



A k -Median Based Online Algorithm for the Stochastic k -Server Problem

Abhijin Adiga¹ , Alexander D. Friedman², and Sharath Raghvendra³ 

¹ Biocomplexity Institute of Virginia Tech, Blacksburg, USA

abhijin@bi.vt.edu

² Department of Mathematics, Virginia Tech, Blacksburg, USA

adfriedm@vt.edu

³ Department of Computer Science, Virginia Tech, Blacksburg, USA

sharathr@vt.edu

Abstract. We consider the k -server problem in the random arrival model and present a simple k -median based algorithm for this problem. Let $\sigma = \langle r_1, \dots, r_n \rangle$ be the sequence of requests. Our algorithm will batch the requests into $\log n$ groups where the $(i+1)^{st}$ group contains requests $\langle r_{2^i+1}, \dots, r_{2^{i+1}} \rangle$. To process the requests of group $i+1$, our algorithm will place the k servers at the k -median centers of the first 2^i requests. When a new request of this group arrives, the algorithm will simply assign the server associated with the nearest k -median center to serve it. We show that this simple algorithm, in the random arrival model, has a competitive ratio of at most $O(\alpha)$ and an additive cost of $O(\Delta k \log n)$, where Δ is the diameter of the requests and α is a lower bound on the competitive ratio of any online algorithm in this model.

For our analysis, we use the following fact: In the random arrival model, the expected cost of serving the next request is minimized when servers are located at the k -median of the requests that have not yet arrived (unprocessed requests). But our algorithm instead uses the k -median of the requests seen so far as a proxy. Using existing analysis techniques, we obtain only a large bound on the difference between k -median of the unprocessed requests and that of the processed ones. In particular, in addition to $O(\alpha)$ times the optimal cost, for some $\epsilon > 0$, existing analysis techniques will also give an additive cost of ϵn for serving n requests. We present a new analysis to show that when the number of processed and unprocessed requests are of comparable sizes, the cost of serving n requests incurs only an additive cost of $k\Delta$ (independent of n and significantly better than the previous methods). We then apply this bound for serving each of the $\log n$ groups and obtain an overall bound which is $O(\alpha)$ times the optimal cost with an additive error of $O(k\Delta \log n)$.

Keywords: k -server problem · Random arrival model
 k -median clustering

A. D. Friedman and S. Raghvendra were supported under grant NSF-CCF 1464276.

A. Adiga was supported by the DTRA CNIMS Contract HDTRA1-11-D-0016-0001.

© Springer International Publishing AG, part of Springer Nature 2018

R. Solis-Oba and R. Fleischer (Eds.): WAOA 2017, LNCS 10787, pp. 176–189, 2018.

https://doi.org/10.1007/978-3-319-89441-6_14

1 Introduction

In this era of instant gratification, consumers demand access to goods and services in real-time. Business ventures, therefore, have to schedule their delivery of goods and services often without the complete knowledge of the future requests or their order of arrival. To assist with this, we need robust and competitive *online algorithms* that immediately and irrevocably allocate resources to requests in real-time with minimal cost.

In this paper, we consider the celebrated *k -server problem*. Initially, we are given a set of locations \mathcal{K} for the k servers. These are locations in a discrete metric space. Requests are generated in the same metric space by an adversary and are revealed one at a time. When a request $r \in \sigma$ is revealed, we have to immediately move one of the k servers to serve this request and incur a cost equal to the distance between their locations. The objective is to design an algorithm that allocates servers to requests that is competitive with respect to the minimum-cost offline solution.

For any algorithm \mathcal{A} , initial configuration of servers \mathcal{K} , and a sequence of requests σ , let $w_{\mathcal{A}}(\sigma, \mathcal{K})$ be the cost incurred when the algorithm \mathcal{A} assigns the requests to servers. Let $w_{\text{OPT}}(\sigma, \mathcal{K})$ be the smallest cost solution generated by an offline algorithm that has complete knowledge of the request sequence σ and assigns servers to requests based on their arrival order. We say that \mathcal{A} is α -competitive if, for a constant $\Phi_0 \geq 0$, the cost incurred by our algorithm satisfies,

$$w_{\mathcal{A}}(\sigma, \mathcal{K}) \leq \alpha w_{\text{OPT}}(\sigma, \mathcal{K}) + \Phi_0$$

for any request set and their arrival order.

In the *adversarial model*, there is an adversary who knows the server locations and the assignments made by the algorithm and generates a sequence to maximize α . In the *random arrival model* [1], the adversary chooses the locations of the requests in σ before the algorithm executes but their arrival order is a permutation chosen uniformly at random from the set of all possible permutations of the requests. In practical situations, it may be useful to assume that the requests are arriving i.i.d. from a known or an unknown distribution \mathcal{D} . Under these (known and unknown) models, the adversary is weaker than in the random arrival model and therefore, the competitive ratio for the random arrival model is an upper bound on its competitive ratio in the known and the unknown distribution models; see [2] for an algorithm in these models. We refer to the k -server problem under the known and unknown distribution as well as the random arrival model as the *stochastic k -server problem*.

For the stochastic k -server problem, the competitive ratio is expressed with respect to the expected costs. More specifically, \mathcal{A} is α -competitive if,

$$\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] \leq \alpha \mathbb{E}[w_{\text{OPT}}(\sigma, \mathcal{K})] + \Phi_0.$$

Previous Work. The k -server problem is central to the theory of online algorithms. The problem was first posted by Manasse *et al.* [3]. In the adversarial

model, the best-known deterministic algorithm for this problem is the $(2k - 1)$ -competitive work function algorithm [4]. It is known that no deterministic algorithm can achieve a competitive ratio better than k and is conjectured that in fact there is a k -competitive algorithm for this problem. This conjecture is popularly called the *k-server conjecture*.

Bansal *et al.* [5] presented an $O(\log^{O(1)} n \log k)$ -competitive randomized algorithm for the k -server problem in the *oblivious adversary model*. This model is similar to the adversarial model; however, the adversary does not have any knowledge of the random choices made by the algorithm. There is also an online algorithm [6] for the closely related online metric matching problem. This algorithm is known to achieve an optimal competitive ratio of $2H_n - 1$ in the random arrival model. To the best of our knowledge, the stochastic k -server problem has not been studied before.

In the stochastic models, we can view any initial set of requests to be a sample that is chosen uniformly at random from σ . In this paper, using this sample, we approximate the k -median of the remaining requests leading to an improved algorithm for the stochastic k -server problem. Independent and uniform random samples have been used before in the context of sub-linear time algorithms for the k -median problem; see [7, 8]. In [7], it has been shown that a random sample of size $\tilde{O}(\Delta/\epsilon^2)$ can be used to approximate the k -median within a constant factor of the optimal k -median with an additional additive cost of ϵn . Meyerson *et al.* [8] show that if all the optimal k -median clusters are dense ($\geq \epsilon n/k$), then a very small random sample of size $\approx k/\epsilon$ can be used to approximate the k -median within a constant factor. In this paper, we will present and analyze a k -median based deterministic algorithm for the stochastic k -server problem.

Our Results. First, we present a simple algorithm which we refer to as the zoned algorithm for the k -server problem in the known distribution model, i.e., the request locations are i.i.d from a distribution \mathcal{D} on the discrete metric space. The zoned algorithm will associate one server to each of the centers of the k -median of \mathcal{D} and when a request arrives, the algorithm simply assigns the server associated with the closest k -median center to this request. The cost of serving any request is lower bounded by the average k -median cost of the distribution \mathcal{D} . Using triangle inequality, we can bound the cost incurred by zoned algorithm by twice the cost incurred by any optimal online algorithm for this problem.

Next, for the unknown distribution model and the random arrival model, we present an adaptive version of the zoned algorithm. Let $\sigma = \langle r_1, \dots, r_n \rangle$ be the request sequence. Our algorithm will batch the requests into $\log n$ groups where the $(i + 1)^{st}$ group (denoted by σ_{i+1}) contains requests $\langle r_{2^i+1}, \dots, r_{2^{i+1}} \rangle$. To process requests group $(i + 1)$, we apply the zoned algorithm using the k -median centers of the first 2^i requests.

In the random arrival model, the first t requests is a uniformly chosen random subset of size t . Using existing bounds [7], a random subset of size $\tilde{O}(\Delta/\epsilon^2)$ can be used to estimate the average k -median cost within a constant factor with an additional additive cost of ϵ . Unfortunately, despite having a large random

subset $(\sigma_0 \cup \sigma_1, \dots, \cup \sigma_{\log n-1})$ of $n/2$ requests (i.e., $\epsilon \approx \sqrt{\Delta/n}$) to estimate the k -median of $n/2$ requests of $\sigma_{\log n}$, we can only bound the average k -median cost within a constant factor with an additional additive cost of $\epsilon n/2 \approx \sqrt{\Delta n}$. Lower bounds on uniform random sample based estimation of k median seem to suggest that there is very little scope for improving this analysis for small-sized random subsets; see [7] for details on the upper and lower bound. In our case, since the sample size is large ($=n/2$), we present a different analysis to show that the k -median of this large random subset is a good proxy for the k -median of σ_{i+1} . Using this analysis, we show that the overall cost incurred in serving requests of σ_{i+1} has an additive cost of only $k\Delta$ (independent of n) and the total additive cost over all the $\log n$ groups to be $O(k\Delta \log n)$ leading to the following theorem (in Sect. 4.2):

Theorem 1. *Let σ be a sequence of n requests from a discrete metric space (X, d) . For any $\alpha > 1.5$, the expected cost of the adaptive zoned algorithm (Algorithm 2) for serving σ is upper bounded as follows.*

$$\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] \leq 2\alpha n \text{medavg}(\sigma) + \left(\frac{8}{e} \left(\frac{2\alpha + 1}{2\alpha - 3} \right)^2 + 1 \right) k\Delta \log n + \Phi_0.$$

Here $\text{medavg}(\sigma)$ is the average k -median cost of all the requests in σ and is formally defined in Sect. 2.

In the random arrival model, the cost of serving the i th request can be lower bounded by the average k -median cost of the remaining (unprocessed) $n - i + 1$ requests. We show (in Sect. 4.3) that the lower bound on the cost of serving all the requests can still be related to the average k -median cost $\text{medavg}(\sigma)$ within an additive cost of $O(k\Delta \log n)$.

Theorem 2. *Let (X, d) be any discrete metric space and σ be a multi-set of n points from X . Let \mathcal{A} be any online algorithm to serve σ under the random arrival model, with initial configuration of servers \mathcal{K} . Let Δ be the diameter of X and for any $0 < \delta < 1$, the expected cost of \mathcal{A} is*

$$\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] \geq \frac{n-1}{2} \delta \text{medavg}(\sigma) - \frac{2\delta}{(1-\delta)^2} (k+2) \Delta \log n.$$

Combining the two theorems, it follows that the adaptive zoned algorithm performs within a constant factor of the cost incurred by the best online algorithm along with an additional additive cost of $O(k\Delta \log n)$.

In Sect. 2, we present the basic terminology required for our algorithm and its analysis. In Sect. 3, we present the zoned algorithm for the known distribution model along with its analysis. In Sect. 4, we present the adaptive zoned algorithm in the random arrival model. We present our analysis of the upper bound of the cost in Sect. 4.2 and lower bound of the cost in Sect. 4.3.

2 Preliminaries

Let P be a multi-set of n points in a given discrete metric space (X, d) . For any point $p \in P$ and a set $K \subset X$, we define $d(p, K)$ to be the distance of p to its

nearest neighbor in K . We define the distance of the set P to the set K denoted by $d(P, K)$.

$$d(P, K) = \sum_{p \in P} d(p, K).$$

The average distance of P from K , denoted as $d_{\text{avg}}(P, K)$ is $d_{\text{avg}}(P, K) = \frac{1}{|P|} \sum_{x \in P} d(x, K)$. We define the k -median of P to be a set of k points $K^* \subseteq X$, given by

$$K^* = \arg \min_{K \subset X, |K|=k} d(P, K).$$

We refer to K^* as the k -median centers of P . The cost of the k -median K^* denoted by $\text{med}(P)$ is $\text{med}(P) = d(P, K^*)$, and the average cost of this k -median, denoted by $\text{medavg}(P)$, is given by

$$\text{medavg}(P) = \frac{\text{med}(P)}{|P|}.$$

In several instances, we denote the k -median of a set A by K^A .

The definition of k -median K^* extends easily to the case where we are given a probability distribution $\mathcal{D}(\cdot)$ on the discrete metric space X :

$$K^* = \arg \min_{K \subset X, |K|=k} \sum_{x \in X} \mathcal{D}(x) d(x, K),$$

and let $\text{medavg}(\mathcal{D}, X) = \sum_{x \in X} \mathcal{D}(x) d(x, K^*)$.

Theorem 3 (Chernoff bounds). *Suppose X_1, \dots, X_n are independent random binary variables, X denotes their sum, and $\mu = \mathbb{E}[X]$. Then*

$$\mathbb{P}[X \geq (1 + \delta)\mu] \leq e^{-\delta^2 \mu / 3}, \quad 0 < \delta < 1, \quad (1a)$$

$$\mathbb{P}[X \geq (1 + \delta)\mu] \leq e^{-\delta \mu / 3}, \quad 1 < \delta, \quad (1b)$$

$$\mathbb{P}[X \leq (1 - \delta)\mu] \leq e^{-\delta^2 \mu / 2}, \quad 0 < \delta < 1. \quad (1c)$$

3 Zoned Algorithm

We begin by introducing the zoned algorithm for the k -server problem in the known distribution (Algorithm 1) and the random arrival model (Algorithm 2). The core idea behind the algorithm is that the discrete metric space (X, d) can be partitioned into k zones each with a single server. For any request in a given zone, the corresponding server of this zone will serve it. For the known distribution, the partition is induced by the k -median of the distribution and is presented below.

Algorithm 1. Zoned algorithm for a known distribution

Data: Metric space (X, d) , probability distribution \mathcal{D} , a sequence of requests $\sigma = (r_1, \dots, r_n)$, k -servers $(\kappa_1, \dots, \kappa_k)$ and their initial positions

Result: Sequence of servers assigned in an online manner

- 1 Compute the k -median centers K^* of the distribution \mathcal{D} ;
- 2 $\phi : K^* \rightarrow (\kappa_1, \dots, \kappa_k)$, minimum-cost bipartite matching from K^* to the initial locations of k -servers;
- 3 **for** r **in** σ **do**
- 4 Find k -median center c with distance $d(r, K^*)$;
- 5 Move server $\phi(c)$ to r ;
- 6 **end**

The following lemma bounds the cost of any online algorithm with the k -median cost.

Lemma 1. *Given a discrete metric space (X, d) , and any request sequence σ of n locations chosen i.i.d. from a known distribution \mathcal{D} on X , the expected cost of any online algorithm \mathbb{A} is*

$$\mathbb{E}[w_{\mathbb{A}}(\sigma, \mathcal{K})] \geq n \cdot \text{medavg}(\mathcal{D}, X).$$

Proof. When a request arrives the algorithm must assign a server to it, so the cost of the algorithm must be at least the distance from the closest server to the request. The expected cost is therefore bounded below by the expected distance of a request to its nearest server. This is minimized if the servers have the configuration of the k -median centers K^* with the expected distance of $\text{medavg}(\mathcal{D}, X)$. The result then follows by linearity of expectation.

Theorem 4. *The zoned algorithm \mathcal{A} has an expected cost that is at most twice the cost incurred by any optimal online algorithm in the known distribution model.*

Proof. Let the initial configuration of the k servers be \mathcal{K} . For every request $r \in \sigma$, let $k \in K^*$ be its closest center. The zoned algorithm will move the server $\phi(k)$ to the request point r . By the triangle inequality this distance is less than if we had moved the server to c first, and then to r after. Every request under such a modification therefore incurs at most two costs, movement from c and movement to c . The expected distance of any request to its closest center is $\text{medavg}(\mathcal{D}, X)$, so using the modification, by linearity of expectation, and by Lemma 1:

$$\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] \leq 2n \cdot \text{medavg}(\mathcal{D}, X) + \Phi_0 \leq 2w_{\mathbb{A}}(\sigma, \mathcal{K}) + \Phi_0$$

for n requests, any online algorithm \mathbb{A} , and where Φ_0 is the cost of the matching of \mathcal{K} and K^* .

4 Random Arrival Model

4.1 Adaptive Zoned Algorithm

We present and analyze a slightly modified zoned algorithm for the random arrival model. We partition the request sequence σ into $\log n$ groups $\sigma = \sigma_0 \frown \sigma_1, \dots, \frown \sigma_{\log n}$ where $|\sigma_0| = 1$ and $|\sigma_i| = 2^{i-1}$. For the request in σ_i we apply the zoned algorithm by using the k -median of the requests in $\sigma_0 \cup \sigma_1, \dots, \cup \sigma_{i-1}$. Note that after serving requests in σ_i , we need to recompute the k -median of all the requests seen so far and move the k -servers to these locations (implicitly through the mapping ϕ). This results in a reconfiguration cost which is bounded by $O(k\Delta \log n)$. The algorithm is presented next.

Algorithm 2. Adaptive zoned algorithm \mathbb{A} that runs in the Random Arrival Model

Data: Metric (X, d) , a sequence of requests $\sigma = (r_1, \dots, r_n)$, k -server $(\kappa_1, \dots, \kappa_k)$ and their initial positions

Result: Sequence of servers assigned in an online manner

```

1 1  $\rightarrow i$ ;
2 1  $\rightarrow j$ ;
3  $K^{(1)}$  is the initial positions of the  $k$ -servers  $(\kappa_1, \dots, \kappa_k)$ ;
4  $\phi$  is the map of locations in  $K^{(1)}$  to the servers that are located there;
5 for  $r$  in  $\sigma$  do
6   Find center  $c \in K^{(j)}$  that minimizes  $d(r, K^{(j)})$ ;
7   Move server  $\phi(c)$  to  $\sigma$ ;
8   if  $i = 2^{j-1}$  then
9     Compute the  $k$ -median centers  $K^{(j)}$  on  $\{r_1, \dots, r_i\}$ ;
10     $\phi : K^{(j)} \rightarrow (\kappa_1, \dots, \kappa_k)$ , minimum-cost bipartite matching
        from  $K^{(j)}$  to the servers at their current locations;
11     $j \leftarrow j + 1$ ;
12  end
13   $i \leftarrow i + 1$ ;
14 end
```

4.2 Upper Bound

We use the following simple lemma in bounding the cost of the adaptive zoned algorithm.

Lemma 2. *Let P be any set of n points and let Q be a subset of P that is chosen uniformly at random from all possible subsets of size t . Then,*

$$\mathbb{E}[d(Q, K^P)] = t \cdot \text{medavg}(P).$$

Proof. Let \mathcal{S}_t be the set of all subsets of P with cardinality exactly t . The number of such subsets is given by $|\mathcal{S}_t| = \binom{n}{t}$. Therefore, the expected value of $d(Q, K^P)$ can be bounded by

$$\mathbb{E}[d(Q, K^P)] = \sum_{Q \in \mathcal{S}_t} \frac{1}{\binom{n}{t}} \sum_{q \in Q} d(q, K^P).$$

Every point $q \in P$ appears in exactly $\binom{n-1}{t-1}$ subsets of \mathcal{S}_t . Therefore, we can rewrite the expected value as

$$\mathbb{E}[d(Q, K^P)] = \frac{\binom{n-1}{t-1}}{\binom{n}{t}} \sum_{q \in P} d(q, K^P) = \frac{t}{n} \sum_{q \in P} d(q, K^P) = t \cdot \text{medavg}(P).$$

Lemma 3. *Let P be a set of n points with diameter Δ where n is a power of 2. For a random permutation of the points of P , let A and B correspond to the first $n/2$ and the last $n/2$ points of this permutation. Let K^P be the k -median centers of P . Given that A has been observed and K^A is the k -median centers of A , and for any $\alpha > 1.5$, the expected value of distance $d(B, K^A)$ is at most $\alpha n \text{medavg}(P) + \frac{4}{e} \left(\frac{2\alpha+1}{2\alpha-3} \right)^2 k\Delta$.*

Proof. Since A is the first $n/2$ points of a random permutation of P , A can be considered to be a subset chosen uniformly at random from the set of all subsets of size $n/2$. For the optimal k -median centers of A , namely K^A , we will bound the expected cost of $d(B, K^A)$. To help with the analysis, we partition P into k clusters based on assigning each point $p \in P$ to its closest k -median center in K^P (the optimal k -median of P). Let C_j^P denote the j th cluster with $k_j^P \in K^P$ as its median center and let $s_j = |C_j^P|$ be the size of this cluster. Let $A_j = A \cap C_j^P$ and $B_j = B \cap C_j^P$. By using triangle inequality (as in [8]), we can bound the distance of median center of cluster C_j^P , i.e., k_j^P to its closest median center in K^A

$$d(k_j^P, K^A) \leq \min_{x \in A_j} (d(k_j^P, x) + d(x, K^A)) \leq \frac{1}{|A_j|} \sum_{x \in A_j} (d(k_j^P, x) + d(x, K^A)).$$

The last inequality follows from the fact that the minimum of a set of numbers is bounded from above by its average. Using this bound, we can bound the distance from any point $y \in C_j^P$ to its closest median center in K^A by

$$d(y, K^A) \leq d(y, k_j^P) + d(k_j^P, K^A) \leq d(y, k_j^P) + \frac{1}{|A_j|} \sum_{x \in A_j} (d(k_j^P, x) + d(x, K^A)).$$

Therefore, the total distance of points in B_j to their closest center in K^A is bounded by

$$d(B_j, K^A) \leq \sum_{y \in B_j} d(y, k_j^P) + \frac{|B_j|}{|A_j|} \sum_{x \in A_j} (d(k_j^P, x) + d(x, K^A)). \quad (2)$$

Next, we will bound the expected value of $d(B_j, K^A)$. Note that the expected value $\mathbb{E}[|A_j|] = \frac{s_j}{2} = \mathbb{E}[|B_j|]$. Consider the event $\mathcal{E} : (|A_j| > (1 - \delta)\mathbb{E}[|A_j|])$, where $0 < \delta < 1$. Applying Chernoff bound (Theorem 3, inequality (1c)), event \mathcal{E} occurs with probability at least $1 - \exp\left(-\frac{\delta^2 s_j}{4}\right)$. When event \mathcal{E} occurs,

$$\frac{|B_j|}{|A_j|} = \frac{s_j - |A_j|}{|A_j|} < \frac{s_j(1 - \frac{1-\delta}{2})}{(1-\delta)\frac{s_j}{2}} = \frac{1+\delta}{1-\delta}.$$

and we can bound $d(B_j, K^A)$ by

$$\sum_{y \in B_j} d(y, k_j^P) + \frac{1+\delta}{1-\delta} \sum_{x \in A_j} (d(k_j^P, x) + d(x, K^A)).$$

When \mathcal{E} does not occur (which has a probability of at most $\exp\left(-\frac{\delta^2 s_j}{4}\right)$), i.e., $|A_j| \leq (1-\delta)\mathbb{E}[|A_j|]$, we use a trivial upper bound of $d(B_j, K^A) \leq s_j \Delta$. Applying this, we have the following upper bound for $\mathbb{E}[d(B_j, K^A)]$, where the expectation is over all possible permutations of P .

$$\begin{aligned} \mathbb{E}[d(B_j, K^A)] &= \Pr(\mathcal{E})\mathbb{E}[d(B_j, K^A) \mid \mathcal{E}] + \Pr(\bar{\mathcal{E}})\mathbb{E}[d(B_j, K^A) \mid \bar{\mathcal{E}}] \\ &< \mathbb{E}\left[\sum_{y \in B_j} d(y, k_j^P) + \frac{1+\delta}{1-\delta} \sum_{x \in A_j} (d(k_j^P, x) + d(x, K^A))\right] \\ &\quad + \exp\left(-\frac{\delta^2 s_j}{4}\right) s_j \Delta. \end{aligned} \quad (3)$$

We set $\delta := 2\sqrt{\frac{\log(s_j/\tau)}{s_j}}$. When $s_j \geq \tau$, we can reduce the last term in (3) to $\exp\left(-\frac{\delta^2 s_j}{4}\right) s_j \Delta \leq \tau \Delta$. When $s_j < \tau$, we can simply bound $d(B_j, K^A) \leq |B_j| \Delta < \tau \Delta$. Also, we note that $\frac{1+\delta}{1-\delta}$ is a monotonically increasing function of δ for $0 < \delta < 1$. For $x > 0$, $\delta = 2\sqrt{\frac{\log(s_j/\tau)}{s_j}}$ attains the maximum value of $\frac{2}{\sqrt{e\tau}}$ at $s_j = e\tau$. Therefore, $\frac{1+\delta}{1-\delta} \leq \frac{\sqrt{e\tau}+2}{\sqrt{e\tau}-2}$. Using these bounds, we rewrite (3) as

$$\mathbb{E}[d(B_j, K^A)] \leq \mathbb{E}\left[\sum_{y \in B_j} d(y, k_j^P) + \frac{\sqrt{e\tau}+2}{\sqrt{e\tau}-2} \sum_{x \in A_j} (d(k_j^P, x) + d(x, K^A))\right] + \tau \Delta.$$

Summing over all clusters,

$$\begin{aligned} \mathbb{E}[d(B, K^A)] &\leq \mathbb{E}\left[\sum_{j=1}^k \sum_{y \in B_j} d(y, k_j^P) + \frac{\sqrt{e\tau}+2}{\sqrt{e\tau}-2} \sum_{j=1}^k \sum_{x \in A_j} (d(k_j^P, x) + d(x, K^A))\right] \\ &\quad + k\tau \Delta \\ &\leq \mathbb{E}\left[\sum_{y \in B} d(y, K^P) + \frac{\sqrt{e\tau}+2}{\sqrt{e\tau}-2} \sum_{x \in A} (d(K^P, x) + d(x, K^A))\right] + k\tau \Delta. \end{aligned}$$

Since $\sum_{x \in A} d(x, K^A) \leq \sum_{x \in A} d(x, K^P)$, we have

$$\mathbb{E}[d(B, K^A)] \leq \mathbb{E}\left[\sum_{y \in B} d(y, K^P) + 2\frac{\sqrt{e\tau} + 2}{\sqrt{e\tau} - 2} \sum_{x \in A} d(x, K^P)\right] + k\tau\Delta.$$

Since A and B are subsets chosen uniformly at random from all possible subsets of P of size $n/2$, by Lemma 2, we have $\mathbb{E}[\sum_{y \in B} d(y, K^P)] = |B| \text{medavg}(P)$ and $\mathbb{E}[\sum_{y \in A} d(y, K^P)] = |A| \text{medavg}(P)$. Therefore,

$$\mathbb{E}[d(B, K^A)] \leq |B| \text{medavg}(P) + 2\frac{\sqrt{e\tau} + 2}{\sqrt{e\tau} - 2} |A| \text{medavg}(P) + k\tau\Delta$$

and,

$$\mathbb{E}[d(B, K^A)] \leq \left(\frac{\sqrt{e\tau} + 2}{\sqrt{e\tau} - 2} + \frac{1}{2}\right)n \text{medavg}(P) + k\tau\Delta.$$

Setting $\alpha = \frac{\sqrt{e\tau} + 2}{\sqrt{e\tau} - 2} + \frac{1}{2}$ we arrive at the final expression.

Proof (Proof of Theorem 1). Let $\sigma = \sigma^{(0)} \circ \sigma^{(1)} \circ \sigma^{(2)} \circ \dots$, where \circ denotes concatenation, so that $|\sigma^{(0)}| = 1$ and $|\sigma^{(i)}| = 2^{i-1}$. Let $T^{(i)} = \sigma^{(0)} \circ \sigma^{(1)} \circ \dots \circ \sigma^{(i)}$ and $K^{(i)}$ denote the k -median centers of $T^{(i)}$. Here, we use the terms multi-set and sequence interchangeably. By algorithm \mathcal{A} , each sequence $\sigma^{(i)}$ is served by $K^{(i-1)}$. Noting that $T^{(i)} = T^{(i-1)} \circ \sigma^{(i)}$ and $|\sigma^{(i)}| = |T^{(i-1)}| = 2^{i-1}$, we apply Lemma 3 with $P = T^{(i)}$, $A = T^{(i-1)}$ and $B = \sigma^{(i)}$.

The expected distance $d(\sigma^{(i)}, K^{(i-1)}) \leq \alpha |T^{(i)}| \text{medavg}(T^{(i)}) + \frac{4}{e} \left(\frac{2\alpha+1}{2\alpha-3}\right)^2 k\Delta$, and the cost to serve $\sigma^{(i)}$ is at most twice this cost. In addition, for each $\sigma^{(i)}$, the cost to move the servers from their position at the end of $\sigma^{(i-1)}$ to $K^{(i)}$ is at most $k\Delta$. Therefore, $\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})]$ can be bounded as follows.

$$\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] \leq 2 \sum_i \alpha |T^{(i)}| \text{medavg}(T^{(i)}) + \sum_i \left(\frac{8}{e} \left(\frac{2\alpha+1}{2\alpha-3}\right)^2 k\Delta + k\Delta \right) + \Phi_0.$$

Since, $\text{medavg}(T^{(i)}) = \frac{1}{|T^{(i)}|} \sum_{x \in T^{(i)}} d(x, K^{(i)}) \leq \frac{1}{|T^{(i)}|} \sum_{x \in T^{(i)}} d(x, K^*)$,

$$\begin{aligned} \mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] &\leq 2\alpha \sum_i \sum_{x \in T^{(i)}} d(x, K^*) + \left(\frac{8}{e} \left(\frac{2\alpha+1}{2\alpha-3}\right)^2 + 1 \right) k\Delta \log n + \Phi_0 \\ &= 2\alpha \sum_{x \in \sigma} d(x, K^*) + \left(\frac{8}{e} \left(\frac{2\alpha+1}{2\alpha-3}\right)^2 + 1 \right) k\Delta \log n + \Phi_0 \\ &= 2\alpha n \text{medavg}(\sigma) + \left(\frac{8}{e} \left(\frac{2\alpha+1}{2\alpha-3}\right)^2 + 1 \right) k\Delta \log n + \Phi_0. \end{aligned}$$

Hence proved.

4.3 Lower Bound

Let \mathcal{S}_j denote the set of all possible subsets of σ of size j . The expected value of $\text{medavg}(A)$ for a set A of size m chosen uniformly at random from \mathcal{S}_m is denoted by $\text{medavg}_{\mathbb{E}}(m)$, and defined as

$$\text{medavg}_{\mathbb{E}}(m) = \frac{1}{|\mathcal{S}_m|} \sum_{A \in \mathcal{S}_m} \text{medavg}(A).$$

We will first prove the following lemmas which bounds the cost of any online algorithm from below by $\sum_{i=1}^j \text{medavg}_{\mathbb{E}}(i)$.

Lemma 4. $\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] \geq \sum_{i=1}^n \text{medavg}_{\mathbb{E}}(i)$.

Proof. Supposing $i - 1$ requests have been served, let A denote the subset of S containing the points yet to be served, i.e., $|A| = n - i + 1$. Let K be the current configuration of the servers. Since every element of A has the same probability of being picked, for any k -set K , the expected distance of the i^{th} request from K is

$$\frac{1}{|A|} \sum_{x \in A} d(x, K) \geq \text{medavg}(A).$$

Since every set $A \in \mathcal{S}_{n-i+1}$ is equiprobable, the cost of serving the i th request is $\text{medavg}_{\mathbb{E}}(n - i + 1)$ and the result follows.

Lemma 5. For any $m > m'$, $\text{medavg}_{\mathbb{E}}(m) \geq \text{medavg}_{\mathbb{E}}(m')$.

Proof. Let $A \in \mathcal{S}_m$ and $B \subset A$ of size m' . Noting that every element of A occurs in exactly $\binom{m-1}{m'-1}$ subsets B , we have

$$\begin{aligned} \text{medavg}(A) &= \frac{1}{|A|} \sum_{x \in A} d(x, K^A) = \frac{1}{m} \frac{1}{\binom{m-1}{m'-1}} \sum_{B \subset A, |B|=m'} d(x, K^A) \\ &\geq \frac{1}{m} \frac{1}{\binom{m-1}{m'-1}} \sum_{B \subset A, |B|=m'} d(x, K^B) \\ &= \frac{m'}{m} \frac{1}{\binom{m-1}{m'-1}} \sum_{B \subset A, |B|=m'} \text{medavg}(B). \end{aligned}$$

Since, every B is a subset of exactly $\binom{n-m'}{m-m'}$ sets of \mathcal{S}_m ,

$$\begin{aligned} \sum_{A \in \mathcal{S}_m} \text{medavg}(A) &\geq \binom{n-m'}{m-m'} \frac{m'}{m} \frac{1}{\binom{m-1}{m'-1}} \sum_{B \in \mathcal{S}_{m'}} \text{medavg}(B) \\ &= \frac{|\mathcal{S}_m|}{|\mathcal{S}_{m'}|} \sum_{B \in \mathcal{S}_{m'}} \text{medavg}(B). \end{aligned}$$

Lemma 6. $\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] \geq \lceil \frac{n}{2} \rceil \text{medav}_{\mathbb{E}}(\lceil \frac{n}{2} \rceil)$.

Proof. From Lemma 5, for any $m > \frac{n}{2}$, $\text{medav}_{\mathbb{E}}(m) \geq \text{medav}_{\mathbb{E}}(\lceil \frac{n}{2} \rceil)$. From Lemma 4,

$$\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] \geq \sum_{i=1}^n \text{medav}_{\mathbb{E}}(i) \geq \sum_{i=1}^{\lceil \frac{n}{2} \rceil} \text{medav}_{\mathbb{E}}(n-i+1) \geq \lceil \frac{n}{2} \rceil \text{medav}_{\mathbb{E}}(\lceil \frac{n}{2} \rceil).$$

Following Meyerson *et al.* [8], for any k -set K of σ , let $\beta(K, b) = (B_1, B_2, \dots, B_b)$ denote the partition of σ induced by K as follows: We order points of σ by their distance to the closest point in K (low to high), and divide this sequence into b bins each containing equal number of points $\frac{n}{b}$. Henceforth, let $b = \frac{n(1-\delta)^2}{4(k+2)\log n}$ for $0 < \delta < 1$.

Lemma 7. *Let A be any random subset of σ of size $\lceil \frac{n}{2} \rceil$. With probability at least $1 - \frac{1}{n}$, every bin of $\beta(K^A, b)$ has at most $\frac{\delta n}{2b}$ points from A .*

Proof. We will show that with high probability the above statement is satisfied for all k -sets K . Since K^A is one among these sets, the bound follows. Let K be any k -set and $\beta(K, b)$ be the induced partition. The expected number of points of A belonging to each bin is $\frac{n}{2b}$. Using Chernoff bound (1c), for $0 < \delta < 1$, the probability that a bin has at most $\frac{\delta n}{2b}$ points is at most $e^{-\frac{n(1-\delta)^2}{4b}}$. Since there are at most n^k sets K and b bins, applying union bound the proof follows.

Proof (Proof of Theorem 2). Let $A \subset \sigma$ of size $\frac{n}{2}$, and $\beta(K^A, b)$ be the partition of σ induced by K^A . Let d_j^{\min} and d_j^{\max} denote the minimum and maximum distances respectively of points from bin B_j . From Lemma 7, with probability at least $1 - \frac{1}{n}$, every bin of $\beta(K^A, b)$ has at most $\frac{\delta n}{2b}$ points of A . Therefore,

$$\sum_{x \in A} d(x, K^A) \geq \sum_{j=1}^b \frac{\delta n}{2b} d_j^{\min} \geq \frac{\delta}{2} \sum_{j=1}^{b-1} \frac{n}{b} d_j^{\max}.$$

The last inequality follows from the fact that $d_j^{\min} \geq d_{j-1}^{\max}$. For each j , since d_j^{\max} is the maximum assigned distance, $\frac{n}{b} d_j^{\max} \geq \sum_{y \in B_j} d(y, K^A)$. Finally, we note that $d_b^{\max} \leq \Delta$, the diameter. Combining,

$$\begin{aligned} \sum_{x \in A} d(x, K^A) &\geq \frac{\delta}{2} \sum_{j=1}^{b-1} \sum_{y \in B_j} d(y, K^A) + \frac{\delta}{2} \sum_{y \in B_b} d(y, K^A) - \frac{\delta n}{2b} \Delta \\ &= \frac{\delta}{2} \sum_{j=1}^b \sum_{y \in B_j} d(y, K^A) - \frac{\delta n}{2b} \Delta \\ &= \frac{\delta}{2} \sum_{y \in \sigma} d(y, K^A) - \frac{\delta n}{2b} \Delta. \end{aligned}$$

Therefore, with probability $(1 - 1/n)$

$$\text{medavg}(A) \geq \frac{\delta}{n} \sum_{y \in \sigma} d(y, K^A) - \frac{\delta \Delta}{b} \geq \delta \text{medavg}(\sigma) - \frac{\delta \Delta}{b}.$$

Hence,

$$\text{medavg}_{\mathbb{E}}\left(\left\lceil \frac{n}{2} \right\rceil\right) \geq \left(1 - \frac{1}{n}\right) \left(\frac{\delta}{n} \sum_{y \in \sigma} d(y, K^A) - \frac{\delta \Delta}{b}\right) + \frac{1}{n} \cdot 0$$

From Lemma 6 and the choice of $b = \frac{n(1-\delta)^2}{4(k+2)\log n}$,

$$\begin{aligned} \mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] &\geq \left\lceil \frac{n}{2} \right\rceil \text{medavg}_{\mathbb{E}}\left(\left\lceil \frac{n}{2} \right\rceil\right) \geq \left\lceil \frac{n}{2} \right\rceil \left(1 - \frac{1}{n}\right) \left(\delta \text{medavg}(\sigma) - \frac{\delta \Delta}{b}\right) \\ &\geq \frac{n-1}{2} \left(\delta \text{medavg}(\sigma) - \frac{\delta \Delta}{b}\right) \\ &= \frac{n-1}{2} \delta \text{medavg}(\sigma) - \frac{2\delta}{(1-\delta)^2} (k+2) \Delta \log n. \end{aligned}$$

Hence proved.

5 Conclusion

In this paper, we presented and analyzed a simple k -median based algorithm for the stochastic k -server problem. Our result is based on proving a new and sharper approximation bound of the k -median of a large random subset of a point set with respect to the k -median of entire point set.

In the random arrival model, the cost of serving the next request is lower bounded by the average k -median cost of the requests that have not yet been processed. Clearly, the k servers cannot always be in the optimal k -median configuration even for the best online algorithm. Therefore, it is conceivable that one can prove a stronger lower bound. In particular, can we prove a stronger lower or upper bound and reduce the additive error from $O(k\Delta \log n)$ to $O(k\Delta)$ (completely independent of n)?

References

1. Mahdian, M., Yan, Q.: Online bipartite matching with random arrivals: an approach based on strongly factor-revealing LPs. In: Proceedings of the 43rd Annual ACM Symposium on Theory of Computing, STOC 2011, pp. 597–606 (2011)
2. Karande, C., Mehta, A., Tripathi, P.: Online bipartite matching with unknown distributions. In: Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing, STOC 2011, pp. 587–596. ACM, New York (2011)
3. Manasse, M.S., McGeoch, L.A., Sleator, D.D.: Competitive algorithms for server problems. *J. Algorithms* **11**(2), 208–230 (1990)

4. Koutsoupias, E., Papadimitriou, C.H.: On the k -server conjecture. *J. ACM* **42**(5), 971–983 (1995)
5. Bansal, N., Buchbinder, N., Madry, A., Naor, J.: A polylogarithmic-competitive algorithm for the k -server problem. In: *Proceedings of the IEEE 52nd Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 267–276, October 2011
6. Raghvendra, S.: A Robust and optimal online algorithm for minimum metric bipartite matching. In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2016*, vol. 60, pp. 18:1–18:16 (2016)
7. Czumaj, A., Sohler, C.: Sublinear-time approximation algorithms for clustering via random sampling. *Random Struct. Algorithms* **30**(1–2), 226–256 (2007)
8. Meyerson, A., O’callaghan, L., Plotkin, S.: A k -median algorithm with running time independent of data size. *Mach. Learn.* **56**(1–3), 61–87 (2004)