A 1.85-approximation algorithm for the fault tolerant facility location problem with equal connectivity requirements

A. Bumb and W. Kern
Faculty of Mathematical Sciences
University of Twente
P.O. Box 217, 7500 AE Enschede
The Netherlands
email a.f.bumb@math.utwente.nl, kern@math.utwente.nl

Abstract

Recently, several approximation algorithms were proposed for the uncapacitated facility location problem (UFLP). We consider the fault tolerant version of this problem, in which every demand point j is required to be connected to r different open facilities, where $r \geq 1$. For this problem, we first propose a 3-approximation algorithm, based on the primal-dual technique. Then we show how the performance guarantee can be reduced to 1.85 by combining scaling with a greedy improvement technique of the solution.

Key words: Combinatorial optimization, Approximation algorithms, Facility location

1 Introduction

The uncapacitated facility location problem, denoted by UFLP, has been extensively studied in the OR-literature. In this problem one has information about the costs of building and opening facilities at a set F of locations and about the demand of a set D of clients (demand points). The goal is to decide where to open facilities and how to assign each client to an open facility such that the total cost of building/opening the facilities and of satisfying all the demand is minimized.

The uncapacitated facility location problem has proved to be a very useful tool in modeling many network design or location problems, such as location of plants or warehouses

([21], [6], [10]), placement of caches [9] and agglomeration of traffic [3].

However, in many situations, it is required that the system should be resistant to failures of nodes and links. In this case, the uncapacitated facility location problem cannot be used, since it does not provide any guarantee regarding the cost of assigning a client to other facilities.

The fault tolerant facility location problem, denoted by TFLP, differs from the UFLP in the fact that each demand point j has to be assigned to a number r_j of open facilities. We will call the number r_j the connectivity requirement of demand point j. The assignment cost of j is the sum of the corresponding r_j assignments. When all the r_j's are 1, one obtains the uncapacitated facility location problem. We will consider only the metric variant of the problem, in which the distance function between demand points and facilities is nonnegative, symmetric and satisfies the triangle inequality.

Throughout this paper, a ρ-approximation algorithm is 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.

In recent years, several approximation algorithms for the UFLP ([2, 4, 11, 13, 14, 17, 18, 19]) and for its variants ([1, 15, 17, 20])) have been proposed. The currently known best performance guarantee for the UFLP is 1.52, due to Mahdian, Ye and Zhang [18]. Guha and Khuller [8] and Sviridenko [22] have proved that the UFLP is Max SNP-hard and that the existence of an ρ-approximation algorithm for $\rho < 1.463$ implies $P = NP$.

Although a very natural extension of the UFLP, the TFLP did nor receive too much attention in the literature. The first approximation algorithm for this problem was developed by Jain and Vazirani [12] and has a performance guarantee of $O(logR)$, where $R = \max_{j\in D} r_j$. The first constant approximation algorithm for the TFLP was developed by Guha, Meyerson and Munagala [9], who proposed a 3.16-approximation algorithm based on the filtering technique developed by Lin and Vitter in [16] and linear programming rounding. Combining filtering with greedy improvement, they reduce the performance guarantee to 2.47. Later on, Mahdian *et al.* [17] proposed a 1.86-approximation algorithm based on dual fitting for the special case of TFLP in which all connectivity requirements are equal.

In this paper, we first propose a 3-approximation algorithm for the TFLP with equal connectivity requirements, *i.e.* $r_j = r$, for every demand point $j \in D$. The algorithm is based on the primal-dual technique end extends the algorithm proposed by Jain and Vazirani [13] for the UFLP. Similar ideas were used by Erlenkotter in the branch and bound algorithm he proposed for the case $r = 1$ in [7]. By combining scaling with the greedy improvement technique in the manner of [4] and [9] we reduce the performance guarantee to 1.85.

2 The fault tolerant facility location problem with equal connectivity requirements

Consider a complete bipartite graph $G = (V, E)$ with $V = F \cup D$ and $E = F \times D$. The set F is the set of possible *facility locations* and D is the set of *demand points.* We are given a *service (transportation) cost vector* $c \in R_+^{|E|}$ (i.e., the cost of transporting a unit of demand from i to j is c_{ij}) and an *opening cost vector* $f \in R_+^{|E|}$. (i.e., opening a facility at $i \in F$ incurs a cost $f_i \geq 0$). We assume that c is induced by a metric on V. Every demand point $j \in D$ must be connected to r open facilities. r is called the *connectivity requirement* of j. The goal is to decide which facilities to open and how to connect the demand points to the required number of facilities such that the total cost (the opening costs plus the connection costs) is minimized.

For simplicity, we will assume that each $j \in D$ demands exactly one unit from each of the r facilities to which it gets connected. All the results presented in this chapter can be easily generalized (by an appropriate scaling) to arbitrary demands. Furthermore, we will assume that $f_i > 0$ for every $i \in F$. This is not a restrictive assumption, since if there is a facility i with no opening cost, we can open it after the algorithm terminates and assign to it all the demand points for which a reassignment to i would mean a decrease in the service cost.

One way of formulating the $TFLP$ as an integer program is the following. Let y_i ($i \in F$) be variables indicating whether facility $i \in F$ is open and let x_{ij} ($i \in F, j \in D$) indicate whether demand point j is connected to facility i. We let

$$c(x) := \sum_{i \in F, j \in D} c_{ij} x_{ij}$$

and

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

The $TFLP$ is then equivalent with

$$
\begin{aligned}
(\mathbf{P}_{\text{int}}) \qquad & \text{minimize } c(x)+f(y) \\
& \text{subject to} \sum_{i\in F} x_{ij} \geq r, \quad \text{for each } j \in D \qquad (1)\\
& x_{ij} \leq y_i, \quad \text{for each } i \in F \text{ and } j \in D \qquad (2)\\
& x_{ij}, y_i \in \{0,1\}, \text{ for each } i \in F \text{ and } j \in D
\end{aligned}
$$

Constraints (1) ensure that each demand point $j \in D$ is connected to at least r facilities and constraints (2) ensure that each of these facilities must be open.

We consider the following LP-relaxation of $(\mathbf{P}_{\text{int}})$:

$$
\begin{aligned}
(\mathbf{P}) \qquad & \text{minimize } c(x) + f(y) \\
& \text{subject to } (1), (2) \\
& x_{ij} \leq 1, \text{ for each } i \in F \text{ and } j \in D \qquad (3)\\
& x_{ij}, y_i \geq 0, \text{ for each } i \in F \text{ and } j \in D
\end{aligned}
$$

Note that constraints (3) are necessary to ensure that for each $j \in D$, the facilities to which j is assigned are distinct. Denote by C_{OPT} and C_{LP}, the optimal values of $(\mathbf{P}_{\text{int}})$, respectively $(\mathbf{P})$. Clearly, $C_{OPT} \geq C_{LP}$.

Let v_j, t_{ij} and z_i, denote the dual variables corresponding to constraints $(1), (2)$ and (3) in $(\mathbf{P})$. The dual program corresponding to the linear programming relaxation $(\mathbf{P})$ is

$$
\begin{aligned}
(\mathbf{D}) \qquad & \text{maximize } \sum_{j\in D} r v_j - \sum_{i\in F}\sum_{j\in D} z_{ij} \\
& \text{subject to } v_j - t_{ij} - z_{ij} \leq c_{ij}, \text{ for each } i \in F \text{ and } j \in D \qquad (4)\\
& \sum_{j\in D} t_{ij} \leq f_i, \text{ for each } i \in F \qquad (5)\\
& v_j, z_{ij}, t_{ij} \geq 0, \text{ for each } i \in F \text{ and } j \in D.
\end{aligned}
$$

To derive feasible primal and dual solutions we will use the complementary slackness conditions for optimal solutions (x^*, y^*), respectively (v^*, t^*, z^*) of $(\mathbf{P})$, respectively $(\mathbf{D})$. The primal and dual complementary slackness conditions are:

(T1) If $x^*_{ij} > 0$ then $v^*_j - t^*_{ij} - z^*_{ij} = c_{ij}$.

(T2) If $y^*_i > 0$ then $\sum_{j \in D} t^*_{ij} = f_i$.

(T3) If $v^*_j > 0$ then $\sum_{i \in F} x^*_{ij} = r$.

(T4) If $t^*_{ij} > 0$ then $x^*_{ij} = y^*_i$.

(T5) If $z^*_{ij} > 0$ then $x^*_{ij} = 1$.

3 A 3-approximation algorithm

The main idea of the 3-approximation algorithm is to construct, in polynomial time, a primal integer feasible solution (x, y) and a dual feasible solution (v, t, z) that satisfy the complementary slackness conditions (T2)-(T5) and a the following relaxation of (T1):

$(T1')$ If $x_{ij} > 0$, then $\frac{1}{3}c_{ij} \le v_j - t_{ij} - z_{ij} \le c_{ij}$.

The cost of such a primal solution (x, y) can then be bounded by

$$\begin{aligned}
c(x) + 3f(y) &= \sum_{i \in F, j \in D} c_{ij}x_{ij} + 3\sum_{i \in F} f_i y_i = \sum_{i \in F, j \in D} c_{ij}x_{ij} + 3\sum_{i \in F, j \in D} t_{ij}x_{ij} \\
&= \sum_{j \in D}\sum_{i \in F}(c_{ij} + 3t_{ij})x_{ij} \le 3\sum_{j \in D}\sum_{i \in F}(v_j - z_{ij})x_{ij} = 3(\sum_{j \in D} rv_j - \sum_{j \in D}\sum_{i \in F} z_{ij}) \\
&\le 3C_{LP} \le 3C_{OPT},
\end{aligned}$$

where the second equality is based on $(T2)$ and $(T4)$, the first inequality is based on $(T1')$ and the last equality on $(T3)$ and $(T5)$.

Before describing the algorithm in detail, we introduce some terminology with respect to a primal, respectively dual feasible solution (x, y), respectively (v, t, z).

A demand point $j \in D$ *reaches* a facility $i \in F$ if

$$c_{ij} \le v_j.$$

Denote by F_j the set of facilities reached by j.

If $t_{ij} > 0$, we say that j is *willing to contribute* with t_{ij} to opening facility i. A facility i is *fully paid* when

$$\sum_{j \in D} t_{ij} = f_i.$$

For every demand point j let the *neighborhood* N_j of j be the set

$$N_j = \left\{ j' \in D | F_j \cap F_{j'} \neq \emptyset \text{ and } v_{j'} \leq v_j \right\}.$$

We will call a demand point j *saturated* when there is a $j' \in N_j$ such that j' reaches r open facilities.

Remark 1 *Note that, complementary slackness conditions (T1) and (T2) imply that, with respect to an optimal primal, respectively optimal dual solution, each demand point reaches the facilities to which it is assigned by the optimal solution and each open facility is fully paid.*

The definition of a saturated demand point is justified by the following lemma.

Lemma 1 *Let* j, j' *be two demand points such that* $j' \in N_j$. *For every* $i \in F_{j'}$ *we have* $c_{ij} \leq 3v_j$.

Proof. Since $j' \in N_j$, $F_j \cap F_{j'} \neq \emptyset$ and $v_{j'} \leq v_j$. Let $i' \in F_{j'} \cap F_j$. The definition of the sets F_j, respectively $F_{j'}$ implies that $c_{i'j} \leq v_j$, $c_{i'j'} \leq v_{j'}$ and $c_{ij'} \leq v_{j'}$. The triangle inequality thus yields

$$c_{ij} \leq c_{i'j} + c_{i'j'} + c_{ij'} \leq 2v_{j'} + v_j \leq 3v_j.$$

■

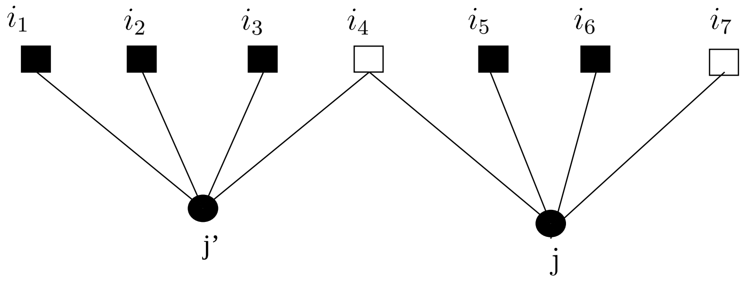
Remark 2 *Lemma 1 is essential for the algorithm. If we connect a demand point* j *to a facility* $i \in F_j$ *the service cost can be bounded by* v_j. *Lemma* (1) *tells us that if we connect* j *to facilities in* N_j *the service cost may increase by at most a factor of* 3. *Hence, a saturated demand point* j *has at least* r *open facilities at distance at most* $3v_j$.

A closed facility i is *blocked* by a demand point j if the following conditions are satisfied

- $\left| \left\{ i' \in F_j | i' \text{ is open and } t_{i'j} + z_{i'j} > 0 \right\} \right| = r$
- $t_{ij} + z_{ij} > 0$.

Observe that if a demand point j blocks a facility i, then j is saturated, since F_j contains r open facilities and $j \in N_j$.

Example 1 *In Figure 1 we give an example of a saturated demand point. In this example,* $r = 3$, *facilities* $\{i_1, i_2, i_3, i_5, i_6\}$ *are open,* $F_{j'} = \{i_1, i_2, i_3, i_4\}$ *and* $F_j = \{i_4, i_5, i_6, i_7\}$. *Since* r *facilities in* $F_{j'}$ *are open, demand point* j' *is saturated. If* $v_{j'} \leq v_j$, *then* $j' \in N_j$ *and* j *is saturated as well. If* $t_{ij'} + z_{ij'} > 0$ *for all* $i \in \{i_1, i_2, i_3, i_4\}$, *then facility* i_4 *is blocked by* j'.



■ are open facilities and □ are closed

Figure 1: Example of saturated demand point

As an indirect consequence of the definitions of the neighborhood and a blocked facility we obtain the following lemma.

Lemma 2 *If i is a facility blocked by a demand point j' then any demand point j with $v_{j'} \leq v_j$ that reaches i is saturated.*

The algorithm can be described as follows.

Start with $v \equiv t \equiv z \equiv 0$. Initially, all the demand points are non-saturated and no facility is open. Until all the demand points become saturated, for all non-saturated demand points we proceed as follows. Increase v_j uniformly until j becomes saturated. For all not-fully paid facilities i reached by j increase t_{ij} uniformly, until either f_i is fully paid or j becomes saturated. For all the fully paid facilities i reached by j, increase z_{ij} uniformly until j becomes saturated. Open all the fully paid and not blocked facilities. In other words, the algorithm proceeds as follows:

```
LET $v \equiv t \equiv z \equiv 0$.
UNTIL all $j \in D$ are saturated DO
    LET $\Delta_1 = \min \left\{ \frac{f_i - \sum\limits_{j \in D} t_{ij}}{|\{j : i \in F_j\}|} | i \in F, \text{i is not fully paid and not blocked} \right\}$
    LET $\Delta_2 = \min\{c_{ij} - v_j | c_{ij} \geq v_j\}$
    LET $\Delta = \min\{\Delta_1, \Delta_2\}$
    FOR all non-saturated $j \in D$ DO
          $v_j = v_j + \Delta$
        FOR all not fully paid facilities $i \in F$ reached by $j$
          $t_{ij} = t_{ij} + \Delta$
        FOR all fully paid facilities $i \in F$ reached by $j$
          $z_{ij} = z_{ij} + \Delta$.
    Open all fully paid, not blocked facilities $i \in F$
```

Remark 3 *If in an iteration* $\Delta = \Delta_1 = \frac{f_{i^*} - \sum\limits_{j \in D} t_{i^*j}}{|\{j : i^* \in F_j\}|}$ *for some* $i^* \in F$ *it means that facility* i^* *becomes fully paid, and if* $\Delta = \Delta_2 = c_{i^*j^*} - v_{j^*}$*, means that demand point* j^* *reaches facility* i^*.

Example 1 We will illustrate the construction of the dual solution on the following example.

Let $|D| = 3$ and $|F| = 6$. The costs of opening the facilities are given by the vector $f = (100, 80, 70, 110, 150, 300)$. The service costs are given in the following table.The connectvity requirement is $r = 2$.

F D	1	2	3	4	5	6
1	20	30	20	150	70	130
2	120	200	30	20	70	140
3	110	150	120	140	70	130

We start with $v \equiv t \equiv z \equiv 0$.

Iteration 1 All demand points are non-saturated. $\Delta_1 = \infty$, as no facility is reached by a demand point, $\Delta_2 = 20$, hence $\Delta = 20$. The v-vector becomes $v = (20, 20, 20)$. First demand point reaches facilities $F_1 = \{1, 3\}$, second demand point reaches $F_2 = \{4\}$ and the third does not reach any facility, hence $F_3 = \emptyset$.

Iteration 2 $\Delta_1 = \min\{f_1, f_3, f_4, \infty\} = \min\{100, 70, 110, \infty\} = 70$ and $\Delta_2 = 10$, hence $\Delta = 10$. The dual solution becomes $v_1 = v_2 = v_3 = 30$ and $t_{11} = t_{31} = t_{42} = 10$. Demand point 1 reaches facility 2 and demand point 2 reaches facility 3, *i.e.*, $F_1 = F_1 \cup \{2\} = \{1, 2, 3\}$ and $F_2 = F_2 \cup \{3\} = \{3, 4\}$.

Iteration 3 $\Delta_1 = \min\left\{f_1 - t_{11}, f_2, \frac{f_3 - t_{31}}{2}, f_4 - t_{42}, \infty\right\} = \min\{90, 80, 30, 100, \infty\}$ and $\Delta_2 = 40$. This implies $\Delta = 30$, and that facility 3 will become fully paid. The v-vector changes in $v_1 = v_2 = v_3 = 60$ and the contributions of demand points to facilities are $t_{11} = t_{31} = t_{42} = 40$ and $t_{21} = t_{32} = 30$. Facility 3 is opened.

Iteration 4 $\Delta_1 = \min\{f_1 - t_{11}, f_2 - t_{21}, f_4 - t_{42}, \infty\} = \min\{60, 50, 70, \infty\} = 50$ and $\Delta_2 = \{10\}$. Hence $\Delta = 10$ and $v_1 = v_2 = v_3 = 70$. Clearly, all the demand points reach facility 5, therefore, $F_1 = \{1, 2, 3, 5\}, F_2 = \{3, 4, 5\}, F_3 = \{5\}$. Moreover, $t_{11} = t_{42} = 50$ and $t_{21} = 40$. Facility 3 being fully paid, $z_{32} = 10$ and $z_{31} = 10$.

Iteration 5 $\Delta_1 = \min\left\{f_1 - t_{11}, f_2 - t_{21}, f_4 - t_{42}, \frac{f_5}{3}, \infty\right\} = \min\{50, 40, 60, 50, \infty\} = 40$ and $\Delta_2 = 50$. Hence, $\Delta = 40$. v becomes $v_1 = v_2 = v_3 = 110$. The contribution to opening facilities changes as follows $t_{11} = 90, t_{21} = 80, t_{42} = 90, t_{51} = t_{52} = t_{53} = 40$. The z-vector becomes $z_{31} = z_{32} = 50$, all the other components being zero. Clearly, facility 2 becomes fully paid and is opened. Demand point 1 is saturated, because it reaches the fully paid facilities 2 and 3. Therefore the other facilities in F_1, namely 1 and 5 are blocked. Since demand points 2 and 3 reach the blocked facility 5, they are saturated as well, cf. Lemma 2.

We summarize the properties of the dual solution (v, t, z) constructed in the following lemma.

Lemma 3 *(i) For every demand point j and facility i reached by j, $v_j - t_{ij} - z_{ij} = c_{ij}$. For every facility i that was not reached by j, $t_{ij} = z_{ij} = 0$.*

(ii) (v, t, z) is a feasible dual solution.

(iii) Every facility i with $\sum_{j \in D} z_{ij} > 0$ is opened.

(iv) For every open facility i and demand points j,j' such that $i \in F_{j'}$ and $j' \in N_j$, the following holds

$$\frac{1}{3} c_{ij} \leq v_j - t_{ij} - z_{ij} \leq c_{ij}. \tag{6}$$

Moreover, for every demand point j, there are at least r open facilities i satisfying (6).

Proof. (i) For all the demand points j and facilities i, first we have increased the variable v_j and maintained $t_{ij} = z_{ij} = 0$ until $v_j = c_{ij}$. Then, until i became fully paid, we increased simultaneously t_{ij} and v_j, thus maintaining $v_j - t_{ij} = c_{ij}$ and $z_{ij} = 0$. If after i became fully paid, j was still non-saturated, we increased simultaneously v_j and z_{ij} and kept t_{ij} fixed.

(ii) We have actually proved in (i) that (v, t, z) satisfies (4). Since for every facility i we stop increasing the dual variables t_{ij} when $\sum\limits_{j \in D} t_{ij} = f_i$, (5) is satisfied as well.

(iii) When i became fully paid, j did not became saturated otherwise we would not have increased z_{ij}. Facility i was not blocked as well, otherwise j would have become saturated, cf. Lemma 2. This implies that facility i was opened.

(iv) If j reached facility i, $v_j - t_{ij} - z_{ij} = c_{ij}$ and (6) holds. If j did not reach i, $t_{ij} = z_{ij} = 0$ and (6) follows from Lemma 1. The second part of (iv) follows from the definition of a saturated point. ■

Now we proceed to construct a primal solution (x, y). For every facility $i \in F$,

$$y_i = \begin{cases} 1, & \text{if facility } i \text{ is opened} \\ 0, & \text{otherwise} \end{cases}.$$

From Lemma 3 we know that there are r facilities satisfying (6) for each $j \in D$. Assign each j to them and set

$$x_{ij} = \begin{cases} 1, & \text{if } j \text{ is assigned to } i. \\ 0, & \text{otherwise} \end{cases}.$$

Example 2 Consider the dual solution constructed in Example 1. The corresponding primal solution is given by $y = (0, 1, 1, 0, 0, 0)$ and by $x_{3j} = x_{2j} = 1$, for all $j \in 1, 2, 3$. All the other components of x are zero.

It is easy to see that the primal and the dual solution constructed satisfy (T1') and (T2)-(T5). Since every demand point j is assigned to facilities for which (6) is true, we conclude that (x, y) and (v, t, z) satisfy (T1'). (T2) is implied by the fact that only fully paid facilities are open. (T3) is verified for each $j \in D$. (T4) follows from the fact that every demand point j was assigned to all open facilities i to which he was willing to contribute. Finally, since every $j \in D$ was assigned to all open facilities $i \in F$ with $z_{ij} > 0$, (T5) is true as well.

Tight example

The following infinite family of examples shows that the analysis of our algorithm is tight. There are $n + 1$ demand points and two facilities 1 and 2. The connectivity requirement is $r = 1$. The costs of opening the facilities are $f_1 = \frac{1}{n}$ and $f_2 = 1 + \frac{2}{n}$. The service costs are as follows: $c_{11} = 1$ and $c_{1j} = 3$ for $j \in D, j > 1$ and $c_{2j} = 1$, for each $j \in D$. Clearly, in the optimal solution facility 2 is opened and all the demand points are assigned to it. The cost of the optimal solution is then $n + 2 + \frac{1}{n}$. The primal dual algorithm, however, will construct the following dual solution: $v_j = 1 + \frac{1}{n}, t_{ij} = \frac{1}{n}, z_{ij} = 0$, for each $i \in F, j \in D$. Facility 1 is fully paid, demand point 1 reaches bothe facility, hence facility 2 is blocked. All the other demand points reach the second facility. Hence, the primal solution will open facility 1 and will assign all the demand points to it. The cost of the primal solution will be $3n + 1 + \frac{1}{n}$. For n large, the ratio between the solution constructed by the algorithm and the optimal solution approaches 3.

Running time

The analysis of the algorithm is similar to the one in [13]. Let $n_d = |D|$, $n_f = |F|$ and $n = n_d + n_f$. Sort all the edges by increasing cost. This gives the order and the time when demand points reach facilities. This operation can be done in time $O(n^2 \log n)$.

For each facility i, we maintain the number of demand points that are currently willing to contribute towards it, and the *anticipated time*, T_i, at which it would become fully paid if no other event happens in between. Initially, all the $T_i's$ are infinite and the contribution of each demand point to each facility is equal to zero. The $T_i's$ are maintained in a binary heap so that we can update each one and find the current minimum in $O(\log n_f)$ time. Two types of events happen and they lead to the following updates.

- A demand point j reaches a facility i

-If j becomes saturated, for all the facilities that had j as contributor we need to decrease the number of contributors and recompute the anticipated time at which it will become fully paid.

-If j does not become saturated, then if facility i is not fully paid, it gets one more demand point willing to contribute towards its cost and we update T_i accordingly. The new T_i can be calculated in constant time.

- A facility i becomes fully paid. For all the demand points that become saturated we will execute the first case of the previous event, i.e., update the not fully paid facilities towards which they were contributing.

Note that every edge (i, j) is considered at most twice. First when j reaches i, and then when demand point j becomes saturated. For each consideration of this edge, we will do $O(\log n_f)$ work. Hence, we have proved that

Theorem 2 *The algorithm presented above constructs in time* $O(n^2 \log n)$ *a solution* (x, y) *of cost* $c(x) + 3f(y) \leq 3C_{LP} \leq 3C_{OPT}$.

4 Improved approximation algorithm

In this section we reduce the performance guarantee of the algorithm presented in Section 3, by using scaling and the greedy improvement technique in a way similar to [4] and [9]. We will use the term *adding* a facility i to a solution for opening i and assigning to i all the demand points for which the assignment cost decreases by this operation, while maintaining the property that each demand point is assigned to r different facilities.

Define the *Gain(i)* of a facility i to be the decrease in total cost (*i.e.*, decrease in assignment cost minus f_i) of the solution if facility i is opened. The *gain ratio* associated to facility i is $\frac{Gain(i)}{f_i}$.

Denote by $(\widehat{x}, \widehat{y})$ the optimal solution to $(\mathbf{P}_{\text{int}})$. First we describe the greedy improvement technique.

The greedy improvement technique

The greedy heuristic proceeds iteratively. Start with a solution (x, y) to $(\mathbf{P}_{\text{int}})$. Reassign, if necessary, the demand points to facilities such that each demand point is assigned to the r closest open facilities. As long as there are closed facilities with positive gain, add to the solution the facility with the biggest gain ratio.

The following Lemma is proved in ([9]), though in a slightly different way.

Lemma 4 *Let (x, y) be a solution of $(\boldsymbol{P}_{int})$ such that each demand point is assigned to the r closest open facilities. There exists a facility with gain ratio larger than* $\frac{c(x)-c(\widehat{x})-f(\widehat{y})}{f(\widehat{y})}$.

Proof. For every demand point j, let S_j, respectively $\widehat{S_j}$ denote the sets of facilities to which j is assigned by (x, y), respectively $(\widehat{x}, \widehat{y})$.
Let $\sigma_j : \widehat{S_j} \mapsto S_j$ be a bijection with the property that $\sigma_j(i) = i$ for each $i \in S_j \cap \widehat{S_j}$. Suppose we add a facility $i_0 \in F$ opened by $(\widehat{x}, \widehat{y})$ (i.e., with $\widehat{y_{i_0}} = 1$) to the set of the facilities opened by (x, y). The gain is obtained by assigning to i_0 demand points j for which this modification is profitable. Hence,

$$Gain(i_0) \geq -f_{i_0} + \sum_{j \in D} (c_{\sigma_j(i_0)j} - c_{i_0 j})\widehat{x_{i_0 j}}, \tag{7}$$

since the second term on the right hand side represents the assignment gain that is obtained by connecting all $j \in D$ to i_0 instead of $\sigma_j(i_0)$, if they are connected to i_0 by $\widehat{x}$ (even though this may increase the current assignment cost for some $j \in D$). Note that the gain might be negative. Furthermore, note that (7) also holds in case i_0 is opened by (x, y), since $Gain(i_0) = 0$ and for every j with $\widehat{x_{i_0 j}} = 1$, $c_{\sigma_j(i_0)j} = c_{i_0 j}$ if $i_0 \in S_j$ or $c_{\sigma_j(i_0)j} \leq c_{i_0 j}$ if $i_0 \notin S_j$, due to the fact that j is assigned to the r closest open facilities.

Summing the gain obtained by adding to (x, y) all facilities opened by $(\widehat{x}, \widehat{y})$, and taking into account that since $(\widehat{x}, \widehat{y})$ is integral, (so, $\widehat{x}_{ij}\widehat{y}_i = \widehat{x}_{ij}$), we obtain

$$\sum_{i_0 \in F} Gain(i_0)\widehat{y_{i_0}} \geq -\sum_{i_0 \in F} f_{i_0}\widehat{y_{i_0}} + \sum_{i_0 \in F}\sum_{j \in D}(c_{\sigma_j(i_0)j} - c_{i_0 j})\widehat{x_{i_0 j}}.$$

Since

$$\sum_{i_0 \in F} c_{\sigma_j(i_0)j}\widehat{x_{i_0 j}} = \sum_{i_0 \in \widehat{S_j}} c_{\sigma_j(i_0)j} = \sum_{i_0 \in S_j} c_{i_0 j} = \sum_{i_0 \in F} c_{i_0 j} x_{i_0 j}$$

the total gain is equal to

$$\sum_{i \in F} Gain(i)\widehat{y}_i \geq c(x) - c(\widehat{x}) - f(\widehat{y}). \tag{8}$$

Suppose now that for every facility opened by $(\widehat{x}, \widehat{y})$, $\frac{Gain(i)}{f(i)} < \frac{c(x)-c(\widehat{x})-f(\widehat{y})}{f(\widehat{y})}$. This would imply that

$$\sum_{i \in F} Gain(i)\widehat{y}_i < \frac{c(x) - c(\widehat{x}) - f(\widehat{y})}{f(\widehat{y})} \sum_{i \in F} f_i \widehat{y}_i = c(x) - c(\widehat{x}) - f(\widehat{y}),$$

contradicting (8). Hence, there is a facility i opened by $(\widehat{x}, \widehat{y})$ with gain ratio larger than $\frac{c(x)-c(\widehat{x})-f(\widehat{y})}{f(\widehat{y})}$. ■

Lemma 5 *([4]) The solution obtained by the greedy improvement will incur a total cost of at most*

$$f(y) + f(\widehat{y}) + c(\widehat{x}) + f(\widehat{y}) \max\left\{0, \ln \frac{c(x) - c(\widehat{x})}{f(\widehat{y})}\right\}.$$

Proof. Denote by C_l respectively F_l the assignment cost, respectively the facility cost, associated with the solution obtained in the $l - th$ iteration of the greedy improvement procedure. Initially, $C_0 = c(x)$ and $F_0 = f(y)$. If $C_0 \leq c(\widehat{x}) + f(\widehat{y})$, the lemma is true. Suppose $C_0 > c(\widehat{x}) + f(\widehat{y})$.

Since in every iteration $l, l \geq 1$ we add the facility with the best gain ratio,

$$\frac{C_{l-1} - C_l - (F_l - F_{l-1})}{F_l - F_{l-1}} \geq \frac{C_{l-1} - c(\widehat{x}) - f(\widehat{y})}{f(\widehat{y})}$$

and therefore the cost of a facility i added in the l-th iteration can be bounded by $f_i \leq \frac{C_{l-1}-C_l}{C_{l-1}-c(\widehat{x})} f(\widehat{y})$. Furthermore,

$$\frac{C_{l-1} - C_l}{C_{l-1} - c(\widehat{x})} = 1 - \frac{C_l - c(\widehat{x})}{C_{l-1} - c(\widehat{x})} \leq \ln \frac{C_{l-1} - c(\widehat{x})}{C_l - c(\widehat{x})},$$

where the last inequality is derived from the fact that for all $x, 0 < x \leq 1$, $1 - x \leq \ln(1/x)$. Hence,

$$f_i \leq f(\widehat{y}) \ln \frac{C_{l-1} - c(\widehat{x})}{C_l - c(\widehat{x})}. \tag{9}$$

Note that, as we open facilities in each step, eventually, $C_l \leq c(\widehat{x}) + f(\widehat{y})$ must occur. (If all facilities are opened, the assignment cost is at most $c(\widehat{x})$. Assume $C_m \leq c(\widehat{x}) + f(\widehat{y})$ happens for the first time after m iterations. Since in all subsequent iterations the cost decreases, it is sufficient to bound the total cost after m iterations.

$$\begin{aligned}
C_m + F_m &= C_m + F_0 + (F_1 - F_0) + ... + (F_m - F_{m-1}) \\
&\leq c(\widehat{x}) + f(\widehat{y}) + f(y) + f(\widehat{y}) \ln \frac{C_{m-1} - c(\widehat{x})}{C_m - c(\widehat{x})} \frac{C_{m-2} - c(\widehat{x})}{C_{m-1} - c(\widehat{x})} \cdots \frac{C_0 - c(\widehat{x})}{C_1 - c(\widehat{x})} \\
&= f(y) + c(\widehat{x}) + f(\widehat{y}) + f(\widehat{y}) \ln \frac{C_0 - c(\widehat{x})}{C_m - c(\widehat{x})} \\
&\leq f(y) + c(\widehat{x}) + f(\widehat{y}) + f(\widehat{y}) \ln \frac{c(x) - c(\widehat{x})}{f(\widehat{y})}.
\end{aligned}$$

■

The *improved approximation algorithm* for the $TFLP$ can be described as follows:

- Scale the facility costs by a factor of δ
- Run for the scaled instance the algorithm proposed in Section 3
- Scale back the solution and apply greedy improvement

The following Lemma was proved in [4] for $r = 1$.

Lemma 6 *The solution obtained by the improved approximation algorithm is within a factor of* $\min_\delta \max\left\{2 - \frac{1}{3\delta}, 1 + \frac{1}{\delta}, 1 + \ln(3\delta)\right\}$ *the optimum.*

Proof. Denote by $(\mathbf{P}_{\rm sc})$ the integer program corresponding to the problem with scaled opening facility costs.

Let $(\widehat{x}, \widehat{y})$ be an optimal solution for $(\mathbf{P}_{\rm int})$ and (x', y') the feasible solution for $(\mathbf{P}_{\rm sc})$ obtained in the second step.

Since $(\widehat{x}, \widehat{y})$ is feasible for $(\mathbf{P}_{\rm sc})$, the optimal solution of $(\mathbf{P}_{\rm sc})$ will have a total cost of at most $c(\widehat{x}) + \delta f(\widehat{y})$.

Being obtained by applying the primal-dual algorithm described in Section 3, (x', y') satisfies

$$c(x') + 3\delta f(y') \leq 3(c(\widehat{x}) + \delta f(\widehat{y})). \tag{10}$$

Clearly, (x', y') is also feasible for $(\mathbf{P}_{\rm int})$ and incurs a total cost of $c(x') + f(y')$.

If $c(x') \leq c(\widehat{x}) + f(\widehat{y})$, then (10) implies

$$\begin{aligned} c(x') + f(y') &= \frac{c(x') + 3\delta f(y')}{3\delta} + (1 - \frac{1}{3\delta})c(x') \\ &\leq \frac{c(\widehat{x}) + \delta f(\widehat{y})}{\delta} + (1 - \frac{1}{3\delta})(c(\widehat{x}) + f(\widehat{y})) \\ &\leq (1 + \frac{1}{\delta})c(\widehat{x}) + (2 - \frac{1}{3\delta})f(\widehat{y}). \end{aligned} \tag{11}$$

Suppose now that $c(x') > c(\widehat{x}) + f(\widehat{y})$. Since

$$c(x') \leq 3c(\widehat{x}) + 3\delta f(\widehat{y}) - 3\delta f(y'),$$

from Lemma 5, the cost of the solution after the greedy improvement can be bounded by

$$f(y') + f(\widehat{y}) + c(\widehat{x}) + f(\widehat{y}) \ln \frac{2c(\widehat{x}) + 3\delta f(\widehat{y}) - 3\delta f(y')}{f(\widehat{y})}.$$

This expression, as function of $f(y')$, is maximized for $f(y') = \frac{1}{\delta}c(\widehat{x})$. Choosing $f(y') = \frac{1}{\delta}c(\widehat{x})$, we obtain the following upper bound for the total cost

$$(1 + \frac{1}{\delta})c(\widehat{x}) + (1 + \ln(3\delta))f(\widehat{y}). \tag{12}$$

Thus, from (11) and (12) we conclude that the solution obtained is within $\min_\delta \max\left\{2 - \frac{1}{3\delta}, 1 + \frac{1}{\delta}, 1 + \ln(3\delta)\right\}$ times the optimum. ∎

Note that $\min_\delta \max\left\{2 - \frac{1}{3\delta}, 1 + \frac{1}{\delta}, 1 + \ln(3\delta)\right\}$ is attained for $\delta = 0.7192$ and is equal to 1.85. Hence we have proved

Theorem 3 *The improved approximation algorithm has a performence guarantee 1.85.*

5 Discussion

In this article we have first presented a 3 approximation algorithm for the fault tolerant facility location problem with equal connectivity requirements. Then we have shown how the performance guarantee of this algorithm can be reduced to 1.85 by using standard scaling and greedy improvement. For the 3-approximation algorithm we could find a tight example to show that the analysis of the algorithm is tight, but for the 1.85 approximation algorithm we were not able to do that.

There are several open questions related with the fault tolerance aspect of facility location problems. For the case with arbitrary connectivity requirements, there is still no combinatorial approximation algorithm known and no approximation algorithm with a performance guarantee better than 2.47. In particular, it is not clear whether an algorithm with the same bound as for the case $r = 1$, namely 1.52, can be designed for the fault tolerant facility location problem.

Acknowledgements The authors are grateful to the referees whose insightful comments led us to a significant improvement of the paper.

References

[1] A. Ageev. Improved approximation algorithms for multilevel facility location problems. Operations Research Letters 30 (5) (2002) 327-332.

[2] V. Arya, N. Garg, R. Khandekar, V. Pandit, A. Meyerson, K. Munagala. Local search heuristics for k-median and facility location problems, in: Pro-

ceedings of the 33rd ACM Symposium on Theory of Computing, 2001, pp. 21-29.

[3] M. Andrews and L. Zhang. The access network design problem, in: Proceedings of the 39th Annual Symposium on Foundations of Computer Science, 1998, pp. 40-49.

[4] M. Charikar and 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.

[5] F. A. Chudak. Improved approximation algorithms for uncapacitated facility location problem, in: R. E. Bixby, E. A. Boyd and R. Z. Rios-Mercado (Eds.), Integer Programming and Combinatorial Optimization, Springer-Verlag LNCS Vol. 1412, 1998, pp. 180-194.

[6] 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.

[7] D. Erlenkotter. A dual-based procedure for uncapacitated facility location problem, Operations Research 26 (6) (1978) 992-1009.

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

[9] S. Guha, A. Meyerson, K. Munagala. Improved algorithms for fault tolerant facility location, in: Proceedings of 12th ACM-SIAM Symposium on Discrete Algorithms, 2001, pp. 636-641.

[10] K. Hidaka and H. Okano. An approximation algorithm for a large-scale facility location problem. Algorithmica 35 (2003) 216-224.

[11] K. Jain, M. Mahdian, A. Saberi. A new greedy approach for facility location problem, in: Proceedings of the 34th Annual ACM Symposium on Theory of Computing, 2002, pp. 731-740.

[12] K. Jain and V. Vazirani. An approximation algorithm for the fault tolerant facility location problem, in: Proceedings of the 3th International Workshop on Approximation Algorithms for Combinatorial Optimization, Springer-Verlag LNCS Vol 1913, 2000, pp. 177-183.

[13] K. Jain and 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.

[14] M. R. Korupolu, C. G. Plaxton and R. Rajaraman. Analysis of a local search heuristic for facility location problems, in: Proceedings of the 9-th Annual ACM-SIAM Symposium on Discrete Algorithms, 1998, pp. 1-10.

[15] P. Krysta and R. Solis-Oba. Approximation algorithms for bounded facility location problems. Journal of Combinatorial Optimization 12 (1) (2001) 233-247.

[16] J. H. Lin and J. S. Vitter. ϵ-approximation with minimum packing constraint violation, in: Proceedings of the 24th Annual ACM Symposium on Theory of Computing, 1992, pp. 771-782.

[17] M. Mahdian, E. Markakis, A. Saberi. V.V. Vazirani. A greedy facility location algorithm analyzed using dual fitting, in: Proceedings of the 4th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, Springer-Verlag LNCS Vol. 2129, 2001, pp. 127-137.

[18] 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, pp. 229-242.

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

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

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

[22] 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).