Deterministic Sampling Algorithms for Network Design

Anke van Zuylen*

School of Operations Research and Information Engineering,
Cornell University, Ithaca, NY 14853.
Fax: (607) 255-9129.
avz2@cornell.edu

Abstract. For several NP-hard network design problems, the best known approximation algorithms are remarkably simple randomized algorithms called Sample-Augment algorithms in [11]. 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. [11].

1 Introduction

For several NP-hard network design problems, the best known approximation algorithms are remarkably simple randomized algorithms. 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. Following [11], 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 [21] 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 [3], the virtual private network design problem [12, 11, 1, 2], 2-stage stochastic Steiner tree problem with independent decisions [13], the *a priori* traveling salesman problem [18], and even the single sink buy-at-bulk problem [12, 11, 9], 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, and further discussions of the known results and Sample-Augment algorithms for them, to the relevant sections. We refer the reader also to the paper by Gupta, Kumar, Pál and Ravi [11], which is the journal version of the papers which first introduced Sample-Augment algorithms [12, 10].

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. [12, 11] we

* Supported by NSF grant CCF-0514628.

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 [4] 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.

What we show is that we *can* find an upper bound on the cost of the Subproblem plus Augmentation Steps that can be efficiently computed. Suppose we can 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 2-stage rooted stochastic Steiner tree problem with independent decisions, the *a priori* traveling salesman problem, the connected facility location problem in which the open facilities need to be connected by a 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. [11] (the only problem in [11] 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 [7, 20] for the single-sink buy-at-bulk problem has integrality gap at most 27.72. We also match the best known bounds for the single source rent-or-buy problem and the connected facility location problem in which open facilities need to be connected by a tree, which were obtained by applying the techniques from Williamson and Van Zuylen [21], which is a special case of our approach. We summarize our results in Table 1.

Problem	randomized	prev. best deterministic	our result
SSRoB	2.92 [3]	4.2 [14], 3.28* [21, 3]	3.28
2-stage Steiner	3.55 [13]	$\log n$ [16]	8
a priori TSP	4 [18], O(1)[6]	8* [18]	6.5
CFL-tree	4 [3]	8.29 [15], 4.23* [3]	4.23
k-CFL-tree	6.85 [3]	6.98* [3]	6.98
CFL-tour	4.12 [3]	-	4.12
VPND	3.55 [2]	$\log n$ [5]	8.02
SSBaB	24.92 [9]	216 [20]	27.72

Table 1. 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 * are obtained by using the method in Williamson and Van Zuylen [21], which is a special case of our approach. 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 [17]: 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 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 Section 3 we illustrate our method using the single source rent-or-buy problem as an example. In Sections 4, 5, 6, 7 and 8 we sketch 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, the virtual private network design problem and the single sink buy-at-bulk problem. We conclude with a brief discussion of some future directions in Section 9.

2 Derandomization of Sample-Augment Algorithms

We give a high-level description of a class of algorithms first introduced by Gupta, Kumar and Roughgarden [12], which were called Sample-Augment algorithms in [11]. Given a (minimization) problem $\mathcal{P}$, the Sample-Augment problem is defined by

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

$\mathcal{P}$-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 $\texttt{Sol}_{aug}(D, \texttt{Sol}_{sub}(D))$.

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

In the following, we let OPT denote the optimal value 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 $\mathbb{E}_p\big[C(\mathbb{D})\big]$ 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 $\mathbb{E}_p\big[C_{SA}(\mathbb{D})\big]$ given that j is included in $\mathbb{D}$ is $\mathbb{E}_{p,p_j \leftarrow 1}\big[C_{SA}(\mathbb{D})\big]$, and the conditional expectation, given that j is not included in $\mathbb{D}$ is $\mathbb{E}_{p,p_j \leftarrow 0}\big[C_{SA}(\mathbb{D})\big]$. By the method of conditional expectation [4], one of these conditional expectations has value at most $\mathbb{E}_p\big[C_{SA}(\mathbb{D})\big]$. 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 $\mathbb{E}_p\big[C_{SA}(\mathbb{D})\big]$.

Unfortunately, this is not very useful to us yet, since it is generally not the case that we can compute $\mathbb{E}_p\big[C_{SA}(\mathbb{D})\big]$. However, as we will show, for many problems and corresponding Sample-Augment algorithms, it is the case that $\mathbb{E}_p\big[C_{aug}(\mathbb{D})\big]$ 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 $U_{sub} : 2^{\mathcal{D}} \rightarrow R$ such that $C_{sub}(D) \leq U_{sub}(D)$ for any $D \subset \mathcal{D}$, and suppose we can efficiently compute $\mathbb{E}_p\big[U_{sub}(\mathbb{D})\big]$ and $\mathbb{E}_p\big[C_{aug}(\mathbb{D})\big]$ for any vector p. If there exists some vector $\hat{p}$ such that

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

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

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

(i) $\mathbb{E}_p\big[C_{aug}(\mathbb{D})\big]$ *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) There exist known 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 $D \subset \mathcal{D}$.
(iv) There exists a vector $\hat{p}$ such that

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

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) $\mathbb{E}_p\big[U_{sub}(\mathbb{D})\big]$ can be efficiently computed for any p, and by (iv) Equation (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, so that there are well-known LP relaxations and algorithms such that $\alpha = 2$ if the subproblem is a Steiner tree problem [8], and $\alpha = 1.5$ if the subproblem is a traveling salesman tour problem [22, 19]. 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 $\mathbb{E}_{\hat{p}}\big[C_{LP}(y(\mathbb{D}))\big]$ is bounded by a constant factor times OPT. Using the analysis for the randomized algorithm to bound $\mathbb{E}_{\hat{p}}\big[C_{aug}(\mathbb{D})\big]$, we can then show that (iv) holds.

Remark. In some cases, $\mathcal{P}_{sub}$ and $\mathcal{P}_{aug}$ are only defined for $D \neq \emptyset$. In such cases, we require that condition (i) holds for all p such that $p_j = 1$ for some j, and that condition (ii) holds for non-empty subsets D. Condition (iv) then asks for $\hat{p}$ such that $\hat{p}_j = 1$ for some j. The derandomization procedure will not change this element, so that the Sample-Augment algorithm is always well defined for the vectors p that we consider.

3 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 [21]; 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 $t_1, \ldots, t_k \in V$, 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 t to t_j. The cost of renting an edge e is c_e and the cost of buying e is Mc_e. For a set $T \subseteq E$, we denote by $c(T) = \sum_{e \in T} c_e$, hence the cost of solution $(B, R_1, \ldots, R_k)$ is $Mc(B) + \sum_{j=1}^{k} c(R_j)$. For $u, v \in V$, we denote by $\ell(u,v)$ the lenght of the shortest path from u to v with respect to costs c, and we let $\ell(u, F) = \min_{v \in F} \ell(u,v)$.

Gupta, Kumar, and Roughgarden [12] propose the random sampling algorithm given below, where they set $p_j = \frac{1}{M}$ for all $j = 1, \ldots, k$.

SSRoB-Sample-Augment$(G = (V,E), c, s, \{t_1, \ldots, t_k\}, p)$

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 $D \cup \{s\}$.

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. [12] 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 2 ([12]). *If $p_j = \frac{1}{M}$ for $j = 1, \ldots, k$, then $\mathbb{E}\big[C_{aug}(\mathbb{D})\big] \le 2OPT$.*

Lemma 3 ([21]). *There exists a deterministic 4-approximation algorithm for SSRoB.*

Proof. We verify that the four conditions of Theorem 1 hold. It is straightforward to show that $\mathbb{E}_p\big[C_{aug}(\mathbb{D})\big]$, the expected cost incurred in the Augmentation Step, can be computed for any vector of sampling probabilities p. Now consider the subproblem on a given subset D of $\{t_1, \ldots, t_k\}$. From Goemans and Bertsimas [8] 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:

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

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 $\mathbb{E}_p\big[C_{LP}(y(\mathbb{D}))\big]$. 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.

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

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 $D \subset \mathcal{D}$ 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 the existence of a vector $\hat{p}$ such that $\mathbb{E}_{\hat{p}}\big[C_{aug}(\mathbb{D})\big] + 2\mathbb{E}_{\hat{p}}\big[C_{LP}(y(\mathbb{D}))\big] \le 4OPT$. Let $\hat{p}_j = \frac{1}{M}$ for every $t_j \in \mathcal{D}$. Then by Lemma 2, the expected cost of the Augmentation Step is at most $2OPT$, and $2\mathbb{E}_{\hat{p}}\big[C_{LP}(y(\mathbb{D}))\big]$ is

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

Hence, applying Theorem 1, we get that there exists a 4-approximation algorithm for SSRoB. ◻

We note that it was shown in [21] that a better deterministic approximation algorithm exists, by using the improved analysis of the randomized algorithm given by Eisenbrand et al. [3], which allows us to more carefully balance the charge against the optimal renting and the optimal buying costs. We refer the reader to [21] for the details.

Lemma 4 ([21, 3]). *There exists a deterministic 3.28-approximation algorithm for the single source rent-or-buy problem.*

4 2-Stage Rooted 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 $c_e \geq 0$, a root s and terminals $t_1, \ldots, t_k$ with activation probabilities $q_1, \ldots, q_k$ 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 [13] 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 $\sigma q_j c_e$.

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.

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

Suppose we could find the optimal Steiner tree on $D \cup \{s\}$ in the Subproblem Step of the Boosted Sampling algorithm. Then Gupta et al. [11] show that the expected cost of the first stage solution is at most OPT, if the sampling probabilities are $\min\{1, q_j \sigma\}$. In addition, they show that the expected cost of the second stage is at most $2OPT$. 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 $\hat{p}_j = q_j$, we find that

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

Combining this with Gupta et al. [11]'s result that under these sampling probabilities, $\mathbb{E}_{\hat{p}}\big[C_{LP}(y(\mathbb{D}))\big] \leq 2OPT$, Theorem 1 allows us to get the following result.

Lemma 5. *There exists a deterministic 8-approximation algorithm for the 2-stage rooted stochastic Steiner tree problem with independent decisions.*

5 *A Priori* Traveling Salesman Problem 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 $t_1, \ldots, t_k$, 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 [18] 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 there is no t_j such that $q_j = 1$, then they need a revised Sampling Step to ensure at least one terminal is marked. We omit the details here.) In the Subproblem Step they find a tour on the marked terminals and finally, in 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 if at least one terminal is marked, then 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 S 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 an 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. $\tilde{q}_j = \frac{q_j(1-\prod_{i \neq j}(1-q_i))}{Q}$. 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 below. We note that the vector of sampling probabilities must have at least one element set to 1, otherwise the Augmentation Step may not be well defined. We will therefore make sure that the vector $\hat{p}$ with which we start the derandomization of APTSP-Sample-Augment has at least one element equal to 1 (in fact, it will have two elements set to 1).

APTSP-Sample-Augment$(G = (V, E), c, Q, \tilde{q}, s, \{t_1, \ldots, t_k\}, p)$

1. *(Sampling Step)* Mark each terminal t_j with probability p_j. Let D be the set of marked terminals.
2. *(Subproblem Step)* Construct a traveling salesman tour on D, and pay 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.

Shmoys and Talwar [18] show that if $\tilde{p}_j = q_j$ for every terminal, and if we were able to find a *minimum cost* solution to the subproblem, then $\mathbb{E}_{\tilde{p}}\left[C_{sub}(\mathbb{D}) \big| |\mathbb{D}| \geq 2\right] \leq OPT$, and $\mathbb{E}_{\tilde{p}}\left[C_{aug}(\mathbb{D}) \big| |\mathbb{D}| \geq 2\right] \leq 2OPT$.

This 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.

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

Note that we do not know t^*, but we can solve APTSP-LP for any $t^* \in \{t_1, \ldots, t_k\}$ 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

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

Note that this satisfies condition (ii) in Theorem 1 with $\alpha = 1.5$ by [22, 19]. 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 let $\tilde{p}_j = q_j$ and consider the expectation of $\mathbb{E}_{\tilde{p}}\big[C_{LP}(y(\mathbb{D}))\big||\mathbb{D}| \geq 2\big]$ and $\mathbb{E}_{\tilde{p}}\big[C_{aug}(\mathbb{D})\big||\mathbb{D}| \geq 2\big]$. From Shmoys and Talwar 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

$$
\begin{aligned}
1.5\mathbb{E}_{\tilde{p}}\big[C_{LP}(y(\mathbb{D}))\big||\mathbb{D}| \geq 2\big] &= 1.5\Big(\sum_{e\in E} Qc_e\hat{b}_e + \sum_{j=1}^{k} Q\tilde{q}_j c_e \hat{r}_e^j\Big) = 1.5\sum_{e\in E}\Big(Qc_e\hat{b}_e + \sum_{j=1}^{k} q_j\big(1 - \prod_{i\neq j}(1-q_i)\big)c_e\hat{r}_e^j\Big) \\
&\leq 1.5\sum_{e\in E}\Big(Qc_e\hat{b}_e + \sum_{j=1}^{k} q_j c_e \hat{r}_e^j\Big) \leq 4.5OPT \qquad (2)
\end{aligned}
$$

where the last inequality holds since we showed that APTSP-LP is a relaxation of the *a priori* Traveling Salesman Problem.

Finally, we want to get rid of the conditioning on $|\mathbb{D}| \geq 2$. By conditioning on the two smallest indices in $\mathbb{D}$ and then using basic properties of conditional expectation, one can show that there must exist two elements, say $j_1 < j_2$ such that if we let $\hat{p}_{j_1} = \hat{p}_{j_2} = 1$, $\hat{p}_j = 0$ for all $j < j_2$ and $\hat{p}_j = q_j$ for all $j > j_2$, then

$$
1.5\mathbb{E}_{\hat{p}}\big[C_{LP}(y(\mathbb{D}))\big] + \mathbb{E}_{\hat{p}}\big[C_{aug}(\mathbb{D})\big] \leq 1.5\mathbb{E}_{\tilde{p}}\big[C_{LP}(y(\mathbb{D}))\big||\mathbb{D}| \geq 2\big] + \mathbb{E}_{\tilde{p}}\big[C_{aug}(\mathbb{D})\big||\mathbb{D}| \geq 2\big].
$$

Hence we can try all possible choices of j_1, j_2, and we will find $\hat{p}$ with at least two elements equal to 1, so that condition (iv) of Theorem 1 holds with $\beta = 6.5$. Hence we get the following result.

Lemma 6. *There exists a deterministic 6.5-approximation algorithm for a priori Traveling Salesman Problem.*

Remark. Shmoys and Talwar [18] use 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 $\alpha\mathbb{E}_{\tilde{p}}\big[C_{LP}(y(\mathbb{D}))\big||\mathbb{D}| \geq 2\big] \leq 6OPT$, instead of what we find in (2), and thus get an 8-approximation algorithm.

6 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 $\mathcal{D} \subset V$ with demands d_j for $j \in \mathcal{D}$, a set of potential facilities $\mathcal{F} \subset V$, 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 edges costs satisfy the triangle inequality. The goal is to find a subset of facilities $F \subseteq \mathcal{F}$ to open such that $|F| \leq k$ (k may be ∞) and a set of edges T so that 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 their closest open facility.

We may assume without loss of generality that $d_j = 1$ for all $j \in \mathcal{D}$; see [12] for details. 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 two open facilities need to be connected by ρ_{cr} distinct paths.

To determine which facilities to open, the Sample-Augment algorithm from Eisenbrand, Grandoni, Rothvoß and Schäfer [3] marks each *client* $j \in \mathcal{D}$ independently with probability p_j and opens the facilities that the marked clients are assigned to in an (approximately optimal) solution to the corresponding 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. In our description of the Sample-Augment algorithm we omit this, but in our derandomization, we will make sure that the vector $\hat{p}$ in condition (iv) of Theorem 1 has at least one element equal to 1. CFL-Sample-Augment gives our version of the algorithm from Eisenbrand et al. [3]:

CFL-Sample-Augment$(G = (V, E), c, \mathcal{D}, \mathcal{F}, f, k, M, CR)$

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. 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 $F = \bigcup_{j \in D} \sigma_U(j)$, 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 to its closest facility in F.

It can be verified that condition (i) of Theorem 1 is satisfied for any sampling probabilities p such that $p_j = 1$ for some j. We define Sub-LP(D) as

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

Condition (ii) of Theorem 1 is satisfied with $\alpha = 2$ if $CR =$ `SteinerTree` [8], or 1.5 if $CR =$ `Tour` [22, 19].

Let $\lambda = a + \frac{M}{|\mathcal{D}|}$, where a is a parameter to be determined later. We assume we know that facility i^* 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 weighing of the renting cost by λ.

$$\begin{aligned} & \min \sum_{e \in E} M c_e b_e + \lambda \rho_{cr} \sum_{j \in \mathcal{D}} \sum_{e \in E} c_e r_e^j \\ \text{(CFL-LP)} \quad & \text{s.t.} \sum_{e \in \delta(S)} (b_e + \rho_{cr} r_e^j) \geq \rho_{cr} \quad \forall S \subset V, i^* \notin S, j \in \mathcal{D} \cap S \\ & r_e^j, b_e \geq 0 \quad \forall e \in E, j \in \mathcal{D} \end{aligned}$$

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. It is easily verified that the optimal value of CFL-LP is at most $B^* + \lambda R^*$. We define $y_e(D) = \hat{b}_e + \rho_{cr} \sum_{j \in D} \hat{r}_e^j$.

Now, if we mark one client chosen uniformly at random, and then mark each client with probability $\frac{a}{M}$ as in Eisenbrand et al. [3], then the probability that j is marked is at most $\frac{a}{M} + \frac{1}{|\mathcal{D}|}$, hence $\mathbb{E}\big[C_{LP}(y(\mathbb{D}))\big] \leq B^* + \lambda \rho_{cr} R^*$. Combined with the bounds given by Eisenbrand et al. on $\mathbb{E}\big[C_{aug}(\mathbb{D})\big]$ under this sampling strategy, we can show that $\mathbb{E}\big[C_{aug}(\mathbb{D})\big] + \alpha \mathbb{E}\big[C_{LP}(y(\mathbb{D}))\big] \leq \beta(a) OPT$, where $\beta(a)$ depends on the variant of the problem we are considering. The details are left for the full version of this paper.

Finally, we only need to remark that by the definition of expectation, there must then also exist some vector $\hat{p}$, such that $\hat{p}_j = 1$ for some j and $\hat{p}_{j'} = \frac{a}{M}$ for all other j' such that $\mathbb{E}_{\hat{p}}\big[C_{aug}(\mathbb{D})\big] + \alpha \mathbb{E}_{\hat{p}}\big[C_{LP}(y(\mathbb{D}))\big] \leq \beta(a) OPT$. Choosing a appropriately and invoking Theorem 1 then gives the following results.

Lemma 7 ([3]). *There exists a deterministic 4.23-approximation algorithm for k-connected facility location with $k = \infty$ and $CR =$ `SteinerTree`.*

Lemma 8 ([3]). *There exists a deterministic 6.98-approximation algorithm for k-connected facility location with $k < \infty$ and $CR =$ `SteinerTree`.*

Lemma 9. *There exists a deterministic 4.12-approximation algorithm for k-connected facility location with $k = \infty$ and $CR =$ `Tour`.*

7 Virtual Private Network Design.

In the virtual private network design problem, we are given a graph $G = (V, E)$ with edge cost $c_e \geq 0$, 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 $\sum_i f_{ij} \leq b_{in}(j)$ and $\sum_i f_{ji} \leq b_{out}(j)$ for every $j \in \mathcal{D}$ we need to have sufficient capacity on the paths, i.e. $\sum_{ij:e\in P_{ij}} f_{ij} \leq u_e$ for every $e \in E$. The objective is to find a solution that minimizes the cost $\sum_{e\in E} c_e u_e$ of installing capacity.

Gupta, Kumar, and Roughgarden [12] 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 [1] and Eisenbrand, Grandoni, Oriolo and Skutella [2]. We will show how Theorem 1 can be used to derandomize the improved algorithm in [2].

As was shown by Gupta et al. [12], 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 $|\mathcal{I}| \leq |\mathcal{J}|$.

The algorithm as described by Eisenbrand et al. [2] partitions $\mathcal{J}$ into $\mathcal{I}$ groups, 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, we just assume we mark each receiver with some probability p_j. We will ensure that we only consider sampling probabilities so that at least one p_j will be 1, since otherwise the Augmentation Step is not well-defined.

The VPN-Sample-Augment algorithm is described below. The algorithm installs the capacities and outputs the Steiner trees found in the Subproblem Step. 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'.

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.
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.

Eisenbrand et al. [2] also show that there exists a (deterministic) $(1+\frac{|\mathcal{J}|}{|\mathcal{I}|})$ approximation algorithm. Using Theorem 1 and the ideas from Section 5 we can show that if $|\mathcal{J}| \geq 7|\mathcal{I}|$, then there exists a deterministic 8.02-approximation algorithm.

Lemma 10. *There exists a deterministic 8.02-approximation algorithm for virtual private network design.*

Proof. (sketch) It is easily verified that condition (i) of Theorem 1 holds for all p with $p_j = 1$ for some j. The Sub-LP for condition (ii) is made up of $|\mathcal{I}|$ different Steiner tree LPs, and has $\alpha = 2$ [8]:

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

Let $\kappa = 1 - e^{-\frac{|\mathcal{J}|}{|\mathcal{I}|}}$. It follows from the analysis of Eisenbrand et al. [2] that the following LP is a relaxation of the virtual private network design problem:

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

Given an algorithm that finds minimum cost Steiner trees in the Subproblem Step, the expected cost of the solution constructed by Eisenbrand et al.[2] is at most $\frac{3}{\kappa}OPT$. Hence there is some non-empty subset D for which the cost is at most $\frac{3}{\kappa}OPT$. We let $b_e^i = 1$ for the edges added in the Subproblem Step, and $r_e^j = 1$ for the edges added in the Augmentation Step. This defines a feasible solution to VPN-LP of cost at most OPT.

Let $\hat{r}, \hat{b}$ be an optimal solution to VPN-LP, then we can let $y_e^i(D) = \hat{b}_e^i + \sum_{j\in D}\hat{r}_e^j$.

By letting $\hat{p}_j = \frac{1}{|\mathcal{I}|}$ we can then show by borrowing from the analysis in [2] that

$$\mathbb{E}_{\hat{p}}\big[C_{aug}(\mathbb{D})\big|\mathbb{D}\neq\emptyset\big] + 2\mathbb{E}_{\hat{p}}\big[C_{LP}(y(\mathbb{D}))\big|\mathbb{D}\neq\emptyset\big] \leq \frac{2+2\times 3/\kappa}{\kappa}$$

Hence there exists some $\tilde{p}$ with one element equal to 1 and the rest equal to $\frac{1}{|\mathcal{I}|}$, for which $\mathbb{E}_{\tilde{p}}\big[C_{aug}(\mathbb{D})\big] + \mathbb{E}_{\tilde{p}}\big[C_{LP}(y(\mathbb{D}))\big] \leq \frac{2+6/\kappa}{\kappa}$. Since $|\mathcal{J}| \geq 7|\mathcal{I}|$, the right hand side is at most 8.02. □

8 Single-Sink Buy-at-Bulk Network Design.

The single sink buy-at-bulk problem is a generalization of the single source rent-or-buy problem. We are again 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 $s_1, \ldots, s_n \in V$ with weight $w_j > 0$ for source s_j. We denote $\{s_1, \ldots, s_n\} = \mathcal{S}$. 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 simultaneously. 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. [12, 11] 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 defer the discussion of the Sample-Augment algorithm and its derandomization using Theorem 1 to the full version of this paper. We remark here only that we need more machinery to tackle this problem, since the decisions in stage k change the input to subsequent stages. However, we show that we can use similar ideas to upper bound the expected cost of future stages, and that this upper bound can be efficiently computed for any decisions made in the current stage. We can thus derandomize the Sample-Augment algorithm from Gupta et al. [12, 11] and obtain a deterministic 80-approximation algorithm. Using the improved Sample-Augment algorithm by Grandoni and Italiano [9], we obtain the following result.

Lemma 11. *There exists a deterministic 27.72-approximation algorithm for single sink buy-at-bulk.*

9 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. [11], 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. If we want to use Theorem 1, we would need to be able to compute $\mathbb{E}_p\left[C_{aug}(\mathbb{D})\right]$ (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, $\mathbb{E}_p\left[C_{aug}(\mathbb{D})\right]$ 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 stochastic optimization problems [13], but here again it is not obvious how to determine $\mathbb{E}_p\left[C_{aug}(\mathbb{D})\right]$.

References

1. F. Eisenbrand and F. Grandoni. An improved approximation algorithm for virtual private network design. In *SODA*, pages 928–932, 2005.
2. F. Eisenbrand, F. Grandoni, G. Oriolo, and M. Skutella. New approaches for virtual private network design. *SIAM J. Comput.*, 37(3):706–721, 2007.
3. F. Eisenbrand, F. Grandoni, T. Rothvoß, and G. Schäfer. Approximating connected facility location problems via random facility sampling and core detouring. In *SODA*, pages 1174–1183, 2008.
4. P. Erdős and J. Spencer. *Probabilistic methods in combinatorics*. Academic Press, 1974.
5. J. Fakcharoenphol, S. Rao, and K. Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *J. Comput. System Sci.*, 69(3):485–497, 2004.
6. N. Garg, A. Gupta, S. Leonardi, and P. Sankowski. Stochastic analyses for online combinatorial optimization problems. In *SODA*, pages 942–951, 2008.
7. N. Garg, R. Khandekar, G. Konjevod, R. Ravi, F. S. Salman, and A. Sinha. On the integrality gap of a natural formulation of the single-sink buy-at-bulk network design problem. In *IPCO*, pages 170–184, 2001.
8. M. X. Goemans and D. J. Bertsimas. Survivable networks, linear programming relaxations and the parsimonious property. *Math. Program.*, 60(2, Ser. A):145–166, 1993.
9. F. Grandoni and G. F. Italiano. Improved approximation for single-sink buy-at-bulk. In *ISAAC*, pages 111–120, 2006.
10. A. Gupta, A. Kumar, M. Pál, and T. Roughgarden. Approximation via cost-sharing: A simple approximation algorithm for the multicommodity rent-or-buy problem. In *FOCS*, pages 606–615, 2003.
11. A. Gupta, A. Kumar, M. Pál, and T. Roughgarden. Approximation via cost sharing: simpler and better approximation algorithms for network design. *J. ACM*, 54(3):11, 2007.
12. A. Gupta, A. Kumar, and T. Roughgarden. Simpler and better approximation algorithms for network design. In *STOC*, pages 365–372, 2003.
13. A. Gupta, M. Pál, R. Ravi, and A. Sinha. Boosted sampling: approximation algorithms for stochastic optimization. In *STOC*, pages 417–426, 2004.
14. A. Gupta, A. Srinivasan, and É. Tardos. Cost-sharing mechanisms for network design. *Algorithmica*, 50(1):98–119, 2008.
15. M. K. Hasan, H. Jung, and K.-Y. Chwa. Approximation algorithms for connected facility location problems. *J. Comb. Optim.*, to appear, 2008.
16. N. Immorlica, D. Karger, M. Minkoff, and V. S. Mirrokni. On the costs and benefits of procrastination: approximation algorithms for stochastic combinatorial optimization problems. In *SODA*, pages 691–700, 2004.
17. P. Raghavan. Probabilistic construction of deterministic algorithms: approximating packing integer programs. *J. Comput. System Sci.*, 37(2):130–143, 1988.
18. D. Shmoys and K. Talwar. A constant approximation algorithm for the a priori traveling salesman problem. In *IPCO*, 2008.
19. D. B. Shmoys and D. P. Williamson. Analyzing the Held-Karp TSP bound: A monotonicity property with application. *Inf. Process. Lett.*, 35:281–285, 1990.
20. K. Talwar. The single-sink buy-at-bulk LP has constant integrality gap. In *IPCO*, pages 475–486, 2002.
21. D. P. Williamson and A. van Zuylen. A simpler and better derandomization of an approximation algorithm for single source rent-or-buy. *Oper. Res. Lett.*, 35(6):707–712, 2007.
22. L. A. Wolsey. Heuristic analysis, linear programming and branch and bound. *Math. Prog. Study*, 13:121–134, 1980.