An approximation algorithm for a facility location problem with stochastic demands

A.F. Bumb and J.C.W. van Ommeren*

Abstract

In this article we propose, for any $\epsilon > 0$, a $2(1+\epsilon)$-approximation algorithm for a facility location problem with stochastic demands. At open facilities, inventory is kept such that arriving requests find a zero inventory with (at most) some pre-specified probability. The incurred costs are the expected transportation costs from the demand points to the facilities, the operating costs of the facilities and the investment in inventory .

Keywords approximation algorithms, stochastic facility location
AMS classification 68W25, 90B06, 60K30

1 Introduction

Facility location problems have been extensively studied in the OR-literature. In a facility location problem, we are given a set of demand points and a set of location where facilities may be opened. The goal is to decide at which location to open facilities and how to assign demand points to facilities such that the total cost of opening facilities and of connecting demand points to facilities is minimized. Variants of this problem can be formulated if one imposes a set of requirements on the set of open facilities or the assignment has to satisfy [9]. Examples of such requirements are a maximum number of facilities that may be opened, a maximum demand that may be served by a facility, or a maximum travel distance from a demand point to an open facility. The facility location problem with its variants has proved to be a very useful tool in modeling many network design or location problems, such as location of plants or warehouses ([26], [9]) and placement of caches [11].

*Corresponding author
Email address: j.c.w.vanommeren@ewi.utwente.nl
Postal address: Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

In this paper we study a variant of the facility location problem where at demand points a stochastic number of requests for items is generated. At open facilities, inventory is kept and, if possible, requests for items are fulfilled immediately. However, since the number of requests is random, it may occur that there is no inventory at the arrival of a request and the request has to be cancelled. An arbitrary request arriving at a facility, should only have a (pre-specified) small probability of being lost. We are interested in the relationship between the problem with stochastic demands and inventory and known facility location problems, in particular from the perspective of approximation algorithms.

We will call a ρ-approximation algorithm a polynomial time algorithm that always finds a feasible solution with **objective function value** within ρ times the optimum. The value ρ is called the *performance (approximation) guarantee* of the algorithm.

The majority of facility location problems for which approximation algorithms are known, are deterministic. The simplest version of a facility location problem, *the metric uncapacitated facility location problem* (UFLP), that is the facility location problem with no restrictions on the facilities or the assignment of demand points and with the transportation costs being a metric, is known to be NP-hard. If the transportation costs are unrestricted, approximating the UFLP is as hard as approximating set cover, and therefore cannot be done better than $O(\log n)$ factor, unless $\mathbf{NP} \subseteq \tilde{\mathbf{P}}$. In this article, we assume, for all the facility locations mentioned, that the transportation costs form a metric. There are several approximation algorithms for the UFLP known in the literature ([1, 8, 10, 13, 14, 15, 18, 26]). The currently known best performance guarantee for the UFLP is 1.52, due to Mahdian, Ye and Zhang [18]. Guha and Khuller [10] and Sviridenko [27] have proved that a better factor than 1.463 for the UFLP is not possible unless $\mathbf{NP} \subseteq \tilde{\mathbf{P}}$.

The problem in which each facility has a certain capacity, but more facilities may be opened at a location if the demand exceeds the capacity of one facility, is known as the *soft capacitated* facility location problem. The approximation algorithms for the soft capacitated facility location problems are usually based on reductions to the uncapacitated version of the problem [14, 13, 18, 19]. The best approximation algorithm for this problem has an approximation ratio of 2 and was proposed by Mahdian, Ye and Zhang [18] in [19]. In ([12]) the authors propose a 1.861-approximation algorithm for the variant in which the cost of facilities are concave functions of the number of demand points served. For the *hard capacitated* facility location problem with splittable demands, where each facility has a certain capacity, only one facility may be open at a location and a demand point may be served by several locations, the best approximation algorithm is due to Zhang, Chen and Ye ([29]), and achieves an approximation ratio between $3 + 2\sqrt{2} - \epsilon$ and $3 + 2\sqrt{2} + \epsilon$, for any given constant $\epsilon > 0$.

Stochastic facility location problems (problem where the demand is stochastic or/and the service offered by facilities is of stochastic nature) were mainly treated in the OR literature ([2, 3, 4, 7, 6, 20, 21, 28]). Several heuristics have been proposed to obtain solutions for these problems. To the best of our knowledge, the first approximation algorithm for a stochastic facility location problem was proposed by Ravi and Sinha in [24] and was improved by Mahdian in ([16]). The latest algorithm is based on the primal-dual technique and has a 3-approximation guarantee. Their approach is scenario-based, *i.e.* in each scenario all the data are known, including the probability with which each scenario takes place.

The paper is organized as follows. In section 2 we describe the stochastic facility location problem in more detail and formulate it such that it can be reduced to a soft capacitated facility location problem. Based on this reduction, we then propose in Section 3, a $2(1+\epsilon)$-approximation algorithm for our problem. We conclude the section by showing that the same ideas can be applied for designing approximation algorithms for a larger class of problems. Finally, we present some conclusions and remarks on the stochastic facility location problem we have analyzed.

2 The facility location problem with stochastic demands

In this section we describe in more detail the stochastic facility location problem in which we are interested. There is a set of demand points $D, |D| = N$ at which requests are generated, and a set of locations, $F, |F| = K$, where facilities may be opened. We assume that the requests at a demand point $j \in D$ are generated according to a Poisson process, independent of the processes at other demand points in D. At each open facility an inventory is kept such that an arriving request finds a zero inventory (and is lost), with probability at most α. We then say that $(1-\alpha)$ is the *fill rate* of the system . The inventories at the open facilities are restored only at fixed points in time and the period between two such points is called a *reorder period*. The holding cost per unit of inventory at an open facility $i \in F$ is c_i and the cost of keeping a facility open at location $i \in F$ during a reorder period is f_i. The transportation cost per unit of demand from facility $i \in F$ to demand point $j \in D$ is c_{ij}. We assume that the transportation costs are proportional to the distances and form a metric.

The goal is to decide at which locations to open facilities, the level of inventory to be installed at each open facility and how to assign demand points to facilities such that the fill rate is at least $1-\alpha$ and the average total cost per reorder period is minimized.

Let X_j denote the number of generated requests at demand point j during a reorder period and let $\lambda_j = E(X_j)$. Denote by V_i the inventory order up to level at facility $i \in F$, *i.e.* the inventory level at the beginning of a reorder period. Let y_i, respectively x_{ij}, be $0-1$ variables indicating if a facility at location $i \in F$ is open, respectively if demand point $j \in D$ is assigned to a facility $i \in F$. The facility location problem with stochastic demands given above, is fully described by the following integer program:

$$\begin{aligned}
\min \quad & \sum_{i\in F}(f_i + c_iV_i)y_i + \sum_{j\in D}\sum_{i\in F}\lambda_j c_{ij}x_{ij} && (1)\\
\text{s.t.} \quad & x_{ij} \le y_i, \qquad i\in F, \quad j\in D, && (2)\\
& \sum_{i\in F} x_{ij} = 1, \qquad j\in D, && (3)\\
& \text{P(an arbitrary arriving requests at facility } i \text{ is lost)} \le \alpha, \quad i\in F, && (4)\\
& x_{ij}, y_i \in \{0,1\}, \quad i\in F, \quad j\in D. && (5)
\end{aligned}$$

The first term in the objective function includes the costs for keeping facilities open and for the maximum inventory at the facilities during a reorder period, while the second term is the expected transportation cost during such a period. Constraints (2), (3) and (5) guarantee that each demand point is assigned to exactly one open facility and constraints (4) guarantee that the fill rate attained at each open location will be at least $1-\alpha$.

Next we will give an equivalent formulation of constraints (4). Let $\tilde{X}_i$ be the total demand assigned to location i. Clearly, $\tilde{X}_i = \sum_{j\in D} x_{ij}X_j$. Since the requests generated at demand points during reorder periods are independent Poisson distributed random variables, $\tilde{X}_i$ has a Poisson distribution with mean $\mathrm{E}(\tilde{X}_i) = \sum_{j\in D} x_{ij}\lambda_j$. From the theory of regenerative processes (see e.g. [25]), it follows that for location i, the following holds:

$$\text{P(an arbitrary arriving requests at facility } i \text{ is lost)} = \frac{\mathrm{E}((\tilde{X}_i - V_i)^+)}{\mathrm{E}(\tilde{X}_i)}, \qquad (6)$$

where $(a)^+ = \max(0,a)$. Condition (4) can be rewritten as

$$\mathrm{E}((\tilde{X}_i - V_i)^+) \le \alpha \mathrm{E}(\tilde{X}_i).$$

For a Poisson distributed random variable Y with $E(Y) = \lambda$, define the inventory $V_\alpha(\lambda)$ by

$$V_\alpha(\lambda) = \min\{n | \left[E((Y-n)^+)) \le \alpha\lambda\right]\}. \qquad (7)$$

Using (6) and (7), our problem can be reformulated as

$$(\mathbf{P}) \quad \begin{aligned}
\min \quad & \sum_{i\in F}(f_i + c_iV_\alpha(\sum_{j\in D} x_{ij}\lambda_j))y_i + \sum_{j\in D}\sum_{i\in F}\lambda_j c_{ij}x_{ij}\\
\text{s.t.} \quad & x_{ij} \le y_i, \qquad i\in F, \quad j\in D,\\
& \sum_{i\in F} x_{ij} = 1, \qquad j\in D,\\
& x_{ij}, y_i \in \{0,1\}, \qquad i\in F, \quad j\in D.
\end{aligned}$$

Note that constraints (4) have moved into the objective function. This will enable us to further reduce the problem to a soft capacitated facility location problem, for which approximation algorithms are known (see e.g. [19]). In the remainder of the paper we will present this reduction in detail.

3 A $2(1+\epsilon)$-approximation algorithm for the stochastic facility location problem

For a facility location problem (P), an instance $\mathcal{I}$ and a feasible solution $\mathcal{S}$ we denote by $cost_{F,\mathcal{I}(P)}(\mathcal{S})$ the cost of opening facilities and by $cost_{T,\mathcal{I}(P)}(\mathcal{S})$ the transportation cost incurred by $\mathcal{S}$. For the sake of simplicity, we will omit the instance from the notation.

Definition 1 *We call a polynomial time reduction $\mathcal{R}$ from problem P_1 to P_2 a (σ_F, σ_T)-reduction if $\mathcal{R}$ maps an instance $\mathcal{I}$ of P_1 to an instance $\mathcal{R}(\mathcal{I})$ of P_2 and it has the following properties:*

a) For any feasible solution $\mathcal{S}_1$ for the instance $\mathcal{I}$ of P_1 there is a corresponding solution $\mathcal{S}_2$ for the instance $\mathcal{I}$ of P_2 with

$$cost_{F,P_2}(\mathcal{S}_2) \leq \sigma_f cost_{F,P_1}(\mathcal{S}_1),$$

and

$$cost_{T,P_2}(\mathcal{S}_2) \leq \sigma_c cost_{T,P_1}(\mathcal{S}_1).$$

b) For any feasible solution $\mathcal{S}_2$ for the instance $\mathcal{R}(\mathcal{I})$ of P_2, there is a feasible solution $\mathcal{S}_1$ for the instance $\mathcal{I}$ of P_1 with

$$cost_{F,P_1}(\mathcal{S}_1) + cost_{T,P_1}(\mathcal{S}_1) \leq cost_{F,P_2}(\mathcal{S}_2) + cost_{T,P_2}(\mathcal{S}_2).$$

Definition 2 *An algorithm is called an (α, β)-approximation algorithm for a facility location problem (P), if for any instance $\mathcal{I}$ of (P), and for any solution $\mathcal{S}$ for $\mathcal{I}$ the cost of the solution found by the algorithm is at most $\alpha cost_{F,P}(\mathcal{S}) + \beta cost_{T,P}(\mathcal{S})$.*

Remark 3 Note that combining a (σ_F, σ_T)-reduction from P_1 to P_2 and an (α, β)-approximation algorithm for P_2 gives an $(\alpha\sigma_F, \beta\sigma_T)$-approximation algorithm for P_1. Moreover, the approximation guarantee of the algorithm for P_1 is $\max\{\alpha\sigma_F, \beta\sigma_T\}$.

The construction of a $2(1+\epsilon)$-approximation algorithm for $(\mathbf{P})$, consists of several steps. First we will study the inventory function $V_\alpha(\lambda)$ given by (7). Based on it's properties, we propose a $(2,1)$-reduction of $(\mathbf{P})$ to a soft capacitated facility location problem, named $(\mathbf{SP_2})$. Finally, we describe a refined soft capacitated problem, $(\mathbf{SP_{1+\epsilon}})$ to which $(\mathbf{P})$ can be $(1+\epsilon, 1)$-reduced and show that this gives $2(1+\epsilon)$-approximation algorithm for $(\mathbf{P})$.

Lemma 4 *The function $V_\alpha(\lambda)$ satisfies*

$$V_\alpha(\lambda_1 + \lambda_2) \leq V_\alpha(\lambda_1) + V_\alpha(\lambda_2).$$

Proof. Suppose that two independent Poisson streams with rate λ_1, respectively λ_2, arrive at a location i and that the inventory level at location i is $V_\alpha(\lambda_1)+V_\alpha(\lambda_2)$. Let Y_1 and Y_2 be the number of arrivals in the first, respectively in the second stream. Since

$$(Y_1 + Y_2 - (V_\alpha(\lambda_1) + V_\alpha(\lambda_2)))^+ \leq (Y_1 - V_\alpha(\lambda_1))^+ + (Y_2 - V_\alpha(\lambda_2))^+,$$

it is readily seen that

$$\begin{aligned} \mathrm{E}(X_1 + X_2 - (V_\alpha(\lambda_1) + V_\alpha(\lambda_2)))^+ &\leq \mathrm{E}(X_1 - V_\alpha(\lambda_1))^+ + \mathrm{E}(X_2 - V_\alpha(\lambda_2))^+ \\ &\leq \alpha(\lambda_1 + \lambda_2). \end{aligned}$$

Hence, $V_\alpha(\lambda_1 + \lambda_2) \leq V_\alpha(\lambda_1) + V_\alpha(\lambda_2)$. ■

Remark 5 Note that $V_\alpha(\lambda)$ is a step function, thus not concave. Therefore we cannot directly use the procedure proposed in Mahdian and Pal ([17], for solving the facility location problem with concave facility cost functions. Moreover, not even the length of the steps is increasing as function of the height , where the length of a step at level n is defined as $\sup\{\lambda | V_\alpha(\lambda) = n\} - \inf\{\lambda | V_\alpha(\lambda) = n\}$. For example, numerical experiments show that, when $\alpha = 0.1$, the length of the steps is increasing up to level 40 and decreasing above this level.

Next we present a reduction of (**P**) to a soft capacitated facility location problem, which we denote by ($\mathbf{SP_2}$). The demand points, their requests and facility locations are the same as in problem (**P**). Let $M = \lceil \log_2(V_\alpha(\sum_{j\in D} \lambda_j) \rceil$ and let $L = \{1, \cdots, M\}$. We define M types of facilities with capacities $u_\ell = \max\{\lambda | V_\alpha(\lambda) \leq 2^\ell\}$, respectively. A facility of type l at location i is denoted by (i, l) and has corresponding cost $f_{il} = f_i + c_i 2^\ell$. At each location $i \in F$, M facilities may be opened.

Let the 0-1 variables y_{il}, x_{ilj}, indicate whether a facility of type l is opened at location i, respectively whether demand point j is assigned to facility (i, l). Then, ($\mathbf{SP_2}$) can be formulated as the integer program:

$$(\mathbf{SP_2}) \quad \begin{aligned} \min \quad & \sum_{j\in D} \sum_{i\in F} \sum_{\ell\in L} \lambda_j c_{ij} x_{i\ell j} + \sum_{i\in F} \sum_{\ell\in L} f_{i\ell} y_{i\ell} \\ \text{s.t.} \quad & \sum_{j\in D} \lambda_j x_{i\ell j} \leq u_\ell y_{il}, \qquad i \in F, \quad \ell \in L, && (8) \\ & \sum_{i\in F} \sum_{\ell\in L} x_{i\ell j} = 1, \qquad j \in D, && (9) \\ & x_{i\ell j}, y_{i\ell} \in \{0,1\}, \qquad i \in F, \quad j \in D, \quad \ell \in L.. && (10) \end{aligned}$$

Constraints (8), (9) and (10) insure that each demand point is assigned to one open facility and that no more then u_ℓ demand points are assigned to a facility of type ℓ.

Remark 6 Note that although formulated as a hard capacitated facility location problem ($y_{il} \in \{0,1\}$), problem ($\mathbf{SP_2}$) is a soft capacitated problem. Suppose that we relax the y variables to be integer. Consider first a $k < M$. The optimal solution of the relaxed version will not choose to open two facilities of type k at a location, since opening a facility of type $k+1$ is cheaper and has the same capacity as two facilities of type k. Since one facility of type M can handle all the demand, there will be always at most one facility of type M open in the optimal solution of the relaxed version of ($\mathbf{SP_2}$). Thus, ($\mathbf{SP_2}$) is a soft capacitated facility location problem.

In the following lemma we describe a $(2,1)$-reduction of ($\mathbf{P}$) to ($\mathbf{SP_2}$).

Lemma 7
(i) For each feasible solution $(\tilde{x},\tilde{y})$ of ($\mathbf{P}$) with facility cost $cost_{F,(\mathbf{P})}(\tilde{x},\tilde{y})$ and transportation cost $cost_{T,(\mathbf{P})}(\tilde{x},\tilde{y})$ there exists a feasible solution (x,y) of ($\mathbf{SP_2}$) with $cost_{F,(\mathbf{SP_2})}(x,y) \le 2cost_{F,(\mathbf{P})}(\tilde{x},\tilde{y})$ and $cost_{T,(\mathbf{SP_2})}(x,y) = cost_{T,(\mathbf{P})}(\tilde{x},\tilde{y})$.
(ii) For each feasible solution (x,y) of ($\mathbf{SP_2}$), there exists a feasible solution $(\tilde{x},\tilde{y})$ of ($\mathbf{P}$) of lower cost.
(iii) There exists a $(2,1)$-reduction of ($\mathbf{P}$) to ($\mathbf{SP_2}$) .

Proof. (i) Consider a solution $(\tilde{x},\tilde{y})$ of ($\mathbf{P}$). For $i \in F$ with $\tilde{y}_i = 1$ and $\ell \in L$ define $\ell_i = \min\{n | \sum_{j\in D} \tilde{x}_{ij}\lambda_j \le u_n\}$, set $y_{i\ell} = 1$ for $\ell = \ell_i$, set $y_{i\ell} = 0$ otherwise and set $x_{i\ell j} = \tilde{x}_{ij} y_{i\ell}$ for $j \in D$. For each $i \in F$ with $\tilde{y}_i = 0$, set $x_{i\ell j} = y_{i\ell} = 0$ for $j \in D$ and $\ell \in \{1,\cdots,M\}$ and define $\ell_i = 1$. It can readily be seen that (x,y) is a feasible solution of ($\mathbf{SP_2}$) with associated costs

$$\text{cost}_{T,(\mathbf{SP_2})}(x,y) = \sum_{j\in D}\sum_{i\in F}\sum_{\ell\in L} \lambda_j c_{ij} x_{i\ell j} = \sum_{j\in D}\sum_{i\in F} \lambda_j c_{ij}\tilde{x}_{ij} = \text{cost}_{T,(\mathbf{P})}(\tilde{x},\tilde{y})$$

and

$$\text{cost}_{F,(\mathbf{SP_2})}(x,y) = \sum_{i\in F}\sum_{\ell\in L} f_{i\ell} y_{i\ell} = \sum_{i\in F}(f_i + 2^{\ell_i}) y_{i\ell_i} \le 2\text{cost}_{F,(\mathbf{P})}(\tilde{x},\tilde{y}),$$

where the inequality follows from the definitions of ℓ_i and u_n.

(ii) For each feasible solution (x,y) of ($\mathbf{SP_2}$), define the vector $(\tilde{x},\tilde{y})$ by $\tilde{x}_{i,j} = \max_{\ell\in\{1,\cdots,M\}}\{x_{i\ell j}\}$ and $\tilde{y}_i = \max_{\ell\in\{1,\cdots,M\}}\{y_{i\ell}\}$. Clearly, $(\tilde{x},\tilde{y})$ is a feasible solution for ($\mathbf{P}$). Moreover, from Lemma 4 follows that $V_\alpha(\sum_{j\in D}\tilde{x}_{ij}\lambda_j) \le \sum_\ell u_\ell y_{i\ell}$ and so $(\tilde{x},\tilde{y})$ has a lower cost than the one incurred by (x,y) for ($\mathbf{SP_2}$).

(iii) Follows from (i) and (ii) of this lemma. ■

In the following, we prove that one can obtain a $(1+\epsilon,1)$-reduction between ($\mathbf{P}$) and a slightly modified version of ($\mathbf{SP_2}$) by the same reasoning as in Lemma 7. We define this modified version ($\mathbf{SP_{1+\epsilon}}$) as follows.

Define for $\epsilon > 0$ the integer sequence $\tilde{v}_{0,0} = 0$; $v_{m0} = \lfloor (1+\epsilon)(1+v_{m-1,0}) \rfloor$ and $v_{mk} = 2^k v_{m0}$ for $m = 1, 2, \cdots$ and $k = 0, 1, \cdots$. Next, define the integer sequence $v_0 = 0$ and $v_\ell = \min\{\tilde{v}_{mk} > v_{\ell-1} | m = 1, 2, \cdots \text{ and } k = 0, 1, \cdots\}$ for $\ell = 1, 2, \cdots$ and define $M = \min\{\ell | v_\ell \geq V_\alpha(\sum_{j\in D} \lambda_j)\}$. Since $\tilde{v}_{m0} \geq (1+\epsilon)v_{m-1,0}$, it is easy to find that, for $\epsilon \in (0,1)$,

$$M \leq \lceil \log_{(1+\epsilon)}(V_\alpha(\sum_{j\in D} \lambda_j)\rceil \lceil \log_2(V_\alpha(\sum_{j\in D} \lambda_j)\rceil \leq \frac{4}{\epsilon} \lceil \log_2(V_\alpha(\sum_{j\in D} \lambda_j)\rceil^2.$$

Furthermore, from the construction of the sequence v_ℓ, we see that $(1+v_\ell) \leq v_{\ell+1} \leq (1+\epsilon)(1+v_\ell)$. Consider a facility location problem with the same demand points, requests and facility locations as in problem $(\mathbf{P})$. At each location $i \in F$, M facilities may be opened, $(i,1), ...(i,M)$, of costs $f_i + c_i v_\ell$ and capacities $u_\ell = \max\{\lambda | V_\alpha(\lambda) \leq v_\ell\}$.

Let the 0-1 variables y_{il}, x_{ilj}, indicate whether a facility of type l is opened at location i, respectively whether demand point j is assigned to facility (i,l). Then, $(\mathbf{SP_{1+\epsilon}})$ can be formulated as an integer program similar to $(\mathbf{SP_2})$.

As in Remark **??**, we note that although formulated as a hard capacitated facility location problem, $(\mathbf{SP_{1+\epsilon}})$ is in fact a soft capacitated facility location problem. In order to show this, we prove that, even if we allow more facilities of the same type to be opened at a location, at most one will be opened in the optimal solution. Assume that in the optimal solution, at least one facility of type k at location i is opened. If the cost of facility (i,k) exceeds $f_i + c_i V_\alpha(\lambda)/2$, then opening facility (i,M) (which can handle all demands) is cheaper than opening two facilities (i,k). If the costs of facility (i,k) equals $f_i + c_i v_k$ with $v_k \leq V_\alpha(\lambda)/2$, we see, by the definition of the sequence v_ℓ, that there is also a facility (i,k') with cost $f_i + 2c_i v_k$. By Lemma 4, the capacity of a type k' facility is at least twice the capacity of a type k facility. Hence, in the optimal solution of the relaxed problem of $(\mathbf{SP_{1+\epsilon}})$, at every location at most one facility of type k is opened. Thus, $(\mathbf{SP_{1+\epsilon}})$ is a soft capacitated facility location problem.

Lemma 8 *For any $\epsilon > 0$, the problem $(\mathbf{P})$ can be $(1+\epsilon,1)$-reduced to $(\mathbf{SP_{1+\epsilon}})$.*

Proof. We follow the proof of Lemma 7. Consider a feasible solution $(\tilde{x},\tilde{y})$ of $(\mathbf{P})$ and construct a feasible solution (x,y) of $(\mathbf{SP_{1+\epsilon}})$ as follows. Open facility (i,ℓ) at location i only if $\sum_{j\in D} \tilde{x}_{ij} = 1$ and $\ell = \min\{n | \sum_{j\in D} \tilde{x}_{ij}\lambda_j \leq u_n\}$. Since the inventory levels are discrete and $\sum_{j\in D} \tilde{x}_{ij}\lambda_j > u_{\ell-1}$, the inventory at location i satisfies $V_\alpha(\sum_{j\in D} x_{ij}\lambda_j) \geq 1 + v_{\ell-1}$ and therefore the cost of opening facilities in $(\mathbf{SP_{1+\epsilon}})$ is at most $(1+\epsilon)$ times the facility costs in $(\mathbf{P})$.

Now consider a solution (x,y) of $(\mathbf{SP_{1+\epsilon}})$ and construct a corresponding solution $(\tilde{x},\tilde{y})$ of $(\mathbf{P})$ by $\tilde{x}_{i,j} = \max_{\ell\in\{1,\cdots,M\}}\{x_{i\ell j}\}$ and $\tilde{y}_i = \max_{\ell\in\{1,\cdots,M\}}\{y_{i\ell}\}$. As in Lemma 7, one can show that $(\tilde{x},\tilde{y})$ is a feasible solution with the same transportation cost as the one incurred by (x,y) and with less opening facility cost than the one incurred by (x,y). ■

Theorem 9 *There is a $2(1+\epsilon)$-approximation algorithm for the facility location problem with stochastic demands* $(\mathbf{P})$.

Proof. The existence of a (2,2)-approximation algorithm for the soft capacitated facility location problem $(\mathbf{SP_{1+\epsilon}})$, implies, by Lemma 8 and Remark 3, the existence of a $2(1+\epsilon)$-approximation algorithm for the stochastic facility location problem $(\mathbf{P})$. Thus, to prove this theorem, it suffices to show that there exists a (2,2)-approximation algorithm for $(\mathbf{SP_{1+\epsilon}})$.

The $(\mathbf{SP_{1+\epsilon}})$ is a soft capacitated facility location problem with general demands. For the soft capacitated facility location problem with unit demands, a (2,2)-approximation algorithm was proposed in ([19]). First, the problem is reduced to a so called linear cost facility location problem, that is a facility location problem where the facility cost at a location is linear in the number of demand points that is assigned to this location. Then the authors prove that the algorithm proposed by Jain *et al.* in [13] for the UFLP can be used for deriving an approximation algorithm for the linear cost facility location problem. Moreover, the approximation ratio for both the transportation cost and opening facilities cost remain the same as for the UFLP.

In the appendix we show that their analysis can be easily extended to general demands, thus implying a (2,2)-approximation algorithm for $(\mathbf{SP_{1+\epsilon}})$. ∎

Generalization At the basis of our algorithm lies the property that, for two demand points j and j', with demand λ_j, respectively $\lambda_{j'}$, the inventory which has to be installed at a facility satisfies $V_\alpha(\lambda_j+\lambda_{j'}) \leq V_\alpha(\lambda_j)+V_\alpha(\lambda_{j'})$, i.e., it is more profitable to look at the joint demand than to treat the demands separately. It is easy to see that the same analysis holds for the metric UFLP with the cost of opening facilities depending on the amount served by a facility and satisfying $f_i(\lambda_j+\lambda_{j'}) \leq f_i(\lambda_j)+f_i(\lambda_{j'})$, for each $i \in F$ and $j, j' \in D$. Clearly, concave facility costs have this property.

Remark 10 The same technique can also be used for the following version of the facility location problem with stochastic demands: at facilities an arbitrary number of servers can be placed, which all work at unit speed. At each facility, there is an upperbound on the probability that a customer has to wait more than some fixed time. The incurred costs are the transportation costs and the facility costs; the cost of a facility is the sum the opening cost and the cost for installing servers, that is linear in the number of installed servers.

We model a facility as an $M/M/c$ queue, that is a queue with c servers and exponential interarrival and service times. Without loss of generality, we assume that the expected service time is 1. Let $WT(M_\lambda/M/c)$ denote the waiting time at such a queue with arrival rate λ. At an open facility i with arrival rate Λ_i and c_i servers, the constraint on the waiting time then is $\mathrm{P}(WT(M_{\Lambda_i}/M/c_i) \geq \tau) \leq \alpha$ for some pre-specified α and τ. An explicit expression for this probability can be found in *e.g.* [**?**], page 73. Define $N_{\alpha,\tau}(\lambda) = \min\{c | \mathrm{P}(WT(M_\lambda/M/c) \geq \tau) \leq \alpha\}$. It can be shown that $N_{\alpha,\tau}(\lambda_1 + \lambda_2) \leq N_{\alpha,\tau}(\lambda_1) + N_{\alpha,\tau}(\lambda_2)$. Thus, applying a similar reduction as the one described in this section, one obtains a $2(1+\epsilon)$-approximation algorithm for this problem as well.

4 Conclusions

In this paper we have introduced a facility location problem with inventory and stochastic demands. We proposed a $2(1+\epsilon)$-approximation algorithm for this model by giving both a $(1+\epsilon, 1)$-reduction to a soft capacitated facility location problem with general demands and a $(2, 2)$-approximating algorithm for this soft capacitated facility location problem. The same analysis is applied for approximating more general problems.

References

[1] V. Arya, N. Garg, R. Khandekar, V. Pandit, A. Meyerson, K. Munagala, Local search heuristics for k-median and facility location problems, in: Proceedings of the 33rd ACM Symposium on Theory of Computing, 2001, pp. 21-29.

[2] R. Batta, Single server queuing-location models with rejection, Transportation Science 22 (1988) 209-216.

[3] R. Batta, A queuing location model with expected service time dependent queuing disciplines, European Journal of Operational Research 39 (1989) 192-205.

[4] O. Berman, R. Larson and S. Chiu, Optimal server location on a network operating as a M/G/1 queue, Operations Research 12 (1985) 746-771.

[5] O. Berman, R. Larson and c. Parkan, The stochastic queue p-median location problem, Transportation Science 21 (1987) 207-216.

[6] O. Berman and K. Sapna, Optimal control of service for facilities holding inventory, Computers & Operations Research 28 (2001) 429-441,

[7] O. Berman and R. Mandowsky, Location-allocation on congested networks, European Journal of Operational Research 26 (1986) 238-250.

[8] F.A. Chudak and D.B. Shmoys, Improved Approximation Algorithms for the Uncapacitated Facility Location Problem, SIAM Journal of Computing 33(1) (2003) 1-25.

[9] G. Cornuejols, G.L. Nemhauser and L.A. Wolsey, The uncapacitated facility location problem, in: P. Mirchandani and R. Francis (Eds.), Discrete Location Theory, John Wiley and Sons, New York, 1990, pp. 119-171.

[10] D. Gross and C.M. Harris, Fundamentals of Queueing Theory, 3d ed., John Wiley & Sons, Inc. (1998).

[11] S. Guha, S. Khuller, Greedy strikes back: Improved facility location algorithms, Journal of Algorithms 31(1) (1999) 228-248.

[12] S. Guha, A. Meyerson, K. Munagala, A constant factor approximation algorithm for the fault-tolerant facility location problem. J. Algorithms 48(2) (2003) 429-440.

[13] M.G. Hajiaghayi, M. Mahdian, V.S Mirrokni, The facility location problem with general cost functions. Networks 42(1) (2003) 42-47.

[14] K. Jain, M. Mahdian, E. Markakis, A. Saberi, V. Vazirani, Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. Journal of the ACM 50(6) (2003) 795-824.

[15] K. Jain, V. Vazirani, Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and Lagrangian relaxation, Journal of the ACM 48 (2001) 274-296.

[16] M. R. Korupolu, C. G. Plaxton, R. Rajaraman. Analysis of a Local Search Heuristic for Facility Location Problems. J. Algorithms 37(1) (2000) 146-188.

[17] M. Mahdian, Facility Location and the Analysis of Algorithms through Factor-Revealing Programs, Ph.D. Thesis, MIT, June 2004. (available at http://www-math.mit.edu/∼mahdian/phdthesis.pdf).

[18] M. Mahdian, M. Pal, Universal Facility Location, ESA 2003, 409-421

[19] M. Mahdian, Y. Ye, J. Zhang, A 1.52 approximation algorithm for the uncapacitated facility location problem, in: Proceedings of the 5th International Workshop on Approximation Algorithms for Combinatorial Optimization, Springer-Verlag LNCS Vol 2462, 2002, 229-242.

[20] M. Mahdian, Y. Ye, J. Zhang, A 2-Approximation Algorithm for the Soft-Capacitated Facility Location Problem, RANDOM-APPROX 2003, 129-140

[21] V. Marianov, D. Serra, Probabilistic, maximal covering location-allocation models for congested systems, Journal of Regional Science 38 (1998) 401-424.

[22] V. Marianov, D. Serra, Location-allocation of multiple-server service centers with constrained queues or waiting times, Annals of Operations Research 111 (2002) 35-50.

[23] R. R. Mettu, C. G. Plaxton, The online median problem, in: Proceedings of the 41st IEEE Symposium on Foundation of Computer Science, 2000, 339-348.

[24] M. Pal, E. Tardos, T. Wexler, Facility location with nonuniform hard capacities, in: Proceedings of the 42nd Annual IEEE Symposium on the Foundations of Computer Science, 2001, 329-338.

[25] R. Ravi, A. Sinha, Hedging Uncertainty: Approximation Algorithms for Stochastic Optimization Problems, IPCO 2004, 101-115.

[26] W.L. Smith Regenerative stochastic processes, Proc. Roy Soc. Ser. A 232 (1955) 6-31.

[27] D. Shmoys, E. Tardos, K. Aardal. Approximation algorithms for facility location problems, in: Proceedings of the 29th ACM Symposium on Theory of Computing, 1997, 265-274.

[28] M. Sviridenko. Personal communication. Cited in S. Guha, Approximation algorithms for facility location problems, PhD thesis, Stanford, 2000, (Downloadable from website http://Theory.Stanford.EDU/∼sudipto).

[29] Q. Wang, R. Batta, C. Rump, Algorithms for a facility location problem with stochastic customer demand and immobile servers. Annals of Operations Research 111 (2002) 17-34.

[30] J. Zhang, B. Chen, Y. Ye, A Multi-exchange Local Search Algorithm for the Capacitated Facility Location Problem, IPCO 2004, 219-233

Appendix: A (2,2)-approximation algorithm for the soft capacitated facility location problem with general demands

Consider an instance of a metric UFLP with general demands given by: a set of demand points D, a set of facility locations F, demands $\{\lambda_j, j \in D\}$ and transportation costs $\{c_{ij}, i \in F, j \in D\}$. If additionally, a capacity u_i is specified for each facility at location $i \in F$, we obtain an instance of the soft capacitated facility location problem. If in a UFLP every facility $i \in F$ has a cost of the form $f_i = a_i\Lambda_i + b_i$, where Λ_i is the demand served by facility i, we obtain a linear cost facility location problem (LCFLP) with general demands.

In the sequel we prove that the algorithm proposed in [19] can be extended for finding good approximate solutions to the soft capacitated facility location problem with general demands. We will follow the same steps: first we give the factor revealing LP for the UFLP with general demands. Next we show that the algorithm proposed in Jain *et al.*[13] is a $(1, 2)$-approximation algorithm for the linear cost facility location problem with general demands, by looking at a factor revealing LP for a suitably chosen UFLP. Finally, we reduce the soft capacitated facility location problem to an LCFLP with general demands. We prove that this is a

$(2, 1)$-reduction. Combining the reduction and the approximation algorithm of Jain *et al.*, we can obtain a (2,2)-approximation algorithm for the soft capacitated facility location problem with general demands. We proceed with the detailed description of each step.

A.1 The factor revealing LP for the UFLP with general demands

In this subsection we will closely follow the terminology and proofs in [13]. Consider a UFLP as described above.

Recall the algorithm Jain *et al.* propose for solving the UFLP (Algorithm 2 in [13]). Based on this algorithm, the authors formulate a factor revealing LP, which finally gives the approximation ratio of the algorithm. Every demand point (city in their case) $j \in D$ has a *budget* α_j, from which he *pays* for transportation and opening of facilities. The algorithm has a notion of time. The budget of each demand point is gradually increased (from zero) at unit rate during the algorithm, until all demand points get connected to some facility. For details on how the budget is increased and which are the criteria for opening facilities and assigning of demand points to them see [13].

In Subsection 8.3, the authors indicate that, for solving the UFLP with general demands, the only modification one has to bring to the algorithm for UFLP is to increase the budget of each demand point $j \in D$ at rate λ_j. They claim that the running time and the analysis of the algorithm remain unchanged. As in Section 5.2 and Section 6 in [13], one can construct the following factor revealing LP for the UFLP with general demands as follows.

Consider a star consisting of a facility having cost f and k demand points numbered from 1 to k. Let d_j be the transportation cost of one unit of demand between demand point j and the facility (so $d_j = c_{mn}$ for some $m \in F$ and $n \in D$) and let α_j be the budget of j at the end of the algorithm. Assume that $\alpha_1 \leq ... \leq \alpha_k$. For every $1 \leq j' \leq k$, consider the situation of the algorithm at time $t = \alpha_{j'} - \delta$, with δ very small. For every demand point j, $j < j'$, if j is connected to some facility, let $r_{j,j'}$ be the transportation cost of one unit of demand from j to that facility, otherwise, let $r_{j,j'} = \frac{\alpha_j}{\lambda_j}$. Note that the definition of the variables $r_{j,j'}$ is slightly different than the one for unit demands, but this has no influence on the analysis.

The factor revealing LP associated with the algorithm for the UFLP with general demands is:

$$
\text{(FR)} \quad
\begin{aligned}
\max \quad & \frac{\sum_{j=1}^{k} \alpha_j - \gamma_f f}{\sum_{j=1}^{k} \lambda_j d_j} \\
\text{s.t.} \quad & \alpha_j \le \alpha_{j+1}, && \forall\ 1 \le j < k : \\
& r_{j,j'} \ge r_{j,j'+1}, && \forall\ 1 \le j < j' < k \\
& \frac{\alpha_{j'}}{\lambda_{j'}} \le r_{j,j'} + d_j + d_{j'}, && \forall\ 1 \le j < j' \le k \qquad (11) \\
& \sum_{j=1}^{j'-1} \max(\lambda_j r_{j,j'} - \lambda_j d_j, 0) + \sum_{j=j'}^{k} \max(\alpha_{j'} - \lambda_j d_j, 0) \le f, && \forall\ 1 \le j' \le k \qquad (12) \\
& \alpha_j, d_j, r_{j,j'}, f \ge 0, && \forall\ 1 \le j < j' \le k.
\end{aligned}
$$

Jain *et al.* prove that for $\gamma_f = 1$, the algorithm is a $(1, 2)$-approximation algorithm for the UFLP with general demands.

A.2 The UFLP and the LCFLP

We will next associate to each LCFLP a UFLP. Let S be a solution of the LCFLP with facility costs $f_i = a_i \Lambda_i + b_i, i \in F$ and transportation costs $(c_{ij})_{i \in F, j \in D}$. Clearly, S can be viewed as a solution of the UFLP with facility costs $f_i = b_i, i \in F$ and transportation costs $\tilde{c}_{ij} = c_{ij} + a_i$, for each $i \in F$ and $j \in D$.

For solving this UFLP, one can employ the algorithm proposed by Jain *et al.* [13], based on the dual fitting technique and the factor revealing LP (**FR**). However, we will show that (**FR**) can be strengthened in the sense that, instead of the costs $\tilde{c}$, one can use the transportation costs c appearing in the LCFLP. This implies that the approximations ratio for the transportation cost and the cost of opening facilities remain the same. Our proof closely follows [19], where the same reasoning is used for unit demands.

Clearly, for solving the LCFLP, one can use the algorithm developed by Jain *et al.* for the associated UFLP with costs defined as in Subsection **??**.

Consider location i with open facility in S and the set of clients assigned to it. Denote these clients by $\{1, ..., k\}$. Then $\Lambda_i = \sum_{j=1}^{k} \lambda_j$. Let d_j, respectively $\tilde{d}_j$ be the transportation cost of one unit of demand from demand point j to the facility at location i in LCFLP, respectively the associated UFLP. Clearly, $d_j = \tilde{d}_j - a_i$. By condition (**??**), we know that $\frac{\alpha_{j'}}{\lambda_{j'}} \le r_{j,j'} + \tilde{d}_j + \tilde{d}_{j'}$. In a similar way to [19] we will strengthen this result to $\frac{\alpha_{j'}}{\lambda_{j'}} \le r_{j,j'} + d_j + d_{j'}$.

If $\alpha_{j'} = \alpha_j$, the claim is true, since this happens only when $r_{jj'} = \frac{\alpha_j}{\lambda_j}$. Otherwise, consider demand points $j < j'$ at time $t = \frac{\alpha_{j'}}{\lambda_{j'}} - \delta$, with δ very small. Let s be the facility to which j is assigned at time t. By the triangle inequality,

$$\tilde{c}_{sj'} = c_{sj'} + a_s \leq c_{sj} + d_j + d_{j'} + a_s = \tilde{c}_{jj'} + d_j + d_{j'} \leq r_{jj'} + d_j + d_{j'}.$$

Moreover, $\frac{\alpha_{j'}}{\lambda_{j'}} \leq \tilde{c}_{sj'}$, otherwise j' could have been connected to s at a time earlier then t.

Next we will strengthen relation condition (**??**) of the factor revealing LP for the associated UFLP. It is known that

$$\sum_{j=1}^{j'-1} \max(\lambda_j r_{j,j'} - \lambda_j \tilde{d}_j, 0) + \sum_{j=j'}^{k} \max(\alpha_{j'} - \lambda_j \tilde{d}_j, 0) \leq b_i.$$

Since $\max(x - a, 0) \geq \max(x, 0) - a$ for $x \geq 0$, then

$$\sum_{j=1}^{j'-1} \max(\lambda_j r_{j,j'} - \lambda_j d_j, 0) + \sum_{j=j'}^{k} \max(\alpha_{j'} - \lambda_j d_j, 0) \leq b_i + a_i \sum_{j=1}^{k} \lambda_j.$$

Hence, $\alpha_j, r_{jj'}, d_j, a$ and b are a feasible solution of the following optimization problem:

$$\begin{array}{lll}
\max & \dfrac{\sum_{j=1}^{k} \alpha_j - \gamma_f(b + a\sum_{j=1}^{k} \lambda_j)}{\sum_{j=1}^{k} \lambda_j d_j} & \\
\text{s.t.} & \alpha_j \leq \alpha_{j+1}, & \forall\ 1 \leq j < k \\
& r_{j,j'} \geq r_{j,j'+1}, & \forall\ 1 \leq j < j' < k \\
& \dfrac{\alpha_{j'}}{\lambda_{j'}} \leq r_{j,j'} + d_j + d_{j'}, & \forall\ 1 \leq j < j' \leq k \\
& \displaystyle\sum_{j=1}^{j'-1} \max(\lambda_j r_{j,j'} - \lambda_j d_j, 0) + \sum_{j=j'}^{k} \max(\alpha_{j'} - \lambda_j d_j, 0) \leq b + a \sum_{j=1}^{k} \lambda_j, & \\
& & \forall\ 1 \leq j \leq k \\
& \alpha_j, d_j, r_{j,j'}, a, b \geq 0, & \forall\ 1 \leq j < j' \leq k.
\end{array}$$

This optimization problem is the same as the factor revealing LP of a UFLP with transportation costs c and opening facility costs $f_i = a_i \Lambda_i + b_i$. Hence, the algorithm of Jain *et al.* will give a $(1, 2)$approximation algorithm for the LCFLP.

A.3 Reduction of the soft capacitated facility location problem with general demands to an LCFLP with general demands

Following the same ideas as in Theorem 4 of [19], one can reduce a soft capacitated facility location problem with general demands to an LCFLP as follows. From an instance of the soft capacitated facility location problem construct an instance of the LCFLP with the same set of demand points and demands, set of locations, transportation costs and the costs of opening facilities given by $f_i' = f_i\left(1 + \Lambda_i u_i^{-1}\right)$, where Λ_i is the demand served by the facility at location i. Note that $\lceil \lambda u_i^{-1} \rceil \leq 1 + \lambda u_i^{-1} \leq 2\lceil \lambda u_i^{-1} \rceil$ for $\lambda > 0$ and $u_i > 0$. This implies that this reduction of the soft capacitated facility location problem to the LCFLP is a $(2,1)$-reduction. In section **??** it is shown that for an LCFLP a $(1,2)$-approximation algorithm exists. Together these results give a (2,2)-approximation algorithm for the soft capacitated facility location problem with general demands.