Integrality ratio for Group Steiner Trees and Directed Steiner Trees

Eran Halperin[*] Guy Kortsarz[†] Robert Krauthgamer[*] Aravind Srinivasan[‡]
Nan Wang[§]

Abstract

We present an $\Omega(\log^2 k)$ lower bound on the integrality ratio of the flow-based relaxation for the Group Steiner Tree problem, where k denotes the number of groups; this holds even for input graphs that are *Hierarchically Well-Separated Trees*, introduced by Bartal [*Symp. Foundations of Computer Science*, pp. 184–193, 1996], in which case this lower bound is tight. This relaxation appears to be the only one that have been studied for the problem, as well as for its generalization, the Directed Steiner Tree problem. For the latter problem, our results imply an $\Omega(\frac{\log^2 n}{(\log\log n)^2})$ integrality ratio, where n is the number of vertices in the graph. For both problems, this is the first known lower bound on the integrality ratio that is superlogarithmic in the input size. We also show algorithmically that the integrality ratio for Group Steiner Tree is much better for certain families of instances, which helps pinpoint the types of instances that appear to be most difficult to approximate.

1 Introduction

Group Steiner Tree is a network design problem that generalizes both Set Cover and the Steiner Tree problem. The Directed Steiner Tree problem is a further generalization of Group Steiner Tree. The polynomial-time approximability of these NP-hard problems is not yet understood. In particular, there is an intriguing gap between algorithms that achieve polylogarithmic approximation ratio (in quasi-polynomial time for the latter problem) and a logarithmic hardness of approximation that immediately follows from Set Cover. The only known relaxation for these problems is a natural flow-based linear programming relaxation. We show a polylogarithmic lower bound on the integrality ratio of this relaxation; this is the first such lower bound that is superlogarithmic in the input size. In fact, our bound is nearly tight in the important special case of input graphs which are tree networks. We also present improved approximation algorithms for certain families of instances of the Group Steiner Tree problem, shedding light on the type of instances that appear to be most difficult for the flow-based relaxation.

Our work unravels a major obstacle for achieving a logarithmic approximation ratio for these problems. We thus hope that it will lead to better approximation algorithms (say by an appropriate strengthening of the relaxation), or alternatively, to improved hardness results. We note that there is a (roughly) similar gap in many other optimization problems; several of these problems (e.g., bandwidth and cutwidth) have relaxations whose integrality ratio is at most polylogarithmic (see e.g. [DV01, BV02]), but no superlogarithmic lower bound is known for the integrality ratios of these relaxations.

(a). The Group Steiner Tree problem. The (undirected) Group Steiner Tree problem is the following. Given an undirected graph $G = (V, E)$, a collection of subsets (called groups) $g_1, g_2, \ldots, g_k$ of V, and a weight $w_e \geq 0$ for each edge $e \in E$, the problem is to construct a minimum-weight tree in G that spans at least one vertex from each group g_i. We can assume without loss of generality that there is a distinguished vertex $r \in V$ (called the root) that must be included in the output tree. The case where $|g_i| = 1$ for all i is just the classical Steiner Tree problem; the case where G is a tree (or even a star) can be used to model the set cover problem. A natural flow-based relaxation for this problem is the following. Come up with a *capacity* $x_e \in [0, 1]$ for each edge $e \in E$ so that the capacities can support one unit of flow from r to g_i, separately for each g_i (as opposed to supporting a unit flow simultaneously for all g_i). Subject to this constraint, we want to minimize $\sum_e w_e x_e$. It is easy to check that the feasible solutions which satisfy $x_e \in \{0, 1\}$ for all e, exactly correspond to feasible solutions for the Group Steiner Tree prob-

[*] International Computer Science Institute, Berkeley, CA 94704, USA and Computer Science Division, University of California, Berkeley, CA 94720, USA. Supported in part by NSF grants CCR-9820951 and CCR-0121555 and DARPA cooperative agreement F30602-00-2-0601. Email: {eran,robi}@cs.berkeley.edu

[†] Department of Computer Sciences, Rutgers University, Camden, NJ 08102, USA. Email: guyk@camden.rutgers.edu

[‡] Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA. E-mail: srin@cs.umd.edu. Supported in part by NSF Award CCR-0208005.

[§] Department of Computer Science, University of Maryland, College Park, MD 20742, USA. Email: nwang@cs.umd.edu lem; hence, the above flow-based relaxation is indeed a valid linear programming (LP) relaxation for the problem. This is the only known relaxation for this problem (and for some of its generalizations), and is the main subject of investigation in this paper.

We start with a useful definition from [Bar96]. (Item (ii) is slightly stronger than the original definition from [Bar96], but can be assumed without loss of generality due to the analyses of [Bar96, Bar98, KRS01].)

Definition 1.1. *Let $c > 1$. A c-*Hierarchically Well-Separated Tree *(c-HST) is a rooted weighted tree such that (i) all leaves are at the same distance from the root, (ii) the edges in the same level are equal-weighted, and (iii) the weight of an edge is exactly $1/c$ times the weight of its parent edge.*

We simply say "HST" when referring to a c-HST for an arbitrary constant $c > 1$.

The first polylogarithmic approximation algorithm for the Group Steiner Tree problem was achieved in the elegant work of [GKR00]. A brief sketch of their $O(\log n \log\log n \log N \log k)$–approximation algorithm, where $n = |V|$ and $N = \max_i |g_i|$, is as follows. First, the powerful results of [Bar98] are used to appropriately reduce the problem to the case where G is a *tree* T, with an $O(\log n \log\log n)$ factor loss in the approximation ratio. T can be furthermore assumed to be a c-HST for any desired constant $c > 1$. Next, solve the flow-based LP relaxation on T and round the fractional solution into an integral solution for T by applying a novel randomized rounding approach that is developed in [GKR00]. It is established in [GKR00] that for any tree T, this randomized rounding leads to an $O(\log N \log k)$–approximation. Thus, for the input graph G, we get an $O(\log n \log\log n \log N \log k)$–approximation. The work of [GKR00] has been extended and expanded in several ways: Their algorithm was derandomized in [CCGG98, Sri01]; an alternative (combinatorial) algorithm is devised in [CEK02]; the loss incurred by the reduction to an HST is improved to $O(\log n \log\log\log n)$ in [BM03]. (We will discuss the Directed Steiner Tree problem below, but just mention for now that the same flow-based relaxation has been shown to have an integrality ratio of $\Omega(\sqrt{k})$ for this problem [ZK02].)

Since the first appearance of a polylogarithmic approximation for the Group Steiner Tree problem (in the conference version of [GKR00] in 1998), there has been much interest in whether the approximation ratio can be improved. One concrete notable question in this regard has been the following: Can we achieve an approximation ratio better than $O(\log N \log k)$ for trees? This is interesting for at least two reasons. First, since [GKR00] shows a reduction to the case of trees as seen above, an improved approximation for trees (or even for the case of c-HSTs for some constant $c > 1$) would directly lead to an improved approximation for general graphs. Further, even the case where G is a star (which is a tree) captures the Set Cover problem for which $o(\log k)$–approximation is hard [Fei98], so there is an intriguing gap even on trees.

Our main technical result is that for any constant $c > 1$, the integrality ratio of the flow-based relaxation for c-HSTs is $\Omega(\log^2 k)$. This bound is in fact tight, since an $O(\log^2 k)$ bound on the integrality ratio holds for c-HSTs; this is an unpublished work, resulting from our discussions with Anupam Gupta and R. Ravi. Recall that the upper bound of [GKR00] for trees in general is $O(\log N \log k)$; our methods show an $\Omega(\log N \log k / \log\log N)$ lower bound on the integrality ratio, even for a class of HSTs. Such log-squared lower bounds had been conjectured by Uri Feige circa 1998. Our integrality ratio lower bound is shown via a random construction. The analysis is somewhat intricate, and requires delving into lower-order terms. We also show that the same lower bound holds also for trees where all weights are the same (i.e., unit-weight trees). Finally, we show randomized rounding algorithms for the flow-based relaxation that lead to improved approximation algorithms for certain special families of HSTs; this sheds light on the type of instances that are most difficult to approximate.

(b). The Directed Steiner Tree problem. This is the directed version of the (undirected) Steiner Tree problem. Given an edge-weighted directed graph that specifies a *root* vertex r and k *terminal* nodes $v_1, v_2, \ldots, v_k$, the goal is to construct a minimum-weight out-branching tree rooted at r, which spans all the terminals v_i. This problem is easily seen to generalize the undirected Group Steiner Tree problem, as well as to be equivalent to the directed Group Steiner Tree problem. Aside of intrinsic interest, this problem is also of current interest, e.g., in the context of multicasting in the Internet (where inter-node distances are often not symmetric). The polynomial-time approximation ratio currently known for this problem is k^ϵ, for any constant $\epsilon > 0$ [CCC+99]; their algorithm extends to a polylogarithmic approximation ratio in quasi-polynomial running time. The flow-based relaxation here is similar: install a capacity $x_e \in [0, 1]$ so that a unit of flow can be shipped from r to v_i, separately for any given i. Intriguingly, it was recently shown in [ZK02] that this relaxation has an integrality ratio of $\Omega(\sqrt{k})$, precluding a polylog(k)–approximation algorithm based on this relaxation. However, the examples constructed in [ZK02] have $k = \Theta(\frac{\log^2 n}{(\log\log n)^2})$; hence, the result of

[ZK02] does not imply an $\omega(\log n)$ integrality ratio. Our lower-bound result above for the Group Steiner Tree problem, implies an $\Omega(\frac{\log^2 n}{(\log\log n)^2})$ lower bound on the integrality gap for the Directed Steiner Tree problem. (Note that the problem is not expected to have a $o(\log n)$–approximation algorithm, since it generalizes Set Cover.)

In summary, this work develops improved/tight lower bounds on the integrality ratio of the only known relaxation for Group Steiner Tree and Directed Steiner Tree; we also prove algorithmically that the integrality ratio for Group Steiner Tree is much better for certain families of instances, pinpointing the type of instances that appear difficult for this relaxation. Our hope is that this will spur new approaches/relaxations for the problem, or alternatively help us determine the limits to its polynomial-time approximability.

2 Lower bounds on the integrality ratio

In this section we start by proving a lower bound of $\Omega(\log^2 k)$ on the integrality ratio of the flow-based relaxation of the Group Steiner Tree problem even on HSTs. In terms of n, the gap is $\Omega(\frac{\log^2 n}{(\log\log n)^2})$. We then point out in Section 2.4 how this immediately leads to a lower bound of $\Omega(\frac{\log^2 n}{(\log\log n)^2})$ on the integrality ratio for the Directed Steiner Tree problem. We only show our Group Steiner Tree lower bound for 2-HSTs; a simple modification leads to the same lower bounds for c-HSTs, for an arbitrary constant $c > 1$.

As in [GKR00], the flow-based relaxation for Group Steiner Tree is as follows:

$$(2.1)\quad \boxed{\begin{array}{ll} \text{Minimize} & \displaystyle\sum_{e\in E} w_e x_e \\[2ex] 0 \le x_e \le 1, & \forall e \in E \\[2ex] \displaystyle\sum_{e\in\delta(S)} x_e \ge 1, & \forall S \subseteq V \text{ s.t. } r \in S \text{ and} \\ & S \cap g_j = \emptyset \text{ for some } g_j \end{array}}$$

Let $\mathbb{T}_n$ be a 2-HST with n nodes and with a collection $\mathcal{G} = \{g_1, g_2, \ldots, g_k\}$ of k groups assigned randomly as follows. The value of k, as well as those of two other parameters H and d, will be defined shortly. The height (i.e., depth) of $\mathbb{T}_n$ is H, and every non-leaf vertex has d children. The root of $\mathbb{T}_n$ is denoted r. As usual, the *level* of a vertex is its depth; r is at level 0, and there are $H + 1$ levels. An edge is said to be at level i iff it connects a vertex at level $i - 1$ to a vertex at level i. Each edge at level i has weight $1/2^i$; thus, for instance, edges incident at r have weight $1/2$. As usual, each group g_j is a subset of the leaves, described as follows. We shall associate a subset $A(\ell) \subseteq \mathcal{G}$ of the groups with each leaf ℓ, and define each group g_j to be the set of leaves ℓ for which $g_j \in A(\ell)$. Thus, by reaching a leaf ℓ by a path from r, we cover all groups in $A(\ell)$. To define $A(\ell)$ for each leaf ℓ, we now recursively and randomly define a set $A(v)$ for each node v in the tree, as follows. Proceed *independently* for each group g_j as follows. We start by letting $g_j \in A(r)$ with probability 1. In general, if $g_j \in A(u)$ for some non-leaf node u, then for each child v of u, we independently put g_j in $A(v)$ with probability $1/2$. Thus, this random process goes top-down in the tree, independently for each group. Note that the number of vertices in $\mathbb{T}_n$ is $n \simeq d^H$, where H is the height of the tree. We set $H = \frac{1}{2}\log k$ and thus $k = 2^{2H} \simeq n^{2/\log d}$. The expected size of every group is $d^H/2^H$.

Parameters and Notation. We will set $d = c_0 \log n = \Theta(\log k \log\log k)$ for some absolute constant $c_0 > 0$; in particular, we take $k = n^{\Theta(1/\log\log n)}$. Throughout, *with high probability* means with probability that is at least, say, $1 - 1/n$. All probabilities refer to the randomness in constructing the instance $\mathbb{T}_n$.

2.1 The fractional solution.

Recall that $d = c_0 \log n$. We start with a couple of propositions which show that if the constant c_0 is sufficiently large, then certain quantities related to our randomly chosen groups stay close to their mean.

PROPOSITION 2.1. *Let c_0 be a sufficiently large constant. Then, with high probability, all groups have size at least $(d/2)^H/3$.*

Proof. Fix j. We now show that if c_0 is large enough, then $\Pr\left[|g_j| < (d/2)^H/3\right] \le 1/n^2$. We may then apply the union bound over all j to conclude the proof.

Let $\delta = 1/4$. Let X_1 be the number of vertices u at level 1 (i.e., children of r) such that $g_j \in A(u)$. Then X_1 has Binomial distribution $X_1 \sim B(d, 1/2)$, so by a Chernoff bound on the lower-tail (see e.g. [MR95]),

$$\Pr\left[X_1 \le (1-\delta)\frac{d}{2}\right] \le e^{-\frac{1}{2}\delta^2\cdot\frac{d}{2}}.$$

Let X_2 be the number of vertices u at level 2 such that $g_j \in A(u)$. Then X_2 has binomial distribution $X_2 \sim B(X_1 \cdot d, 1/2)$. Suppose that $X_1 > (1-\delta)\mathbb{E}[X_1] = (1-\delta)\frac{d}{2}$. Then, it is immediate that X_2 stochastically dominates a random variable $X_2' \sim B((1-\delta)\frac{d}{2}\cdot d, 1/2)$, i.e., $\Pr[X_2 \le t] \le \Pr[X_2' \le t]$ for all t. By applying the Chernoff bound on X_2' we get

$$\Pr\left[X_2' \le (1-\frac{\delta}{2})(1-\delta)(\frac{d}{2})^2\right] \le e^{-\frac{1}{2}(\frac{\delta}{2})^2\cdot(1-\delta)(\frac{d}{2})^2}$$

Continue similarly for $i = 3, \ldots, H$, by defining X_i to be the number of vertices u at level i such that $g_j \in A(u)$, and by assuming that $X_i > (1 - \frac{\delta}{2^{i-1}}) \cdot \ldots \cdot (1 - \frac{\delta}{2})(1 - \delta)(\frac{d}{2})^i$. We get by the Chernoff bound that

$$\Pr\left[X'_i \leq (1 - \frac{\delta}{2^{i-1}}) \cdot \ldots \cdot (1 - \frac{\delta}{2})(1 - \delta)(\frac{d}{2})^i\right]$$
$$\leq e^{-\frac{1}{2}(\frac{\delta}{2^{i-1}})^2 \cdot (1 - \frac{\delta}{2^{i-2}}) \cdot \ldots \cdot (1 - \frac{\delta}{2})(1-\delta)(\frac{d}{2})^i}.$$

For any $0 < \delta' \leq \frac{1}{2}$ we have $1 - \delta' \geq \frac{1}{1+2\delta'} \geq e^{-2\delta'}$. Thus, $(1 - \frac{\delta}{2^{i-1}}) \cdot \ldots \cdot (1 - \frac{\delta}{2})(1 - \delta) \geq e^{-\frac{\delta}{2^{i-2}} - \ldots - \delta - 2\delta} \geq e^{-4\delta} > \frac{1}{3}$. It follows that the tail-bound obtained by applying the Chernoff bound on X'_i is at most $e^{-\Omega((d/8)^i)}$. Applying the union bound on these H events we get that with high probability none of them happens (if the constant c_0 is sufficiently large), and in particular, $X_H \geq (1 - \frac{\delta}{2^{H-1}}) \cdot \ldots \cdot (1 - \frac{\delta}{2})(1 - \delta)(\frac{d}{2})^H \geq \frac{1}{3}(\frac{d}{2})^H$. This concludes the proof of Proposition 2.1. □

The following proposition has a very similar proof; the main difference is that we will now employ Chernoff bounds on the upper-tail.

PROPOSITION 2.2. *Suppose that the constant c_0 is large enough. Then with high probability, the following holds for every level i and every group g_j: If a vertex u at level i is such that $g_j \in A(u)$, then the number of leaves ℓ in the subtree rooted at u which satisfy $g_j \in A(\ell)$, is at most $3(d/2)^{H-i}$.*

Proof. Fix a pair (i, j) and a vertex u at level i s.t. $g_j \in A(u)$. Let $L(u)$ be the set of leaves of the subtree rooted at u, and $A(L(u)) = \bigcup_{v \in L(u)} A(v)$. We now show that if c_0 is large enough, then $\Pr\left[|g_j \bigcap A(L(u))| > 3(d/2)^{H-i}\right] \leq 1/n^3$. We then apply a union bound over all (i, j, u) to conclude the proof.

Let $\delta = 1/4 < \frac{\ln 3}{2}$. Let X_1 be the number of vertices v at level 1 of the subtree rooted at u (i.e., children of u) such that $g_j \in A(v)$. Then X_1 has Binomial distribution $X_1 \sim B(d, 1/2)$, so by a Chernoff bound on the upper-tail (see e.g. [MR95]),

$$\Pr\left[X_1 \geq (1 + \delta)\frac{d}{2}\right] \leq e^{-\frac{\delta^2}{3} \cdot \frac{d}{2}}.$$

Let X_2 be the number of vertices u at level 2 of the subtree rooted at u such that $g_j \in A(u)$. Then X_2 has binomial distribution $X_2 \sim B(X_1 \cdot d, 1/2)$. Suppose that $X_1 < (1 + \delta)\mathbb{E}[X_1] = (1 + \delta)\frac{d}{2}$. Then, it is immediate that X_2 is stochastically dominated by a random variable $X'_2 \sim B((1+\delta)\frac{d}{2} \cdot d, 1/2)$, i.e., $\Pr[X_2 \geq t] \leq \Pr[X'_2 \geq t]$ for all t. By applying the Chernoff bound on X'_2 we get

$$\Pr\left[X'_2 \geq (1 + \frac{\delta}{2})(1 + \delta)(\frac{d}{2})^2\right] \leq e^{-\frac{1}{3}(\frac{\delta}{2})^2 \cdot (1+\delta)(\frac{d}{2})^2}$$

Continue similarly for $l = 3, \ldots, H - i$, by defining X_l to be the number of vertices v at level l of the subtree rooted at u such that $g_j \in A(v)$, and by assuming that $X_l < (1 + \frac{\delta}{2^{l-1}}) \cdot \ldots \cdot (1 + \frac{\delta}{2})(1 + \delta)(\frac{d}{2})^l$. We get by the Chernoff bound that

$$\Pr\left[X'_l \geq (1 + \frac{\delta}{2^{l-1}}) \cdot \ldots \cdot (1 + \frac{\delta}{2})(1 + \delta)(\frac{d}{2})^l\right]$$
$$\leq e^{-\frac{1}{3}(\frac{\delta}{2^{l-1}})^2 \cdot (1 + \frac{\delta}{2^{l-2}}) \cdot \ldots \cdot (1 + \frac{\delta}{2})(1+\delta)(\frac{d}{2})^l}.$$

Thus, the tail-bound obtained by applying the Chernoff bound on X'_l is at most $e^{-\Omega((d/8)^l)}$. Applying the union bound on these $H - i \leq H$ events we get that with probability at least $1 - 1/n^3$ none of these events happen, if the constant c_0 is sufficiently large; in particular, $X_{H-i} \leq (1 + \frac{\delta}{2^{H-i-1}}) \cdot \ldots \cdot (1 + \frac{\delta}{2})(1 + \delta)(\frac{d}{2})^{H-i} \leq 3(\frac{d}{2})^{H-i}$. (This is because of the following. For any δ' we have $1 + \delta' \leq e^{\delta'}$. Thus, $(1 + \frac{\delta}{2^{l-1}}) \cdot \ldots \cdot (1 + \frac{\delta}{2})(1 + \delta) \leq e^{\frac{\delta}{2^{l-1}} + \ldots + \delta} \leq e^{2\delta} < 3$.) This concludes the proof of Proposition 2.2. □

We now upper bound the value of LP (2.1) for the tree $\mathbb{T}_n$ by exhibiting a feasible solution for it: let each edge e at each level i have value $\hat{x}_e = 9 \cdot (2/d)^i$.

LEMMA 2.1. *With high probability, $\hat{x}$ is a feasible solution to LP (2.1). Its value is $9H$.*

Proof. Observe that $\hat{x}$ satisfies the constraints of LP (2.1) if (see also [GKR00]), for every group g_j, every cut $(S, \bar{S})$ separating r from all the vertices of g_j has capacity at least 1, where the capacity of each edge e is $\hat{x}_e$. By the (single-source) max-flow min-cut theorem (or, say, weak duality) it suffices to show that for every group g_j, a unit of flow can be shipped from the root r to the vertices of g_j while obeying the "capacity" $\hat{x}_e$ of each edge e. To this end, fix a group g_j and define the flow f as follows. For every vertex v in g_j (i.e., for every leaf v such that $g_j \in A(v)$), ship $3 \cdot (2/d)^H$ units of flow along the unique simple path from r to v. By Proposition 2.1, the total flow shipped to g_j is at least $|g_j| \cdot 3 \cdot (2/d)^H \geq 1$. Next, consider a node u at some level i, for which $g_j \in A(u)$. By Proposition 2.2, the total flow shipped through u is at most $3(d/2)^{H-i} \cdot 3(2/d)^H = 9(2/d)^i$, obeying the capacity of the edge between u and its parent. We conclude that with high probability $\hat{x}$ is a feasible solution.

The value of the solution $\hat{x}$ is $\sum_{i=1}^{H} d^i \cdot 1/2^i \cdot 9(2/d)^i = 9H$ since each level i contains d^i edges of weight $1/2^i$. □

2.2 The integral solution. We now show that with high probability (over the random choice of the groups), all integral solutions have value $\Omega(H^2 \log k)$. Whenever we say that some T' is a subtree of $\mathbb{T}_n$, we allow T' to be an arbitrary connected subgraph of $\mathbb{T}_n$. Since $\mathbb{T}_n$ is rooted, any subtree T' of $\mathbb{T}_n$ is also thought of as rooted in the obvious way: the node in T' of the smallest depth is the root of T' (and is denoted $\mathrm{root}(T')$). Also, when we say that some T' is a subtree of $\mathbb{T}_n$ with root u, we allow T' to be an arbitrary connected subgraph of $\mathbb{T}_n$ with root u.

Let $M(c)$ be the number of subtrees of $\mathbb{T}_n$ which are rooted at r and have total weight at most c. Fix $g_j \in \mathcal{G}$. For any given subtree T' of $\mathbb{T}_n$, let $p(T')$ be the probability that no leaf of T' belongs to the group g_j, conditioned on the event that $g_j \in A(\mathrm{root}(T'))$. We now define a key value $f(H,i,c)$ as follows. Choose an arbitrary vertex u at level i. Then $f(H,i,c)$ is the minimum value of $p(T')$, taken over all possible subtrees T' that are rooted at u and have total weight at most c. (If there is no such T', then $f(H,i,c) = 1$. Also, it is easy to see by symmetry that $f(H,i,c)$ does not depend upon the choice of j or u.) Let P_c be the probability that there exists an integral solution of weight c. We wish to show that $P_c = o(1)$ for c that is smaller than a certain threshold of the order $H^2 \log k$. Using the independence between the different groups and applying a union bound over all possible subtrees rooted at r that have total weight c, we obtain

$$P_c \le M(c)(1 - f(H,0,c))^k. \tag{2.2}$$

We now have to lower bound f and upper bound M. We employ the following crude bound on $M(c)$. Note that it suffices to count only subtrees of $\mathbb{T}_n$ that are minimal with respect to containment; each such tree is defined by the leaves that it spans (since the groups g_i contain only leaves). Observing that $\mathbb{T}_n$ has d^H leaves, and a subtree of total weight at most c spans at most $c2^H$ leaves (since each spanned leaf requires a distinct edge at level H), we get that

$$M(c) \le \binom{d^H}{c2^H} \le d^{cH2^H}.$$

Bounding $f(H,h,c)$**.** We start with some preliminaries. The main technical result is Lemma 2.2 below.

Proposition 2.3. *Let $l \ge 2$ and $\beta > 0$. Then the minimum of $\sum_{S \subseteq \{1,\dots,l\}} \prod_{i\in S} e^{-\beta x_i}$ over all $(x_1,\dots,x_l)$ with a given $\sum_{i=1}^{l} x_i$ is attained when all x_i are equal.*

Proof. The minimum is clearly attained at some point $(x_1,\dots,x_l)$, so assume to the contrary that at this point not all x_i are equal, say without loss of generality that $x_1 > \sum_i x_i/l > x_2$. We will show that changing both x_1 and x_2 to $\frac{x_1+x_2}{2}$ decreases the above sum while maintaining $\sum_i x_i$, which contradicts the assumption that $(x_1,\dots,x_l)$ is a minimum point. Actually, it suffices to prove that

$$\sum_{S' \subseteq \{1,2\}} \prod_{i\in S'} e^{-\beta x_i} > \sum_{S' \subseteq \{1,2\}} \prod_{i\in S'} e^{-\beta\cdot\frac{x_1+x_2}{2}}\ , \tag{2.3}$$

since multiplying (2.3) by $\prod_{i\in S''} e^{-\beta x_i}$ and summing over all $S'' \subseteq \{3,\dots,l\}$ shows that changing x_1, x_2 indeed decreases the above-mentioned sum. To prove (2.3), observe that it simplifies to

$$e^{-\beta x_1} + e^{-\beta x_2} > 2e^{-\beta(x_1+x_2)/2}$$

which follows from the arithmetic mean-geometric mean inequality since $x_1 \ne x_2$. This completes the proof of Proposition 2.3. □

Next, note (say, by Taylor's Theorem) that there exists a constant $0 < B_0 \le \frac{1}{2}$, such that for all $B \le B_0$

$$e^{-B} \ge 1 - B + \frac{B^2}{2} - \frac{B^3}{6}. \tag{2.4}$$

Proposition 2.4. *There exists a constant $\delta > 0$ such that for all $B \ge B_0$ we have $\frac{1+e^{-B}}{2} \ge e^{-\frac{B}{2+\delta}}$.*

Proof. We first make sure that the inequality holds at B_0. By the arithmetic mean-geometric mean inequality $\frac{1+e^{-B_0}}{2} > e^{-B_0/2}$ (since $B_0 > 0$) so a sufficiently small $\delta > 0$ satisfies $\frac{1+e^{-B_0}}{2} > e^{-B_0/(2+\delta)}$. It now suffices to make sure that for all $B \ge B_0$ the derivative of the lefthand side is at least that of the righthand side, i.e., that $-\frac{1}{2}e^{-B} \ge -\frac{1}{2+\delta}e^{-B/(2+\delta)}$. This holds for any $0 < \delta < B_0$ since $\frac{2+\delta}{2} = 1 + \delta/2 \le 1 + B_0/2 \le e^{B_0/2} \le e^{B/2} \le e^{B-B/(2+\delta)}$, completing the proof of Proposition 2.4. □

Lemma 2.2. *Let γ be a sufficiently large constant. Then $f(H,h,c) \ge \exp(-\frac{\gamma c 2^h}{(H-h)^2})$ for all $c > 0$ and all $0 \le h \le H-1$.*

Proof. The proof is by backward induction on h, i.e., we assume that the claim holds for $h+1$ and prove it for h, where $h \le H-2$. (We will consider the base case of the induction later on.) In order to bound f, we derive a recurrence relation for $f(H,h,c)$. Recall the definition of $f(H,h,c)$: fix an arbitrary vertex u at level h, and take the minimum value of $p(T')$, over all possible subtrees T' rooted at u such that the total weight of T' is at most c. We bound $f(H,h,c)$ by considering all possibilities of u having $l = 1,2,\dots,d$ children and all possible partitions $\vec{x}^{(l)} = (x_1,x_2,\dots,x_l)$ of the weight c to (the subtrees under) these l children; since the edge from u to each of its children has weight $\frac{1}{2^{h+1}}$, we get that $\sum_{i=1}^{l} x_i = c - \frac{l}{2^{h+1}}$. We then get that

$$f(H,h,c) \geq \min_{1\leq l\leq d} \min_{\vec{x}^{(l)}\in D} \psi_1(\vec{x}^{(l)}),$$

where $D = \{\vec{x}^{(l)} \geq 0 : \sum_{i=1}^{l} x_i = c - \frac{l}{2^{h+1}}\}$ and

$$\psi_1(\vec{x}^{(l)}) = \frac{1}{2^l} \sum_{S\subseteq\{1,\ldots,l\}} \prod_{i\in S} f(H,h+1,x_i);$$

since once the l children of u are chosen, we only need to consider the subset S in them that contain g_j in their $A(\cdot)$ set. (Each such set S occurs with probability $1/2^l$.) Plugging the induction hypothesis in, we get that

$$f(H,h,c) \geq \min_{1\leq l\leq d} \min_{\vec{x}^{(l)}\in D} \psi_2(\vec{x}^{(l)}), \tag{2.5}$$

where

$$\psi_2(\vec{x}^{(l)}) = \frac{1}{2^l} \sum_{S\subseteq\{1,\ldots,l\}} \prod_{i\in S} \exp(-\frac{\gamma x_i 2^{h+1}}{(H-h-1)^2}).$$

For any l, we have by Proposition 2.3 that the righthand side of (2.5) is minimized when all x_i are equal to $\frac{c}{l} - \frac{1}{2^{h+1}}$. We thus get that

$$\begin{aligned} & f(H,h,c) \\ \geq\ & \min_{1\leq l\leq d} \frac{1}{2^l} \sum_{S\subseteq\{1,\ldots,l\}} \left(\exp\left(-\frac{\gamma(\frac{c}{l}-\frac{1}{2^{h+1}})2^{h+1}}{(H-h-1)^2}\right)\right)^{|S|} \\ =\ & \min_{1\leq l\leq d} \frac{1}{2^l} \sum_{i=0}^{l} \binom{l}{i} \left(\exp\left(-\frac{\gamma(\frac{c}{l}-\frac{1}{2^{h+1}})2^{h+1}}{(H-h-1)^2}\right)\right)^{i} \\ =\ & \min_{1\leq l\leq d} \left(\frac{1+\exp\left(-\frac{\gamma(\frac{c}{l}-\frac{1}{2^{h+1}})2^{h+1}}{(H-h-1)^2}\right)}{2}\right)^{l}. \end{aligned}$$

Fix l arbitrarily such that $1 \leq l \leq d$. Let $B = \dfrac{\gamma(\frac{c}{l}-\frac{1}{2^{h+1}})2^{h+1}}{(H-h-1)^2}$ and $C = \dfrac{\gamma\frac{c}{l}2^h}{(H-h)^2}$. To complete the induction, we want to prove that $\left(\frac{1+e^{-B}}{2}\right)^l \geq e^{-Cl}$, i.e., that

$$\frac{1+e^{-B}}{2} \geq e^{-C}. \tag{2.6}$$

We have four cases.

Case 1: In this case we assume that $C \geq \frac{B}{2}$. By the arithmetic mean-geometric mean inequality we have that $\frac{1+e^{-B}}{2} \geq e^{-B/2} \geq e^{-C}$, which proves (2.6).

Case 2: In this case we assume that $C \leq \frac{B}{2}$ and $B < B_0$. Then by (2.4) we have $\frac{1+e^{-B}}{2} \geq 1 - \frac{B}{2} + \frac{B^2}{4} - \frac{B^3}{12}$. Since $C \geq 0$, we have (by Taylor's Theorem) that $e^{-C} \leq 1 - C + \frac{C^2}{2}$. Thus, it suffices to prove that

$$1 - \frac{B}{2} + \frac{B^2}{4} - \frac{B^3}{12} \geq 1 - C + \frac{C^2}{2}.$$

Since $B < B_0 \leq \frac{1}{2}$ we have that $\frac{B^3}{12} \leq \frac{B^2}{24}$, and then since $2C \leq B$, we have that $\frac{B^2}{4} - \frac{B^3}{12} \geq \frac{5B^2}{24} \geq \frac{5C^2}{6}$. It therefore suffices to prove that

$$C + \frac{C^2}{3} \geq \frac{B}{2}.$$

Note that

$$\frac{B}{2} - C \leq \frac{\gamma 2^{h+1}\frac{c}{l}}{(H-h-1)^2(H-h)} - \frac{\gamma}{2(H-h-1)^2}.$$

Plugging in the values of B and C and simplifying we get that it suffices to prove that

$$\frac{\gamma 2^{h+1}\frac{c}{l}}{(H-h-1)^2(H-h)} \leq \frac{\gamma}{2(H-h-1)^2} + \frac{\gamma^2\frac{c^2}{l^2}2^{2h}}{3(H-h)^4}.$$

If $\frac{2^{h+2}\frac{c}{l}}{H-h} \leq 1$, then the desired inequality indeed holds since $\frac{\gamma 2^{h+1}\frac{c}{l}}{(H-h-1)^2(H-h)} \leq \frac{\gamma}{2(H-h-1)^2}$. Otherwise, the inequality holds for any $\gamma \geq 96$, since then,

$$\begin{aligned} \frac{\gamma^2\frac{c^2}{l^2}2^{2h}}{3(H-h)^4} &= \frac{\gamma}{6}\cdot\frac{2^{h+2}\frac{c}{l}}{H-h}\cdot\frac{\gamma\frac{c}{l}2^{h-1}}{(H-h)^3} \\ &\geq 16\frac{\gamma\frac{c}{l}2^{h-1}}{(H-h)^3} \\ &\geq \frac{\gamma 2^{h+1}\frac{c}{l}}{(H-h-1)^2(H-h)}. \end{aligned}$$

Case 3: In this case we assume that $B \geq B_0$ and $\frac{B}{2+\delta} \leq C \leq \frac{B}{2}$. Then we have from Proposition 2.4 that $\frac{1+e^{-B}}{2} \geq e^{-\frac{B}{2+\delta}} \geq e^{-C}$, which proves (2.6).

Case 4: In this case we assume that $C < \frac{B}{2+\delta}$. Note that for $h \leq H-2$,

$$\begin{aligned} 2+\delta \leq \frac{B}{C} &= 2\frac{\frac{c}{l}-\frac{1}{2^{h+1}}}{\frac{c}{l}}\cdot\frac{(H-h)^2}{(H-h-1)^2} \\ &\leq 2\frac{(H-h)^2}{(H-h-1)^2} \\ &\leq 2+\frac{6}{H-h-1}. \end{aligned}$$

Thus, $h \geq H - 1 - \frac{6}{\delta}$. Since $\delta > 0$ is a constant, this is really the base case of the induction, which we shall prove directly. Consider a subtree T' of weight at most c that is rooted at a vertex u at level h. Since u has at most $c2^{h+1}$ children in T', each *not* having the group

g_j in its $A(\cdot)$ set independently with probability 1/2, with probability at least $2^{-c2^{h+1}}$ the subtree T' does not cover g_j. Thus, $f(H,h,c) \geq e^{-c2^{h+1}}$. Choosing a constant $\gamma \geq 2(1+\frac{6}{\delta})^2$, we get that $\gamma \geq 2(H-h)^2$, and thus

$$f(H,h,c) \geq e^{-c2^{h+1}} \geq \exp(-\frac{c2^h \cdot \gamma}{(H-h)^2}).$$

This concludes the proof of Lemma 2.2. □

Bounding the weight of an integral solution. We have from Lemma 2.2 that $f(H,0,c) \geq e^{-\gamma \frac{c}{H^2}}$. Plugging into (2.2) we get that

$$\begin{aligned} P_c &\leq M(c) \cdot \exp\{-k \cdot f(H,0,c)\} \\ &\leq \exp\{cH2^H \log d - ke^{-\gamma\frac{c}{H^2}}\}. \end{aligned}$$

Now, suppose that $c \leq \frac{1}{4\gamma}H^2 \ln k$. Then $cH2^H = O(2^H H^3 \log k)$. Recalling that $H = \frac{1}{2}\log k$, we have

$$P_c \leq \exp\{\tilde{O}(\sqrt{k}) - \Theta(k^{3/4})\} = o(1).$$

We conclude that with high probability no subtree of weight at most $\frac{1}{4\gamma}H^2 \log k$ covers all the groups, and thus an optimal integral solution has value at least $\Omega(H^2 \log k)$. Since LP (2.1) has a fractional feasible solution of value $9H$, the integrality ratio is $\Omega(\log^2 k)$. Note that in terms of N, k, the integrality gap is $\Omega(\log k \log N / \log\log N)$ and in terms of n it is $\Omega(\frac{\log^2 n}{(\log\log n)^2})$.

2.3 Integrality ratio for unit-weight trees.

The above analysis gives a lower bound on the integrality gap for HSTs. A consequent interesting question is whether the LP is tighter for unit-weight trees. We show here that a slight modification of the trees described above gives the same integrality ratio for unit-weight trees. The basic idea is very simple. Recall that in our random construction, edges at level i had weight $1/2^i$; replacing each such edge by a path of 2^{H-i} unit-weight edges does not really change our integrality ratio argument. We now formally prove this.

Consider first the 2-HST $\mathbb{T}_n$ defined above. The fractional solution for $\mathbb{T}_n$ is at most $9H$, and the integral solution is at least $\Omega(H^2 \log k)$, where $H = \frac{\log k}{2}$. We construct from $\mathbb{T}_n$ a unit weight tree $\mathbb{T}'_n$ in the following way. Replace each edge at level i in $\mathbb{T}_n$ by a path of 2^{H-i} unit weight edges. In the resulting tree $\mathbb{T}'_n$, all the groups are still in the leaves, and they are actually in the same leaves they were in $\mathbb{T}_n$. For each edge e in $\mathbb{T}'_n$, we say that e is in *original* level i if e is on a path originating from an edge at level i in $\mathbb{T}_n$.

The fractional solution is the following. For every edge $e \in \mathbb{T}'_n$ in *original* level i, we set $\hat{x}_e = 9(\frac{2}{d})^i$. It is easy to see that this is indeed a feasible solution since every flow from the root to a group in $\mathbb{T}'_n$ which satisfies the capacity constraints, corresponds to a flow in $\mathbb{T}_n$ under the capacity constraints where $x_e = 9(\frac{2}{d})^i$ for a level i edge e. This is true, since whenever a flow enters a path, it can push the flow down the path without violating any constraints, since on the path all the capacities are the same. It is also easy to check that the value of the fractional solution is $9H2^H$.

We now lower bound any integral solution in $\mathbb{T}'_n$. Consider an optimal integral solution *OPT'* for $\mathbb{T}'_n$. Consider a path $e_1, \ldots, e_t$ in $\mathbb{T}'_n$ that originates at an edge $e \in \mathbb{T}_n$. Clearly, either *OPT'* contains all the edges of the path or it contains none of them. For each edge $e \in \mathbb{T}_n$, let X_e be an indicator to the event that the path corresponding to e (in $\mathbb{T}'_n$) is in *OPT'*. Thus, if e is of original level i, the contribution of its corresponding path to *OPT'* is $2^{H-i}X_e$. Consider the integral solution to $\mathbb{T}_n$ formed by taking all edges e with $X_e = 1$. The value of this solution is $INT = \sum_{e \in E} w_e X_e \geq \Omega(H^2 \log k)$. Note that the contribution to INT of an edge e at original level i is $X_e w_e = \frac{X_e}{2^i}$. Thus, $OPT' = 2^H INT = \Omega(2^H H^2 \log k)$, and the integrality ratio is still $\Omega(H \log k) = \Omega(\log^2 k)$.

2.4 Integrality ratio for Directed Steiner Tree.

The above results immediately lead to a lower bound of $\Omega(\frac{\log^2 n}{(\log\log n)^2})$ on the integrality ratio for the Directed Steiner Tree problem. Let I be an instance as described above with $\Omega(\frac{\log^2 n}{(\log\log n)^2})$ integrality ratio for Group Steiner Tree, and construct a Directed Steiner Tree instance as follows. Orient all the edges of I away from the root r. Then, introduce new nodes $v_1, v_2, \ldots, v_k$, and for each j and each $u \in g_j$, introduce a zero-weight arc from u to v_j. This defines a Directed Steiner Tree instance I' which is essentially the same as I: fractional and integral solutions for the problems map bijectively, with identical total weights. Observe that the number of vertices in the resulting graph is $n+k \leq 2n$, and thus the lower bound of $\Omega(\frac{\log^2 n}{(\log\log n)^2})$ on the integrality ratio for I holds also for I'.

3 Improved approximations for certain families of trees

What are the Group Steiner Tree instances (in particular, trees) which are difficult to approximate better than within ratio $O(\log k \log N)$? We partially answer this question by presenting a significantly better approximation ratio for a certain family of trees, which differs from the trees constructed in Section 2 in a crucial way. Fix a

Group Steiner Tree instance on an arbitrary tree T, and an (optimal) solution to its flow-based relaxation. Define z_i^* to be the total contribution of the edges at level i (of T) to the objective function of the relaxation. We show that the relationship between the different z_i^* plays a crucial role in the strength/weakness of the LP: If for some constant $\alpha > 1$ we have $z_{i+1}^* \geq \alpha z_i^*$ for all i, then we can achieve an $O(\log k \cdot \log\log(kN))$–approximation.

This approximation ratio may suggest that instances with $z_i^* \simeq z_{i+1}^*$ for all/most i are among the worst cases for the relaxation. Indeed, the instances $\mathbb{T}_n$ and $\mathbb{T}'_n$ constructed in Section 2 have the same z_i^* values for all i. This approximation ratio of $O(\log k \cdot \log\log(kN))$ also elucidates a disparity between the integrality ratio of a relaxation and the performance of a rounding procedure for the relaxation. It is relatively straightforward to show that the rounding procedure of [GKR00] produces integral solutions that are within factor $\Omega(\log k \log N)$ of the relaxation not only on $\mathbb{T}_n$ and $\mathbb{T}'_n$ but also on their "simpler" variants with all edges having unit weight. However, these instances satisfy $z_{i+1}^* = 2z_i^*$ and thus do not yield the desired $\Omega(\log^2 k)$ integrality ratio.

The following lemma proves the improved approximation ratio for the case where the z_i^* values increase (at least) by a factor of 2. The argument easily extends to any constant factor greater than 1. We sometimes refer to a valid (integral) Group Steiner tree simply as a cover.

Lemma 3.1. *If $z_{i+1}^* \geq 2z_i^*$ for all i then we can find a cover of size $O(z^* \cdot \log k \cdot \log\log(kN))$, where z^* denotes the optimal LP value.*

Proof. Note that $z^* = \sum_i z_i^* \leq 2z_H^*$. It is straightforward to assume that all groups contain only leaves of T, by adding zero weight edges. Let L_i be the set of edges at level i. Let $h = 2\log\log N$. Let $U = \{e : e \in L_i \text{ for } i \leq H - h\}$, and $L = \{e : e \in L_i \text{ for } i > H - h\}$. We first construct a new tree T' in the following way. For every $e \in U$, let y_e be x_e rounded upwards to the nearest power of 2, increasing the LP value by a factor of at most 2. Let t be the smallest value such that $y_e \geq 1/2^t$ for every edge e in L_{H-h} and $t > c_1 \log\log(kN)$ for a sufficiently large absolute constant. Let $e \in L_{H-h}$ be such that $y_e > 1/2^t$. Let T_e be the subtree of e. Duplicate T_e (including the edge e) and let T'_e be the copy of T_e. Let both T_e and T'_e be rooted at the same vertex where T_e was rooted, adding the new edges to U,L and L_i. For every edge in T_e and T'_e (including the edge e and its copy) we halve its y value. We continue this procedure until all $e \in L_{H-h}$ have the same y value, i.e., $y_e = 1/2^t$. Let T' be the resulting tree.

The fractional solution in T extends to T' (with the same LP value), and, furthermore, any cover of T' can be translated to a cover of T: Given a solution for T', let $S_e \subset T_e$ and $S'_e \subset T'_e$ be its restriction to the copies of T_e, and then in T we take $S_e \cup S'_e \subset T_e$ be the solution. Clearly, the solution in T is less expensive than the solution in T', and thus it suffices to find an integral solution in T'.

We now find a small cover in T' as follows. For every $e \in U$, assign $\hat{x}_e = \min\{1, y_e \cdot \log k \log^2 N\}$, and use one iteration of the rounding scheme presented in [GKR00] to solve the problem in U. The expected total weight of this solution is $z^*(U) \log k \log^2 N \leq z^* \log k$, where $z^*(U) = \sum_{i=1}^{H-h} z_i^*$ is the total contribution to z^* of the edges in U. Consider a group g. Let $e_1, \ldots, e_m$ be the leaves of (the subtree induced on) U that "lead" to g (i.e., g contains at least one of their descendants in T'). Note that if c_1 is large enough then $\hat{x}_{e_i} = \frac{\log k \log^2 N}{2^t}$ for all $1 \leq i \leq m$. Let $f_1, \ldots, f_m$ be the flows to g on the edges $e_1, \ldots, e_m$ under the original LP values x_e. Clearly, $\sum_i f_i \geq 1$. Partition the m flows, letting $A_i = \{j : \frac{1}{2^{t+i}} < f_j \leq \frac{1}{2^{t+i-1}}\}$. Let $B(g) = \{i : |A_i| \geq 2^{t-2}\}$ consist of "big" sets A_i. It is easy to see that the total flow in the remaining sets A_i is at most $\frac{1}{2}$, and thus $\sum_{i \in B(g)} \frac{|A_i|}{2^{i+t}} \geq \frac{1}{2}$. Let V_i be the set of vertices of A_i chosen by the [GKR00] procedure. For every $i \in B(g)$ we have that The expectation of $|V_i|$ is $\mu_i = \sum_{e \in A_i} \hat{x}_e = \frac{|A_i| \log k \log^2 N}{2^t} \geq \frac{\log k \log^2 N}{4}$. By Janson's inequality,

$$\Pr(|V_i| \leq \mu_i/2) \leq e^{-\Omega(\frac{\mu_i}{2+\Delta_i/\mu_i})},$$

where $\Delta_i = \sum_{e \sim e'} \Pr[e \text{ and } e' \text{ are chosen}]$; here, the sum is over pairs of distinct edges e and e' whose events of being chosen are not independent. By the proof in [GKR00], and by the fact that $|A_i| \geq 2^{t-2}$, it is easy to see that $\frac{\mu_i}{2+\Delta_i/\mu_i} \geq \Omega(\log k \log N)$, and thus,

$$\Pr(|V_i| \leq \mu_i/2) \leq e^{-\Omega(\log k \log N)}.$$

For any group, one can bound the number of sets A_i by a fixed polynomial in kN, even in T' (i.e., after the duplication of edges at level $H - h$ and below). We thus get by the union bound that with high probability, for every group g and every $i \in B(g)$, $|V_i| = \Omega(\mu_i) = \Omega(\sum_{e \in A_i} \hat{x}_e)$. Thus, the total flow that can be shipped into g using only the leaves of U chosen by the [GKR00] procedure is at least

$$\begin{aligned} \sum_{i \in B(g)} \frac{|V_i|}{2^{i+t}} &= \Omega\left(\sum_{i \in B(g)} \frac{|A_i| \log k \log^2 N}{2^{i+2t}}\right) \\ &= \Omega\left(\frac{\log k \log^2 N}{2^t} \sum_{i \in B(g)} \frac{|A_i|}{2^{i+t}}\right) \end{aligned}$$

$$= \quad \Omega\left(\frac{\log k \log^2 N}{2^t}\right).$$

For every $e \in L \cup L_{H-h}$, set $\hat{x}_e = \frac{2^t}{\log k \log^2 N} x_e$. Now apply the rounding algorithm of [GKR00] to L with the values $\hat{x}_e$, starting from every chosen vertex of U. Clearly, $\{\hat{x}_e\}$ satisfies the LP constraints since the flow to every group is $\Omega(1)$ (due to the above equality and the fact that $\frac{1}{2^t} \cdot \log k \log^2 N \leq 1$). It is proven in [GKR00], that after $O(h \log k)$ iterations of the rounding scheme, with high probability all the groups are covered. We now claim that the expected size of each such iteration is at most z^*. Consider an edge e, and let e' be its lowest ancestor in U. The probability that e will be chosen is the probability that e' will be chosen times the probability of choosing e given that e' is chosen. The probability that e' will be chosen in the first part of the algorithm is $\Theta(\hat{x}_{e'}) = \Theta(\frac{\log k \log^2 N}{2^t})$. The probability of choosing e given that e' is chosen is $\hat{x}_e = \frac{2^t}{\log k \log^2 N} x_e$. Thus, $\hat{x}_{e'} \hat{x}_e = x_e$. The claim now follows by the linearity of expectation.

Therefore, the expected cost of this solution is $O(z^* \max\{h \log k, \log k \cdot \log\log(kN)\}) = O(z^* \log k \cdot \log\log(kN))$. □

4 Discussion

Our results improve the current understanding of the integrality ratio of the flow-based relaxation for the Group Steiner Tree problem, but some very intriguing gaps still remain. Although for HSTs our $\Omega(\log^2 k)$ lower bound is tight, for general trees there is a slight slackness between our $\Omega(\log k \log N / \log\log N)$ lower bound and the $O(\log k \log N)$ upper bound of [GKR00]. Interestingly, an $O(\log^2(kN)/\log\log(kN))$–approximation by a quasi-polynomial time algorithm is devised in [CEK02]; their algorithm is combinatorial (i.e., not LP-based). Does their algorithm hint that the known upper bound on the integrality ratio in trees is not tight? Or maybe it hints that there is a separation between polynomial and quasi-polynomial (approximation) algorithms?

A possible step towards closing this gap (in the integrality ratio on trees) is to analyze the following instance suggested by Uri Feige circa 1998: Take a complete tree of arity 4 (i.e., every non-leaf vertex has 4 children) and height $\log_2 k$; now generate k groups, each containing k leaves, by an independent randomized branching process that starts from the root and picks two out of four children until the leaves are reached.

For general graphs, there is an even bigger slackness, as the known upper bound is $\tilde{O}(\log n \log k \log N)$ [GKR00] and the lower bound is just the lower bound for trees described above. It is worth noting that a significantly better upper bound can be achieved in (general) graphs of small diameter. In particular, an $O(\log k)$ upper bound for expander graphs is shown in [BM03]; this bound is tight since expanders contain a large star metric. We therefore set forth the following question, which was formulated together with Yair Bartal: What is the integrality ratio for the Group Steiner Tree problem on a (say two-dimensional) grid?

The shortest-path metric of a grid contains, up to constant distortion, an HST which is a complete regular tree (see e.g. [BBM01]). This tree is similar to our tree $\mathbb{T}_n$ (and to Feige's tree described above), but differs in parameters like arity and weight; thus, one may suspect that the integrality ratio in grids is at least as large as in HSTs. In comparison, the best upper bound that we are aware of for two-dimensional grids is $O(\log n \log k \log N)$, by employing the [GKR00] approach with a specialized reduction of the grid to HST (using e.g. [KRS01]).

Acknowledgments. Our work has been influenced by discussions with several people. We thank Moses Charikar, Chandra Chekuri, Uri Feige, Anupam Gupta, Sanjeev Khanna, Adam Meyerson, Seffi Naor, R. Ravi, and Leonid Zosin for very helpful conversations. We also thank the SODA 2003 reviewers for their helpful comments.

References

[Bar96] Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *37th Annual Symposium on Foundations of Computer Science*, pages 184–193. IEEE, 1996.

[Bar98] Y. Bartal. On approximating arbitrary metrics by tree metrics. In *30th Annual ACM Symposium on Theory of Computing*, pages 161–168. ACM, 1998.

[BBM01] Y. Bartal, B. Bollobás, and M. Mendel. A Ramsey-type theorem for metric spaces and its applications for metrical task systems and related problems. In *42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 396–405, October 2001.

[BM03] Y. Bartal and M. Mendel. Multi-embedding and path approximation of metric spaces. These proceedings, 2003.

[BV02] C. F. Bornstein and S. Vempala. Flow metrics. In *LATIN 2002: Theoretical Informatics, 5th Latin American Symposium*, pages 516–527. Springer, 2002.

[CCC$^+$99] M. Charikar, C. Chekuri, T. Cheung, Z. Dai, A. Goel, S. Guha, and M. Li. Approximation algorithms for directed Steiner problems. *J. Algorithms*, 33(1):73–91, 1999.

[CCGG98] M. Charikar, C. Chekuri, A. Goel, and S. Guha. Rounding via trees: deterministic approximation algorithms for group Steiner trees and k-median. In *30th Annual ACM Symposium on Theory of Computing*, pages 114–123. ACM, New York, 1998.

[CEK02] C. Chekuri, G. Even, and G. Kortsarz. An approximation algorithm for the group Steiner problem. Manuscript, 2002. (Preliminary version: G. Even and G. Kortsarz, An approximation algorithm for the group Steiner problem, *ACM-SIAM Symposium on Discrete Algorithms*, pages 49–58, 2002.)

[DV01] J. Dunagan and S. Vempala. On Euclidean embeddings and bandwidth minimization. In *Randomization, approximation, and combinatorial optimization*, pages 229–240. Springer, 2001.

[Fei98] U. Feige. A threshold of $\ln n$ for approximating set cover. *J. ACM*, 45(4):634–652, 1998.

[GKR00] N. Garg, G. Konjevod, and R. Ravi. A polylogarithmic approximation algorithm for the group Steiner tree problem. *J. Algorithms*, 37(1):66–84, 2000.

[KRS01] G. Konjevod, R. Ravi, and F. S. Salman. On approximating planar metrics by tree metrics. *Inform. Process. Lett.*, 80(4):213–219, 2001.

[MR95] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

[Sri01] A. Srinivasan. New approaches to covering and packing problems. In *13th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 567–576, 2001.

[ZK02] L. Zosin and S. Khuller. On directed Steiner trees. In *13th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 59–63, 2002.