Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computer and System Sciences 72 (2006) 890–921
JOURNAL OF COMPUTER AND SYSTEM SCIENCES
www.elsevier.com/locate/jcss

Ramsey-type theorems for metric spaces with applications to online problems ☆

Yair Bartal [a,*,1], Béla Bollobás [b], Manor Mendel [a,2]

[a] Hebrew University, Jerusalem, Israel
[b] University of Memphis, Memphis, TN 38152, USA

Received 1 February 2002; received in revised form 2 May 2004
Available online 28 February 2006

Abstract

A nearly logarithmic lower bound on the randomized competitive ratio for the *metrical task systems* problem is presented. This implies a similar lower bound for the extensively studied K-server problem. The proof is based on Ramsey-type theorems for metric spaces, that state that every metric space contains a large subspace which is approximately a *hierarchically well-separated tree* (and in particular an ultrametric). These Ramsey-type theorems may be of independent interest.

Keywords: Metric task systems; Online servers problem; Metric Ramsey theory

☆ A preliminary version, entitled "A Ramsey-type theorem for metric spaces and its applications for metrical task systems and related problems", appeared in Proceedings of the 42nd Annual Symposium on Foundations of Computer Science, 2001.

* Corresponding author.
E-mail addresses: yair@cs.huji.ac.il (Y. Bartal), bollobas@msci.memphis.edu (B. Bollobás), mendelma@uiuc.edu, mendelma@yahoo.com (M. Mendel).

[1] Supported in part by a grant from the Israeli Science Foundation (195/02).

[2] Work mostly done while the author was a PhD student in Tel-Aviv University, under the supervision of Prof. A. Fiat. Author's current address: Department of Computer Science, University of Illinois, Urbana, IL 61801, USA. Supported in part by a grant from the Israeli Science Foundation (195/02).

0022-0000/$ – see front matter
doi:10.1016/j.jcss.2005.05.008

1. Introduction

This paper deals with the analysis of the performance of randomized online algorithms in the context of two fundamental online problems—metrical task systems and the K-server problem.

A *metrical task system* (MTS), introduced by Borodin, Linial, and Saks [BLS92], is a system that may be in one of a set of n internal states. The aim of the system is to perform a given sequence of *tasks*. The performance of each task has a certain cost that depends on the task and the state of the system. The system may switch states; the cost of such a switch is the distance between the states in a metric space defined on the set of states. After a switch, the cost of the service is the one associated with the new state.

In the *K-server problem*, defined by Manasse, McGeoch, and Sleator [MMS90], K mobile servers reside in points of a given metric space. A sequence of requests for points in the space is presented to the servers. To satisfy a request, one of the K servers must be moved to the point associated with the request. The cost of an algorithm for serving a sequence of requests is the total distance traveled by the servers.

An online algorithm receives requests one by one and must serve them immediately without knowledge of future requests. A randomized online algorithm is called *r-competitive* if on every sequence its expected cost is at most r times the optimal offline cost plus an optional constant additive term.

The MTS and K-server problems have been studied extensively with the aim of determining the best competitive ratio of online algorithms. Borodin et al. [BLS92] have shown that the deterministic competitive ratio for MTS on an n-point metric space is exactly $2n-1$. Manasse et al. [MMS90] proved a lower bound of K on the competitive ratio of deterministic K-server algorithms. The best upper bound for arbitrary metric spaces and any K is currently $2K-1$ [KP95].

The randomized competitive ratio for these problems is not as well understood. For the uniform metric space, where all distances are equal, the randomized competitive ratio is known to within a constant factor, and is $\Theta(\log n)$ [BLS92,IS98] for MTS and $\Theta(\log K)$ [FKL+91,MS91,ACN00] for the K-server problem. In fact, it has been conjectured that, in any metric space, the randomized competitive ratio is $\Theta(\log n)$ for MTS and $\Theta(\log K)$ for the K-server problem. Previous lower bounds were $\Omega(\log\log n)$ [KRR94], and $\Omega(\sqrt{\log n/\log\log n})$ [BKRS00] for MTS and similar lower bounds for the K-server problem in metric spaces with more than K points. The upper bound for MTS was improved in a sequence of papers [Bar96,BBBT97,Bar98,FM03,BM03,FRT03], and is currently $O(\log^2 n\log\log n)$. The upper bound for MTS implies a similar bound for the K-server problem on $K+c$ points, when c is a constant. However, no "general" randomized upper bound for the K-server problem better than $2K-1$ [KP95] is currently known. Seiden [Sei01] has a result in this direction, showing sublinear bounds for certain spaces with certain number of servers.

In this paper we give lower bounds on the randomized competitive ratios that get closer to the conjectured bounds. We prove that, in any n-point metric space, the randomized competitive ratio of the MTS problem is $\Omega(\log n/\log^2\log n)$. For the K-server problem, we prove that the randomized competitive ratio is $\Omega(\log K/\log^2\log K)$ for any metric space on more than K points. Slightly better bounds are obtained for specific metric spaces such as ℓ-dimensional meshes. We also prove for any $\varepsilon > 0$, a lower bound of $\Omega(\log K)$ for the K-server problem in any n-point metric space where $n \geqslant K^{\log^{\varepsilon} K}$, improving a lower bound from [KRR94] of $\Omega(\min\{\log K, \log\log n\})$. We note that the improved lower bounds for the K-server problem also imply improved lower bounds for the *distributed paging problem*, as shown in [ABF93]. The lower bounds for the K-server problem follow from a general reduction from MTS on a metric space of $K + 1$ points to the K-server problem in the same metric space. The rest of the discussion is therefore in terms of the MTS problem.

In [KRR94,BKRS00,Bar96] it is observed that the randomized competitive ratio for MTS is conceptually easier to analyze on "decomposable spaces," spaces that are composed of subspaces with small diameter compared to that of the entire space. Bartal [Bar96] introduced a class of decomposable spaces he called *hierarchically well-separated trees* (HST). A k-HST is a metric space defined on the leaves of a tree such that, for each level of the tree, the diameters of the subtrees decrease by a factor of k between the levels. Consider a particular level of an HST. The distances to all subtrees are approximately the same and thus it is natural to use a recursive solution for the HST where the problem at a particular level is essentially on a uniform space.

In order to analyze the competitive ratio for a specific metric space M, it is helpful to consider how close it is to a simpler metric space N (such as HST). We say that N α-approximates M if the distances in N are within a factor α from those in M. Clearly, if there is a r-competitive algorithm for N then there is αr-competitive algorithm for M. This notion can be generalized to a probabilistic metric approximation [Bar96] by considering a set of approximating metric spaces that dominate the original metric space and bounding the expectation of the distances. Any metric space on n points can be $O(\log n)$-*probabilistically approximated* by HSTs [Bar96,Bar98,FRT03], thus reducing the problem of devising algorithm for MTS on any metric space to devising an algorithm for HSTs only [BBBT97,FM03]. HSTs and their probabilistic approximation of metric spaces have found many other applications in online and approximation algorithms, for example [Bar96, AA97,KT99]. See [Ind01, Sections 2.4 and 5] for a survey on this topic.

The first step toward obtaining a lower bound for arbitrary metric spaces is showing that a lower bound for HSTs implies a lower bound for arbitrary metric spaces. Probabilistic approximations are not useful for this purpose. One of the reasons for this is that the approximation bound is at least logarithmic, and therefore a logarithmic lower bound for HSTs would not imply any non-trivial lower bound for arbitrary metrics. What makes the reduction in this paper possible is the observation that a lower bound for a subspace implies a lower bound for the entire space. Therefore, in order to get a lower bound for a specific metric space M, we need to find a large subspace which is a good approximation of an HST. Such theorems are called *Ramsey-type theorems for metric spaces* [KRR94]. The main Ramsey-type theorem in this paper states that in any metric space on n points there exists a subspace of size $n^{\Omega(\log^{-1} k)}$ points that $O(\log\log n)$-approximates a k-HST. In fact, we further show that the approximated k-HST can have the additional property that any internal vertex of the underlying tree of the HST, either has only two children or all the children's subtrees are of almost equal size (in terms of the number of leaves). It is worth noting that HSTs are ultrametrics and thus embed isometrically in ℓ_2. Therefore, our Ramsey-type theorems give subspaces in Euclidean space. Previously, Bourgain et al. [BFM86], Karloff et al. [KRR94] and Blum et al. [BKRS00] proved other Ramsey-type theorems, showing the existence of special types of HSTs on significantly smaller subspaces. In Section 7 we elaborate on these results and relate our constructions to their constructions. Subsequent work is discussed in Section 1.1. Different Ramsey-type problems for metric spaces appear in [Mat92].

The lower bound for HST spaces follows a general framework originated in [BKRS00] and explicitly formulated in [Sei99,BBBT97]: The recursive structure of the HST is modeled via the *unfair metrical task system* (UMTS) problem [Sei99,BBBT97] on a uniform metric space. This concept is presented in greater details in Section 2. For readers already familiar with the concept, the rest of this paragraph provides a brief summary of our results for this model. In a UMTS problem, every point v_i of the metric space is associated with a *cost ratio* r_i which multiplies the online costs for processing tasks in that point. Offline costs remain as before. The cost ratio r_i roughly corresponds to the competitive ratio of the online algorithm in a subspace of the HST. We prove a lower bound for the randomized competitive ratio of a UMTS on the uniform metric space *for the entire range of cost ratios* $(r_i)_i$, $r_i \geqslant 1$. This lower bound is tight assuming the conjectured $\Theta(\log n)$ competitive ratio for MTS. Previously, tight lower bounds (in the above sense) were only known for two point spaces [BKRS00,Sei99] and fair MTS, where $r_1 = r_2 = \cdots = r_n = 1$ [BLS92]. *Upper bounds* for UMTS problems were given for two point spaces [BKRS00,Sei99,BBBT97] and when all the cost ratios are equal $r_1 = r_2 = \cdots = r_n$ [BBBT97]. Our lower bound matches these upper bounds in these cases.

By making use of the lower bounds for UMTSs on uniform metric spaces, we compose lower bounds to obtain a lower bound of $\Omega(\log n)$ for HSTs. Our main lower bound result follows from the lower bound on HST and the Ramsey-type theorem.

1.1. Subsequent work

Subsequent to this paper, metric Ramsey problems have been further studied in a sequence of papers [BLMN04a,BLMN03b,BLMN04b,BLMN03a]. The main theorem in [BLMN04a] states that any n-point metric space contains a subspace of size $n^{1-(c\log\alpha)/\alpha}$ which α-approximates a 1-HST for $\alpha > 2$ and an appropriate value $c > 0$. Since a 1-HST is equivalent to an ultrametric which isometrically embeds in ℓ_2 this theorem gives nearly tight Ramsey-type theorem for embedding metrics spaces in Euclidean space. The proof of the theorem uses techniques developed in this paper, but is more involved and requires new ingredients as well.

It follows from [BLMN04a] that the main Ramsey theorem in this paper (Theorem 2) can be improved to the following: there exists $c > 0$ such that any n-point metric space contains a subset of size $n^{c/\log(2k)}$ which 3-approximates a k-HST. Together with the lower bounds for k-HSTs in the current paper (Theorem 3), the lower bound on the randomized competitive ratio for the MTS problem on n-point metric space (Theorem 1) is improved to $\Omega(\log n/\log\log n)$, and the lower bound for the K-server problem is improved to $\Omega(\log K/\log\log K)$. Also, the results of Section 7 are complemented in [BLMN03b], where tight bounds on these metric Ramsey problems are given.

Also related is work on multi-embeddings [BM03] which studies a concept that can be viewed as dual to the Ramsey problem. In a multi-embedding a metric space is embedded in a larger metric space where points embed into multiple points. That paper also uses techniques very similar to the ones developed here. It is shown there how this concept can be used to obtain upper bounds for the MTS problem.

1.2. Outline of the paper

In Section 2 the problems and the main concepts are formally defined along with an outline of the proof of the lower bound. Section 3 is devoted to present our main Ramsey-type theorem for metric spaces. In Section 4 we prove a lower bound for UMTSs on a uniform metric spaces, and use it in Section 5 to deduce a lower bound for HSTs. In Section 6 we apply these lower bounds to the K-server problem. In Section 7 we discuss additional Ramsey-type theorems and tight examples. We also relate our work to previous known constructions. Finally, in Section 8, we present a number of open problems that arise from the paper.

2. Overview and definitions

In this section we outline the proof of the lower bounds for the metrical task systems problem on arbitrary metric spaces. We start with defining the MTS problem.

Definition 1. A metric space $M = (V, d)$ consists of a set of points V and a metric distance function $d : V \times V \to \mathbb{R}^+$ such that d is symmetric, satisfies the triangle inequality and $d(u, v) = 0$ if and only if $u = v$.

For $\alpha > 0$, we denote by αM the metric space obtained from M by scaling the distances in M by a factor α.

Definition 2. A *metrical task system* (MTS) [BLS92] is a problem defined on a metric space $M = (V, d_M)$ that consists of $|V| = b$ points, $v_1, \ldots, v_b$. The associated online problem is defined as follows. Points in the metric space represent internal states of an online algorithm A. At each step, the algorithm A occupies a point $v_i \in M$. Given a task, the algorithm may move from v_i to a point v_j in order to minimize costs. A task is a vector $(c_1, c_2, \ldots, c_b) \in (\mathbb{R}^+ \cup \{\infty\})^b$, and the cost for algorithm A associated with servicing the task is $d_M(v_i, v_j) + c_j$. The *cost* for A associated with servicing a sequence of tasks σ, denoted by $\mathrm{cost}_A(\sigma)$, is the sum of costs for servicing the individual tasks of the sequence consecutively. An online algorithm makes its decisions based only upon the tasks seen so far.

We define $\mathrm{cost}_{\mathrm{OPT}}(\sigma)$ to be the minimum cost, for any off-line algorithm, to start at the initial state and process σ. A randomized online algorithm A for an MTS is an online algorithm that decides upon the next state using a random process. The expected cost of a randomized algorithm A on a sequence σ is denoted by $\mathrm{E}[\mathrm{cost}_A(\sigma)]$.

Definition 3. [ST85,KMRS88,BDBK+94] A randomized online algorithm is called r-*competitive against an oblivious adversary* if there exists a constant c such that for every task sequence σ, $\mathrm{E}[\mathrm{cost}_A(\sigma)] \leqslant r \cdot \mathrm{cost}_{\mathrm{OPT}}(\sigma) + c$.

The main result of this paper is the following theorem.

Theorem 1. *Given a metric space M on n points, the competitive ratio* (*against oblivious adversaries*) *of any randomized online algorithm for the MTS defined on M, is at least* $\Omega(\log n/\log^2 \log n)$.

We first observe the fact that a lower bound for a subspace of M implies a lower bound for M. Therefore if we have a class of metric spaces $\mathcal{S}$ for which we have a lower bound we can get a lower bound for a metric space M if it contains a metric space, $M' \in \mathcal{S}$ as a subspace. This may also be done if the subspace approximates the metric space M'.

Definition 4. A metric space M over V α-approximates a metric space M' over V if for all $u, v \in V$, $d_{M'}(u, v) \leqslant d_M(u, v) \leqslant \alpha d_{M'}(u, v)$.

Note that Definition 4 is essentially symmetric in a sense that if M α-approximates M', then M' α-approximates $\alpha^{-1}M$.

Proposition 1. *Given a metric space M on V that α-approximates a metric space M' on V, a lower bound of r' for the MTS on M' implies a lower bound of r'/α for the MTS on M.*

Proof. Assume there exists an r-competitive algorithm A for M. Let A' be the algorithm that simulates A on M'. Let B' be an optimal algorithm for M' and let B be its simulation in M. Then

$$\mathrm{E}\big[\mathrm{cost}_{A'}(\sigma)\big] \leqslant \mathrm{E}\big[\mathrm{cost}_A(\sigma)\big] \leqslant r\,\mathrm{cost}_B(\sigma) + c \leqslant \alpha r\,\mathrm{cost}_{B'}(\sigma) + \alpha c.$$

Therefore A' is αr competitive. □

Next, we define the class of metric spaces for which we will construct lower bounds for the MTS problem. Following Bartal [Bar96],[3] we define the following class of metric spaces.

Definition 5. For $k \geqslant 1$, a *k-hierarchically well-separated tree* (k-HST) is a metric space defined on the leaves of a rooted tree T. To each vertex $u \in T$ there is associated a label $\Delta(u) \geqslant 0$ such that $\Delta(u) = 0$ if and only if u is a leaf of T. The labels are such that if a vertex u is a child of a vertex v then $\Delta(u) \leqslant \Delta(v)/k$. The distance between two leaves $x, y \in T$ is defined as $\Delta(\mathrm{lca}(x, y))$, where $\mathrm{lca}(x, y)$ is the least common ancestor of x and y in T. Clearly, this function is a metric on the set of vertices. We call a vertex with exactly one child, a *degenerate vertex*. For a non-degenerate vertex u, $\Delta(u)$ is the diameter of the

[3] The definition given here for k-HST differs slightly from the original definition in [Bar96]. For $k > 1$ the metric spaces given by these two definitions approximate each other to within a factor of $k/(k-1)$.

subspace induced on the subtree rooted by u. Any k-HST can be transformed into a k-HST without degenerate vertices and with the same metric.

Any k-HST is also a 1-HST. We use the term *HST* to denote any 1-HST. An HST is usually referred to as *ultrametric*, but note that for $k > 1$, a k-HST is a stronger notion.

In Section 3 we prove a generalized form of the following Ramsey-type theorem for metric spaces.

Theorem 2. *Given a metric space $M = (V, d)$ on $|V| = n$ points and a number $k \geqslant 2$, there exists a subset $S \subseteq V$ such that $|S| \geqslant n^{\Omega(1/\log k)}$ and the metric space (S, d) $O(\log\log n)$-approximates a k-HST.*

It follows that it suffices to give lower bounds for the MTS problem on HST metric spaces. A natural method for doing that is to recursively combine lower bounds for subspaces of the HST into a lower bound for the entire metric space. Consider an internal vertex u at some level of the HST. Let $v_1, v_2, \dots, v_b$ be its children and assume we have lower bounds of $r_1, r_2, \dots, r_b$ on the competitive ratio for the subspaces rooted at the v_is. We would like to combine the lower bounds for these subspaces into a lower bound of r for the subspace rooted at u. Recall that the distances between points in the subspaces associated with different v_is are equal to $\Delta(u)$. We would like to think of such a subspace rooted at v_i as being replaced by a single point and the subspace rooted at u being a uniform metric space. Given a task of cost δ at the point associated with the subspace rooted at v_i the cost charged to the online algorithm is at least $r_i\delta$. Informally speaking, given a lower bound for this metrical task system with unfair costs on a uniform metric space we can obtain a lower bound for the subspace rooted at u. This serves as a motivation for the following definition.

Definition 6. [BKRS00,Sei99,BBBT97] An *unfair metrical task system* (UMTS) $U = (M; r_1, \dots, r_b; s)$ consists of a metric space M on b points, $v_1, \dots, v_b$, with a metric d_M, a sequence of *cost ratios* $r_1, r_2, \dots, r_b \in \mathbb{R}^+$, and a *distance ratio* $s \in \mathbb{R}^+$. For $s = 1$, we omit the parameter s from the notation.

The UMTS problem differs from the regular MTS problem in that the cost of the online algorithm for servicing a task $(c_1, c_2, \dots, c_b)$ by switching from v_i to v_j is $s \cdot d_M(v_i, v_j) + r_j c_j$, whereas the offline cost remains as before.

Observation 2. It is sufficient to analyze UMTSs with distance ratio equals one since a UMTS $U = (M; r_1, \dots, r_b; s)$ has a competitive ratio r if and only if $U' = (M; r_1 s^{-1}, \dots, r_b s^{-1}; 1)$ has a competitive ratio rs^{-1}. This is so since the adversary costs in U and U' are the same, whereas the online costs in U are s times larger than in U'.

Our goal is to obtain lower bounds for the UMTS problem on a uniform metric space (where all distances between different points are equal). Consider attempting to prove an $\Omega(\log n)$ lower bound for fair MTS problem on HST metric. If our abstraction is correct, it is reasonable to expect that for UMTS $U = (\mathcal{U}_b^{\Delta}; r_1, \dots, r_b)$ (where $\mathcal{U}_b^{\Delta}$ is the uniform metric space on b points with distance Δ), if $r_i \geqslant c \log n_i$ then there is a lower bound of r

for U such that $r \geqslant c\log(\sum_i n_i)$. Indeed we prove such a claim in Section 4 (Lemma 13). In Section 5 we combine the lower bounds for the uniform UMTS and obtain an $\Omega(\log n)$ lower bound on a k-HST along the lines outlined above. In order to avoid interference between the levels this applies only for $k = \Omega(\log^2 n)$. We prove

Theorem 3. *Given an $\Omega(\log^2 n)$-HST M on n points, the competitive ratio (against oblivious adversaries) of any randomized online algorithm for the MTS defined on M, is at least $\Omega(\log n)$.*

Theorems 3, 2 and Proposition 1 imply a lower bound of $\Omega(\log n/\log^2\log n)$ for any metric space, which concludes Theorem 1.

3. Ramsey-type theorems for metric spaces

Lemma 3. *Given a metric space $M = (V, d)$ on $|V| = n$ points, and $\beta > 1$, there exists a subset $S \subseteq V$, such that $|S| \geqslant n^{1/\beta}$ and (S, d) $O(\log_\beta \log n)$-approximates a 1-HST.*

Proof. We may assume that $\beta \leqslant \log n$ and $n > 2$ (otherwise the claim is trivial). Let Δ be the diameter of M, and let $t = \lceil \log_\beta \log n + 1\rceil$. Choose an endpoint of the diameter $x \in M$. Define a series of sets $A_i = \{y \in M \mid d(x, y) \leqslant \Delta(i/(2t+1))\}$, for $i = 0, 1, 2, \ldots, 2t+1$, and "shells" $S_0 = \{x\}$, $S_i = A_i \setminus A_{i-1}$. Choose S_i, $1 \leqslant i \leqslant 2t$, and delete it. Denote by $B = A_{i-1}$ and $C = V \setminus A_i$. The root of the 1-HST is associated with label $\Delta/(2t+1)$, the two subtrees are built recursively by applying the same procedure on $(B, d|_B)$ and $(C, d|_C)$. Let S be the resulting set of points that are left at the end of the recursive process. Since distances in the 1-HST are at most $1/(2t+1)$ smaller than those in M we get that the subspace (S, d) indeed $2t+1 = O(\log_\beta \log n)$-approximates the resulting 1-HST. We are left to show how to choose i such that $|S| \geqslant n^{1/\beta}$.

Let $\varepsilon_i = A_i/n$. Note that $n^{-1} \leqslant \varepsilon_0 \leqslant \varepsilon_{2t} \leqslant 1 - n^{-1}$. Without loss of generality we may assume that $\varepsilon_t \leqslant 1/2$, since otherwise we may consider the sequence $A'_i = V \setminus A_{2t-i}$ and $\varepsilon'_i = 1 - \varepsilon_{2t-i}$. After deleting S_i we are left with two subspaces, $|B| = \varepsilon_{i-1}n$, and $|C| = (1-\varepsilon_i)n$. Inductively, assume that the recursive selection leaves at least $(\varepsilon_{i-1}n)^{1/\beta}$ points in B and at least $((1-\varepsilon_i)n)^{1/\beta}$ points in C. So $|S| \geqslant (\varepsilon_{i-1}^{1/\beta} + (1-\varepsilon_i)^{1/\beta})n^{1/\beta}$ points. To finish the proof it is enough to show the existence of i_0 for which $\varepsilon_{i_0-1}^{1/\beta} + (1-\varepsilon_{i_0})^{1/\beta} \geqslant 1$. If exists $0 \leqslant i_0 < t$ for which $\varepsilon_{i_0-1} \geqslant \varepsilon_{i_0}^\beta$ then $\varepsilon_{i_0-1}^{1/\beta} + (1-\varepsilon_{i_0})^{1/\beta} \geqslant \varepsilon_{i_0} + (1-\varepsilon_{i_0}) = 1$ and we are done. Otherwise, we have that $\varepsilon_{i-1} < \varepsilon_i^\beta$ for all $0 \leqslant i < t$, and since $\varepsilon_t \leqslant 1/2$, we conclude by induction on i that $\varepsilon_i \leqslant (1/2)^{\beta^{(t-i)}}$. But then

$$\varepsilon_0 \leqslant \left(\frac{1}{2}\right)^{\beta^t} < \left(\frac{1}{2}\right)^{\beta^{\log_\beta \log n}} = \frac{1}{n},$$

which contradicts $\varepsilon_0 \geqslant 1/n$. □

We also need the following lemma from [Bar98].

Lemma 4. *For any $\ell > 1$, any 1-HST ℓ-approximates some ℓ-HST.*[4]

Proof *(sketch).* Let T be a 1-HST. We construct a ℓ-HST by incrementally changing T as follows. Scan the vertices of T in top–down fashion. For any non-root vertex v, and its father u, if $\Delta(v) \geqslant \Delta(u)/\ell$ then delete v and connects v's children directly to u. The resulting tree is clearly an ℓ-HST and a ℓ-approximation of T. □

Next, we show how to prune an ℓ-HST on n leaves, to get a subtree which is a k-HST with $n^{1/\lceil \log_\ell k \rceil}$ leaves. This follows from the following combinatorial lemma for arbitrary rooted trees. Recall that a vertex in a rooted tree is called non-degenerate if the number of its children is not one.

Definition 7. A rooted tree is *h-sparse* if the number of edges along the path between any two non-degenerate vertices is at least h.

Lemma 5. *Given a rooted tree T on n leaves, there exists a subtree T' with at least $n^{1/h}$ leaves that is h-sparse.*

Proof. For a tree T and $i \in \{0, 1, \ldots, h-1\}$ let $f_i(T)$ be the maximum number of leaves in h-sparse subtree of T for which any vertex of depth less than i has out degree at most one. Clearly $f_0(T) = \max_i f_i(T)$.

We prove by induction on the height of T that $\prod_{i=0}^{h-1} f_i(T) \geqslant n$, and thus $f_0(T) \geqslant n^{1/h}$. The base of the induction is a tree T of height 0, for which $f_i(T) = 1$ for any i, as required. For T with height at least 1, denote by $\{T_j\}_j$ the subtrees of T rooted at the children of the root of T. Assume that T_j has n_j leaves, and $n = \sum_j n_j$. One possible way to obtain an h-sparse subtree of T would be to include the root in the tree and the union of the solutions of $f_{h-1}(T_j)$. Therefore

$$f_0(T) \geqslant \sum_j f_{h-1}(T_j).$$

Consider the case $i > 0$. Let v_j be a child of the root and let T_j be the subtree rooted at v_j. Let S_j be an h-sparse subtree of T_j, with maximum number of leaves, for which any vertex of depth less than $i-1$ has out degree at most one. Construct a subtree S by concatenating the edge from the root to v_j with the subtree S_j. This results in an h-sparse subtree of T for which any vertex of depth less than i has out degree at most one. Hence

$$f_i(T) = \max_j f_{i-1}(T_j) \quad \forall i \in \{1, \ldots, h-1\}.$$

Thus

$$\prod_{i=0}^{h-1} f_i(T) \geqslant \left(\sum_j f_{h-1}(T_j)\right) \cdot \prod_{i=1}^{h-1} \max_j f_{i-1}(T_j) \geqslant \sum_j \left(f_{h-1}(T_j) \cdot \prod_{i=1}^{h-1} f_{i-1}(T_j)\right)$$

[4] In [Bar98], 1-HST is referred as "hierarchical partition metric."

$$= \sum_j \prod_{i=0}^{h-1} f_i(T_j) \geqslant \sum_j n_j = n.$$

The last inequality follows from the induction hypothesis. □

Lemma 6. *Given a* 1*-HST N on n points there exists a subspace of N on $n^{1/\lceil \log_\ell k \rceil}$ points which ℓ-approximates a k-HST.*

Proof. As a first step we construct, using Lemma 4, an ℓ-HST M that is ℓ approximated by N.

Let $h = \lceil \log_\ell k \rceil$. Let T be the underlying tree of M. Applying Lemma 5 on T we get a subtree S of T which is h-sparse. Let S' be the tree resulting from coalescing pairs of edges with a common degenerate vertex in S. Consider the metric space M' defined on the leaves of S' with the associated labels. Clearly, M' is a subspace of M. Consider any internal node u in S' and let v be a child of u in S'. If v is a leaf then $\Delta(v) = 0$. Otherwise both u and v are non-degenerate and therefore the number of edges on the path in T between u and v is at least h. This implies that $\Delta(u)/\Delta(v) \geqslant \ell^h \geqslant k$. Thus M' is a k-HST. □

Theorem 4. *For any metric space $M = (V, d)$ on $|V| = n$ points, any $\beta > 1$, any $k > 1$, and any $1 < \ell \leqslant k$ there exists a subset $S \subseteq V$, such that $|S| \geqslant n^{1/(\beta \lceil \log_\ell k \rceil)}$ and (S, d) $O(\ell \log_\beta \log n)$-approximates a k-HST.*

Proof. Given a metric space M on n points, from Lemma 3, we get a subspace of M with $n^{1/\beta}$ points that $O(\log_\beta \log n)$ approximates an 1-HST S. We then apply Lemma 6 to obtain a subspace of S on $n^{1/(\beta \lceil \log_\ell k \rceil)}$ points which $O(\ell \log_\beta \log n)$-approximates a k-HST. □

Theorem 2 is a corollary of Theorem 4 when substituting $\beta = \ell = 2$.

As discussed in Section 1.1, Lemma 3 has been recently improved in [BLMN04a]. Lemma 6 is tight as shown in Proposition 29. Furthermore we show in Proposition 26 that in order to get a Ramsey-type theorem with a constant approximation for HSTs, the subspace's size must be at most n^c for some constant $c \in (0, 1)$.

For specific metric spaces, better approximations are possible. Here we consider the ℓ-dimensional mesh. The result is based on the Gilbert–Varshamov bound from coding theory (see [MS77, Chapter 17, Theorem 30]).

Lemma 7 *(Gilbert–Varshamov bound). For any $h \in \mathbb{N}$, and $\alpha \in (0, 0.5)$, there exists a binary code $C \subset \{0, 1\}^h$ on h-bit words such that the minimum Hamming distance between any two codewords is at least αh, and $|C| \geqslant 2^{h(1-H_2(\alpha))}$, where $H_2(x) = -(x \log_2 x + (1 - x)\log_2(1 - x))$ is the binary entropy.*

Lemma 8. *Given an h-dimensional mesh $M = [s]^h = \{0, 1, \ldots, s-1\}^h$ with the ℓ_p-norm ($p \in [1, \infty]$) on $n = s^h$ points. Then, there exists a subspace $S \subset [s]^d$ that* 12*-approximates a* 9*-HST, and $|S| \geqslant n^c$ for a constant $c = 0.08 \log_9 2$.*

Proof. We construct an HST T by induction on s. For $s = 1$, T is simply one point.

For $s > 1$ we construct T as follows. Fix $\alpha = 1/3$. By Lemma 7, there exists an h-bit binary code C with a minimum Hamming distance of $h/3$, and $|C| \geqslant 2^{h(1-H_2(1/3))} \geqslant 2^{0.08h}$. For each codeword $w = (a_1, \ldots, a_h) \in C$, we choose a submesh of size $\lceil s/9 \rceil^h$ with a corner located at $(s-1)w$. More specifically,

$$S_w = \left(a_1\left(s - \left\lceil \frac{s}{9} \right\rceil\right) + \left[\left\lceil \frac{s}{9} \right\rceil\right]\right) \times \left(a_2\left(s - \left\lceil \frac{s}{9} \right\rceil\right) + \left[\left\lceil \frac{s}{9} \right\rceil\right]\right) \times \cdots \times \left(a_h\left(s - \left\lceil \frac{s}{9} \right\rceil\right) + \left[\left\lceil \frac{s}{9} \right\rceil\right]\right),$$

where for a set of numbers Y and a number x, $x + Y = \{x + y \mid y \in Y\}$ is the Minkowski sum.

Let $x \in S_w$ and $y \in S_{w'}$ where $w, w' \in C$ and $w \neq w'$. Obviously $d(x, y) \leqslant \sqrt[p]{h}\,(s-1)$, but also, by the triangle inequality, $d(x, y)$ is at least

$$\begin{aligned} d(x, y) &\geqslant \Big(\max_{a \in S_w,\, b \in S_{w'}} d(a, b)\Big) - \operatorname{diam}(S_w) - \operatorname{diam}(S_{w'}) \\ &\geqslant \sqrt[p]{\frac{h}{3}}\,(s-1) - 2\sqrt[p]{h}\left(\left\lceil \frac{s}{9} \right\rceil - 1\right) \\ &\geqslant \begin{cases} \frac{\sqrt[p]{h}}{3}(s-1) - 2\sqrt[p]{h}\frac{s}{9} \geqslant \frac{\sqrt[p]{h}\,(s-3)}{9} \geqslant \frac{\sqrt[p]{h}\,(s-1)}{12}, & s \geqslant 9, \\ \frac{\sqrt[p]{h}}{3}(s-1), & 2 \leqslant s < 9. \end{cases} \end{aligned}$$

Hence, the distances between points in different subspaces are approximately the same, up-to a factor of 12. T has a root labeled with $\sqrt[p]{h}(s-1)/12$. Its children correspond to the subspaces S_w for $w \in C$. For each subspace an HST is constructed inductively with $s \leftarrow \lceil s/9 \rceil$.

From the construction, T is a 9-HST, and from the previous discussion, the distances in T are 12 approximated by the original distances in the mesh. T is also a complete and balanced tree. Its height is at least $\log_9 s$, and the out-degree of each internal vertex is $|C|$. Hence, the number of leaves in T is at least $|C|^{\log_9 s} \geqslant 2^{0.08h \log_2 s \log_9 2} = n^c$. $\square$

In Proposition 26 we show the above lemma to be tight.

4. Lower bounds for uniform UMTS

Our goal is to construct a lower bound on HSTs. This is done in the next section by combining lower bounds for subtrees of the HST, using a lower bound for a corresponding unfair MTS problem on a uniform metric space. In this section we formally define the type of the lower bounds we use, and prove such a lower bound for the uniform metric space.

Our lower bounds are based on Yao's principle (Theorem 5), by which adversaries produce a distribution over sequences against deterministic algorithms. However, since the adversaries for (sub)spaces would be part of a larger adversary, we need to be more careful about their structure. In particular, since the expected cost of the adversary on the distribution would serve as the task for UMTS abstracting a higher level view of the space, and since the lower bounds for UMTS rely crucially on the tasks being relatively small, we need to maintain upper bounds on the expected cost of the optimal offline algorithm. We formalize it in the following definitions.

Given an algorithm A_U for UMTS $U=(M;r_1,\dots,r_b)$, define $\mathrm{cost}_{A_U}(\sigma,u_0)$ to be the cost of A_U on the task sequence σ when starting from point $u_0\in M$. Let OPT^0 be the optimal offline algorithm for servicing a task sequence and returning to the starting point. An elementary task (v,δ), where $v\in M$, is a task that assigns cost δ to the point v and 0 to every other point. Our lower bound argument uses only elementary tasks.

Definition 8. Given a UMTS $U=(M;r_1,\dots,r_b)$ on a metric space with diameter $\Delta>0$, define an (r,β)-*adversary* $\mathcal{D}$ to be a distribution on finite elementary task sequences for U such that

- $\min_{u_0\in M}\mathrm{E}_{\sigma\in\mathcal{D}}[\mathrm{cost}_{\mathrm{OPT}^0}(\sigma,u_0)]\leqslant\beta\Delta$;
- for any online algorithm A, $\min_{u_0\in M}\mathrm{E}_{\sigma\in\mathcal{D}}[\mathrm{cost}_A(\sigma,u_0)]\geqslant r\beta\Delta$.

Yao's principle (cf. [BLS92,BEY98]), as applied to (unfair) metrical task systems implies the following result.

Theorem 5. *If there exists an (r,β)-adversary for a UMTS U, then r is a lower bound on the randomized competitive ratio for U against oblivious adversaries.*

Proof. The proof is standard and can be found, e.g., in [BEY98]. The only issue here is to generate a sequence of unbounded cost for the online algorithm. As we can repeatedly and independently sample from the same distribution over and over again, we can make the cost of the online unbounded. Note that the offline costs indeed sum up as required since OPT^0 always return to the same point. □

Our basic adversaries can only use discrete tasks. We formalize it in the following definition.

Definition 9. Given a UMTS $U=(M;r_1,\dots,r_b)$ on a metric space with diameter Δ, an $(r,\beta;\alpha_1,\dots,\alpha_b)$-*discrete adversary* is an (r,β)-adversary that uses only tasks of the form $(v_i,\alpha_i\Delta)$.

Observation 9. For $\gamma>0$, denote by γM a metric in which the distances are scaled by a factor of γ compared to M. A UMTS $U=(M;r_1,\dots,r_b;s)$ and a UMTS $U'=(\gamma M;r_1,\dots,r_b;s)$ have the same competitive ratio. Moreover (r,β)-adversary and $(r,\beta;\alpha_1,\dots,\alpha_b)$-discrete adversary for U are easily transformed into (r,β)-adversary and $(r,\beta;\alpha_1,\dots,\alpha_b)$-discrete adversary (respectively) for U' by scaling the tasks by a factor of γ.

Lemma 10. *There exist constants*[5] $\lambda_3, \rho > 0$ *such that for any UMTS* $U = (\mathcal{U}_b^{\Delta}; r_1, \ldots, r_b)$, $r_1 \geqslant r_2 \geqslant \cdots \geqslant r_b \geqslant 1$ *satisfying* $r_1 \geqslant \frac{1}{4}\ln b$, *there exists an* $(r, \beta; r_1^{-1}, r_2^{-1}, \ldots, r_b^{-1})$*-discrete adversary, where* $\beta \leqslant \lambda_3 r_1$, *and*

$$r \geqslant \rho \ln\left(\sum_{i=1}^{b} e^{\rho^{-1} r_i}\right). \tag{1}$$

Formula (1) is better understood in the following context. Let $n_i = e^{\rho^{-1} r_i}$, where n_i should be thought of as a (lower bound) estimate on the number of points in the subspace that corresponding to v_i, and "generates" a lower bound on the competitive ratio of $r_i = \rho \log n_i$. Let $n = \sum_i n_i$, and thus formula (1) implies a lower bound of $\rho \ln n$ on the competitive ratio for the whole space, represented by U. This is the recursive argument we need in order to prove a $\rho \log n$ lower bound.

Without loss of generality (due to Observation 9), we may assume that $\Delta = 1$. To prove Lemma 10 we use the following distribution. Let m be a parameter to be determined later. A task sequence of length m is generated by repeatedly and independently picking a random point v_i and generating an elementary task (v_i, r_i^{-1}). The expected cost of any online algorithm on this distribution is at least $\mu = m/n$.

We give an upper bound for OPT^0 on this sequence by presenting the following offline algorithm PHASE. PHASE starts at $v_1 \in \mathcal{U}_b^1$. It chooses in hindsight a point v_i, moves to v_i at the beginning of σ, stay there for the entire duration of σ, and at the end returns to v_1. The point v_i is chosen so as to minimize the cost of PHASE, i.e., the local cost on v_i during σ plus zero if $i = 1$ and plus two if $i > 1$. Denote by X_i the number of tasks given to point v_i. Thus the total local cost for v_i is X_i/r_i, and the expected cost of PHASE (which is an upper bound on the cost of OPT^0) is

$$E\left[\min\left\{\frac{X_1}{r_1}, 2 + \min_{i \geqslant 2} \frac{X_i}{r_i}\right\}\right]. \tag{2}$$

The analysis of formula (2) is rather complicated. Fortunately, in order to prove inequality (1), it is sufficient to establish it in only two cases: when $b = 2$ and when $r_1 = \cdots = r_b$. This is due to the following proposition.

Proposition 11. *Given a non-increasing sequence of positive real numbers* $(n_i)_{i \geqslant 1}$. *Denote by* $n = \sum_i n_i$ *and assume* $n < \infty$. *Then either* $\sqrt{n_1} + \sqrt{n_2} \geqslant \sqrt{n}$, *or there exists* $\ell \geqslant 3$ *such that* $\ell \cdot \sqrt{n_\ell} > \sqrt{n}$.

Proof. We first normalize by setting $x_i = n_i/n$. Thus, $\sum_i x_i = 1$, and we need to prove that either $\sqrt{x_1} + \sqrt{x_2} \geqslant 1$ or there exists $\ell \geqslant 3$ such that $\ell\sqrt{x_\ell} > 1$.

Assume that the second condition does not holds, i.e., $\forall \ell \geqslant 3$, $x_\ell \leqslant \ell^{-2}$. We will prove that $\sqrt{x_1} + \sqrt{x_2} \geqslant 1$. Let $b = \lfloor {x_2}^{-0.5} \rfloor$. We may assume that $x_2 \leqslant 1/4$ (otherwise $\sqrt{x_1} + \sqrt{x_2} \geqslant 1$), and therefore $b \geqslant 2$. Hence

[5] The constant ρ we achieve is quite small. We have made no serious attempt to optimize it, and preferred simplicity whenever possible.

$$\sum_{i=b+1}^{\infty} x_i \leqslant \sum_{i=b+1}^{\infty} i^{-2} \leqslant x_2\bigl(x_2^{-0.5} - b\bigr) + \int_{x_2^{-0.5}}^{\infty} z^{-2}\,dz$$
$$= x_2\bigl(x_2^{-0.5} - b\bigr) + \sqrt{x_2} = 2\sqrt{x_2} - bx_2.$$

So,

$$x_1 = 1 - \sum_{i=2}^{\infty} x_i \geqslant 1 - (b-1)x_2 - \bigl(2\sqrt{x_2} - bx_2\bigr) = 1 - 2\sqrt{x_2} + x_2 = \bigl(1 - \sqrt{x_2}\bigr)^2.$$

That is $\sqrt{x_1} + \sqrt{x_2} \geqslant 1$, as needed. □

Proof of Lemma 10. We will use in the proof some elementary probabilistic arguments. For the sake of completeness, we include their proofs in Appendix A. We derive an upper bound on formula (2) as follows. Fix $\delta_1 \in [0, 1]$, and denote by Y the event "$\exists i \geqslant 2$, $X_i/r_i \leqslant (1-\delta_1)\mu/r_1$," i.e., one of the points in $\{v_2, \ldots, v_n\}$ has a local cost of at most $(1-\delta_1)\mu/r_1$. Let $\hat{p} = \Pr[Y]$. We can bound the cost for PHASE as follows: If Y does not happen, PHASE can stay in v_1, otherwise it moves to v_i with a local cost at most $(1-\delta_1)\mu/r_1$. Hence its cost is at most $(1-\hat{p})E[X_1 \mid \neg Y]/r_1 + \hat{p}(2 + (1-\delta_1)\mu/r_1)$. By Proposition A.5 in Appendix A, $E[X_1 \mid \neg Y] \leqslant E[X_1] = \mu$, and so we derive the following bound.

$$\mathrm{E}\bigl[\mathrm{cost}_{\mathrm{OPT}^0}(\sigma, v_1)\bigr] \leqslant (1-\hat{p})\frac{\mu}{r_1} + \hat{p}\left(\frac{(1-\delta_1)\mu}{r_1} + 2\right) = \frac{\mu}{r_1}(1 - \hat{p}\delta_1) + 2\hat{p}. \tag{3}$$

Assuming $1 \geqslant \delta_1 \geqslant 4r_1/\mu$, we have the following bound

$$\mathrm{E}\bigl[\mathrm{cost}_{\mathrm{OPT}^0}(\sigma, v_1)\bigr] \leqslant \frac{\mu}{r_1}\left(1 - \hat{p}\frac{\delta_1}{2}\right) = \beta.$$

The lower bound on the competitive ratio we achieve is

$$r \geqslant \frac{\mu}{\frac{\mu}{r_1}(1 - \hat{p}\frac{\delta_1}{2})} \geqslant r_1\left(1 + \hat{p}\frac{\delta_1}{2}\right).$$

Clearly, we need a lower bound on $\hat{p}$. We define $p_i = \Pr[X_i \leqslant (1-\delta_1)\mu/r_1]$, and analyze the lower bound in two special cases.

- In case $r_1 = \cdots = r_b$, we have $p_1 = p_2 = \cdots = p_b$. We bound $\hat{p}$ in terms of p_1.

  $$1 - \hat{p} \leqslant (1-p_1)^{b-1} \leqslant \exp\bigl(-(b-1)p_1\bigr) \leqslant 1 - \min\left\{\frac{1}{4}, \frac{1}{2}(b-1)p_1\right\}.$$

  The first inequality follows from Proposition A.4 in Appendix A, and the last inequality follows since $e^{-\tau} \leqslant \max\{0.75, 1 - 0.5\tau\}$ for $\tau \geqslant 0$. Thus $\hat{p} \geqslant \min\{0.25, 0.5(b-1)p_1\}$.
  To bound p_1, we use a lower bound estimate on the tail probability of a binomial variable. Lemma A.1 in Appendix A states that there exist constants $\lambda_2 \geqslant 1 \geqslant \lambda_1 > 0$ such that, $p_1 \geqslant \lambda_1 e^{-\lambda_2 \delta_1^2 \mu}$, provided that $\mu \geqslant 4$. Thus $\hat{p} \geqslant \min\{0.25, 0.5(b-1)\lambda_1 e^{-\lambda_2 \delta_1^2 \mu}\}$.

Fix $\tilde{\mu} = \frac{16r_1^2\lambda_2}{\ln b}$. Note that $\tilde{\mu} \geqslant 4$, since $r_1 \geqslant (\ln b)/4$. We want to set $\mu \approx \tilde{\mu}$, however, we need to maintain $m = n\mu \in \mathbb{N}$, so we choose $\mu = \lceil \tilde{\mu} \rceil \leqslant \frac{5}{4}\tilde{\mu}$. In order to satisfy the constraint on δ_1, we choose $\delta_1 = \frac{4r_1}{\tilde{\mu}} = \frac{\ln b}{4\lambda_2 r_1}$, so $1 \geqslant \delta \geqslant \frac{4r_1}{\mu}$.
Since $\delta_1 = \sqrt{\frac{\ln b}{\tilde{\mu}\lambda_2}}$, we have $\hat{p} \geqslant 0.5\lambda_1 e^{-5/4} \geqslant \lambda_1/8$. Thus, the lower bound we show is

$$r_1 + \hat{p}\frac{\delta_1}{2}r_1 \geqslant r_1 + \frac{\lambda_1}{8}\frac{\ln b}{8\lambda_2 r_1}r_1 = r_1 + \frac{\lambda_1}{64\lambda_2}\ln b \geqslant \lambda \ln\left(be^{\lambda^{-1}r_1}\right),$$

for $\lambda \leqslant \frac{\lambda_1}{64\lambda_2}$. Note that $\beta \leqslant \frac{\mu}{r_1} \leqslant 20\lambda_2 r_1$.

- In case $b = 2$, let $\delta_2 \in [0, 1]$ such that $(1 - \delta_1)\frac{\mu}{r_1} = (1 - \delta_2)\frac{\mu}{r_2}$. We fix $\delta_1 = \frac{r_1 - r_2 + (20\lambda_2)^{-1}}{r_1}$, so

$$\delta_2 = \frac{r_1 - r_2}{r_1} + \delta_1\frac{r_2}{r_1} = \frac{r_1 - r_2}{r_1} + \frac{r_1 - r_2 + (20\lambda_2)^{-1}}{r_1}\frac{r_2}{r_1}$$
$$\leqslant 2\frac{r_1 - r_2 + (20\lambda_2)^{-1}}{r_1}.$$

In order to satisfy the constraint on δ_1, we choose $\mu = \lceil \tilde{\mu} \rceil$, where $\tilde{\mu} = \frac{4r_1}{\delta_1} = \frac{4r_1^2}{r_1 - r_2 + (20\lambda_2)^{-1}}$, so $\mu \leqslant \frac{5}{4}\tilde{\mu}$. In this case, by applying Lemma A.1,

$$\hat{p} = p_2 \geqslant \lambda_1 e^{-\lambda_2\delta_2^2\mu} \geqslant \lambda_1 e^{-16\frac{5}{4}\lambda_2(r_1 - r_2 + (20\lambda_2)^{-1})} = \frac{\lambda_1}{e}e^{-20\lambda_2(r_1 - r_2)}.$$

Assuming $\lambda \leqslant \frac{\lambda_1}{40e\lambda_2}$, the lower bound we show is

$$r_1 + \hat{p}\frac{\delta_1}{2}r_1$$
$$\geqslant r_1 + \frac{\lambda_1}{e}e^{-20\lambda_2(r_1 - r_2)}\frac{r_1 - r_2 + (20\lambda_2)^{-1}}{2r_1}r_1 \geqslant r_1 + \frac{\lambda_1}{40e\lambda_2}e^{-20\lambda_2(r_1 - r_2)}$$
$$\geqslant r_1 + \lambda e^{\lambda^{-1}(r_2 - r_1)} \geqslant r_1 + \lambda\ln\left(1 + e^{\lambda^{-1}(r_2 - r_1)}\right) = \lambda\ln\left(e^{\lambda^{-1}r_1} + e^{\lambda^{-1}r_2}\right).$$

Note that $\beta \leqslant \frac{\mu}{r_1} \leqslant 100\lambda_2 r_1$.

In the general case, let $\rho = \lambda/2$, $n_i = e^{\rho^{-1}r_i}$, and $n = \sum_{i=1}^{b} n_i$. Applying Proposition 11, we get one of the following two possible cases:

- $\exists \ell$ such that $\ell\sqrt{n_\ell} \geqslant \sqrt{n}$. Note that $\min\{\frac{X_1}{r_\ell}, 2 + \min_{i:\, \ell \geqslant i \geqslant 2}\frac{X_i}{r_\ell}\}$ is an upper bound on formula (2). Thus, our lower bound for ℓ equal cost ratios ($= r_\ell$) applies here, and we get a lower bound of

$$\lambda\ln\left(\ell e^{\lambda^{-1}r_\ell}\right) = \lambda\ln\left(\ell\sqrt{n_\ell}\right) \geqslant \lambda\ln\sqrt{n} = \rho\ln n.$$

- $\sqrt{n_1} + \sqrt{n_2} \geqslant \sqrt{n}$. Again, $\min\{\frac{X_1}{r_1}, 2 + \frac{X_2}{r_2}\}$ is an upper bound on formula (2). Thus, our lower bound for $b = 2$ applies here, and we get a lower bound

$$\lambda \ln\left(e^{\lambda^{-1} r_1} + e^{\lambda^{-1} r_2}\right) = \lambda \ln\left(\sqrt{n_1} + \sqrt{n_2}\right) \geqslant \lambda \ln \sqrt{n} = \rho \ln n.$$

We conclude that the claim is proved with the constants $\lambda_3 = 100\lambda_2$, and $\rho = \frac{\lambda_1}{2\cdot 64 e \lambda_2}$. $\square$

Still, Lemma 10 requires $r_1 \geqslant \frac{1}{4}\ln b$. For a small r_1 we use the standard (fair) MTS lower bound.[6]

Lemma 12. *For a UMTS $U = (\mathcal{U}_b^{\Delta}; 1, \ldots, 1)$ there exists an $(\frac{H_b}{2}, 2; 1, 1, \ldots, 1)$-discrete adversary, where $H_n = \sum_{i=1}^{b} i^{-1}$.*

Proof. Without loss of generality, assume $\Delta = 1$. The sequence is determined by a random permutation π of the points in the space. Then, $\sigma = \tau_1 \tau_2 \cdots \tau_b$, where $\tau_i = (v_{\pi(1)}, 1)(v_{\pi(2)}, 1) \cdots (v_{\pi(i)}, 1)$.

Obviously, OPT^0's cost is at most 2, because it can move at the beginning of σ to $v_{\pi(b)}$, and return at the end of σ. The expected cost of the online on the other hand is at least $\frac{1}{b-i+1}$ in τ_i, and thus at least $\sum_{i=1}^{b} i^{-1}$ in σ. $\square$

Lemma 13. *There exist constants λ_3, $\rho > 0$ satisfying the following. Given a UMTS $U = (\mathcal{U}_b^{\Delta}; r_1, \ldots, r_b)$, with $r_1 \geqslant r_2 \geqslant \cdots \geqslant r_b \geqslant 1$, and $(n_i)_i$ satisfying $r_i = \rho(1 + \ln n_i)$, there exists an $(r, \beta; \alpha_1, \ldots, \alpha_b)$-discrete adversary such that $r \geqslant \rho(1 + \ln(\sum_i n_i))$, $\beta \leqslant \lambda_3 r_1$, and $\min_i \alpha_i \geqslant r_b^{-1}$.*

Proof. Let $\rho \leqslant 1/4$ and $\lambda_3 \geqslant 2$ be the constants from Lemma 10. If $r_1 \geqslant \frac{\ln b}{4}$ then the claim follows from Lemma 10.

For $r_1 \leqslant \frac{\ln b}{4}$, we use the adversary from Lemma 12. Thus,

$$r \geqslant 0.5 H_b \geqslant r_1 + \frac{\ln b}{4} \geqslant \rho(1 + \ln n_1) + \rho \ln b \geqslant \rho\left(1 + \ln(n_1 b)\right).$$

Also, $\beta = 2 \leqslant \lambda_3 r_1$ and $\min_i \alpha_i = 1 \geqslant r_b^{-1}$. $\square$

5. Combining adversaries on HSTs

In this section we prove a lemma for combining adversaries for subspaces using the discrete adversary of Lemma 13 as the combining adversary. We then construct adversaries for HSTs by inductively combining adversaries for subtrees. When attempting to combine (r, β) adversaries, we still have the following problem. The adversary of Lemma 13 can only use specific task sizes, but the tasks we have from subtrees' adversaries are not necessarily of these sizes. Our solution is to inductively maintain "flexible" adversaries that

[6] The adversary of Lemma 10 actually works for $r_1 \leqslant \frac{\ln b}{4}$ as well, by choosing $\mu \approx \ln b$, $\delta = 1$, and a simple bound of $\Pr[X = 0] \geqslant 4^{-\mu}$. We choose to present this lower bound using a different adversary, since the analysis is simpler, and the bound is better.

can generate lower bound sequences with associated optimal cost of value that may vary arbitrarily in a specified range.

Definition 10. Given a UMTS $U = (M; r_1, \ldots, r_b)$ an $(r, \beta; \eta)$-*flexible* adversary for $\eta \in [0, 1]$ is defined as a collection $\mathcal{A}$ of (r, β')-adversaries, for all $\beta' \in [\eta\beta, \beta]$.

Definition 11. Given a UMTS U, an $(r, \beta; \eta; \alpha_1, \ldots, \alpha_b)$-*flexible discrete adversary* is a collection $\mathcal{A}$ of discrete adversaries for U such that $\forall \beta' \in [\eta\beta, \beta]$, $\exists (\alpha'_i)_i$ such that $\alpha'_i \geqslant \alpha_i$ and $\mathcal{A}$ includes an $(r, \beta'; \alpha'_1, \ldots, \alpha'_b)$-discrete adversary. Obviously, $\mathcal{A}$ is an $(r, \beta; \eta)$-flexible adversary.

We start by showing how to transform a discrete adversary into a flexible discrete adversary with only a small loss in the lower bound obtained.

Lemma 14. *Denote the UMTSs $U_s = (M; r_1, \ldots, r_b; s)$ with $\Delta(M) = \Delta$, and assume there exists an $(r, \eta\beta; \alpha_1, \ldots, \alpha_b)$ discrete adversary $\mathcal{D}_\eta$ for U_η, then there exists $(r, \beta; \eta; \alpha_1, \ldots, \alpha_b)$ flexible discrete adversary $\mathcal{A}$ for U_1.*

Proof. Denote by $U_{\eta,\alpha} = (\alpha M; r_1, \ldots, r_b; \eta)$ (so $U_\eta = U_{\eta,1}$). $\mathcal{D}_\eta$ is $(r, \eta\beta; \alpha_1, \ldots, \alpha_b)$ discrete adversary for U_η. Observation 9 implies the existence of $(r, \eta\beta; \alpha_1, \ldots, \alpha_b)$ discrete adversary $\mathcal{D}_{\eta,\alpha^{-1}}$ for $U_{\eta,\alpha^{-1}}$ that replaces each task $(v_i, \alpha_i \Delta)$ of $\mathcal{D}_\eta$ with $(v_i, \alpha_i \alpha^{-1} \Delta)$.

Consider the adversary $\mathcal{D}_{\eta,\alpha^{-1}}$, for $\alpha \in [\eta, 1]$ when applied to U_1,

$$\begin{aligned}\min_{u_0 \in M} \mathrm{E}_{\sigma \in \mathcal{D}_{\eta,\alpha^{-1}}}\big[\mathrm{cost}_{\mathrm{OPT}^0_{U_1}}(\sigma, u_0)\big] &\leqslant \min_{u_0 \in M} \mathrm{E}_{\sigma \in \mathcal{D}_{\eta,\alpha^{-1}}}\big[\mathrm{cost}_{\mathrm{OPT}^0_{U_{\eta,\alpha^{-1}}}}(\sigma, u_0)\big] \\ &\leqslant \eta\beta\alpha^{-1}\Delta.\end{aligned}$$

The first inequality follows since the distances in $U_{\eta,\alpha^{-1}}$ are larger than in U_1.

On the other hand, for any online algorithm A_{U_1} for U_1, consider $A_{U_{\eta,\alpha^{-1}}}$ the simulation of A_{U_1} on $U_{\eta,\alpha^{-1}}$. The moving costs for online algorithms in $U_{\eta,\alpha^{-1}}$ are smaller than in U_1, since $\eta\alpha^{-1} \leqslant 1$. So we have,

$$\min_{u_0 \in M} \mathrm{E}_{\sigma \in \mathcal{D}_{\eta,\alpha^{-1}}}\big[\mathrm{cost}_{A_{U_1}}(\sigma, u_0)\big] \geqslant \min_{u_0 \in M} \mathrm{E}_{\sigma \in \mathcal{D}_{\eta,\alpha^{-1}}}\big[\mathrm{cost}_{A_{U_{\eta,\alpha^{-1}}}}(\sigma, u_0)\big] \geqslant r\eta\beta\alpha^{-1}\Delta.$$

Hence, $\mathcal{D}_{\eta,\alpha^{-1}}$ is $(r, \eta\alpha^{-1}\beta)$ adversary for U_1. Note that for v_i, $\mathcal{D}_{\eta,\alpha^{-1}}$ uses the tasks $(v_i, \alpha_i\alpha^{-1}\Delta)$, and thus it is actually $(r, \eta\alpha^{-1}\beta; \alpha^{-1}\alpha_1, \alpha^{-1}\alpha_2, \ldots, \alpha^{-1}\alpha_b)$-discrete adversary for U_1. Thus $\mathcal{A} = \{\mathcal{D}_{\eta,\alpha^{-1}} \mid \alpha \in [\eta, 1]\}$ is an $(r, \beta; \eta; \alpha_1, \ldots, \alpha_b)$ flexible discrete adversary for U_1. □

Lemma 15. *The existence of an $(r\eta^{-1}, \eta\beta; \alpha_1, \ldots, \alpha_b)$-discrete adversary for $U' = (M; r_1\eta^{-1}, \ldots, r_b\eta^{-1})$ implies the existence of an $(r, \beta; \eta; \alpha_1, \ldots, \alpha_b)$-flexible discrete adversary for $U = (M; r_1, \ldots, r_b)$.*

Proof. Apply Observation 2 to deduce that the same adversary is an $(r, \eta\beta; \alpha_1, \dots, \alpha_b)$ discrete adversary for $U'' = (M; r_1, \dots, r_b; \eta)$ and then apply Lemma 14 on U'' to get an $(r, \beta; \eta; \alpha_1, \dots, \alpha_b)$ flexible discrete adversary for $U = (M; r_1, \dots r_b; 1)$. $\square$

Corollary 16. *Given a UMTS $U = (\mathcal{U}_b^\Delta; r_1, \dots, r_b)$, $r_1 \geqslant \dots \geqslant r_b \geqslant 1$, and $(n_i)_i$ satisfying $r_i = 0.5\rho(1 + \ln n_i)$, there exists an $(r, \beta; 0.5; \alpha_1, \dots, \alpha_b)$-flexible discrete adversary such that $r \geqslant 0.5\rho(1 + \ln \sum_i n_i)$, $\beta \leqslant 4\lambda_3 r_1$, and $\min_i \alpha_i \geqslant 0.5 r_1^{-1}$.*

Proof. Fix $\eta = 0.5$. Let $\bar{r}_i = r_i \eta^{-1}$. By Lemma 13 we have an $(\bar{r}, \bar{\beta}; \alpha_1, \dots, \alpha_b)$ discrete adversary for $(\mathcal{U}_b^\Delta; \bar{r}_1, \dots, \bar{r}_b)$, satisfying $\min_i \alpha_i \geqslant \bar{r}_1^{-1}$, and $\bar{\beta} \leqslant \lambda_3 \bar{r}_1$. By Lemma 15, there exists $(r, \beta; \eta; \alpha_1, \dots, \alpha_b)$ flexible discrete adversary for U, where $r = \eta\bar{r}$ and $\beta = \bar{\beta}\eta^{-1}$. Since $\bar{r}_i = \eta^{-1} r_i = \eta^{-1}\eta\rho(1 + \ln n_i)$, by Lemma 13, $r = \eta\bar{r} \geqslant \eta\rho \ln(1 + \sum_i n_i)$. $\square$

Next we show how to combine flexible adversaries.

Lemma 17 *(Combining Lemma). Let $U = (M; \bar{r}_1, \dots, \bar{r}_n)$ be an UMTS, where M is a k-HST metric space of diameter δ on n points, and denote the root vertex of the HST by u. Let $(M_1, M_2, \dots, M_b)$ be the partition of M to subspaces corresponding to the children u.*

Let U_j be the UMTS induced by U on M_j. Assume that for each $j \in \{1, \dots, b\}$ there exists $(r_j, \beta_j; \eta)$-flexible adversary $\mathcal{A}_j$ for U_j. Let $\hat{U} = (\mathcal{U}_b^\Delta; r_1, \dots, r_b)$ be the "combining UMTS". Assume there exists an $(r, \beta; \eta; \alpha_1, \dots, \alpha_b)$-flexible discrete adversary $\hat{\mathcal{A}}$ for $\hat{U}$.

If $k \geqslant \frac{\eta}{1-\eta} \max_j \frac{\beta_j}{\alpha_j}$, then there exists a $(r, \beta; \eta)$-flexible adversary $\mathcal{A}$ for U.

Proof. We fix $\beta' \in [\eta\beta, \beta]$, and the goal is to construct (r, β') adversary $\mathcal{D}$ for U.

We start by choosing a $(r, \beta'; \alpha'_1, \dots, \alpha'_b)$-discrete adversary $\hat{D}$ from $\hat{\mathcal{A}}$. Denote $\Delta = \Delta(M)$, and $\Delta_j = \Delta(M_j)$ the diameters of M and M_j, respectively. Then, for each j, we choose $\beta'_j \in (\eta\beta_j, \beta_j]$ such that $t_j = \alpha'_j \Delta / (\beta'_j \Delta_j)$ is a natural number. This is possible since

$$\frac{\alpha'_j \Delta}{\eta\beta_j \Delta_j} - \frac{\alpha'_j \Delta}{\beta_j \Delta_j} \geqslant \frac{\alpha_j \Delta}{\eta\beta_j \Delta_j} - \frac{\alpha_j \Delta}{\beta_j \Delta_j} \geqslant \frac{\alpha_j}{\beta_j} \frac{1-\eta}{\eta} k \geqslant 1.$$

Let $\mathcal{D}'_j$ be an (r_j, β'_j)-adversary from $\mathcal{A}_j$.

We construct a distribution $\mathcal{D}$ on elementary task sequences for U as follows: first we sample $\hat{\sigma} \in \hat{\mathcal{D}}$; then we replace each task $(z_j, \alpha'_j \Delta)$ in $\hat{\sigma}$ with $\sigma_j = \sigma_j^{(1)} \sigma_j^{(2)} \cdots \sigma_j^{(t_j)}$ where each $\sigma_j^{(i)}$ is independently sampled from $\mathcal{D}'_j$.

Next, we bound OPT_U^0. Let z_q be the point in $\hat{U}$ that minimizes $\mathrm{E}_{\hat{\sigma} \in \hat{D}}[\mathrm{cost}_{\mathrm{OPT}_{\hat{U}}^0}(\hat{\sigma}, z_q)]$. Let $v_0 \in M_q$ the point that minimizes $\mathrm{E}_{\sigma \in (\mathcal{D}'_q)}[\mathrm{cost}_{\mathrm{OPT}_{U_q}^0}(\sigma, v_0)]$. Consider the following offline strategy B for serving $\sigma \in \mathcal{D}$: the algorithm starts and finishes at v_0. B maintains the invariant that if $\mathrm{OPT}_{\hat{U}}^0$ is at a point z_i then B is at some point in M_i. Consider some task $(z_j, \alpha'_j \Delta)$ in $\hat{\sigma}$. It is replaced by sequence σ_j as described above. If $\mathrm{OPT}_{\hat{U}}^0$ moves to a point different from z_j it incurs a cost of Δ. In this case B moves out of M_j ahead of the task sequence σ_j incurring a cost of Δ as well. If $\mathrm{OPT}_{\hat{U}}^0$ is not at z_j then its cost for the task is 0 and the cost for B on σ_j is also 0. Otherwise, $\mathrm{OPT}^0_{\hat{U}}$ incurs a cost of $\alpha'_j \Delta$ for the task. In this case B uses $\mathrm{OPT}^0_{U_j}$ to serve σ_j in M_j. The expected cost of $\mathrm{OPT}^0_{U_j}$ for each subsequence $\sigma_j^{(i)}$ of σ_j is at most $\beta'_j \Delta_j$, and therefore the cost of B equals

$$\min_{u_0 \in M_j} \mathrm{E}_{\sigma_j \in (\mathcal{D}'_j)^{t_j}} \left[\mathrm{cost}_{\mathrm{OPT}^0_{U_j}}(\sigma_j, u_0)\right] \leqslant t_j \beta'_j \Delta_j = \alpha'_j \Delta.$$

It follows that the expected cost of B for serving σ_j is bounded from above by the cost of $\mathrm{OPT}^0_{\hat{U}}$ on the task $(z_j, \alpha'_j \Delta)$. Hence

$$\min_{u_0 \in M} \mathrm{E}_{\sigma \in \mathcal{D}}\left[\mathrm{cost}_{\mathrm{OPT}^0_U}(\sigma, u_0)\right] \leqslant \mathrm{E}_{\sigma \in \mathcal{D}}\left[\mathrm{cost}_B(\sigma, v_0)\right] \leqslant \mathrm{E}_{\hat{\sigma} \in \hat{\mathcal{D}}}\left[\mathrm{cost}_{\mathrm{OPT}^0_{\hat{U}}}(\hat{\sigma}, z_q)\right] \leqslant \beta' \Delta.$$

It is left to show a lower bound on online algorithms for U. Let A be an online algorithm for U. We can naturally define an online algorithm $\hat{A}$ for $\hat{U}$ as follows. Consider a distribution on sequences $\sigma \in \mathcal{D}$ generated as described above from a sequence $\hat{\sigma} \in \hat{\mathcal{D}}$. Consider a task $(z_j, \alpha'_j \Delta)$ in $\hat{\sigma}$ and let σ_j be the corresponding sequence generated above for M_j. Whenever A moves between subspaces into a point in subspace M_i during the service of σ_j, if $i \neq j$ then $\hat{A}$ makes a move to the corresponding point z_i before serving the task. $\hat{A}$ serves the task in the last such z_i and finally moves to the point corresponding to the subspaces in which A ends the service of σ_j. Obviously, the moving cost of $\hat{A}$ is bounded from above by the cost A incurs on moves between subspaces. If A does move between subspaces during the service of σ_j then $\hat{A}$ incurs zero local cost for the task and therefore its cost for the task is at most that of A on σ_j. Otherwise, if A is in a subspace M_i, $i \neq j$, during the entire sequence σ_j then the cost of $\hat{A}$ for the task is 0. If A is in M_j during σ_j then we have

$$\min_{u_0 \in M_j} \mathrm{E}_{\sigma_j \in (\mathcal{D}'_j)^{t_j}} \left[\mathrm{cost}_A(\sigma_j, u_0)\right] \geqslant t_j \cdot r_j \beta'_j \Delta_j = r_j \alpha'_j \Delta,$$

which is the cost for $\hat{A}$. It follows that in all cases the expected cost of A on σ_j is at least the cost of $\hat{A}$ on the task. Thus, we get that for any online algorithm A,

$$\min_{u_0 \in M} \mathrm{E}_{\sigma \in \mathcal{D}}\left[\mathrm{cost}_A(\sigma, u_0)\right] \geqslant \min_{z_0 \in \hat{M}} \mathrm{E}_{\hat{\sigma} \in \hat{D}}\left[\mathrm{cost}_{\hat{A}}(\hat{\sigma}, z_0)\right] \geqslant r\beta' \Delta. \qquad \square$$

Proof of Theorem 3. Fix constants $c_2 = 0.5\rho$, $c_3 = 4\lambda_3$, and $c_1 = 2c_3$. Consider an arbitrary $(c_1(1 + \ln N)^2)$-HST on N points. We construct by induction on the height of a subtree T_u rooted at u, a $(r_u, \beta_u; 0.5)$-flexible adversary for a subtree with n_u leaves, such that $r_u \geqslant \max\{1, c_2(1 + \ln n_u)\}$, and $\beta_u \leqslant c_3(1 + \ln n_u)$.

The base case are trees on of height 1 for which we can apply the adversary of Lemma 12 with $r_i = 1$.

For height larger than one, assume an internal vertex u of the HST has n points in its subspace, and b children. Inductively, assume that each T_i, a tree rooted at the children of u, has $(r_i, \beta_i; 0.5)$ flexible adversary, such that $\beta_i \leqslant c_3(1 + \ln n_i)$ and $r_i \geqslant \max\{1, c_2(1 + \ln n_i)\}$. Note that $(r, \beta; \eta)$ flexible adversary implies $(r', \beta; \eta)$ flexible adversary for $r' \leqslant r$, and therefore we may assume that $r_i = \max\{1, c_2(1 + \ln n_i)\}$.

We use the flexible discrete adversary from Corollary 16 as the combining adversary in Lemma 17. Here $\beta_i/\alpha_i \leqslant 2r_i\beta_i \leqslant 2\max\{c_2(1+\ln n_u),1\}c_3(1+\ln n_u) \leqslant c_1(1+\ln N)^2$, and thus we get an $(r,\beta;0.5)$ flexible adversary for T_u with $r \geqslant c_2(1+\ln n_u)$, and $\beta \leqslant 4\lambda_3 \max_i r_i \leqslant c_3(1+\ln n_u)$. $\square$

Corollary 18. *The randomized competitive ratio of the MTS problem in n-point ℓ-dimensional mesh is $\Omega(\frac{\log n}{\log\log n})$.*

Proof. Combining Lemma 8 with Lemma 6, using $k = \Theta(\log^2 n)$, we deduce that the mesh contains a subspace of size $n^{\Omega(1/\log\log n)}$ that $O(1)$ approximates a $\Omega(\log^2 n)$-HST. Next we apply the lower bound of Theorem 3 on that HST. $\square$

6. Lower bounds for K-server

Theorem 1 also implies a lower bound for the K-server problem. This follows from the following general reduction from the MTS problem on an n-point metric space to the $(n-1)$-server problem on the same metric space.

Lemma 19. *An r-competitive randomized algorithm for the $(n-1)$-servers problem on an n-point metric space against oblivious adversaries implies a $(2r+1)$ upper bound on the randomized competitive ratio for MTS on the same metric space.*

Proof. We will prove the implication only for MTS problems in which the tasks are elementary. For the purpose of establishing a lower bound for the K-server this is sufficient since our lower bound for MTS uses only elementary tasks. However, there is also a general reduction [BBBT97] from an upper bound for any tasks to an upper bound for elementary tasks.

Given a metric space M on n points with metric d and diameter Δ, denote by S the $(n-1)$-servers problem on M and by T the MTS problem on M. For a request sequence σ in S, and point $i \in M$, we denote by $w^S_\sigma(i)$ the optimal offline cost for servicing σ and end without a server in i. Similarly for task sequence τ in T we denote by $w^T_\tau(i)$ the optimal cost for servicing τ and ending in state i (these are called *work functions*). Note that for any τ,σ,i,j, $w^S_\sigma(i) - w^S_\sigma(j) \leqslant d(i,j)$ and $w^T_\tau(i) - w^T_\tau(j) \leqslant d(i,j)$.

Given a randomized algorithm A_S for S, we construct an algorithm A_T for T. A_T transforms a task sequence τ into a sequence σ for S as follows. Assume the sequence is $\tau' = \tau e$ where τ has been already transformed into σ. A_T transforms the elementary task $e = (i,\delta_i)$ using the following rule. If

$$w^T_\tau(i) + \delta_i \geqslant \min_{j:\, j\neq i} w^T_\tau(j) + d(i,j), \tag{4}$$

it gives a request for i in S, otherwise no request is given. A_T simulates A_S and maintains its state in the point where A_S does not have a server.[7] Note that the request se-

[7] Without loss of generality, we may assume that no two servers of A_S are at the same point.

quence σ was constructed oblivious to the random bits of A_S, and thus $\mathrm{E}[\mathrm{cost}_{A_S}(\sigma)] \leqslant r\,\mathrm{cost}_{\mathrm{OPT}_S}(\sigma) + C$.

Next, we prove by induction on the sequence that for any τ and any i, $w^S_\sigma(i) \leqslant w^T_\tau(i)$. For $\tau = \varepsilon$, it is obvious that for all i, $w^S_\varepsilon(i) = w^T_\varepsilon(i)$. Tasks in T that do not generate tasks in S, obviously maintain the inductive invariant. Otherwise, let $e = (l, \delta_l)$ be a task in T that generates a request e' in S for l. w^S has the following update rules. In point l,

$$w^S_{\sigma e'}(l) = \min_{j:\ j \neq l} \bigl(w^S_\sigma(j) + d(l, j)\bigr) \leqslant \min_{j:\ j \neq l} \bigl(w^T_\tau(j) + d(l, j)\bigr) = w^T_{\tau e}(l).$$

The last equality follows from (4). For $i \neq l$, $w^S_{\sigma e'}(i) = w^S_\sigma(i) \leqslant w^T_\tau(i) \leqslant w^T_{\tau e}(i)$. Therefore $\mathrm{cost}_{\mathrm{OPT}_S}(\sigma) \leqslant \mathrm{cost}_{\mathrm{OPT}_T}(\tau)$.

Denote by $\mathrm{lcost}_A(\tau)$ and $\mathrm{mcost}_A(\tau)$ the local cost and the movement cost of algorithm A on sequence τ. Since A_T moves similarly to A_S, $\mathrm{mcost}_{A_T}(\tau) = \mathrm{mcost}_{A_S}(\sigma)$. To bound the local cost of A_T, we prove that

$$\mathrm{lcost}_{A_T}(\tau) \leqslant \mathrm{mcost}_{A_T}(\tau) + w^T_\tau(i_c), \tag{5}$$

where i_c is the current state of A_T. Consider a task $e = (i, \delta_i)$. If A_T was not in a state i, no local cost was generated. If A_T was in a state i and did not move in response to task e, its local cost is δ_i. On the other hand, for any j, $w^T_\tau(i) + \delta_i - w^T_\tau(j) \leqslant d(i, j)$, so $w^T_{\tau e}(i) = w^T_\tau(i) + \delta_i$, hence Eq. (5) is maintained. If A_T moves to state j then its local cost is 0. In this case the right side of Eq. (5) is changed by $d(i, j) + w^T_\tau(j) - w^T_\tau(i) \geqslant 0$, and therefore Eq. (5) is maintained. We conclude that $\mathrm{cost}_{A_T}(\tau) \leqslant 2\,\mathrm{cost}_{A_S}(\sigma) + \mathrm{cost}_{\mathrm{OPT}_T}(\tau) + \Delta$. To summarize

$$\begin{aligned}
\mathrm{E}\bigl[\mathrm{cost}_{A_T}(\tau)\bigr] &\leqslant 2\mathrm{E}\bigl[\mathrm{cost}_{A_S}(\sigma)\bigr] + \mathrm{cost}_{\mathrm{OPT}_T}(\tau) + \Delta \\
&\leqslant 2r\,\mathrm{cost}_{\mathrm{OPT}_S}(\sigma) + \mathrm{cost}_{\mathrm{OPT}_T}(\tau) + \Delta + 2C \\
&= (2r+1)\,\mathrm{cost}_{\mathrm{OPT}_T}(\tau) + C',
\end{aligned}$$

where $C' = 2C + \Delta$ is a constant. □

We remark that the technique of Lemma 19 can also be applied to deterministic algorithms, but not directly to randomized algorithms in the *adaptive online adversary model* [BDBK+94]. In [MMS90], a different reduction from the MTS problem to the servers problem is given. Their reduction applies to all adversary models and is more efficient. However, it reduces an MTS problem to a servers problem in a *different metric space*, and therefore inappropriate for our purposes.

When applying Lemma 19 on Theorem 1 we deduce following.[8]

Theorem 6. *The randomized competitive ratio against oblivious adversaries of the K-server problem on a metric space with more than K points is $\Omega(\log K/\log^2\log K)$.*

Using Corollary 18 we have

[8] A direct way to argue Theorem 6 without using Lemma 19 is to observe that the adversary in the proof of Theorem 3 uses tasks that if replaced with task size infinity will increase OPT^0's cost by at most a factor of two.

Corollary 20. *The randomized competitive ratio against oblivious adversaries of the K-server problem on h-dimensional mesh with more than K points is $\Omega(\log K/\log\log K)$.*

Proof *(sketch).* Let M be an h-dimensional mesh, $[s]^h$. We first remove points from M to obtain a maximal submesh, M', of M of size $m \leqslant K$. It easy to observe that $m \geqslant \sqrt{K}$. It follows from Lemma 19 and Corollary 18 that M' has a lower bound of $\Omega(\log m/\log\log m)$ for $m-1$ servers. To get a lower bound for M we pick $K-m+1$ points not in M' and modify the adversary for M' by inserting repeated requests to these points between its original requests to make sure that $K-m+1$ servers will have to stay at these points. □

For $n \gg K$, it is possible to get a better lower bound.

Theorem 7. *Fix a constant $\varepsilon > 0$. Then for any K and any metric space M on $n \geqslant K^{\log^\varepsilon K}$ points, the K-server problem on M has a lower bound of $\Omega(\log K)$ on the competitive ratio for randomized online algorithms against oblivious adversaries.*

Proof *(sketch).* Assume K is large enough. Let $f = K^{\log^\varepsilon K}$. We take an arbitrary subspace with f points. Using Theorem 4 with $\beta = \log^{\varepsilon/2} K$, $\ell = 2$, and $k = \Theta(\log^2 K)$, we find a subspace that $O(\log_\beta \log f) = O(\varepsilon^{-1})$ approximates a k-HST and has $f^{(\beta\lceil\log k\rceil)^{-1}} > K$ points. We further delete arbitrary points from this subspace to get exactly $K+1$ points. From Theorem 3 we have a lower bound of $\Omega(\log K)$ MTS in this space. We conclude the claim by using Proposition 1 and Lemma 19. □

7. Additional Ramsey-type theorems

In this section we prove additional Ramsey-type theorems, and relate our constructions to those of [BFM86,KRR94,BKRS00]. In a subsequent paper [BLMN03b] these metric Ramsey problems are further studied, and tight bounds are given.

Definition 12. A vertex u in a rooted tree is called *balanced* if the difference between the number of leaves of any two subtrees rooted at u's children, is at most one. The following is a decreasing hierarchy of HST subclasses.

(1) A *“binary/balanced” k-HST* is a k-HST with the property that every internal vertex is either balanced or has at most two children.
(2) A *“binary/uniform” k-HST* is a k-HST with the property that every internal vertex u either has at most two children or all its children are leaves.
(3) A *“BKRS” k-HST* is a “binary/uniform” k-HST such that an internal vertex with exactly two children is either balanced or one of the children is a leaf.
(4) A *“BFM” HST* is a 1-HST whose underlying tree is binary and for each vertex at most one child is not a leaf.

(5) A *"KRR" k-HST*, for $k > 1$, is either a uniform space or a k-super-increasing metric space, where a k-super-increasing space is a k-HST in which every internal vertex has at most two children, and at most one of them is not a leaf.

Bourgain et al. [BFM86], Karloff et al. [KRR94] and Blum et al. [BKRS00] essentially prove the following Ramsey-type theorems.

Theorem 8. *For any $k \geqslant 4$ and any metric space $M = (S, d)$ on n points*:

(1) [BFM86] *There exists a subspace $S' \subseteq S$ such that $|S'| \geqslant C(\varepsilon)\log n$ and (S', d) is $(1+\varepsilon)$-approximated by a "BFM" HST.*[9]
(2) [KRR94] *There exists a subspace $S' \subseteq S$ such that $|S'| = \Omega(\frac{\log n}{\log\log n})$ and (S', d) is $O(k^2)$-approximated by a "KRR" k-HST.*[10]
(3) [BKRS00] *There exists a subspace $S' \subseteq S$ such that $|S'| = 2^{\Omega(\sqrt{\log_k n} - \log^2 k)}$, and (S', d) is 4-approximated by a "BKRS" k-HST.*[11]

"Binary/balanced" HSTs are of special interest for us. Our lower bound on the competitive ratio of HST is actually proved for this class of spaces, with Proposition 11 as the key argument for applying it on arbitrary HST (see Lemma 10). Here we explicitly construct "binary/balanced" HSTs.

Lemma 21. *In any HST on n leaves there exist a subset of the leaves of size $\sqrt{n}$ on which the induced HST is a "binary/balanced" HST.*

Proof. The lemma is proved inductively by applying Proposition 11 as the inductive argument. The only issue here is how to maintain subtrees with the same number of leaves. This is done using a dynamic programming approach.

Formally, we prove by induction on h that for any rooted tree T with a root r, of height h, and with n leaves, and for any $m \in \{0, 1, \ldots, \lceil\sqrt{n}\rceil\}$, T contains a subtree rooted at r on m leaves.

For $h = 0$ the claim is trivial. Otherwise, let $T_1, \ldots, T_b$ be the subtrees rooted at the children of r. Denote by $n_i = |T_i|$, so $n = \sum_{i=1}^{b} n_i$. Assume without loss of generality that $n_1 \geqslant n_2 \geqslant \cdots \geqslant n_b > 0$. If $b = 1$ the claim follows by the inductive hypothesis on T_1. Otherwise, fix an integer m, $\lceil\sqrt{n}\rceil \geqslant m \geqslant 0$. By Proposition 11, one of the following holds:

(1) $\sqrt{n} \leqslant \sqrt{n_1} + \sqrt{n_2}$. In this case we choose integers $m_1 \leqslant \lceil\sqrt{n_1}\rceil$ and $m_2 \leqslant \lceil\sqrt{n_2}\rceil$ such that $m = m_1 + m_2$. By the inductive hypothesis there exist T_1', a "binary/balanced" subtree of T_1 with m_1 leaves and T_2' a "binary/balanced" subtree of T_1 with m_2 leaves.

[9] This is statement is only implicit in [BFM86]. They are interested in embedding a subspace inside ℓ_2. Embedding in ℓ_2 is achieved by observing that a "BFM" HST is isomorphic to a subset of ℓ_2.
[10] Using Lemma 4, it is possible to improve the theorem to $O(k)$ approximation by a "KRR" k-HST.
[11] The definition of "BKRS" HST, the statement of this claim, and its proof are only implicit in [BKRS00]. In particular, they only consider the case $k = \log^3 n$.

The tree T'—rooted at r with the two subtrees T_1' and T_2' as the children—is a "binary/balanced" tree with m leaves.

(2) $\exists \ell \in \{2, \ldots, b\}$ such that $\sqrt{n} \leqslant \ell \sqrt{n_\ell}$. Let $m' = m/\ell$. Note that $m' \leqslant \lceil \sqrt{n_\ell} \rceil$, since $m \leqslant \lceil \sqrt{n} \rceil \leqslant \ell \lceil \sqrt{n_\ell} \rceil$. Thus, by the inductive hypothesis, for $i \leqslant \ell$, it is possible to extract from T_i trees with $\lfloor m' \rfloor$ and $\lceil m' \rceil$ leaves. By choosing a combination of trees T_i' of sizes $\lfloor m' \rfloor$ or $\lceil m' \rceil$, it is possible to get trees T_i' such that $|T_i'| - |T_j'| \in \{-1, 0, 1\}$ and $\sum_i |T_i'| = m$. Combining these subtrees with the root r, gives a "binary/balanced" subtree T' with m leaves. $\square$

Lemma 21 combined with Theorem 4, is the strongest Ramsey-type theorem presented in this paper.

Theorem 9. *For any metric space $M = (V, d)$ on $|V| = n$ points, any $\beta > 1$, any $k > 1$, and any $1 < \ell \leqslant k$ there exists a subset $S \subseteq V$, such that $|S| \geqslant n^{1/(2\beta \lceil \log_\ell k \rceil)}$ and (S, d) $O(\ell \log_\beta \log n)$-approximates a binary/balanced k-HST.*

Proposition 22. *In any k-HST on n leaves there exist a subset of the leaves of size $\Omega(\frac{\log n}{\log \log n})$ on which the induced HST is a "KRR" k-HST.*

Proof. Let T be the given HST and assume it does not have degenerate vertices. Either T has an internal vertex u with at least $\log n$ children, and in this case, by taking one descendant leaf from each child of u, we get a uniform space. Otherwise, T must have a vertical path of length at least $\log_{\log n} n$. Take this path and add for each internal vertex along the path another child as a leaf. The resulting HST is super-increasing. $\square$

Proposition 23. *In any HST on n leaves there exists a subset of the leaves of size $\Omega(\log n)$ on which the induced metric space is a "BFM" HST.*

Proof. We first observe that any HST can be transformed into a 1-HST whose underlying tree is binary without degenerate vertices. We then take the longest vertical path p in S—its length is at least $\log n$—and adjoin for each internal vertex u along p, a leaf from the subtree of the child of u not on p. $\square$

Proposition 24. *Given a sequence $n_1 \geqslant n_2 \geqslant \cdots \geqslant n_b > 0$, and $n = \sum_{i=1}^b n_i$, then* $\max\{b, 2n_2^{1/(2\sqrt{\log n})}, n_1^{1/(2\sqrt{\log n})} + 1\} \geqslant 2^{\sqrt{\log n}/2}$.

Proof. Assume that $\max\{b, 2n_2^{1/(2\sqrt{\log n})}\} < 2^{\sqrt{\log n}/2}$. Then

$$n_1 \geqslant n - bn_2 \geqslant n - 2^{\sqrt{\log n}/2} \frac{n}{2^{2\sqrt{\log n}}} \geqslant n\left(1 - \frac{1}{2^{\sqrt{\log n}}}\right).$$

Therefore,

$$\begin{aligned} n_1^{1/(2\sqrt{\log n})} &\geqslant n^{1/(2\sqrt{\log n})} \left(1 - \frac{1}{2^{\sqrt{\log n}}}\right)^{1/(2\sqrt{\log n})} \geqslant 2^{\sqrt{\log n}/2}\left(1 - \frac{1}{2^{\sqrt{\log n}}}\right) \\ &\geqslant 2^{\sqrt{\log n}/2} - 1. \qquad \square \end{aligned}$$

Proposition 25. *In any HST on n leaves there exists a subset of the leaves of size at least $2^{\sqrt{\log n}/2}$ on which the induced HST is a "BKRS" HST.*

Proof. We prove, by induction on the height of the tree, that for any tree T with n leaves and for any $m \leqslant \lceil 2^{\sqrt{\log n}/2} \rceil$, T contains a "BKRS" subtree T' with m leaves.

Let r be the root of T and $T_1, \ldots, T_b$ the subtrees rooted at the children of r. Denote $n_i = |T_i|$, and apply Proposition 24.

If $b \geqslant 2^{\sqrt{\log n}/2}$, we construct T' by connecting r to one leaf from each T_i, for $1 \leqslant i \leqslant m$.

If $2n_2^{1/(2\sqrt{\log n})} \geqslant 2^{\sqrt{\log n}/2}$, then we construct T' by connecting r to T_1' and T_2', where T_1' is a subtree of T_1 with $\lceil m/2 \rceil$ leaves, and T_2' is a subtree of T_2 with $\lfloor m/2 \rfloor$ leaves.

If $n_1^{1/(2\sqrt{\log n})} + 1 \geqslant 2^{\sqrt{\log n}/2}$, then we construct T' by connecting r to T_1' and one leaf from T_2, where T_1' is a subtree of T_1 with $m-1$ leaves. □

Propositions 22, 23, and 25, when combined with Theorem 4, give corresponding Ramsey-type theorems. These results, however, are slightly weaker than Theorem 8, as the approximation factor is $O(\log\log n)$ instead of a constant.[12] We include them to demonstrate the simplicity of their proof, when using HST.

We end the section with some impossibility examples. The first one deals with subspaces of equally spaced points on the line.

Proposition 26. *For any $\alpha \geqslant 1$ there exists $c < 1$, such that any subset of n equally spaced points on the line that is α-approximated by an HST, is of size at most $O(n^c)$.*

Proof. Let $M = \{v_1, v_2, \ldots, v_n\}$ be the metric space on n points such that $d_M(v_i, v_j) = |i - j|$. Let $S \subseteq M$ be a subspace that is α approximated by an HST T. We prove by induction on n that $|S| \leqslant 2(\alpha+1)n^c$, where $c = c(\alpha) < 1$ will be chosen later.

Without loss of generality, we may assume that T is a binary tree without degenerate vertices. Let $n' = \max\{d_M(u, v) \mid u, v \in S\} + 1 \leqslant n$. Without loss of generality, assume that $v_1, v_{n'} \in S$ are the two extreme points in S. For $n' \leqslant 2(\alpha+1)$ the inductive claim is trivially true. Otherwise, let $u = \mathrm{lca}_T(v_1, v_{n'})$, so $\Delta(u) \geqslant n' - 1$. Denote by S_1 and S_2 the two subspaces induced by the children of u. Since for any $v \in S_1$ and $v' \in S_2$, $d_M(v, v') \geqslant (n'-1)/\alpha$, we can partition the interval $\{v_1, \ldots, v_{n'}\}$ into $2\ell+1$ subintervals $I_1, \ldots, I_{2\ell+1}$, such that for any $i \in \{1, \ldots, \ell\}$, $|I_{2i}| \geqslant \frac{n'-1}{\alpha} - 1$ and $I_{2i} \cap S = \emptyset$; for $i \geqslant 0$, $I_{4i+1} \cap S \subseteq S_1$ and $I_{4i+3} \cap S \subseteq S_2$. Denote by $n_i = |I_i|$. Thus $\sum_{i=0}^{\ell} n_{2i+1} + \ell(\alpha(n'-1) - 1) \leqslant n'$. The induced HST on $S \cap I_{2i+1}$ α-approximates the original distances, and therefore by the inductive hypothesis $|S \cap I_{2i+1}| \leqslant 2(\alpha+1)n_{2i+1}^c$.

Assume $\ell = 1$, then $n_1 + n_3 \leqslant n' - (\frac{n'-1}{\alpha} - 1) \leqslant n'(1 - \frac{1}{2\alpha})$, the last inequality follows since $n' \geqslant 2(\alpha+1)$. By concavity, the maximum of $n_1^c + n_3^c$ is reached when $n_1 = n_3 \leqslant (n'(1 - \frac{1}{2\alpha}))/2$. Thus

[12] This is when using a constant β. Alternatively, when choosing $\beta = \log^{\varepsilon} n$, we get a constant approximation but of slightly smaller subspaces.

$$|S| \leqslant 2(\alpha+1)2\left(\frac{n'(1-\frac{1}{2\alpha})}{2}\right)^c \leqslant 2(\alpha+1)2\left(\frac{(1-\frac{1}{2\alpha})}{2}\right)^c n'^c \leqslant 2(\alpha+1)n'^c.$$

The last inequality follows since it is possible to choose $c < 1$ such that $2(\frac{(1-\frac{1}{2\alpha})}{2})^c \leqslant 1$.

The proof for $\ell > 1$ follows by induction on ℓ. Denote $J_1 = \bigcup_{i=1}^{2\ell-1} I_i$, $J_2 = I_{2\ell}$, and $J_3 = I_{2\ell+1}$. Also denote $N_1 = |J_1|$ and $N_3 = |J_3|$. By the inductive hypothesis, $|S \cap J_1| \leqslant 2(\alpha+1)N_1^c$, and $|S \cap J_3| \leqslant 2(\alpha+1)N_3^c$. Applying the argument above, we conclude that $N_1^c + N_3^c \leqslant n'^c$. □

Next we show examples that prove that Lemma 6, Theorem 8, Propositions 22, 23 and 25 are all essentially tight. Before presenting the examples we need the following claims.

Proposition 27. *Assume that an HST T is ℓ-approximated by a k-HST W, and $\ell < k$. Then for any four (not necessarily distinct) points a, b, c, d in the space,*

$$\mathrm{lca}_T(a,b) = \mathrm{lca}_T(c,d) \quad \Rightarrow \quad \mathrm{lca}_W(a,b) = \mathrm{lca}_W(c,d).$$

Proof. Assume $\mathrm{lca}_T(a,b) = \mathrm{lca}_T(c,d)$. Denote $u' = \mathrm{lca}_W(a,b)$, and $v' = \mathrm{lca}_W(c,d)$. Assume for the sake of contradiction that $u' \neq v'$. First we observe that u' cannot be a proper ancestor of v', since otherwise $d_W(a,b) > \ell d_W(c,d)$, and this means that W does not ℓ approximates T. From the same reason v' is not a proper ancestor of u'. This implies that $\mathrm{lca}_W(a,c)$ is a proper ancestor of $\mathrm{lca}_W(a,b)$, and so $d_W(a,c) > \ell d_W(a,b)$, whereas in T it must be that $\mathrm{lca}_T(a,c)$ is a descendant of $\mathrm{lca}_T(a,b)$, and thus $d_T(a,c) \leqslant d_T(a,b)$. Again, this means that W does not ℓ-approximate T, a contradiction. □

Lemma 28. *Assume that a k-HST T is ℓ-approximated by a k-HST W. If both T and W do not have degenerate vertices and $\ell < k$, then the underlying trees of T and W are isomorphic.*

Proof. It is sufficient to show that for any four (not necessarily distinct) points a, b, c, d in the space, $\mathrm{lca}_T(a,b) = \mathrm{lca}_T(c,d)$ if and only if $\mathrm{lca}_W(a,b) = \mathrm{lca}_W(c,d)$. This is so since we can define $f: T \to W$, by $f(\mathrm{lca}_T(a,b)) = \mathrm{lca}_W(a,b)$. It is easy to check that f is well defined injective and bijective. Also, if u is ancestor of v in T, then $f(u)$ is an ancestor of $f(v)$ in W. To see this, let a, b two descendant leaves of v in T such that $\mathrm{lca}_T(a,b) = v$, and let c be a descendant leaf of u such that $\mathrm{lca}_T(a,c) = \mathrm{lca}_T(b,c) = u$, but then $\mathrm{lca}_W(a,c) = \mathrm{lca}_W(b,c)$, and this implies that $\mathrm{lca}_W(a,c)$ is an ancestor of $\mathrm{lca}_W(a,b)$.

In order to prove that $\forall a, b, c, d$, $\mathrm{lca}_T(a,b) = \mathrm{lca}_T(c,d)$ if and only if $\mathrm{lca}_W(a,b) = \mathrm{lca}_W(c,d)$, we apply Proposition 27 in two directions (and noting that the approximation relation is essentially symmetric, see the discussion after Definition 4). □

Proposition 29. *Let $k > \ell > 1$. There are infinitely many values of n for which there exist HSTs (collectively denoted by T) with n leaves such that*:

(1) *Any k-HST that is ℓ-approximated by a subspace of T, has at most $n^{1/\log_\ell k}$ points.*
(2) *Any "binary/uniform" k-HST that is ℓ-approximated by a subspace of T, has at most $2^{2\sqrt{\log n/\log_\ell k}}$ points.*

(3) *Any "KRR" k-HST that is ℓ-approximated by a subspace of T, has at most $O(\frac{\log n}{\log_\ell k \log\log n})$ points.*
(4) *Any "BFM" HST that is ℓ-approximated by a subspace of T, has at most $O(\log n)$ points.*

Proof. The examples will all have the same basic structure. Fix a small constant $\varepsilon > 0$ to be determined later, and let $k > \ell' = (1+\varepsilon)\ell$. Consider an ℓ'-HST T such that an internal vertex v with edge depth i has diameter $\Delta(v) = \ell'^{-i}$. Let $h \in \mathbb{N}$ be a parameter of the size of T.

(1) In this case, T is a complete binary tree of height $h\lceil \log_{\ell'} k\rceil$ with $n = 2^{h\lceil \log_{\ell'} k\rceil}$ leaves. Let $R \subseteq S$ be a subset of the points that ℓ approximates a k-HST W. Let T' be the subtree of T that its leaves are exactly the subset R. It follows from Lemma 28 that the edge distance in T' between any two non-degenerate vertices u and v is at least $\lceil \log_{\ell'} k\rceil$. Hence, when coalescing pair of edges with common degenerate vertex in T', the resulting tree is a binary tree of height at most h with the same set of leaves, R, and so $|R| \leqslant 2^h \leqslant n^{1/\lceil \log_{\ell'}\rceil k}$. Choosing $\varepsilon > 0$ small enough implies that $|R| < n^{1/\log_\ell k} + 1$.
(2) In this case, T is a complete tree of height $h\lceil \log_{\ell'} k\rceil$ and the out-degree of each internal vertex is 2^h. Hence $n = 2^{h^2\lceil \log_{\ell'} k\rceil}$, so $h = \sqrt{\log n/\lceil \log_{\ell'} k\rceil}$. Let R be a subset of points approximating a "binary/uniform" k-HST W, and let T' be the subtree of T whose set of leaves is exactly R. By Lemma 28, T' is also a "binary/uniform" HST. As before on any vertical path in T' there are only h non-degenerate vertices. After removing degenerate vertices from T' (by coalescing pair of edges with common degenerate vertex), it is easy to show by induction on the levels, that a vertex in level i in T' cannot have more than $2^{h-i}2^h$ leaves, and therefore T' has no more than 2^{2h} leaves. By choosing $\varepsilon > 0$ small enough we conclude the claim.
(3) In this case, T is a complete tree of height $h\lceil \log_{\ell'} k\rceil$ and the out-degree of each internal vertex is h. Assume also that $h \geqslant \lceil \log_{\ell'} k\rceil$. Hence $n = h^{h\lceil \log_{\ell'} k\rceil}$, so
$$h = \Theta\left(\frac{\log n}{\lceil \log_{\ell'} k\rceil(\log\log n - \log\lceil \log_{\ell'} k\rceil)}\right) = \Theta\left(\frac{2\log n}{\lceil \log_{\ell'} k\rceil \log\log n}\right).$$
The last inequality follows since $\log\log n \geqslant 2\log\lceil \log_{\ell'} k\rceil$.
Let R be a subset of points approximating a "KRR" k-HST W, and let T' be the subtree of T that its leaves are exactly R. By Lemma 28, T' is also a "KRR" k-HST. Either T' is a uniform metric, and then the leaves are all children of one vertex (after removing degenerate vertices), and therefore there at most h such leaves. Otherwise, T' is k-super-increasing. Only h vertices on any vertical path in T' are non-degenerate. and so T' has at most $h+1$ leaves. Again, the claim follows by taking $\varepsilon > 0$ small enough.
(4) In this case, T is a complete binary tree of height h with $n = 2^h$ leaves. Let R be a subset of points approximating a "BFM" HST W, and let T' be the subtree of T that its leaves are exactly R. T' is a binary tree, since it is a subtree of a binary tree.
We want to prove that T' is a "BFM" HST. Assume for the sake of contradiction that T' is not a "BFM" HST. This implies the existence of four distinct leaves a, b, c, d,

with the following "pairing" property: There exists a partition of a, b, c, d into two pairs $\{a, b\}$ and $\{c, d\}$, such that $\mathrm{lca}_{T'}(a, c) = \mathrm{lca}_{T'}(a, d) = \mathrm{lca}_{T'}(b, c) = \mathrm{lca}_{T'}(b, d)$ (equals u), but $\mathrm{lca}_{T'}(a, b) \neq u$ and $\mathrm{lca}_{T'}(c, d) \neq u$. However, in a "BFM" HST S, for any subset of leaves A, there exists $x \in A$ such that for any $x \notin \{y, z\} \subset A$, $\mathrm{lca}_S(x, y) = \mathrm{lca}_S(x, z)$. So without loss of generality, $\mathrm{lca}_W(a, b) = \mathrm{lca}_W(a, c) = \mathrm{lca}_W(a, d)$. By Proposition 27, it implies that $\mathrm{lca}_{T'}(a, b) = \mathrm{lca}_{T'}(a, c) = \mathrm{lca}_{T'}(a, d)$. Therefore any partition of a, b, c, d to two pairs contradicts the "pairing" property.

So T' is a "BFM" tree, and therefore has at most $h + 1$ leaves. □

8. Concluding remarks

As mentioned before, the lower bound on the competitive ratio of the MTS problem in n-point metric spaces was improved in [BLMN04a] to $\Omega(\log n/\log\log n)$. It is an interesting challenge to achieve $\Omega(\log n)$ lower bound for any metric space. A plausible way to do it is proving a lower bound for k-HSTs *with constant* k. This was done in the context of proving upper bounds on the competitive ratio for the MTS problem in [FM03] using "fine grained" combining technique.

Lemma 13 is a tight lower bound for UMTS on uniform metric when assuming the conjecture of $\Theta(\log n)$ competitive ratio for MTS. An interesting problem is to find a matching upper bound. This would improve the general upper bound for MTS by a factor of $\log\log n$. A harder problem is to improve the upper bound for MTS to $o(\log^2 n)$.

For the K-server problem in arbitrary metric spaces, no sublinear upper bound on the randomized competitive ratio is known.

Acknowledgments

We thank Noga Alon, Amos Fiat, Guy Kindler, Nati Linial, Yuri Rabinovich, Mike Saks, Steve Seiden, and Amit Singer for many discussions and suggestions. In particular, Nati helped in simplifying and improving earlier versions of Lemma 10 and Proposition 11.

Appendix A. Some probabilistic calculations

In this section we present some probabilistic arguments needed in the proof of Lemma 10.

Lemma A.1. *There exist constants* $\lambda_2 \geqslant 1 \geqslant \lambda_1 > 0$ *such that for any binomial random variable* X *with* $p \leqslant 0.5$ *and mean* $\mu \geqslant 4$ *and any* $\delta \in [0, 1]$ *we have*

$$\Pr\bigl[X \leqslant (1-\delta)\mu\bigr] \geqslant \lambda_1 e^{-\lambda_2 \delta^2 \mu}.$$

Lemma A.1 is easily realized for most of the range of p, δ, μ using the Poisson and normal approximations of binomial distribution (cf. [Bol85, Chapter 1]). Here we give an elementary proof.

Set $f(x) = (1-x)^{1/x}$. Clearly, f is increasing as x decreases to 0, and its limit is e^{-1}.

Proposition A.2. *Let $X \sim \mathcal{B}(m, p)$ be a binomial random variable, $p + q = 1$, $p \leqslant 1/2$, $\mu = pm$, $k = (1 - \eta)\mu \in [m]$, and $\eta \in (0, 1)$. Then*

$$\Pr[X = k] \geqslant \frac{f(\eta)^{\eta^2\mu}}{3\sqrt{\mu}}.$$

Proof. Recall that by Stirling formula (cf. [Bol85, p. 4]),

$$\Pr[X = k] = \binom{m}{k} p^k q^{m-k} \geqslant \frac{1}{e^{1/6}\sqrt{2\pi k}} \left(\frac{pm}{k}\right)^k \left(\frac{qm}{m-k}\right)^{(m-k)}$$
$$\geqslant \frac{1}{3\sqrt{\mu}} \left(\frac{pm}{k}\right)^k \left(\frac{qm}{m-k}\right)^{(m-k)}.$$

Also,

$$\left(\frac{pm}{k}\right)^k = \left(\frac{\mu}{(1-\eta)\mu}\right)^{(1-\eta)\mu} = (1-\eta)^{-(1-\eta)\mu} = f(\eta)^{-\eta(1-\eta)\mu}$$
$$= f(\eta)^{-\eta\mu} f(\eta)^{\eta^2\mu},$$
$$\left(\frac{qm}{m-k}\right)^{(m-k)} = \left(1 - \frac{\eta p}{q + \eta p}\right)^{m(q+\eta p)} = f\left(\frac{\eta p}{q + \eta p}\right)^{\frac{\eta p}{q+\eta p} m(q+\eta p)}$$
$$= f\left(\frac{\eta p}{q + \eta p}\right)^{\eta\mu}.$$

Note that $\eta \geqslant (\eta p)/(q + \eta p)$, so $f(\eta) \leqslant f((\eta p)/(q + \eta p))$, and the claim is proved. □

Proposition A.3. *Given a binomial random variable X with mean μ, $\delta \leqslant 1/3$, and $\delta\mu \geqslant 4$, then*

$$\Pr\bigl[X \leqslant (1 - \delta)\mu\bigr] \geqslant \frac{\delta\sqrt{\mu}}{6} e^{-7\delta^2\mu}.$$

Proof. Applying Proposition A.2,

$$\Pr\bigl[X \leqslant (1-\delta)\mu\bigr] \geqslant \sum_{k=\lceil(1-2\delta)\mu\rceil}^{\lfloor(1-\delta)\mu\rfloor} \Pr[X = k] \geqslant (\delta\mu - 2) \Pr\bigl[X = \lceil(1-2\delta)\mu\rceil\bigr]$$
$$\geqslant \frac{\delta\mu}{2} \frac{(f(2\delta))^{2^2\delta^2\mu}}{3\sqrt{\mu}} \geqslant \frac{\delta\sqrt{\mu}}{6} 3^{-1.5\cdot 4\delta^2\mu} \geqslant \frac{\delta\sqrt{\mu}}{6} e^{-7\delta^2\mu}. \qquad \square$$

Proof of Lemma A.1. For $\delta > 1/3$:

$$\Pr\bigl[X \leqslant (1-\delta)\mu\bigr] \geqslant \Pr[X = 0] \geqslant (1-p)^m = \bigl((1-p)^{p^{-1}}\bigr)^{\mu} \geqslant 4^{-\mu} \geqslant e^{-13\delta^2\mu}.$$

For $4 \leqslant \mu \leqslant 12^2$ and $\delta \leqslant 1/3$: There exists $\delta' \in [\delta, \delta + \mu^{-1})$ such that $(1 - \delta')\mu \in \mathbb{N}$, so $\delta' \leqslant \delta + 1/4 \leqslant 2/3$. Applying Proposition A.2,

$$\begin{aligned}\Pr\bigl[X \leqslant (1-\delta)\mu\bigr] &\geqslant \Pr\bigl[X = (1-\delta')\mu\bigr]\\ &\geqslant \frac{(1/3)^{1.5\delta'^2\mu}}{36} \geqslant \frac{e^{-1.7(\delta^2\mu+2\delta+\mu^{-1})}}{36} \geqslant \frac{e^{-1.7\delta^2\mu}}{5\cdot 36}.\end{aligned}$$

For $\mu \geqslant 12^2$ and $\frac{1}{3} \geqslant \delta \geqslant \frac{1}{3}\mu^{-0.5}$: we note that $\delta\mu \geqslant 4$, so applying Proposition A.3,

$$\Pr\bigl[X \leqslant (1-\delta)\mu\bigr] \geqslant \frac{1}{18}e^{-7\delta^2\mu}.$$

For $\mu \geqslant 12^2$ and $\frac{1}{3}\mu^{-0.5} \geqslant \delta$: let $\delta' = \frac{1}{3}\mu^{-0.5}$. Note that $1/3 \geqslant \delta' \geqslant \delta$, so applying Proposition A.3,

$$\Pr\bigl[X \leqslant (1-\delta)\mu\bigr] \geqslant \Pr\bigl[X \leqslant (1-\delta')\mu\bigr] \geqslant \frac{e^{-7/9}}{18} \geqslant \frac{1}{40}.$$

We conclude that $\Pr[X \leqslant (1-\delta)\mu] \geqslant \frac{1}{180}e^{-13\delta^2\mu}$. □

Proposition A.4. *Consider the following experiment: m balls are randomly put in n bins. Let X_i be the number of balls in the ith bin. Then, for any $1 \leqslant \ell \leqslant n$ and any integer sequence $(\alpha_i)_{1\leqslant i\leqslant \ell}$,*

$$\Pr\left[\bigwedge_{i=1}^{\ell}(X_i > \alpha_i)\right] \leqslant \prod_{i=1}^{\ell}\Pr[X_i > \alpha_i].$$

Proof. Let E_i be the event $X_i > \alpha_i$. Fixing $i > 1$, let $a_j = \Pr[E_1 \wedge \cdots \wedge E_{i-1} \mid X_i = j]$. It is elementary to check that a_j is monotonic non-increasing in j. Thus,

$$\Pr[E_1 \wedge \cdots \wedge E_{i-1} \mid E_i] = \sum_{j>\alpha_i} a_j \frac{\Pr[X_i = j]}{\Pr[E_i]} \leqslant \sum_{j\leqslant \alpha_i} a_j \frac{\Pr[X_i = j]}{1-\Pr[E_i]},$$

and so

$$\begin{aligned}&\Pr[E_1 \wedge \cdots \wedge E_{i-1} \mid E_i]\\ &\quad\leqslant \Pr[E_i]\sum_{j>\alpha_i} a_j \frac{\Pr[X_i = j]}{\Pr[E_i]} + \bigl(1-\Pr[E_i]\bigr)\sum_{j\leqslant\alpha_i} a_j \frac{\Pr[X_i = j]}{1-\Pr[E_i]}\\ &\quad= \sum_j a_j \Pr[X_i = j] = \Pr[E_1 \wedge \cdots \wedge E_{i-1}].\end{aligned} \tag{A.1}$$

We conclude by induction on i that $\Pr[E_1 \wedge \cdots \wedge E_n] \leqslant \Pr[E_1]\Pr[E_2]\cdots\Pr[E_i]$, since

$$\begin{aligned}\Pr[E_1 \wedge \cdots \wedge E_i] &= \Pr[E_1 \wedge \cdots \wedge E_{i-1} \mid E_i]\Pr[E_i]\\ &\leqslant \Pr[E_1 \wedge \cdots \wedge E_{i-1}]\Pr[E_i] \leqslant \Pr[E_1]\Pr[E_2]\cdots\Pr[E_i].\end{aligned}$$

The last inequality follows from the induction hypothesis. □

Proposition A.5. *Under the conditions of Proposition* A.4*, given $\alpha > 0$, denote by $Z = \bigwedge_{i=2}^{n}(X_i > \alpha)$, then $E[X_1 \mid Z] \leqslant E[X_1]$.*

Proof. From Eq. (A.1) *in the proof of* Proposition A.4,

$$\Pr[X_1 > j \mid Z] = \frac{\Pr[(X_1 > j) \wedge \bigwedge_{i=2}^{n}(X_i > \alpha)]}{\Pr[\bigwedge_{i=2}^{n}(X_i > \alpha)]} \leqslant \Pr[X_1 > j].$$

In general, for integer non-negative variable W, we have that

$$E[W] = \sum_{j=0}^{\infty} j \Pr[W = j] = \sum_{j=0}^{\infty} j\big(\Pr[W > j-1] - \Pr[W > j]\big) = \sum_{j=0}^{\infty} \Pr[W > j],$$

so in our case,

$$E[X_1 \mid Z] = \sum_{j=0}^{\infty} \Pr[X_1 > j \mid Z] \leqslant \sum_{j=0}^{\infty} \Pr[X_1 > j] = E[X_1]. \qquad \square$$

References

[AA97] B. Awerbuch, Y. Azar, Buy-at-bulk network design, in: 38th Annual Symposium on Foundations of Computer Science, 1997, pp. 542–547.

[ABF93] B. Awerbuch, Y. Bartal, A. Fiat, Heat & dump: Competitive distributed paging, in: Proceedings of the 34th Annual Symposium on Foundations of Computer Science, Palo Alto, CA, 1993, pp. 22–31.

[ACN00] D. Achlioptas, M. Chrobak, J. Noga, Competitive analysis of randomized paging algorithms, Theoret. Comput. Sci. 234 (2000) 203–218.

[Bar96] Y. Bartal, Probabilistic approximations of metric space and its algorithmic application, in: 37th Annual Symposium on Foundations of Computer Science, October, 1996, pp. 183–193.

[Bar98] Y. Bartal, On approximating arbitrary metrics by tree metrics, in: Proceedings of the 30th Annual ACM Symposium on Theory of Computing, 1998, pp. 183–193.

[BBBT97] Y. Bartal, A. Blum, C. Burch, A. Tomkins, A polylog(n)-competitive algorithm for metrical task systems, in: Proceedings of the 29th Annual ACM Symposium on Theory of Computing, May, 1997, pp. 711–719.

[BDBK+94] S. Ben-David, A. Borodin, R. Karp, G. Tardos, A. Wigderson, On the power of randomization in on-line algorithms, Algorithmica 11 (1) (1994) 2–14.

[BEY98] A. Borodin, R. El-Yaniv, Online Computation and Competitive Analysis, Cambridge Univ. Press, Cambridge, UK, 1998.

[BFM86] J. Bourgain, T. Figiel, V. Milman, On hilbertian subsets of finite metric spaces, Israel J. Math. 55 (1986) 147–152.

[BKRS00] A. Blum, H. Karloff, Y. Rabani, M. Saks, A decomposition theorem and lower bounds for randomized server problems, SIAM J. Comput. 30 (5) (2000) 1624–1661.

[BLMN03a] Y. Bartal, N. Linial, M. Mendel, A. Naor, Limitations to Fréchet metric embedding method, Isreal J. Math. 151 (2006) 111–124.

[BLMN03b] Y. Bartal, N. Linial, M. Mendel, A. Naor, On metric Ramsey-type dichotomies, J. London Math. Soc. 71 (2) (2005) 289–303.

[BLMN04a] Y. Bartal, N. Linial, M. Mendel, A. Naor, On metric Ramsey-type phenomena, Ann. of Math. 162 (2) (2005) 643–709.

[BLMN04b] Y. Bartal, N. Linial, M. Mendel, A. Naor, On some low distortion metric Ramsey problems, Discrete Comput. Geom. 33 (1) (2005) 25–41.

[BLS92] A. Borodin, N. Linial, M. Saks, An optimal online algorithm for metrical task systems, J. Assoc. Comput. Mach. 39 (4) (1992) 745–763.

[BM03] Y. Bartal, M. Mendel, Multi-embeddings and path-approximation of metric spaces, in: Proceedings of the 14th Annual ACM–SIAM Symposium on Discrete Algorithms, 2003, pp. 424–433.

[Bol85] B. Bollobás, Random Graphs, Academic Press, London, 1985.

[FKL+91] A. Fiat, R. Karp, M. Luby, L.A. McGeoch, D.D. Sleator, N.E. Young, Competitive paging algorithms, J. Algorithms 12 (1991) 685–699.

[FM03] A. Fiat, M. Mendel, Better algorithms for unfair metrical task systems and applications, SIAM J. Comput. 32 (6) (2003) 1403–1422.

[FRT03] J. Fakcharoenphol, S. Rao, K. Talwar, A tight bound on approximating arbitrary metrics by tree metrics, J. Comput. System Sci. 69 (3) (2004) 485–497.

[Ind01] P. Indyk, Algorithmic applications of geometric embeddings, in: Proceedings of the 42nd Annual Symposium on Foundations of Computer Science, Las-Vegas, NV, 2001.

[IS98] S. Irani, S. Seiden, Randomized algorithms for metrical task systems, Theoret. Comput. Sci. 194 (1–2) (1998) 163–182.

[KMRS88] A. Karlin, M. Manasse, L. Rudolph, D.D. Sleator, Competitive snoopy caching, Algorithmica 3 (1) (1988) 79–119.

[KP95] E. Koutsoupias, C. Papadimitriou, On the k-server conjecture, J. Assoc. Comput. Mach. 42 (5) (1995) 971–983.

[KRR94] H. Karloff, Y. Rabani, Y. Ravid, Lower bounds for randomized k-server and motion-planning algorithms, SIAM J. Comput. 23 (2) (1994) 293–312.

[KT99] J.M. Kleinberg, E. Tardos, Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields, J. ACM 49 (5) (2002) 616–639.

[Mat92] J. Matoušek, Ramsey-like properties for bi-Lipschitz embeddings of finite metric spaces, Comment. Math. Univ. Carolin. 33 (3) (1992) 451–463.

[MMS90] M. Manasse, L.A. McGeoch, D. Sleator, Competitive algorithms for server problems, J. Algorithms 11 (1990) 208–230.

[MS77] F.J. MacWilliams, N.J.A. Sloane, The Theory of Error-Correcting Codes, Elsevier Science, 1977.

[MS91] L. McGeoch, D. Sleator, A strongly competitive randomized paging algorithm, J. Algorithms 6 (1991) 816–825.

[Sei99] S.S. Seiden, Unfair problems and randomized algorithms for metrical task systems, Inform. and Comput. 148 (2) (1999) 219–240.

[Sei01] S.S. Seiden, A general decomposition theorem for the k-server problem, Inform. and Comput. 174 (2) (2002) 193–202.

[ST85] D.D. Sleator, R.E. Tarjan, Amortized efficiency of list update and paging rules, Comm. ACM 28 (1985) 202–208.