J Sched (2014) 17:385–398
DOI 10.1007/s10951-014-0367-z

Frequency capping in online advertising

Niv Buchbinder · Moran Feldman · Arpita Ghosh · Joseph Naor

Received: 3 October 2012 / Accepted: 2 January 2014 / Published online: 23 January 2014

Abstract We study the following online problem. There are n advertisers. Each advertiser a_i has a total demand d_i and a value v_i for each supply unit. Supply units arrive one by one in an online fashion, and must be allocated to an agent immediately. Each unit is associated with a user, and each advertiser a_i is willing to accept no more than f_i units associated with any single user (the value f_i is called the *frequency cap* of advertiser a_i). The goal is to design an online allocation algorithm maximizing the total value. We first show a deterministic $3/4$-competitiveness upper bound, which holds even when all frequency caps are 1, and all advertisers share identical values and demands. A competitive ratio approaching $1 - 1/e$ can be achieved via a reduction to a different model considered by Goel and Mehta (WINE '07: Workshop on Internet and Network, Economics: 335–340, 2007). Our main contribution is analyzing two $3/4$-competitive greedy algorithms for the cases of equal values, and arbitrary valuations with equal integral demand to frequency cap ratios. Finally, we give a primal-dual algorithm which may serve as a good starting point for improving upon the ratio of $1 - 1/e$.

Keywords Competitive analysis · Frequency capping · Advertising · Online allocation

A preliminary version of this paper appeared in the Proceedings of the Algorithms and Data Structures Symposium (WADS), New York, NY, 2011.

N. Buchbinder
Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel
e-mail: niv.buchbinder@gmail.com

M. Feldman (✉)
Institute of Core Computing Science, EPFL, Lausanne, Switzerland
e-mail: moran.feldman@epfl.ch

A. Ghosh
Information Science Department, Cornell University, Ithaca, NY, USA
e-mail: arpitaghosh@cornell.edu

J. Naor
Computer Science Department, Technion, Haifa, Israel
e-mail: naor@cs.technion.ac.il

1 Introduction

Display advertising, consisting of graphic or text-based ads embedded in webpages, constitutes a large portion of the revenue from Internet advertising, totaling billions of dollars in 2008. Display, or brand, advertising is typically sold by publishers or ad networks on a pay-per-impression basis, with the advertiser specifying the total number of impressions she wants (the demand) and the price she is willing to pay-per-impression.[1]

Since display ads are sold on a pay-per-impression rather than on a pay-per-click or pay-per-action basis, *effective delivery* of display ads is very important to maximize advertiser value—each impression that an advertiser pays for must be shown to *as valuable a user* as possible. One aspect of effectively delivering display ads, which has been widely studied, is good targeting—matching ads to users who are likely to be responsive to the content of the ad. Another very important, but less studied, aspect is *limiting user exposure* to an ad—displaying the same ad to a user multiple times diminishes value to the advertiser, since the incremental benefit from repeatedly displaying the same ad to a user is likely

[1] In contrast, sponsored search advertisers typically pay-per-click or per action, and usually have budgets, rather than a demand, or quota, on the total number of impressions.

to be small (a user is unlikely to react to an ad after he has seen it a few times).

The notion of limiting the number of times a user is exposed to a particular ad is called *frequency capping*,[2] and is often cited as a way to avoid banner ad burnout. That is, frequency capping prevents ads from being displayed repeatedly to the point where visitors are being overexposed, and response drops.[3] Serving frequency capped ads is a very real requirement to maximize the value delivered to display advertisers, particularly in the pay-per-impression structure of the display advertising market. This is recognized by a number of publishers and ad networks (for instance, RightMedia, Google, and Yahoo!) who already offer, or implicitly implement, frequency capping for their display advertisers.

Serving display ads subject to a frequency capping constraint poses an *online assignment* problem since the supply of users, or impressions, is not known to the ad server in advance. How should the ad server allocate impressions to advertisers when it does not know which users' impressions will arrive in the future? In this paper, we study the simplest abstractions of the assignment problems motivated by frequency capping.

Problem statement There are n advertisers. Advertiser i has value per impression v_i, which is the price she is willing to pay for an impression, and a demand d_i, which is the maximum number of impressions she is interested in. In addition, she also has a *frequency cap* f_i, which is the maximum number of times her ad can be displayed to the same user. That is, advertiser a_i pays v_i only for impressions from users who have not seen her ad more than f_i times. The set of advertisers, and their parameters, is known to the ad server in advance.

Impressions from users arrive online. We say an advertiser is *eligible* for an impression if she still has leftover demand, and has not yet exhausted her frequency cap for the user associated with this impression. When an impression arrives, the ad server must immediately decide which ad of which advertiser, among the set of eligible advertisers, to display for that impression. The total revenue obtained by an algorithm is the sum of the revenues from all impressions it allocates. We want to design algorithms that are competitive against the optimal offline allocation, which knows the supply of impressions (with their associated users) in advance. We note that this problem is captured by the model of Goel and Mehta (2007), see Sect. 1.1 for more details.

[2] See, e.g., http://www.marketingterms.com/dictionary/frequency_cap.

[3] While it might be argued that displaying an ad more than once to a user reinforces the advertiser's message, repeated display without an upper limit clearly diminishes value.

In the absence of the frequency capping constraints ($f_i = \infty$), the natural greedy algorithm, assigning each arriving impression to the eligible advertiser with the highest per-impression value v_i, is optimal. However, with the frequency capping constraint, the ad server faces a tradeoff between assigning an arriving impression to an advertiser with high v_i but large frequency cap (since the supply can stop anytime) and a lower value advertiser with a smaller frequency cap (since small f_i means this advertiser needs to be assigned to many distinct users). In fact, even when all advertisers have identical values (with arbitrary tie breaking), the greedy algorithm is not optimal, as the following example shows: there are two advertisers, the first with $v_1 = 1$, $f_1 = n$, and the second with $v_2 = 1 - \epsilon$ and $f_2 = 1$; both advertisers have demand n (the $1 - \epsilon$ is used for tie breaking). The sequence of users is $u_1, \ldots, u_n, u_{n+1}, \ldots, u_{n+1}$, where the last user appears n times (n impressions). The greedy allocation gets a value of $n + 1$, whereas the optimal offline allocation gets $2n$.

As the next example shows, however, it is not even the different frequency caps that lead to the suboptimality of the greedy algorithm: suppose there are $n + 1$ advertisers each with $f_i = 1$. The first n advertisers have value 1 and demand 1, and the last advertiser has value $1 - \epsilon$ and demand n. With the same arrival sequence of users, a greedy allocation, again, gets a value of $n + 1$, whereas the optimal value is $2n$. In fact, as we will show in Sect. 3, even when *all* values and demands are equal and *all* frequency caps are 1, no deterministic algorithm can have a competitive ratio better than $3/4$.

Distinction from online matching Finding a matching in a bipartite graph, where one side is known and the other side is exposed one vertex at a time, is known as *online matching*. While the problem of online allocation with frequency capping constraints appears to be similar to online matching, they are actually quite different. In the frequency capping problem, a-priori each impression can be assigned to any of the advertisers. Now, as the impressions arrive, in the language of online matching, the existence of an edge between an advertiser and an arriving impression *depends* on the previous assignments made by the algorithm because of the frequency capping constraint. Specifically, if the algorithm has already assigned enough impressions from user j to advertiser i, or has exhausted i's demand, there is no edge between advertiser i and a newly arrived impression of user j; otherwise, there is an edge. This means that an adversary can no longer control the set of edges hitting each new impression; instead, the online algorithm determines the set of edges using indirect means. While we expect this property to translate into better competitive ratios for the frequency capping problem, taking advantage of it is not easy; a fact which is demonstrated by the involved analysis for the natural greedy algorithm for the problem.

Results Our online assignment problem can also be stated abstractly as follows. There are n agents, each one has a total demand d_i and a value v_i for items. Items of different types arrive one by one in an online fashion, and must be allocated to an agent immediately. Agent i wants no more than f_i copies of any single type of item. How should an online algorithm assign each arriving item to agents to maximize value? This abstract statement suggests the following simpler algorithmic questions.

– *Equal values, arbitrary d_i, f_i* Suppose agents (advertisers) have identical values for items (impressions), that is, $v_i = 1$ for all i. Now, the goal of the online algorithm is simply to assign as many items as possible. Our main technical contribution is the analysis of a novel greedy algorithm, proving that it is 3/4-competitive; which is optimal for a deterministic algorithm. The first step towards this result is to show that we can assume without loss of generality that every advertiser has frequency cap of 1, i.e., wants no more than one impression from each user (the reduction is independent of advertisers having the same value, and also applies when advertisers have arbitrary values). This reduction is simple, yet crucial—for each of the cases we study, designing algorithms directly, with arbitrary frequency caps, turns out to be rather hard.
  We then analyze our greedy algorithm, which assigns arriving impressions in decreasing order of *total* demand amongst eligible advertisers, for instances with unit frequency cap. (Assigning greedily according to maximum *residual demand* does not work; this algorithm cannot do better than 2/3.) The unit frequency cap means that an advertiser is eligible for an impression if she has leftover demand, and has not yet been assigned an impression of this user. We first prove that any *non-lazy* algorithm has a competitive ratio of 3/4 when all demands are equal (in addition to the equal values); then we build on this analysis to account for the fact that advertisers have unequal demands.
  Combinatorial analysis of online algorithms is usually done via a potential function argument which shows that at each step, the change in the potential function plus the algorithm's revenue are comparable to the gain of the optimal solution. Surprisingly, our analysis considers only the final assignment, disregarding the way in which it is reached. This allows us to avoid coming up with a potential function (which in many cases seems to come "out of nowhere"), and skip the tedious consideration of each possible step.
  Our result is especially interesting in light of the known upper bounds for unweighted online matching: 0.5 and $1 - 1/e \approx 0.63$ for deterministic and randomized algorithms, respectively (Karp et al. 1990).
– *Arbitrary values, equal integral d_i/f_i* The ideas used in the analysis of the equal values case can be extended to analyze the case where advertisers have different values, but the same integral ratio of demand to frequency cap. We show here that the natural greedy algorithm, which assigns in decreasing order of value, has a competitive ratio of 3/4 for this case; again, this is optimal in the sense that no deterministic algorithm can do better.
– *Arbitrary values, d_i and f_i* Finally, for the general case with arbitrary values, demands and frequency caps, we design a primal-dual algorithm whose competitive ratio approaches $1-1/e \approx 0.63$ for $d_i/f_i \gg 1$; we also show an upper bound of $1/\sqrt{2} \approx 0.707$ for this case. A competitive ratio of $1 - 1/e$ for the general case is already known from the work of Goel and Mehta (2007), also under an assumption similar to ours.
  Our online primal-dual algorithm has an interesting feature: during the execution of the algorithm, it both increases and decreases primal variables. (We refer to our online problem as the dual problem.) The same algorithm and competitive ratio also apply when advertisers have *target sets*, i.e., they have value v_i for impressions from a set S_i of users, and value of 0 for other impressions. For this case, we have a matching upper bound for deterministic online algorithms, using the upper bound on online b-matching (Kalyanasundaram and Pruhs 2000). (See Sect. 1.1 for a discussion regarding Goel and Mehta (2007) and online primal-dual algorithms.)

1.1 Related work

The maximization of revenue in online ad auctions has received much attention in recent years (Blum et al. 2004; Blum and Hartline 2005; Mehta et al. 2007; Mahdian and Saberi 2006; Bansal et al. 2010; Buchbinder et al. 2007; Feldman et al. 2009a). The problem of designing online algorithms to maximize advertising revenue was introduced by the *adwords model* (Mehta et al. 2007): advertisers have budgets, and bids for different keywords. Keywords arrive online, and the goal is to match advertisers to keywords to maximize revenue, while respecting the advertisers' budget constraints. Goel and Mehta (2007) extend the adwords model, allowing advertisers to specify bids for keywords which are decreasing functions of the number of impressions (of the keyword) already assigned to the advertiser. Our frequency capping problem is, in fact, a special case of the model of (Goel and Mehta 2007) (but not of the adwords model of Mehta et al. (2007)), where keywords correspond to users, and the decreasing function takes the form of a step function with a cutoff equal to the frequency cap f_i of the advertiser. Hence, the $(1-1/e)$-competitive online algorithm of Goel and Mehta (2007) applies to our problem as well. On the other hand, the upper bounds in Goel and Mehta (2007) do not apply to our problem since the model of Goel and Mehta (2007) also captures online matching. Improving upon the ratio of $1-1/e$ in special cases is posed as an open problem in Goel and Mehta (2007).

Our greedy algorithms in Sects. 3 and 4 obtain a ratio of $3/4$, improving upon this ratio of $1-1/e$. While the competitive ratio of our algorithm in Sect. 5 is the same as that in Goel and Mehta (2007), the algorithms are quite different. Moreover, our model does not inherit the upper bound of $1-1/e$, and in fact, the best upper bound[4] for the case without target sets is $1/\sqrt{2}$. Also, while the most general problem we solve in this paper remains within the model of Goel and Mehta (2007), the most general and realistic version of the frequency capping problem (Sect. 6) cannot be stated as a special case of the model of Goel and Mehta (2007). For this model the question of both a competitive algorithm and an upper bound (tighter than $1-1/e$) are open.

The primal-dual framework for online problems, first introduced by Buchbinder and Naor (2009), has been shown to be useful in many online scenarios including ad auctions, see Bansal et al. (2012a,b), Alon et al. (2006, 2009), Buchbinder and Naor (2006), Buchbinder et al. (2007). Unlike these primal-dual algorithms, which update the primal variables monotonically in each round, our primal-dual algorithm is novel in that it reassigns primal variables several times during the execution of the algorithm; hence, the primal variables do not necessarily increase monotonically with each round of new supply.

Feldman et al. (2009b) consider frequency capping in a stochastic model, but they leave open the question of improving upon the $1-1/e$ ratio in this model. Finally, the work in Abrams and Vee (2007) also addresses user fatigue in the context of sponsored search; however, the model and algorithms substantially differ from ours.

2 Preliminaries

We denote by $ALG(\sigma)$ the revenue of algorithm ALG on a sequence σ of arrivals of impressions, and by $OPT(\sigma)$ the revenue of the optimal offline algorithm, which knows σ in advance. Our goal is to design an online algorithm ALG that assigns each impression, immediately upon arrival, to produce a feasible allocation whose total value $ALG(\sigma)$ is competitive against $OPT(\sigma)$ for any arrival sequence σ of impressions. The *natural greedy algorithm* for the problem, denoted by $GREEDY_V$, allocates each arriving impression to the eligible advertiser with the highest value (breaking ties arbitrarily, but consistently). The examples in the introduction show that $GREEDY_V$ is no better than $1/2$-competitive. The next theorem shows that these examples are tight. The proof of the theorem is based on matching every impression assigned by $GREEDY_V$ to up to two impressions of OPT.

Theorem 1 *The competitive ratio of $GREEDY_V$ is 1/2.*

Proof For a particular sequence of arrivals, let $o(t)$ and $g(t)$ be the advertisers for which an impression t is assigned by OPT and $GREEDY_V$, respectively. Let f denote the number of impressions assigned by both OPT and $GREEDY_V$ to the same advertiser, i.e., $|\{t \mid o(t) = g(t)\}|$.

If $v(o(t)) > v(g(t))$, and the reason $GREEDY_V$ does not allocate impression t to advertiser $o(t)$ is that $GREEDY_V$ has already allocated the maximum allowed number of impressions of the same user to $o(t)$, then there must be $t' \neq t$ such that t and t' are impressions of the same user, $g(t') = o(t)$ and $o(t') \neq o(t)$. Swapping the assignment of t and t' in OPT (i.e., allocating t to advertiser $o(t')$ and t' to advertiser $o(t)$) does not change its value, and increases the value of f by 1. Repeat this process for all such impressions. Since f is upper bounded by the total number of impressions, the process is repeated only a finite number of times. At the end of the process, if $v(o(t)) > v(g(t))$ for some impression t, the reason that $GREEDY_V$ does not allocate t to $o(t)$ must be that $GREEDY_V$ has exhausted the demand of $o(t)$.

For each impression t with $v(o(t)) > v(g(t))$, we map t to a distinct impression allocated, by $GREEDY_V$, to advertiser $o(t)$. Since $GREEDY_V$ exhausts the demand of $o(t)$, such a mapping is one-to-one. Therefore, each allocation of $GREEDY_V$ is counted at most twice to cover the allocation of OPT, so $GREEDY_V$ is 1/2-competitive. □

We now establish a reduction from general frequency caps to unit frequency caps which greatly simplifies our algorithms. The following theorem allows us to assume $f_i = 1$ in the rest of the paper.

Theorem 2 (Reduction to unit frequency cap) *For every frequency capping instance there is an equivalent instance where all frequency caps are* 1. *Moreover, any solution to the equivalent instance can be transformed in an online fashion to an equivalent solution of the original instance.*

Proof For any given instance I (v_i, d_i, f_i), we construct a new instance I', so that every feasible allocation of I can be mapped to a feasible allocation of I', and vice versa. We replace each advertiser a_i in I having $d_i > f_i$ with f_i new dummy advertisers with value v_i, frequency cap 1, and demand either $\lfloor d_i/f_i \rfloor$ or $\lceil d_i/f_i \rceil$ such that the sum of demands of these f_i advertisers is d_i. If $d_i \leq f_i$, replace a_i by d_i dummy advertisers with value v_i and demand and frequency cap 1 each.

[4] While $1-1/e$ is the best possible competitive factor for the model of Goel and Mehta (2007) since this model captures the adwords model of Mehta et al. (2007), the frequency capping problem does not generalize the adwords model of Mehta et al. (2007). Therefore, it does not follow that $1-1/e$ is an upper bound for our problem.

Given a feasible allocation in I', allocate all impressions assigned to the copies of advertiser a_i in I' to a_i in *I*. This allocation is feasible—for an advertiser with $d_i > f_i$, there are f_i dummy advertisers each with frequency cap 1 in I', so a_i does not receive more than f_i impressions from any single user since I' is a feasible allocation. Also, since the sum of the demands of the dummy advertisers in I' is equal to the demand d_i, the allocation in *I* does not exceed demand either. A similar argument applies for advertisers with $d_i < f_i$. This process can be done in an online fashion.

Conversely, given an allocation in *I*, construct an allocation in I' as follows. Assuming $f_i \leq d_i$, let $a_i(1), \ldots, a_i(f_i)$ denote the dummy new advertisers in I' corresponding to $a(i)$ in *I* (the other case, $d_i > f_i$, is similar), ordered according to non-increasing demand. Note that the demand difference between any two dummy advertisers is at most 1. Let $Q(i)$ be the set of impressions allocated to $a(i)$ in *I*. Order the impressions in $Q(i)$ such that all impressions that belong to the same user are subsequent. Then allocate the impressions to the advertisers $a_i(1), \ldots, a_i(f_i)$ one by one in a cyclic way. That is, allocate the *j*-th impression in $Q(i)$ to advertiser $j \bmod f_i$. Since no user occurs in $Q(i)$ more than f_i times, no user is assigned to any dummy advertiser more than once. Finally, this cyclic allocation maintains the invariant that the difference in residual demand of any two dummy advertisers is at most 1. Moreover, the allocation method always allocates the next impression to one of the dummy advertisers that have maximal residual demand. Thus, if the *j*-th impression in $Q(i)$ cannot be allocated due to demand overflow it means that all advertisers have residual demand 0. However, this cannot happen since $|Q(i)| \leq d_i$, i.e., $Q(i)$ is not larger than the total demand of the dummy advertisers. Thus, the instances *I* and I' are equivalent. □

3 Identical valuations

In this section, we assume all advertisers have identical valuations, i.e., for each advertiser a_i, $v_i = 1$. Let us begin with an upper bound on any deterministic online algorithm.

Theorem 3 *No deterministic online algorithm is better than 3/4-competitive, even if all advertisers have identical values, demands, and frequency caps.*

Proof Suppose there are two advertisers a_1 and a_2 with demand 2 each. Consider any deterministic online algorithm ALG. Assume the first three impressions belong to users u_1, u_2, and u_3. If ALG refuses to allocate any one of these impressions, then it cannot be better than 2/3-competitive, since OPT accepts them all. Otherwise, necessarily, one of the advertisers was assigned two impressions, and the other one was assigned only one impression. Assume, w.l.o.g., u_1 is the user whose impression was assigned on its own to an advertiser. Let the fourth impression belong to user u_1, then ALG cannot allocate the last impression, while OPT allocates all four impressions, yielding a ratio of 3/4. □

We now turn to online algorithms. A natural greedy algorithm is one that assigns an arriving impression to an eligible advertiser with the maximum residual demand. However, assigning according to residual demand, breaking ties arbitrarily, cannot have a competitive ratio better than 2/3, as the following example shows. There are two advertisers, with $d_1 = 1$ and $d_2 = 2$, with ties broken in favor of a_1. The sequence of arrivals is u_1, u_2, u_1. The residual demand algorithm allocates only two impressions: the first impression to a_2 and then the second impression to a_1. The optimal assignment, however, can assign all three impressions.

We show that an alternative greedy algorithm, named $GREEDY_D$, which assigns according to *total* demand, has a competitive ratio of 3/4. Hereby is algorithm $GREEDY_D$:

1. Sort advertisers $a_1, \ldots, a_n$ in a non-decreasing demand order ($d_1 \geq \cdots \geq d_n$).
2. Upon arrival of a new impression, assign it to the first eligible advertiser according to the above order.

We need the following notation. Let y_i denote the number of impressions assigned by $GREEDY_D$ to advertiser a_i, and let $y^* = \min_i y_i$. Let k denote the number of advertisers whose demand is exhausted by $GREEDY_D$. In Sect. 3.1, we analyze the case of equal demands (and $v_i = 1$), and in Sect. 3.2 we build on this analysis to deal with the case where demands are arbitrary. We include the proof of the equal demands case since it is simpler, yet gives some insight into the proof of the general case.

3.1 Equal demand case

Algorithm $GREEDY_D$ has the property that it is *non-lazy*, i.e., it allocates every impression it receives, unless no advertiser is eligible for it. We show that any non-lazy algorithm, including $GREEDY_D$, is 3/4-competitive when all advertisers have equal demand, denoted by d.

Theorem 4 *Let ALG be a non-lazy algorithm, and let σ be a sequence of input impressions. Then,* $\frac{ALG(\sigma)}{OPT(\sigma)} \geq 3/4$.

Before going into the proof of Theorem 4, consider the example depicted in Fig. 1. The rectangle is divided into three areas: R_1 is the total allocation of advertisers who have exhausted their demand, R_2 is the total allocation of advertisers who have not exhausted their demand, and R_3 is "unused" demand. We use two upper bounds on $OPT(\sigma) - ALG(\sigma)$:

$$|R_3| \leq (d - y^*) \cdot (n - k) \leq |R_2| \cdot (d - y^*)/y^*,$$

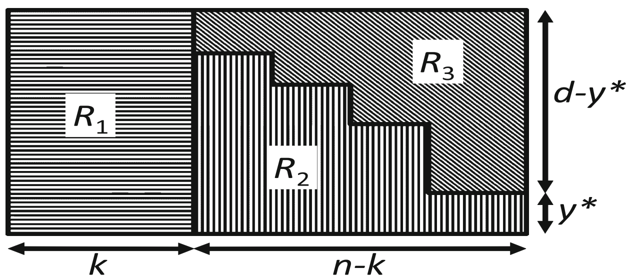

Fig. 1 An assignment constructed by the online algorithm ALG. Each *column* is an advertiser and each *row* corresponds to a unit demand

and

$k \cdot y^* \leq |R_1| \cdot y^*/d.$

Note that $y^* > 0$, since an advertiser who has received no impressions can always be assigned at least one impression without violating the frequency cap constraint. The theorem now follows from these bounds together with the observation $|ALG(\sigma)| = |R_1| + |R_2|$.

Let us now begin the formal proof of Theorem 4. Let A be the set of impressions allocated by OPT, and let $B \subseteq A$ be of size $OPT(\sigma) - ALG(\sigma)$. Associate each impression of B with an advertiser, such that up to $d - y_i$ impressions of B are associated with each advertiser a_i. This is possible since

$$\sum_{i=1}^{n}(d-y_i)=nd-ALG(\sigma) \geq OPT(\sigma)-ALG(\sigma)=|B|.$$

Lemma 1 $|B| = OPT(\sigma) - ALG(\sigma) \leq y^* k.$

Proof Let a_{i^*} be an advertiser for which $y_{i^*} = y^*$. If $y^* = d$, then $ALG(\sigma) = nd = OPT(\sigma)$, so we can assume $y^* < d$. Thus, each impression that ALG fails to allocate belongs to a user that already has an impression allocated to a_{i^*} (otherwise, ALG could have assigned it to a_{i^*}). Hence, there are at most y^* users having impressions that are not allocated to all advertisers. Each such user u can have at most k more impressions allocated by OPT than by ALG (if u has an unassigned impression, all $n-k$ advertisers with non-exhausted demands must have been assigned an impression of u). □

Next, for the purpose of analysis, we define two types of payments received by each impression $x \in B$. Suppose impression x is associated with advertiser a_i. The first payment x gets is $p_x = y_i/(d - y_i)$, and the second payment is $p'_x = d/y^*$.

Lemma 2 *The total payment received by all impressions of B is at most $ALG(\sigma)$.*

Proof Let E denote the set of advertisers whose demand is not exhausted by ALG (i.e., $|E| = n - k$). Let $a_i \in E$. For each impression x associated with a_i, we have $p_x = y_i/(d - y_i)$ and the number of such impressions is at most $d - y_i$. Therefore, the first type of payment received by impressions associated with a_i sums up to at most y_i. Adding up over all advertisers of E, the sum of the first type payments to all impressions in B is at most $\sum_{a_i \in E} y_i$. Since payments of the second type are all equal, they add up to

$$|B| \cdot \frac{d}{y^*} \leq y^* k \cdot \frac{d}{y^*} = dk.$$

Note that $dk + \sum_{a_i \in E} y_i = ALG(\sigma)$, since $a_i \notin E \Rightarrow y_i = d$, completing the proof. □

Lemma 3 *For each impression $x \in B$, $p_x + p'_x \geq 3$.*

Proof Suppose x is associated with an advertiser a_i. The total payment received by x is:

$$\frac{y_i}{d - y_i} + \frac{d}{y^*} \geq \frac{y^*}{d - y^*} + \frac{d}{y^*} = \frac{y^{*2} + d(d - y^*)}{y^*(d - y^*)}$$
$$= 3 + \frac{(2y^* - d)^2}{y^*(d - y^*)} \geq 3.$$

□

Corollary 1 $ALG(\sigma) \geq 3|B|$.

The proof of Theorem 4 is now immediate:

$$\frac{ALG(\sigma)}{OPT(\sigma)} = \frac{ALG(\sigma)}{ALG(\sigma) + |B|} \geq \frac{3|B|}{3|B| + |B|} = \frac{3}{4}. \quad (1)$$

3.2 General case

In this section we prove the main result of our paper. Unfortunately, the proof from the previous section does not readily generalize; the core of the difficulty is that it is no longer possible to sort the advertisers in a non-decreasing demand order in which all exhausted advertisers appear before the non-exhausted advertisers. Instead, exhausted and non-exhausted advertisers might be interleaved in every non-decreasing demand ordering of the advertisers. Thus, it is hard to guarantee the extent to which impressions of exhausted advertisers can be charged. A simple approach to overcome this difficulty is to split the advertisers into blocks, making sure that within each block the exhausted advertisers appear before the non-exhausted ones. However, this fails since OPT and $GREEDY_D$ may place impressions in different blocks. To circumvent this problem we consider subsets of advertisers having demand above a given threshold. The proof then makes a connection between the difference in number of impressions allocated by OPT and $GREEDY_D$ to a subset of the advertisers and the number of exhausted advertisers in the subset, yielding a lower bound on the payment that can be extracted from the impressions of the exhausted advertisers.

The next theorem shows that the competitive ratio of $GREEDY_D$ is 3/4 even when the demands are arbitrary.

Theorem 5 *Let σ be a sequence of input impressions. Then,* $\dfrac{GREEDY_D(\sigma)}{OPT(\sigma)} \geq 3/4$.

The rest of this section is devoted for proving Theorem 5. Let k_i be an indicator for the event that the demand of advertiser a_i is exhausted by the algorithm, i.e., $k_i = 1$ if $d_i = y_i$, and $k_i = 0$ otherwise. Note that $k = \sum_{i=1}^{n} k_i$. Let $OPT_i(\sigma)$ denote the number of impressions allocated by OPT to advertiser a_i.

We first prove a convenient property of the optimal solution. We construct a flow network N where there is a node for each user u_j and advertiser a_i, together with source node s and sink node t. The edges of the network are as follows:

- For each user u_j with m_j impressions, there is an edge of capacity m_j from s to u_j.
- For each advertiser a_i with demand d_i, there is an edge of capacity d_i from a_i to t.
- For each pair of user u_j and advertiser a_i, there is an edge of capacity 1 from u_j to a_i, representing the frequency cap which we assume, by Theorem 2, is 1.

Observation 1 *Every feasible integral flow f in the network N corresponds to a legal assignment of impressions to advertisers with revenue $|f|$, such that the number of impressions assigned to advertiser a_i is equal to the flow on the edge (a_i, t). The reverse direction holds as well.*

Using this observation we can prove the next lemma:

Lemma 4 *For any input sequence σ of impressions, and a feasible assignment S of impressions to advertisers, there is an optimal assignment S' such that for each advertiser a_i, S' assigns to a_i at least as many impressions as S does.*

Proof Let S'' be any optimal integral assignment. Using Observation 1, assignments S and S'' define flow functions f and f'', respectively, in the network N. Flow function $f'' - f$ is a legal flow function in the residual network of f, and it can be decomposed into a collection of augmenting cycles and simple augmenting paths from s to t. Therefore, every augmenting path in $f'' - f$ can only increase the flow on the edges going from the advertisers to the sink t. Let f' be the flow resulting from f by the addition of all augmentation paths in $f'' - f$, and let S' be the corresponding assignment. Inevitably, S' is optimal, and has at least as many impressions assigned to each advertiser as in S. □

Applying Lemma 4 to the solution produced by $GREEDY_D$, we can assume that for each advertiser a_i, $OPT_i(\sigma) \geq y_i$.

Observation 2 *If $i < j$, then $y_j \leq y_i$. That is, the number of impressions allocated by $GREEDY_D$ to the advertisers is non-increasing.*

Proof If $y_i = d_i$ the observation holds, since $y_j \leq d_j \leq d_i$. Else, each impression $GREEDY_D$ allocates to a_j must belong to a user that already has an impression allocated to a_i. □

We now extend Lemma 1 to the general demands setting.

Lemma 5 *For every advertiser a_i, if $k_i = 0$, then* $\sum_{j=1}^{i} \left[OPT_j(\sigma) - y_j\right] \leq y_i \cdot \sum_{j=1}^{i} k_j$.

Proof Every impression that $GREEDY_D$ fails to allocate to one of the advertisers $a_1, \ldots, a_i$ must belong to a user already having an impression allocated to a_i (since $k_i = 0$). Therefore, there are at most y_i users who have impressions that are not allocated to advertisers $a_1, \ldots, a_i$. Each such user can have at most $\sum_{j=1}^{i} k_j$ more impressions allocated to advertisers $a_1, \ldots, a_i$ by OPT than by $GREEDY_D$, completing the proof. □

We now order the impressions allocated by $OPT(\sigma)$ to the advertisers. An impression x assigned to advertiser $a'(x)$ (by $OPT(\sigma)$) precedes an impression y assigned to advertiser $a'(y)$ (again, by $OPT(\sigma)$) if $a'(x)$ appears before $a'(y)$ in the order define by $GREEDY_D$. If x and y are assigned to the same advertiser, the order between them is arbitrarily chosen. Let A be the set of impressions allocated by OPT. Define $B \subseteq A$ as the following subset of impressions. For each advertiser a_i, all impressions OPT allocates to a_i belong to B, except for the first y_i impressions (if $OPT_i(\sigma) = y_i$, then none of the impressions OPT allocates to a_i belongs to B). See Fig. 2 for a graphical representation of the relationship between A, B, $OPT(\sigma)$ and $GREEDY_D(\sigma)$. For any impression $x \in B$, let $B_{\leq x}$ denote the set of impressions in B which appear before x in the above order, including x.

Observation 3 *If impression $x \in B$ is allocated to advertiser a_i by $OPT(\sigma)$, then $k_i = 0$.*

Proof Otherwise, $y_i = d_i$, and no impression allocated by OPT to a_i belongs to B. □

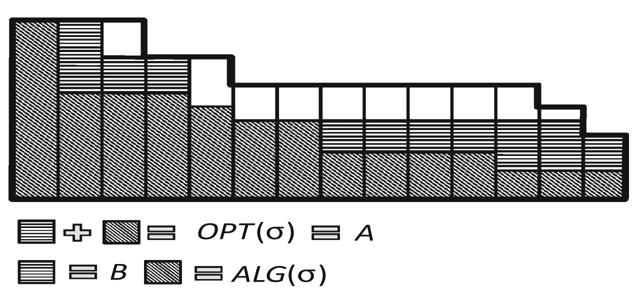

Fig. 2 An assignment constructed by $GREEDY_D$. The set A contains the impressions assigned by $OPT(\sigma)$. The set B contains $OPT_i(\sigma) - y_i$ impressions of A from every advertiser a_i (recall that we assumed $OPT_i(\sigma) \geq y_i$). Thus, B and $GREEDY_V(\sigma)$ add up to $OPT(\sigma)$

Observation 4 $OPT(\sigma) = GREEDY_D(\sigma) + |B|$.

We now define, for analysis purposes, two types of payments made by the impressions allocated by $GREEDY_D$ to the impressions of B. Consider an impression $x \in B$ allocated to advertiser a_i by OPT. Then, the impressions allocated to a_i by $GREEDY_D$ contribute $p_x = y_i/(d_i - y_i)$ to x (recall that $d_i - y_i$ is not zero since $x \in B$, and therefore $k_i = 0$). Let h be the minimal integer such that: $y_i \cdot \sum_{j=1}^{h} k_j \geq |B_{\leq x}|$.[5] Then, the impressions allocated to a_h by $GREEDY_D$ contribute $p'_x = d_h/y_i$ to x.

Lemma 6 *For each impression $x \in B$ allocated by OPT to a_i, The payment p'_x is relieved from impressions that are allocated by $GREEDY_D$ to an advertiser a_h with $h < i$ and $k_h = 1$.*

Proof Since $x \in B$ is assigned to a_i, it must hold that:

$$|B_{\leq x}| \leq \sum_{j=1}^{i} \left[OPT_j(\sigma) - y_j\right].$$

Also, by Lemma 5,

$$\sum_{j=1}^{i} \left[OPT_j(\sigma) - y_j(\sigma)\right] \leq y_i \cdot \sum_{j=1}^{i} k_j.$$

Combining both inequalities, we get $|B_{\leq x}| \leq y_i \cdot \sum_{j=1}^{i} k_j$, and can conclude $h \leq i$. Notice that $k_h = 1$, otherwise h could be replaced by $h - 1$ without changing $y_i \cdot \sum_{j=1}^{h} k_j$. On the other hand, $k_i = 0$ since $x \in B$, and therefore, $h \neq i$. □

Lemma 7 *For each impression $x \in B$, $p_x + p'_x \geq 3$.*

Proof Suppose OPT allocates x to advertiser a_i. Let a_h be the advertiser whose impressions contribute to p'_x. Clearly, $p_x + p'_x = y_i/(d_i - y_i) + d_h/y_i$. By Lemma 6, $h < i$, and therefore, $d_h \geq d_i$. Hence,

$$p_x + p'_x = \frac{y_i}{d_i - y_i} + \frac{d_h}{y_i} \geq \frac{y_i}{d_i - y_i} + \frac{d_i}{y_i} \geq 3, \tag{2}$$

where the last inequality follows from the same arguments used in the proof of Lemma 3. □

For an advertiser a_h, we use the term the *total contribution* made by the impressions allocated to a_h to denote the total payment that is associated with the impressions that were allocated to it by $GREEDY_D$.

Lemma 8 *Let a_h be an advertiser with $k_h = 0$, then the total contribution made by the impressions allocated to a_h (by $GREEDY_D$) is at most y_h.*

Proof By Lemma 6, the impressions allocated to a_h contribute, for any impression x, only to p_x (and not to p'_x). Each contribution is $y_h/(d_h - y_h)$. A contribution is made only to impressions allocated by OPT to a_h which are in B, and the number of such impressions is $OPT_h(\sigma) - y_h \leq d_h - y_h$. □

Lemma 9 *Let a_h be an advertiser with $k_h = 1$, then the total contribution made by the impressions allocated to a_h (by $GREEDY_D$) is at most $y_h = d_h$.*

Proof Let $x \in B$ be the last impression whose p'_x is contributed by the impressions allocated to a_h by $GREEDY_D$, and let a_i be the advertiser to which OPT allocated x. Consider any impression $z \in B$ such that $|B_{\leq z}| \leq |B_{\leq x}| - y_i$. Necessarily, z is allocated to an advertiser $a_{i'}$ with $i' \leq i$, hence, $y_i \leq y_{i'}$. Implying:

$$\begin{aligned}|B_{\leq z}| &\leq |B_{\leq x}| - y_i \leq y_i \left(\sum_{j=1}^{h} k_j\right) - y_i \\ &= y_i \cdot \sum_{j=1}^{h-1} k_j \leq y_{i'} \cdot \sum_{j=1}^{h-1} k_j.\end{aligned}$$

Therefore, by definition, z gets its p'_z contribution from the impressions allocated to advertiser a_{h-1} or advertisers preceding it. By the choice of x and z this implies that there are at most y_i impressions in B which get their p'_x contribution from the impressions allocated to a_h by $GREEDY_D$. The total contribution the impressions of a_h contribute is, therefore, at most $y_i \cdot (d_h/y_i) = d_h$. □

Corollary 2 $GREEDY_D(\sigma) \geq 3|B|$.

Proof By Lemmata 8 and 9, the total contribution of all the impressions assigned to some advertiser by $GREEDY_D$ is at most $GREEDY_D(\sigma)$. On the other hand, by Lemma 7, the contribution for each impression of B is at least 3. □

The proof of Theorem 5 is now immediate:

$$\begin{aligned}\frac{GREEDY_D(\sigma)}{OPT(\sigma)} &\geq \frac{GREEDY_D(\sigma)}{GREEDY_D(\sigma) + |B|} \\ &\geq \frac{3|B|}{3|B| + |B|} = \frac{3}{4}.\end{aligned}$$

4 Equal demands, arbitrary valuations

In this section, we assume advertisers have different values, but the same integer ratio of demand to frequency cap (this can happen, for example, when each advertiser has frequency cap f_i and wants to advertise to the same number of distinct users u, so that $d_i = f_i u$). The reduction to unit frequency cap makes this equivalent to the assumption that demands are equal ($d_i = d$) and all frequency caps are 1.

[5] Following Lemma 5, every advertiser with an exhausted demand can be "blamed" for at most y_i impressions in B. The definition of h attempts to isolate the advertiser a_h which is to be blamed for the fact that $x \in B$.

The following theorem shows that the natural greedy algorithm $GREEDY_V$, which assigns in decreasing order of value (breaking ties arbitrarily), has a competitive ratio of $3/4$. Note that by Theorem 3, this ratio is optimal.

Theorem 6 *Let σ be a sequence of input impressions. Then, under the above assumptions,*

$$\frac{GREEDY_V(\sigma)}{OPT(\sigma)} \geq 3/4.$$

The rest of this section is devoted for proving Theorem 6. Algorithm $GREEDY_V$ sorts the advertisers in a non-decreasing value order, $a_1, a_2, \ldots, a_n$. Let y_i denote the number of impressions $GREEDY_V$ allocates to advertiser a_i. Let k denote the number of advertisers whose demand was exhausted (i.e., $k = |\{i | y_i = d\}|$). Let OPT_i denote the number of impressions allocated by OPT to a_i and let $v(Z)$ denote the total value that a set z of impressions contributes to $OPT(\sigma)$.

Observation 5 *If $i < j$, then $y_j \leq y_i$. That is, the number of impressions allocated by $GREEDY_V$ to the advertisers is non-increasing.*

Proof Each impression $GREEDY_V$ allocates to a_j must belong to a user that already has an impression allocated to a_i. Therefore, $y_j \leq y_i$. □

Corollary 3 *The k advertisers whose demand was fulfilled by $GREEDY_V$ are the first k advertisers.*

Lemma 10 *For every advertiser a_i with $i > k$, $\sum_{j=1}^{i} OPT_j \leq y_i k + \sum_{j=1}^{i} y_j$.*

Proof Each impression that $GREEDY_V$ fails to allocate to one of the advertisers $a_1, \ldots, a_i$ must belong to a user which already has an impression allocated to a_i by $GREEDY_V$ (because $i > k$, and therefore, there is still demand left for a_i). Therefore, there are at most y_i users who have impressions that are not allocated to the advertisers $a_1, \ldots, a_i$. Each such user can have at most k more impressions allocated to advertisers $a_1, \ldots, a_i$ by OPT, than by $GREEDY_V$. □

Let A be the set of impressions allocated by OPT (i.e., $v(A) = OPT(\sigma)$). We order the impressions of A according to the order of the advertisers defined by $GREEDY_V$, breaking ties arbitrarily (i.e., the impressions of A that OPT allocates to a_1 appear first in the order, then the impressions of A that OPT allocates to a_2 and so forth.). Define B as the following set of impressions. For each advertiser a_i, all impressions OPT allocates to a_i belong to B, except for the first y_i impressions. (If $OPT_i(\sigma) \leq y_i$, then none of the impressions OPT allocates to a_i belongs to B.)

For every set $Z \subseteq A$, let $Z_{<x}$ denote the set of impressions in z which appear before x in the above order, excluding x, and let $Z_{\leq x}$ denote the same set including x if $x \in Z$. Notice that if $x \notin Z$ then $Z_{<x} = Z_{\leq x}$. Similarly, let us define $Z_{>x}$ and $Z_{\geq x}$.

We define $C \subseteq B$ as following. An impression $x \in B$ allocated by OPT to advertiser a_i is in C if it satisfies the following inequality:

$$\sum_{j=1}^{i} y_j < |(A - B)_{<x}| + |B_{\leq x}| - |C_{<x}|.$$

Although this definition uses the set C, it is not cyclic because the membership of each impression x in C depends only on the membership in C of impressions that appear before it in the above order. The intuition behind the set C is that it contains the impressions that give the "real" difference between the solutions of OPT and $GREEDY_V$, i.e., C is a subset of B of size equal to the difference between the number of impressions assigned by OPT and $GREEDY_V$. The set C was not necessary in the proof of Theorem 5 because there we could assume $y_i \leq OPT_i$, which would have implied $B = C$ given the above definition of C.

Observation 6 *If impression $x \in C$ is allocated by OPT to a_i, then $i > k$ (otherwise $y_k = d$, and no impression allocated by OPT to a_i could have been in B).*

Let x_i be the last impression of A that is allocated by OPT to one of the advertisers $a_1, \ldots, a_i$. x_i may not exist, but then no impressions arrived, and any algorithm is optimal; so it is safe to assume x_i exists.

Lemma 11 *For $1 \leq i \leq n$, $\sum_{j=1}^{i} \left(OPT_j - y_j\right) \leq |C_{\leq x_i}|$.*

Proof Let z_i be the last impression of $(B - C)_{\leq x_i}$. If there is no z_i then $B_{\leq x_i} = C_{\leq x_i}$, and the proof is complete since by definition $\sum_{j=1}^{i} \left(OPT_j - y_j\right) \leq |B_{\leq x_i}|$. Therefore, we can assume that z_i exists. Let a_h be the advertiser that z_i was assigned to (by OPT). Since $z_i \notin C$,

$$\sum_{j=1}^{h} y_j \geq |(A - B)_{<z_i}| + |B_{\leq z_i}| - |C_{<z_i}|,$$

and therefore,

$$\begin{aligned}\sum_{j=1}^{i} OPT_j &\leq \sum_{j=1}^{h} y_j + \sum_{j=1}^{i} OPT_j \\ &\quad - |(A - B)_{<z_i}| - |B_{\leq z_i}| + |C_{<z_i}| \\ &= \sum_{j=1}^{h} y_j + |A_{>z_i} \cap A_{\leq x_i}| + |C_{<z_i}|.\end{aligned}$$

Notice that by the definition of B we have: $|(A - B)_{>z_i} \cap (A - B)_{\leq x_i}| \leq \sum_{j=h+1}^{i} y_j$. In addition, due to the way z_i was chosen, each impression of $A_{>z_i} \cap A_{\leq x_i}$ is either in C, or not in B,therefore:

$$|A_{>z_i} \cap A_{\le x_i}| = |(A-B)_{>z_i} \cap (A-B)_{\le x_i}| + |C_{>z_i} \cap C_{\le x_i}| \le \sum_{j=k+1}^{i} y_j + |C_{\le x_i}| - |C_{\le z_i}|.$$

Plugging this into the previous inequality gives:

$$\sum_{j=1}^{i} OPT_j \le \sum_{j=1}^{h} y_j + \left(\sum_{j=h+1}^{i} y_j + |C_{\le x_i}| - |C_{\le z_i}| \right) + |C_{<z_i}| = \sum_{j=1}^{i} y_j + |C_{\le x_i}|.$$

□

Lemma 12 $OPT(\sigma) \le GREEDY_V(\sigma) + v(C)$

Proof Let us define $v_{n+1} = 0$ for consistency, then one can easily observe that $OPT = \sum_{i=1}^{n} \left[(v_i - v_{i+1}) \cdot \sum_{j=1}^{i} OPT_j \right]$. Using Lemma 11, we get:

$$\begin{aligned} OPT(\sigma) &\le \sum_{i=1}^{n} \left[(v_i - v_{i+1}) \cdot \left(\sum_{j=1}^{i} y_j + |C_{\le x_i}| \right) \right] \\ &= GREEDY_V(\sigma) + \sum_{i=1}^{n} \left[(v_i - v_{i+1}) \cdot |C_{\le x_i}| \right] \\ &= GREEDY_V(\sigma) + v(C), \end{aligned}$$

where the inequality holds since $v_i - v_{i+1} \ge 0$, and the last equality follows since $|C_{\le x_i}|$ is the number of impressions of C allocated to advertisers $a_1, \ldots, a_i$ by OPT. □

We now define two types of payments that the impressions allocated by $GREEDY_V$ pay to the impressions of C. Consider an impression $x \in C$ allocated to advertiser a_i by OPT. The impressions allocated to a_i by $GREEDY_V$ pay $p_x = v_i y_i/(d - y_i)$ to x (recall that $d - y_i$ is not zero because $x \in C$, and therefore, $i > k$). Let h be the minimal integer such that $hy_i \ge |C_{\le x}|$, then the impressions allocated to a_h by $GREEDY_V$ pay $p'_x = v_h d_h / y_i$ to x.

Lemma 13 *For each impression $x \in C$ allocated to a_i by OPT, p'_x is contributed by the impressions allocated (by $GREEDY_V$) to an advertiser a_h with $h \le k < i$.*

Proof Since $x \in C$:

$$\sum_{j=1}^{i} y_j < |(A-B)_{<x}| + |B_{\le x}| - |C_{<x}|.$$

Firstly notice that $|C_{<x}| = |C_{\le x}| - 1$, and also $\sum_{j=1}^{i} OPT_j \ge |A_{\le x}| = |(A-B)_{<x}| + |B_{\le x}|$. Plugging this into the previous inequality gives:

$$\sum_{j=1}^{i} y_j < \sum_{j=1}^{i} OPT_j - |C_{\le x}| + 1 \Rightarrow |C_{\le x}| \le \sum_{j=1}^{i} OPT_j - \sum_{j=1}^{i} y_j.$$

From Lemma 10, $\sum_{j=1}^{i} OPT_j \le y_i k + \sum_{j=1}^{i} y_j$. Combining the last two inequalities we get $|C_{\le x}| \le y_i k$, and we can immediately conclude $h \le k$. In addition, since $x \in C$, Observation 6 implies $k < i$. □

Lemma 14 *For each impression $x \in C$, $p_x + p'_x \ge 3v_i$.*

Proof Suppose OPT allocates x to advertiser a_i, and let a_h be the advertiser whose impressions contribute to p'_x. Then $p_x + p'_x = v_i y_i/(d - y_i) + v_h d / y_i$. By Lemma 13, $h < i$, and therefore, $v_h \ge v_i$. Hence:

$$p_x + p'_x = \frac{v_i y_i}{d - y_i} + \frac{v_h d}{y_i} \ge v_i \cdot \left(\frac{y_i}{d - y_i} + \frac{d}{y_i} \right) \ge 3v_i.$$

Where the last inequality follows from the same arguments used in the proof of Lemma 3. □

Lemma 15 *Let a_h be an advertiser with $h > k$, then the total contribution made by the impressions allocated to a_h (by $GREEDY_V$) is at most $v_h y_h$.*

Proof By Lemma 13, the impressions allocated to a_h contribute, for any impression x, only to p_x (and not to p'_x). Each contribution is $v_h y_h/(d_h - y_h)$. A contribution is made only to impressions allocated by OPT to a_h which are in $C \subseteq B$, and the number of such impressions cannot exceed $d_h - y_h$. □

Lemma 16 *Let a_h be an advertiser with $h \le k$, then the total contribution made by the impressions allocated to a_h (by $GREEDY_V$) is at most $v_h y_h = v_h d$.*

Proof Let $x \in C$ be the last impression whose p'_x is contributed by the impressions allocated to a_h, and let a_i be the advertiser x is allocated to by OPT. Consider any impression $z \in C$ such that $|C_{\le z}| \le |C_{\le x}| - y_i$. Necessarily, z is allocated to an advertiser $a_{i'}$ with $i' \le i$, and therefore $y_i \le y_{i'}$. This implies:

$$|C_{\le z}| \le |C_{\le x}| - y_i \le y_i h - y_i = y_i(h-1) \le y_{i'}(h-1).$$

Therefore, by definition, z gets its p'_z payment from the impressions allocated to advertiser a_{h-1} or an advertiser before it. By the choice of x and z this implies that there are at most y_i impressions in C whose second payment is contributed by the impressions allocated to a_h by $GREEDY_V$. The total contribution of the impressions of a_h is, therefore, at most: $y_i \cdot (v_h d / y_i) = v_h d$. □

Corollary 4 $GREEDY_V(\sigma) \ge 3v(C)$

Proof By Lemmata 15 and 16, the total contribution of all the impressions assigned by $GREEDY_V$ is no greater than $GREEDY_V(\sigma)$. On the other hand, by Lemma 14, the payment received by each impression $x \in C$ is at least $3v(x)$. □

The proof of Theorem 6 is now immediate:

$$\frac{GREEDY_V(\sigma)}{OPT(\sigma)} \geq \frac{GREEDY_V(\sigma)}{GREEDY_V(\sigma)+v(C)} \geq \frac{3v(C)}{3v(C)+v(C)} = \frac{3}{4}.$$

5 Arbitrary valuations

We now consider arbitrary valuations v_i. We first prove an improved upper bound for this case.

Theorem 7 *No deterministic algorithm has a competitive ratio better than* $1/\sqrt{2} \approx 0.707$.

Proof Consider any deterministic algorithm ALG and the following instance: There are two advertisers: a_1 with $v_1 = 1$, $d_1 = 1$ and a_2 with $v_2 = 1/\sqrt{2}$, $d_2 = 2$. At step one, a user u_1 comes. If ALG assigns to a_2, the process stops. The optimal algorithm will assign to a_1. The ratio of the algorithm is $1/\sqrt{2}$. If ALG assigns to a_1 at step one, a new user u_2 comes with two impressions. ALG can only assign one of them to a_2, giving a total revenue of $1+1/\sqrt{2}$. However, the optimal algorithm assigns the three impressions to a_2, a_1, a_2, respectively, earning a total revenue of $1+2/\sqrt{2}$. Hence, the ratio of the algorithm is $\frac{1+1/\sqrt{2}}{1+2/\sqrt{2}} = 1/\sqrt{2}$. □

5.1 A primal-dual algorithm

In order to apply the primal-dual approach to the problem, we first formulate the *offline* allocation problem as a linear program as follows. We refer to the allocation problem as the *dual* problem, since it is a maximization problem. Let A be the set of advertisers. Let B be the set of users. Finally, for each user $j \in B$, let $K(j)$ be the number of impressions of user j. We define variables $y(i, j, k)$ indicating that the k-th impression of user j is assigned to advertiser a_i.

$$\begin{aligned}
\max \; & \sum_{a_i \in A} v_i \sum_{j \in B} \sum_{k=1}^{K(j)} y(i,j,k) \quad (D) \\
\text{s.t.} \; & \sum_{j \in B} \sum_{k=1}^{K(j)} y(i,j,k) \leq d_i \quad && \forall a_i \in A \\
& \sum_{k=1}^{K(j)} y(i,j,k) \leq f_i && \forall a_i \in A, j \in B \\
& \sum_{a_i \in A} y(i,j,k) \leq 1 && \forall j \in B, k \in \{1,2\ldots,K(j)\} \\
& y(i,j,k) \geq 0
\end{aligned}$$

The first set of constraints guarantees that at most d_i impressions are assigned to advertiser a_i. The second set of constraints guarantees the frequency cap of each advertiser. Finally, the last set of constraints guarantees that each impression is assigned only once. For consistency with previous work (Buchbinder and Naor 2009), we refer to the maximization problem as the dual problem. We now define the primal problem. We have variable $x(i)$ for each advertiser a_i, a variable $w(i, j)$ for each pair of advertiser a_i and user j and variable $z(j, k)$ for the k-th impression of user j.

$$\begin{aligned}
\min \; & \sum_{a_i \in A} d_i x(i) + \sum_{a_i \in A, j \in B} f_i w(i,j) + \sum_{j \in B, k} z(j,k) \quad (P) \\
\text{s.t.} \; & x(i) + w(i,j) + z(j,k) \geq v_i \forall a_i \in A, j \in B, k \\
& x, w, z \qquad\qquad\qquad \geq 0
\end{aligned}$$

The allocation algorithm is as follows. We assume that the reduction to the case where the frequency cap of each advertiser is 1 has already been applied.

Allocation Algorithm: Upon arrival of impression k of user j:

- Let $S(j)$ be the set of advertisers not yet assigned impressions of user j, and let $\overline{S(j)} = A \setminus S(j)$.
- Let $m_1 \in S(j)$ be the advertiser maximizing $v_i - x(i)$. Let $m_2 \in S(j) \setminus m_1$ be the advertiser maximizing $v_i - x(i)$.[a]

1. Assign impression k to advertiser m_1.
2. For each advertiser $i \in \overline{S(j)} \cup m_1$ set: $w(i, j) \leftarrow \max\{0, (v_i - x(i)) - \left(v_{m_2} - x(m_2)\right)\}$.
3. For each advertiser $i \in S(j) \setminus m_1$ set: $w(i, j) \leftarrow 0$.
4. For each impression $\ell \leq k$ of user j set: $z(j, \ell) \leftarrow v_{m_2} - x(m_2)$.
5. For advertiser m_1: $x(m_1) \leftarrow x(m_1)\left(1+\frac{1}{d_i}\right) + \frac{v_{m_1}}{c \cdot d_i}$ (where c is a constant to be determined later).

[a] If $\max_{S(j)}(v_i - x(i)) \leq 0$, or $S(j) = \emptyset$, no assignment is made and no variables are updated. If there is no m_2, we view $v_{m_2} - x(m_2)$ as equal to 0.

Notice that this algorithm differs from the standard online primal-dual approach because it both increases and decreases primal variables.

Theorem 8 *The algorithm is* $(1-(c+1)^{-1})$*-competitive, for* $c = \left(1+\frac{1}{d_{\min}}\right)^{d_{\min}} - 1$, *where* $d_{\min}$ *is the minimum demand of any advertiser.*

Proof The assignment of impressions to advertisers defines a solution to the dual program. The algorithm generates a primal solution online.

We start with a few useful observations. Note first that the variables in the primal solution are always non-negative. Next, observe that for each advertiser i, $x(i)$ is monotonically increasing during the execution of the algorithm, and $v_i - x(i)$ is monotonically decreasing. Also, observe that if at some point of time in which an impression of user j arrives, $v_{m_2} - x(m_2) = v'$, then in the next time in which an impression of user j arrives, $v_{m_2} - x(m_2) \le v'$, i.e., $v_{m_2} - x(m_2)$ is also monotonically decreasing. This follows since m_2 maximizes $v_i - x(i)$ among advertisers i that have not yet been assigned an impression of user j (i.e., advertisers in $S(j)$), $v_i - x(i)$ is monotonically decreasing and the set $S(j)$ can only shrink over time.

We now prove the following claims from which the theorem readily follows.

1. The primal solution produced by the algorithm is feasible.
2. In each round (arrival of an impression) the change in the dual profit is at least $1 - (c+1)^{-1}$ times the change in the primal cost.
3. The dual solution produced is feasible.

Proof of (1) We prove feasibility by induction on the steps of the algorithm. Initially, there are no primal constraints and the primal solution is trivially feasible. Consider the step in which impression k of user j arrives. The algorithm does not change the value of variables $w(i, j')$ and $z(j', k)$ corresponding to any user $j' \neq j$. Thus, since $x(i)$ only increases, all the constraints corresponding to users $j' \neq j$ and impression ℓ remain feasible, i.e., $x(i) + w(i, j') + z(j', \ell) \ge v_i$.

Next, consider the primal constraint corresponding to user j, advertiser i, and impression $\ell \le k$: $x(i) + w(i, j) + z(j, \ell) \ge v_i$. If advertiser $i \in \overline{S(j)} \cup m_1$, i.e., i has already been assigned an impression of user j, then:

$$\begin{aligned} & w(i, j) + z(j, \ell) && (3)\\ &= \max\{0, (v_i - x(i)) - (v_{m_2} - x(m_2))\} + v_{m_2} - x(m_2) \\ &\ge v_i - x(i)\ . \end{aligned}$$

and the constraint is satisfied. If advertiser $i \in S(j) \setminus m_1$, i.e., i has not been assigned an impression of user j, then:

$$w(i, j) + z(j, \ell) = 0 + v_{m_2} - x(m_2) \ge v_i - x(i). \quad (4)$$

The last inequality follows since m_2 is the advertiser that maximizes the quantity $v_i - x(i)$ in the set $S(j) \setminus m_1$, yielding that the constraint is satisfied.

Proof of (2) We analyze the change in the primal cost and the dual profit in the current round. Consider first the contribution to the primal cost due to the change in the w-variables and z-variables, excluding advertiser m_1 and impression (j, k):

$$\sum_{i \in S(j)} w(i, j) + \sum_{\ell < k} z(j, \ell). \quad (5)$$

The first sum has $k-1$ terms corresponding to advertisers that were already assigned impressions belonging to user j, and the second sum has also $k-1$ terms corresponding to the first $k-1$ impressions of user j. We now match each advertiser $i \in S(j)$ with the variable $z(j, \ell)$ corresponding to the impression ℓ assigned to i. The new values of the variables satisfy:

$$w(i, j) + z(j, \ell) = \max\{v_i - x(i), v_{m_2} - x(m_2)\}. \quad (6)$$

Since both terms above can only decrease during the execution of the algorithm, the contribution to the primal cost of advertiser $i \in S(j)$ and impression $\ell < k$ in the current round is not higher than the contribution to the primal cost of those advertisers and impressions in the previous round.

The only increase in the primal cost is, thus, due to the increase in the variables $w(m_1, j)$, $z(j, k)$ and $x(m_1)$. Their contribution is:

$$\begin{aligned} & w(m_1, j) + z(j, k) + d_{m_1} \cdot \Delta x(m_1) && (7)\\ &= v_{m_1} - x(m_1) + x(m_1) + \frac{v_{m_1}}{c} \\ &= v_{m_1}(1 + 1/c)\ . \end{aligned}$$

The dual profit in this round is v_{m_1}, so the ratio between the dual profit and the primal cost is $(1 + 1/c)^{-1} = c/(c+1) = 1 - (c+1)^{-1}$.

Proof of (3) The algorithm never assigns more than one impression of user j to advertiser i. We only need to make sure that the algorithm does not assign a total of more than d_i impressions to advertiser i. It is easy to verify that after assigning ℓ impressions to advertiser i:

$$x(i) = \frac{v_i}{c} \cdot \left(\left(1 + \frac{1}{d_i}\right)^{\ell} - 1 \right). \quad (8)$$

Thus, after d_i rounds $x(i) \ge v_i$ if $c = (1 + \frac{1}{d_{\min}})^{d_{\min}} - 1$. At this point all primal constraints corresponding to advertiser i are satisfied and the algorithm is guaranteed not to assign any further impressions to advertiser i. □

Targeting constraints We assumed thus far that advertisers valued all users equally. In practice, however, when buying display ad space, advertisers can provide *targeting* information, specifying which subset of impressions is acceptable. That is, advertisers have value v_i for acceptable impressions that meet the targeting constraints and value of 0 for others (contracts for display ads typically specify a single price-per-impression that does not vary across the set of acceptable impressions, i.e., v_i does not take on different non-zero values).

Suppose targeting information is user-dependent only, i.e., an advertiser may value only a subset of users with certain characteristics (age, gender, location, etc.), but does not distinguish between different impressions (e.g., when visiting different webpages) from the same user. In this case, advertiser values have the following form: $v(i, j)$ is either v_i or

0 (i.e., a_i finds users with $v(i, j) = v_i$ *acceptable*, and the rest unacceptable). We observe that the above algorithm also works for this more general setting. The only change is that the sets $S(j)$ and $\overline{S(j)}$ include only advertisers that accept user j. This implies the following.

Theorem 9 *For* $c = \left(1 + \frac{1}{d_{\min}}\right)^{d_{\min}} - 1$, *the algorithm is* $(1 - (c + 1)^{-1})$*-competitive also when* $v(i, j) \in \{0, v_i\}$ *for all* i, j.

Proof Sketch In the dual objective function, advertisers do not sums up their contribution over all users, but rather only over their acceptable users. This corresponds in the primal program to constraints that only contain an advertiser i and an acceptable user j. The same analysis shows now that all primal constraints remain feasible during the execution of the algorithm.

The relation between the change in primal cost and dual profit also follows by noting that if impression k of user j is assigned, then *every* previous impression $l \leq k$ is also assigned. So we can find an advertiser i to match every $l \leq k$ as before. □

Theorem 10 *In the presence of targeting constraints, no deterministic algorithm has a competitive ratio higher than* $1 - 1/e$*, even when demands are large.*

Proof Online b-matching, which generalizes online matching, is defined as follows: requests arrive online; a request can be matched to any server to which it has an edge, but no server can be used more than b times. An instance of the b-matching problem can be mapped to an instance of the frequency capping problem with targeting constraints: advertisers correspond to servers, and each request corresponds to a new user j; $j \in S_i$ for the set of servers to which the request has edges. All advertisers have demand $d = b$, $f_i = 1$ and $v_i = 1$. This reduction implies an upper bound of $1 - \frac{1}{\left(1+\frac{1}{d}\right)^d}$, which follows from the upper bound on the online b-matching problem given by Kalyanasundaram and Pruhs (2000). For $d \gg f_i = 1$, this upper bound tends to $1 - 1/e$. □

6 Further directions

The frequency capping problem is an important practical problem which imposes interesting algorithmic challenges. Here are two main directions for further work.

- *Improving 1-1/e for arbitrary valuations* There is a gap between the best upper bound of $1/\sqrt{2}$ and the best algorithm $(1-1/e)$ for the case of arbitrary valuations without targeting constraints, discussed in Sect. 5. The targeting constraints are to be blamed for the "matching" aspects, leading to the upper bound of $1 - 1/e$ in Theorem 10. By removing these constraints, the difference between our problem and online matching resurfaces, and the upper bound of $1 - 1/e$ does not hold anymore. We believe that our primal-dual algorithm is an excellent starting point for a future online algorithm for frequency capping with arbitrary values that will go beyond $1 - 1/e$.
- *Content-based targeting specifications* Targeting specifications may be not only user-based, but also depend on the webpage's content. For instance, an advertiser might want to display her ads only to males (user targeting) when they browse a sports related webpage (content targeting); targeting constraints are often of this form. So, advertisers now have valuations of the form $v(i, j, k) \in \{0, v_i\}$, i.e., the value of the k-th impression of the j-th user to advertiser i is either v_i or 0 depending on what page the user was surfing on his k-th impression. Note that the model of Goel and Mehta (2007) does not capture this problem, which entangles a matching aspect with frequency capping. The questions of designing a good online algorithm and finding the smallest upper bound (of course, $1 - 1/e$ is a trivial upper bound since this problem generalizes arbitrary valuations with targeting) are both open.

Acknowledgments We are extremely grateful to Ning Chen for several helpful discussions, and for first suggesting the total demand algorithm. Research supported in part by ISF Grant 954/11, and by BSF Grant 2010426. Research supported in part by the Google Inter-University Center for Electronic Markets, by ISF Grant 954/11, and by BSF Grant 2010426.

References

Abrams, Z., & Vee, E. (2007). Personalized ad delivery when ads fatigue: An approximation algorithm. In *WINE '07: Workshop on Internet and Network, Economics* (pp. 535–540).

Alon, N., Awerbuch, B., Azar, Y., Buchbinder, N., & Naor, J. (2006). A general approach to online network optimization problems. *ACM Transactions on Algorithms*, *2*(4), 640–660.

Alon, N., Awerbuch, B., Azar, Y., Buchbinder, N., & Naor, J. (2009). The online set cover problem. *SIAM Journal on Computing*, *39*(2), 361–370.

Bansal, N., Chen, N., Cherniavsky, N., Rudra, A., Schieber, B., & Sviridenko, M. (2010). Dynamic pricing for impatient bidders. *ACM Transactions on Algorithms*, *6*(2), 726–735.

Bansal, N., Buchbinder, N., & Naor, J. (2012a). A primal-dual randomized algorithm for weighted paging. *Journal of the ACM*, *59*(4), 19.

Bansal, Nl, Buchbinder, N., & Naor, J. (2012b). Randomized competitive algorithms for generalized caching. *SIAM Journal of Computing*, *41*(2), 391–414.

Blum, A., & Hartline, J. (2005). Near-optimal online auctions. In *SODA '05: Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms* (pp. 1156–1163).

Blum, A., Kumar, V., & Rudra, A. (2004). Online learning in online auctions. *Theoretical Computer Sciences*, *324*(2–3), 137–146.

Buchbinder, N., & Naor, J. (2006). Improved bounds for online routing and packing via a primal-dual approach. In *FOCS '06: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer, Science* (pp. 293–304).

Buchbinder, N., & Naor, J. (2009). Online primal-dual algorithms for covering and packing problems. *Mathematics of Operations Research*, *34*(2), 270–286.

Buchbinder, N., Jain, K., & Naor, J. (2007). Online primal-dual algorithms for maximizing ad-auctions revenue. In *ESA '07: Proceedings of the 15th Annual European Symposium* (pp. 253–264).

Feldman, J., Korula, N., Mirrokni, V., Muthukrishnan, S., & Pál, M. (2009a). Online ad assignment with free disposal. In *WINE '09: Proceedings of the 5th International Workshop on Internet and Network, Economics* (pp. 374–385).

Feldman, J., Mehta, A., Mirrokni, V. S., & Muthukrishnan, S. (2009b). Online stochastic matching: Beating 1–1/*e*. In *50th Annual IEEE Symposium on Foundations of Computer Science*.

Goel, G., & Mehta, A. (2007). Adwords auctions with decreasing valuation bids. In *WINE '07: Workshop on Internet and Network, Economics* (pp. 335–340).

Kalyanasundaram, B., & Pruhs, K. R. (2000). An optimal deterministic algorithm for online *b*-matching. *Theoretical Computer Science*, *233*(1–2), 319–325.

Karp, R. M., Vazirani, U. V., & Vazirani, V. V. (1990). An optimal algorithm for on-line bipartite matching. In *STOC '90: Proceedings of the 22nd Annual ACM Symposium on Theory of Computing* (pp. 352–358).

Mahdian, M., & Saberi, A. (2006). Multi-unit auctions with unknown supply. In *EC '06: Proceedings of the 7th ACM Conference on Electronic Commerce* (pp. 243–249).

Mehta, A., Saberi, A., Vazirani, U., & Vazirani, V. (2007). Adwords and generalized online matching. *Journal of the ACM*, *54*(5), 22.