Oblivious Medians Via Online Bidding
(Extended Abstract)

Marek Chrobak[1,⋆], Claire Kenyon[2], John Noga[3], and Neal E. Young[1]

[1] Department of Computer Science, University of California, Riverside
[2] Computer Science Department, Brown University
[3] Department of Computer Science, California State University, Northridge

Abstract. Following Mettu and Plaxton [22, 21], we study oblivious algorithms for the k-medians problem. Such an algorithm produces an incremental sequence of facility sets. We give improved algorithms, including a $(24 + \epsilon)$-competitive deterministic polynomial algorithm and a $2e \approx 5.44$-competitive randomized non-polynomial algorithm. Our approach is similar to that of [18], which was done independently.

We then consider the competitive ratio with respect to *size*. An algorithm is *s-size-competitive* if, for each k, the cost of F_k is at most the minimum cost of any set of k facilities, while the size of F_k is at most sk. We present optimally competitive algorithms for this problem.

Our proofs reduce oblivious medians to the following *online bidding* problem: faced with some unknown threshold $T \in \mathbb{R}^+$, an algorithm must submit "bids" $b \in \mathbb{R}^+$ until it submits a bid $b \geq T$, paying the sum of its bids. We describe optimally competitive algorithms for online bidding.

Some of these results extend to approximately metric distance functions, oblivious fractional medians, and oblivious bicriteria approximation.

When the number of medians takes only two possible values k or l, for $k < l$, we show that the optimal cost-competitive ratio is $2 - 1/l$.

1 Introduction and Summary of Results

An instance of the k-median problem is specified by a finite set $\mathcal{C}$ of customers, a finite set $\mathcal{F}$ of facilities, and, for each customer u and facility f, a distance $d_{uf} \geq 0$ from u to f representing the cost of serving u from f. The cost of a set of facilities $X \subseteq \mathcal{F}$ is $cost(X) = \sum_{u \in \mathcal{C}} d_{uX}$, where $d_{uX} = \min_{f \in X} d_{uf}$. For a given k, the *(offline) k-median problem* is to compute a *k-median*, that is, a set $X \subseteq \mathcal{F}$ of cardinality k for which $cost(X) = opt_k$ is minimum (among all sets of cardinality k). *Metric* k-median refers to the case where the distance function is metric (the shortest u-to-f path has length d_{uf} for each u and f).

The k-median problem is a well-known NP-hard facility location problem. Substantial work has been done on efficient approximation algorithms that, given k, find a set F_k of k medians of approximately minimum cost [2, 1, 6, 5, 13, 12, 24]. In particular, for the metric version Arya et al. show that, for any $\epsilon > 0$, a set F_k of cost at most $(3 + \epsilon)\,opt_k$ can be found in polynomial time [2].

⋆ Research supported by NSF Grant CCR-0208856.

J.R. Correa, A. Hevia, and M. Kiwi (Eds.): LATIN 2006, LNCS 3887, pp. 311–322, 2006.

problem:	**cost-competitive metric**		**size-competitive**		**bidding**
time:	polynomial	non-polynomial	polynomial	non-polynomial	polynomial
deterministic	$24+\epsilon$	8	$\mathbf{O(\log n)}$	**4**	**4**
randomized	$6e+\epsilon < 16.31$	$2e < 5.44$	$\mathbf{O(\log n)}$	$\mathbf{e} < 2.72$	$\mathbf{e} < 2.72$

Fig. 1. Competitive ratios shown for oblivious medians and online bidding. Ratios in bold are optimal.

Oblivious medians is an online version of the k-median problem where k is not specified in advance [22, 21]. Instead, authorizations for additional facilities arrive over time. A (possibly randomized) *oblivious algorithm* produces a sequence $\bar{F} = (F_1, F_2, \ldots, F_n)$ of facility sets which must satisfy the oblivious constraint $F_1 \subseteq F_2 \subseteq \cdots \subseteq F_n \subseteq \mathcal{F}$. In general, in an oblivious solution, the F_k's cannot all simultaneously have minimum cost. The algorithm is said to be *c-cost-competitive*, or to have *cost-competitive ratio* of c, if it produces a (possibly random) sequence $\bar{F}$ of sets which is c-cost-competitive, that is, such that for each k, the set F_k has size at most k and (expected) cost at most $c \cdot opt_k$. For offline solutions we use the term "approximate" instead of "competitive".

Mettu and Plaxton [22, 21] give a c-cost-competitive linear time oblivious algorithm with $c \approx 30$. Our first contribution is to improve this ratio. The problem is difficult both because (1) the solution must be oblivious, and (2) even the offline problem is NP-hard. To study separately the effects of the two difficulties, we consider both polynomial and non-polynomial algorithms.

Theorem 1. (a) *Oblivious metric medians has non-polynomial deterministic and randomized algorithms that are* 8*-cost-competitive and* $2e$*-cost-competitive, respectively.* **(b)** *If metric k-median has a polynomial c-cost-approximation algorithm, then the oblivious problem has polynomial deterministic and randomized algorithms that are* $8c$*-cost-competitive and* $2ec$*-cost-competitive, respectively.*

As it is known that there is a polynomial $(3+\epsilon)$-cost-approximation algorithm for the offline metric medians [2], Theorem 1 implies the cost-competitive ratios shown in Fig. 1. Theorem 1 was recently and independently discovered by Lin, Nagarajan, Rajaraman and Williamson [18]. For polynomial algorithms, they improved the result further using a Lagrangian-multiplier-preserving approximation algorithm for facility location; they obtained 16-cost-competitive and randomized $4e$-competitive polynomial algorithms for metric medians.

We also consider here oblivious algorithms that are *s-size-competitive*: they are allowed to use extra medians, but must achieve the optimal cost for each k. An algorithm is *s-size-competitive* if it produces a sequence $\bar{F}$ such that each set F_k has cost at most opt_k and size at most sk. (If the algorithm is randomized, it must produce a random sequence such that each set F_k costs at most opt_k and has *expected* size at most sk.)

To our knowledge, size-competitive algorithms for oblivious medians have not been studied, although other online problems have been analyzed in an analogous setting of *resource augmentation* (e.g. [14, 7, 17]). We completely characterize the optimal size-competitive ratios for oblivious medians:

Theorem 2. **(a)** *Oblivious medians has non-polynomial deterministic and randomized oblivious algorithms that are* 4*-size-competitive and* e*-size-competitive, respectively.* **(b)** *No deterministic or randomized oblivious algorithm is less than* 4*-size-competitive or* e*-size-competitive, respectively.* **(c)** *If offline* k*-median has a polynomial* c*-size-competitive algorithm, then the oblivious problem has polynomial deterministic and randomized algorithms that are* $4c$*-size-competitive and* ec*-size-competitive, respectively.*

The upper and lower bounds in Theorem 2 hold for both the metric and non-metric problems. Part (c) on polynomial algorithms is included for completeness, as is the following result for offline k-medians (proof omitted):

Theorem 3. *Offline* k*-medians has a polynomial* $O(\log(n))$*-size-approximation algorithm.*

This improves the best previous result — a bicriteria approximation algorithm that finds a facility set of size $\ln(n+n/\epsilon)k$ and cost $(1+\epsilon)\,opt_k$ [24]. Our algorithm finds a true (not bicriteria) approximate solution: a facility set of size $O(\log k)$ and cost at most opt_k.

Theorems 2 and 3 imply the size-competitive ratios shown in Fig. 1. Note also that no polynomial algorithm (oblivious or offline) is $o(\log n)$-size-competitive unless P=NP, even for the metric case.

To analyze oblivious medians, we reduce the size- and cost-competitive oblivious problems to the following folklore "*online bidding problem*": An algorithm repeatedly submits "bids" $b \in \mathbb{R}^+$, until it submits a bid b that is at least as large as some unknown threshold $T \in \mathbb{R}^+$. Its cost is the total of the submitted bids. The algorithm is β-competitive if, for any $T \in \mathbb{R}^+$, its cost is at most βT (or, if the algorithm is randomized, its expected cost is at most βT). More generally, the algorithm may be given in advance a closed universe $\mathcal{U} \subseteq \mathbb{R}^+$, with a guarantee that the threshold T is in $\mathcal{U}$ and a requirement that all bids be in $\mathcal{U}$.

For $\mathcal{U} = \mathbb{R}^+$, it is known that an optimal deterministic strategy bids increasing powers of 2, and that there is a better randomized strategy which bids (randomly translated) powers of e. We complete this characterization by proving that the randomized strategy is optimal.

Theorem 4. **(a)** *Online bidding has deterministic and randomized algorithms that are* 4*-competitive and* e*-competitive, respectively. Furthermore, if* $\mathcal{U}$ *is finite, the algorithms run in time polynomial in* $|\mathcal{U}|$*.* **(b)** *No deterministic or randomized algorithm is less than* 4*-competitive or* e*-competitive, respectively, even when restricted to instances of the form* $\mathcal{U} = \{1, 2, ..., n\}$ *for some integer* n*.*

Weighted medians. All of our results extend to the weighted version, where we allow the facilities and the customers to have non-negative weights w. In this case, for a facility set X, one constrains the total weight $\sum_{f \in X} w(f)$ to be at most k, and one takes $cost(X) = \sum_{u \in \mathcal{C}} w(u) d_{uX}$.

Approximate triangle inequality. Mettu and Plaxton show that their oblivious median algorithm also works in "λ-approximate" metric spaces, achieving cost-competitive-ratio $O(\lambda^4)$ [22, 21]. We reduce this ratio to $O(\lambda^2)$. We say that the cost function d is a *λ-relaxed metric* if $d_{fy} \le \lambda(d_{fx} + d_{xg} + d_{gy})$ for any $f, g \in \mathcal{F}$ and $x, y \in \mathcal{C}$. (This condition is somewhat less restrictive than the one in [22, 21]. A related concept was studied in [10].) Theorem 1 generalizes as follows (proofs omitted):

Theorem 5. **(a)** *Oblivious λ-relaxed metric medians has (non-polynomial) deterministic and randomized algorithms that are $8\lambda^2$-cost-competitive and $2e\lambda^2$-cost-competitive, respectively.* **(b)** *If offline λ-relaxed metric k-median has a polynomial c-cost-approximation algorithm, then the oblivious problem has deterministic and randomized polynomial algorithms that are $8\lambda^2 c$-cost-competitive and $2e\lambda^2 c$-cost-competitive, respectively.*

The kl-medians problem. A natural question to ask is whether better competitive ratios are possible if the number of medians can take only some limited number of values. As shown in [22, 21], no algorithm can be better than 2-competitive even when there are only two possible numbers of medians, either 1 or k, for some large k. Here, we solve the deterministic kl-median problem (where the number of medians is either k or $l > k$).

Theorem 6. *For any $k < l$, there is a deterministic oblivious algorithm for kl-medians with competitive ratio $2 - 1/l$, and no better ratio is possible.*

Oblivious fractional medians. A *fractional k-median* is a solution to the linear program which is the relaxation of the standard integer program for the k-median problem. The natural *oblivious* version of this fractional problem is to find a $c \ge 1$ and, for every integer $k \in [n]$ simultaneously, a pair $(x_{if}^{(k)}), (y_f^{(k)})$ meeting the constraints of the linear program, as well as $y_f^{(k)} \le y_f^{(k+1)}$ (for all f) and $\sum_u \sum_f x_{uf} d_{uf} \le c \cdot opt_k$ (where opt_k is the minimum cost of any fractional k-median). The goal is to minimize the competitive ratio c.

The proof of the theorem below (omitted) extends the proof of Theorem 1, along with the observation that the randomized algorithm for the fractional problem can be derandomized without increasing the competitive ratio.

Theorem 7. *Oblivious fractional metric medians has a deterministic polynomial algorithm that is $2e$-cost-competitive.*

Bicriteria approximations. Combining Theorem 2, Theorem 8, and offline bicriteria results from [2, 19, 20, 16], we can obtain oblivious, polynomial algorithms with the following bicriteria (c, s)-competitiveness guarantees for oblivious metric medians. The first quantity c is the cost-competitive ratio and the second quantity s is the size-competitive ratio: (a) $(3+\epsilon, 4)$, for any $\epsilon > 0$, (b) $(2+\epsilon, 4(1+2\epsilon^{-1}))$, for any $\epsilon > 0$, (c) $(1+\epsilon, 4(3+5\epsilon^{-1}))$, for any $\epsilon > 0$.

Notation. Throughout we use the following terminology for online bidding. Given the universe $\mathcal{U}$, the algorithm outputs a *bid set* $\mathcal{B} \subseteq \mathcal{U}$. Against a particular threshold T, the algorithm pays for the bids $\{b \in \mathcal{B} : b \leq T^+\}$, where $T^+ = \min\{b \in \mathcal{B} : b \geq T\}$. The bid set $\mathcal{B}$ is *β-competitive* if, for any $T \in \mathcal{U}$, this payment is at most βT. Also, $\mathbb{R}^+$ denotes the set of non-negative reals, $\mathbb{Z}$ the set of integers, and $\mathbb{N}^+$ the set of positive integers. For $n \in \mathbb{N}^+$, let $[n] = \{1, 2, \ldots, n\}$.

Plan of the paper. We prove our upper bounds on competitive algorithms for oblivious medians (Theorem 1 for cost-competitive algorithms and Theorem 2(a) for size-competitive algorithms) by reducing oblivious medians to online bidding (Theorem 8, below) and then proving the upper bounds for online bidding (Theorem 4). We prove our lower bounds on size-competitive algorithms for oblivious medians (Theorem 2(b)) by reducing online bidding to size-competitive medians (Theorem 9, below) and then proving the lower bounds for online bidding in Theorem 4. We prove the reductions in Section 2 and analyze online bidding in Section 3. In Section 4 we prove Theorem 6.

2 Oblivious Medians and Online Bidding

We start by showing that oblivious medians can be reduced to online bidding. We show that (a) $2c\beta$-cost-competitive oblivious metric medians reduces (in polynomial time) to β-competitive online bidding and c-cost-approximate offline medians, and (b) $s\beta$-size-competitive oblivious medians reduces (in polynomial time) to β-competitive online bidding and s-size-approximate offline medians.

Note that part (b) holds even for non-metric medians. Also, if allowing non-polynomial time, one can take F_k^* to be the optimal k-median in Theorem 8, which is both 1-cost-approximate and 1-size-approximate; then the oblivious solution $\bar{F}$ is (a) 2β-cost-competitive or (b) β-size-competitive.

Theorem 8. *Let $\beta \geq 1$ and assume that there exists a polynomial β-competitive algorithm for online bidding. Fix an instance of k-median.*

(a) *In the metric case, suppose that for each $i \in [n]$ we have a set of facilities F_i^* with $|F_i^*| \leq i$ and cost$(F_i^*) \leq c \cdot opt_i$. Then in polynomial time we can compute an* oblivious *solution $(F_i)_i$ where $|F_i| \leq i$ and cost$(F_i) \leq 2c\beta \cdot opt_i$.*

(b) *Suppose that for each $i \in [n]$, we have a set of facilities F_i^* with $|F_i^*| \leq s \cdot i$ and cost$(F_i^*) \leq opt_i$. Then in polynomial time we can compute an* oblivious *solution $(F_i)_i$ where $|F_i| \leq s\beta \cdot i$ and cost$(F_i) \leq opt_i$.*

If the algorithm for online bidding is randomized, then the computations in (a) and (b) are also randomized.

Proof. We first prove part (a) of Theorem 8 in the deterministic case. The proof in the randomized setting is similar and we omit it.

For convenience, we introduce distances between facilities: given two $f, g \in \mathcal{F}$, let $d'_{fg} = \min_{x \in \mathcal{C}}(d_{fx} + d_{xg})$. This extension satisfies the triangle inequality. By assumption, each F^*_k is c-cost-approximate: $|F^*_k| \le k$ and $cost(F^*_k) \le c \cdot opt_k$. Assume without loss of generality that $cost(F^*_k) \le cost(F^*_{k+1})$ for all k.

The algorithm constructs the oblivious solution $(F_i)_i$ from $(F^*_i)_i$ in several steps. First, fix some index set $\mathcal{K} \subseteq [n]$, with $1 \in \mathcal{K}$, by a method to be described later, and let $\kappa(1), \kappa(2), \ldots, \kappa(m)$ denote the indices in $\mathcal{K}$ in increasing order. Next, compute F_k just for $k \in \mathcal{K}$. Start by defining $F_{\kappa(m)} = F^*_{\kappa(m)}$. Then, working backwards, inductively define $F_{\kappa(i)}$ to contain the facilities within $F_{\kappa(i+1)}$ that are "closest" to $F^*_{\kappa(i)}$.

More precisely, given two subsets A, B of $\mathcal{F}$, let $\Gamma(A,B)$ denote a subset Γ of B, minimal with respect to inclusion, and such that $d'_{\mu\Gamma} = d'_{\mu B}$ for all $\mu \in A$ (breaking ties arbitrarily). Obviously, $|\Gamma(A,B)| \le |A|$, and $\Gamma(A,B)$ can be computed in polynomial time given A and B. Then $F_{\kappa(i)} = \Gamma(F^*_{\kappa(i)}, F_{\kappa(i+1)})$.

Finally, define F_k for $k \in [n] \setminus \mathcal{K}$ as follows. Let $k^- = \max\{i \in \mathcal{K} : i \le k\}$ (it is well defined, since $1 \in \mathcal{K}$.) Define $F_k = F_{k^-}$. To complete the construction, it remains to describe how to compute $\mathcal{K}$, which we momentarily defer.

To analyze the size, note that $|F_k| \le k$, because for $k \in \mathcal{K}$, by definition of Γ we have $|F_k| \le |F^*_k| \le k$, while for $k \notin \mathcal{K}$, we have $|F_k| = |F_{k^-}| \le k^- < k$.

To analyze the cost, we use the following lemma. (The proof can be found in [8] and is also implicit in [13].)

Lemma 1. *Assume that the distance function is metric. Consider two sets $A, B \subseteq \mathcal{F}$ and let $\Gamma = \Gamma(A,B)$. Then for every $x \in X$ we have $c_{x\Gamma} \le 2c_{xA} + c_{xB}$.*

We now claim that

$$cost(F_k) \;\le\; 2 \sum_{\ell \ge k^-, \ell \in \mathcal{K}} cost(F^*_\ell). \tag{1}$$

Indeed, for indices $k \in \mathcal{K}$, we have $k = k^-$, and (1) follows from Lemma 1 summed over all x and from the construction of F_k (for $k = \kappa(m), \ldots, \kappa(1)$). For $k \notin \mathcal{K}$, inequality (1) holds as well, simply because $F_k = F_{k^-}$, the bound holds for $k = k^-$, and $(k^-)^- = k^-$.

Since $cost(F^*_k) \le c\, opt_k$, to make F $2c\beta$-cost-competitive we will choose $\mathcal{K}$ so that, for all k,

$$\sum_{\ell \ge k^-, \ell \in \mathcal{K}} cost(F^*_\ell) \;\le\; \beta\, cost(F^*_k). \tag{2}$$

To compute the set $\mathcal{K}$, let $\mathcal{U} = \{cost(F^*_n), cost(F^*_{n-1}), \ldots, cost(F^*_1)\}$ and take $\mathcal{B}$ to be any β-competitive bid set for universe $\mathcal{U}$. Define $\mathcal{K} = \{\kappa(m), \kappa(m-1), \ldots, \kappa(1)\}$ to be a minimal set (containing 1) such that the bid set is $\mathcal{B} = \{cost(F^*_{\kappa(m)}), cost(F^*_{\kappa(m-1)}), \ldots, cost(F^*_{\kappa(1)})\}$. Then the left-hand side of (2) is exactly the sum of the bids paid from the bid set for threshold $T = cost(F^*_k)$. Since the bid set is β-competitive, this is at most $\beta\, cost(F^*_k)$, so (2) holds. This completes the proof of part (a).

We now prove part (b) of Theorem 8. By assumption each F^*_k is s-size-approximate, that is, $|F^*_k| \le sk$ and $cost(F^*_k) \le opt_k$.

Fix some β-competitive bid set $\mathcal{B}$. Let $\mathcal{B}_k$ be the set of bids in $\mathcal{B}$ paid against threshold $T = k$ with $\mathcal{U} = [n]$. Define $F_k = \bigcup_{b \in \mathcal{B}_k} F^*_b$. Then $\bar{F} = (F_1, F_2, ..., F_n)$

is an oblivious solution because $\mathcal{B}_k \subseteq \mathcal{B}_\ell$ for $\ell \geq k$. Further, $cost(F_k) \leq opt_k$ because F_k contains F_b^* for some $b \geq k$, so $cost(F_k) \leq cost(F_b^*) \leq opt_b \leq opt_k$. Since $\mathcal{B}$ is β-competitive, we have $|F_k| \leq \sum_{b \in \mathcal{B}_k} |F_b^*| \leq \sum_{b \in \mathcal{B}_k} sb \leq s\beta k$.

Our next reduction shows that competitive online bidding reduces to size-competitive oblivious medians. Note that, together with Theorem 8(b), this implies that online bidding and size-competitive oblivious medians are equivalent.

Theorem 9. *Let $s \geq 1$ and assume that, for oblivious medians (metric or not), there is a (possibly randomized) s-size-competitive algorithm. Then, for any integer n, there is a (randomized) s-competitive algorithm for online bidding with $\mathcal{U} = [n]$.*

Proof. We give the proof in the deterministic setting. (The proof in the randomized setting is similar and we omit it.) For any arbitrarily large m, we construct sets $\mathcal{C}$ of customers and $\mathcal{F}$ of facilities, a metric distance function d_{uf}, for $u \in \mathcal{C}$ and $f \in \mathcal{F}$. The facility set $\mathcal{F}$ will be partitioned into sets $M_1, M_2, \ldots, M_m$, where $|M_k| = k$ for each k, with the following properties: (i) For all k, $cost(M_k) > cost(M_{k+1})$, and (ii) For all k, and for every set F of facilities, if $cost(F) \leq cost(M_k)$ then there exists $\ell \geq k$ such that M_ℓ is contained in F. These conditions imply that each M_k is the unique optimum k-median.

Assume for the moment that there exists such a metric space, and consider an s-size-competitive oblivious median $\bar{F}$ for it. Let $\mathcal{B} = \{k : M_k \subseteq F_k\}$. We show that $\mathcal{B}$ is an s-competitive bid set for universe $\mathcal{U} = [m]$. Against any threshold $T \in [m]$, the total of the bids paid will be

$$X = \sum \{k : k < T, M_k \subseteq F_k\} \;+\; \min\{\ell : \ell \geq T, M_\ell \subseteq F_\ell\} \tag{3}$$

Now, $\sum\{k : k < T, M_k \subseteq F_k\} \leq \sum\{k : k < T, M_k \subseteq F_T\}$ since $\bar{F}$ is a nested sequence. Similarly, we have

$$\min\{\ell : \ell \geq T, M_\ell \subseteq F_\ell\} \leq \min\{\ell : \ell \geq T, M_\ell \subseteq F_T\}$$

(By (ii), $M_\ell \subseteq F_T$ for some $\ell \geq T$, so the minimum on the right is well-defined for $T \in [m]$.) Thus:

$$\begin{aligned}
X &\leq \sum\{k : k < T, M_k \subseteq F_T\} + \min\{\ell : \ell \geq T, M_\ell \subseteq F_T\} \\
&= \sum\{|M_k| : k < T, M_k \subseteq F_T\} + \min\{|M_\ell| : \ell \geq T, M_\ell \subseteq F_T\} \;\; \text{since } |M_k| = k \\
&\leq \sum\{|M_k| : M_k \subseteq F_T\} \\
&\leq |F_T| \quad \text{since the } M_k\text{'s are disjoint} \\
&\leq sT \quad \text{since } \bar{F} \text{ is } s\text{-size-competitive.}
\end{aligned}$$

Thus, the bid set $\mathcal{B}$ is s-competitive for universe $\mathcal{U} = [m]$.

We now present the construction of the metric space satisfying conditions (i) and (ii). Let $\mathcal{C}$ be the set of integer vectors $\bar{u} = (u_1, u_2, \ldots, u_m)$ where $u_\ell \in [1, \ell]$ for all $\ell = 1, 2, \ldots, m$. For each $\ell \in [1, m]$, introduce a set $M_\ell = \{\mu_{\ell,1}, \mu_{\ell,2}, \ldots, \mu_{\ell,\ell}\}$, and for each node $\bar{u}$ in $\mathcal{C}$, connect $\bar{u}$ to μ_{ℓ,u_ℓ} with an edge of length $\delta_\ell = 1 + (m!)^{-\ell}$. The set of facilities is $\mathcal{F} = \bigcup_\ell M_\ell$. All distances between points in $\mathcal{C} \cup \mathcal{F}$ other than those specified above are determined by shortest-path distances. The resulting distance function satisfies the triangle inequality.

We have $cost(M_j) = m!\delta_j$ for each $j \in [1, m]$, so (i) holds. We prove (ii) by contradiction. Fix some index j and consider a set $F \subseteq \mathcal{F}$ that does not contain M_ℓ for any $\ell \geq j$: for each $\ell \geq j$ there is $i_\ell \leq \ell$ such that $\mu_{\ell,i_\ell} \notin F$. Define a customer $\bar{u}$ as follows: $u_i = 1$ for $\ell = 1, \ldots, j-1$ and $u_i = i_\ell$ for $\ell = j, \ldots, m$. Then the facility $\mu_{\ell,i} \in F$ serving this $\bar{u}$ must have $\ell < j$ or $i \neq i_\ell$. Either way, it is at distance at least δ_{j-1} from $\bar{u}$. Since each other customers pays strictly more than 1, we get $cost(F) > m! - 1 + \delta_{j-1} = m!\delta_j = cost(M_j)$ – a contradiction.

3 Online Bidding

In this section we prove Theorem 4. For completeness, we give proofs of the (folklore) deterministic and randomized upper bounds and deterministic lower bound. The upper bound uses a doubling algorithm that has been used in several papers, first in [15, 23] and later in [11, 3, 4, 9]. Our main new contribution in this section is a new randomized lower bound that matches the upper bound. (The proof of Lemma 3 was communicated to us by Yossi Azar.)

Lemma 2. *For online bidding, there is a deterministic 4-competitive algorithm. If $\mathcal{U}$ is finite, the algorithm runs in time polynomial in $|\mathcal{U}|$.*

Proof. First consider the case $\mathcal{U} = \mathbb{R}^+$. Define the algorithm to produce the set of bids $\{0\} \cup \{2^j : j \in \mathbb{N}\}$. Let $i = \lceil \log_2 T \rceil$, where $T > 0$ is the threshold: the algorithm pays $\sum_{j \leq i} 2^j = 2^{i+1} \leq 4T$, hence is 4-competitive.

Next, we reduce the general case to the case $\mathcal{U} = \mathbb{R}^+$. Knowing that $T \in \mathcal{U}$, the algorithm, when it would have bid $b \notin \mathcal{U}$, will instead bid the next smaller bid in $\mathcal{U}$ (if there is one, and otherwise the bid is skipped). This only decreases the cost the algorithm pays against any threshold $T \in \mathcal{U}$. Note that the modified algorithm can be implemented in time polynomial in $|\mathcal{U}|$ if $\mathcal{U}$ is finite.

Lemma 3. *For online bidding, no deterministic algorithm can be better than 4-competitive, even for $\mathcal{U} = \mathbb{N}^+$.*

Proof. let x_n be the nth bid, $s_n = \sum_1^n x_i$ and $y_n = s_{n+1}/s_n$. Suppose, for a contradiction, that there exists $a < 4$ such that $s_{n+1}/x_n < a$ for all n. Rewriting, we get $y_{n+1} \leq (1 - 1/y_n)a$. Since $1 - 1/z < z/4$, this implies $y_{n+1} < (y_n/4)a$; thus $y_n < (a/4)^n y_0$, and so eventually $s_{n+1} < s_n$, which is a contradiction.

Lemma 4. *For online bidding, there is a randomized e-competitive algorithm. If $\mathcal{U}$ is finite, then the algorithm runs in time polynomial in $|\mathcal{U}|$.*

Proof. First we consider the case $\mathcal{U} = \mathbb{R}^+$. Pick a real number $\xi \in (0, 1]$ uniformly at random, then choose the set of bids $\mathcal{B} = \{0\} \cup \{e^{i+\xi} : i \in \mathbb{N}\}$.

For the analysis, let random variable b be the largest bid paid by the algorithm against threshold $T > 0$. The total paid by the algorithm is less than $\sum_{i=0}^{\infty} be^{-i} = be/(e-1)$. Since b/T is distributed like e^{ξ} where ξ is distributed uniformly in $[0,1)$, the expectation of b is $T\int_0^1 e^z\,dz = T(e-1)$. Thus, the expected total payment is eT, and the algorithm is e-competitive.

The general case reduces to the case $\mathcal{U} = \mathbb{R}^+$ just as in the proof of Lemma 2.

Lemma 5. *Fix any $n \in \mathbb{N}^+$. Suppose $\mu : [n] \to \mathbb{R}^+$ and $\pi : [n] \to \mathbb{R}^+$ satisfy*

$$\sum_{T=t}^{n} \frac{1}{T}\pi(T) \;\geq\; \frac{1}{b}\sum_{T=t}^{b}\mu(T) \quad (\forall b, t : 1 \leq t \leq b \leq n). \tag{4}$$

For online bidding with $\mathcal{U} = [n]$, there is no randomized algorithm with competitive ratio better than $\sum_{T=1}^{n}\mu(T) \;/\; \sum_{T=1}^{n}\pi(T)$.

Proof. Consider a random set $\mathcal{B}$ of bids generated by any β-competitive randomized algorithm when $\mathcal{U} = [n]$. Without loss of generality, the maximum bid in $\mathcal{B}$ is n.

Let $\mathcal{B} = \{b_1, b_2, \ldots, b_m = n\}$ be the ordered sequence of bids in $\mathcal{B}$. Consider the sequence of intervals $([1, b_1], [b_1+1, b_2], [b_2+1, b_3], \ldots, [b_{m-1}+1, b_m])$, which exactly covers the points $1, 2, \ldots, n$. Let $x(t,b)$ denote the probability (over all random $\mathcal{B}$) that $[t,b]$ is one of these intervals. The algorithm pays bid b against threshold T if and only if, for some integer $t \leq T$, $[t,b]$ is one of these intervals. Thus, for any threshold T and bid b, $\sum_{t=1}^{T} x(t,b)$ is the probability that bid b is made against threshold T. (We will use this below.)

We claim that β, x form a feasible solution to the following linear program (LP):

$$\text{minimize}_{\beta,x}\ \beta \ \text{ subject to } \begin{cases} \beta - \displaystyle\sum_{b=1}^{n}\frac{b}{T}\sum_{t=1}^{T} x(t,b) \geq 0 \ (\forall T \in [n]) \\ \displaystyle\sum_{b=T}^{n}\sum_{t=1}^{T} x(t,b) \geq 1 \ (\forall T \in [n]) \\ x(t,b) \geq 0 \ (\forall t, b \in [n]). \end{cases}$$

The first constraint is met because, for any threshold T, $\sum_{t \leq T; b} b\, x(t,b)$ is the expected sum of the bids made by the algorithm if T is the threshold. This is at most βT because the algorithm has competitive ratio β. The second constraint is met because for any threshold T, the algorithm must have at least one bid above the threshold, hence at least one $[t,b]$ with $t \leq T \leq b$.

Thus, the value of this linear program (LP) is a lower bound on the optimal competitive ratio of the randomized algorithm. To get a lower bound on the value of (LP), we use the dual (DLP) (where the dual variables $\mu(T)$ correspond to the first set of constraints and $\pi(T)$ to the second set of constraints):

$$\text{maximize}_{\mu,\pi} \sum_{T=1}^{n} \mu(T) \quad \text{subject to} \begin{cases} \displaystyle\sum_{T=1}^{n} \pi(T) \leq 1 \\ \displaystyle\sum_{T=t}^{b} \mu(T) - \sum_{T=t}^{n} \frac{b}{T}\pi(T) \leq 0 \ (\forall t, b \in [n]) \\ \mu(T), \pi(T) \geq 0 \ (\forall T \in [n]). \end{cases}$$

Now, given any μ and π meeting the condition of the lemma, if we scale μ and π by dividing by $\sum_T \pi(T)$, we get a feasible dual solution whose value is $\sum_T \mu(T) \ / \sum_T \pi(T)$. Since the value of any feasible dual solution is a lower bound on the value of any feasible solution to the primal, it follows that the competitive ratio β of the randomized algorithm is at least $\sum_T \mu(T) \ / \sum_T \pi(T)$.

Lemma 6. *There exists $\mu : [n] \to \mathbb{R}^+$ and $\pi : [n] \to \mathbb{R}^+$ satisfying Condition (4) of Lemma 5 and such that $\sum_T \mu(T) / \sum_T \pi(T) \geq (1 - o(1))e$.*

Proof. Fix U arbitrarily large and let $n = \lceil U^2 \log U \rceil$. Let $\alpha > 0$ be a constant to be determined later: We will choose α so that Condition (4) holds, and then show that the corresponding lower bound is $e(1 - o(1))$ as $U \to \infty$. Define

$$\mu(T) = \begin{cases} \alpha/T & \text{if } U \leq T \leq U^2 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \pi(T) = \begin{cases} 1/T & \text{if } U \leq T \leq U^2 \log U \\ 0 & \text{otherwise.} \end{cases}.$$

If $T \geq U^2$, then the right-hand side of Condition (4) has value 0, so the condition holds trivially. On the other hand, since $\pi(T)$ and $\mu(T)$ are zero for $T < U$, if the condition holds for $T = U$, then it also holds for $T < U$. So, we need only verify the condition for T in the range $U \leq T \leq U^2$. The expression on the left-hand side of (4) then has value

$$\sum_{T=t}^{U^2 \log U} \frac{1}{T^2} \geq \int_t^{1+U^2 \log U} \frac{1}{T^2}\, dT = \frac{1}{t} - \frac{1}{1 + U^2 \log U} \geq \frac{1}{t}(1 - o(1)).$$

In comparison, the expression on the right-hand side has value at most

$$\max_{b \geq t} \frac{1}{b} \sum_{T=t}^{b} \frac{\alpha}{T} \leq \alpha \max_{b \geq t} \frac{1}{b} \int_{t-1}^{b} \frac{1}{T}\, dT = \alpha \max_{b \geq t} \frac{1}{b} \ln \frac{b}{t-1} = \frac{\alpha}{e\, t(1 - o(1))}.$$

(The second equation follows by calculus, for the maximum occurs when $b = e(t-1)$.) Thus, Condition (4) is met for $\alpha = (1 - o(1))e$. Then, Lemma 5 gives a lower bound on the competitive ratio of

$$\frac{\sum_T \mu(T)}{\sum_T \pi(T)} = \frac{\sum_{T=U}^{U^2} \alpha/T}{\sum_{T=U}^{U^2 \log U} 1/T} = (1 - o(1))\alpha \frac{\ln(U^2/U)}{\ln((U^2 \log U)/U)} = (1 - o(1))e.$$

Theorem 4 follows directly from Lemmas 2, 3, 4, 5, and 6.

4 Oblivious Algorithms for *kl*-Medians

In this section we sketch the proof of Theorem 6. Formally, in the kl-median problem we need to compute two sets $F_k \subseteq F_l$ with $|F_k| = k$ and $|F_l| = l$, minimizing the competitive ratio $R = \max\{cost(F_k)/opt_k, cost(F_l)/opt_l\}$.

The lower bound is a slight refinement of the one in [22, 21]. The metric space contains l customers, where customers j is connected to facility g_j by an edge of length $\delta = 1/l$. All customers are also connected to a facility f with edges of length 1.

Let $G = \{g_1, \ldots, g_l\}$. Then G is the optimal l-median. We have $cost(f) = l$, $cost(G) = l\delta$, $cost(g_i) = \delta + (l-1)(2+\delta)$, and $cost(G - g_i + f) = (l-1)\delta + 1$. So for $\delta = 1/l$, we get:

$$\frac{cost(g_i)}{cost(f)} = 2 - 1/l \qquad \text{and} \qquad \frac{cost(G - g_i + f)}{cost(G)} = 2 - 1/l.$$

The upper bound is achieved as follows. Let F and G denote, respectively, the optimum k-median and the optimum l-median. The algorithm choosese the better of two options: either (a) $F_k = F$ and $F_l = F \cup G - X$, where $X \subseteq G$ is a set of cardinality k that minimizes $cost(F \cup G - X)$, or (b) $F_k = Y$, where $Y \subseteq G$ is a set of cardinality k that minimizes $cost(Y)$, and $F_l = G$.

The competitive analysis of this algorithm is based on a probabilistic argument and will appear in the full version of this paper.

Acknowledgments. We are grateful to anonymous referees for suggestions to improve the presentation. We also wish to thank Yossi Azar for pointing out references to previous work on online bidding and simplifying the proof of Lemma 3.

References

1. A. Archer, R. Rajagopalan, and D.B. Shmoys. Lagrangian relaxation for the k-median problem: new insights and continuity properties. In *Proc. 11th European Symp. on Algorithms (ESA)*, pages 31–42, 2003.
2. V. Arya, N. Garg, R. Khandekar, K. Munagala, and V. Pandit. Local search heuristic for k-median and facility location problems. In *Proc. 33rd Symp. Theory of Computing (STOC)*, pages 21–29. ACM, 2001.
3. S. Chakrabarti, C.A. Phillips, A.S. Schulz, D.B. Shmoys, C. Stein, and J. Wein. Improved scheduling algorithms for minsum criteria. In *Automata, Languages and Programming*, pages 646–657, 1996.
4. M. Charikar, C. Chekuri, T. Feder, and R. Motwani. Incremental clustering and dynamic information retrieval. In *Proc. 29th Symp. Theory of Computing (STOC)*, pages 626–635. ACM, 1997.
5. M. Charikar and S. Guha. Improved combinatorial algorithms for the facility location and k-median problems. In *Proc. 40th Symp. Foundations of Computer Science (FOCS)*, pages 378–388. IEEE, 1999.
6. M. Charikar, S. Guha, E. Tardos, and D.B. Shmoys. A constant-factor approximation algorithm for the k-median problem. In *Proc. 31st Symp. Theory of Computing (STOC)*, pages 1–10. ACM, 1999.

7. C. Chekuri, A. Goel, S. Khanna, and A. Kumar. Multi-processor scheduling to minimize flow time with ϵ-resource augmentation. In *Proc. 36th Symp. Theory of Computing (STOC)*, pages 363–372. ACM, 2004.
8. M. Chrobak, C. Kenyon, and N.E. Young. The reverse greedy algorithm for the k-median problem. *Information Processing Letters*, 97:68–72, 2006.
9. S. Dasgupta and P.M. Long. Performance guarantees for hierarchical clustering. *Journal of Computer and System Sciences*, 70(4):555–569, 2005.
10. R. Fagin and L. Stockmeyer. Relaxing the triangle inequality in pattern matching. *IJCV: International Journal of Computer Vision*, 30:219–231, 1998.
11. M. Goemans and J. Kleinberg. An improved approximation ratio for the minimum latency problem. In *Proc. 7th Symp. on Discrete Algorithms (SODA)*, pages 152 – 158. ACM/SIAM, 1996.
12. K. Jain, M. Mahdian, and A. Saberi. A new greedy approach for facility location problems. In *Proc. 34th Symp. Theory of Computing (STOC)*, pages 731–740. ACM, 2002.
13. K. Jain and V.V. Vazirani. Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and lagrangian relaxation. *Journal of ACM*, 48:274–296, 2001.
14. B. Kalyanasundaram and K. Pruhs. Speed is as powerful as clairvoyance. *J. ACM*, 47:214–221, 2000.
15. M. Kao, J.H. Reif, and S. Tate. Searching in an unknown environment: An optimal randomized algorithm for the cow-path problem. *Information and Computation*, 131(1):63–80, 1996. Preliminary version appeared in the Proceedings of the Symp. on Discrete Algorithms, Austin, TX, Jan 1993.
16. M.R. Korupolu, C.G. Plaxton, and R. Rajaraman. Analysis of a local search heuristic for facility location problems. *Journal of Algorithms*, 37:146–188, 2000.
17. E. Koutsoupias. Weak adversaries for the k-server problem. In *Proc. 40th Symp. Foundations of Computer Science (FOCS)*, pages 444–449. IEEE, 1999.
18. G. Lin, C. Nagarajan, R. Rajamaran, and D.P. Williamson. A general approach for incremental approximation and hierarchical clustering. In *Proc. 17th Symp. on Discrete Algorithms (SODA)*. ACM/SIAM, 2006.
19. J-H. Lin and J.S. Vitter. Approximation algorithms for geometric median problems. *Information Processing Letters*, 44:245–249, 1992.
20. J-H. Lin and J.S. Vitter. ϵ-approximations with minimum packing constraint violation (extended abstract). In *Proc. 24th Symp. Theory of Computing (STOC)*, pages 771–782. ACM, 1992.
21. R. Mettu and C.G. Plaxton. The online median problem. *SIAM Journal on Computing*, 32:816–832, 2003.
22. R.R. Mettu and C.G. Plaxton. The online median problem. In *Proc. 41st Symp. Foundations of Computer Science (FOCS)*, pages 339–348. IEEE, 2000.
23. R. Motwani, S. Phillips, and E. Torng. Nonclairvoyant scheduling. *Theoretical Computer Science*, 130(1):17–47, 1994.
24. N.E. Young. K-medians, facility location, and the Chernoff-Wald bound. In *Proc. 11th Symp. on Discrete Algorithms (SODA)*, pages 86–95. ACM/SIAM, January 2000.