



## A new approximation algorithm for the multilevel facility location problem

Adriana F. Gabor<sup>a,\*</sup>, Jan-Kees C.W. van Ommeren<sup>b</sup>

<sup>a</sup> *Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands*

<sup>b</sup> *Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands*

### ARTICLE INFO

#### Article history:

Received 25 September 2007

Received in revised form 5 November 2009

Accepted 17 November 2009

Available online 16 December 2009

#### Keywords:

Facility location

Approximation algorithms

Randomized algorithms

### ABSTRACT

In this paper we propose a new integer programming formulation for the multilevel facility location problem and a novel 3-approximation algorithm based on LP-rounding. The linear program that we use has a polynomial number of variables and constraints, thus being more efficient than the one commonly used in the approximation algorithms for these types of problems.

© 2009 Elsevier B.V. All rights reserved.

### 1. Introduction

Facility location problems have been extensively studied in the OR and theoretical computer science literature [10,20]. In a facility location problem the following data are given: a set of demand points  $D$ , a set of locations  $F$  where facilities may be opened, the costs of opening facilities and the transportation costs from demand points to facilities. One has to decide where to open facilities and how to assign the demand points to them, such that the total cost (of opening facilities and transportation) is minimized.

In this paper, we study the *multilevel facility location problem*, (MFLP) where facilities are organized on  $n$  levels  $F = V_1 \cup \dots \cup V_n$  and each demand point  $k \in D$  has to be assigned to a path  $p \in V_1 \times \dots \times V_n$  of open facilities passing each level. The demand of each demand point  $k$  is  $d_k$ . The cost of opening a facility  $i \in F$  is  $f_i$  and the cost of transporting one unit of demand from facility  $i$  to facility  $j$  is the same for all demand points, namely  $c_{ij}$ . The cost of transporting a unit of demand from a demand point  $k$  to a facility  $i \in V_1$  is  $c_{ki}$ . We assume that each facility can serve an unlimited demand and that the transportation costs form a metric. One has to decide where to open facilities and how to assign a demand point to a path of open facilities such that the total cost is minimized. The metric MFLP is encountered in supply chains and the placement of servers in internet [11].

For  $n = 1$ , the MFLP reduces to the classical *uncapacitated* facility location problem (UFLP). Since the UFLP is NP-hard, the MFLP is NP-hard as well. The focus of our paper will be on approximation algorithms for the metric MFLP. We call a  $\rho$ -approximation algorithm a polynomial time algorithm which gives a solution of cost at most  $\rho$  times the cost of an optimal solution.  $\rho$  is called the approximation guarantee (factor) of the algorithm. For the metric UFLP, a series of approximation algorithms have been developed in recent years, encompassing a broad range of techniques, such as: LP-rounding [21,9], greedy algorithms [12], local search [16,5], primal–dual [14,8] and dual fitting [17,15]. Until recently, the best approximation ratio for the UFLP has been 1.517 and it is attained by the algorithm proposed by Mahdian, Ye and Zhang [17]. In [7],

\* Corresponding author. Tel.: +31 10 408 2378; fax: +31 40 246 5995.

E-mail addresses: [gabor@ese.eur.nl](mailto:gabor@ese.eur.nl) (A.F. Gabor), [j.c.w.vanommeren@ewi.utwente.nl](mailto:j.c.w.vanommeren@ewi.utwente.nl) (J.-K.C.W. van Ommeren).

Byrka modifies the approximation algorithm proposed by Chudak and Shmoys in [9] and improves the approximation guarantee to 1.5. Guha and Khuller proved in [12] that there is no  $\rho$ -approximation algorithm with  $\rho < 1.463$ , unless  $NP \subseteq DTIME(n^{\log \log n})$ .

For the MFLP with  $n = 2$ , the first constant approximation algorithm was developed by Shmoys, Tardos and Aardal in [21] and was based on LP-rounding. In [2], Aardal, Chudak and Shmoys extend the algorithm proposed in [21] to an arbitrary number of levels and improve the approximation guarantee to 3. Although it has the best known approximation guarantee, their algorithm has the drawback of having to solve a linear program with an exponential number of variables. In the search of more efficient algorithms, several combinatorial algorithms have been developed in the recent years. The first such algorithm was developed by Meyerson, Munagala and Plotkin [18] and has an approximation guarantee of  $O(\ln(|D|))$ . Subsequently, Guha, Meyerson and Munagala [13] improved the approximation guarantee to 9.2. Bumb and Kern [6] used the primal–dual technique to improve the approximation factor to 6. In [3], Ageev proves an important result for the development of approximation algorithms for the MFLP, namely that any  $\rho$ -approximation algorithm for the UFLP leads to a  $3\rho$ -approximation algorithm for the MFLP. The reduction used by Ageev is similar to the one proposed by Edwards in [11]. By improving the reduction procedure, Ageev, Ye and Zhang [4] obtain a performance guarantee of 3.27, the best known performance guarantee obtained by a combinatorial algorithm for the metric MFLP. Zhang shows in [22] that for  $n = 2$ , a 1.77-approximation algorithm can be obtained by combining techniques such as randomized rounding, dual fitting and a greedy procedure.

The first contribution of this paper is a new integer programming formulation for the MFLP. Our integer program can be seen as an extension to more levels of the integer program introduced in [1] for the maximization version of the two level facility location problems. The difference between the integer program we are using and the commonly used integer program in the approximation algorithms for MFLP, is that instead of assigning demand points to paths, we assign them to adjacent edges between consecutive levels. The integer program thus preserves the “level structure” of the MFLP. As a consequence, the number of variables in its linear programming relaxation is decreased from an exponential one ( $|D||V_1| \times \dots \times |V_n| + |F|$ , as in [2,6]) to a polynomial one ( $|F| + |D||V_1| + |D| \sum_{l=1}^{n-1} |V_l||V_{l+1}|$  in this paper). The number of constraints is however higher, but still polynomial:  $|D| + 2|D| \sum_{l=1}^{n-1} |V_l| + |D||V_n|$  constraints versus  $|D| + |F|$  in [2,6]. The second contribution of the paper is a novel 3-approximation algorithm based on randomized rounding. For  $n = 1$ , our algorithm reduces to the 3-approximation algorithm of Chudak and Shmoys described in [9]. For  $n > 1$ , the algorithm is more elaborated, due to the fact that for each demand point, one has to insure a path of open facilities passing all levels. The algorithm exploits the “level structure” preserved by the integer program: if one knows which facilities to open on the lowest  $m$  levels ( $m \geq 1$ ) in order to insure optimality, the problem is reduced to a facility level problem on  $n - m$  levels. In each level, facilities are opened according to a procedure similar to the one used in [9] for the one level problem. Due to the fact that the integer program formulated in this paper allows the decomposition of MFLP on levels, we hope that it could be useful in designing an algorithm with an approximation ratio less than 3.

The paper is organized as follows: Section 2 contains the new integer program and some properties of its LP-relaxation. Section 3 contains the algorithm and its analysis. In Section 4 we present conclusions and further research ideas.

## 2. On an integer formulation of the MFLP and its LP-relaxation

In this section we describe a new integer programming formulation for the multilevel facility location problem. Our formulation is inspired by the one introduced in [1] for the maximization version of the two level facility location problems.

Unless otherwise specified, we will call a path  $p$  an  $n$ -tuple  $(i_1, \dots, i_n) \in V_1 \times \dots \times V_n$ . We will indicate that  $i$  is a component of  $p$  by  $i \in p$ .

The integer programming formulation most commonly used in approximation algorithms for MFLP models naturally the description of the problem (see [2,6]). The assignment of a demand point  $k \in D$  to a path  $p$  is indicated by a 0–1 variable  $x_{kp}$  and the opening of a facility  $i \in F$  through the 0–1 variable  $y_i$ . The constraints require that each demand point is assigned to one path (i.e.  $\sum_{p \in V_1 \times \dots \times V_n} x_{kp} = 1$ ) and that all the facilities on a path  $p$  to which a demand point was assigned are opened (i.e.  $\sum_{p': i \in p'} x_{kp'} \leq y_i$ , for each  $i \in F$ ). Although straightforward, this formulation has  $|D| \times |V_1| \times \dots \times |V_n| + |F|$  variables and requires extra technical details in solving it (see [2,11]).

Instead of assigning demand points to paths, we will assign demand points to edges, such that each demand point is assigned to an edge between each two consecutive levels of facilities and the edges have a vertex in common. For modeling this, we introduce the following 0–1 variables:

- $y_i, i \in F$  indicate whether  $i \in F$  is open,
- $x_{ki}, i \in V_1, k \in D$ , indicate whether demand point  $k$  is assigned to facility  $i \in V_1$
- $z_{kij}, (i, j) \in V_l \times V_{l+1}$ , for  $l = 1, \dots, n - 1$  indicate whether demand point  $k$  uses the edge  $(i, j)$ .

We denote the transportation costs by

$$c(x, z) := \sum_{k \in D} \sum_{i \in V_1} d_k c_{ki} x_{ki} + \sum_{k \in D} \sum_{l=1}^{n-1} \sum_{(i,j) \in V_l \times V_{l+1}} d_k c_{ij} z_{kij}$$

and the costs for opening facilities by

$$f(y) := \sum_{i \in F} f_i y_i.$$

We formulate the MFLP as the integer program  $(\mathbf{P}_{\text{int}})$  (see Fig. 1).

$$\begin{aligned}
 & \text{minimize } c(x, z) + f(y) \\
 & \text{subject to } \sum_{i \in V_1} x_{ki} = 1, \quad k \in D, \tag{1} \\
 & \sum_{j \in V_2} z_{kij} = x_{ki}, \quad i \in V_1, k \in D, \tag{2} \\
 & \sum_{j \in V_{l+1}} z_{kij} = \sum_{j' \in V_{l-1}} z_{kj'i}, \quad i \in V_l, 2 \leq l \leq n-1, k \in D, \tag{3} \\
 & x_{ki} \leq y_i, \quad k \in D, i \in V_1, \tag{4} \\
 & \sum_{j \in V_{l-1}} z_{kji} \leq y_i, \quad 2 \leq l \leq n, i \in V_l, k \in D, \tag{5} \\
 & y_i \in \{0, 1\}, \quad i \in F, \\
 & x_{ki} \in \{0, 1\}, \quad \forall k \in D, i \in V_1, \\
 & z_{kij} \in \{0, 1\}, \quad (i, j) \in V_l \times V_{l+1}, 1 \leq l \leq n-1, k \in D.
 \end{aligned}$$

$(\mathbf{P}_{\text{int}})$

Fig. 1. The integer program  $(\mathbf{P}_{\text{int}})$ .

Constraints (1) ensure that each demand point  $k \in D$  gets connected to exactly one facility on the first level. Constraints (2) say that demand point  $k$  uses an edge  $(i, j) \in V_1 \times V_2$  only if  $k$  is assigned to facility  $i \in V_1$ , i.e.,  $x_{ki} = 1$ . Constraints (3) ensure that demand point  $k$  uses an edge  $(i, j) \in V_l \times V_{l+1}, 2 \leq l \leq n-1$  only if  $k$  uses an edge  $(j', i)$ , with  $j' \in V_{l-1}$ . Finally, constraints (4), respectively, (5) say that a demand point  $k$  will be assigned to a facility  $i \in V_1$ , respectively will use an edge  $(j, i) \in V_{l-1} \times V_l$ , for  $2 \leq l \leq n$ , only if facility  $i$  is open. Denote by  $C_{\text{OPT}}$  the optimal value of  $(\mathbf{P}_{\text{int}})$ .

Note that the variables  $x_{ki}$  can be eliminated from the above integer program and constraints (1) and (2) replaced by  $\sum_{i \in V_1} \sum_{j \in V_2} z_{kij} = 1$ , as it is done for 2 levels in [1]. Although  $(\mathbf{P}_{\text{int}})$  is not the most compact formulation, we prefer to use it, as it is more suitable for the description of the approximation algorithm we propose.

In the remaining of the paper we will heavily make use of the Linear Programming relaxation of  $(\mathbf{P}_{\text{int}})$  described in Fig. 2.

$$\begin{aligned}
 & \text{minimize } c(x, z) + f(y) \\
 & \text{subject to } \sum_{i \in V_1} x_{ki} = 1, \quad k \in D, \tag{6} \\
 & \sum_{j \in V_2} z_{kij} = x_{ki}, \quad i \in V_1, k \in D, \tag{7} \\
 & \sum_{j \in V_{l+1}} z_{kij} = \sum_{j' \in V_{l-1}} z_{kj'i}, \quad i \in V_l, 2 \leq l \leq n-1, k \in D, \tag{8} \\
 & x_{ki} \leq y_i, \quad k \in D, i \in V_1, \tag{9} \\
 & \sum_{j \in V_{l-1}} z_{kji} \leq y_i, \quad 2 \leq l \leq n, i \in V_l, k \in D, \tag{10} \\
 & y_i \geq 0, \quad i \in F, \\
 & x_{ki} \geq 0, \quad \forall k \in D, i \in V_1, \\
 & z_{kij} \geq 0 \quad (i, j) \in V_l \times V_{l+1}, 1 \leq l \leq n-1, k \in D.
 \end{aligned}$$

$(\mathbf{P}_{\text{LP}})$

Fig. 2. The linear program  $(\mathbf{P}_{\text{LP}})$ .

First observe that the LP-program  $(\mathbf{P}_{\text{LP}})$  has  $|F| + |D||V_1| + |D| \sum_{l=1}^{n-1} |V_l||V_{l+1}|$  variables and  $|D| + 2|D| \sum_{l=1}^{n-1} |V_l| + |D||V_n|$  constraints.

Remark that it is not necessary to impose in  $(\mathbf{P}_{\text{LP}})$  that  $x_{ki} \leq 1$ , for  $k \in D$  and  $i \in V_1$  since this is insured by constraint (6). Furthermore, we conclude from (7) that  $\sum_{i \in V_1} \sum_{j \in V_2} z_{kij} = 1$  and, using (8) iteratively, that  $\sum_{i \in V_l} \sum_{j \in V_{l+1}} z_{kij} = 1$ . Therefore,  $z_{kij} \leq \sum_{j \in V_{l+1}} z_{kij} \leq 1$  for each  $1 \leq l \leq n-1, i \in V_l, j \in V_{l+1}, k \in D$ .

Moreover, in an optimal solution  $(x, y, z)$  of  $(P_{LP})$ , for each  $i \in V_1, k \in D, x_{ki} \leq 1$  which implies that  $y_i \leq 1$ . Finally, in an optimal solution, from  $\sum_{j \in V_{l-1}} z_{kji} \leq 1$  follows that  $y_i \leq 1$ , for  $2 \leq l \leq n, i \in V_l$  and  $k \in D$ .

Denote by  $C_{LP}$  the optimum value to  $(P_{LP})$ . Clearly,  $C_{LP} \leq C_{OPT}$ .

The results in next section will heavily rely on the optimal dual solution of  $(P_{LP})$  and the primal complementary slackness conditions. Let  $v_k$  be the dual variables corresponding to constraints (6),  $t_{ki}$  the dual variables corresponding to (7) for  $i \in V_1$  and (8) for  $i \in V_l, l \geq 2$  and  $u_{ki}$  the dual variables corresponding to (9) for  $i \in V_1$ , respectively (10) for  $i \in V_l$ , with  $l \geq 2$ . The dual  $(D_{LP})$  is described in Fig. 3.

$$\begin{aligned}
 & \text{maximize } \sum_{k \in D} v_k \\
 & \text{subject to } v_k - t_{ki} - u_{ki} \leq d_k c_{ki}, \quad k \in D, \quad i \in V_1, \\
 & \quad t_{ki} - t_{kj} - u_{kj} \leq d_k c_{ij}, \quad k \in D, \quad i \in V_l, \quad j \in V_{l+1}, \quad 1 \leq l \leq n-2, \\
 (D_{LP}) \quad & \quad t_{ki} - u_{kj} \leq d_k c_{ij}, \quad k \in D, \quad i \in V_{n-1}, \quad j \in V_n \\
 & \quad \sum_{k \in D} u_{ki} \leq f_i, \quad i \in F, \\
 & \quad u_{ki} \geq 0, \quad k \in D, \quad i \in V_1.
 \end{aligned}$$

Fig. 3. The dual program  $(D_{LP})$ .

Let  $(x^*, y^*, z^*)$ , respectively  $(v^*, t^*, u^*)$  be optimal solutions for  $(P_{LP})$ , respectively  $(D_{LP})$ . The primal complementary slackness constraints give the following relations between the two optimal solutions:

- (C1)  $\forall k \in D$  and  $i \in V_1, x_{ki}^* > 0$  implies  $v_k^* - t_{ki}^* - u_{ki}^* = d_k c_{ki}$
- (C2)  $\forall (i, j) \in V_1 \times V_{l+1}, 1 \leq l \leq n-2$ , and  $k \in D, z_{kij}^* > 0$  implies  $t_{ki}^* - t_{kj}^* - u_{kj}^* = d_k c_{ij}$
- (C3)  $\forall (i, j) \in V_{n-1} \times V_n$  and  $k \in D, z_{kij}^* > 0$  implies  $t_{ki}^* - u_{kj}^* = d_k c_{ij}$
- (C4)  $\forall i \in F, y_i^* > 0$  implies  $\sum_{k \in D} u_{ki}^* = f_i$ .

Next we will present some properties of the optimal solutions  $(x^*, y^*, z^*)$ , respectively  $(v^*, t^*, u^*)$ .

We say that a demand point  $k \in D$  is LP-assigned to a path  $(i_1, \dots, i_n) \in V_1 \times \dots \times V_n$ , in the optimal solution  $(x^*, y^*, z^*)$  of  $(P_{LP})$ , if  $x_{ki_1}^* > 0, z_{ki_1 i_2}^* > 0, \dots, z_{ki_{n-1} i_n}^* > 0$ . Note that a demand point may be assigned to more than one path.

**Lemma 1.** *If a demand point  $k$  is LP-assigned to path  $(i_1, \dots, i_n) \in V_1 \times \dots \times V_n$  in an optimal solution  $(x^*, y^*, z^*)$ , then  $c_{ki_1} + \sum_{l=1}^{n-1} c_{i_l i_{l+1}} \leq \frac{v_k^*}{d_k}$ .*

**Proof.** Based on (C1)–(C2), it follows that:

$$\begin{aligned}
 v_k^* - t_{ki_1}^* - u_{ki_1}^* &= d_k c_{ki_1} \\
 t_{ki_l}^* - t_{ki_{l+1}}^* - u_{ki_{l+1}}^* &= d_k c_{i_l i_{l+1}}, \quad \text{for } 1 \leq l \leq n-2 \\
 t_{ki_{n-1}}^* - u_{ki_n}^* &= d_k c_{i_{n-1} i_n}.
 \end{aligned}$$

By summing up these equalities, we obtain that

$$d_k \left( c_{ki_1} + \sum_{l=1}^{n-1} c_{i_l i_{l+1}} \right) = v_k^* - \sum_{l=1}^n u_{ki_l}^*.$$

Since  $u_{ki_l}^* \geq 0$ , the claim follows. ■

In other words, we have shown that the transportation costs along any path to which  $k$  is assigned in the primal optimal solution, cannot exceed  $v_k^*$ .

### 3. A 3-approximation algorithm for MFLP

In this section we will describe a 3-approximation algorithm for the MFLP based on randomized rounding. The algorithm aims to construct a random solution  $(X, Y, Z)$  for  $(P_{int})$  such that  $E(c(X, Z) + f(Y)) \leq 3C_{LP} \leq 3C_{OPT}$ .

Before presenting the algorithm, we will introduce some definitions and notations.

Let  $(x^*, y^*, z^*)$ , respectively  $(v^*, t^*, u^*)$  be optimal solutions to  $(P_{LP})$ , respectively  $(D_{LP})$ . For each demand point  $k$ , denote by  $C_k$  the transportation costs incurred by  $k$  in the optimal solution, i.e.,

$$C_k = \sum_{i \in V_1} d_k c_{ki} x_{ki}^* + \sum_{l=1}^{n-1} \sum_{(i,j) \in V_l \times V_{l+1}} d_k c_{ij} z_{kij}^*.$$

Denote by  $N(k)$  the neighborhood of  $k$ , i.e. the set of facilities  $i \in V_1$  with  $x_{ki}^* > 0$  and  $i \in V_l, 2 \leq l \leq n-1$  for which there exists a  $j \in V_{l-1}$  such that  $z_{kij}^* > 0$ . Clearly, if  $i \in N(k) \cap V_l$  for some  $k \in D, 1 \leq l \leq n-1$ , (8) imply that  $\sum_{j \in V_{l+1}} z_{kij}^* > 0$ .

In the next Lemma we present some properties of an optimal solution of  $(P_{LP})$ .

**Lemma 2.** (a) For each  $i \in N(k) \cap V_l, l \geq 2$  the set  $\{j \in V_{l-1} | z_{kji}^* > 0\} \subseteq N(k)$ .

(b) For each  $k \in D$  and  $i \in V_1 \cap N(k)$ , there exists a path  $p$  in  $N(k)$  such that  $i \in p$  and  $k$  is LP-assigned to  $p$ .

(c) For each  $k \in D$  and  $i \in N(k) \cap V_l, 2 \leq l \leq n$ , there exists a path  $p$  in  $N(k)$  such that  $i \in p$  and  $k$  is LP-assigned to  $p$ .

**Proof.** (a) Consider a  $j \in V_{l-1}$  such that  $z_{kji}^* > 0$ . If  $l = 2$ , respectively  $l > 2$ , constraints (7), respectively constraints (8) imply that  $x_{kj}^* > 0$ , respectively that there exists an  $i_{l-2} \in V_{l-2}$  such that  $z_{kij_{l-2}}^* > 0$ . In both cases,  $j \in N(k)$ .

(b) From constraint (7) follows that if  $x_{ki}^* > 0$ , there exists a facility  $i_2 \in V_2$  such that  $z_{kii_2}^* > 0$ . Clearly,  $i_2 \in N(k)$ . The claim then follows by using (8) in an induction procedure on the level  $l$ .

(c) Follows from (b) and (a). ■

**Approximation algorithm**

- Order the demand points in increasing order of  $\frac{v_k^* + C_k}{d_k}$ .
- Declare all the demand points *unclustered* and let the set of clustered points be  $Cl = \emptyset$ .
- Repeat until  $Cl \supseteq D$  (that is all points are clustered).
  - Choose among the unclustered demand points the demand point  $k$  with the smallest value of  $\frac{v_k^* + C_k}{d_k}$ .
  - Declare  $k$  a cluster center.
  - Choose an index  $i \in V_1$  with probability  $x_{ki}^*$ .
  - Iteratively, perform the following: for each level  $l, 1 \leq l \leq n-1$ , if facility  $i \in V_l$  was opened, open facility  $j \in V_{l+1}$  with probability  $\frac{z_{kij}^*}{\sum_{j \in V_{l+1}} z_{kij}^*}$ .
  - Assign to the cluster centered at  $k, Cl_k$ , all facilities in  $N(k)$  and the unclustered demand points  $k'$  with  $N(k) \cap N(k') \neq \emptyset$ . Set  $Cl = Cl \cup Cl_k$  (that is declare these points *clustered*).
  - Assign all the demand points in  $Cl_k$  to the path of opened facilities in  $Cl_k$ .

Denote by  $CC$  the set of cluster centers. Lemma 2 together with constraints (6) and the fact that  $\sum_{j \in V_{l+1}} \frac{z_{kij}^*}{\sum_{j \in V_{l+1}} z_{kij}^*} = 1$ , imply that the probabilities used in the algorithm are well defined.

Before analyzing the solution returned by the algorithm, note the following important property of cluster centers.

**Lemma 3.** (a) The neighborhoods of any two cluster centers are disjoint.

(b) In the neighborhood of any cluster center, there exists a path of open facilities.

**Proof.** (a) Consider two cluster centers  $k$  and  $k'$ . Suppose that  $\frac{v_k^* + C_k}{d_k} \leq \frac{v_{k'}^* + C_{k'}}{d_{k'}}$ . If there was an  $i \in N(k) \cap N(k')$ , then  $k'$  would belong to the cluster centered at  $k$  and  $k'$  would not be a cluster center. Hence,  $N(k) \cap N(k') = \emptyset$ .

(b) Follows from the definition of the neighborhood and the way of opening facilities in the algorithm. ■

Since each demand point is contained in exactly one cluster and in each cluster there is one path of open facilities, each demand point will be assigned to one path. Thus, we have obtained the following random solution  $(X, Y, Z)$  to  $(P_{int})$ :

for each  $i \in F$ ,

$$Y_i = \begin{cases} 1, & \text{if } i \text{ was opened} \\ 0, & \text{otherwise;} \end{cases}$$

for each  $(i, k) \in V_1 \times D$ ,

$$X_{ki} = \begin{cases} 1, & \text{if } i \text{ is on the path to which } k \text{ was assigned} \\ 0, & \text{otherwise;} \end{cases}$$

and for each  $(i, j, k) \in V_l \times V_{l+1} \times D, 1 \leq l \leq n-1$ ,

$$Z_{kij} = \begin{cases} 1, & \text{if } (i, j) \text{ is on the path to which } k \text{ was assigned} \\ 0, & \text{otherwise.} \end{cases}$$

**Remark 4.** For a demand point  $k' \in Cl_k$  and a facility  $i \in V_1, X_{k'i} = 1$  if and only if  $X_{ki} = 1$ . Moreover, for each  $(i, j) \in V_l \times V_{l+1}, 1 \leq l \leq n-1, Z_{k'ij} = 1$  if and only if  $Z_{kij} = 1$ .

It remains to prove that  $E(c(X, Z) + f(Y)) \leq 3C_{LP}$ .

**Lemma 5.** (a) For each  $k \in CC$ , a facility  $i \in N(k) \cap V_l$  will be opened with probability  $x_{ki}^*$ , if  $l = 1$ , and with probability  $\sum_{j \in V_{l-1}} z_{kji}^*$ , if  $2 \leq l \leq n$ .

(b) The expected cost of opening facilities satisfies:  $E(f(Y)) \leq \sum_{i \in F} f_i y_i^*$ .

**Proof.** (a) Recall that the algorithm opens only facilities which are in the neighborhood of some cluster center. In Lemma 3 we have proved that each facility is in at most one cluster. Consider a cluster center  $k \in CC$ . For facilities on the first level the claim follows directly from the algorithm. The probability of opening a facility  $i$  in  $N(k) \cap V_2$  is:

$$\begin{aligned} P(i \text{ is opened}) &= \sum_{j \in V_1 \cap N(k)} P(Y_i = 1 | Y_j = 1) P(Y_j = 1) \\ &= \sum_{j \in V_1} \frac{z_{kji}^*}{\sum_{i \in V_2} z_{kji}^*} x_{kj}^* = \sum_{j \in V_1} z_{kji}^*, \end{aligned}$$

where for the last equality we have used (7).

Suppose that each facility  $i \in N(k) \cap V_l$  on a level  $2 \leq l < n$  is opened with probability  $\sum_{j \in V_{l-1}} z_{kji}^*$  and consider a facility  $i'$  in  $N(k)$  on level  $l + 1$ . This facility is opened with probability:

$$\begin{aligned} P(i' \text{ is opened}) &= \sum_{i \in V_l \cap N(k)} P(Y_{i'} = 1 | Y_i = 1) P(Y_i = 1) \\ &= \sum_{i \in V_l} \frac{z_{kii'}^*}{\sum_{i' \in V_{l+1}} z_{kii'}^*} \sum_{j \in V_{l-1}} z_{kji}^* = \sum_{i \in V_l} z_{kii'}^*, \end{aligned}$$

where for the last equality we have used (8).

(b) Since the neighborhoods of two cluster centers are disjoint, each facility is opened at most once. Constraints (9) and (10), together with (a) imply that for each facility  $i \in F$ ,  $P(Y_i = 1) \leq y_i^*$ . The expected cost for opening facilities can then be bounded by:

$$E(f(Y)) = \sum_{i \in F} f_i P(Y_i = 1) \leq \sum_{i \in F} f_i y_i^*. \quad \blacksquare$$

Next we will bound the transportation costs.

**Lemma 6.** (a) The probability that the edge  $(i, j) \in V_l \times V_{l+1}$ ,  $1 \leq l \leq n - 1$  is used by a cluster center  $k$  is  $P(Z_{kij} = 1) = z_{kij}^*$ .

(b) For a cluster center  $k \in D$ , the expected transportation costs are  $C_k$ .

For a demand point  $k' \in (Cl_k \cap D) \setminus \{k\}$ , the expected transportation costs are at most  $2v_{k'} + C_{k'}$ .

**Proof.** (a) Let  $k \in CC$ . Lemma 5 together with (8) imply that the probability that edge  $(i, j) \in V_l \times V_{l+1}$ ,  $1 \leq l \leq n - 1$  is used by  $k$  can be calculated as follows:

$$\begin{aligned} P(Z_{kij} = 1) &= P(Y_j = 1 | Y_i = 1) P(Y_i = 1) \\ &= \frac{z_{kij}^*}{\sum_{j \in V_{l+1}} z_{kij}^*} \sum_{j \in V_{l+1}} z_{kij}^* = z_{kij}^*. \end{aligned}$$

(b) For a cluster center  $k \in D$ , the expected transportation costs are

$$\begin{aligned} E \left( \sum_{i \in V_1} d_k c_{ki} X_{ki} + \sum_{l=1}^{n-1} \sum_{(i,j) \in V_l \times V_{l+1}} d_k c_{ij} Z_{kij} \right) &= \sum_{i \in V_1} d_k c_{ki} P(X_{ki} = 1) + \sum_{l=1}^{n-1} \sum_{(i,j) \in V_l \times V_{l+1}} d_k c_{ij} P(Z_{kij} = 1) \\ &= \sum_{i \in V_1} d_k c_{ki} x_{ki}^* + \sum_{l=1}^{n-1} \sum_{(i,j) \in V_l \times V_{l+1}} d_k c_{ij} z_{kij}^* = C_k. \end{aligned}$$

(c) Consider a demand point  $k' \in (Cl_k \cap D) \setminus \{k\}$ . By the definition of a cluster, there exists a facility  $i_l \in N(k) \cap N(k')$ .

From the definition of a neighborhood and Lemma 2 it follows that there exist two paths  $p = (i_1, \dots, i_n)$  and  $p' = (i'_1, \dots, i'_n)$  such that  $i_l \in p$ ,  $i'_l \in p'$ ,  $k$  is assigned to  $p$  and  $k'$  is assigned to  $p'$ . The transportation costs till facility  $i_l$  along these paths, can be bounded by using Lemma 1:

$$c_{ki_1} + \dots + c_{i_{l-1}i_l} \leq \frac{v_k^*}{d_k} \tag{11}$$

and

$$c_{k'i'_1} + \dots + c_{i'_{l-1}i_l} \leq \frac{v_{k'}^*}{d_{k'}}. \tag{12}$$

Denote by  $dist_{kk'}$  the distance between  $k$  and  $k'$ . By using the triangle inequality,  $dist_{kk'}$  can be bounded by:

$$dist_{kk'} \leq c_{ki_1} + \sum_{s=1}^{l-1} c_{i_s i_{s+1}} + c_{k'i'_1} + \sum_{s=1}^{l-1} c_{i'_s i'_{s+1}} \leq \frac{v_k^*}{d_k} + \frac{v_{k'}^*}{d_{k'}}.$$

The transportation cost of  $k'$  can now be bounded by:

$$\begin{aligned} E\left(\sum_{l=1}^{n-1} \sum_{(i,j) \in V_l \times V_{l+1}} d_{k'} c_{ij} Z_{k'ij} + \sum_{i \in V_1} d_{k'} c_{k'i} X_{k'i}\right) &= \sum_{l=1}^{n-1} \sum_{(i,j) \in V_l \times V_{l+1}} d_{k'} c_{ij} P(Z_{k'ij} = 1) + \sum_{i \in V_1} d_{k'} c_{k'i} P(X_{k'i} = 1) \\ &= \sum_{l=1}^{n-1} \sum_{(i,j) \in V_l \times V_{l+1}} d_{k'} c_{ij} P(Z_{kij} = 1) + \sum_{i \in V_1} d_{k'} c_{k'i} P(X_{ki} = 1) \end{aligned} \tag{13}$$

$$\leq \sum_{l=1}^{n-1} \sum_{(i,j) \in V_l \times V_{l+1}} d_{k'} c_{ij} P(Z_{kij} = 1) + \sum_{i \in V_1} d_{k'} (c_{ki} + dist_{kk'}) P(X_{ki} = 1) \tag{14}$$

$$\begin{aligned} &= \frac{d_{k'}}{d_k} C_k + d_{k'} dist_{kk'} \leq \frac{d_{k'}}{d_k} C_k + d_{k'} \left(\frac{v_k}{d_k} + \frac{v_{k'}}{d_{k'}}\right) \\ &\leq C_{k'} + 2v_{k'}^*, \end{aligned} \tag{15}$$

where for (13) we have used Remark 4, for (14) we have used the triangle inequality, and for (15) we have used that  $\frac{C_k + v_k^*}{d_k} \leq \frac{C_{k'} + v_{k'}^*}{d_{k'}}$ , which follows from the fact that  $k' \in C_k$  and from the way clusters were constructed. ■

We are able now to bound the expected costs of  $(X, Y, Z)$ .

**Theorem 7.** *The expected costs of the solution  $(X, Y, Z)$  found by our algorithm satisfy:*

$$E(c(X, Z) + f(Y)) \leq 3C_{LP} \leq 3C_{OPT}.$$

**Proof.** In Lemma 5 we have proved that:

$$E(f(Y)) \leq \sum_{i \in F} f_i y_i^*.$$

From Lemma 6 and the fact that each demand point is assigned to the path opened in the cluster to which it belongs, follows that the transportation costs can be bounded by

$$\begin{aligned} E(c(X, Z)) &= \sum_{k \in CC} \sum_{k' \in C_k \cap D} E\left(\sum_{l=1}^{n-1} \sum_{(i,j) \in V_l \times V_{l+1}} d_{k'} c_{ij} Z_{k'ij} + \sum_{i \in V_1} d_{k'} c_{k'i} X_{k'i}\right) \\ &\leq \sum_{k \in CC} \left[ C_k + \sum_{k' \in (C_k \cap D) \setminus k} (C_{k'} + 2v_{k'}^*) \right] \\ &\leq \sum_{k \in D} (C_k + 2v_k^*). \end{aligned}$$

Since  $\sum_{k \in D} C_k + \sum_{i \in F} f_i y_i^* = \sum_{k \in D} v_k^* = C_{LP}$ , we conclude that

$$\begin{aligned} E(c(X, Z) + f(Y)) &= E(c(X, Z)) + E(f(Y)) \\ &\leq \sum_{k \in D} C_k + \sum_{i \in F} f_i y_i^* + 2 \sum_{k \in D} v_k^* = 3C_{LP} \leq 3C_{OPT}. \quad \blacksquare \end{aligned}$$

Theorem 7 implies that the algorithm we proposed is a 3-approximation (randomized) algorithm.

*Derandomization.* The 3-approximation algorithm described above can be derandomized, while maintaining the approximation guarantee. A technique often used in derandomization is the method of conditional probabilities (see e.g. [19] for an extensive presentation of the method). The main idea behind the derandomization is to find a solution of lower cost than the expected value. In our problem, we have calculated the expected cost as the sum of the expected costs of all clusters.

Since in each cluster  $Cl_k$ ,  $k \in CC$  only facilities along one path  $p$  were opened, the costs incurred by the cluster were the costs incurred for opening facilities along  $p$  and the transportation costs of each demand point in the cluster along the respective path (we will shortly call these costs *the cost of  $p$* ). We have shown that in a cluster, the expected cost is bounded by

$$\sum_{i \in Cl_k \cap F} f_i y_i^* + \sum_{k' \in Cl_k \cap D} (C_{k'} + 2v_{k'}^*). \quad (16)$$

Clearly, in each cluster  $Cl_k$ , there must exist a path  $p'$  such that the transportation costs of all demand points in the cluster along  $p'$  and the costs of opening facilities on  $p'$  are no larger than the bound in (16). One can find such a path in polynomial time via a shortest path algorithm in the graph defined on  $F \cap Cl_k \cup \{k\}$  with the distances  $\tilde{c}_{ki} = f_i + \sum_{k' \in Cl_k \cap D} c_{k'i}$  for each  $i \in V_1$  and  $\tilde{c}_{ij} = c_{ij} + f_j$  for each  $(i, j) \in V_1 \times V_{l+1}$ ,  $l = 1, \dots, n-1$ . The solution obtained by opening facilities along these shortest paths in every cluster, and by assigning all the demand points in a cluster to the corresponding path, yields lower cost than the expected value. Thus, we have a deterministic 3-approximation algorithm for the MFLP.

#### 4. Conclusions

In this paper we have proposed a new integer programming formulation for the MFLP, which has an LP-relaxation with a polynomial number of constraints and variables. We have also shown how one can use this formulation to design a 3-approximation algorithm for the MFLP. Since many algorithms for facility location problems use LP based techniques, (LP-rounding, primal–dual, dual fitting), it would be interesting to further investigate if the new LP-relaxation may be used in decreasing the approximation guarantee for the MFLP.

#### Acknowledgements

We are indebted to the referees for their constructive remarks, which improved considerably the presentation of the paper.

In the Netherlands, the 3 universities of technology have formed the 3TU.Federation. This article is the result of the joint research in the 3TU.Centre of Competence NIRICT (Netherlands Institute for Research on ICT).

The first author's research was carried out while the author was affiliated with the Eindhoven University of Technology.

#### References

- [1] K. Aardal, M. Labbé, J. Leung, M. Queyranne, On the two-level uncapacitated facility location problem, *INFORMS J. Comput.* 8 (1996) 289–301.
- [2] K. Aardal, F.A. Chudak, D.B. Shmoys, A 3-approximation algorithm for the  $k$ -level uncapacitated facility location problem, *Inf. Process. Lett.* 72 (5–6) (1999) 161–167.
- [3] A. Ageev, Improved approximation algorithms for multilevel facility location problems, *Oper. Res. Lett.* 30 (5) (2002) 327–332.
- [4] A.A. Ageev, Y. Ye, J. Zhang, Improved combinatorial approximation algorithms for the  $k$ -level facility location problem, *SIAM J. Discrete Math.* 18 (1) (2004) 207–217.
- [5] 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.
- [6] A.F. Bumb, W. Kern, A simple dual ascent algorithm for the multilevel facility location problem, in: *Proceedings of the 4th International Workshop on Approximation Algorithms for Combinatorial Optimization*, in: LNCS, vol. 2129, Springer, 2001, pp. 55–62.
- [7] J. Byrka, An optimal bifactor approximation algorithm for the metric uncapacitated facility location problem, in: *Proceedings of the 10th International Workshop APPROX 2007 and 11th International Workshop RANDOM 2007*, in: LNCS, vol. 4627, Springer, 2007, pp. 29–43.
- [8] M. Charikar, S. Guha, Improved combinatorial algorithms for the facility location problem and  $k$ -median location problems, in: *Proceedings of the 40th Annual Symposium on Foundations of Computer Science*, 1999, pp. 378–388.
- [9] F.A. Chudak, D.B. Shmoys, Improved approximation algorithms for the uncapacitated facility location problem, *SIAM J. Comput.* 33 (1) (2003) 1–25.
- [10] G. Cornuejols, G.L. Nemhauser, L.A. Wolsey, The uncapacitated facility location problem, in: P. Mirchandani, R. Francis (Eds.), *Discrete Location Theory*, John Wiley and Sons, New York, 1990, pp. 119–171.
- [11] N. Edwards, Approximation algorithms for the multilevel facility location problem, Ph.D. Thesis, School of operations Research and Industrial engineering, Cornell University, Ithaca, NY, 2001.
- [12] S. Guha, S. Khuller, Greedy strikes back: Improved facility location algorithms, *J. Algorithms* 31 (1) (1999) 228–248.
- [13] S. Guha, A. Meyerson, K. Munagala, Hierarchical placement and network design problems, in: *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer science*, 2000, pp. 603–612.
- [14] K. Jain, V. Vazirani, Approximation algorithms for metric facility location and  $k$ -median problems using the primal-dual schema and Lagrangian relaxation, *J. ACM* 48 (2001) 274–296.
- [15] K. Jain, M. Mahdian, E. Markakis, A. Saberi, V.V. Vazirani, Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP, *J. ACM* 50 (6) (2003) 795–824.
- [16] M.R. Korupolu, C.G. Plaxton, R. Rajaraman, Analysis of a local search heuristic for facility location problems, *J. Algorithms* 37 (2000) 146–188.
- [17] M. Mahdian, Y. Ye, J. Zhang, Approximation algorithms for metric facility location problems, *SIAM J. Comput.* 36 (2) (2006) 411–432.
- [18] A. Meyerson, K. Munagala, S. Plotkin, Cost-distance: Two-metric network design, in: *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science*, 2000, pp. 624–630.
- [19] R. Motwani, P. Raghavan, *Randomized Algorithms*, Cambridge University Press, 1995.
- [20] D. Shmoys, Approximation algorithms for facility location problems, in: *Proceedings of 3rd International Workshop of Approximation Algorithms for Combinatorial Optimization*, in: Springer-Verlag LNCS, vol. 1913, 2000, pp. 27–33.
- [21] D. Shmoys, E. Tardos, K. Aardal, Approximation algorithms for facility location problems, in: *Proceedings of the 29th ACM Symposium on Theory of Computing*, 1997, pp. 265–274.
- [22] J. Zhang, Approximating the two-level facility location problem via a quasi-greedy approach, *Math. Program.* 108 (1) (2006) 159–176.