel ℓ supports rate r for user i , if $r_{i\ell} \geq r$.

Our optimization problem amounts to maximizing the weighted sum of rates of all users, a.k.a. total utility, subject to having at most one user per channel, and the above common-rate and rate interval constraints. We refer to the problem as channel allocation with common rate (CACR).

Remark 1. Throughout the paper, without loss of generality, we assume that every channel supports the lowest rate for at least one user. Moreover, for each user, there exist some rate and channels supporting the rate, such that the total rate is within the given interval. \square

We present an ILP model to mathematically formalize CACR. For user $i \in \mathcal{I}$, a binary variable x_{ir} is defined to represent if rate r is selected or not. Another set of binary variables encode channel allocation as well as the rate on each allocated channel. An individual variable of this set, $y_{ir\ell}$, is one if channel ℓ is allocated to user i with rate r .

In the model presented below, parameters $\underline{n}_{ir} = \lfloor \underline{d}_i/v_r \rfloor$ and $\bar{n}_{ir} = \lfloor \bar{d}_i/v_r \rfloor$. Thus, the number of channels to be allocated to user i is confined by interval $[\underline{n}_{ir}, \bar{n}_{ir}]$, if rate r is selected. We define utility coefficient $u_{ir} = w_i v_r$; this is the rate value achieved by allocating one channel of rate r to user i , scaled by the user's weight. In addition, notation \mathcal{L}_{ir} is used to represent the subset of channels supporting rate r for user i .

$$\text{(SEP)} \quad \max \sum_{i \in \mathcal{I}} \sum_{r \in \mathcal{R}} \sum_{\ell \in \mathcal{L}_{ir}} u_{ir} y_{ir\ell} \quad (1a)$$

$$\text{s.t.} \quad \sum_{r \in \mathcal{R}} x_{ir} = 1, \quad i \in \mathcal{I} \quad (1b)$$

$$y_{ir\ell} \leq x_{ir}, \quad \ell \in \mathcal{L}_{ir}, i \in \mathcal{I}, r \in \mathcal{R} \quad (1c)$$

$$\sum_{i \in \mathcal{I}} \sum_{r \in \mathcal{R}: \ell \in \mathcal{L}_{ir}} y_{ir\ell} \leq 1, \quad \ell \in \mathcal{L} \quad (1d)$$

$$\underline{n}_{ir} x_{ir} \leq \sum_{\ell \in \mathcal{L}_{ir}} y_{ir\ell} \leq \bar{n}_{ir} x_{ir}, \quad i \in \mathcal{I}, r \in \mathcal{R} \quad (1e)$$

$$x_{ir} \in \{0, 1\}, \quad i \in \mathcal{I}, r \in \mathcal{R} \quad (1f)$$

$$y_{ir\ell} \in \{0, 1\}, \quad \ell \in \mathcal{L}_{ir}, i \in \mathcal{I}, r \in \mathcal{R} \quad (1g)$$

We refer to the above model as SEP since rate selection and the utility induced by channel allocation are encoded by two separate sets of variables. Constraint (1b) states the condition of selecting one rate for each user. By (1c), a channel may be allocated to a user with a data rate, only if this rate is selected for the user and the channel does support the rate. The next constraint ensures at most one user per channel. Finally, by (1e), for each user and rate, the number of channels allocated must comply to the demand interval, if the rate is selected. We remark that the inclusion of (1c) is for strengthening the LP relaxation; it is redundant for any integer solution due to the presence of (1e).

4. Problem complexity

Proving the NP-hardness of CACR is not the core of this paper. Nevertheless, it is of significance in understanding problem structure, in particular as this complexity result is novel with respect to the available literature. Moreover, we provide the structural insight that, if the rates are selected, channel allocation reduces to a network flow problem. This has implications to mathematical modeling and deriving integer solutions via LP relaxations.

Proposition 1. CACR is NP-hard.

Proof. We use a reduction from the MAX-3SAT problem where each clause contains exactly three literals and each literal appears exactly twice. This version of MAX-3SAT is NP-hard and in fact hard to approximate (Berman et al., 2003). Let N and K denote respectively the numbers of Boolean variables and clauses in an instance of this type of MAX-3SAT problem.

We define a CACR instance as follows. There are $2N + 3K$ users. One pair of users, called literal users, is defined for each variable $i, i = 1, \dots, N$, and three users, called clause users, are defined for each clause. The rate set is composed by $\{\frac{1}{6}, \frac{1}{3}, \frac{2}{3}, \frac{4}{3}\}$. For each pair of literal users, we introduce five channels. We call one of these channels the high channel. The other four channels, called literal channels, are formed by two pairs, where each pair corresponds to the two occurrences of a literal in the MAX-3SAT instance. Both literal users achieve rate $\frac{4}{3}$ on the high channel. For a literal user and any of the user's two literal channels, the rate is $\frac{1}{3}$. Next, for each clause, we define two additional channels, called clause channels. For the corresponding three clause users, they are symmetric with respect to the two clause channels. Specifically, they all achieve rate $\frac{4}{3}$ on one clause channel, and rate $\frac{2}{3}$ on the other clause channel. Moreover, each of the three clause users has rate $\frac{1}{6}$ on exactly one channel of the three literals appearing in the clause of the MAX-3SAT instance, with a one-to-one mapping. Any user and channel combination that has not been mentioned has rate zero. The weight of any literal user equals $8K$, whereas the weight of any clause user is one. There is no lower or upper bound in rate allocation.

Consider the three clause users of any clause. Let us examine the total rate that may be achieved by them.

1. If all three literal channels of the clause are used by literal users, then obviously the best possible allocation is to let (any) two of the three clause users use the two clause channels respectively with a total rate of $\frac{4}{2} + \frac{2}{3} = 2$. (The third user will have rate zero.) This is because giving both clause channels to the same clause user gives a lower total rate of $\frac{4}{3}$, with $\frac{2}{3}$ on each clause channel.
2. If one of the three literal channels is available, then the optimal allocation is to give this literal channel to the corresponding clause user, and let the other two users use the two clause users as above, with a total rate of $2 + \frac{1}{6} = \frac{13}{6}$. Note that it is sub-optimal to let any clause user use more than one channel (due to the decrease of per-channel rate).
3. Suppose two or three literal channels are available. From the above reasoning, it is not difficult to see that the optimal allocation is to allocate one of the available literal channels to its corresponding clause user, and let the other two use the two clause channels respectively. Hence only one of the literal channels will be used. The total utility is again $\frac{13}{6}$.

To summarize, for each clause, the three clause users will together achieve a total rate of $\frac{13}{6}$ if at least one of the three literal channels is available, otherwise the total rate is 2. Thus the total utility of all clause users together will never exceed $\frac{13}{6}K$. Note that the weight of a literal user is $8K$ and the two possible (positive) rates on a channel are $\frac{4}{3}$ and $\frac{1}{3}$. Hence, priority will be given to the literal users in channel allocation.

Let us consider any pair of literal users and the five channels defined for the pair. Suppose one user is allocated with the high channel, then it is not optimal to use the other two literal channels as well, because then the rate on each channel becomes $\frac{1}{3}$, with a total of 1.0, whereas the high channel by itself provides a rate of $\frac{4}{3}$. Therefore, at optimum, one literal user of the pair is allocated the high channel only, leaving its two literal channels unoccupied, and the other is allocated its two literal channels, or vice versa. This corresponds to the value assignment

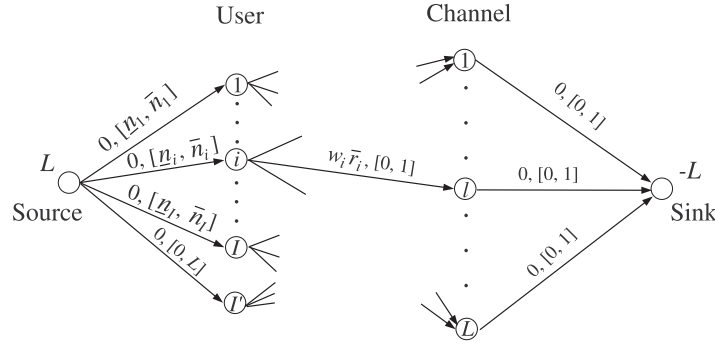


Fig. 1. An illustration of the network flow problem for given rate selection.

of the Boolean variables in MAX-3SAT. Moreover, at optimum the total utility of all literal users is a constant of $16NK$. For the clause users, by the above analysis, the total utility equals $2K + \frac{m}{6}$ where m is the number of clauses for which at least one clause channel is available for use by clause users. Thus solving the CACR instance also solves the MAX-3SAT instance, hence the proof. \square

Remark 2. In the above proof of NP-hardness, there is no lower or upper bound on the total rate to be allocated to a user in the CACR instance. Thus a source of hardness is the common-rate requirement in rate selection. Having lower or upper bound is another source of difficulty. Suppose we relax the common-rate requirement, but there is lower or upper bound on user rate. The problem is also NP-hard via a simple reduction from the partitioning problem. For a partitioning problem instance of k integers, consider k channels and two identical users of weight one, where each user has the integer numbers of the partitioning instance as the rates achievable on the channels. Let the lower (or the upper) bound of the total rate of each user be half of the sum of the integers. The question is then if the CACR instance is feasible (in case of lower bound) or if the optimal total rate achieves the sum, and the answer solves the partitioning instance. \square

A solution of CACR is naturally viewed in a bipartite graph where the nodes are defined for the users and channels. This structure will be utilized later for the analysis of mathematical formulations. Here, in the context of complexity, we prove that CACR becomes a network flow problem (and hence can be solved efficiently) if the rates of users are given.

Proposition 2. For given rate selection, CACR reduces to a network flow problem in an acyclic graph with polynomial-time complexity.

Proof. Let \bar{r}_i denote the selected rate of user $i \in I$. Then the subset of eligible channels for i is $\{\ell \in \mathcal{L} : r_{i\ell} \geq \bar{r}_i\}$. Moreover, let $n_i = \lfloor d_i / v_{\bar{r}_i} \rfloor$ and $\bar{n}_i = \lfloor \bar{d}_i / v_{\bar{r}_i} \rfloor$. The number of channels to be allocated to user i is then confined by interval $[n_i, \bar{n}_i]$. We construct a graph, shown in Fig. 1. The nodes consist of users and channels, a source, a sink, and an auxiliary node denoted by I' . On an arc, the first value is the utility, followed by flow lower and upper bounds. The total supply equals the number of channels L . Note that an arc from user node i and channel node ℓ is present if and only if $r_{i\ell} \geq \bar{r}_i$. Also, there is an arc from I' to every channel node with zero utility and flow bounds $[0, 1]$.

By the construction, for each user i , the number of flow units, which represent the number of channels allocated, is constrained to be within the interval. Having one flow unit from user node i to channel node ℓ yields the utility of allocating ℓ to i . In addition, the upper bound of one on each arc ending at the sink ensures that a channel is used by at most one user. Finally, channels that are not used by any user will be taken by the flow through I' . One can easily see that there is a one-to-one mapping between feasible channel allocations and integer flow solutions with equal objective function value. Moreover, the graph

is acyclic by construction. Thus solving the maximum-weighted flow problem (or, minimum-cost flow with negative costs) in this acyclic graph leads to the optimum of CACR for given rate selection. \square

We further remark that if for each user the channels are identical in the rates they support, the problem becomes tractable. This is because the rate selection becomes simple — each user will use the highest rate supported by all channels, and the result follows then from Proposition 2.

5. ILP formulations

5.1. Alternative compact models

SEP in Section 3 is a compact model for CACR. In the following, we present alternative compact models. It is observed that the total rate of a user and hence its utility is fully determined by rate and the number of channels to use. Using the observation, we can formulate the problem without the need of modeling the utility on individual channels (as done in SEP). To this end, we define set D_i , containing all combinations of rate and number of channels satisfying the rate demand interval of user i . That is, $D_i = \{(r, n), r \in \mathcal{R}, n \in \{1, \dots, L\} : \underline{d}_i \leq nv_r \leq \bar{d}_i\}$. We refer to D_i as the domain of user i . For $(r, n) \in D_i$, we use $u_{irn} = w_i nv_r$ to denote the total utility.

We use binary variable x_{irn} to represent if $(r, n) \in D_i$ is chosen for i . Moreover, binary variable $y_{i\ell}$ is used to indicate if a specific channel ℓ is allocated to i . The resulting model is as follows. We call the model COMB, to refer to that, for each user, one set of variables combines the decisions of selecting a rate and the number of channels to be allocated.

$$(COMB) \quad \max \sum_{i \in I} \sum_{(r,n) \in D_i} u_{irn} x_{irn} \quad (2a)$$

$$\text{s.t.} \quad \sum_{(r,n) \in D_i} x_{irn} = 1, \quad i \in I \quad (2b)$$

$$\sum_{i \in I: \ell \in \mathcal{L}_{i1}} y_{i\ell} \leq 1, \quad \ell \in \mathcal{L} \quad (2c)$$

$$\sum_{(h,n) \in D_i: h \geq r} n x_{ihn} \leq \sum_{\ell \in \mathcal{L}_{ir}} y_{i\ell}, \quad i \in I, r \in \mathcal{R} \quad (2d)$$

$$x_{irn} \in \{0, 1\}, \quad (r, n) \in D_i, i \in I \quad (2e)$$

$$y_{i\ell} \in \{0, 1\}, \quad \ell \in \mathcal{L}_{i1}, i \in I \quad (2f)$$

By (2b), exactly one (eligible) combination of rate and number of channels must be selected for each user. The next constraint states that a channel may be allocated to at most one user among the relevant ones (i.e., users for which the channel can provide the minimum candidate rate). Constraint (2d) sets the relation of the two sets of variables. Note that, for user i and each $(r, n) \in D_i$, $n x_{irn} \leq \sum_{\ell \in \mathcal{L}_{ir}} y_{i\ell}$ will ensure that if x_{irn} equals one, then n channels will be allocated. As one x -variable per user will equal one by (2b), we have $\sum_{(r,n) \in D_i} n x_{irn} \leq \sum_{\ell \in \mathcal{L}_{ir}} y_{i\ell}$. However, it is not difficult to see that we can strengthen it, by including also rates that are higher than r in the left-hand side, leading to (2d).

Remark 3. Variable notation x and y are reused in COMB. We have chosen this reuse to avoid introducing additional notation, and for the reason that the variables, when used, are accompanied either with subscripts or a model name, and hence there shall be no risk of ambiguity. \square

The next compact model, called COMB', is a variation of COMB. COMB' uses the same decision variables as COMB. However, the relation between them is formulated differently.

$$(COMB') \max \sum_{i \in I} \sum_{(r,n) \in D_i} u_{irn} x_{irn} \quad (3a)$$

s.t. (2b), (2c)

$$\sum_{(r,n) \in D_i} n x_{irn} \leq \sum_{\ell \in \mathcal{L}_{i1}} y_{i\ell}, \quad i \in I \quad (3b)$$

$$y_{i\ell} \leq \sum_{(r,n) \in D_i: \ell \in \mathcal{L}_{ir}} x_{irn}, \quad \ell \in \mathcal{L}_{i1}, i \in I \quad (3c)$$

(2e), (2f)

Instead of (2d), the above model uses two sets of constraints linking together x and y . Constraint (3b) is, to some extent, similar to (2d), as it states that sufficiently many channels are allocated with respect to the number indicated by the x -variables. However, unlike (2d), for one user there is only one constraint of (3b), where the left-hand side considers all possible rates. For this reason, it cannot be strengthened as done in (2d). Moreover, since (3b) does not consider if a channel allocated does match the rate selected, we need (3c). By this additional constraint, channel ℓ may be allocated only if it supports the rate selection.

Remark 4. It may be more intuitive to state constraints (2d) and (3b) as equality. It is easily realized that the use of inequality does not alter the integer optimal value or that of the LP relaxation of COMB and COMB'. From a solution point of view, inequality permits a solution in which more channels than that indicated by the x -variables are allocated (i.e., the inequality is strict), though reducing this solution to one with equality is trivial. We choose inequality as it simplifies some of the proofs later on. \square

5.2. Non-compact models

We now turn our attention to non-compact formulations of CACR. We first derive a model that does not explicitly formulate channel allocation. Consider the x -variables defined in COMB. If $x_{irn} = 1$, obviously there must exist at least n channels, all supporting rate r for user i . Taking one step further, consider two variables of two users, say $x_{i_1 n_1 r_1}$ and $x_{i_2 n_2 r_2}$. A necessary condition for having both variables being one is the existence of at least $n_1 + n_2$ channels, such that each of them supports either r_1 for i_1 , or r_2 for i_2 , or both. In other words, each of these channels is eligible for being allocated to at least one of the two users.

We can generalize the above to an arbitrary subset of users, with some specific tuple (r, n) for each of them. Suppose that the total number of relevant channels for the user subset is \tilde{n} . The corresponding condition states that there must be a minimum of \tilde{n} channels, each being eligible for at least one of the users. Leveraging the notion gives us a model with an exponential number of rows. To present the model, let \mathcal{M} denote the set containing all non-empty subsets of users. Moreover, for any $m \in \mathcal{M}$, denote by \mathcal{V}_m the set of vectors, each of which contains one combination of rates of the users in m . We use r_i^v to denote the rate of user i in vector $v \in \mathcal{V}_m$. For m and v , the set of channels supporting at least one user and its rate is denoted by \mathcal{L}_{mv} .

$$(CUT) \max \sum_{i \in I} \sum_{(r,n) \in D_i} u_{irn} x_{irn} \quad (4a)$$

s.t. (2b)

$$\sum_{i \in m} \sum_{(h,n) \in D_i: h \geq r_i^v} n x_{ihn} \leq |\mathcal{L}_{mv}|, \quad v \in \mathcal{V}_m, m \in \mathcal{M} \quad (4b)$$

(2e)

In (4b), the second sum in the left-hand side follows the same rationale as that in (2d). One can easily verify this is valid and makes the LP tighter. We refer to the model in (4) as CUT, because it in fact resembles cut-based models for network optimization problems (cf. Fig. 1). Because (4b) ensures that there is a sufficient number of channels for any subset of users and their rates, it is reasonable to infer that CUT is a correct model for CACR, even though there is no explicit channel allocation. Mathematically, however, this result is not immediate. Therefore we formalize the validity of CUT below with a proof.

Proposition 3. CUT is a valid model for CACR.

Proof. The constraints in CUT are obviously all valid. Thus showing the validity of CUT amounts to proving that they together are also sufficient. That is, for any feasible x -solution of CUT, there exists a channel allocation, such that each user indeed is given the number of channels with the rate indicated by the x -solution. Consider any solution of CUT, and suppose in the solution $x_{i r_i n_i} = 1, i \in I$.

With respect to the x -solution, consider a bipartite graph, and denote its two node sets by \mathcal{N}'_I and \mathcal{N}'_{II} , respectively. For each user i, n_i nodes are introduced to be in \mathcal{N}'_I . Hence $|\mathcal{N}'_I| = \sum_{i \in I} n_i$. Node set \mathcal{N}'_{II} contains one node for each channel. For a node in \mathcal{N}'_I for user i , there is an edge to channel node $\ell \in \mathcal{N}'_{II}$, if and only if $\ell \in \mathcal{L}_{ir_i}$ (i.e., the channel supports rate r_i). A feasible channel allocation corresponds to a matching covering all nodes in \mathcal{N}'_I .

Consider any subset \mathcal{N}'_I of \mathcal{N}'_I . For user i , set \mathcal{N}'_I contains up to n_i nodes defined for this user. Let $m \subseteq I$ denote the set of users having at least one node in \mathcal{N}'_I . We now examine how many channel nodes in \mathcal{N}'_{II} are connected to the nodes in \mathcal{N}'_I . Note that, for any user i , the n_i nodes are all connected to the same set of channel nodes (i.e., nodes defined for channels in \mathcal{L}_{ir_i}). Thus, for user subset m , the most stringent case is to put all the n_i nodes in \mathcal{N}'_I , for all $i \in m$. User subset m , together with rates $r_i, i \in m$, corresponds to a constraint of (4b), and the term $n_i x_{i r_i n_i}$ is part of the left-hand side. Moreover, by definition, every channel in \mathcal{L}_{mv} , where v is the combination of rates $r_i, i \in m$, is connected to at least one node in \mathcal{N}'_I . Hence there are at least $\sum_{i \in m} n_i$ nodes in \mathcal{N}'_{II} connected to the nodes in \mathcal{N}'_I . By Hall's theorem for matching (Hall, 1935), for the defined bipartite graph, there is a matching of size $|\mathcal{N}'_I| = \sum_{i \in I} n_i$, and the result follows. \square

CUT is a row-oriented model. The next non-compact model does the opposite by being column oriented. It is observed that in terms of channel allocation, a solution to CACR resembles a partition of channels among the users. (It is not an exact partition, since there may be leftover channels.) We consider, for each user, candidate subsets of channels for allocation. Denote by S_i the set of all valid subsets of channels for user i . We denote by $r(s)$ the rate used for $s \in S_i$, and let $n(s) = |s|$. A subset $s \in S_i$ is called a valid subset if the following two conditions hold.

- Rate $r(s)$ is supported by all channels in s , and consequently the total rate achieved by the subset equals $n(s)r(s)$. Thus for $s \in S_i$, the utility $u_{is} = w_i n(s)r(s)$. In other words, all channels in the subset are used. Making this restriction is valid, because if some channels of s are not used, then there is a smaller subset $s' \subset s$ satisfying the condition with the same utility.
- The tuple $(r(s), n(s))$ for $s \in S_i$ meets the rate interval of user i , i.e., $(r(s), n(s)) \in D_i$.

Let z_{is} be a binary variable to represent if subset $s \in S_i$ is chosen, and $a_{i\ell}$ be a binary parameter to indicate if channel ℓ is in s . We can then formulate CACR as follows.

$$(COL) \max \sum_{i \in I} \sum_{s \in S_i} u_{is} z_{is} \quad (5a)$$

$$\text{s.t. } \sum_{s \in S_i} z_{is} = 1, \quad i \in I \quad (5b)$$

$$\sum_{i \in I} \sum_{s \in S_i} a_{s\ell} z_{is} \leq 1, \quad \ell \in \mathcal{L} \quad (5c)$$

$$z_{is} \in \{0, 1\}, \quad s \in S_i, i \in I \quad (5d)$$

Constraints (5b) and (5c) state, respectively, that one subset of channels has to be selected per user, and a channel appears at most once in the selected channel subsets. As the formulation has an exponential number of columns, we name it COL.

We remark that the first condition defining a valid subset has no impact on the LP optimal value. Suppose we relax this condition, and denote by \tilde{z} the LP optimal solution. Then for subsets s and s' above, we can increase $\tilde{z}_{i,s'}$ by an amount of $\tilde{z}_{i,s}$ and then set $\tilde{z}_{i,s} = 0$, without affecting the objective function value. Therefore, without loss of generality, we can restrict the consideration to valid subsets in analyzing COL.

6. Analysis of LP relaxations

In this section we examine the LP bounds of the presented optimization models. We use LP with a superscript containing a model's name, e.g., LP^{COMB} , to denote the optimal LP objective function value of the model. When comparing the values, we use " $<$ " to denote that " \leq " holds and there exist instances with strict inequality. Note that since CACR is maximization, lower LP value means a tighter bound.

As our first result, we show that the non-compact model COL is stronger than the compact model COMB.

Proposition 4. $LP^{COL} < LP^{COMB}$.

Proof. We prove that a feasible LP solution of COL can be mapped to a feasible solution in the LP of COMB with the same objective function value, and hence $LP^{COL} \leq LP^{COMB}$. We then provide an instance for which strict inequality holds.

Denote by \tilde{z}_{is} a feasible LP solution of COL. We define a solution to COMB as follows.

$$\begin{aligned} \tilde{x}_{irm} &= \sum_{s \in S_i: r(s)=r, n(s)=n} \tilde{z}_{is}, (r, n) \in D_i, i \in I \\ \tilde{y}_{i\ell} &= \sum_{s \in S_i} a_{s\ell} \tilde{z}_{is}, \ell \in \mathcal{L}_{ir}, i \in I \end{aligned} \quad (6)$$

For the solution defined in (6), constraint (2b) is satisfied. This is because, for each i and $s \in S_i$, $(r(s), n(s))$ is unique, that is, the value of variable \tilde{z}_{is} appears once in the left-hand side of (2b). The conclusion follows then from (5b) in COL. Next, for user i , $\sum_{i \in I: \ell \in \mathcal{L}_{i1}} \tilde{y}_{i\ell} = \sum_{i \in I: \ell \in \mathcal{L}_{i1}} \sum_{s \in S_i} a_{s\ell} \tilde{z}_{is} \leq 1$ by (5c).

We now examine the left-hand side of (2d). Plugging in (6), we obtain the following.

$$\sum_{(h,n) \in D_i: h \geq r} n \tilde{x}_{ihn} = \sum_{(h,n) \in D_i: h \geq r} n \sum_{s \in S_i: r(s)=h, n(s)=n} \tilde{z}_{is} = \sum_{s \in S_i: r(s) \geq r} n(s) \tilde{z}_{is} \quad (7)$$

The second equality is implied by that, for any $(h, n) \in D_i$ (with $h \geq r$), \tilde{z}_{is} appears in the definition of one \tilde{x}_{ihn} that is multiplied with n , if $r(s) = h$ and $n(s) = n$. Note that in this sum we do not need to require that $(r(s), n(s)) \in D_i$, because this is always the case, by the definition of S_i . For the right-hand side of (2d), by using again (6) and swapping the two sums, we obtain $\sum_{s \in S_i} \sum_{\ell \in \mathcal{L}_{ir}} a_{s\ell} \tilde{z}_{is}$. Thus, (2d) becomes the following inequality.

$$\sum_{s \in S_i: r(s) \geq r} n(s) \tilde{z}_{is} \leq \sum_{s \in S_i} \sum_{\ell \in \mathcal{L}_{ir}} a_{s\ell} \tilde{z}_{is} \quad (8)$$

It is clear that each \tilde{z}_{is} in the left-hand side of (8) also appears in the right-hand side. Note that $n(s)$ is the number of channels in s and since s is valid, these channels all support rate $r(s)$, and hence they also support rate r . Hence $\sum_{\ell \in \mathcal{L}_{ir}} a_{s\ell} \geq n(s)$, meaning that (8) and consequently (2d) holds. Therefore, $LP^{COL} \leq LP^{COMB}$.

Next, consider an instance of two users and three channels. The set of rates is $\{0.545, 0.814, 0.960\}$. The two users have weights 0.44

and 0.36 respectively. The first user has its demand interval $[0.6, 3.0]$, and that of the second user is $[0.9, 3.0]$. The achievable rates, given as two vectors for the two users respectively, are $(0.814, 0.960, 0.960)$ and $(0.545, 0.545, 0.960)$. One can verify that for this instance we have strict inequality, and the proposition follows. \square

Our next result states the relationship between COMB and the first non-compact model CUT. By this result, they are equally strong (or weak) in LP bound. We establish the result by showing $LP^{COMB} \leq LP^{CUT}$ and then $LP^{CUT} \leq LP^{COMB}$.

Lemma 5. $LP^{COMB} \leq LP^{CUT}$.

Proof. We prove that a feasible LP solution of COMB can be mapped to a feasible solution of the LP relaxation of CUT. Denote by \tilde{x}_{irm} and $\tilde{y}_{i\ell}$ a feasible LP solution of COMB. Now consider a generic user subset m and rate selection $v \in \mathcal{V}_m$ in CUT. For \tilde{x} , (2d) holds for any user $i \in m$ with rate r_i^v . Summing up both sides of the inequalities, we obtain (9) below:

$$\sum_{i \in m} \sum_{(h,n) \in D_i: h \geq r_i^v} n \tilde{x}_{ihn} \leq \sum_{i \in m} \sum_{\ell \in \mathcal{L}_{ir_i^v}} \tilde{y}_{i\ell} = \sum_{\ell \in \mathcal{L}} \sum_{i \in m} e_{\ell r_i^v} \tilde{y}_{i\ell} \leq \sum_{\ell \in \mathcal{L}} \max_{i \in m} e_{\ell r_i^v}, \quad (9)$$

where $e_{\ell r_i^v}$ is a binary indicator: $e_{\ell r_i^v} = 1$ if $\ell \in \mathcal{L}_{ir_i^v}$, otherwise $e_{\ell r_i^v} = 0$. The last inequality above is because $\sum_{i \in m} \tilde{y}_{i\ell} \leq 1$. Next, observe that if $\max_{i \in m} e_{\ell r_i^v} = 1$, then channel ℓ supports rate r_i^v for at least one user $i \in m$, and hence this channel contributes to \mathcal{L}_{mv} . Therefore, $\sum_{\ell \in \mathcal{L}} \max_{i \in m} e_{\ell r_i^v} \leq |\mathcal{L}_{mv}|$, and the lemma follows. \square

To show $LP^{COMB} \leq LP^{CUT}$, the underlying idea has similarity to proving that CUT is a valid model for CACR. The graph is however more complex than that outlined in the proof of Proposition 3, as we are dealing with a fractional solution, not an integer solution. Specifically, we construct a graph, and show that for any feasible solution of the LP relaxation of CUT, if the maximum flow equals the total capacity (based on an LP solution of CUT) on the arcs originating from the source, then the flow leads to a y -solution satisfying constraints (2c)–(2d) in the LP relaxation of COMB. If this is not the case, we prove there is a contradiction, namely at least one of (4b) is violated.

Lemma 6. $LP^{CUT} \leq LP^{COMB}$.

Proof. Denote by \tilde{x} a feasible solution of the LP relaxation of CUT. We construct a graph, shown in Fig. 2. Apart from the source and sink, there is one node defined for each combination of a user and a candidate rate. These are called user-rate nodes. For user i and rate r , the node is denoted by $\langle i, r \rangle$. In addition, a node is defined per channel. The arcs and their capacities are as follows. For clarity, we do not draw the arcs except for a representative one.

- There is an arc from the source to each user-rate node. For the arc ending at $\langle i, r \rangle$, the capacity is $\sum_{n: (r,n) \in D_i} n \tilde{x}_{irn}$.
- From node $\langle i, r \rangle$, there is one arc to each channel supporting rate r , with capacity one.
- Finally, from each channel node, there is an arc to the sink with capacity one.

For the maximum flow from the source to the sink, denote by $\tilde{f}_{\langle i, r \rangle \ell}$ the flow on arc $(\langle i, r \rangle, \ell)$, and for user i and channel $\ell \in \mathcal{L}_{i1}$, let $\tilde{y}_{i\ell} = \sum_{r \in R: \ell \in \mathcal{L}_{ir}} \tilde{f}_{\langle i, r \rangle \ell}$. Due to the unit capacity from channel node ℓ to the sink, $\sum_{i \in I} \tilde{y}_{i\ell} \leq 1$, i.e., constraint (2c) is satisfied.

Suppose the maximum flow reaches $\sum_{i \in I} \sum_{r \in R} \sum_{n: (r,n) \in D_i} n \tilde{x}_{irn}$. In this case, the minimum cut consists of all arcs connecting the source to the user-rate nodes. Therefore, the flow from the source to $\langle i, r \rangle$ equals $\sum_{n: (r,n) \in D_i} n \tilde{x}_{irn}$. By flow balance, this amount of flow equals $\sum_{\ell \in \mathcal{L}_{ir}} \tilde{f}_{\langle i, r \rangle \ell}$.

Consider (2d) defined for user i and rate r' . (We use r' instead of r as the latter is used as a running index in defining \tilde{y} .) The left-hand side is

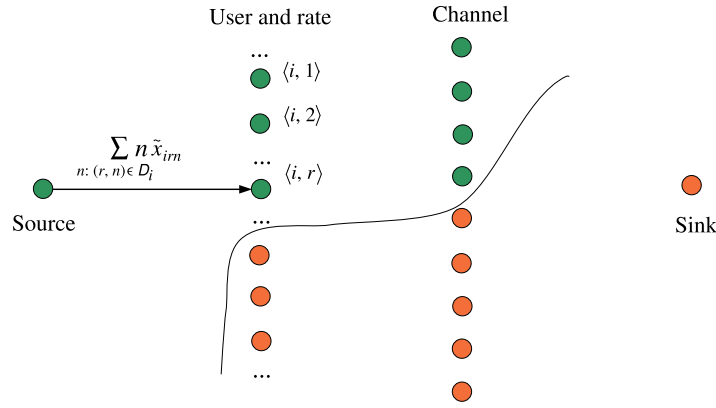


Fig. 2. The graph used to show $LP^{CUT} \leq LP^{COMB}$.

$\sum_{(h,n) \in D_i: h \geq r'} n \tilde{x}_{ihn} = \sum_{h \geq r'} \sum_{n: (h,n) \in D_i} n \tilde{x}_{ihn} = \sum_{h \geq r'} \sum_{\ell \in \mathcal{L}_{ih}} \tilde{f}_{(i,h)\ell}$. Note that $\ell \in \mathcal{L}_{ih}$ implies $\ell \in \mathcal{L}_{ir'}$ for $h \geq r'$. Therefore, a channel ℓ is relevant to the terms in the sum, if and only if $\ell \in \mathcal{L}_{ir'}$. This leads to (10), showing that constraint (2d) is satisfied. In conclusion, \tilde{x} and \tilde{y} form a feasible solution to the LP of COMB.

$$\sum_{h \geq r'} \sum_{\ell \in \mathcal{L}_{ih}} \tilde{f}_{(i,h)\ell} = \sum_{\ell \in \mathcal{L}_{ir'}} \sum_{h \geq r': \ell \in \mathcal{L}_{ih}} \tilde{f}_{(i,h)\ell} \leq \sum_{\ell \in \mathcal{L}_{ir'}} \sum_{r: \ell \in \mathcal{L}_{ir}} \tilde{f}_{(i,r)\ell} = \sum_{\ell \in \mathcal{L}_{ir'}} \tilde{y}_i \ell \quad (10)$$

Let us assume the maximum flow is strictly less than the case analyzed above. Denote by \mathcal{N}_1 and \mathcal{N}_2 the subsets of nodes containing the source and sink, respectively, for a minimum cut. First, \mathcal{N}_1 will not be a singleton with the source only, as that cut has capacity $\sum_{i \in I} \sum_{r \in R} \sum_{n: (r,n) \in D_i} n \tilde{x}_{irn}$, contracting the current assumption. Second, \mathcal{N}_2 will not be a singleton of the sink. Suppose the opposite, then the cut capacity is L and this value cannot be less than $\sum_{i \in I} \sum_{r \in R} \sum_{n: (r,n) \in D_i} n \tilde{x}_{irn}$. This is because there is a constraint of (4b) for all users and their lowest possible rates, for which $\sum_{i \in I} \sum_{r \in R} \sum_{n: (r,n) \in D_i} n \tilde{x}_{irn}$ is the left-hand side and the right-hand side equals L . Moreover, the minimum cut will not be formed by all the arcs connecting the user-rate nodes to the channel nodes either, as the number of arcs in this cut is at least as large as L . Therefore, the minimum cut has the structure shown in Fig. 2, where the green and red colors are used to indicate the nodes in \mathcal{N}_1 and \mathcal{N}_2 , respectively. Next, we make and prove a series of claims.

1. There is a minimum cut, such that there is no arc from a green user-rate node to a red channel node. Suppose the opposite, and say arc $(\langle i, r \rangle, \ell)$ is in the cut, with green (i, r) and red ℓ . Consider moving ℓ to \mathcal{N}_1 instead. The capacity of the new cut increases by one due to the arc from ℓ to the sink, but decreases by at least one as $(\langle i, r \rangle, \ell)$ is no longer in the cut.
2. For the minimum cut above, for each green channel node, there is at least one arc from a green user-rate node. To see this, suppose the opposite for a green channel node ℓ . If we move this node to \mathcal{N}_2 , the new cut would have lower capacity, because the arc from ℓ to the sink is no longer in the cut, and there is no new arc added to the cut.
3. The flow from the source to a red user-rate node $\langle i, r \rangle$ equals $\sum_{n: (r,n) \in D_i} n \tilde{x}_{irn}$, because the arc is a forward arc in the minimum cut and hence must be saturated. Moreover, the flow on an arc from a red user-rate node to a green channel node is zero, as this is a backward arc in the minimum cut.
4. If a user-rate node $\langle i, r \rangle$ is green, then any node $\langle i, h \rangle$ is also green, i.e., they are in \mathcal{N}_1 , if $h > r$. To see this, suppose $\langle i, r \rangle$ is green but $\langle i, h \rangle$ is red. Consider moving $\langle i, h \rangle$ from \mathcal{N}_2 to \mathcal{N}_1 . Then, an amount of $\sum_{n: (h,n) \in D_i} \tilde{x}_{ihn}$ is no longer part of the cut capacity. By the first property, there is no arc from $\langle i, r \rangle$ to

any red channel node, this holds true also for $\langle i, h \rangle$ because any channel supporting h will also support r . Hence the cut capacity, after the update, will decrease by $\sum_{n: (h,n) \in D_i} \tilde{x}_{ihn} \geq 0$.

Denote by m the set of users, for which there is at least one green user-rate node. For $i \in m$, denote by r_i the lowest rate for which node $\langle i, r_i \rangle$ is green. By the last property above, all nodes $\langle i, h \rangle$ with $h > r_i$ are green as well. Denote further by L_g the number of green channels.

The cut capacity consists of two parts. The first is the total arc capacity from the source to the red user-rate nodes, and this capacity equals $\sum_{i \in I} \sum_{r=1}^{r_i-1} \sum_{n: (r,n) \in D_i} \tilde{x}_{irn}$. The second part is from the L_g green channel nodes to the sink. By the assumption that the cut capacity is smaller than $\sum_{i \in I} \sum_{r \in R} \sum_{n: (r,n) \in D_i} n \tilde{x}_{irn}$, we obtain the following.

$$\sum_{i \in I} \sum_{r=1}^{r_i-1} \sum_{n: (r,n) \in D_i} \tilde{x}_{irn} + L_g < \sum_{i \in I} \sum_{r \in R} \sum_{n: (r,n) \in D_i} n \tilde{x}_{irn} \Rightarrow L_g < \sum_{i \in m} \sum_{r=r_i}^R \sum_{n: (r,n) \in D_i} \tilde{x}_{irn} \quad (11)$$

The right-hand side of the second inequality in (11) is the left-hand side of a constraint of (4b) for user set m and rate $r_i, i \in m$. Moreover, by the first and second properties, L_g equals the right-hand side of this constraint, meaning that the constraint is violated. This contradiction completes the proof. \square

Lemmas 5–6 together give the relationship of LP bounding via COMB and CUT. We formalize the result below.

Corollary 7. $LP^{COMB} = LP^{CUT}$.

The next question is the strength of the compact model SEP given in Section 3. The answer is that it provides the same LP bound as that of COL. A direct proof, based on showing that any solution of the LP of one of them maps to a solution in the other LP, and vice versa, turned out to be difficult, however. We take a different approach in our proof, using the notion of Lagrangian relaxation and Dantzig–Wolfe decomposition.

Lemma 8. Applying Lagrangian relaxation to (1d), the optimal value of the Lagrangian dual function equals the LP optimum of COL.

Proof. Denote by $\pi_\ell \geq 0, \ell \in \mathcal{L}$, the Lagrangian multipliers. The Lagrangian subproblem decomposes by user. For a generic user, the subproblem is as follows. For clarity, we omit the user index.

$$\max \sum_{r \in R} \sum_{\ell \in \mathcal{L}_r} (u_r - \pi_\ell) y_{r\ell} \quad (12a)$$

$$\text{s.t.} \quad \sum_{r \in R} x_r = 1, \quad (12b)$$

$$y_{r\ell} \leq x_r, \quad \ell \in \mathcal{L}_r, r \in R \quad (12c)$$

$$\underline{n}_r x_r \leq \sum_{\ell \in \mathcal{L}_r} y_{r\ell} \leq \bar{n}_r x_r, \quad r \in R \quad (12d)$$

$$x_r \in \{0, 1\}, \quad r \in \mathcal{R} \quad (12e)$$

$$y_{r\ell} \in \{0, 1\}, \quad \ell \in \mathcal{L}_r, r \in \mathcal{R} \quad (12f)$$

At the optimum of the subproblem, the solution for a user is a subset of channels and its utility equals the selected rate multiplied by the cardinality of the subset. Hence, the solution for user i corresponds to an element in S_i . By the properties of Lagrangian relaxation for integer programming, the optimal value of the Lagrangian dual is defined by the optimal solution obtained over the space defined as the intersection of the convex hull of the feasible solutions of the subproblem and the relaxed constraint (that is Dantzig–Wolfe reformulation, [Wolsey \(2020\)](#)). As the subproblem decomposes by user, the convex hull is the Cartesian product of those over the users, cf. (5c). It is rather straightforward to see that this Dantzig–Wolfe reformulation is in fact the LP of COL (after any necessary scaling), and the result follows. \square

Remark 5. For solving (12), we can consider the possible rates one by one. For each rate r , we sort the channels in \mathcal{L}_r in descending order of $(u_r - \pi_\ell)$. The first \underline{n}_r channels are selected. Next, we continue selecting the remaining channels in the sorted sequence, and stops once the composite objective coefficient is non-positive, or \bar{n}_r channels are selected. It is rather obvious that the solution procedure leads to the optimum of the subproblem. \square

Lemma 9. *Subproblem (12) has integrality property.*

Proof. Let $\mu_{r\ell} = u_r - \pi_\ell, \ell \in \mathcal{L}_r, r \in \mathcal{R}$. Consider the LP relaxation of subproblem (12), and some rate $r \in \mathcal{R}$. For rate r and any $0 < x_r \leq 1$, the optimization in the y -variables of this rate is as follows. Without loss of generality, we assume that the channel indices are sorted in descending order of μ_r , i.e., $\mu_{r1} \geq \mu_{r2} \geq \dots \geq \mu_{r|\mathcal{L}_r|}$. This order is assumed throughout the proof.

$$\max \sum_{\ell \in \mathcal{L}_r} \mu_{r\ell} y_{r\ell} \quad (13a)$$

$$\text{s.t. } y_{r\ell} \leq x_r, \quad \ell \in \mathcal{L}_r \quad (13b)$$

$$\underline{n}_r x_r \leq \sum_{\ell \in \mathcal{L}_r} y_{r\ell} \leq \bar{n}_r x_r \quad (13c)$$

$$0 \leq y_{r\ell} \leq 1, \quad \ell \in \mathcal{L}_r \quad (13d)$$

Note that in (13), the value of x_r is given. Denote by $y_{r\ell}^*, \ell \in \mathcal{L}_r$, the optimum of (13). We will show that there is an optimal solution, consisting of $y_{r\ell}^* = x_r$ for the first n_r^* channels for some integer $n_r^* \geq 1$, for any $0 < x_r \leq 1$.

Suppose this is not the case. Starting from channel one and following the channel indices, assume k is the first channel, such that $y_{rk}^* < x_r$. Note that either y_{rk}^* is the only variable being strictly less than x_r , or there are more such y -variables. In the latter case, we continue scanning the variables and stop at the next channel, say channel k' , with $y_{rk'}^* < x_r$. We increase the value of y_{rk}^* up to x_r , while reducing $y_{rk'}^*$ with the same amount. Note that doing so will not worsen the objective function value, because $\mu_{rk} \geq \mu_{rk'}$. Moreover, the constraints remain satisfied, because the individual upper bounds are adhered to, and the sum $\sum_{\ell \in \mathcal{L}_r} y_{r\ell}^*$ is not altered. The update has two possible outcomes. In the first case, $y_{rk}^* = x_r$ and $y_{rk'}^* > 0$ even though the value is reduced. We let $k = k'$ and repeat the process for the updated index k . In the second case, after the update we still have $y_{rk}^* < x_r$, and $y_{rk'}^* = 0$. For this case, we continue the scan of the remaining indices to look for the next variable being strictly less than x_r .

Applying the above process, repeatedly if necessary, we end up with an optimal LP solution where there is at most one index k with $0 < y_{rk}^* < x_r$, such that $y_{r1}^* = \dots = y_{r,k-1}^* = x_r$, and $y_{r,k+1}^* = \dots = y_{r|\mathcal{L}_r|}^* = 0$. Suppose there is such a variable y_{rk}^* . Let us examine the sign of μ_{rk} . First, assume $\mu_{rk} \geq 0$. Note that $\sum_{\ell=1}^{k-1} y_{r\ell}^* = (k-1)x_r$. Hence $k-1 < \bar{n}_r$ as otherwise having $y_{rk}^* > 0$ violates the upper limit $\bar{n}_r x_r$. Therefore, we

can increase the value of y_{rk}^* to x_r . The objective function value will not decrease as $\mu_{rk} \geq 0$ (in fact $\mu_{rk} = 0$ as otherwise the optimality of y^* is contradicted). Now suppose $\mu_{rk} < 0$. Note that then $k-1 \geq \underline{n}_r$, i.e., the lower limit is met by the first $k-1$ variables, because otherwise the lower limit cannot be satisfied with $y_{rk}^* < x_r$. Hence we can decrease the value of y_{rk}^* and improve the objective value, a contradiction. Finally, for the special case where $\underline{n}_r = \bar{n}_r$, it is easy to see that having one $y_{rk}^* < x_r$ will not occur, as either the upper or lower limit in (13c) is violated.

By the above arguments, we conclude that there is an optimal LP solution in the y -variables, such that the first k variables all equal x_r , and the remaining variables are zero. To be complete, we still need to show that this number k is invariant to the value of x_r . Observing the constraints of (13), it is clear that if for some x_r , the optimal number is k , then the solution, by setting $y_{r1} = \dots = y_{rk} = x_r$, is feasible for any value of $x_r \in [0, 1]$, with the objective function value being scaled by x_r .

Consider two numbers k_1 and k_2 , both leading to feasible solutions by setting the first k_1 and k_2 elements of y to be x_r . The objective function values are then $x_r \sum_{\ell=1}^{k_1} \mu_{r\ell}$ and $x_r \sum_{\ell=1}^{k_2} \mu_{r\ell}$, respectively. Suppose $x_r \sum_{\ell=1}^{k_1} \mu_{r\ell} \geq x_r \sum_{\ell=1}^{k_2} \mu_{r\ell}$. It is obvious that the inequality holds for all $x_r \in [0, 1]$. Therefore, for rate r , there exists an optimal LP solution, such that there is one single number n_r^* , and the first n_r^* elements of the y -variables equal x_r and thus they increase linearly in x_r .

Based on the above, the LP relaxation of subproblem (12) is equivalent to the following problem in the x -variables.

$$\max \sum_{r \in \mathcal{R}} \left(\sum_{\ell=1}^{n_r^*} \mu_{r\ell} \right) x_r \quad (14a)$$

$$\text{s.t. } \sum_{r \in \mathcal{R}} x_r = 1 \quad (14b)$$

$$0 \leq x_r \leq 1, \quad r \in \mathcal{R} \quad (14c)$$

The above problem (14) obviously has an optimal solution that is integer, and the proof is complete. \square

Proposition 10. $LP^{SEP} = LP^{COL}$.

Proof. LP^{SEP} equals the Lagrangian dual optimum due to the integrality property (Lemma 9). The result then follows from Lemma 8. \square

What remains to examine is model COMB'. It turns out that there is no deterministic relation between its LP bound and that of COMB. Of this pair, one may outperform the other, depending on problem instance. Thus we can obtain a strengthened version of COMB (and COMB'), by integrating the two into one model. We denote this model by COMB+.

Lemma 11. *Applying Lagrangian relaxation to (2c) of COMB+, the optimal value of the Lagrangian dual function equals the LP optimum of COL.*

Proof. The subproblem decomposes by user in the Lagrangian relaxation, and the solution of a user's subproblem again is a subset of channels. The lemma is established by applying the same arguments as in the proof of Lemma 8. \square

For COMB+, the Lagrangian subproblem formulation does not exhibit the integrality property, meaning that the Lagrangian dual optimum may be strictly better than the LP relaxation, and this is verified by numerical tests. Therefore we have the following result.

Proposition 12. $LP^{SEP} < LP^{COMB+}$.

Remark 6. We used a direct proof for Proposition 4. The proposition can be alternatively established via Lagrangian relaxation and the lack of integrality property of the subproblem of COMB. \square

Table 1
Parameters used for generating CACR instances. (The rates are in megabits per second (Mbps)).

Parameters	Value(s)
Number of users I	{20, 30, 40, 50}
Number of channels L	100
Set of rates \mathcal{R}	{0, 0.158, 0.212, 0.305, 0.433, 0.545, 0.650, 0.758, 0.814, 0.960}
Rate lower bound $(\underline{d}_i, i \in I)$	Uniformly zero, or random within [0, 20]
Rate upper bound $(\bar{d}_i, i \in I)$	Random within [20, 100]
User weight $(w_i, i \in I)$	Random within [10, 100]
Channel quality	COST-231-HATA plus randomly generated interference

To summarize, COMB and CUT are equivalent in LP bound, and these two formulations are weaker than SEP and COL. For SEP and COL, they form another LP-equivalent pair. Moreover, none of COMB and COMB' dominates the other. The combination of these two, COMB+, remains weaker than SEP and COL.

7. Numerical results

7.1. Experimental setup

We have generated three groups of CACR instances using parameters that are very typical for performance evaluation of mobile communication systems. For each group, the number of users varies from 20 to 50. The number of channels is set to be 100 for all three groups. There are 100 instances per group. The users are located randomly within the coverage area of the BS, of which the radius is 1 Km. The path loss between the BS and a user follows a widely used model, known as the COST-231-HATA model (Damosso and Correia, 1999). The interference on a channel is modeled by assuming some additional surrounding BSs. Each channel is used by a randomly selected subset of these BSs. Thus the interference varies by channel. The maximum achievable rate on a channel is then determined by the channel bandwidth (assumed to be 180 kHz) and signal-to-interference-and-noise ratio (SINR³). A specification of instance generation is given in Table 1.

The purpose of our numerical experiments is two-fold. First, we are interested in making numerical observations of the optimization formulations in their LP bounds. Second, we would like to use numerical results to shed some light on the performance of LP-assisted problem solving. Note that, by the analysis in Section 6, it is sufficient to confine the evaluation to the compact models, namely, SEP, COMB, COMB', and COMB+. Moreover, in light of Proposition 2, it is reasonable to use LP to derive the user rates, as then an integer solution can be obtained via solving a network flow problem. In SEP, the rate of user i is represented by variable x_{ir} . In the other three compact models, this is given by $\sum_{n:(r,n) \in D_i} x_{irn}$. We consider the following schemes for approaching an integer solution.

- **LP + simple rounding:** After solving the LP, we fix the rate for every user, by finding the largest value of the above entities over the rates.
- **LP + iterative rounding:** This is a more graceful version for setting the rate. In one iteration, one user and rate with the largest fractional value is selected, and the rate is fixed for this user. The LP is then solved again, until the rate is determined for all users.
- As a reference solution, we consider the integer solution found by a solver (Gurobi version 9.0) at the root node of branch-and-bound tree. This scheme is denoted by ILP⁰. For ILP⁰, we consider models SEP and COMB+ as they generally lead to better performance.

³ This is the ratio between received signal strength (i.e., the transmission power on a channel scaled by the path loss), and the total received interference plus noise. The transmission power is set to 800 mW, and the noise is -174 dBm/Hz multiplied by the bandwidth of a channel.

We pay attention to the following two aspects in generating the instances. First, with the presence of a lower bound on the total rate to be delivered to the users, any heuristic may fail to find an integer feasible solution, even if such solutions exist. Second, the amount of variation of interference will most likely impact problem difficulty. This is because if the interference tends to be uniform over the channels, then from a user's perspective, the channels are largely invariant, i.e., there are many channels being identical in the rates supported. Because in our experiments a random number of the surrounding BSs is selected as interference source on a channel, the variation in channel quality can be steered by varying the number and locations of surrounding BSs. Having the two aspects in mind, the three instance groups differ as follows.

- For Group I, the rate lower bound is uniformly set to zero for all users. Consequently, constructing an integer solution is guaranteed to be successful, and hence the models are compared using all 100 instances of the group. As for interference, we assume there are six surrounding BSs.
- For Group II, rate lower bound is present. Hence whether or not a feasible solution is found is part of performance evaluation. The number of surrounding BSs is set to be two.
- The last instance group, Group III, also uses rate lower bound. Moreover, the number of surrounding BSs is six. Hence this instance group is expected to be more difficult than Group II, as a result of larger variation in interference.

We remark that all instances of Groups II and III are feasible, i.e., integer solutions do exist. We use the following metrics for performance comparison.

- **Gap:** For LP, this is the gap of the LP and the integer optimum. For LP-based problem-solving schemes, the gap value refers to the difference of the integer solution (if found) and the integer optimum.
- **Success ratio:** For Groups II and III, this is the percentage of instances for which feasible integer solutions are found.

Remark 7. For the LP gap values, we use box plot for presenting the distribution to provide a comprehensive performance picture. For the (heuristic) integer solutions, showing distribution is less justified, mainly because for Groups II and III, the schemes perform very differently on achieving feasible integer solutions over the instances, and the common subset of instances for which all the schemes lead to feasible solutions may be quite small. Moreover, the amount of results will be excessive as there are ten schemes in the comparison. We therefore provide tables showing the average results. Moreover, for Groups II and III, the comparison of gap of the integer solutions is based on a common subset of instances for which feasible solutions are found by all models and schemes. If the cardinality of this common set is less than ten (out of 100 instances), the model or solution scheme with the lowest success ratio is iteratively excluded, until the cardinality is at least ten. □

The focus of our computational experiments is not on solution time. This is because, from the time perspective, the solution schemes cannot be used as is in practice, where CACR need to be solved in millisecond timescale. In fact, even solving the LP once would exceed by far the

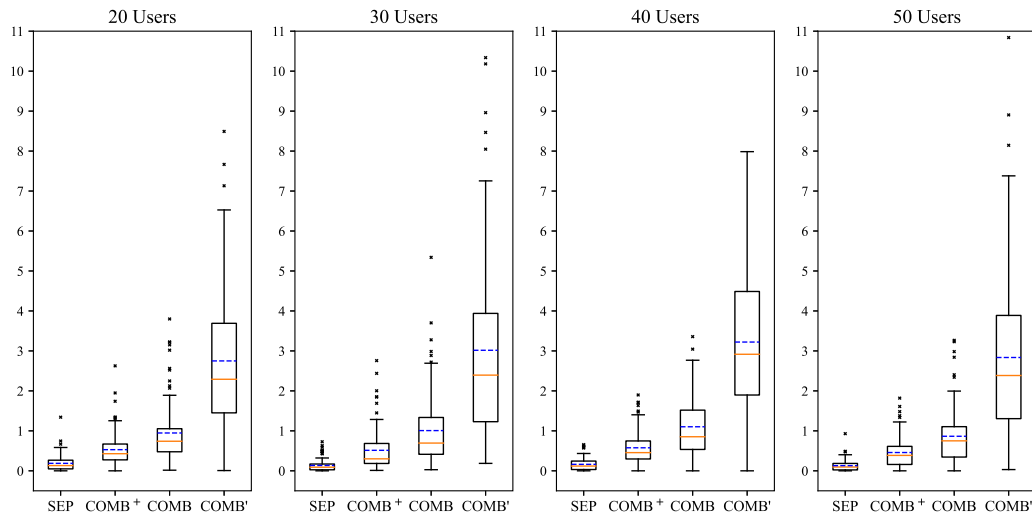


Fig. 3. LP gap (in percentage) for Group I.

time limit in real systems. Our focus is to understand the numerical performance of LP bounding and LP-based problem solving with respect to solution quality. The purpose is to set a ground for developing tailored methods in future works. Such methods, utilizing some of the optimization models in the current paper, are likely to be designed to enable massive parallel computation. Even though solution time is not the central aspect, we will provide some time results and comparisons. The results also highlight the need of forthcoming work.

7.2. Performance results for group I

For Group I, the comparison of LP gap is presented using box plots in Fig. 3. The bottom and top of each box are defined by the first and third quartiles, respectively. The average and median values are shown by the blue and orange lines, respectively. The two ends of a vertical line are obtained by the largest and smallest value samples, respectively, within a distance of 1.5 times of the interquartile range, from quartiles three and one. Finally, the outliers are represented with individual points.

We observe that COMB numerically has smaller gap than COMB' and the amount of difference is quite considerable, even though in theory there is no deterministic relation between the two models in LP bounding. This is probably attributed to that (2d) in COMB is a strengthened form of (3b) in COMB'. Combination COMB+ reduces the average gap of COMB by approximately 50%. Thus even if COMB' is inferior, its structure is helpful for improving COMB. For SEP, we know it is strongest in theory. Numerically, it significantly outperforms the other models in bounding, with an average LP gap of being close to zero. That SEP performs better is also clearly seen from the vertical span of the boxes. In fact, the average gap is consistent with the span; a model with higher average gap also exhibits a larger spread of the gap values over the instances. The coherence also applies to outliers, namely in general a larger average gap is coupled with outliers further away from the average. Note that, for the median values (that reduce the impact of outliers compared to the average values), the relative performance between the models largely remain as observed above. Finally, the LP gap does not exhibit an increasing or decreasing trend in the number of users, and this holds for all models. A likely reason is the absence of a lower bound on the user rate — the system has the full flexibility of choosing which users to serve and how much, and having more users even improves this flexibility.

Let us now consider the performance of integer solutions presented in Table 2. For the two rounding schemes, the relative performance of the models are fully consistent with their respective LP strengths, which is quite expected. Note that COMB' performs poorly in solution

quality. Thus the level of accuracy of LP does matter for LP-assisted problem solving. This is also the case for ILP⁰, as the results of using SEP are clearly better than those derived via COMB. Moreover, similar to the LP gap, there is no clear trend in the quality of integer solutions with respect to the number of users.

Comparing the two rounding schemes, the iterative one outperforms simple rounding for all models and problem sizes. The improvement is most significant for COMB and COMB'. Best performance is obtained by using the solver — the integer solutions are extremely close to optimum, in particular for model SEP. This demonstrates the capability of sophisticated heuristics that are implemented in a state-of-the-art integer solver. We will observe later, though, that the good performance comes with more computing time.

7.3. Performance results for group II

The instances in Group II have more uniform channel conditions than those in Group I, which hints better results. However, this is counter-balanced by the presence of lower bound of user rate.

The LP gap values, shown in Fig. 4, lead to similar observations as made earlier. Namely, the relative performance of the models more or less remain as for Group I, in terms of the average, median, box span, and outliers. There are however also differences. First, the improvement by combining COMB and COMB' is now smaller, and the average LP gap as well as the gap span have a somewhat increasing trend with respect to the number of users. Both may be linked to the presence of rate lower bound. Moreover, in comparison to Fig. 3, the LP gap is generally smaller, and we believe this has to do with that the channel conditions are more uniform for Group II.

Table 3 shows the results of integer solutions. Let us first observe the results of the two rounding schemes. COMB' constantly fails in terms of reaching feasible solutions for most cases. Note that COMB' is the only model having an LP gap being greater than one percent. Apparently, this has a very large impact on the success ratio. For the other three models, there is no significant difference in success ratio. For these models, the success ratio is equal or close to 100% for 20 users, and then gradually drops when the number of users grows as there is more competition for the channels. It is interesting to note that the two rounding schemes have comparable performance in success ratio. This can be explained by that the LP solutions are driven by the objective function, and none of the schemes is actively targeting maximizing feasibility. For the optimality gap, using SEP gives the best results. Moreover, the benefit of iterative rounding is smaller (or even negative in one case) than that of Group I, possibly because the flexibility of resource allocation is low due to the rate lower bound.

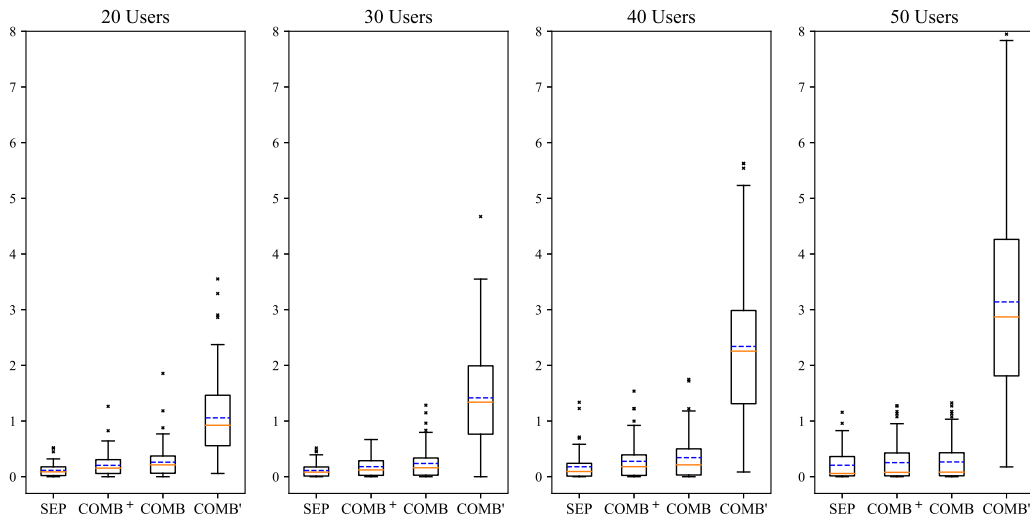


Fig. 4. LP gap (in percentage) for Group II.

Table 2
Average optimality gap of integer solutions for Group I.

Number of users	LP + simple rounding				LP + iterative rounding				ILP ⁰	
	SEP	COMB+	COMB	COMB'	SEP	COMB+	COMB	COMB'	SEP	COMB+
20	0.51%	1.31%	2.56%	11.22%	0.24%	0.41%	0.52%	4.83%	0.01%	0.08%
30	0.30%	1.44%	2.83%	11.45%	0.14%	0.43%	0.56%	4.22%	0.00%	0.08%
40	0.25%	1.31%	3.29%	12.12%	0.14%	0.37%	0.55%	3.76%	0.01%	0.04%
50	0.33%	1.26%	2.46%	11.10%	0.10%	0.37%	0.47%	3.17%	0.01%	0.05%

Table 3
Success ratio and the average gap (within parentheses) of integer solutions for Group II.

Number of users	LP + simple rounding				LP + iterative rounding				ILP ⁰	
	SEP	COMB+	COMB	COMB'	SEP	COMB+	COMB	COMB'	SEP	COMB+
20	100%	97%	100%	34%	100%	99%	100%	42%	100%	100%
	(0.28%)	(0.47%)	(0.61%)	(4.74%)	(0.19%)	(0.44%)	(0.72%)	(4.82%)	(0.00%)	(0.04%)
30	98%	96%	99%	6%	97%	97%	99%	6%	100%	100%
	(0.36%)	(0.45%)	(0.51%)	(-)	(0.26%)	(0.40%)	(0.44%)	(-)	(0.00%)	(0.02%)
40	96%	95%	96%	1%	97%	97%	98%	1%	100%	100%
	(0.23%)	(0.43%)	(0.46%)	(-)	(0.20%)	(0.34%)	(0.41%)	(-)	(0.01%)	(0.04%)
50	88%	87%	87%	0%	88%	87%	87%	0%	100%	100%
	(0.28%)	(0.28%)	(0.27%)	(-)	(0.24%)	(0.28%)	(0.27%)	(-)	(0.00%)	(0.07%)

Compared to the rounding schemes, ILP⁰ delivers the best results. Specifically, it manages to find feasible solutions for all instances, with very small or even zero gap in optimality. Again, a model with a better LP bound (SEP) also means better integer solutions by the solver. As the solver derives integer solutions based on LP, the results in fact show that LP-based models do have the potential of approaching good solutions for CACR, and the LP strength does matter.

7.4. Performance results for group III

For Group III, the performance in LP gap is presented in Fig. 5. Recall that the difference between Groups II and III is that the channels are less uniform in the latter, meaning that the instances are harder. Inspecting Fig. 5, this is indeed the case; the average LP gap as well as its range are now clearly higher. Moreover, the difference in LP gap becomes larger across the models. Thus problem difficulty does translate into LP bound. As before, a better average value corresponds to lower value span. Note that COMB' is now becoming poor in LP bound, however adding its constraints to COMB still helps. In addition, the LP gap boxes move upward in the number of users for all models except for COMB, for which the LP gap stays at about one percent. We do not have a definite answer to the underlying reason, though

a possible explanation is that the structure of (2d), constructed to strengthen the LP, is not present in SEP or COMB'.

That the instances of Group III are harder than those of Group II is manifested not only by the LP gap, but also the results of integer solutions shown in Table 4. For the two rounding schemes, the success ratio drops for all models. SEP remains the model giving highest success ratio, though the values are considerably lower than those for Group II. In contrast, ILP⁰ is able to keep its 100% success ratio (albeit with more computing cost, as will be shown later). We remark that both rounding schemes are oblivious to the issue of infeasibility, and hence a feasibility-oriented scheme, e.g., giving priority to users with large rate requirement in resource allocation, will probably improve the success ratio.

With respect to optimality gap, the values are larger than those in Table 4, again confirming that a bigger variation in channel quality results in harder instances. This holds true also for ILP⁰. Finally, as one could expect, using the strongest model implies the best solution quality for all schemes.

7.5. Computing time

We provide additional results of computing time, based on a notebook with Intel i7-1165G7, using Gurobi optimizer version 9.5.2 with

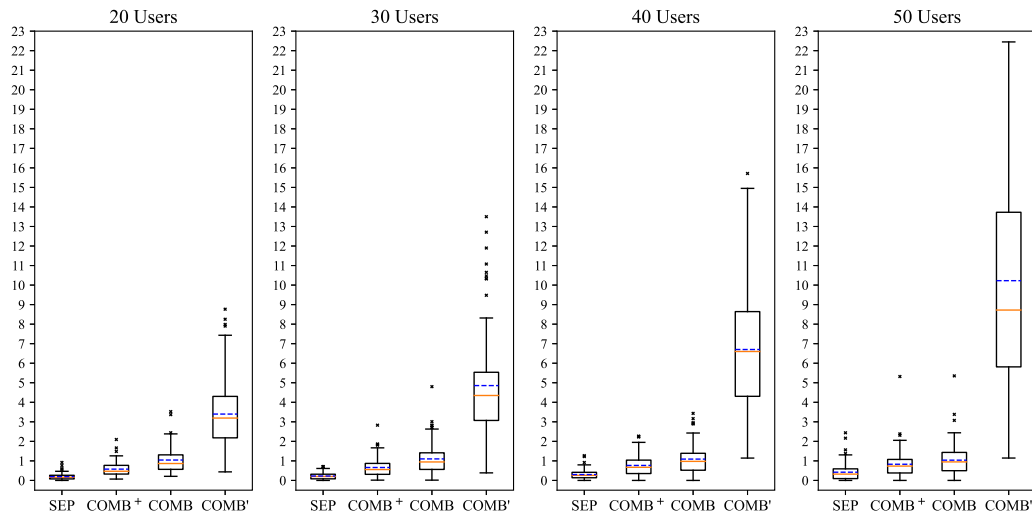


Fig. 5. LP gap (in percentage) for Group III. There is one outlier of value 34% for COMB' and 50 users that is not shown in the plot.

Table 4
Success ratio and average gap (within parentheses) of integers solutions for Group III.

Number of users	LP + simple rounding				LP + iterative rounding				ILP ⁰	
	SEP	COMB+	COMB	COMB'	SEP	COMB+	COMB	COMB'	SEP	COMB+
20	97% (0.54%)	94% (1.84%)	91% (2.55%)	16% (13.18%)	99% (0.40%)	96% (1.04%)	95% (1.65%)	22% (8.88%)	100% (0.02%)	100% (0.17%)
30	93% (0.49%)	79% (0.93%)	77% (1.47%)	1% (-)	92% (0.25%)	87% (0.76%)	85% (1.21%)	1% (0.25%)	100% (0.02%)	100% (0.24%)
40	83% (0.66%)	68% (1.03%)	68% (1.21%)	0% (-)	88% (0.48%)	80% (0.92%)	73% (1.02%)	0% (-)	100% (0.03%)	100% (0.20%)
50	73% (0.44%)	54% (0.93%)	55% (0.83%)	0% (-)	83% (0.49%)	59% (0.77%)	53% (0.99%)	0% (-)	100% (0.03%)	100% (0.18%)

Table 5
Average solution time in milliseconds.

	LP		LP + simple rounding		LP + iterative rounding		ILP		ILP ⁰	
	SEP	COMB+	SEP	COMB+	SEP	COMB+	SEP	COMB+	SEP	COMB+
	Group I	286	446	295	461	489	690	2451	3976	2391
Group II	213	314	225	326	389	596	1919	1654	1912	1641
Group II	347	371	357	381	536	666	3043	3067	2998	2813

one core and dual simplex as LP method. The instances of 50 users are considered, as computing time is more relevant for these instances than the smaller ones. Among the models, SEP and COMB+ are chosen because they perform better in the LP bound as well as the quality of integer solutions found. Also, we report average values for each group, since the variation in time is quite small across the instances.

The results of solution time (in milliseconds) are presented in Table 5, giving the following findings. In the table, ILP denotes to solve CACR to integer optimum.

- All schemes (including just solving the LP) present solution times that are at least a couple of orders of magnitude higher than what would be acceptable for real systems. Hence, as is, none of them can be for practical use (which is in fact expected), highlighting the need of additional research for highly tailored methods that admit massive parallel computation.
- SEP generally leads to lower solution times than COMB+, with the likely reason that the latter has a weaker LP to start with. Moreover, the most difficult instance group (Group III) is also the most time-demanding one.
- Iterative rounding takes, as expected, more time than simple rounding. However the increase is not dramatic. Moreover, both schemes are significantly faster than ILP and ILP⁰. Thus, that

the built-in solver heuristics are able to approach very good solutions comes at the cost of more time. One can observe that the difference in time for ILP⁰ and ILP is very small. Hence for the solver, the time bottleneck is finding integer solutions rather than performing branching.

7.6. Summary

To summarize, SEP has the best numerical performance, in terms of the LP bound, and the success ratio as well as the quality of LP-based integer solutions, whereas COMB' is on the other end of performance line. As the combination of COMB and COMB', COMB+'s improvement over COMB for the rounding schemes is noticeable but not dramatic, in particular for instances with rate lower bound. The use of built-in solver heuristics leads to clearly better integer solutions than the rounding schemes, at the cost of higher computing time. Finally, LP strength has a large impact on the quality of integer solutions derived based on LP relaxations.

8. Conclusions and future work

We have studied mathematical modeling of a resource optimization problem with common-rate channel allocation in mobile communication systems. We have demonstrated that the problem admits an array

of integer programming formulations. The study shows that modeling does matter, analytically as well as numerically, in terms of performance in bounding and LP-based problem solving. Moreover, there is a clear correlation between the accuracy of LP and the quality of integer solution derived thereby.

As we have observed, straightforward schemes of constructing integer solutions are not practical in solution time, whereas for our resource allocation problem, the ultimate target is real-time optimization. To approach this goal with model-based optimization, we make several observations. First, to deliver a solution rapidly, the amount of computation has to be very small, and thus the LP relaxation may have to be solved approximately, followed by some primal heuristic (both admitting massive parallel computation). Second, non-compact models are also of interest, as restricting the number of columns or rows leads to approximation of the problem (and the LP relaxation). Finally, with presence of rate lower bound, obtaining solution feasibility is of clear significance, and one may need to trade solution quality against feasibility. Developing and implementing algorithmic notions along these lines form interesting topics for forthcoming research.

CRedit authorship contribution statement

Yi Zhao: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Di Yuan:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Data availability

Data will be made available on request.

Acknowledgments

The research work has been funded by Huawei Technologies Sweden AB under contract number YBN2020085046, and the Swedish Research Council.

References

- 3GPP, 2018a. Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures. Technical Report TS 36.213, version 15.0.0, 3GPP, <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2427/>.
- 3GPP, 2018b. NR; Physical layer procedures for data. Technical Report TS 38.214, version 15.0.0, 3GPP.
- Al-Kanj, L., El-Beaino, W., El-Hajj, A.M., Dawy, Z., 2015. Optimized joint cell planning and BS on/off switching for LTE networks. *Wirel. Commun. Mob. Comput.* 16, 1537–1555. <http://dx.doi.org/10.1002/wcm.2610>.
- Ali, H., Liu, J., Ejaz, W., 2020. Planning capacity for 5G and beyond wireless networks by discrete fireworks algorithm with ensemble of local search methods. *EURASIP J. Wireless Commun. Networking* 2020, <http://dx.doi.org/10.1186/s13638-020-01798-y>.
- Berman, P., Karpinski, M., Scott, A.D., 2003. Approximation hardness of short symmetric instances of MAX-3SAT. *Electron. Colloquium Comput. Complex.* TR03-049, [arXiv:TR03-049](https://eccc.weizmann.ac.il/eccc-reports/2003/TR03-049/index.html), URL <https://eccc.weizmann.ac.il/eccc-reports/2003/TR03-049/index.html>.
- Bioardi, S., Capone, A., Sansó, B., 2013. Radio planning of energy-aware cellular networks. *Comput. Netw.* 57, 2564–2577. <http://dx.doi.org/10.1016/j.comnet.2013.05.003>.
- Calabrese, F.D., Wang, L., Ghadimi, E., Peters, G., Hanso, L., Soldati, P., 2018. Learning radio resource management in 5G networks: Framework, opportunities and challenges. *IEEE Commun. Mag.* 56, 138–145. <http://dx.doi.org/10.1109/MCOM.2018.1701031>.
- Chen, Y., Wu, Y., Hou, Y.T., Lou, W., 2021. mCore: Achieving sub-millisecond scheduling for 5G MU-MIMO systems. In: *Proceedings of IEEE INFOCOM*. <http://dx.doi.org/10.1109/INFOCOM42981.2021.9488684>.
- Chen, L., Yuan, D., 2010. Solving a minimum-power covering problem with overlap constraint for cellular network design. *European J. Oper. Res.* 203, 714–723. <http://dx.doi.org/10.1016/j.ejor.2009.09.018>.
- Cohen, R., Katz, L., 2010. Computational analysis and efficient algorithms for micro and macro OFDMA downlink scheduling. *IEEE/ACM Trans. Netw.* 18, 15–26. <http://dx.doi.org/10.1145/1816288.1816290>.
- Damosso, E., Correia, L.M. (Eds.), 1999. *Digital Mobile Radio Towards Future Generation Systems*. COST 231, COST 231 Final Report.
- D'Andreagiovanni, F., Mannino, C., Sassano, A., 2012. GUB covers and power-indexed formulations for wireless network design. *Manage. Sci.* 54, 142–156. <http://dx.doi.org/10.1287/mnsc.1120.1571>.
- Dreifuerst, R.M., Daulton, S., Qian, Y., Varkey, P., Balandat, M., Kasturia, S., Tomar, A., Yazdan, A., Ponnampalam, V., Heath, R.W., 2021. Optimizing coverage and capacity in cellular networks using machine learning. In: *Proceedings of 2021 IEEE International Conference on Acoustics, Speech and Signal Processing*. ICASSP, <http://dx.doi.org/10.1109/ICASSP39728.2021.9414155>.
- Ecker, G., Yuan, D., Koster, A.M.C.A., Schmeink, A., 2019. Accurate optimization models for interference constrained bandwidth allocation in cellular networks. *Comput. Oper. Res.* 101, 1–12. <http://dx.doi.org/10.1016/j.cor.2018.08.011>.
- Fallgren, M., 2010. On the complexity of maximizing the minimum Shannon capacity in wireless networks by joint channel assignment and power allocation. In: *Proceedings of IEEE International Workshop on Quality of Service*. IWQoS, pp. 1–7. <http://dx.doi.org/10.1109/IWQoS.2010.5542766>.
- Hall, P., 1935. On representatives of subsets. *J. Lond. Math. Soc.* 10, 26–30. <http://dx.doi.org/10.1112/jlms/s1-10.37.26>.
- Huang, Y., Li, S., Hou, Y.T., Lou, W., 2021. GPF+: A novel ultrafast GPU-based proportional fair scheduler for 5G NR. *IEEE/ACM Trans. Netw.* 30, 601–615. <http://dx.doi.org/10.1109/TNET.2021.3118005>.
- Joung, J., Ho, K.-C., Yuan, D., Sun, S., 2012. Energy efficient network-flow-based algorithm for multiuser multicarrier systems. *IET Networks* 1, 66–73. <http://dx.doi.org/10.1049/iet-net.2012.0007>.
- Kuang, Q., Utschick, W., Dotzler, A., 2016. Optimal joint user association and multi-pattern resource allocation in heterogeneous networks. *IEEE Trans. Signal Process.* 64, 3388–3401. <http://dx.doi.org/10.1109/TSP.2016.2548998>.
- Letchford, A., Ni, Q., Zhong, Z., 2017. An exact algorithm for a resource allocation problem in mobile wireless communications. *Comput. Optim. Appl.* 68, 193–208. <http://dx.doi.org/10.1007/s10589-017-9914-9>.
- Lin, Y.B., Chiu, T.H., Su, Y.T., 2009. Optimal and near-optimal resource allocation algorithms for OFDMA networks. *IEEE Trans. Wireless Commun.* 8, 4066–4077. <http://dx.doi.org/10.1109/TWC.2009.080221>.
- Liu, Y.F., 2014. Complexity analysis of joint subcarrier and power allocation for the cellular downlink OFDMA system. *IEEE Wireless Commun. Lett.* 3, 661–664. <http://dx.doi.org/10.1109/LWC.2014.2352655>.
- Liu, Y.-F., Dai, Y.-H., 2014. On the complexity of joint subcarrier and power allocation for multi-user OFDMA systems. *IEEE Trans. Signal Process.* 62 (3), 583–596. <http://dx.doi.org/10.1109/TSP.2013.2293130>.
- Lopez-Perez, D., Chu, X., Vasilakos, A.V., Clausen, H., 2014. Power minimization based resource allocation for interference mitigation in OFDMA femtocell networks. *IEEE J. Sel. Areas Commun.* 32, 333–344. <http://dx.doi.org/10.1109/JSAC.2014.141213>.
- Mehrotra, A., Trick, M.A., 1995. A column generation approach for graph coloring. *INFORMS J. Comput.* 8, 344–354. <http://dx.doi.org/10.1287/ijoc.8.4.344>.
- Ng, D.W.K., Lo, E.S., Schober, R., 2012. Energy-efficient resource allocation in multicell OFDMA systems with limited backhaul capacity. *IEEE Trans. Wireless Commun.* 11, 3618–3631. <http://dx.doi.org/10.1109/TWC.2012.083112.111951>.
- Ng, C.Y., Sung, C.W., 2008. Low complexity subcarrier and power allocation for utility maximization in uplink OFDMA systems. *IEEE Trans. Wireless Commun.* 7 (5), 1667–1675. <http://dx.doi.org/10.1109/TWC.2008.060723>.
- Tse, D., Viswanath, P., 2005. *Fundamentals of Wireless Communication*. Cambridge University Press.
- Wang, F., Chen, W., Tang, H., Wu, Q., 2017. Joint optimization of user association, subchannel allocation, and power allocation in multi-cell multi-association OFDMA heterogeneous networks. *IEEE Trans. Commun.* 65, 2672–2684. <http://dx.doi.org/10.1109/TCOMM.2017.2678986>.
- Wolsey, L.A., 2020. *Integer Programming, second ed.* Wiley.
- Yuan, D., Joung, J., Ho, C.K., Sun, S., 2012. On tractability aspects of optimal resource allocation in OFDMA systems. *IEEE Trans. Veh. Technol.* 62, 863–873. <http://dx.doi.org/10.1109/TVT.2012.2225854>.