Algorithmica (2011) 60: 110–151
DOI 10.1007/s00453-009-9344-x

Deterministic Sampling Algorithms for Network Design

Anke van Zuylen

Received: 17 November 2008 / Accepted: 13 July 2009 / Published online: 25 July 2009

Abstract For several NP-hard network design problems, the best known approximation algorithms are remarkably simple randomized algorithms called Sample-Augment algorithms in Gupta et al. (J. ACM 54(3):11, 2007). The algorithms draw a random sample from the input, solve a certain subproblem on the random sample, and augment the solution for the subproblem to a solution for the original problem. We give a general framework that allows us to derandomize most Sample-Augment algorithms, i.e. to specify a specific sample for which the cost of the solution created by the Sample-Augment algorithm is at most a constant factor away from optimal. Our approach allows us to give deterministic versions of the Sample-Augment algorithms for the connected facility location problem, in which the open facilities need to be connected by either a tree or a tour, the virtual private network design problem, 2-stage rooted stochastic Steiner tree problem with independent decisions, the a priori traveling salesman problem and the single sink buy-at-bulk problem. This partially answers an open question posed in Gupta et al. (J. ACM 54(3):11, 2007).

Keywords Approximation algorithms · Derandomization · Network design

1 Introduction

For several NP-hard network design problems, the best known approximation algorithms are remarkably simple randomized algorithms. The algorithms draw a random

A preliminary version of this paper [28] appeared in the Proceedings of the 16th European Symposium on Algorithms, 2008.

This research was conducted while the author was at Cornell University and was supported in part by NSF grant CCF-0514628, the National Natural Science Foundation of China Grant 60553001, and the National Basic Research Program of China Grant 2007CB807900,2007CB807901.

A. van Zuylen (✉)
Institute for Theoretical Computer Science, Tsinghua University, Beijing 100084, P.R. China
e-mail: anke@tsinghua.edu.cn

sample from the input, solve a certain subproblem on the random sample, and augment the solution for the subproblem to a solution for the original problem. Following [18], we will refer to this type of algorithm as a Sample-Augment algorithm. We give a general framework that allows us to derandomize most Sample-Augment algorithms, i.e. to specify a specific sample for which the cost of the solution created by the Sample-Augment algorithm is at most a constant factor away from optimal. The derandomization of the Sample-Augment algorithm for the single source rent-or-buy problem in Williamson and Van Zuylen [29] is a special case of our approach, but our approach also extends to the Sample-Augment algorithms for the connected facility location problem, in which the open facilities need to be connected by either a tree or a tour [5], the virtual private network design problem [3, 4, 15, 18], 2-stage stochastic Steiner tree problem with independent decisions [16], the a priori traveling salesman problem [24], and even the single sink buy-at-bulk problem [13, 15, 18], although for this we need to further extend our framework.

Generally speaking, the problems we consider are network design problems: they feature an underlying undirected graph $G = (V, E)$ with edge costs $c_e \geq 0$ that satisfy the triangle inequality, and the algorithm needs to make decisions such as on which edges to install how much capacity or at which vertices to open facilities. The Sample-Augment algorithm proceeds by randomly marking a subset of the vertices, solving some subproblem that is defined on the set of marked vertices, and then augmenting the solution for the subproblem to a solution for the original problem. We defer definitions of the problems we consider to the relevant sections.

As an example, in the single source rent-or-buy problem, we are given a source $s \in V$, a set of sinks $t_1, \ldots, t_k \in V$ and a parameter $M > 1$. An edge e can either be *rented* for sink t_j in which case we pay c_e, or it can be bought and used by any sink, in which case we pay Mc_e. The goal is to find a minimum cost set of edges to buy and rent so that for each sink t_j the bought edges plus the edges rented for t_j contain a path from t_j to s. In the Sampling Step of the Sample-Augment algorithm in Gupta et al. [15, 18] we mark each sink independently with probability $\frac{1}{M}$. Given the set of marked sinks D, the Subproblem Step finds a Steiner tree on $D \cup \{s\}$ and buys the edges of this tree. In the Augmentation Step, the subproblem's solution is augmented to a feasible solution for the single source rent-or-buy problem by renting edges for each unmarked sink t_j to the closest vertex in $D \cup \{s\}$.

To give a deterministic version of the Sample-Augment algorithm, we want to find a set D such that for this set D the cost of the Subproblem Step plus the Augmentation Step is at most the expected cost of the Sample-Augment problem. A natural approach is to try and use the method of conditional expectation [6] to achieve this. However, in order to do this we would need to be able to compute the conditional expectation of the cost of the Sample-Augment problem, conditioned on including/not including $t_j \in D$. Unfortunately, we do not know how to do this for any of the problems for which good Sample-Augment algorithms exist.

We will see however that we can get around this problem by using a good upper bound to provide an estimate of the conditional expectations required. We give more details behind our approach in Sect. 1.2, but first discuss some related work.

1.1 Related Work

Sample-Augment algorithms were first introduced by Gupta, Kumar and Roughgarden [15]. They use the framework to give new approximation algorithms for the single source rent-or-buy, virtual private network design and single sink buy-at-bulk problems. The main principle behind the analysis of the Sample-Augment algorithms is that under the right sampling strategy (i) it is not too difficult to bound the expected subproblem cost in terms of the optimal cost, and (ii) the expected augmentation cost is bounded by the expected subproblem cost.

Gupta, Kumar, Pál and Roughgarden [18] extend this framework, and show how to obtain an improved constant factor approximation algorithm for the multicommodity rent-or-buy problem. The key new ingredient is the notion of cost shares. If D is the set of marked vertices in the Sample-Augment algorithm, then a cost sharing method gives a way of allocating the cost of the subproblem's solution on D to the vertices in D. By imposing a "strictness" requirement on the cost sharing method, they ensure that the expected cost incurred for vertex j in the augmentation step is approximately equal to j's expected cost share. It is again not difficult to bound the expected cost of the subproblem in terms of the optimal cost, and hence the strictness of the cost shares implies that we can also bound the expected augmentation cost.

The ideas of strict cost shares and sampling algorithms have since then been successfully generalized and applied to give approximation algorithms for certain *stochastic* optimization problems. The Boosted Sampling algorithm for two-stage stochastic optimization problems was introduced by Gupta, Pál, Ravi and Sinha [16], and it was extended to multi-stage stochastic optimization problems by the same authors in [17].

As an example, consider the two-stage rooted stochastic Steiner tree problem, of which we will consider a special case in Sect. 3.2. Given a graph $G=(V,E)$ with edge costs $c_e \geq 0$, we are given a root s and terminals $t_1,\dots,t_k$ and a parameter $\sigma > 1$. A solution can be constructed in two stages. In the first stage we do not know which terminals need to be connected to the root, and we can buy edges at cost c_e. In the second stage, we do know which terminals need to connect to the root (we will call these *active*) and we can buy edges at cost σc_e. We assume the probability distribution from which the set of active terminals is drawn is known, either explicitly or as a black box from which we can sample. Examples of explicit probability distributions that have been considered in the literature are the case when there is a polynomial number of possible scenarios or the case when terminals are active independently with known probabilities. The Boosted Sampling algorithm is very similar to the Sample-Augment algorithms: we draw a random sample from the terminals, we buy a Steiner tree on these vertices in the first stage, and then we augment the solution in the second stage to connect the active terminals. However the sampling distribution according to which we sample terminals is now determined by the given probability distribution on the terminals.

In summary, the simple ideas underlying the Sample-Augment algorithms and Boosted Sampling algorithms have given rise to the best approximation algorithms for a great variety of problems. We refer the reader to the relevant sections below for references for the best known sampling algorithms for the problems we consider.

The Sample-Augment algorithm for the single source rent-or-buy problem, the connected facility location problem where the open facilities need to be connected by a tree, and the a priori traveling salesman problem with independent decisions have been derandomized prior to this work: Gupta, Srinivasan and Tardos [19] derandomize the Sample-Augment algorithm for single source rent-or-buy using the following idea. Rather than sampling the sinks independently at random, the sinks are sampled with limited dependence. Gupta et al. show that under this sampling strategy, the Sample-Augment algorithm is a 4.2-approximation algorithm. Then, since this sampling strategy has a small sample space, the algorithm can be derandomized by considering all points in the sample space. Williamson and Van Zuylen [29] give an alternative derandomization of the Sample-Augment algorithm for single source rent-or-buy which, in combination with the improved analysis of Eisenbrand, Grandoni, Rothvoßand Schäfer [5], results in a deterministic 3.28-approximation algorithm. Their approach is also used by Eisenbrand et al. [5] to derandomize the Sample-Augment algorithm for connected facility location where the open facilities need to be connected by a tree and by Shmoys and Talwar [24] for the a priori traveling salesman problem with independent decisions. The approach proposed by Williamson and Van Zuylen [29] is in fact a special case of the derandomization method we describe here.

For some of the problems we consider there exist deterministic algorithms that are not based on derandomizations of Sample-Augment algorithms. Swamy and Kumar [26] give a primal-dual 8.55-approximation algorithm for the connected facility location problem. Their analysis was recently refined to give a slightly better approximation guarantee of 6.55 [20]. Talwar [27] gives a constructive proof that a linear programming relaxation of the single sink buy-at-bulk problem introduced by Garg, Khandekar, Konjevod, Ravi, Salman and Sinha [9] has an integrality gap of at most 216.

Finally, Goyal, Gupta, Leonardi and Ravi [12] recently proposed a primal-dual 8-approximation algorithm for the rooted stochastic Steiner tree problem with a polynomial number of scenarios. However, in Sect. 3.2 we consider the version of the problem with independent decisions, for which no deterministic constant factor approximation algorithm was known.

1.2 Our Results

We give deterministic versions of the Sample-Augment algorithms: in particular, we show how to find a subset of the vertices D such that for this set D the cost of the Subproblem Step plus the Augmentation Step is at most the expected cost of the Sample-Augment problem.

Our approach is based on the method of conditional expectations [6]. We iterate through the vertices and decide whether or not to include the vertex in D depending on which choice gives a lower expected cost. Since we do not know how to compute the conditional expectation of the cost of the Sample-Augment problem, conditioned on including/not including the vertex in D, we need to use an estimate of these conditional expectations. What we show is that we can find an upper bound on the cost of the Subproblem Step plus Augmentation Step that can be efficiently computed. In addition, we show that the expectation of the upper bound under the sampling strategy of the randomized Sample-Augment algorithm is at most βOPT, where OPT is the optimal value and $\beta > 1$ is some constant. Then we can use this upper bound and the method of conditional expectation to find a set D such that the upper bound on the cost of the Subproblem Step plus the Augmentation Step is not more than the expected upper bound for the randomized Sample-Augment algorithm, and hence at most βOPT as well.

Our upper bound on the cost of the Subproblem Step will be obtained from a particular feasible solution to a linear programming (LP) relaxation of the subproblem. We then use well-known approximation algorithms to obtain a solution to the subproblem that comes within a constant factor of the subproblem LP. We do not need to solve the LP relaxation of the subproblem: instead we show that the optimal solution to an LP relaxation *of the original problem* defines a set of feasible solutions to the *subproblem*'s LP relaxation. We note that for some of the problems we consider, for example the virtual private network design problem, this requires us to "discover" a new LP relaxation of the original problem.

Using this technique, we derive the best known deterministic approximation algorithms for the single source rent-or-buy problem, 2-stage rooted stochastic Steiner tree problem with independent decisions, the a priori traveling salesman problem with independent decisions, the connected facility location problem in which the open facilities need to be connected by a Steiner tree or traveling salesman tour, the virtual private network design problem and the single sink buy-at-bulk problem. We thus partially answer an open question in Gupta et al. [18] (the only problem in [18] that we do not give a deterministic algorithm for is the multicommodity rent-or-buy problem). In addition, our analysis implies that the integrality gap of an (even) more natural LP relaxation than the one considered in [9, 27] for the single sink buy-at-bulk problem has integrality gap at most 27.72. We summarize our results in Table 1. The table uses the following abbreviations: SSRoB (single source rent-or-buy problem), 2-stage Steiner (2-stage rooted stochastic Steiner tree problem with independent decisions), a priori TSP (a priori traveling salesman problem with independent decisions), CFL-tree (connected facility location problem in which open facilities need to be connected by a tree), CFL-tour (connected facility location problem in which open facilities need to be connected by a tour), k-CFL-tree (connected facility location problem in which at most k facilities can be opened and the facilities need to be connected by a tree), CPND (virtual private network design problem) and SSBaB (single sink buy-at-bulk problem). The first column contains the best known approximation guarantees for the problems, which are obtained by randomized Sample-Augment algorithms. The second column gives the previous best known approximation guarantee by a deterministic algorithm. Entries marked with * were obtained based on the work of Williamson and Van Zuylen [29] that describes a special case of the approach in this paper. The third column shows the approximation guarantees in this paper.

We remark that our method is related to the method of pessimistic estimators of Raghavan [23]: Raghavan also uses an efficiently computable upper bound in combination with the method of conditional expectation to derandomize a randomized algorithm, where he first proves that the expected "cost" of the randomized algorithm is small. (We note that in the problem he considers, the cost of the algorithm is either

Table 1 Summary of Best Known Approximation Guarantees

Problem	Randomized	Prev. best deterministic	Our result
SSRoB	2.92 [5]	4.2 [19], 3.28* [5, 29]	3.28
2-stage Steiner	3.55 [16]	$\log n$ [21]	8
A priori TSP	4 [24], O(1)[10]	8* [24]	6.5
CFL-tree	4 [5]	6.55 [20], 4.23* [5]	4.23
k-CFL-tree	6.85 [5]	6.98* [5]	6.98
CFL-tour	4.12 [5]	–	4.12
VPND	3.55 [4]	$\log n$ [7]	8.02
SSBaB	24.92 [13]	216 [27]	27.72

0 (the solution is "good") or 1 (the solution is "bad").) However, in Raghavan's work the probabilities in the randomized algorithm depend on a solution to a linear program, but the upper bounds are obtained by a Chernoff-type bound. In our work, the probabilities in the randomized algorithm are already known from previous works, but we demonstrate *upper bounds* on the conditional expectations that depend on linear programming relaxations.

In the next section, we will give a general description of a Sample-Augment algorithm, and give a set of conditions under which we can give a deterministic variant of a Sample-Augment algorithm. In Sect. 3.1 we illustrate our method using the single source rent-or-buy problem as an example. In Sects. 3.2, 3.3, 3.4, and 3.5 we show how to obtain deterministic versions of the Sample-Augment algorithms for the 2-stage rooted stochastic Steiner tree with independent decisions, the a priori traveling salesman problem, connected facility location problems and the virtual private network design problem. In Sect. 4 we show how to extend the ideas from Sect. 2 to give a deterministic algorithm for the single sink buy-at-bulk problem. We conclude with a brief discussion of some future directions in Sect. 5.

2 General Framework

We give a high-level description of a class of algorithms first introduced by Gupta et al. [15], which were called Sample-Augment algorithms in [18]. Given a (minimization) problem $\mathcal{P}$, a Sample-Augment problem is defined by

(i) a set of elements $\mathcal{D} = \{1, \dots, n\}$ and sampling probabilities $p = (p_1, \dots, p_n)$,
(ii) a subproblem $\mathcal{P}_{sub}(D)$ defined for any $D \subseteq \mathcal{D}$, and
(ii) an augmentation problem $\mathcal{P}_{aug}(D, \texttt{Sol}_{sub}(D))$ defined for any $D \subseteq \mathcal{D}$ and solution $\texttt{Sol}_{sub}(D)$ to $\mathcal{P}_{sub}(D)$.

The Sample-Augment algorithm samples from $\mathcal{D}$ independently according to the sampling probabilities p, solves the subproblem and augmentation problem for the random subset, and returns the union of the solutions given by the subproblem and augmentation problem. We give a general statement of the Sample-Augment algorithm in Fig. 1.

We remark that we will consider Sample-Augment algorithms in which the Augmentation Step only depends on D, and not on $\texttt{Sol}_{sub}(D)$.

$\mathcal{P}$-**Sample-Augment**($\mathcal{D}$, p, $\mathcal{P}_{sub}$, $\mathcal{P}_{aug}$)
1. (*Sampling Step*) Mark each element $j \in \mathcal{D}$ independently with probability p_j. Let D be the set of marked elements.
2. (*Subproblem Step*) Solve $\mathcal{P}_{sub}$ on D. Let $\texttt{Sol}_{sub}(D)$ be the solution found.
3. (*Augmentation Step*) Solve $\mathcal{P}_{aug}$ on D, $\texttt{Sol}_{sub}(D)$. Let $\texttt{Sol}_{aug}(D, \texttt{Sol}_{sub}(D))$ be the solution found.
4. Return $\texttt{Sol}_{sub}(D)$ and $\texttt{Sol}_{aug}(D, \texttt{Sol}_{sub}(D))$.

Fig. 1 Sample-Augment algorithm

In the following, we let OPT denote the optimal cost of the problem we are considering. Let $C_{sub}(D)$ be the cost of $\texttt{Sol}_{sub}(D)$, and let $C_{aug}(D)$ be the cost of $\texttt{Sol}_{aug}(D, \texttt{Sol}_{sub}(D))$. Let $C_{SA}(D) = C_{sub}(D) + C_{aug}(D)$. We will use blackboard bold characters to denote random sets. For a function $C(D)$, let $\mathbb{E}_p[C(\mathbb{D})]$ be the expectation of $C(\mathbb{D})$ if $\mathbb{D}$ is obtained by including each $j \in \mathcal{D}$ in $\mathbb{D}$ independently with probability p_j.

Note that, since the elements are included in $\mathbb{D}$ independently, the conditional expectation of $\mathbb{E}_p[C_{SA}(\mathbb{D})]$ given that j is included in $\mathbb{D}$ is $\mathbb{E}_{p,p_j\leftarrow 1}[C_{SA}(\mathbb{D})]$, and the conditional expectation, given that j is not included in $\mathbb{D}$ is $\mathbb{E}_{p,p_j\leftarrow 0}[C_{SA}(\mathbb{D})]$. By the method of conditional expectations [6], one of these conditional expectations has value at most $\mathbb{E}_p[C_{SA}(\mathbb{D})]$. Hence if we could compute the expectations for different vectors of sampling probabilities, we could iterate through the elements and transform p into a binary vector (corresponding to a deterministic set D) without increasing $\mathbb{E}_p[C_{SA}(\mathbb{D})]$.

Unfortunately, this is not very useful to us yet, since it is generally not the case that we can compute $\mathbb{E}_p[C_{SA}(\mathbb{D})]$. However, as we will show, for many problems and corresponding Sample-Augment algorithms, it is the case that $\mathbb{E}_p[C_{aug}(\mathbb{D})]$ can be efficiently computed for any vector of probabilities p, and does not depend on the solution $\texttt{Sol}_{sub}(\mathbb{D})$ for the subproblem, but only on the set $\mathbb{D}$. The expected cost of the subproblem's solution is more difficult to compute. What we therefore do instead is replace the cost of the subproblem by an upper bound on its cost: Suppose there exists a function $U_{sub} : 2^{\mathcal{D}} \to R$ such that $C_{sub}(D) \le U_{sub}(D)$ for any $D \subseteq \mathcal{D}$, and suppose we can efficiently compute $\mathbb{E}_p[U_{sub}(\mathbb{D})]$ and $\mathbb{E}_p[C_{aug}(\mathbb{D})]$ for any vector p. If there exists a known vector $\hat{p}$ such that

$$\mathbb{E}_{\hat{p}}[U_{sub}(\mathbb{D})] + \mathbb{E}_{\hat{p}}[C_{aug}(\mathbb{D})] \le \beta OPT, \tag{1}$$

then we can use the method of conditional expectation to find a set D such that $U_{sub}(D) + C_{aug}(D) \le \beta OPT$, and hence also $C_{sub}(D) + C_{aug}(D) \le \beta OPT$.

In particular, the upper bounds that we will consider will all be given by solutions to an LP relaxation of the subproblem.

Theorem 1 *Given a minimization problem $\mathcal{P}$ and an algorithm $\mathcal{P}$-Sample-Augment, suppose the following four conditions hold*:

(i) $\mathbb{E}_p[C_{aug}(\mathbb{D})]$ *depends only on* $\mathbb{D}$, *not on* $\texttt{Sol}_{sub}(\mathbb{D})$, *and can be efficiently computed for any* p.

(ii) *There exists an LP relaxation Sub-LP*(D) *of* $\mathcal{P}_{sub}(D)$ *and an algorithm for* $\mathcal{P}_{sub}(D)$ *that is guaranteed to output a solution to* $\mathcal{P}_{sub}(D)$ *that costs at most a factor* α *times the cost of any feasible solution to Sub-LP*(D).

(iii) *We can compute vectors* b *and* $r(j)$ *for* $j = 1, \ldots, n$ *such that* $y(D) = b + \sum_{j\in D} r(j)$ *is a feasible solution to Sub-LP*(D) *for any* $D \subseteq \mathcal{D}$.

(iv) *There exists a known vector* $\hat{p}$ *such that*

$$\mathbb{E}_{\hat{p}}[C_{aug}(\mathbb{D})] + \alpha\mathbb{E}_{\hat{p}}[C_{LP}(y(\mathbb{D}))] \leq \beta OPT,$$

where $C_{LP}(y(D))$ *is the objective value of* $y(D)$ *for Sub-LP*(D).

Then there exists a deterministic β*-approximation algorithm for* $\mathcal{P}$.

Proof Let $U_{sub}(D) = \alpha C_{LP}(y(D))$. If we use the algorithm from (ii) in the Subproblem Step of $\mathcal{P}$-Sample-Augment, then by (ii), $C_{sub}(D) \leq U_{sub}(D)$. By (iii) $\mathbb{E}_p[U_{sub}(\mathbb{D})]$ can be efficiently computed for any p, and by (iv) (1) is satisfied. Hence we can use the method of conditional expectation to find a set D such that $C_{sub}(D) + C_{aug}(D) \leq U_{sub}(D) + C_{aug}(D) \leq \beta OPT$. □

In many cases, (i) is easily verified. In the problems we are considering here, the subproblem looks for a Steiner tree or a traveling salesman tour. It was shown by Goemans and Bertsimas [11] that the cost of the minimum cost spanning tree is at most twice the optimal value of the Steiner tree LP relaxation, and hence the minimum cost spanning tree costs at most twice the objective value of any feasible solution to this LP. For the traveling salesman problem, it was shown by Wolsey [30], and independently by Shmoys and Williamson [25], that the Christofides algorithm [2] gives a solution that comes within a factor of 1.5 of the subtour elimination LP. The solution $y(D) = b + \sum_{j\in\mathcal{D}} r(j)$ will be defined by using the optimal solution to an LP relaxation *of the original problem*, so that for appropriately chosen probabilities $\mathbb{E}_{\hat{p}}[C_{LP}(y(\mathbb{D}))]$ is bounded by a constant factor times OPT. Using the analysis for the randomized algorithm to bound $\mathbb{E}_{\hat{p}}[C_{aug}(\mathbb{D})]$, we can then show that (iv) holds.

2.1 Conditioning on the Size of $\mathbb{D}$

In some cases, $\mathcal{P}_{sub}$ and $\mathcal{P}_{aug}$ are only defined for $|D| \geq k$ for some small $k > 0$. Different algorithms deal with this in different ways, but one possible approach to ensure that $|D| \geq k$ is to redo the Sampling Step of the randomized Sample-Augment algorithm until the set of marked elements has size at least k. We note that this does not necessarily give algorithms that run in polynomial time, but that it has been shown that such sampling strategies can be implemented efficiently (see for example [24]). To derandomize these algorithms, we will use the following modified version of Theorem 1.

Theorem 2 *Given a minimization problem* $\mathcal{P}$, *an algorithm* $\mathcal{P}$*-Sample-Augment which repeats the Sampling Step until it outputs* D *with* $|D| \geq k$ *for some constant* k, *suppose condition* (i) *of Theorem* 1 *holds conditioned on* $|\mathbb{D}| \geq k$, *conditions* (ii) *and* (iii) *of Theorem* 1 *hold for all* $|D| \geq k$ *and suppose we have a vector* q *such that*

$$\mathbb{E}_q[C_{aug}(\mathbb{D}) \mid |\mathbb{D}| \geq k] + \alpha \mathbb{E}_q[C_{LP}(y(\mathbb{D})) \mid |\mathbb{D}| \geq k] \leq \beta OPT,$$

then there exists a deterministic β-approximation algorithm for $\mathcal{P}$.

Proof We show that we can find in polynomial time a vector $\hat{p}$ with $|\{j : \hat{p}_j = 1\}| \geq k$ such that

$$\mathbb{E}_{\hat{p}}[C_{aug}(\mathbb{D})] + \alpha \mathbb{E}_{\hat{p}}[C_{LP}(y(\mathbb{D}))] \leq \beta OPT. \tag{2}$$

We can then use the method of conditional expectation as before, and we will be guaranteed that we only consider vectors p with $|\{j : p_j = 1\}| \geq k$, i.e. probability distributions over sets D with $|D| \geq k$.

For ease of notation, we let $C(D) = C_{aug}(D) + C_{LP}(y(D))$. Let $f(\mathbb{D})$ be the k elements in $\mathbb{D}$ with the smallest indices, and let $\mathcal{F}$ be the set of all subsets of $\mathcal{D}$ with exactly k elements. Then

$$\mathbb{E}_q[C(\mathbb{D}) \mid |\mathbb{D}| \geq k] = \sum_{F \in \mathbb{F}} \mathbb{E}_q[C(\mathbb{D}) \mid |\mathbb{D}| \geq k, f(\mathbb{D}) = F] \mathbb{P}[f(\mathbb{D}) = F].$$

Hence there exists some F such that $\mathbb{E}_q[C(\mathbb{D}) \mid |\mathbb{D}| \geq k, f(\mathbb{D}) = F] \leq \mathbb{E}_q[C(\mathbb{D}) \mid |\mathbb{D}| \geq k]$. Now, let $\hat{p}_j = 1$ if $j \in F$, $\hat{p}_j = 0$ if $j \notin F$ and there exists $i \in F$ with $i < j$ and $\hat{p}_j = q_j$ otherwise. Then $\mathbb{E}_q[C(\mathbb{D}) \mid |\mathbb{D}| \geq k, f(\mathbb{D}) = F] = \mathbb{E}_{\hat{p}}[C(\mathbb{D})]$ and $\hat{p}$ satisfies (2).

We can find the right set F by trying all sets in $\mathcal{F}$ and computing $\mathbb{E}_{\hat{p}}[C(\mathbb{D})]$ for the corresponding vector $\hat{p}$. By our assumptions, we can compute these expectations efficiently, and the vector $\hat{p}$ which gives the smallest expectation satisfies (2). □

3 Derandomization of Sample-Augment Algorithms

In this section, we show how Theorems 1 and 2 give the results in Table 1. We will use the following notation. Given an undirected graph $G = (V, E)$ with edge costs $c_e \geq 0$ for $e \in E$, we denote by $\ell(u, v)$ the length of the shortest path from $u \in V$ to $v \in V$ with respect to costs c. For $S \subseteq V$ we let $\ell(u, S) = \min_{v \in S} \ell(u, v)$. For $T \subseteq E$, we will use the short hand notation $c(T)$ for $\sum_{e \in T} c_e$ for $T \subseteq E$. Finally, for a subset $S \subset V$, we let $\delta(S) = \{\{i, j\} \in E : i \in S, j \in V \backslash S\}$.

3.1 Single Source Rent-or-Buy

We illustrate Theorem 1 by showing how it can be used to give a deterministic algorithm for the single source rent-or-buy problem. We note that this was already done in [29]; however, we repeat this here because this is arguably the simplest application of Theorem 1 and hence provides a nice illustration of the more general approach.

In the single source rent-or-buy problem, we are given an undirected graph $G = (V, E)$, edge costs $c_e \geq 0$ for $e \in E$, a source $s \in V$ and a set of sinks $t_1, \dots, t_k \in V$, and a parameter $M > 1$. A solution is a set of edges B to *buy*, and for each sink t_j a set of edges R_j to *rent*, so that $B \cup R_j$ contains a path from s to t_j. The cost of renting an edge e is c_e and the cost of buying e is Mc_e. We want to find a solution $(B, R_1, \dots, R_k)$ that minimizes $Mc(B) + \sum_{j=1}^{k} c(R_j)$.

SSRoB-Sample-Augment$(G=(V,E),c,s,\{t_1,\dots,t_k\},p)$
1. (*Sampling Step*) Mark each sink t_j with probability p_j. Let D be the set of marked sinks. 2. (*Subproblem Step*) Construct a Steiner tree on $D\cup\{s\}$ and *buy* the edges of the tree. 3. (*Augmentation Step*) Rent the shortest path from each unmarked sink to the closest terminal in $D\cup\{s\}$.

Fig. 2 Sample-Augment algorithm for single source rent-or-buy

Gupta et al. [15] propose the random sampling algorithm given in Fig. 2, where they set $p_j=\frac{1}{M}$ for all $j=1,\dots,k$.

Note that the expected cost of the Augmentation Step of SSRoB-Sample-Augment does not depend on the tree bought in the Subproblem Step. Gupta et al. [15] show that if each sink is marked independently with probability $\frac{1}{M}$ then the expected cost of the Augmentation Step can be bounded by $2OPT$.

Lemma 3 [15] *If* $p_j=\frac{1}{M}$ *for* $j=1,\dots,k$, *then* $\mathbb{E}[C_{aug}(\mathbb{D})]\le 2OPT$.

Theorem 4 [29] *There exists a deterministic* 4-*approximation algorithm for SSRoB.*

Proof We verify that the four conditions of Theorem 1 hold. We begin by showing that $\mathbb{E}_p[C_{aug}(\mathbb{D})]$, the expected cost incurred in the Augmentation Step, can be computed for any vector of sampling probabilities p. Fix a sink $t\in\{t_1,\dots,t_k\}$. We label the terminals in $\{s,t_1,\dots,t_k\}$ as $r_0,\dots,r_k$ such that $\ell(t,r_0)\le\ell(t,r_1)\le\cdots\le\ell(t,r_k)$. If we define $p_s=1$, then the expected cost incurred for t in the Augmentation Step is

$$\sum_{i=0}^{k}\ell(t,r_i)p_{r_i}\prod_{j<i}(1-p_{r_j}),$$

and $\mathbb{E}_p[C_{aug}(\mathbb{D})]$ is the sum over these values for each $t\in\{t_1,\dots,t_k\}$.

Now consider the subproblem on a given subset D of $\{t_1,\dots,t_k\}$. From Goemans and Bertsimas [11] we know that we can efficiently find a Steiner tree on $D\cup\{s\}$ of cost at most twice the optimal value (and hence the objective value of any feasible solution) of the following Sub-LP:

$$\begin{aligned}
&\min && \sum_{e\in E} Mc_e y_e\\
\text{(Sub-LP}(D))\quad &\text{s.t.} && \sum_{e\in\delta(S)} y_e\ge 1\quad \forall S\subset V: s\notin S, D\cap S\ne\emptyset,\\
& && y_e\ge 0\quad \forall e\in E.
\end{aligned}$$

We now want to define a feasible solution $y(D)$ to Sub-LP(D) for any $D\subset\mathcal{D}$, such that $y(D)$ can be written as $b+\sum_{t_j\in D}r(j)$, since this form will allow us to efficiently compute $\mathbb{E}_p[C_{LP}(y(\mathbb{D}))]$. To do this, we use an LP relaxation of the single source rent-or-buy problem. Let b_e be a variable that indicates whether we buy edge e, and let r_e^j indicate whether we rent edge e for sink t_j.

$$\text{(SSRoB-LP)}\quad \begin{aligned} \min \quad & \sum_{e\in E} Mc_e b_e + \sum_{e\in E}\sum_{j=1}^{k} c_e r_e^j \\ \text{s.t.} \quad & \sum_{e\in\delta(S)} (b_e + r_e^j) \geq 1 \quad \forall S\subset V : t_j \in S, s \notin S, \\ & b_e, r_e^j \geq 0 \quad \forall e\in E, j=1,\ldots,k. \end{aligned}$$

SSRoB-LP is a relaxation of the single source rent-or-buy problem, since the optimal solution to the single source rent-or-buy problem is feasible for SSRoB-LP and has objective value OPT. Let $\hat{b}, \hat{r}$ be an optimal solution to SSRoB-LP. For a given set $D \subseteq \mathcal{D}$ and edge $e \in E$ we let

$$y_e(D) = \hat{b}_e + \sum_{t_j\in D} \hat{r}_e^j.$$

Clearly, $y(D)$ is a feasible solution to Sub-LP(D) for any D.

Finally, we show that $2\mathbb{E}_{\hat{p}}[C_{LP}(y(\mathbb{D}))] + \mathbb{E}_{\hat{p}}[C_{aug}(\mathbb{D})] \leq 4OPT$ if we let $\hat{p}_j = \frac{1}{M}$ for every $t_j \in \mathcal{D}$: by Lemma 3, the expected cost of the Augmentation Step is at most $2OPT$, and $2\mathbb{E}_{\hat{p}}[C_{LP}(y(\mathbb{D}))]$ is

$$2\sum_{e\in E} Mc_e\left(\hat{b}_e + \sum_{j=1}^{k} \frac{1}{M}\hat{r}_e^j\right) \leq 2OPT.$$

Hence, applying Theorem 1, we get that there exists a 4-approximation algorithm for the single sink rent-or-buy problem. □

As was shown by [5, 29], a better deterministic approximation algorithm can be obtained by using the improved analysis of the randomized algorithm given by Eisenbrand, Grandoni, Rothvoß and Schäfer [5], which allows us to more carefully balance the charge against the optimal renting and the optimal buying costs. For a given optimal solution, let B^* be the buying cost and R^* the renting cost. We need the following lemma from Eisenbrand et al. [5].

Lemma 5 [5] *If $p_j = \frac{a}{M}$ for $j = 1,\ldots,k$ then $\mathbb{E}_p[C_{aug}(\mathbb{D})] \leq \frac{0.807}{a} B^* + 2R^*$.*

Note that if we mark each t_j with probability $\frac{a}{M}$, then $\mathbb{E}_p[C_{LP}(y(\mathbb{D}))] = \sum_{e\in E} Mc_e\hat{b}_e + a\sum_{e\in E}\sum_{j=1}^{k} c_e\hat{r}_e^j$. We would like to claim that this is at most $B^* + aR^*$, but this is not necessarily the case. However, it is true if we replace the objective of SSROB-LP by

$$\min \sum_{e\in E} Mc_e b_e + a\sum_{e\in E}\sum_{j=1}^{k} c_e r_e^j.$$

Hence if we use the optimal solution to SSROB-LP with the modified objective to define $y(D)$, then for $\hat{p} = \frac{a}{M}$, we get that

$$\mathbb{E}_{\hat{p}}[C_{aug}(\mathbb{D})] + 2\mathbb{E}_{\hat{p}}[C_{LP}(y(\mathbb{D}))] \le \frac{0.807}{a} B^* + 2R^* + 2B^* + 2aR^*$$
$$= \left(\frac{0.807}{a} + 2\right) B^* + (2 + 2a) R^*.$$

Choosing $a = 0.636$, we get the following result.

Theorem 6 [5, 29] *There exists a deterministic* 3.28*-approximation algorithm for Single Source Rent-or-Buy.*

3.2 2-Stage Stochastic Steiner Tree with Independent Decisions

The input of the 2-stage rooted stochastic Steiner tree problem with independent decisions consists of a graph $G = (V, E)$ with edge costs $c_e \ge 0$, a root s and terminals $t_1, \dots, t_k$ with activation probabilities $q_1, \dots, q_k$ and a parameter $\sigma > 1$. A solution can be constructed in two stages. In the first stage we do not know which terminals need to be connected to the root, and we can install edges at cost c_e. In the second stage, we do know which terminals need to connect to the root (we will call these *active*) and we can install edges at cost σc_e. Each terminal t_j is active independently with probability q_j.

The Boosted Sampling algorithm proposed in [16] is very similar to the SSRoB-Sample-Augment algorithm. We first sample from the terminals, where terminal t_j is chosen independently with probability $\min\{1, \sigma q_j\}$. Let D be the set of terminals selected. The first stage solution is a Steiner tree on $D \cup \{s\}$. In the second stage, we augment the first stage solution by adding shortest paths from each active terminal to the closest terminal in $D \cup \{s\}$. We are interested in the expected cost of the algorithm's solution, and hence we can replace the Augmentation Step by adding shortest path from each terminal t_j to the closest terminal in $D \cup \{s\}$ with edge costs $\sigma q_j c_e$ as this gives the same expected cost. Hence the Boosted Sampling algorithm for 2-stage rooted stochastic Steiner tree problem with independent decisions is the same as the SSRoB-Sample-Augment algorithm with $M = 1$, except that in the Augmentation Step, the renting cost for renting edge e for terminal j is $\sigma q_j c_e$.

We begin by repeating bounds on the first stage and second stage costs of this algorithm that follow from Theorem 6.2 in [16] and the Prim cost shares in Example 2.8 of [18].

Lemma 7 [16, 18] *If* $p_j = \min\{1, \sigma q_j\}$ *for* $j = 1, \dots, k$ *and if we were able to find a minimum cost solution to the subproblem, then* $\mathbb{E}_p[C_{sub}(\mathbb{D})] \le OPT$, *and* $\mathbb{E}_p[C_{aug}(\mathbb{D})] \le 2OPT$.

We derandomize this algorithm using Theorem 1. It is clear that condition (i) of Theorem 1 is again met. For condition (ii) we can use the same Sub-LP as in the previous section (with $M = 1$), and we again have $\alpha = 2$. Now, we need a good LP relaxation to define the solutions $y(D)$ to the Sub-LP. We claim that the optimal value of the following LP is at most OPT:

$$\begin{aligned} &\min \quad \frac{1}{3}\sum_{e\in E}\left(c_e b_e + \sum_{j=1}^{k}\sigma q_j c_e r_e^j\right) \\ &\text{(2-stage-LP)} \quad \text{s.t.} \quad \sum_{e\in\delta(S)} (b_e + r_e^j) \geq 1 \quad \forall S\subset V : s\notin S, t_j\in S, \\ &\qquad b_e, r_e^j \geq 0 \quad \forall e\in E, j=1,\dots,k. \end{aligned}$$

To see that this is indeed a relaxation of the problem, suppose we could find the optimal Steiner tree on $D\cup\{s\}$ in the Subproblem Step of the Boosted Sampling algorithm. Then it follows from Lemma 7 that the expected cost of the solution constructed by the Boosted Sampling algorithm is at most $3OPT$. Hence there exists some sample D such that the cost of the optimal Steiner tree on $D\cup\{s\}$ plus the cost of the Augmentation Step is at most $3OPT$. Letting $b_e = 1$ for the first stage edges in this solution, and $r_e^j = 1$ for the second stage edges, thus gives a solution to 2-stage-LP of cost at most OPT.

Given an optimal solution $\hat{b}, \hat{r}$ to 2-stage-LP, we define $y_e(D) = \hat{b}_e + \sum_{t_j\in D}\hat{r}_e^j$ as before, and taking $\hat{p}_j = \min\{1, q_j\sigma\}$, we find that

$$2\mathbb{E}_{\hat{p}}[C_{LP}(y(\mathbb{D}))] \leq 2\sum_{e\in E}\left(c_e\hat{b}_e + \sum_{j=1}^{k}\sigma q_j c_e \hat{r}_e^j\right) \leq 6OPT.$$

Combining this with the bound on the second stage cost from Lemma 7, Theorem 1 allows us to get the following result.

Theorem 8 *There exists a deterministic* 8*-approximation algorithm for the* 2*-stage rooted stochastic Steiner tree problem with independent decisions.*

3.3 A Priori Traveling Salesman with Independent Decisions

In the a priori traveling salesman problem with independent decisions, we are given a graph $G = (V, E)$ with edge costs $c_e \geq 0$ and a set of terminals $t_1, \dots, t_k$, where terminal t_j is active independently of the other terminals with probability q_j. The goal is to find a so-called *master tour* on the set of all terminals, such that the expected cost of shortcutting the master tour to the set of active terminals is minimized.

Shmoys and Talwar [24] recently showed that a Sample-Augment type algorithm for this problem is a 4-approximation algorithm. In the Sampling Step, they randomly mark the terminals, where each terminal t_j is marked independently with probability $p_j = q_j$. If fewer than 2 terminals are marked, we redo the marking step, until we have a set of marked terminals of size at least 2. We note that Shmoys and Talwar [24] show how to implement this sampling strategy in polynomial time; however, since we will just be concerned with derandomizing the algorithm, we omit the details of this here. In the Subproblem Step they find a tour on the marked terminals and finally, in

APTSP-Sample-Augment$(G = (V, E), c, Q, \tilde{q}, s, \{t_1, \dots, t_k\}, p)$
1. (*Sampling Step*) Mark each terminal t_j with probability p_j. Let D be the set of marked terminals. If $\|D\| < 2$ then remove all markings and repeat the Sampling Step.
2. (*Subproblem Step*) Construct a traveling salesman tour on D, and incur cost Qc_e for each edge on the tour.
3. (*Augmentation Step*) Add two copies of the shortest path from each unmarked terminal t_j to the closest terminal in D and incur cost $\tilde{q}_j c_e$ for each edge.

Fig. 3 Sample-Augment algorithm for the a priori traveling salesman problem

the Augmentation Step they add two copies of the shortest path from each unmarked terminal to the closest marked terminal.

It is not hard to see that the Sample-Augment algorithm finds an Euler tour on the terminals, and we can shortcut the Euler tour to give the traveling salesman tour that will be the master tour.

To evaluate the expected cost of the shortcut tour on a set of active terminals A, Shmoys and Talwar upper bound the cost of shortcutting the master tour on A by assuming that for any A of size at least 2 we *always* traverse the edges found in the Subproblem Step, and we traverse the edges found in the Augmentation Step only for the active terminals. If $|A| < 2$, then the cost of the shortcut master tour is 0.

Since we are interested in upper bounding the expected cost of the shortcut tour, we can just consider the expectation of this upper bound. Let Q be the probability that at least 2 terminals are active, and let $\tilde{q}_j$ be the probability that t_j is active conditioned on the fact that at least 2 terminals are active, i.e. $\frac{q_j(1-\prod_{i \neq j}(1-q_i))}{Q}$. The expected cost for an edge e in the tour constructed by the Subproblem Step is Qc_e and the expected cost for an edge e that is added for terminal j in the Augmentation Step is $\tilde{q}_j c_e$. Hence we can instead analyze the algorithm APTSP-Sample-Augment given in Fig. 3.

We will use the following bounds on the expected cost of the algorithm that follow from Shmoys and Talwar [24].

Lemma 9 [24] *If $p_j = q_j$ for every terminal, and if we were able to find a minimum cost solution to the subproblem, then $\mathbb{E}_q[C_{sub}(\mathbb{D}) \mid |\mathbb{D}| \geq 2] \leq OPT$, and $\mathbb{E}_q[C_{aug}(\mathbb{D}) \mid |\mathbb{D}| \geq 2] \leq 2OPT$.*

We note that the bound on $\mathbb{E}_q[C_{sub}(\mathbb{D}) \mid |\mathbb{D}| \geq 2]$ in Lemma 9 does not occur in this form in Shmoys and Talwar [24]: they show that $\mathbb{E}_q[2MST(\mathbb{D}) \mid |\mathbb{D}| \geq 2] \leq 2OPT$ but it is straightforward to adapt their analysis to show that the expected cost of the optimal TSP tour on $\mathbb{D}$, conditioned on $|\mathbb{D}| \geq 2$, is at most OPT.

Lemma 9 implies that there is some non-empty set D^* such that $C_{sub}(D^*) + C_{aug}(D^*) \leq 3OPT$. Let t^* be one of the terminals in D^*, and set $b_e = 1$ for each of the edges in the (minimum cost) subproblem's solution on D^*, and let $r_e^j = 1$ for the edges added for terminal j in the Augmentation Step. Then b, r defines a feasible solution to the following LP with objective value at most OPT and hence APTSP-LP is an LP relaxation of the a priori Traveling Salesman Problem:

$$\text{(APTSP-LP)}\quad \begin{aligned} \min\quad & \frac{1}{3}\sum_{e\in E}\left(Qc_eb_e+\sum_{j=1}^{k}\tilde{q}_jc_er_e^j\right)\\ \text{s.t.}\quad & \sum_{e\in\delta(S)}(b_e+r_e^j)\geq 2\quad \forall S\subset V: t^*\notin S, t_j\in S,\\ & b_e, r_e^j\geq 0\quad \forall e\in E, j=1,\dots,k. \end{aligned}$$

Note that we do not know t^*, but we can solve APTSP-LP for any $t^*\in\{t_1,\dots,t_k\}$ and use the LP with the smallest objective value. Let $\hat{b},\hat{r}$ be an optimal solution to that LP.

We let the Sub-LP on D be

$$\text{(Sub-LP}(D))\quad \begin{aligned} \min\quad & \sum_{e\in E}Qc_ey_e\\ \text{s.t.}\quad & \sum_{e\in\delta(S)}y_e\geq 2\quad \forall S\subset V: D\backslash S\neq\emptyset, D\cap S\neq\emptyset,\\ & y_e\geq 0\quad \forall e\in E. \end{aligned}$$

Note that this satisfies condition (ii) in Theorem 2 with $\alpha=1.5$ by [25, 30]. To define solutions $y(D)$ to Sub-LP(D), we let $y_e(D)=\hat{b}_e+\sum_{t_j\in D}\hat{r}_e^j$.

We now consider the expectation of $\mathbb{E}_q[C_{LP}(y(\mathbb{D}))\mid|\mathbb{D}|\geq 2]$ and $\mathbb{E}_q[C_{aug}(\mathbb{D})\mid|\mathbb{D}|\geq 2]$. From Lemma 9 we know that the second term is at most $2OPT$. Also, since the probability that t_j is in $\mathbb{D}$ conditioned on $\mathbb{D}$ having at least 2 elements is $\tilde{q}_j$, we get

$$\begin{aligned} 1.5\mathbb{E}_q[C_{LP}(y(\mathbb{D}))\mid|\mathbb{D}|\geq 2] &= 1.5\left(\sum_{e\in E}Qc_e\hat{b}_e+\sum_{j=1}^{k}Q\tilde{q}_jc_e\hat{r}_e^j\right)\\ &= 1.5\sum_{e\in E}\left(Qc_e\hat{b}_e+\sum_{j=1}^{k}q_j\left(1-\prod_{i\neq j}(1-q_i)\right)c_e\hat{r}_e^j\right)\\ &\leq 1.5\sum_{e\in E}\left(Qc_e\hat{b}_e+\sum_{j=1}^{k}q_jc_e\hat{r}_e^j\right)\leq 4.5OPT, \end{aligned}\tag{3}$$

where the last inequality holds since we showed that APTSP-LP is a relaxation of the a priori Traveling Salesman Problem. Hence we find that $1.5\mathbb{E}_q[C_{LP}(y(\mathbb{D}))\mid|\mathbb{D}|\geq 2]+\mathbb{E}_q[C_{aug}(\mathbb{D})\mid|\mathbb{D}|\geq 2]\leq 6.5OPT$ Hence the conditions of Theorem 2 hold with $\beta=6.5$ and we get the following result.

Theorem 10 *There exists a deterministic* 6.5*-approximation algorithm for a priori Traveling Salesman Problem.*

Remark 11 The deterministic 8-approximation algorithm obtained by Shmoys and Talwar [24] uses similar techniques but uses the Steiner tree LP as the Sub-LP. Since we can get a traveling salesman tour of cost at most twice the cost of a Steiner tree, $\alpha = 4$. They show that for the Steiner Sub-LP $\mathbb{E}_q[C_{LP}(y(\mathbb{D})) \mid |\mathbb{D}| \geq 2] \leq 1.5OPT$. Hence $\alpha\mathbb{E}_q[C_{LP}(y(\mathbb{D})) \mid |\mathbb{D}| \geq 2] \leq 6OPT$ instead of what we find in (3).

3.4 Connected Facility Location Problems

The connected facility location problems that we consider have the following form. We are given an undirected graph $G = (V, E)$ with edge costs $c_e \geq 0$ for $e \in E$, a set of clients $\mathcal{D} \subseteq V$ with demands d_j for $j \in \mathcal{D}$, a set of potential facilities $\mathcal{F} \subseteq V$, with opening cost $f_i \geq 0$ for $i \in \mathcal{F}$, a connectivity requirement $CR \in \{\texttt{Tour}, \texttt{SteinerTree}\}$, a parameter $M > 1$, and a parameter $k > 1$. We assume that the edge costs satisfy the triangle inequality. The goal is to find a subset of facilities $F \subseteq \mathcal{F}$ to open and a set of edges T such that $|F| \leq k$ (k may be ∞) and T is a CR on F that minimizes

$$\sum_{i \in F} f_i + Mc(T) + \sum_{j \in \mathcal{D}} \ell(j, F).$$

We will say that we *buy* the edges of the set T that connect the open facilities, and that we *rent* the edges connecting each client to its closest open facility.

For ease of exposition we assume that $d_j = 1$ for all $j \in \mathcal{D}$. It is not hard to adapt the analysis to the general case, as was shown in [15]. We will make a remark about this at the end of this section. In the following, we denote by $\rho_{cr} = 1$ if $CR = \texttt{SteinerTree}$ and $\rho_{cr} = 2$ if $CR = \texttt{Tour}$, which basically indicates the requirement that any two open facilities need to be connected by ρ_{cr} edge-disjoint paths.

To determine which facilities to open, the Sample-Augment algorithm of Eisenbrand et al. [5] first uses an approximation algorithm to determine a good solution to the facility location problem in which we drop the requirement that the facilities need to be connected. They then mark each *client* $j \in \mathcal{D}$ independently with probability p_j and open the facilities that the marked clients are assigned to in the solution to the unconnected facility location problem. Of course, any feasible solution must have at least 1 open facility, hence we need to mark at least one client. To achieve this, Eisenbrand et al. first mark one client chosen uniformly at random.

To connect the open facilities by bought edges, the algorithm buys a *CR on the marked clients*, and extends this to a *CR* on the open facilities by adding ρ_{cr} copies of the shortest path from each facility to the marked client that caused it to be opened. Finally, we need to rent edges to connect the other clients to their closest open facility.

Let j^* be the client marked by choosing one client uniformly at random. To make the algorithm fit into our framework, we let j^* be part of the input. In addition, we reorder the steps, so that the Subproblem Step only finds the *CR* on the marked clients, and the Augmentation Step contains all the other steps of the algorithm. We give our variant of the Sample-Augment algorithm from Eisenbrand et al. [5] in Fig. 4.

To show that we can derandomize the CFL-Sample-Augment algorithm, we first fix the input variable j^* to be an arbitrary client and we will show that conditions

CFL-Sample-Augment$(G=(V,E), c, \mathcal{D}, \mathcal{F}, f, k, CR, p, j^*)$
1. (*Sampling Step*) Mark every client j in $\mathcal{D}$ independently at random with probability p_j. Let D be the set of marked clients. 2. (*Subproblem Step*) Construct a *CR* solution on the set $D \cup \{j^*\}$. Buy the edges of this solution. 3. (*Augmentation Step*) Compute an (approximately optimal) solution to the corresponding unconnected k-facility location problem. Let F_U be the facilities opened, and for $j \in \mathcal{D}$ let $\sigma_U(j)$ be the facility j is assigned to. Let $F = \bigcup_{j \in D \cup \{j^*\}} \sigma_U(j)$, and open the facilities in F. Rent the edges from each client $j \in \mathcal{D}$ to their closest open facility, and, in addition to the edges bought in Step 2, buy ρ_{cr} copies of the edges on the shortest path from each client j in $D \cup \{j^*\}$ to its closest facility in F.

Fig. 4 Sample-Augment algorithm for connected facility location

(i), (ii) and (iii) of Theorem 1 are satisfied. We then show that we can efficiently find a choice for j^* so that condition (iv) for the required approximation factor is satisfied.

It is not hard to verify that condition (i) of Theorem 1 is satisfied for any sampling probabilities p: in the Augmentation Step the set of facilities we open depends only on the set $D \cup \{j^*\}$, and hence the cost of renting edges between each client and its closest open facility, and the cost of buying edges between the clients in $D \cup \{j^*\}$ and their closest open facility all do not depend on the Steiner tree on $D \cup \{j^*\}$.

We define Sub-LP(D) as

$$\begin{aligned} &\min \quad \sum_{e \in E} M c_e y_e \\ &(\text{Sub-LP}(D)) \quad \text{s.t.} \quad \sum_{e \in \delta(S)} y_e \geq \rho_{cr} \quad \forall S \subset V : D \cup \{j^*\} \not\subseteq S, (D \cup \{j^*\}) \cap S \neq \emptyset, \\ &\qquad y_e \geq 0 \quad \forall e \in E. \end{aligned}$$

Condition (ii) of Theorem 1 is satisfied with $\alpha = 2$ if $CR = \texttt{SteinerTree}$ [11], or 1.5 if $CR = \texttt{Tour}$ [25, 30].

Let $\gamma = \frac{M}{|\mathcal{D}|}$, and let a be a parameter to be determined later. We assume we know some facility i^* that is open in the optimal solution. (We can drop this assumption by taking i^* to be the facility for which the following LP gives the lowest optimal value.) We use the following LP to define the Sub-LP solutions. We note that this is almost an LP relaxation of the connected facility location problem, except for the weighting of the renting cost by $(a+\gamma)\rho_{cr}$

$$\min \quad \sum_{e \in E} M c_e b_e + (a+\gamma)\rho_{cr} \sum_{j \in \mathcal{D}} \sum_{e \in E} c_e r_e^j$$

$$\text{(CFL-LP)} \quad \text{s.t.} \quad \sum_{e\in\delta(S)} (b_e + \rho_{cr} r_e^j) \geq \rho_{cr} \quad \forall S \subset V, i^* \notin S, j \in \mathcal{D} \cap S,$$

$$r_e^j, b_e \geq 0 \quad \forall e \in E, j \in \mathcal{D}.$$

Let $\hat{b}, \hat{r}$ be an optimal solution to CFL-LP. Given an optimal solution to the original problem, let B^*, R^* be the total buying and renting cost. We also define O^* as the facility opening cost in the optimal solution. It is easily verified that the optimal value of CFL-LP is at most $B^* + (a+\gamma)\rho_{cr} R^*$. We define $y_e(D) = \hat{b}_e + \rho_{cr}\hat{r}_e^{j^*} + \rho_{cr}\sum_{j\in D}\hat{r}_e^j$, which satisfies condition (iii).

To show that there exists j^* and $\hat{p}$ such that condition (iv) holds, let $\tilde{\mathbb{E}}_p[C_{aug}(\mathbb{D})]$ denote the expectation of $\mathbb{E}_p[C_{aug}(\mathbb{D})]$ if we run CFL-Sample-Augment with the input client j^* chosen uniformly at random and similarly define $\tilde{\mathbb{E}}_p[C_{LP}(y(\mathbb{D}))]$. We claim that if we can find $\hat{p}$ such that $\tilde{\mathbb{E}}_{\hat{p}}[C_{aug}(\mathbb{D})] + \alpha\tilde{\mathbb{E}}_{\hat{p}}[C_{LP}(y(\mathbb{D}))] \leq \beta OPT$, then this implies that we can construct a deterministic β-approximation algorithm: By definition of $\tilde{\mathbb{E}}_{\hat{p}}[\cdot]$ there exists some j^* for which condition (iv) of Theorem 1 holds with the same $\hat{p}$ and β. Since we can compute $\mathbb{E}_{\hat{p}}[C_{aug}(\mathbb{D})] + \alpha\mathbb{E}_{\hat{p}}[C_{LP}(y(\mathbb{D}))]$ efficiently for any choice of j^*, it remains to choose as j^* the client for which this value is smallest, and then we can use Theorem 1 to derandomize the CFL-Sample-Augment algorithm.

We now show that $\tilde{\mathbb{E}}_{\hat{p}}[C_{aug}(\mathbb{D})] + \alpha\tilde{\mathbb{E}}_{\hat{p}}[C_{LP}(y(\mathbb{D}))] \leq \beta OPT$ for appropriately chosen $\hat{p}$ and β. Let $\hat{p}_j = \frac{a}{M}$ for every $j \in \mathcal{D}$, then the probability that we add $\rho_{cr}\hat{r}_e^j$ to $y_e(\mathbb{D}) = \hat{b}_e + \rho_{cr}\hat{r}_e^{j^*} + \rho_{cr}\sum_{j\in\mathbb{D}}\hat{r}_e^j$ is the probability that $j \in \mathbb{D} \cup \{j^*\}$ which is at most $\frac{a}{M} + \frac{1}{|\mathcal{D}|}$. Hence $\tilde{\mathbb{E}}_{\hat{p}}[C_{LP}(y(\mathbb{D}))] \leq B^* + (a+\gamma)\rho_{cr}R^*$. Depending on whether the connectivity requirement is a tour or a tree, and whether k is finite or infinite, Eisenbrand et al. [5] give different lemmas bounding $\tilde{\mathbb{E}}_{\hat{p}}[C_{aug}(\mathbb{D})]$ in terms of B^*, R^* and O^*. We will state these bounds below in Lemmas 12, 14 and 16. Combining these bounds with $\tilde{\mathbb{E}}_{\hat{p}}[C_{LP}(y(\mathbb{D}))] \leq B^* + (a+\gamma)\rho_{cr}R^*$, we can obtain bounds on $\tilde{\mathbb{E}}_{\hat{p}}[C_{aug}(\mathbb{D})] + \alpha\tilde{\mathbb{E}}_{\hat{p}}[C_{LP}(y(\mathbb{D}))]$ in terms of $OPT = B^* + R^* + O^*$.

Before we proceed to give the results we can thus obtain, we note that we can assume that γ is very small: Eisenbrand et al. [5] show that if $\frac{1}{\gamma} = \frac{|\mathcal{D}|}{M} < C$ for some constant C, then there exists a (deterministic) polynomial-time approximation scheme (PTAS) for the connected facility location problem. Hence we can choose a small constant $1/C$ and use the PTAS for values of γ that are larger than $1/C$.

For the first result (which was also shown by Eisenbrand et al. [5]) we need the following lemma.

Lemma 12 [5] *Let $k = \infty$ and $CR =$* `SteinerTree`. *In the Augmentation Step of CFL-Sample-Augment, use a bifactor approximation algorithm* [22] *that returns a solution such that $\sum_{i\in F_U} f_i + \sum_{j\in D}\ell(j,\sigma_U(j)) \leq (1.11 + \ln\delta)O^* + (1 + \frac{0.78}{\delta})R^*$. Then*

$$\tilde{\mathbb{E}}_{\hat{p}}[C_{aug}(\mathbb{D})] \leq 2R^* + \frac{0.807}{a}B^* + (1+a+\gamma)\left((1.11+\ln\delta)O^* + \left(1+\frac{0.78}{\delta}\right)R^*\right).$$

Theorem 13 [5] *There exists a deterministic* 4.23-*approximation algorithm for k-connected facility location with* $k = \infty$ *and* $CR = \texttt{SteinerTree}$.

Proof By Lemma 12, and because $\alpha = 2$, $\rho_{\texttt{SteinerTree}} = 1$ in this case, we get that

$$\begin{aligned}&\tilde{\mathbb{E}}_{\hat{p}}[C_{aug}(\mathbb{D})] + \alpha\tilde{\mathbb{E}}_{\hat{p}}[C_{LP}(y(\mathbb{D}))]\\&\quad\le (1+a+\gamma)\left(3+\frac{0.78}{\delta}\right)R^* + \left(2+\frac{0.807}{a}\right)B^*\\&\qquad + (1+a+\gamma)(1.11+\ln\delta)O^*.\end{aligned}$$

By taking $a = 0.361885$, $\delta = 7.359457$ and γ sufficiently small, we find that this is at most $4.23OPT$ and by the discussion above, this means that there exists a deterministic 4.23-approximation algorithm. □

The second result was also shown by Eisenbrand et al. [5]. To derive it using our framework, we need the following lemma.

Lemma 14 [5] *Let* $k < \infty$ *and* $CR = \texttt{SteinerTree}$, *and suppose we use a* ρ_{kfl}-*approximation algorithm to find a solution to the unconnected k-facility location problem in the Augmentation Step of CFL-Sample-Augment*, *then*

$$\tilde{\mathbb{E}}_{\hat{p}}[C_{aug}(\mathbb{D})] \le 2R^* + \frac{0.807}{a}B^* + (1+a+\gamma)\rho_{kfl}(R^* + O^*).$$

Theorem 15 [5] *There exists a deterministic* 6.98-*approximation algorithm for k-connected facility location with* $k < \infty$ *and* $CR = \texttt{SteinerTree}$.

Proof By Lemma 14, and because $\alpha = 2$, $\rho_{\texttt{SteinerTree}} = 1$, we get that

$$\begin{aligned}&\tilde{\mathbb{E}}_{\hat{p}}[C_{aug}(\mathbb{D})] + \alpha\tilde{\mathbb{E}}_{\hat{p}}[C_{LP}(y(\mathbb{D}))]\\&\quad\le (1+a+\gamma)(2+\rho_{kfl})R^* + \left(2+\frac{0.807}{a}\right)B^* + (1+a+\gamma)\rho_{kfl}O^*.\end{aligned}$$

Using a 4-approximation algorithm for the (unconnected) k-facility location problem [1] in the Augmentation Step, we have $\rho_{kfl} = 4$. Taking $a = 0.1623$ and γ sufficiently small, we find that $\tilde{\mathbb{E}}_{\hat{p}}[C_{aug}(\mathbb{D})] + \alpha\tilde{\mathbb{E}}_{\hat{p}}[C_{LP}(y(\mathbb{D}))] \le 6.98OPT$. □

Eisenbrand et al. [5] do not give a deterministic algorithm for connected facility location where the facilities need to be connected by a tour. Using the following lemma and our analysis, the existence of a deterministic algorithm readily follows.

Lemma 16 [5] *Let* $k = \infty$ *and* $CR = \texttt{Tour}$. *In the Augmentation Step of CFL-Sample-Augment*, *use a bifactor approximation algorithm* [22] *that returns a solution such that* $\sum_{i\in F_U} f_i + \sum_{j\in D}\ell(j,\sigma_U(j)) \le (1.11+\ln\delta)O^* + (1+\frac{0.78}{\delta})R^*$. *Then*

$$\tilde{\mathbb{E}}_{\hat{p}}[C_{aug}(\mathbb{D})] \le 2R^* + \frac{1}{2a}B^* + (1+2(a+\gamma))\left((1.11+\ln\delta)O^* + \left(1+\frac{0.78}{\delta}\right)R^*\right).$$

Theorem 17 *There exists a deterministic* 4.12*-approximation algorithm for k-connected facility location with* $k=\infty$ *and* $CR=\texttt{Tour}$.

Proof By Lemma 16, and the fact that $\alpha=1.5$ and $\rho_{\texttt{Tour}}=2$, we get that

$$\begin{aligned}
&\tilde{\mathbb{E}}_{\hat{p}}[C_{aug}(\mathbb{D})]+\alpha\tilde{\mathbb{E}}_{\hat{p}}[C_{LP}(y(\mathbb{D}))]\\
&\quad\le\left(0.5+(1+2(a+\gamma))\left(2.5+\frac{0.78}{\delta}\right)\right)R^*\\
&\qquad+\left(1.5+\frac{1}{2a}\right)B^*+(1+2(a+\gamma))(1.11+\ln\delta)O^*.
\end{aligned}$$

Taking $a=0.19084$, $\delta=6.5004$ and γ sufficiently small, we find that $\tilde{\mathbb{E}}_{\hat{p}}[C_{aug}(\mathbb{D})]+\alpha\tilde{\mathbb{E}}_{\hat{p}}[C_{LP}(y(\mathbb{D}))]\le 4.12OPT$. □

Finally, we mention that the results given above can easily be extended to the case when the demands are not necessarily all equal to 1. We now let $\gamma=\frac{M}{\sum_{j'\in\mathcal{D}}d_{j'}}$. It is again the case that there exists a PTAS for the connected facility location problem if $\gamma>\frac{1}{C}$ for some constant C and hence we can assume γ is very small [5]. Now, the first client that is chosen is client j with probability $\frac{d_j}{\sum_{j'\in\mathcal{D}}d_{j'}}$, and we initialize the vector of sampling probabilities by $\hat{p}_j=\frac{a}{M}d_j$ for all $j\in\mathcal{D}$. Then Lemmas 12, 14 and 16 again hold. The objective of CFL-LP can be changed to $\sum_{e\in E}Mc_eb_e+(a+\gamma)\rho_{cr}\sum_{j\in\mathcal{D}}d_j\sum_{e\in E}c_er_e^j$, and we will again have that for an optimal solution $\hat{b},\hat{r}$ to CFL-LP we have $\sum_{e\in E}Mc_e\hat{b}_e+(a+\gamma)\rho_{cr}\sum_{j\in\mathcal{D}}d_j\sum_{e\in E}c_e\hat{r}_e^j\le B^*+(a+\gamma)\rho_{cr}R^*$. Therefore the same definition of $y_e(D)$ as above will ensure that $\tilde{\mathbb{E}}_{\hat{p}}[C_{LP}(y(\mathbb{D}))]\le B^*+(a+\gamma)\rho_{cr}R^*$, and hence we have the exact same inequalities that we needed to prove Theorems 13, 15, and 17.

3.5 Virtual Private Network Design

In the virtual private network design problem, we are given a graph $G=(V,E)$ with edge costs $c_e\ge 0$, and a set of demands $\mathcal{D}\subseteq V$. Each demand $j\in\mathcal{D}$ has thresholds $b_{in}(j), b_{out}(j)$ on the amount of traffic that can enter and leave j.

A feasible solution is a set of paths P_{ij} for every ordered pair $i,j\in\mathcal{D}$ and capacity u_e on the edges so that there is sufficient capacity for any *traffic pattern* $\{f_{ij}\}_{i,j\in\mathcal{D}}$: For any $\{f_{ij}\}_{i,j\in\mathcal{D}}$ such that $\sum_i f_{ij}\le b_{in}(j)$ and $\sum_i f_{ji}\le b_{out}(j)$ for every $j\in\mathcal{D}$ we need to have sufficient capacity on the paths, i.e. $\sum_{ij:e\in P_{ij}}f_{ij}\le u_e$ for every $e\in E$. The objective is to find a solution that minimizes the cost $\sum_{e\in E}c_eu_e$ of installing capacity.

Gupta et al. [15] proposed a random sampling algorithm for the virtual private network design problem that is very similar to the algorithm for single source rent-or-buy. The algorithm and analysis were improved by Eisenbrand and Grandoni [3] and Eisenbrand, Grandoni, Oriolo and Skutella [4]. We will show how Theorem 1 can be used to derandomize the improved algorithm in [4].

VPN-Sample-Augment$(G=(V,E),c,\mathcal{J},\mathcal{I},p)$

1. (*Sampling Step*) Mark each receiver j independently with probability p_j. Let D be the set of marked receivers. If $|D|=0$ then repeat the Sampling Step.
2. (*Subproblem Step*) For each sender $i\in\mathcal{I}$, construct a Steiner tree $T(i)$ on $D\cup\{i\}$ and add one unit of capacity to each edge of $T(i)$.
3. (*Augmentation Step*) Install one unit of capacity on the shortest path from each receiver $j\in\mathcal{J}$ to the closest receiver in D.

Fig. 5 Sample-Augment algorithm for virtual private network design

As was shown by Gupta et al. [15], we assume without loss of generality that each $j\in\mathcal{D}$ is either a *sender* ($b_{in}(j)=0, b_{out}(j)=1$) or a *receiver* ($b_{in}(j)=1$, $b_{out}(j)=0$). Let $\mathcal{J}$ be the set of receivers, and $\mathcal{I}$ be the set of senders. By symmetry, we assume without loss of generality that $|\mathcal{I}|\le|\mathcal{J}|$.

The algorithm as described by Eisenbrand et al. [4] partitions $\mathcal{J}$ into $\mathcal{I}$ groups, by assigning each receiver to a randomly chosen sender, and chooses one non-empty group, say D, at random. In the Subproblem Step, we add one unit of capacity on a Steiner tree spanning $\{i\}\cup D$ for each sender i, and finally, in the Augmentation Step we install one unit of capacity on the shortest path from each receiver j to the closest receiver in D.

For our derandomization, our starting point is a slightly different algorithm which marks each receiver independently with some probability p_j and repeats the Sampling Step if the resulting set D is empty. If we set $p_j=\frac{1}{|\mathcal{I}|}$, then the outcome of our algorithm has the same probability distribution as the algorithm described by Eisenbrand et al. [4].

The VPN-Sample-Augment algorithm is described in Fig. 5. The algorithm installs capacities and outputs the Steiner trees found in the Subproblem Step. The Steiner trees are used to determine the paths P_{ij}: if j' is the receiver in D that is closest to j, then P_{ij} is obtained by concatenating the unique path from j' to i in $T(i)$ and the shortest path from j to j'.

The following lemma follows from Lemmas 5 and 6 in [4].

Lemma 18 [4] *If* $p_j=\frac{1}{|\mathcal{I}|}$ *for every terminal, and if we were able to find a minimum cost solution to the subproblem, then* $\mathbb{E}_p[C_{sub}(\mathbb{D})\mid|\mathbb{D}|\ge 1]\le\frac{1}{1-e^{-|\mathcal{J}|/|\mathcal{I}|}}OPT$, *and* $\mathbb{E}_p[C_{aug}(\mathbb{D})\mid|\mathbb{D}|\ge 1]\le\frac{2}{1-e^{-|\mathcal{J}|/|\mathcal{I}|}}OPT$.

Eisenbrand et al. [4] also show that there exists a (deterministic) $(1+\frac{|\mathcal{J}|}{|\mathcal{I}|})$-approximation algorithm. This gives an 8-approximation algorithm for $|\mathcal{J}|\le 7|\mathcal{I}|$. Using Theorem 2 we can show that if $|\mathcal{J}|\ge 7|\mathcal{I}|$, then there exists a deterministic 8.02-approximation algorithm.

Theorem 19 *There exists a deterministic* 8.02-*approximation algorithm for virtual private network design.*

Proof It is easily verified that condition (i) of Theorem 2 holds for all p with $p_j = 1$ for some j. The Sub-LP for condition (ii) is the sum of $|\mathcal{I}|$ different Steiner tree LPs, and has $\alpha = 2$ [11]:

$$\begin{array}{lll} & \min & \sum_{e\in E} c_e \sum_{i\in\mathcal{I}} y_e^i \\ \text{Sub-LP}(D) & \text{s.t.} & \sum_{e\in\delta(S)} y_e^i \ge 1 \quad \forall i\in\mathcal{I}, \forall S\subset V: i\in S, D\cap S\ne\emptyset, \\ & & y_e^i \ge 0 \quad \forall i\in\mathcal{I}, e\in E. \end{array}$$

Let $\kappa = 1 - e^{-\frac{|\mathcal{J}|}{|\mathcal{I}|}}$. It follows from Lemma 18 that the following LP is a relaxation of the virtual private network design problem.

$$\begin{array}{lll} & \min & \frac{\kappa}{3}\sum_{e\in E} c_e \left(\sum_{i\in\mathcal{I}} b_e^i + \sum_{j\in\mathcal{J}} r_e^j\right) \\ \text{(VPN-LP)} & \text{s.t.} & \sum_{e\in\delta(S)} (b_e^i + r_e^j) \ge 1 \quad \forall S\subseteq V: i\in S\cap\mathcal{I}, j\in\mathcal{J}\backslash S, \\ & & r_e^j, b_e^i \ge 0 \quad \forall e\in E, j\in\mathcal{J}, i\in\mathcal{I}. \end{array}$$

Let $\hat{r}, \hat{b}$ be an optimal solution to VPN-LP. We let $y_e^i(D) = \hat{b}_e^i + \sum_{j\in D}\hat{r}_e^j$, which gives a feasible solution to Sub-LP(D).

If we include each $j \in \mathcal{J}$ in $\mathbb{D}$ independently with probability $\hat{p}_j = \frac{1}{|\mathcal{I}|}$, then $\mathbb{P}[j \in \mathbb{D} \mid |\mathbb{D}| \ge 1] = \hat{p}_j/(1 - \prod_k(1-\hat{p}_k)) = \hat{p}_j/(1-(1-\frac{1}{|\mathcal{I}|})^{|\mathcal{J}|}) \le \hat{p}_j/(1-e^{-|\mathcal{J}|/|\mathcal{I}|}) = \frac{1}{|\mathcal{I}|}\frac{1}{\kappa}$. Hence

$$\begin{aligned} \mathbb{E}_{\hat{p}}[C_{LP}(y(\mathbb{D})) \mid |\mathbb{D}| \ge 1] &= \sum_{e\in E} c_e \sum_{i\in\mathcal{I}} \left(\hat{b}_e^i + \sum_{j\in\mathcal{J}} \mathbb{P}[j\in\mathbb{D} \mid |\mathbb{D}|\ge 1]\hat{r}_e^j\right) \\ &\le \sum_{e\in E} c_e \sum_{i\in\mathcal{I}} \left(\hat{b}_e^i + \sum_{j\in\mathcal{J}} \frac{1}{\kappa|\mathcal{I}|}\hat{r}_e^j\right) \\ &= \sum_{e\in E} c_e \left(\sum_{i\in\mathcal{I}} \hat{b}_e^i + \frac{1}{\kappa}\sum_{j\in\mathcal{J}} \hat{r}_e^j\right) \\ &\le \frac{1}{\kappa}\sum_{e\in E} c_e \left(\sum_{i\in\mathcal{I}} \hat{b}_e^i + \sum_{j\in\mathcal{J}} \hat{r}_e^j\right) \\ &\le \frac{3}{\kappa^2} OPT. \end{aligned}$$

Also, we know from Lemma 18 that

$$\mathbb{E}_{\hat{p}}[C_{aug}(\mathbb{D}) \mid |\mathbb{D}| \geq 1] \leq \frac{2}{\kappa} OPT.$$

Therefore

$$\mathbb{E}_{\hat{p}}[C_{aug}(\mathbb{D}) \mid |\mathbb{D}| \geq 1] + 2\mathbb{E}_{\hat{p}}[C_{LP}(y(\mathbb{D})) \mid |\mathbb{D}| \geq 1] \leq \frac{2 + 2 \times 3/\kappa}{\kappa} OPT.$$

Hence we have shown that if $\hat{p}_j = \frac{1}{|\mathcal{I}|}$ for all $j \in \mathcal{J}$, then

$$\mathbb{E}_{\hat{p}}[C_{aug}(\mathbb{D}) \mid |\mathbb{D}| \geq 1] + 2\mathbb{E}_{\hat{p}}[C_{LP}(y(\mathbb{D})) \mid |\mathbb{D}| \geq 1] \leq \frac{2 + 6/\kappa}{\kappa} OPT.$$

Recall that there exists a (deterministic) $(1 + \frac{|\mathcal{J}|}{|\mathcal{I}|})$-approximation algorithm, hence to prove the theorem we only need to consider $\frac{|\mathcal{J}|}{|\mathcal{I}|} \geq 7$. Since $\kappa = 1 - e^{-\frac{|\mathcal{J}|}{|\mathcal{I}|}}$, we get that $\frac{2+6/\kappa}{\kappa} \leq 8.02$. Therefore by Theorem 2 there exists a deterministic 8.02-approximation algorithm. □

4 Single-Sink Buy-at-Bulk Network Design

Sampling algorithms have been successfully used for various multi-stage stochastic optimization problems. In a multi-stage sampling algorithm, we mark a subset of the vertices and solve a subproblem on the marked vertices *in each stage of the algorithm*. Clearly, a difficulty in derandomizing such an algorithm using our approach is that the cost incurred by the algorithm in future stages often depends on the decisions made in the current stage, and are hence difficult to get a handle on.

One multi-stage sampling algorithm where it is possible to use the techniques from Sect. 2 is the Sample-Augment algorithm for the single sink buy-at-bulk problem [18]. The single sink buy-at-bulk problem is similar to the single source rent-or-buy problem, but instead of just having the option of either renting or buying an edge, we now have a choice of K different cable types, where each cable type has a certain capacity and price per unit length. This algorithm has stages corresponding to the cable types, and in stage k we install cables of type k and $k+1$ only. The cables installed in the current stage are then used to (randomly) redistribute the demands in the network, which means that the input to the next stage is not deterministic.

There are three key properties that allow us to derandomize this algorithm. First of all, it turns out that the Sampling Step of each stage does not influence the expected cost of the future stages (although it does influence its distribution). Hence we can almost directly use the techniques from Sect. 2 to derandomize the Sampling Step. Secondly, as we will see, the random redistributing of the demands has only a small sample space, so we can enumerate all possible outcomes. Thirdly, we can give an efficiently computable upper bound for the cost of future stages, hence allowing us to choose a good outcome among these possible outcomes. We will show that we can thus obtain a deterministic 27.72-approximation algorithm for the single-sink buy-at-bulk problem.

We note that our approximation guarantee is obtained with respect to an LP-relaxation of the single-sink buy-at-bulk problem. Hence we also show that this LP relaxation has integrality gap of at most 27.72. To the best of the author's knowledge, no previous result was known about the integrality gap of this LP relaxation. In [27] it was shown that a different LP relaxation, that was first proposed in [9], has an integrality gap of at most 216.

In Sect. 4.1 we describe the Sample-Augment algorithm of Gupta, Kumar, Pál and Roughgarden [18] and explain how to derandomize this algorithm to obtain a deterministic 80-approximation algorithm. We will suppress some of the proofs, which we will give in Sect. 4.2. Finally, in Sect. 4.3, we discuss an improved Sample-Augment algorithm for the single-sink buy-at-bulk network design problem by Grandoni and Italiano [13], and show how this can be derandomized to give a deterministic 27.72-approximation algorithm.

4.1 Derandomization of the Sample-Augment Algorithm for Single-Sink Buy-at-Bulk

The single sink buy-at-bulk problem is a generalization of the single source rent-or-buy problem. We are given an undirected graph $G = (V, E)$, edge costs $c_e \geq 0$ for $e \in E$, a sink $t \in V$ and a set of sources $s_1, \dots, s_n \in V$ with weight $w_j > 0$ for source s_j. We denote $\{s_1, \dots, s_n\} = \mathcal{S}$. In addition, there are K cable types, where the k-th cable type has capacity u_k and cost σ_k per unit length. The goal is to install sufficient capacity at minimum cost so that we can send w_j units from s_j to t for $j = 1, \dots, n$ simultaneously. We assume without loss of generality that $u_1 < u_2 < \dots < u_K$ and that $\sigma_1 < \sigma_2 < \dots < \sigma_K$, since if $u_k \leq u_\ell$ and $\sigma_k \geq \sigma_\ell$, then we can replace each cable of type k by a cable of type ℓ without increasing the cost of the solution [18]. Note that the single source rent-or-buy problem is the special case where $K = 2$ and $u_1 = 1, u_2 = \infty$ and $\sigma_1 = 1, \sigma_2 = M$.

After a preprocessing step, the Sample-Augment algorithm proposed by Gupta et al. [18] proceeds in stages, where in the k-th stage, it will install cables of type k and $k + 1$. At the beginning of stage k, enough capacity has already been installed to move the weights through the cables and gather the weights into a subset of the sources, so that each source has weight either 0 or u_k. We thus think of the weights at the beginning of stage k as being concentrated in $S_k \subseteq \mathcal{S}$, where each $s \in S_k$ has weight u_k. The final step installs cables of type K from S_K to the sink t.

As in [18], we first round the parameters so that each parameter u_k, σ_k is a factor of 2. It was shown by Guha, Meyerson and Munagala [14] that we can round each u_k down to the nearest power of 2, and each σ_k up to the nearest power of 2, and increase the value of the optimal solution by at most a factor 4. Note that we may assume without loss of generality that the *rounded parameters* satisfy that $\frac{\sigma_{k+1}}{u_{k+1}} < \frac{\sigma_k}{u_k}$, since we may otherwise replace each cable of type $k + 1$ by $\frac{u_{k+1}}{u_k}$ cables of type k. We are guaranteed that $\frac{u_{k+1}}{u_k}$ is integer because numerator and denominator are powers of 2.

$\mathbf{w}' = \text{Redistribute}(T, w, U)$
Let $\underline{w}_v = \lfloor \frac{w_v}{U} \rfloor U$ and $w_v^{frac} = w_v - \underline{w}_v$ for every $v \in T$. Double the edges of T and find an Euler tour on the edges. Number the vertices $v_1, v_2, \dots, v_n$ according to their first appearance on the Euler tour started from an arbitrarily chosen vertex v_1. Let $W_j = \sum_{k=1}^{i} w_{v_k}^{frac}$ for $j = 1, \dots, n$. Choose $Y \in (0, U]$ uniformly at random. Let $w'_{v_j} = \begin{cases} \underline{w}_{v_j} + U & \text{if } Y \in (W_{i-1} - \lfloor \frac{W_{i-1}}{U} \rfloor U, W_i - \lfloor \frac{W_{i-1}}{U} \rfloor U] \\ \underline{w}_{v_j} & \text{otherwise.} \end{cases}$

Fig. 6 The redistribute subroutine

Following [18], we also rescale the parameters so that $u_1 = \sigma_1 = 1$. Note however that after rescaling the capacities, we are not guaranteed that the weights w_j are integer. Gupta et al. [18] therefore use a subroutine to redistribute the weights. The subroutine takes a tree with weights w_v on the vertices $v \in T$, and a parameter U and redistributes the weights along the edges of the tree, so that vertex v's weight becomes either $\lfloor \frac{w_v}{U} \rfloor U$ or $\lceil \frac{w_v}{U} \rceil U$.

In the following, we will use bold lower case letters to indicate (vectors of) random variables (and we continue to use blackboard bold capitalized letters to indicate random sets). Our description of the subroutine is given in Fig. 6. The following lemma is a reformulation of Lemma 5.1 in [18].

Lemma 20 [18] *Given a tree T, a parameter $U > 0$ and weights $w_v \geq 0$ for every $v \in T$, such that $\sum_{v \in T} w_v$ is a multiple of U, let $\underline{w}_v = \lfloor \frac{w_v}{U} \rfloor U$. The subroutine* Redistribute(T, w, U) *outputs weights $\mathbf{w}'_v$ for $v \in T$ so that*:

(i) $\mathbb{P}[\mathbf{w}'_v = \underline{w}_v + U] = \frac{w_v - \underline{w}_v}{U}$, *and* $\mathbb{P}[\mathbf{w}'_v = \underline{w}_v] = 1 - \frac{w_v - \underline{w}_v}{U}$.
(ii) *With probability* 1, *there exists a flow on T such that the net flow into $v \in T$ is $\mathbf{w}'_v - w_v$ and the flow on each edge of T is at most U.*

In the preprocessing step, Gupta et al. [18] find a Steiner tree T_0 on $\mathcal{S} \cup \{t\}$, install a cable of type 1 on each edge of T_0 and let $\mathbf{w}' = \text{Redistribute}(T_0, w, 1)$. We let $S_0 = \mathcal{S}$ and $\mathbb{S}_1 = \{s_j \in S_0 : \mathbf{w}'_j > 0\}$, where we note that $\mathbb{S}_1$ is a random set, since $\mathbf{w}'$ is a vector of random variables. For ease of exposition, we assume that $w_j \leq 1$ for each $s_j \in \mathcal{S}$, so that after redistributing each source in S_1 has weight 1. This assumption is not necessary, as was noted in [18]. In addition, we assume that $\sum_{s_j \in \mathcal{S}} w_j$ is a power of 2 and is at least u_K. This assumption can be made without loss of generality, as we can add a dummy source located at the sink, with weight $u_K - \sum_{s_j \in \mathcal{S}} w_j$ (see [18]).

For ease of notation, we will refer to the preprocessing step as stage 0, the intermediate stages as stages $1, \dots, K-1$ and the final stage as stage K. At the beginning of stage k, for $k \geq 1$, there is sufficient capacity installed in the previous stages to move the weights from S_0 to some subset $S_k \subset S_0$ in such a way that each source in S_k gets exactly u_k weight. We will say that "the weights are located at S_k at the start of stage k". In the k-th stage of the algorithm, we will install capacity to move the weights from S_k to some subset $S_{k+1} \subset S_k$: First we obtain a sample D_k from S_k. For

SSBaB-Sample-Augment$(G=(V,E),c,t,\mathcal{S},u_1,\ldots,u_K,\sigma_1,\ldots,\sigma_K)$

(*Rounding Step*)
Round down all capacities u_k and round up all prices σ_k to the nearest power of 2. Remove cable type k if $\frac{\sigma_k}{u_k} \geq \frac{\sigma_{k-1}}{u_{k-1}}$. Let K be the remaining number of cable types.
(*Preprocessing Step*)
Find a Steiner tree T_0 on $\mathcal{S} \cup \{t\}$, install a cable of type 1 on each edge of T_0. Let $w' = \text{Redistribute}(T_0, w, 1)$. Let $S_1 = \{s_j \in \mathcal{S} : w'_j > 0\}$.
(*Intermediate Stages*)
For $k = 1, \ldots, K-1$

$$\hat{p}^k_j = \frac{\sigma_k}{\sigma_{k+1}} \text{ for } s_j \in S_k,$$

$$S_{k+1} = \text{SSBaB-Stage-}k(G, c, t, S_k, u_k, u_{k+1}, \hat{p}^k).$$

(*Final Stage*)
For every $s \in S_K$, install cables of type K on the shortest path from s to t.

Fig. 7 Sample-Augment algorithm for single-sink buy-at-bulk

each source $s \in S_k$, let $f(s)$ be the closest node in $D_k \cup \{t\}$. We install type k cables from $s \in S_k$ to $f(s)$ and we install type $k+1$ cables on a Steiner tree on $D_k \cup \{t\}$. Since each source in S_k has weight u_k, we can route the weights to $D_k \cup \{t\}$ along the type k cables. We can then use the Redistribute-subroutine to redistribute the weights along the Steiner tree so that the weight w'_v at each node v in $D_k \cup \{t\}$ is an integer multiple of u_{k+1}. Finally, we need to route the weights back to a subset of the sources in S_k. For each vertex in $v \in D_k \cup \{t\}$, we divide the weight at v into $\frac{w'_v}{u_{k+1}}$ packets of size u_{k+1}. We then choose a subset $S'(v)$ of size $\frac{w'_v}{u_{k+1}}$ at random from the sources in S_k that previously sent their weight to v, i.e. $\{s \in S_k : f(s) = v\}$. We install type $k+1$ cables on the shortest paths from v to $S'(v)$ and send u_{k+1} weight to each node in $S'(v)$. If $k < K$ then the next stage has the weights located at $S_{k+1} = \sum_{v \in D_k \cup \{t\}} S'(v)$. We give a complete description of the Sample-Augment algorithm from Gupta et al. [18] in Figs. 7 and 8. Note that we have structured the description in Fig. 8 so that if we consider only a single stage, the three steps fit the framework in Theorem 1.

We begin by showing that we can replace the costs incurred by the algorithm by certain upper bounds. We will see that these upper bounds allow us to bound the expected cost incurred by the algorithm, and that they will have an easy form that will help in derandomizing the algorithm. We will need several lemmas, but to keep the flow of the arguments we defer some of the proofs to Sect. 4.2.

The following lemma is similar to Lemma 5.2 in [18] and will be useful throughout this section.

Lemma 21 *For any $k = 1, \ldots, K-1$, let $\mathbb{S}_k$ be the (random) set of sources at which the weights are located at the beginning of stage k. Given $S_k \subset \mathcal{S}$ and $k < \ell \leq K$,*

$$\mathbb{P}[s \in \mathbb{S}_\ell | \mathbb{S}_k = S_k] = \begin{cases} \frac{u_k}{u_\ell} & \text{if } s \in S_k, \\ 0 & \text{otherwise,} \end{cases}$$

$\mathbb{S}_{k+1} =$ **SSBaB-stage-**$k(G, c, t, S_k, u_k, u_{k+1}, p^k)$

1. (*Sampling Step*) Mark each source $s_j \in S_k$ independently with probability p_j^k. Let D_k be the set of marked sources.
2. (*Subproblem Step*) Construct a Steiner tree T_k on $D_k \cup \{t\}$ and install cables of type $k+1$ on the edges of T_k.
3. (*Augmentation Step*)
   For each $s \in S_k$, let $f(s) = \arg\min_{v \in D_k \cup \{t\}} \ell(s, v)$. Install cables of type k on the shortest path from each $s \in S_k$ to $f(s)$.
   For each $v \in D_k \cup \{t\}$, let $S(v) = \{s \in S_k : f(s) = v\}$ and let $w_v = \sum_{s \in S(v)} u_k$.
   Let $w' = \text{Redistribute}(T_k, w, u_{k+1})$.
   For each $v \in D_k \cup \{t\}$, let $S'(v)$ be a set of $\frac{w'_v}{u_{k+1}}$ sources chosen uniformly at random from $S(v)$. Install a cable of type $k+1$ on the shortest path from v to each source in $S'(v)$.
   Return $S_{k+1} = \sum_{v \in D_k \cup \{t\}} S'(v)$.

Fig. 8 The k-th stage of the Sample-Augment algorithm for single-sink buy-at-bulk

independent of the sampling probabilities in stage k.

Lemma 22 *For any* $k = 0, \ldots, K$, *let* $\mathbb{S}_k$ *be the random set of terminals at which the weight is located at the start of stage* k *(where* $\mathbb{S}_0 \equiv \mathcal{S}$*). For any* $S_k \subseteq \mathcal{S}$, *let* $\mathbb{E}_{\hat{p}^k}[C_k(\mathbb{S}_k)|\mathbb{S}_k = S_k]$ *be the expected cost of the cables installed in stage* k, *given that the weights are located in* S_k *at the start of stage* k, *and the sampling probabilities in stage* k *are given by* $\hat{p}_j^k = \frac{\sigma_k}{\sigma_{k+1}}$ *if* $1 \le k \le K-1$. *There exist values* B_k, $R_k(j)$ *for* $j = 1, \ldots, n$ *such that*

$$\mathbb{E}_{\hat{p}^k}[C_k(\mathbb{S}_k)|\mathbb{S}_k = S_k] \le B_k + \sum_{s_j \in S_k} R_k(j),$$

and

$$B_0 + \sum_{s_j \in \mathcal{S}} R_0(j) + \sum_{k=1}^{K} \left(B_k + \sum_{s_j \in \mathcal{S}} \frac{w_j}{u_k} R_k(j) \right) \le 80 OPT.$$

These two lemmas immediately show that the SSBaB-Sample-Augment algorithm is a *randomized* 80-approximation algorithm. This guarantee is worse than the guarantee of 76.8 in Gupta et al. [18]. However, our analysis will be helpful in the derandomization of the algorithm.

Corollary 23 *There exists a randomized* 80-*approximation algorithm for the single sink buy-at-bulk problem.*

Proof By Lemma 21 we know that $\mathbb{P}[s_j \in \mathbb{S}_k|\mathbb{S}_1 = S_1] = \frac{u_1}{u_k} = \frac{1}{u_k}$, if $s_j \in S_1$ and $\mathbb{P}[s_j \in \mathbb{S}_k|\mathbb{S}_1 = S_1] = 0$ if $s_j \notin S_1$, and by Lemma 20, we have that $\mathbb{P}[s_j \in \mathbb{S}_1] = w_j$, hence $\mathbb{P}[s_j \in \mathbb{S}_k] = \frac{w_j}{u_k}$. By linearity of expectation and Lemma 22, we can thus upper bound the expected cost incurred in stage k by $B_k + \sum_{s_j \in \mathcal{S}} \frac{w_j}{u_k} R_k(j)$. Since $\mathbb{P}[s_j \in S_0] = 1$ for $s_j \in \mathcal{S}$ we can upper bound the expected cost incurred in stage 0 by $B_0 + \sum_{s_j \in \mathcal{S}} R_0(j)$. Hence Lemma 22 implies that the randomized algorithm in Figure 7 is an 80-approximation algorithm. □

Starting with our upper bound $B_0 + \sum_{s_j \in \mathcal{S}} R_0(j) + \sum_{k=1}^{K} (B_k + \sum_{s_j \in \mathcal{S}} \frac{w_j}{u_k} R_k(j)) \leq 80OPT$, we would now would like to iterate through the random decisions made by the algorithm and turn them into deterministic decisions, without increasing the overall upper bound on the (conditional) expected cost.

The first random decisions made are those in the preprocessing step, where the Redistribute algorithm is called. These are easy to deal with because of the following lemma.

Lemma 24 *If n is the number of vertices in T, the Redistribute subroutine on T has only $2n + 1$ different possible outcomes.*

Proof For each $v \in T$ there is an interval $(a_v, b_v]$ such that $w'_v = \underline{w}_v + U$ exactly if the random variable Y is in this interval and otherwise $w'_v = \underline{w}_v$. If we think of the values a_v, b_v for all $v \in T$ as points on a line $[0, U]$ then each different outcome corresponds to a segment between two consecutive points (including the endpoints 0 and U). □

By Lemma 24, we can consider all different outcomes of the Redistribute subroutine directly. Each outcome gives a set S_1, and by Lemma 21 we can update the upper bound on the cost as $B_0 + \sum_{s_j \in \mathcal{S}} R_0(j) + \sum_{k=1}^{K} (B_k + \sum_{s_j \in S_1} \frac{1}{u_k} R_k(j))$. By properties of conditional expectation, if we choose the outcome S_1 for which this upper bound is smallest, we will maintain that $B_0 + \sum_{s_j \in \mathcal{S}} R_0(j) + \sum_{k=1}^{K} (B_k + \sum_{s_j \in S_1} \frac{1}{u_k} R_k(j)) \leq 80OPT$.

The next random decisions of the SSBaB-Sample-Augment algorithm are made when marking the sources in the Sampling Step of stage 1. Since by Lemma 21 the probability that $s \in \mathbb{S}_k$ is $\frac{1}{u_k}$ for $k > 1$ and does not depend on the sampling probabilities in stage 1, we can modify the probabilities according to which we sample, and we will not change the expected upper bound on the future stages $\sum_{k=2}^{K} (B_k + \sum_{s_j \in S_1} \frac{1}{u_k} R_k(j))$. We can thus consider only the current stage, and use a similar approach to the derandomization of single-stage Sample-Augment algorithms. We need the following lemma, which combined with Theorem 1 ensures that we can derandomize the Sampling Step of stage 1, while maintaining that the expected total cost of the cables installed in stage 1 is at most $B_1 + \sum_{s_j \in S_1} R_1(j)$.

Lemma 25 *For any $k = 1, \ldots, K - 1$, let B_k, $R_k(j)$, $j = 1, \ldots, n$ satisfy the conditions in Lemma 22. For any $S_k \subset \mathcal{S}$, the following holds for the k-stage of SSBaB-Sample-Augment*:

(i) *The expected cost of the Augmentation Step depends only on $\mathbb{D}_k$, and not on the Subproblem Step, and can be efficiently computed for any p^k.*

(ii) *There exists an LP relaxation Sub-LP*$_k(D_k)$ *for the minimum cost Steiner tree problem on* $D_k \cup \{t\}$ *in the Subproblem Step and an algorithm for finding a Steiner tree on* $D_k \cup \{t\}$ *that finds a solution that costs at most twice the cost of any feasible solution to Sub-LP*$_k(D_k)$.

(iii) *We can compute vectors* b^k *and* $r^k(j)$ *for* $j = 1, \dots, n$ *such that* $y^k(D_k) = b^k + \sum_{s_j \in D_k} r^k(j)$ *is a feasible solution to Sub-LP*$_k(D_k)$ *for any* $D_k \subset S_k$.

(iv) *If* $\hat{p}_j^k = \frac{\sigma_k}{\sigma_{k+1}}$ *for all* $s_j \in S_k$, *then the expectation of twice the objective value of* $y^k(D_k)$ *to Sub-LP*$_k(D_k)$ *plus the expected cost of the cables installed in the Augmentation Step is at most* $B_k + \sum_{s_j \in S_k} R_k(j)$.

Once we have deterministic sample D_1, the Augmentation Step of stage 1 still has two random processes. The first one is the Redistribute subroutine. Since $|D_1 \cup \{t\}| \le n + 1$, by Lemma 24 there are at most $2n + 3$ different outcomes of the Redistribute subroutine. Since D_1 is fixed, we know $f(s)$ for every $s \in S_1$ and hence we also know w_v for every $v \in D_1 \cup \{t\}$. Each outcome of the Redistribute subroutine gives a vector w'. Since $S'(v)$ is obtained by choosing $\frac{w'_v}{u_2}$ sources uniformly at random from the $\frac{w_v}{u_1}$ sources in $S(v)$, we know that the probability that s will be in $\mathbb{S}_2$ is $\frac{w'_{f(s)}}{w_{f(s)}} \frac{u_1}{u_2}$. Hence for each w' we can compute the conditional expectation of the cost of the type 2 cables installed in the Augmentation Step as $\sum_{v \in D_1 \cup \{t\}} \sum_{s_j \in S(v)} \mathbb{P}[s_j \in \mathbb{S}_2 | \mathbb{D}_1 = D_1, \mathbf{w}' = w'] \sigma_2 \ell(s_j, v) = \sum_{v \in D_1 \cup \{t\}} \sum_{s_j \in S(v)} \frac{w'_{f(s_j)}}{w_{f(s_j)}} \frac{u_1}{u_2} \sigma_2 \ell(s_j, v)$ and we can compute the expected upper bound on the cost of the remaining stages as

$$\sum_{k=2}^{K} \Bigg(B_k + \sum_{s_j \in S_1} \mathbb{P}[s_j \in \mathbb{S}_k | s_j \in \mathbb{S}_2] \mathbb{P}[s_j \in \mathbb{S}_2 | \mathbb{D}_1 = D_1, \mathbf{w}' = w'] R_k(j) \Bigg)$$

$$= \sum_{k=2}^{K} \Bigg(B_k + \sum_{s_j \in S_1} \frac{u_1}{u_k} \frac{w'_{f(s_j)}}{w_{f(s_j)}} R_k(j) \Bigg). \tag{4}$$

We thus evaluate all $2n + 3$ possible outcomes of w' and choose the one that gives the smallest value for the expectation of the Augmentation Step cost plus the upper bound on the cost of the remaining stages.

Finally, we can deterministically choose S_2, by iterating through the vertices s in S_1 and computing the conditional expectation of the Augmentation Step plus the future stages conditioned on including/not including s in S_2. For each $v \in D_1 \cup \{t\}$, we let $A_v \subset S(v)$ be the set of sources that we have already chosen to be included in $S'(v)$ and $B_v \subset S(v)$ the sources that we have chosen not to include in $S'(v)$. Initially, $A_v = B_v = \emptyset$ for all $v \in D_1 \cup \{t\}$. We iterate through the sources, and compute the expected cost of the type $k + 1$ cables in the Augmentation Step plus the expected upper bound on the cost of the remaining stages if we add s to $A_{f(s)}$ or $B_{f(s)}$, and add s to the set that gives the smaller expected total cost. By the definition of conditional expectation, this does not increase the expected total cost. Since $S(v)$ contains $\frac{w_v}{u_1}$ sources and we choose $\frac{w'_v}{u_2}$ of these for $S'(v)$, we can compute the conditional probability that $s \in \mathbb{S}_2$ if $s \notin A_{f(s)} \cup B_{f(s)}$ as

$$\mathbb{P}[s \in \mathbb{S}_2 | \mathbb{D}_1 = D_1, \{A_v, B_v\}_{v \in D_1 \cup \{t\}}, \mathbf{w}' = w'] = \frac{w'_{f(s)}/u_2 - |A_{f(s)}|}{w_{f(s)}/u_1 - |A_{f(s)} \cup B_{f(s)}|}.$$

Hence the conditional expected cost of the future stages is

$$\sum_{k=2}^{K} \left(B_k + \sum_{s_j \in S_1, s_j \in A_{f(s_j)}} \frac{u_2}{u_k} R_k(j) \right.$$

$$\left. + \sum_{s_j \in S_1, s_j \notin A_{f(s_j)} \cup B_{f(s_j)}} \frac{u_2}{u_k} \frac{w'_{f(s_j)}/u_2 - |A_{f(s_j)}|}{w_{f(s_j)}/u_1 - |A_{f(s_j)} \cup B_{f(s_j)}|} R_k(j) \right),$$

and the conditional expected cost of the type 2 cables installed in the Augmentation Step is

$$\sum_{s_j \in S_1, s_j \in A_{f(s_j)}} \sigma_2 \ell(s_j, f(s_j))$$

$$+ \sum_{s_j \in S_1, s_j \notin A_{f(s_j)} \cup B_{f(s_j)}} \frac{w'_{f(s_j)}/u_2 - |A_{f(s_j)}|}{w_{f(s_j)}/u_1 - |A_{f(s_j)} \cup B_{f(s_j)}|} \sigma_2 \ell(s_j, f(s_j)).$$

We have shown how to derandomize the preprocessing step and the first stage of SSBaB-Sample-Augment, without increasing the upper bound on the expected cost. We can use the same approach to iterate through the stages $2, \ldots, K-1$. We thus obtain the following result.

Theorem 26 *There exists a deterministic* 80*-approximation algorithm for the single sink buy-at-bulk problem.*

4.2 Proofs of Lemmas 21, 22 and 25

Proof of Lemma 21 Similar to Lemma 5.2 in [18], we can prove this by induction. If $s \notin S_k$, then it is clear that $s \notin S_{k+1}$, and hence $s \notin S_\ell$ for any $\ell > k$.

Let $s \in S_k$, and suppose the set of marked terminals in the Sampling Step is D_k, and let $f(s)$, w_v, $\mathbf{w}'_v$ and $S(v)$ be defined as in the description of the k-th stage of the algorithm. Given $\mathbf{w}'_v = w'_v$, the probability that s is in $\mathbb{S}_{k+1}$ is $\frac{w'_v/u_{k+1}}{w_v/u_k} = \frac{w'_v}{w_v}\frac{u_k}{u_{k+1}}$. By Lemma 20, $\mathbf{w}'_v = \underline{w}_v + u_{k+1}$ with probability $\frac{w_v - \underline{w}_v}{u_{k+1}}$ and $\mathbf{w}'_v = \underline{w}_v$ otherwise. Hence

$$\begin{aligned}
&\mathbb{P}[s \in \mathbb{S}_{k+1} | \mathbb{S}_k = S_k, \mathbb{D}_k = D_k, f(s) = v] \\
&= \frac{\underline{w}_v}{w_v}\frac{u_k}{u_{k+1}} + \left(\frac{\underline{w}_v + u_{k+1}}{w_v}\frac{u_k}{u_{k+1}} - \frac{\underline{w}_v}{w_v}\frac{u_k}{u_{k+1}} \right) \frac{w_v - \underline{w}_v}{u_{k+1}} \\
&= \frac{u_k}{u_{k+1}}. \qquad (5)
\end{aligned}$$

Since the right hand side is the same for any D_k and $f(s)$, it will also hold unconditionally that $\mathbb{P}[s \in \mathbb{S}_{k+1} | \mathbb{S}_k = S_k] = \frac{u_k}{u_{k+1}}$.

Now, let $s \in S_k$ and suppose the lemma holds for some $\ell > k$. Note that $\mathbb{P}[s \in \mathbb{S}_{\ell+1} | \mathbb{S}_k = S_k] = \sum_{S_\ell : s \in S_\ell} \mathbb{P}[s \in \mathbb{S}_{\ell+1} | \mathbb{S}_\ell = S_\ell] \mathbb{P}[\mathbb{S}_\ell = S_\ell | \mathbb{S}_k = S_k]$, since if $s \notin S_\ell$, then $\mathbb{P}[s \in \mathbb{S}_{\ell+1} | \mathbb{S}_\ell = S_\ell] = 0$. Now, from the (base case of the) induction, we know that $\mathbb{P}[s \in \mathbb{S}_{\ell+1} | \mathbb{S}_\ell = S_\ell] = \frac{u_\ell}{u_{\ell+1}}$ for all S_ℓ such that $s \in S_\ell$. Hence $\sum_{S_\ell : s \in S_\ell} \mathbb{P}[s \in \mathbb{S}_{\ell+1} | \mathbb{S}_\ell = S_\ell] \mathbb{P}[\mathbb{S}_\ell = S_\ell | \mathbb{S}_k = S_k] = \frac{u_\ell}{u_{\ell+1}} \sum_{S_\ell : s \in S_\ell} \mathbb{P}[\mathbb{S}_\ell = S_\ell | \mathbb{S}_k = S_k] = \frac{u_\ell}{u_{\ell+1}} \mathbb{P}[s \in \mathbb{S}_\ell | \mathbb{S}_k = S_k]$. Using the induction hypothesis, we have that $\mathbb{P}[s \in \mathbb{S}_\ell | \mathbb{S}_k = S_k] = \frac{u_k}{u_\ell}$, and hence $\mathbb{P}[s \in \mathbb{S}_{\ell+1} | \mathbb{S}_k = S_k] = \frac{u_\ell}{u_{\ell+1}} \frac{u_k}{u_\ell} = \frac{u_k}{u_{\ell+1}}$. □

We note that the (joint) distribution of $\mathbb{S}_\ell$ *does* indeed depend on the sampling probabilities in stage k, but that by Lemma 21 the *marginal* probability $\mathbb{P}[s \in \mathbb{S}_\ell | \mathbb{S}_k = S_k]$ does not depend on it. Because of the special form of the upper bounds and the linearity of expectation, we do not need to know the joint distribution in order to compute the expectation of the upper bound.

We will now first give the proof of Lemma 25 and then give the proof of Lemma 22.

Proof of Lemma 25 Given set S_k, and sampling probabilities p^k, let $\mathbb{E}_{p^k}[\ell(s, \mathbb{D}_k \cup \{t\})]$ be the expectation of the distance from $s \in S_k$ to the closest terminal in $D_k \cup \{t\}$. By Lemma 21, the expected cost of the Augmentation Step of the k-th stage is

$$\sum_{s \in S_k} \mathbb{E}_{p^k}[\ell(s, \mathbb{D}_k \cup \{t\})] \left(\sigma_k + \sigma_{k+1} \frac{u_k}{u_{k+1}} \right),$$

which does not depend on the outcome of the Subproblem Step and can be efficiently computed for any sampling probabilities p^k.

For $k = 1, \dots, K-1$, the Sub-LP$_k(D_k)$ is the linear programming relaxation of the Steiner tree problem on $D_k \cup \{t\}$:

$$\begin{aligned} & \min && \sum_{e \in E} \sigma_{k+1} c_e y_e \\ \text{Sub-LP}_k(D_k) \quad & \text{s.t.} && \sum_{e \in \delta(S)} y_e \geq 1 \quad \text{for all } S \subset V, t \notin S, D_k \cap S \neq \emptyset, \\ & && y_e \geq 0 \quad \forall e \in E. \end{aligned}$$

By [11], we know that Sub-LP$_k(D_k)$ satisfies the second condition of the lemma.

Similar to our previous approach, we will use a linear programming relaxation of the original problem, to show that the third condition of Lemma 25 is satisfied.

Let $\{u_k\}_{k=1,\dots,K}$, $\{\sigma_k\}_{k=1,\dots,K}$ be the *rounded* cable capacities and costs that we obtain after executing the Rounding Step of SSBaB-Sample-Augment. Let z_e^k indicate whether we install a cable of type k on edge e. Let $x_e^{j,k}$ indicate the amount of flow sent from s_j to t that passes through a cable of type k on edge e

$$\text{(SSBaB-LP)}\quad \begin{aligned} \min \quad & \frac{1}{4}\sum_{e\in E}\sum_{k=1}^{K}\sigma_k c_e z_e^k \\ \text{s.t.} \quad & \sum_{e\in\delta(S)}\sum_{k=1}^{K} x_e^{j,k} \geq w_j \quad \text{for all } S\subset V, j=1,\ldots,n: s_j\in S, t\notin S, \\ & \sum_{j=1}^{n} x_e^{j,k} \leq u_k z_e^k \quad \text{for all } e\in E, k=1,\ldots,K, \\ & \frac{x_e^{j,k}}{w_j} \leq z_e^k \quad \text{for all } e\in E, k=1,\ldots,K, j=1,\ldots,n, \\ & x_e^{j,k}\geq 0, z_e^k \geq 0 \quad \text{for all } e\in E, k=1,\ldots,K, j=1,\ldots,n. \end{aligned}$$

The first set of constraints enforces that s_j sends w_j units to t by enforcing that at least w_j units cross any cut that separates s_j from t. The second and third sets of constraints ensure that there is enough capacity installed to support the flow. In particular, the second set of constraints ensures that we install a sufficient number of cables of type k on edge e to support the flow sent on edge e on a type k cable jointly by all sources. The third set of constraints is implied by the second set of constraints if we constrain z_e^k to be integer, but strengthens the LP relaxation by enforcing that if source s_j sends a $\frac{x_e^{j,k}}{w_j}$ fraction of its flow on a cable of type k on edge e, then there should be at least a $\frac{x_e^{j,k}}{w_j}$ fraction of a type k cable installed on the edge.

To see that SSBaB-LP is a relaxation of the single sink buy-at-bulk problem, consider an optimal solution to the single sink buy-at-bulk problem. We let $x_e^{j,k}$ be the amount of flow from s_j to t that is routed on a cable of type k on edge e. Because we rounded down the cable capacities to the nearest power of 2 to write SSBaB-LP, we need to let $z_e^k = \mathbf{2}$ for every cable of type k that is installed on edge e in this optimal solution. Since we also rounded up the cable costs to the nearest power of 2, we get that $\sum_{e\in E}\sum_{k=1}^{K}\sigma_k c_e z_e^k \leq 4OPT$.

Let $\hat{x}, \hat{z}$ be an optimal solution to SSBaB-LP. Now, for any stage $k=1,\ldots,K-1$ we define

$$b_e^k = \sum_{\ell=k+1}^{K} \hat{z}_e^\ell, \tag{6}$$

$$r_e^k(j) = \sum_{\ell=1}^{k} \frac{\hat{x}_e^{j,\ell}}{w_j}. \tag{7}$$

We let $y_e^k(D_k) = b_e^k + \sum_{s_j\in D_k} r_e^k(j)$ for any $D_k \subseteq S_k$. We need to show that $y^k(D_k)$ is feasible for Sub-LP$_k(D_k)$ for $k=1,\ldots,K-1$.

Note that $\frac{\hat{x}_e^{j,\ell}}{w_j} \leq \hat{z}_e^\ell$, hence for any $S \subset V$ such that $t \in S, s_j \in D_k \backslash S$, we have that $\sum_{e\in\delta(S)} y_e^k(D_k) \geq \sum_{e\in\delta(S)}(\sum_{\ell=k+1}^K \hat{z}_e^\ell + \sum_{\ell=1}^k \frac{\hat{x}_e^{j,\ell}}{w_j}) \geq \sum_{e\in\delta(S)} \sum_{\ell=1}^K \frac{\hat{x}_e^{j,\ell}}{w_j} \geq 1$. Hence $y^k(D_k)$ is a feasible solution to the Sub-LP$_k(D_k)$.

Finally, we will choose $B_k, R_k(j)$ for $j = 1, \dots, n$ so that condition (iv) is satisfied. We will then show in the proof of Lemma 22 that this choice will also satisfy Lemma 22. Let $C_{MST}(D_k)$ be the cost (with respect to the edge costs c_e) of a minimum cost spanning tree on $D_k \cup \{t\}$. Let $\mathbb{E}_{\hat{p}^k}[C_{MST}(\mathbb{D}_k)|\mathbb{S}_k = S_k]$ be the expectation of $C_{MST}(D_k)$ if the sampling probabilities are given by $\hat{p}_j^k = \frac{\sigma_k}{\sigma_{k+1}}$ for $s_j \in S_k$, and $\hat{p}_j^k = 0$ if $s_j \notin S_k$. Similarly let $\mathbb{E}_{\hat{p}^k}[\sum_{e\in E} c_e y_e^k(\mathbb{D}_k)|\mathbb{S}_k = S_k]$ be the expectation of $\sum_{e\in E} c_e y_e^k(\mathbb{D}_k)$. Since $y^k(D_k)$ is feasible for Sub-LP$_k(D_k)$, we know from Goemans and Bertsimas [11] that $C_{MST}(D_k) \leq 2\sum_{e\in E} c_e y_e^k(D_k)$ for every $D_k \subset \mathcal{D}$.

It follows from Lemma 5.4 in [18] that for $k = 1, \dots, K-1$, the expected cost of the type k cables installed in the Augmentation Step of stage k, given S_k and sampling probabilities $\hat{p}^k$ can be bounded by

$$\left(1 - \frac{\sigma_k}{\sigma_{k+1}}\right)\sigma_{k+1}\mathbb{E}_{\hat{p}^k}[C_{MST}(\mathbb{D}_k)|\mathbb{S}_k = S_k].$$

Since Gupta et al. [18] do not derive the above bound explicitly, we give an explicit proof here. Let $X_j(D_k)$ be the cost of connecting s_j to the closest terminal in $D_k \cup \{t\}$ using type k cables. For a given set D_k, consider the minimum cost spanning tree on $D_k \cup \{t\}$ rooted at t, and for $s_j \in D_k$, let $Y_j(D_k)$ be the length of the edge connecting s_j to its parent in this tree, and let $Y_j(D_k) = 0$ if $s_j \notin D_k$. Then $\mathbb{E}_{\hat{p}^k}[\sum_{s_j\in S_k} Y_j(\mathbb{D}_k)|\mathbb{S}_k = S_k] = \mathbb{E}_{\hat{p}^k}[C_{MST}(\mathbb{D}_k)|\mathbb{S}_k = S_k]$. Also

$$\mathbb{E}_{\hat{p}^k}\Bigg[\sum_{s_j\in S_k} X_j(\mathbb{D}_k)\big|\mathbb{S}_k = S_k\Bigg]$$

$$= \sigma_k \sum_{s_j\in S_k} \mathbb{E}_{\hat{p}^k}[\ell(s_j, \mathbb{D}_k \cup \{t\})|\mathbb{S}_k = S_k] \tag{8}$$

$$= \sigma_k \sum_{s_j\in S_k} (1 - \hat{p}_j^k)\mathbb{E}_{\hat{p}^k}[\ell(s_j, \mathbb{D}_k \cup \{t\})|\mathbb{S}_k = S_k, s_j \notin \mathbb{D}_k] \tag{9}$$

$$= \sigma_k \sum_{s_j\in S_k} (1 - \hat{p}_j^k)\mathbb{E}_{\hat{p}^k}[\ell(s_j, \mathbb{D}_k\backslash\{s_j\} \cup \{t\})|\mathbb{S}_k = S_k, s_j \in \mathbb{D}_k] \tag{10}$$

$$\leq \sigma_k \sum_{s_j\in S_k} (1 - \hat{p}_j^k)\mathbb{E}_{\hat{p}^k}[Y_j(\mathbb{D}_k)|\mathbb{S}_k = S_k, s_j \in \mathbb{D}_k] \tag{11}$$

$$= \sigma_k \sum_{s_j\in S_k} (1 - \hat{p}_j^k)\frac{1}{\hat{p}_j^k}\mathbb{E}_{\hat{p}^k}[Y_j(\mathbb{D}_k)|\mathbb{S}_k = S_k] \tag{12}$$

$$= \left(1 - \frac{\sigma_k}{\sigma_{k+1}}\right)\sigma_{k+1}\mathbb{E}_{\hat{p}^k}[C_{MST}(\mathbb{D}_k)|\mathbb{S}_k = S_k]. \tag{13}$$

Equation (8) follows from the definition of $X_j(D_k)$, and (9) follows since $X_j(D_k) = 0$ if $s_j \in D_k$. Equation (10) follows because the sources are marked independently, so that for $s_j \in S_k$, $\mathbb{E}_{\hat{p}^k}[\ell(s_j, \mathbb{D}_k \cup \{t\})|\mathbb{S}_k = S_k, s_j \notin \mathbb{D}_k] = \mathbb{E}_{\hat{p}^k}[\ell(s_j, \mathbb{D}_k \backslash \{s_j\} \cup \{t\})|\mathbb{S}_k = S_k, s_j \in \mathbb{D}_k]$. Inequality (11) follows since by the definition $Y_j(D_k) \geq \ell(s_j, D_k \backslash \{s_j\} \cup \{t\})$, and (12) follows since $Y_j(D_k) = 0$ if $s_j \notin D_k$. Finally, (13) follows by substituting $\hat{p}_j^k = \frac{\sigma_k}{\sigma_{k+1}}$ and $\mathbb{E}_{\hat{p}^k}[\sum_{s_j \in S_k} Y_j(\mathbb{D}_k)|\mathbb{S}_k = S_k] = \mathbb{E}_{\hat{p}^k}[C_{MST}(\mathbb{D}_k)|\mathbb{S}_k = S_k]$.

Also, by Lemma 21, each path on which type k cables are installed in the Augmentation Step of stage k will also have type $k+1$ cables installed in the Augmentation Step with probability $\frac{u_k}{u_{k+1}}$. Hence the expected cost of the type $k+1$ cables installed in the Augmentation Step is $\frac{\sigma_{k+1}/u_{k+1}}{\sigma_k/u_k}$ times the expected cost of the type k cables installed in the Augmentation Step. The total expected cost of the Augmentation Step of stage k, given S_k and sampling probabilities $\hat{p}^k$ can thus be bounded by

$$\left(1 + \frac{\sigma_{k+1}/u_{k+1}}{\sigma_k/u_k}\right)\left(1 - \frac{\sigma_k}{\sigma_{k+1}}\right)\sigma_{k+1}\mathbb{E}_{\hat{p}^k}[C_{MST}(\mathbb{D}_k)|\mathbb{S}_k = S_k]. \tag{14}$$

Now, since by our assumptions $1 - \frac{\sigma_k}{\sigma_{k+1}} \leq 1$, $\frac{\sigma_{k+1}/u_{k+1}}{\sigma_k/u_k} \leq \frac{1}{2}$, and because by Goemans and Bertsimas [11] $\mathbb{E}_{\hat{p}^k}[C_{MST}(\mathbb{D}_k)|\mathbb{S}_k = S_k] \leq 2\mathbb{E}_{\hat{p}^k}[\sum_{e \in E} c_e y_e^k(\mathbb{D}_k)|\mathbb{S}_k = S_k]$ we get that the expected cost of the Augmentation Step of stage k, given S_k and sampling probabilities $\hat{p}^k$ is at most

$$3\sigma_{k+1}\mathbb{E}_{\hat{p}^k}\left[\sum_{e \in E} c_e y_e^k(\mathbb{D}_k)\big|\mathbb{S}_k = S_k\right]. \tag{15}$$

Because the cost of the solution created in the Subproblem Step on D_k is at most $2\sum_{e \in E} \sigma_{k+1} c_e y_e^k(D_k)$ for any $D_k \subseteq \mathcal{D}$, we get that, given $\mathbb{S}_k = S_k$, the total expected cost of stage k is at most $5\sigma_{k+1}\mathbb{E}_{\hat{p}^k}[\sum_{e \in E} c_e y_e^k(\mathbb{D}_k)|\mathbb{S}_k = S_k]$. Now, noting that $\mathbb{E}_{\hat{p}^k}[y_e^k(\mathbb{D})|\mathbb{S}_k = S_k] = b_e^k + \sum_{s_j \in S_k} \hat{p}_j^k r_e^k(j) = b_e^k + \sum_{s_j \in S_k} \frac{\sigma_k}{\sigma_{k+1}} r_e^k(j)$, it follows that condition (iv) of Lemma 25 is satisfied for $k = 1, \ldots, K-1$ if we let

$$B_k = 5\sum_{e \in E} \sigma_{k+1} c_e b_e^k, \tag{16}$$

$$R_k(j) = 5\sum_{e \in E} \sigma_k c_e r_e^k. \tag{17}$$

□

Finally, we show that we can define B_0, B_K and $R_0(j)$, $R_K(j)$ for $j = 1, \ldots, n$ so that these values combined with the values defined in (16, 17) satisfy the conditions in Lemma 22.

Proof of Lemma 22 We already saw that B_k, $R_k(j)$ as defined in (16, 17) satisfy the first inequality in Lemma 22 for $k = 1, \ldots, K-1$.

We now define $b_e^0, r_e^K(j)$ according to the definition in (6), (7) (where we note that these definitions set $b_e^K = 0, r_e^0(j) = 0$). We claim that b_e^0 is a feasible solution to the Steiner LP on $\mathcal{S} \cup \{t\}$. Indeed consider some $s_j \in \mathcal{S}$ and take $S \subset V$ such that $s_j \notin S, t \in S$. By the third set of constraints of SSBaB-LP, $b_e^0 = \sum_{\ell=1}^K \hat{z}_e^\ell \geq \sum_{\ell=1}^K \frac{\hat{x}_e^\ell}{w_j}$. Also, the first set of constraints ensures that $\sum_{e\in\delta(S)} \sum_{\ell=1}^K \frac{\hat{x}_e^\ell}{w_j} \geq 1$. Hence by [11], we know that we can find a tree T_0 on $\mathcal{S} \cup \{t\}$ and install cables of type 1, at cost at most $2\sum_{e\in E} \sigma_1 c_e b_e^0$.

Hence we can define

$$B_0 = 2\sum_{e\in E} \sigma_1 c_e b_e^0, \quad \text{and} \quad R_0(j) = 0, j = 1, \ldots, n. \tag{18}$$

For the final stage, we need to give a bound on $\sum_{s_j\in S_K} \sigma_K \ell(s_j, t)$. From the first set of constraints of SSBaB-LP, it is clear that $\sum_{e\in E} c_e \sum_{\ell=1}^K \frac{\hat{x}_e^{j,\ell}}{w_j}$ is an upper bound on the length of the shortest path from s_j to t, hence $\sum_{s_j\in S_K} \sigma_K \ell(s_j, t) \leq \sum_{s_j\in S_K} \sigma_K \sum_{e\in E} c_e \sum_{\ell=1}^K \frac{\hat{x}_e^{j,\ell}}{w_j} = \sum_{e\in E} \sigma_K c_e \sum_{s_j\in S_K} r_e^K(j)$. So we can define

$$B_K = 0, \quad \text{and} \quad R_K(j) = \sum_{e\in E} \sigma_K c_e r_e^K(j), j = 1, \ldots, n. \tag{19}$$

Now, some algebra similar to that in Gupta et al. [18] shows that the second inequality of the lemma holds. For ease of notation, we first write B_K, B_0 and $R_K(j), R_0(j)$ in the same form as $B_k, R_k(j)$ for $1 \leq k \leq K-1$. To do this, we define $\sigma_0 = \frac{1}{2}\sigma_1, u_0 = 1, \sigma_{K+1} = 2\sigma_K, u_{K+1} = u_K$, and increase our upper bounds $B_0, R_0(j), B_K, R_K(j)$ so that for every $k \in \{0, \ldots, K\}$ and $j \in \{1, \ldots, n\}$:

$$B_k = 5\sum_{e\in E} \sigma_{k+1} c_e b_e^k, \quad \text{and} \quad R_k(j) = 5\sum_{e\in E} \sigma_k c_e r_e^k.$$

(Note that we could have chosen any finite values for σ_0, σ_{K+1}, since b_e^K and r_e^0 are 0.)

Then for $k = 1, \ldots, K$,

$$\begin{aligned}
\sum_{s_j\in\mathcal{S}} \frac{w_j}{u_k} R_k(j) &= 5\sum_{e\in E} \sigma_k c_e \sum_{s_j\in\mathcal{S}} \frac{w_j}{u_k} r_e^k(j) \\
&= 5\frac{\sigma_k}{u_k} \sum_{e\in E} c_e \sum_{s_j\in\mathcal{S}} \sum_{\ell=1}^k \hat{x}_e^{j,\ell} \\
&\leq 5\frac{\sigma_k}{u_k} \sum_{e\in E} c_e \sum_{\ell=1}^k u_\ell \hat{z}_e^\ell \quad \text{by the second set of constraints of SSBaB-LP.}
\end{aligned}$$

Also, note that $R_0(j) = 0$, hence $\sum_{s_j\in\mathcal{S}} R_0(j) = 0 \leq 5\frac{\sigma_0}{u_0} \sum_{e\in E} c_e \sum_{\ell=1}^0 u_\ell \hat{z}_e^\ell$.

Now, since $B_k = 5\sum_{e\in E}\sigma_{k+1}c_e b_e^k = 5\sum_{e\in E}\sigma_{k+1}c_e\sum_{\ell=k+1}^{K}\hat{z}_e^\ell$ for $k = 0,\ldots,K$, we get that

$$
\begin{aligned}
&\sum_{k=0}^{K} B_k + \sum_{s_j\in\mathcal{S}}\left(R_0(j) + \sum_{k=1}^{K}\frac{w_j}{u_k}R_k(j)\right)\\
&\quad\le 5\sum_{k=0}^{K}\sum_{e\in E}c_e\left(\sum_{\ell=k+1}^{K}\sigma_{k+1}\hat{z}_e^\ell + \sum_{\ell=1}^{k}\sigma_k\frac{u_\ell}{u_k}\hat{z}_e^\ell\right)\\
&\quad= 5\sum_{\ell=1}^{K}\sum_{e\in E}c_e\hat{z}_e^\ell\left(\sum_{k=\ell}^{K}\sigma_k\frac{u_\ell}{u_k} + \sum_{k=0}^{\ell-1}\sigma_{k+1}\right)\\
&\quad= 5\sum_{\ell=1}^{K}\sum_{e\in E}\sigma_\ell c_e\hat{z}_e^\ell\left(\sum_{k=\ell}^{K}\frac{\sigma_k/u_k}{\sigma_\ell/u_\ell} + \sum_{k=1}^{\ell}\frac{\sigma_k}{\sigma_\ell}\right)\\
&\quad= 5\sum_{\ell=1}^{K}\sum_{e\in E}\sigma_\ell c_e\hat{z}_e^\ell\left(\sum_{k=0}^{K-\ell}\frac{\sigma_{\ell+k}/u_{\ell+k}}{\sigma_\ell/u_\ell} + \sum_{k=0}^{\ell-1}\frac{\sigma_{\ell-k}}{\sigma_\ell}\right)\\
&\quad\le 20\sum_{\ell=1}^{K}\sum_{e\in E}\sigma_\ell c_e\hat{z}_e^\ell\\
&\quad\le 80\mathit{OPT}.
\end{aligned}
$$

The second to last inequality follows from the fact that $\frac{\sigma_k}{\sigma_{k+1}} \le \frac{1}{2}$ and $\frac{\sigma_{k+1}/u_{k+1}}{\sigma_k/u_k} \le \frac{1}{2}$ for $k = 0,\ldots,K-1$, since then $\sum_{k=0}^{K-\ell}\frac{\sigma_{\ell+k}/u_{\ell+k}}{\sigma_\ell/u_\ell} \le \sum_{k=0}^{K-\ell}(\frac{1}{2})^k \le 2$ and $\sum_{k=0}^{\ell-1}\frac{\sigma_{\ell-k}}{\sigma_\ell} \le \sum_{k=0}^{\ell-1}(\frac{1}{2})^k \le 2$. □

4.3 Improved Analysis

We can improve the approximation guarantee of the algorithm using the ideas of Grandoni and Italiano [13]. Rather than rounding down the cable capacities and rounding up the prices, they carefully select a subset of the cables, which allows them to significantly improve the approximation ratio. They require the cable types to satisfy economies of scale: the cost $\frac{\sigma_k}{u_k}$ per unit capacity must decrease as the capacity of the cable increases. It turns out that we can drop this assumption; see Remark 28 at the end of this section.

It is not hard to adapt our upper bounds to their algorithm and analysis, instead of the algorithm of Gupta et al. [18]. We give the revised algorithm in Fig. 9. We replaced the Rounding Step by a Cable Selection Step, which takes two parameters α and β. We note that in the Intermediate Stages, the algorithm SBBaB-stage-$i(q)$ for $i(q) = k$ is the same as SSBaB-stage-k given in Fig. 8, except that we replace $k+1$ by $i(q+1)$ (in other words, we only use the cable types $i(1),\ldots,i(K')$). If we use $\alpha = 3.059$, $\beta = 2.475$ in the algorithm given in Fig. 9, we get the following result.

SSBaB-Sample-Augment$(G=(V,E),c,t,\mathcal{S},u_1,\ldots,u_K,\sigma_1,\ldots,\sigma_K)$
(*Cable Selection Step*) Let $i(1)=1$, and given $i(q)$, $1<i(q)<K$, let $i(q+1)$ be the smallest index such that $\sigma_{i(q+1)+1}\geq\alpha\sigma_{i(q)}$ and $\frac{\sigma_{i(q)}}{u_{i(q)}}\geq\beta\frac{\sigma_{i(q+1)}}{u_{i(q+1)}}$, or, if no such index exists, $i(q+1)=i(K')=K$. (*Preprocessing Step*) Find a Steiner tree T_0 on $\mathcal{S}\cup\{t\}$, install a cable of type $i(1)$ on each edge of T_0. Let $w'=\text{Redistribute}(T_0,w,u_{i(1)})$. Let $S_{i(1)}=\{s_j\in\mathcal{S}:w'_j>0\}$. (*Intermediate Stages*) For $q=1,\ldots,K'-1$ $\hat{p}_j^{i(q)}=\frac{\sigma_{i(q)}}{\sigma_{i(q+1)}}$ for $s_j\in S_{i(q)}$, $S_{i(q+1)}=\text{SSBaB-Stage-}i(q)\left(G,c,t,S_{i(q)},u_{i(q)},u_{i(q+1)},\hat{p}^{i(q)}\right).$ (*Final Stage*) For every $s\in S_{i(K')}$, install cables of type $i(K')$ on the shortest path from s to t.

Fig. 9 Sample-Augment algorithm for single-sink buy-at-bulk with generalized cable selection rule

Theorem 27 *There exists a deterministic* 27.72*-approximation algorithm for the single sink buy-at-bulk problem.*

Proof We adapt and strengthen our previous analysis; in particular, we slightly change the definitions of $b^k, r^k(j)$ and $B_k, R_k(j)$ that we used in the proofs of Lemmas 22 and 25. Since we now have stages $i(0),\ldots,i(K')$, we define values $b^{i(q)}, r^{i(q)}(j)$ and $B_{i(q)}, R_{i(q)}(j)$. We show that $b^{i(q)}, r^{i(q)}$ satisfy Lemma 25, and that if $\hat{p}_j^{i(q)}=\frac{\sigma_{i(q)}}{\sigma_{i(q+1)}}$ for $s_j\in S_{i(q)}$, then the total expected cost of the cables installed in stage $i(q)$ is at most $B_{i(q)}+\sum_{s_j\in S_{i(q)}}R_{i(q)}(j)$. This part of the analysis is basically the same as before. The main difference occurs in Lemma 22. We will show that if $\alpha=3.059$, $\beta=2.475$, then

$$B_{i(0)}+\sum_{s_j\in\mathcal{S}}R_{i(0)}(j)+\sum_{q=1}^{K'}\left(B_{i(q)}+\sum_{s_j\in\mathcal{S}}\frac{w_j}{u_{i(q)}}R_{i(q)}(j)\right)\leq 27.72 OPT.$$

We again define $b^{i(q)}, r^{i(q)}(j)$ and $B_{i(q)}, R_{i(q)}(j)$ based on an LP relaxation of the single sink buy-at-bulk problem. We use a slightly different LP relaxation. Recall that previously, SSBaB-LP was defined using the rounded cable capacities and cable costs. We now redefine SSBaB-LP on the original parameters, so that we can let the objective value be

$$\min\sum_{e\in E}\sum_{k=1}^{K}\sigma_k c_e z_e^k.$$

In other words, if we define the LP using the original cable parameters, we do not have to divide the objective value by 4 in order to have SSBaB-LP be a relaxation of the single sink buy-at-bulk problem.

Now, since we assume that $\frac{\sigma_{k+1}}{u_{k+1}} \le \frac{\sigma_k}{u_k}$ and $\sigma_k \le \sigma_{k+1}$ for all $k = 1, \dots, K-1$, it is the case that for every fixed q, $\frac{\sigma_{i(q)}/u_{i(q)}}{\sigma_\ell/u_\ell}$ is increasing in ℓ and $\frac{\sigma_{i(q+1)}}{\sigma_\ell}$ is decreasing in ℓ. Hence there exists some $\ell(q)$ such that $\frac{\sigma_{i(q)}/u_{i(q)}}{\sigma_\ell/u_\ell} \le \frac{\sigma_{i(q+1)}}{\sigma_\ell}$ for $\ell \le \ell(q)$, and $\frac{\sigma_{i(q)}/u_{i(q)}}{\sigma_\ell/u_\ell} \ge \frac{\sigma_{i(q+1)}}{\sigma_\ell}$ for $\ell > \ell(q)$. Given this definition of $\ell(q)$ we define

$$r_e^{i(q)}(j) = \sum_{\ell=1}^{\ell(q)} \frac{\hat{x}_e^{j,\ell}}{w_j}, \tag{20}$$

$$b_e^{i(q)} = \sum_{\ell=\ell(q)+1}^{K} \hat{z}_e^\ell, \tag{21}$$

where $\hat{z}, \hat{x}$ is the optimal solution to the SSBaB-LP.

Then it is not hard to show that Lemma 25 holds for stages $i(1), \dots, i(K'-1)$, if we define the values $B_{i(q)}, R_{i(q)}(j)$ to be

$$B_{i(q)} = \gamma_{i(q)} \sum_{e\in E} \sigma_{i(q+1)} c_e b_e^{i(q)}, \qquad R_{i(q)}(j) = \gamma_{i(q)} \sum_{e\in E} \sigma_{i(q)} c_e r_e^{i(q)},$$

for appropriately chosen values of $\gamma_{i(q)}$. Instead of setting $\gamma_{i(q)} = 5$ as we did before, we refine the analysis and set $\gamma_{i(q)} = 2(1 + \frac{\sigma_{i(q+1)}/u_{i(q+1)}}{\sigma_{i(q)}/u_{i(q)}})(1 - \frac{\sigma_{i(q)}}{\sigma_{i(q+1)}}) + 2$. From (14) and the arguments following it we see that this setting for $\gamma_{i(q)}$ rather than 5 is sufficient for Lemma 25 to hold for stage $i(q)$.

We refer to the preprocessing step as stage $i(0)$ and the final stage as $i(K')$ and we define

$$B_{i(0)} = 2\sum_{e\in E} \sigma_{i(1)} c_e \sum_{\ell=1}^{K} \hat{z}_e^\ell b_e^{i(0)}, \qquad R_{i(0)}(j) = 0,$$

$$B_{i(K')} = 0, \qquad R_{i(K')}(j) = \sum_{e\in E} \sigma_{i(K')} c_e \sum_{\ell=1}^{K} \frac{\hat{x}_e^{j,\ell}}{w_j},$$

which is the same as (18) and (19). If we let $\gamma_{i(0)} = 2, \gamma_{i(K')} = 1$ and $\ell(0) = 0, \ell(K') = K$, then we can write $B_{(i(q)}, R_{i(q)}$ in a single format for all $q = 0, \dots, K'$ as

$$B_{i(q)} = \gamma_{i(q)} \sum_{e\in E} \sigma_{i(q+1)} c_e \sum_{\ell=\ell(q)+1}^{K} \hat{z}_e^\ell,$$

$$R_{i(q)}(j) = \gamma_{i(q)} \sum_{e\in E} \sigma_{i(q)} c_e \sum_{\ell=1}^{\ell(q)} \frac{\hat{x}_e^{j,\ell}}{w_j}.$$

Now, we can follow a similar analysis as in the proof of Lemma 22. First, it follows from the same arguments as in the proof of Lemma 22 that for $q = 0, \dots, K'$,

$$\sum_{s_j \in \mathcal{S}} \frac{w_j}{u_{i(q)}} R_{i(q)}(j) \le \gamma_{i(q)} \frac{\sigma_{i(q)}}{u_{i(q)}} \sum_{e \in E} c_e \sum_{\ell=1}^{\ell(q)} u_\ell \hat{z}_e^\ell.$$

Therefore

$$\begin{aligned}
&B_{i(0)} + \sum_{s_j \in \mathcal{S}} R_{i(0)}(j) + \sum_{q=1}^{K'} \left(B_{i(q)} + \sum_{s_j \in \mathcal{S}} \frac{w_j}{u_{i(q)}} R_{i(q)}(j) \right) \\
&\le \sum_{q=0}^{K'} \gamma_{i(q)} \sum_{e \in E} c_e \left(\sum_{\ell=1}^{\ell(q)} \sigma_{i(q)} \frac{u_\ell}{u_{i(q)}} \hat{z}_e^\ell + \sum_{\ell=\ell(q)+1}^{K} \sigma_{i(q+1)} \hat{z}_e^\ell \right) \\
&= \sum_{\ell=1}^{K} \sum_{e \in E} \sigma_\ell c_e \hat{z}_e^\ell \left(\sum_{q:\ell(q)<\ell} \gamma_{i(q)} \frac{\sigma_{i(q+1)}}{\sigma_\ell} + \sum_{q:\ell(q)\ge\ell} \gamma_{i(q)} \frac{\sigma_{i(q)}/u_{i(q)}}{\sigma_\ell/u_\ell} \right) \\
&= \sum_{\ell=1}^{K} \sum_{e \in E} \sigma_\ell c_e \hat{z}_e^\ell \Bigg(\gamma_{i(0)} \frac{\sigma_{i(1)}}{\sigma_\ell} + \sum_{q=1}^{K'-1} \gamma_{i(q)} \min\left\{ \frac{\sigma_{i(q)}/u_{i(q)}}{\sigma_\ell/u_\ell}, \frac{\sigma_{i(q+1)}}{\sigma_\ell} \right\} \\
&\quad + \gamma_{i(K')} \frac{\sigma_{i(K')}/u_{i(K')}}{\sigma_\ell/u_\ell} \Bigg),
\end{aligned}$$

where the first equality follows by changing the order of summation, and the second one follows from the definition of $\ell(q)$.

Plugging in $\gamma_{i(0)} = 2$, $\gamma_{i(K')} = 1$ and

$$\begin{aligned}
\gamma_{i(q)} &= 2 \left(1 + \frac{\sigma_{i(q+1)}/u_{i(q+1)}}{\sigma_{i(q)}/u_{i(q)}} \right) \left(1 - \frac{\sigma_{i(q)}}{\sigma_{i(q+1)}} \right) + 2 \\
&= \left(\left(2 + 2 \frac{\sigma_{i(q+1)}/u_{i(q+1)}}{\sigma_{i(q)}/u_{i(q)}} \right) \left(1 - \frac{\sigma_{i(q)}}{\sigma_{i(q+1)}} \right) + 2 \right)
\end{aligned}$$

for $q = 1, \dots, K' - 1$, and noting that $\frac{\sigma_{i(K')}/u_{i(K')}}{\sigma_\ell/u_\ell} \le 1$, we see that $\sum_{e \in E} \sigma_\ell c_e \hat{z}_e^\ell$ is thus charged at most

$$\begin{aligned}
&2 \frac{\sigma_{i(1)}}{\sigma_\ell} + \sum_{q=1}^{K'-1} \left(\left(2 + 2 \frac{\sigma_{i(q+1)}/u_{i(q+1)}}{\sigma_{i(q)}/u_{i(q)}} \right) \left(1 - \frac{\sigma_{i(q)}}{\sigma_{i(q+1)}} \right) + 2 \right) \\
&\quad \times \min\left\{ \frac{\sigma_{i(q+1)}}{\sigma_\ell}, \frac{\sigma_{i(q)}/u_{i(q)}}{\sigma_\ell/u_\ell} \right\} + 1
\end{aligned}$$

times. This expression is the same as the expression $apx(\rho, \ell)$ in Lemma 1 in [13], if we take $\rho = 2$:

$$apx(\rho, \ell) = 1 + \rho\frac{\sigma_{i(1)}}{\sigma_\ell} + \sum_{q=1}^{K'-1}\left(\left(2 + 2\frac{\sigma_{i(q+1)}/u_{i(q+1)}}{\sigma_{i(q)}/u_{i(q)}}\right)\left(1 - \frac{\sigma_{i(q)}}{\sigma_{i(q+1)}}\right) + \rho\right) \times \min\left\{\frac{\sigma_{i(q+1)}}{\sigma_\ell}, \frac{\sigma_{i(q)}/u_{i(q)}}{\sigma_\ell/u_\ell}\right\}.$$

Now, Grandoni and Italiano show how to bound $apx(\rho, \ell)$ if $i(1), \dots, i(K')$ are chosen according to the Cable Selection Step, i.e. if $i(1) = 1$, and if, given $i(q)$, $1 < i(q) < K$, $i(q+1)$ is the smallest index such that $\sigma_{i(q+1)+1} \geq \alpha\sigma_{i(q)}$ and $\frac{\sigma_{i(q)}}{u_{i(q)}} \geq \beta\frac{\sigma_{i(q+1)}}{u_{i(q+1)}}$, or, if no such index exists, $i(q+1) = i(K') = K$. Equations (7–12) in the proof of Theorem 2 in [13] state that if $\alpha > 1$, $\beta > 1$ and $\beta \leq 3 + \frac{\rho}{2}$ then

$$apx(\rho, \ell) \leq \max\begin{Bmatrix} 1 + (2 + \frac{2}{\beta} + \rho)(1 + \beta + \frac{\beta}{\beta-1}) + \rho\frac{2}{\alpha-1}, \\ 1 + (2 + \frac{2}{\beta} + \rho)(\alpha + \frac{\beta}{\beta-1}) + \rho\frac{\alpha+1}{\alpha-1}, \\ 1 + (2 + \frac{2}{\beta} + \rho)(1 + \frac{\beta}{\beta-1}) + \rho\frac{\alpha+1}{\alpha-1}, \\ 1 + (2 + \frac{2}{\beta} + \rho) + \rho\frac{2}{\alpha-1} + (4 + \rho)\beta, \\ 1 + \rho\frac{\alpha+1}{\alpha-1} + (4 + \rho)\alpha, \\ 1 + \rho\frac{\alpha+1}{\alpha-1} + (4 + \rho). \end{Bmatrix}$$

Hence if we set $\alpha = 3.059$, $\beta = 2.475$, then $apx(2, \ell) \leq 27.72$ for $\ell = 1, \dots, K$. Hence we have shown that

$$B_{i(0)} + \sum_{s_j \in \mathcal{S}} R_{i(0)}(j) + \sum_{q=1}^{K'}\left(B_{i(q)} + \sum_{s_j \in \mathcal{S}} \frac{w_j}{u_{i(q)}} R_{i(q)}(j)\right) \leq 27.72 \sum_{\ell=1}^{K}\sum_{e \in E} \sigma_\ell c_e \hat{z}_e^\ell \leq 27.72 OPT.$$

The arguments leading up to Theorem 26 show that these observations are enough to show that there exists a deterministic 27.72-approximation algorithm for the single sink rent-or-buy problem. □

Remark 28 We note that for our analysis, the assumption that the cables satisfy economies of scale is *without loss of generality*, because we use the optimal value of SSBaB-LP as an upper bound: if $\frac{\sigma_{k+1}}{u_{k+1}} > \frac{\sigma_k}{u_k}$, then the optimal solution to SSBaB-LP will not use cable type $k+1$, since it can instead use $\frac{u_{k+1}}{u_k}$ cables of type k which results in a lower cost. Grandoni and Italiano [13] use the (integer) optimum as an upper bound, and since $\frac{u_{k+1}}{u_k}$ may not be integer, they do need the additional assumption of economies of scale.

5 Conclusion

We propose a specific method for derandomizing Sample-Augment algorithms, and we successfully apply this method to all but one of the Sample-Augment algorithms in Gupta et al. [18], and to the a priori traveling salesman problem and the 2-stage rooted stochastic Steiner tree problem with independent decisions.

The question whether the Sample-Augment algorithm for multicommodity rent-or-buy problem can be derandomized remains open. The multicommodity rent-or-buy problem is a generalization of the single source rent-or-buy problem: instead of one source s and sinks $t_1, \ldots, t_k$, we are given k source-sink pairs $(s_1, t_1), \ldots, (s_k, t_k)$ and need to construct a network so that each source-sink pair is connected. The Sample-Augment algorithm for this problem [8, 18] marks each source-sink pair with probability $\frac{1}{M}$ and buys a Steiner forest on the marked terminals in the Subproblem Step. In the Augmentation Step, we contract the bought edges, and rent the shortest path connecting each terminal pair in the contracted graph. If we want to use Theorem 1, we would need to be able to compute $\mathbb{E}_p[C_{aug}(\mathbb{D})]$ (or a good upper bound for it) efficiently and it is unclear how to do this for the multicommodity rent-or-buy algorithm, because unlike in the algorithms we discussed here, $\mathbb{E}_p[C_{aug}(\mathbb{D})]$ does depend on the subproblem solution, and not just on $\mathbb{D}$.

It may also be possible to extend our approach to the Boosted Sampling algorithms for two-stage stochastic optimization problems [16], especially for the special case of independent decisions, but except for the rooted Steiner tree problem it is not obvious how to determine $\mathbb{E}_p[C_{aug}(\mathbb{D})]$. There is a similar but even larger obstacle if we want to use our techniques to derandomize the Boosted Sampling algorithms for multi-stage stochastic optimization problems, because here we would also need to be able to compute (an upper bound on) the expected cost of future stages.

Acknowledgements The author would like to thank David P. Williamson and Frans Schalekamp for helpful comments on earlier drafts of this paper.

References

1. Arya, V., Garg, N., Khandekar, R., Meyerson, A., Munagala, K., Pandit, V.: Local search heuristic for k-median and facility location problems. In: Proceedings of the 33rd Annual ACM Symposium on Theory of Computing, pp. 21–29 (2001)
2. Christofides, N.: Worst case analysis of a new heuristic for the traveling salesman problem. Report 388, Graduate School of Industrial Administration, Carnegie Mellon University (1976)
3. Eisenbrand, F., Grandoni, F.: An improved approximation algorithm for virtual private network design. In: Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 928–932 (2005)
4. Eisenbrand, F., Grandoni, F., Oriolo, G., Skutella, M.: New approaches for virtual private network design. SIAM J. Comput. **37**(3), 706–721 (2007)
5. Eisenbrand, F., Grandoni, F., Rothvoß, T., Schäfer, G.: Approximating connected facility location problems via random facility sampling and core detouring. In: Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1174–1183 (2008)
6. Erdős, P., Spencer, J.: Probabilistic Methods in Combinatorics. Academic Press, San Diego (1974)
7. Fakcharoenphol, J., Rao, S., Talwar, K.: A tight bound on approximating arbitrary metrics by tree metrics. J. Comput. Syst. Sci. **69**(3), 485–497 (2004)

8. Fleischer, L., Könemann, J., Leonardi, S., Schäfer, G.: Simple cost sharing schemes for multicommodity rent-or-buy and stochastic Steiner tree. In: Proceedings of the 38th Annual ACM Symposium on Theory of Computing, pp. 663–670 (2006)
9. Garg, N., Khandekar, R., Konjevod, G., Ravi, R., Salman, F.S., Sinha, A.: On the integrality gap of a natural formulation of the single-sink buy-at-bulk network design problem. In: Proceedings of the 8th International Integer Programming and Combinatorial Optimization Conference, pp. 170–184 (2001)
10. Garg, N., Gupta, A., Leonardi, S., Sankowski, P.: Stochastic analyses for online combinatorial optimization problems. In: Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 942–951 (2008)
11. Goemans, M.X., Bertsimas, D.J.: Survivable networks, linear programming relaxations and the parsimonious property. Math. Program. Ser. A **60**(2), 145–166 (1993)
12. Goyal, V., Gupta, A., Leonardi, S., Ravi, R.: Pricing tree access networks with connected backbones. In: Proceedings of the 15th Annual European Symposium on Algorithms, pp. 498–509 (2007)
13. Grandoni, F., Italiano, G.F.: Improved approximation for single-sink buy-at-bulk. In: Proceedings of the 17th International Symposium on Algorithms and Computation, pp. 111–120 (2006)
14. Guha, S., Meyerson, A., Munagala, K.: A constant factor approximation for the single sink edge installation problems. In: Proceedings of the 33rd Annual ACM Symposium on Theory of Computing, pp. 383–388 (2001)
15. Gupta, A., Kumar, A., Roughgarden, T.: Simpler and better approximation algorithms for network design. In: Proceedings of the 35th Annual ACM Symposium on Theory of Computing, pp. 365–372 (2003)
16. Gupta, A., Pál, M., Ravi, R., Sinha, A.: Boosted sampling: approximation algorithms for stochastic optimization. In: Proceedings of the 36th Annual ACM Symposium on Theory of Computing, pp. 417–426 (2004)
17. Gupta, A., Pál, M., Ravi, R., Sinha, A.: What about Wednesday? Approximation algorithms for multistage stochastic optimization. In: Approximation, Randomization and Combinatorial Optimization. Lecture Notes in Computer Science, vol. 3624, pp. 86–98. Springer, Berlin (2005)
18. Gupta, A., Kumar, A., Pál, M., Roughgarden, T.: Approximation via cost sharing: simpler and better approximation algorithms for network design. J. ACM **54**(3), 11 (2007)
19. Gupta, A., Srinivasan, A., Tardos, É.: Cost-sharing mechanisms for network design. Algorithmica **50**(1), 98–119 (2008)
20. Jung, H., Hasan, M.K., Chwa, K.-Y.: Improved primal-dual approximation algorithm for the connected facility location problem. In: COCOA 2008: Proceedings of the 2nd International Conference on Combinatorial Optimization and Applications. Lecture Notes in Computer Science, vol. 5165, pp. 265–277. Springer, Berlin (2008)
21. Immorlica, N., Karger, D., Minkoff, M., Mirrokni, V.S.: On the costs and benefits of procrastination: approximation algorithms for stochastic combinatorial optimization problems. In: Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 691–700 (2004)
22. Mahdian, M., Ye, Y., Zhang, J.: Approximation algorithms for metric facility location problems. SIAM J. Comput. **36**(2), 411–432 (2006) (electronic)
23. Raghavan, P.: Probabilistic construction of deterministic algorithms: approximating packing integer programs. J. Comput. Syst. Sci. **37**(2), 130–143 (1988)
24. Shmoys, D., Talwar, K.: A constant approximation algorithm for the a priori traveling salesman problem. In: Proceedings of the 13th International Integer Programming and Combinatorial Optimization Conference (2008)
25. Shmoys, D.B., Williamson, D.P.: Analyzing the Held-Karp TSP bound: A monotonicity property with application. Inf. Process. Lett. **35**, 281–285 (1990)
26. Swamy, C., Kumar, A.: Primal-dual algorithms for connected facility location problems. Algorithmica **40**(4), 245–269 (2004)
27. Talwar, K.: The single-sink buy-at-bulk LP has constant integrality gap. In: Proceedings of the 9th International Integer Programming and Combinatorial Optimization Conference, pp. 475–486 (2002)
28. van Zuylen, A.: Deterministic sampling algorithms for network design. In: Proceedings of the 16th Annual European Symposium. Lecture Notes in Computer Science, vol. 5193, pp. 830–841. Springer, Berlin (2008)
29. Williamson, D.P., van Zuylen, A.: A simpler and better derandomization of an approximation algorithm for single source rent-or-buy. Oper. Res. Lett. **35**(6), 707–712 (2007)
30. Wolsey, L.A.: Heuristic analysis, linear programming and branch and bound. Math. Program. Study **13**, 121–134 (1980)