J Syst Sci Complex (2015) 28: 1102–1114

Approximation Algorithms for the Priority Facility Location Problem with Penalties*

WANG Fengmin · XU Dachuan · WU Chenchen

DOI: 10.1007/s11424-014-2157-2
Received: 30 July 2012 / Revised: 13 November 2012

Abstract This paper considers the priority facility location problem with penalties. The authors develop a primal-dual 3-approximation algorithm for this problem. Combining with the greedy augmentation procedure, the authors further improve the previous ratio 3 to 1.8526.

Keywords Approximation algorithm, facility location problem, greedy augmentation, primal-dual.

1 Introduction

The facility location problem (FLP) is one of the classical NP-hard problems. Uncapacitated facility location problem (UFLP) is the most basic FLP. In the UFLP, given a facility set $\mathcal{F}$ and a client set $\mathcal{C}$, each facility $i \in \mathcal{F}$ has an opening cost f_i, each client $j \in \mathcal{C}$ has a service demand d_j (often assumed to be 1), and there is a connection cost c_{ij} indicating the cost of per unit demand facility i provide for client j. Generally, the UFLP is assumed to be metric, that is, the connection costs satisfy non-negativity, symmetric and the triangle inequality. The objective is to open some facilities such that each client $j \in \mathcal{C}$ is assigned to an open facility with the minimum total opening and connection cost. From the point of view of approximation algorithm, there are three important results for the UFLP: The first one was the 3.16-approximation algorithm by Shmoys, et al.[1], which gave us the first constant factor; the second one was the 1.488-approximation algorithm by Li[2], which gave us the currently best factor; and the third one was 1.463 given by Guha and Khuller[3], which is the lower bound of the factor.

WANG Fengmin · XU Dachuan (Corresponding author)
College of Applied Sciences, Beijing University of Technology, Beijing 100124, *China.*
Email: wfm@emails.bjut.edu.cn; xudc@bjut.edu.cn.
WU Chenchen
College of Science, Tianjin University of Technology, Tianjin 300384, *China.*
Email: wu_chenchen_tjut@163.com.
*This research was supported by the National Natural Science Foundation of China under Grant No. 11371001.
◇ *This paper was recommended for publication by Editor DAI Yuhong.*

Due to the broad applications, variants of the UFLP arise naturally (see [4–12]). Facility location problem with penalties (FLPWP), first studied by Charikar, et al.[5], is one of the variants. The difference between the FLPWP and the UFLP is, in the former problem, not all clients are required to be serviced and there is a penalty cost for the rejected clients. The objective is to minimize the total cost including the opening cost, the connection cost and the penalty cost. According to the type of the penalty cost function, the FLPWP can be classified into either facility location problem with linear penalties (FLPLP) or facility location problem with submodular penalties (FLPSP). For the FLPLP, Charikar, et al.[13] gave a primal-dual based 3-approximation algorithm. Later, Xu and Xu[14, 15] achieved an LP-rounding based $(2 + 2/\mathrm{e})$-approximation algorithm, and then, combining with the cost scaling technique and the greedy augmentation procedure, they designed a primal-dual based 1.8526-approximation algorithm. Besides, Hayrapetyan, et al.[16] presented an LP-rounding based $(1 + \rho)$-approximation algorithm, where ρ is a constant parameter. For the FLPSP, Chudak and Nagano[17] gave a convex program rounding based $(1 + \varepsilon)(1 + \rho)$-approximation algorithm. Du, et al.[12] designed a primal-dual based 3-approximation algorithm.

The priority facility location problem (PFLP), first proposed by Ravi and Sinha[18] as a special case of multicommodity facility location problem (MFLP), is another variant of the UFLP. The differences between the PFLP and the UFLP are, in the PFLP, each client has a level-of-service requirement, and each facility has a non-decreasing cost function that specifies the cost of opening the facility at the level-of-service. The requirement of all clients must be satisfied. The objective is to minimize the total cost including the opening cost and the connection cost. Mahdian[19] offered a primal-dual based 3-approximation algorithm for the PFLP in his Ph. D. thesis. Li, et al.[20] presented a primal-dual based 3-approximation algorithm for the stochastic version of the PFLP. Using the greedy augmentation procedure, they further improved the ratio to 1.8526 which is the best ratio for the PFLP and its stochastic version.

In this paper, combining with the above two variants of the UFLP, we consider the priority facility location problem with penalties (PFLPWP). Different from the UFLP, in the PFLPWP, each client $j \in \mathcal{C}$ has a level-of-service requirement $l_j \in \{1, 2, \cdots, L\}$, and each facility $i \in \mathcal{F}$ has a non-decreasing cost function $f_i(l)$ that specifies the cost of opening the facility i at the level-of-service $l (l = 1, 2, \cdots, L)$. Not all clients are required to be serviced and there is a penalty cost p_j for the rejected client j. The objective is to minimize the total cost including the opening cost, the connection cost and the penalty cost.

Our 3-approximation algorithm for the PFLPWP is an extension of the primal-dual algorithm by Jain and Vazirani[21] for the UFLP. Since the problem is equipped with different level-of-service requirements and penalties, we carefully arrange the order of the level-of-service similar to the method in [19] and accurately identify the outliers. Later, by opening a virtual facility for clients[15], and then combining with the greedy augmentation[3, 5], we further improve the approximation ratio from 3 to 1.8526.

The rest of this paper is organized as follows. In Section 2, we present the integer program, the linear programming relaxation and the dual program for the PFLPWP. In Section 3, we design and analyze the primal-dual algorithm. We offer the improved algorithm with the approximation ratio 1.8526 in Section 4. Finally, the conclusions are given in Section 5.

2 Priority Facility Location Problem with Penalties

In the PFLPWP, given a facility set $\mathcal{F}$ and a client set $\mathcal{C}$, each client j has a level-of-service requirement $l_j \in \{1, 2, \cdots, L\}$ and a penalty cost p_j. The opening cost of facility $i \in \mathcal{F}$, at the level-of-service l, is a non-decreasing function $f_i(l)$ $(l = 1, 2, \cdots, L)$. The connection cost, between client $j \in \mathcal{C}$ and the facility $i \in \mathcal{F}$ satisfying the level-of-service requirement of client j, is c_{ij}. Note that a facility can not be opened at just one level-of-service; otherwise, the level-of-service requirements of some clients will not be satisfied. For convenience, we let $\mathscr{F} = \{(i;l) | i \in \mathcal{F}, l = 1, 2, \cdots, L\}$, where $(i;l)$ is facility-level pair. For short, we call it facility. Our objective is to determine an opening facility set $\widehat{\mathscr{F}} \subseteq \mathscr{F}$, while selecting a penalty client set $\widehat{\mathcal{P}} \subseteq \mathcal{C}$, and then connect the clients in $\mathcal{C} \backslash \widehat{\mathcal{P}}$ to the opening facilities in $\widehat{\mathscr{F}}$, such that the total cost including the opening cost $\sum_{(i;l) \in \widehat{\mathscr{F}}} f_i(l)$, the connection cost $\sum_{j \in \mathcal{C} \backslash \widehat{\mathcal{P}}} c_{\widehat{\theta}(j)j}$, and the penalty cost $\sum_{j \in \widehat{\mathcal{P}}} p_j$ is minimized, where $\widehat{\theta}(j)$ is the closest facility in $\widehat{\mathscr{F}}$ that can satisfy client j's level-of-service requirement.

To derive an integer program formulation for the PFLPWP, we introduce three types of $\{0, 1\}$ variables: $y_i(l)$ indicating whether facility $(i;l)$ is opened at level-of-service l; x_{ij} indicating whether client j is connected to facility i; and z_j indicating whether client j is penalized. The PFLPWP is formulated as

$$\begin{aligned} \min \quad & \sum_{(i;l) \in \mathscr{F}} f_i(l) y_i(l) + \sum_{i \in \mathcal{F}} \sum_{j \in \mathcal{C}} c_{ij} x_{ij} + \sum_{j \in \mathcal{C}} p_j z_j \\ \text{s.t.} \quad & \sum_{i \in \mathcal{F}} x_{ij} + z_j \geq 1, \quad \forall j \in \mathcal{C}, \\ & x_{ij} \leq \sum_{l=l_j}^{L} y_i(l), \quad \forall i \in \mathcal{F}, j \in \mathcal{C}, \\ & x_{ij}, y_i(l), z_j \in \{0, 1\}, \quad \forall (i;l) \in \mathscr{F}, j \in \mathcal{C}. \end{aligned} \tag{1}$$

In the above program, the first constraints denote that each client $j \in \mathcal{C}$ is either connected to a facility or rejected; the second constraints ensure that if client j is connected to facility i, then this facility must open at the level-of-service between l_j and L. Furthermore, in an optimal solution to (1), a facility can only be opened at one level-of-service; otherwise, if there is a facility i opened at two level-of-service l' and l such that $l \geq l'$, then we can just open i at level-of-service l, implying that the total facility cost is decreased by $f(l')$ without increasing the connection cost. Relaxing the last constraints, we obtain

$$\begin{aligned}
\min \quad & \sum_{(i;l)\in\mathscr{F}} f_i(l)y_i(l) + \sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{C}} c_{ij}x_{ij} + \sum_{j\in\mathcal{C}} p_j z_j \\
\text{s.t.} \quad & \sum_{i\in\mathcal{F}} x_{ij} + z_j \geq 1, \quad \forall j \in \mathcal{C}, \\
& x_{ij} \leq \sum_{l=l_j}^{L} y_i(l), \quad \forall i \in \mathcal{F}, j \in \mathcal{C}, \\
& x_{ij}, y_i(l), z_j \geq 0, \quad \forall (i;l) \in \mathscr{F}, j \in \mathcal{C}.
\end{aligned} \tag{2}$$

Introducing the dual variables α_j and β_{ij}, we obtain the dual of the program (2)

$$\begin{aligned}
\max \quad & \sum_{j\in\mathcal{C}} \alpha_j \\
\text{s.t.} \quad & \alpha_j - \beta_{ij} \leq c_{ij}, \quad \forall i \in \mathcal{F}, j \in \mathcal{C}, \\
& \sum_{j\in\mathcal{C}:l\geq l_j} \beta_{ij} \leq f_i(l), \quad \forall (i;l) \in \mathscr{F}, \\
& \alpha_j \leq p_j, \quad \forall j \in \mathcal{C}, \\
& \alpha_j, \beta_{ij} \geq 0, \quad \forall i \in \mathcal{F}, j \in \mathcal{C},
\end{aligned} \tag{3}$$

where α_j can be regarded as the budget of client j, and β_{ij} as the contribution of the client j to the facility $(i;l)$ with $l \geq l_j$.

3 Primal-Dual Algorithm

In this section, we will first propose a primal-dual algorithm for the PFLPWP, then analyze the algorithm to obtain the approximation ratio of 3.

3.1 The Primal-Dual Algorithm

We now give the primal-dual algorithm for the PFLPWP.

Algorithm 1 (The primal-dual algorithm)

Stage 1 (Constructing a dual feasible solution to (3))

Step 1 First introduce a concept of time, denoted by t. The algorithm starts at time $t = 0$. Initially all the dual variables are zero, all the facilities are closed, and all clients are unfrozen. In the process of the algorithm, when the dual variable α_j stop to increase, we call client j is frozen. Let $\widetilde{\mathscr{F}}$ denote the temporarily open facility set, U denote the unfrozen client set, and $\widetilde{\mathcal{P}} = \{j \in \mathcal{C} | \alpha_j = p_j\}$. At the beginning of the algorithm, set $\widetilde{\mathscr{F}} := \emptyset$, $U := \mathcal{C}$, $\widetilde{\mathcal{P}} := \emptyset$.

Step 2 For the unfrozen client $j \in U$, we increase α_j at the same rate with time t. As time t goes, one of the following events will occur.

Event 1 There is a client $j \in U$, such that $\alpha_j = p_j$. Freeze j, and update $\widetilde{\mathcal{P}} := \widetilde{\mathcal{P}} \cup \{j\}$ and $U := U \setminus \{j\}$.

Event 2 There is a client $j \in U$, $(i;l) \in \mathscr{F} \setminus \widetilde{\mathscr{F}}$, $l \geq l_j$, such that $\alpha_j = c_{ij}$. We say that the facility-client pair (i,j) is tight. After that, the corresponding dual variable β_{ij} will increase in accordance with α_j. Define $\beta_{ij} = \max\{0, t - c_{ij}\}$.

Event 3 There is a facility $(i;l) \in \mathscr{F} \setminus \widetilde{\mathscr{F}}$, such that $\sum_{j\in\mathcal{C}:l\geq l_j} \beta_{ij} = f_i(l)$. We call facility $(i;l)$ is temporarily open. Update $\widetilde{\mathscr{F}} := \widetilde{\mathscr{F}} \cup \{(i;l)\}$, and define $N(i;l) = \{j|\beta_{ij} > 0, l \geq l_j\}$ to be the neighbor of the facility $(i;l)$. Now, freeze client $j \in N(i;l) \cap U$, and connect (directly) this client to the facility $(i;l)$, which is declared the connecting witness of client j. For the convenience of the algorithm analysis, let $\widehat{N}(i;l) := N(i;l)$. Update $U := U \setminus N(i;l)$.

Event 4 There is a client $j \in U$, $(i;l) \in \widetilde{\mathscr{F}}$, $l \geq l_j$, such that $\alpha_j = c_{ij}$. Freeze j and connect (directly) this client to facility $(i;l)$, which is declared the connecting witness of client j. Update $\widehat{N}(i;l) := \widehat{N}(i;l) \cup \{j\}$, $U := U \setminus \{j\}$.

If all the events happen at the same time, the algorithm executes them in an arbitrary order. When $U = \emptyset$, go to Stage 2.

Stage 2 (Constructing a primal integer feasible solution to (1))

Step 1 Let $\widehat{\mathscr{F}}$ denote the finally open facility set, and $\widehat{\mathcal{P}}$ denote the penalty client set. Set $\widehat{\mathscr{F}} := \emptyset$, $\widehat{\mathcal{P}} := \emptyset$.

Step 2 Determine open facilities. Sort the temporarily open facility in a decreasing level-of-service. According to this order, consider each facility $(i;l) \in \widetilde{\mathscr{F}}$. If there is $(i',l') \in \widehat{\mathscr{F}}$ such that $l' \geq l$ and $N(i;l) \cap N(i',l') \neq \emptyset$, then consider the next facility in $\widetilde{\mathscr{F}}$; otherwise update $\widehat{\mathscr{F}} := \widehat{\mathscr{F}} \cup \{(i;l)\}$. If $(i;l) \in \widehat{\mathscr{F}}$, we say that facility i is open at level l.

Step 3 Determine penalty clients. Let $\widehat{\mathcal{P}} := \widetilde{\mathcal{P}} \setminus \bigcup_{(i;l)\in\widehat{\mathscr{F}}} \widehat{N}(i;l)$.

Step 4 Connect client $j \in \mathcal{C} \setminus \widehat{\mathcal{P}}$ to the closest facility $(i;l) \in \widehat{\mathscr{F}}$ opening at level $l \geq l_j$.

Let (α, β) be the dual solution obtained by Stage 1, and $(\widehat{x}, \widehat{y}, \widehat{z})$ be the primal solution obtained by Stage 2, respectively. It is easy to prove that the two solutions obtained by Algorithm 1 are both feasible. Let F, C and P be the opening cost, connection cost and penalty cost of the solution $(\widehat{x}, \widehat{y}, \widehat{z})$.

Next we present the analysis of Algorithm 1.

3.2 Analysis of Algorithm 1

We now describe the analysis of Algorithm 1. We first give three lemmas to bound the opening cost F, connection cost C and penalty cost P respectively, and then bound the total cost of the solution $(\widehat{x}, \widehat{y}, \widehat{z})$ to obtain the approximation ratio 3.

Since for all facilities $(i;l)$, the neighbors $N(i;l)$ are disjoint, we have the following lemma.

Lemma 3.1

$$F = \sum_{(i;l)\in\widehat{\mathscr{F}}} f_i(l) = \sum_{(i;l)\in\widehat{\mathscr{F}}} \sum_{j\in N(i;l)} \beta_{ij}.$$

Note that clients in $\mathcal{C} \setminus \widehat{\mathcal{P}}$ can be divided into three groups: The clients contributing to the finally open facilities, i.e., clients in $\mathcal{C}_1 = \bigcup_{(i;l)\in\widehat{\mathscr{F}}} N(i;l)$; the clients directly connected but not contributing to the finally open facilities, i.e., clients in $\mathcal{C}_2 = \bigcup_{(i;l)\in\widehat{\mathscr{F}}} \left(\widehat{N}(i;l) \setminus N(i;l)\right)$; and the clients switched its connection eventually because its connecting witness are not open at Stage 2 in Algorithm 1, i.e., clients in $\mathcal{C}_3 = (\mathcal{C} \setminus \widehat{\mathcal{P}}) \setminus \bigcup_{(i;l)\in\widehat{\mathscr{F}}} \widehat{N}(i;l)$. Thus we obtain the following lemma.

Lemma 3.2

$$C \le \sum_{(i;l)\in\widehat{\mathscr{F}}} \sum_{j\in N(i;l)} c_{ij} + \sum_{j\in\mathcal{C}_2} \alpha_j + 3\sum_{j\in\mathcal{C}_3} \alpha_j.$$

Proof For any client $j \in \mathcal{C} \setminus \widehat{\mathcal{P}}$, $(i;l) \in \widehat{\mathscr{F}}$, we consider the following three possibilities.

(i) For client $j \in \mathcal{C}_1$, connect (directly) j to its connecting witness $(i;l) \in \widehat{\mathscr{F}}$. The connection cost is c_{ij}.

(ii) For client $j \in \mathcal{C}_2$, connect (directly) j to its connecting witness $(i;l) \in \widehat{\mathscr{F}}$. The connection cost is $c_{ij} = \alpha_j$.

(iii) For client $j \in \mathcal{C}_3$, denote its connecting witness by $(\widetilde{i};\widetilde{l})$. Since $(\widetilde{i};\widetilde{l})$ is not open at Stage 2 in Algorithm 1, there exist $(i;l) \in \widehat{\mathscr{F}}(l \ge \widetilde{l})$ and client j', such that j' contributs to $(\widetilde{i};\widetilde{l})$ and $(i;l)$, namely $j' \in N(\widetilde{i};\widetilde{l}) \cap N(i;l)$. Connect (indirectly) j to facility $(i;l)$. See Figure 1.

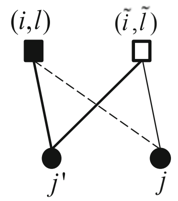

Figure 1 The evaluation of connection cost for client $j \in \mathcal{C}_3$. Heavy line corresponds to the client has contribution to the open facility, thin line corresponds to the connection between client and facility, dashed line corresponds to the reconnection of client j to facility $(i;l)$

And because $(\widetilde{i};\widetilde{l})$ is the connecting witness of client j, $j \notin \widetilde{\mathcal{P}}$, thus $\alpha_j \ge \max\{c_{\widetilde{i}j}, t(\widetilde{i};\widetilde{l})\}$. By the triangle inequality, we get

$$c_{ij} \le c_{ij'} + c_{\widetilde{i}j'} + c_{\widetilde{i}j} \le 2t(\widetilde{i};\widetilde{l}) + \alpha_j \le 3\alpha_j.$$

Summing up the connection costs of all clients, we obtain the lemma. ∎

From the definition of $\widehat{\mathcal{P}}$, we have the following lemma.

Lemma 3.3

$$P = \sum_{j\in\widehat{\mathcal{P}}} \alpha_j.$$

The main result of this subsection is the following theorem.

Theorem 3.4 *Algorithm 1 is a 3-approximation algorithm for the PFLPWP.*

Proof Combining Lemmas 3.1, 3.2, and 3.3, we obtain

$$\begin{aligned}3F+3P+C &\le 3\sum_{(i;l)\in\widehat{\mathscr{F}}}\sum_{j\in N(i;l)}\beta_{ij}+3\sum_{j\in\widehat{\mathcal{P}}}\alpha_j+\sum_{(i;l)\in\widehat{\mathscr{F}}}\sum_{j\in N(i;l)}c_{ij}+\sum_{j\in\mathcal{C}_2}\alpha_j+3\sum_{j\in\mathcal{C}_3}\alpha_j\\ &\le 3\Bigg(\sum_{(i;l)\in\widehat{\mathscr{F}}}\sum_{j\in N(i;l)}\beta_{ij}+\sum_{(i;l)\in\widehat{\mathscr{F}}}\sum_{j\in N(i;l)}c_{ij}\Bigg)+3\sum_{j\in\widehat{\mathcal{P}}}\alpha_j+\sum_{j\in\mathcal{C}_2}\alpha_j+3\sum_{j\in\mathcal{C}_3}\alpha_j\\ &= 3\sum_{j\in\mathcal{C}_1}\alpha_j+3\sum_{j\in\widehat{\mathcal{P}}}\alpha_j+\sum_{j\in\mathcal{C}_2}\alpha_j+3\sum_{j\in\mathcal{C}_3}\alpha_j.\end{aligned}$$

Therefore

$$3F+3P+C\le 3\sum_{j\in\mathcal{C}}\alpha_j. \tag{4}$$

The proof is finished. ∎

4 The Improved 1.8526-Approximation Algorithm

In this section, we use the greedy augmentation technique in [5] to improve the approximation ratio from 3 to 1.8526.

4.1 The Improved Algorithm

We now present our greedy improvement schema.

Algorithm 2 (The improved algorithm)

Stage 1 (Constructing and solving a new instance)

Step 1 For each facility $(i;l)\in\mathscr{F}$ and constant $\gamma>0$, let $f_i'(l):=\gamma f_i(l)$ to obtain a new instance $\mathcal{I}$.

Step 2 Solve $\mathcal{I}$ by Algorithm 1 to get the feasible solution $(\widehat{x}',\widehat{y}',\widehat{z}')$. Note that this solution is also feasible to (2). Assume that solution $(\widehat{x}',\widehat{y}',\widehat{z}')$ opens facilities set $\widehat{\mathscr{F}'}$ and penalty clients set $\widehat{\mathcal{P}}'$ in the instance $\mathcal{I}$. Let $\widehat{\mathscr{F}_g}:=\widehat{\mathscr{F}'}$ and $\widehat{\mathcal{P}}_g:=\widehat{\mathcal{P}}'$ denote the open facilities set and the penalty clients set of solution $(\widehat{x}',\widehat{y}',\widehat{z}')$ in (2).

Stage 2 (Greedy improvement)

Step 1 Consider (2). For the clients in $\mathcal{C}$, add a facility $(i_0;L)$ to the feasible solution $(\widehat{x}',\widehat{y}',\widehat{z}')$ with opening cost $f_{i_0}(L)=0$ and connection cost $c_{i_0j}=p_j$, $\forall j\in\mathcal{C}$. For each client $j\in\mathcal{C}$, let $\theta(j)$ be the closest facility in solution $(\widehat{x}',\widehat{y}',\widehat{z}')$ that can satisfy its level-of-service requirement. Let $\mathscr{F}:=\mathscr{F}\cup\{(i_0;L)\}$, and $\widehat{\mathscr{F}_g}:=\widehat{\mathscr{F}_g}\cup\{(i_0;L)\}$.

Step 2 For each facility $(i;l)\in\mathscr{F}\setminus\widehat{\mathscr{F}_g}$, compute

$$\text{gain}(i;l):=\sum_{j\in\mathcal{C}:c_{\theta(j)j}\ge c_{ij},l\ge l_j}(c_{\theta(j)j}-c_{ij})-f_i(l),$$

and let

$$(i_g;l_g):=\arg\max_{(i;l)\in\mathscr{F}\setminus\widehat{\mathscr{F}_g}}\frac{\text{gain}(i;l)}{f_i(l)}.$$

Step 3 If $\text{gain}(i_g, l_g) > 0$, update $\widehat{\mathscr{F}_g} := \widehat{\mathscr{F}_g} \cup \{(i_g, l_g)\}$, and $\widehat{\mathcal{P}}_g := \widehat{\mathcal{P}}_g \setminus \{j | c_{\theta(j)j} \geq c_{i_g j}\}$, and goto Step 2; otherwise goto Step 4.

Step 4 Connect $j \in \mathcal{C} \setminus \widehat{\mathcal{P}}_g$ to the closest facility $(i;l) \in \widehat{\mathscr{F}_g}$, where $l \geq l_j$. Denote the solution by $(\widehat{x}_g, \widehat{y}_g, \widehat{z}_g)$. See Figure 2.

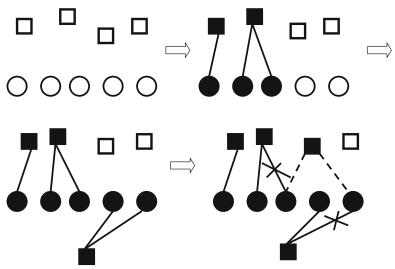

Figure 2 The construction of integer feasible solution(Algorithm 2). Heavy line corresponds to the client connect to the open facility, dashed line corresponds to the reconnection of client j to the newly open facility

Let F_g, C_g and P_g denote the opening cost, connection cost and penalty cost respectively in solution $(\widehat{x}_g, \widehat{y}_g, \widehat{z}_g)$. Let F^*, C^* and P^* denote the opening cost, connection cost and penalty cost respectively in the optimal solution $OPT = (x^*, y^*, z^*)$ of (1). Let $\mathscr{F}^*$ and $\mathcal{P}^*$ be the open facility set and the penalty client set respectively in OPT.

4.2 Analysis of Algorithm 2

We first develop some lemmas, and then present the approximation ratio of Algorithm 2.

Lemma 4.1 *The connection cost and penalty cost of the solution obtained from Algorithm* 2 *satisfies*

$$C_g + P_g \leq F^* + C^* + P^*. \tag{5}$$

Proof To prove by contradiction, we assume that $C_g + P_g > F^* + C^* + P^*$. Similar to Step 1 in Stage 2 of Algorithm 2, for each client $j \in \mathcal{C}$, add a facility $(i_0; L)$ to $\mathscr{F}^*$ with opening cost $f_{i_0}(L) = 0$ and connection cost $c_{i_0 j} = p_j$ for any $j \in \mathcal{C}$. Let $\mathscr{F}^* := \mathscr{F}^* \cup \{(i_0; L)\}$. For client j, let $\theta^*(j)$ be the closest facility in $\mathscr{F}^*$ that can satisfy its level-of-service requirement.

Note that, Step 2 in Stage 2 of Algorithm 2 is an iterative process. Assume that at the beginning of the iteration, the connection cost and penalty cost are C and P respectively. Let us consider the following restricted operation which results in a less greedy gain$'$. The facilities $(i;l)$ are added to $\widehat{\mathscr{F}_g}$ if and only if $\theta^*(j) = i$, implying that client j is connected to facility i. Define

$$\text{gain}'(i;l) := \sum_{j \in \mathcal{C}: \theta^*(j) = i, l \geq l_j} (c_{\theta(j)j} - c_{ij}) - f_i(l).$$

Obviously, for any $(i;l) \in \mathscr{F}^*$, we have $\text{gain}(i;l) \geq \text{gain}'(i;l)$. Hence,

$$\sum_{(i;l) \in \mathscr{F}^*} \text{gain}(i;l) \geq \sum_{(i;l) \in \mathscr{F}^*} \text{gain}'(i;l).$$

And because

$$\begin{aligned}\sum_{(i;l)\in\mathscr{F}^*}\text{gain}'(i;l) &= \sum_{(i;l)\in\mathscr{F}^*}\left(\sum_{j\in\mathcal{C}:\theta^*(j)=i,l\geq l_j}(c_{\theta(j)j}-c_{ij})-f_i(l)\right)\\ &= \sum_{j\in\mathcal{C}}c_{\theta(j)j}-\sum_{j\in\mathcal{C}}c_{\theta^*(j)j}-\sum_{(i;l)\in\mathscr{F}^*}f_i(l)\\ &= C+P-(C^*+P^*)-F^*,\end{aligned}$$

we have

$$\sum_{(i;l)\in\mathscr{F}^*}\text{gain}(i;l)\geq C+P-C^*-P^*-F^*.$$

So, if the iteration starts with the solution $(\widehat{x}_g,\widehat{y}_g,\widehat{z}_g)$, then

$$\sum_{(i;l)\in\mathscr{F}^*}\text{gain}(i;l)\geq C_g+P_g-C^*-P^*-F^*.$$

According to the assumption of contradictory, we have

$$\sum_{(i;l)\in\mathscr{F}^*}\text{gain}(i;l)>0,$$

which indicates that there exists a facility $(i;l)$ such that $\text{gain}(i;l)>0$. Thus, the algorithm will carry out the next iteration. Therefore, we must have $C_g+P_g\leq F^*+C^*+P^*$. ∎

Assume that after the kth iteration in Algorithm 2, the opening cost, connection cost and penalty cost of the obtained solution are F_k, C_k and P_k, respectively. From the proof of Lemma 4.1, we have the following corollary.

Corollary 4.2 *If*

$$C_k+P_k>F^*+C^*+P^*,$$

Algorithm 2 *will carry out the* $(k+1)$*th iteration. During the* $(k+1)$*th iteration, there exists a facility* $(i;l)\in\mathscr{F}^*$ *such that*

$$\frac{\text{gain}(i;l)}{f_i(l)}\geq\frac{C_k+P_k-F^*-C^*-P^*}{F^*}.\tag{6}$$

Assume F, C and P are the opening cost, connection cost, and penalty cost of the solution $(\widehat{x}',\widehat{y}',\widehat{z}')$ in the original instance, respectively. Noting that Algorithm 2 carries out the iteration at the solution $(\widehat{x}',\widehat{y}',\widehat{z}')$, we set $F_0:=F$, $C_0:=C$, $P_0:=P$. We have the following lemma.

Lemma 4.3 *If* $C+P>F^*+C^*+P^*$, *then the opening cost of the solution obtained from Algorithm* 2 *satisfies*

$$F_g+C_g+P_g\leq F+F^*\ln\left(\frac{3\gamma F^*-3\gamma F+2C^*+2P^*}{F^*}\right)+F^*+C^*+P^*.\tag{7}$$

Proof By Corollary 4.2, there exists an integer $K > 0$ such that $C_K + P_K \le F^* + C^* + P^*$ and $C_k + P_k > F^* + C^* + P^*$ for any $0 \le k \le K - 1$. Let us consider the $(k+1)$th iteration, where $0 \le k \le K - 1$. From (6), we have

$$\frac{C_k + P_k - C_{k+1} - P_{k+1} - (F_{k+1} - F_k)}{F_{k+1} - F_k} \ge \frac{C_k + P_k - F^* - C^* - P^*}{F^*}.$$

It is easy to see that

$$F_{k+1} - F_k \le F^* \left(\frac{C_k + P_k - C_{k+1} - P_{k+1}}{C_k + P_k - C^* - P^*} \right).$$

Therefore,

$$\begin{aligned} F_K &= F + \sum_{k=1}^{K} (F_k - F_{k-1}) \\ &\le F + F^* \sum_{k=1}^{K} \left(\frac{C_{k-1} + P_{k-1} - C_k - P_k}{C_{k-1} + P_{k-1} - C^* - P^*} \right) \end{aligned}$$

implying that

$$\begin{aligned} F_g + C_g + P_g &\le F_K + C_K + P_K \\ &\le F + F^* \sum_{k=1}^{K} \left(\frac{C_{k-1} + P_{k-1} - C_k - P_k}{C_{k-1} + P_{k-1} - C^* - P^*} \right) + C_K + P_K. \end{aligned} \tag{8}$$

The derivation with respect to $C_K + P_K$ on the right side of (8) is

$$1 - \frac{F^*}{C_{K-1} + P_{K-1} - C^* + P^*} \ge 0,$$

implying that the right side of (8) is a monotone increasing function with respect to $C_K + P_K$, and by (5), it achieves its maximum at $C_K + P_K = F^* + C^* + P^*$. In the following, we assume $C_K + P_K = F^* + C^* + P^*$. Since

$$\begin{aligned} \frac{C_{k-1} + P_{k-1} - C_k - P_k}{C_{k-1} + P_{k-1} - C^* - P^*} &= 1 - \frac{C_k + P_k - C^* - P^*}{C_{k-1} + P_{k-1} - C^* - P^*} \\ &\le \ln \left(\frac{C_{k-1} + P_{k-1} - C^* - P^*}{C_k + P_k - C^* - P^*} \right), \end{aligned}$$

we have

$$\begin{aligned} F_g + C_g + P_g &\le F + F^* \sum_{k=1}^{K} \ln \left(\frac{C_{k-1} + P_{k-1} - C^* - P^*}{C_k + P_k - C^* - P^*} \right) + C_K + P_K \\ &= F + F^* \ln \left(\frac{C + P - C^* - P^*}{C_K + P_K - C^* - P^*} \right) + C_K + P_K \\ &= F + F^* \ln \left(\frac{C + P - C^* - P^*}{F^*} \right) + F^* + C^* + P^*. \end{aligned} \tag{9}$$

On the other hand, by (4) and the weak duality theorem, we have

$$3\gamma F + C + P \leq 3\gamma F + C + 3P \leq 3(\gamma F^* + C^* + P^*) \tag{10}$$

which indicates that

$$C + P - C^* - P^* \leq 3\gamma F^* - 3\gamma F + 2C^* + 2P^*.$$

We complete the proof by combining the above inequality and (9). ∎

By Lemma 4.3, we obtain

Lemma 4.4 *If $C + P > F^* + C^* + P^*$, then the cost of the solution obtained from Algorithm 2 is no more than*

$$(1 + \ln(3\gamma))F^* + \left(1 + \frac{2}{3\gamma}\right)(C^* + P^*).$$

Proof Taking the derivation with respect to F on the right side of (7), we have

$$1 - \frac{3\gamma F^*}{3\gamma F^* - 3\gamma F + 2C^* + 2P^*}.$$

Therefore the right side of (7) achieves its maximum at $F = \frac{2}{3\gamma}(C^* + P^*)$. Setting $F := \frac{2}{3\gamma}(C^* + P^*)$, we obtain

$$\begin{aligned} F_g + C_g + P_g &\leq \frac{2}{3\gamma(C^* + P^*)} + F^* \ln(3\gamma) + F^* + C^* + P^* \\ &= (1 + \ln(3\gamma))F^* + \left(1 + \frac{2}{3\gamma}\right)(C^* + P^*). \end{aligned}$$

∎

Lemma 4.5 *If $C + P \leq F^* + C^* + P^*$, then the cost of the solution obtained from Algorithm 2 is no more than*

$$\left(2 - \frac{1}{3\gamma}\right)F^* + \left(1 + \frac{2}{3\gamma}\right)(C^* + P^*).$$

Proof By (10) and $C + P \leq F^* + C^* + P^*$, we have

$$\begin{aligned} F + C + P &= \frac{3\gamma F + C + P}{3\gamma} + \left(1 - \frac{1}{3\gamma}\right)(C + P) \\ &\leq \frac{3(\gamma F^* + C^* + P^*)}{3\gamma} + \left(1 - \frac{1}{3\gamma}\right)(F^* + C^* + P^*) \\ &= \left(2 - \frac{1}{3\gamma}\right)F^* + \left(1 + \frac{2}{3\gamma}\right)(C^* + P^*). \end{aligned}$$

Since Algorithm 2 does not increase the total cost, we have

$$F_g + C_g + P_g \leq \left(2 - \frac{1}{3\gamma}\right)F^* + \left(1 + \frac{2}{3\gamma}\right)(C^* + P^*).$$

∎

Finally, we obtain the following main result in this section.

Theorem 4.6 *Setting* $\gamma := 0.7192$*, Algorithm* 2 *is a* 1.8526*-approximation algorithm for the PFLPWP.*

Proof It follows from Lemmas 4.4 and 4.5 that the approximation ratio of Algorithm 2 is

$$\max\left\{1+\ln(3\gamma), 2-\frac{1}{3\gamma}, 1+\frac{2}{3\gamma}\right\} \approx 1.8526.$$

∎

5 Conclusion

In this paper, we propose the PFLPWP along with a primal-dual 3-approximation algorithm which is the first constant (combinatorial) approximation algorithm for the PFLPWP. Furthermore, by integrating the scaling and greedy augmentation techniques (see [3, 5]), we obtain an improved ratio of 1.8526. It would be interesting to further improve the approximation ratio for the PFLPWP.

Acknowledgements We would like to thank the anonymous referees for helping us improve the presentation of this manuscript significantly. In particular, one of the referees pointed out a technical error in an earlier version of our proof of Lemma 4.3.

References

[1] Shmoys D B, Tardos É, and Aardal K I, Approximation algorithms for facility location problems, *Proceedings of STOC*, 1997, 265–274.

[2] Li S, A 1.488-approximation algorithm for the uncapacitated facility location problem, *Proceedings of ICALP, Part* II, 2011, 77–88.

[3] Guha S and Khuller S, Greedy strikes back: Improved facility location algorithms, *Proceedings of SODA*, 1998, 649–657.

[4] Ageev A, Ye Y, and Zhang J, Improved combinatorial approximation algorithms for the k-level facility location problem, *SIAM J. Discrete Math.*, 2003, **18**: 207–217.

[5] Charikar M and Guha S, Improved combinatorial algorithms for facility location problems, *SIAM J. Comput.*, 2005, **34**: 803–824.

[6] Zhang J, Chen B, and Ye Y, A multiexchange local search algorithm for the capacitated facility location problem, *Math. Oper. Res.*, 2005, **30**: 389–403.

[7] Zhang J, Approximating the two-level facility location problem via a quasi-greedy approach, *Math. Program.*, 2006, **108**: 159–176.

[8] Zhang P, A new approximation algorithm for the k-facility location problem, *Theor. Comput. Sci.*, 2007, **384**: 126–135.

[9] Chen X and Chen B, Approximation algorithms for soft-capacitated facility location in capacitated network design, *Algorithmica*, 2009, **53**: 263–297.

[10] Du D, Wang X, and Xu D, An approximation algorithm for the k-level capacitated facility location problem, *J. Comb. Optim.*, 2010, **20**: 361–368.

[11] Shu J, An efficient greedy heuristic for warehouse-retailer network design optimization, *Transport. Sci.*, 2010, **44**: 183–192.

[12] Du D, Lu R, and Xu D, A primal-dual approximation algorithm for the facility location problem with submodular penalties, *Algorithmica*, 2012, **63**: 191–200.

[13] Charikar M, Khuller S, Mount D M, and Narasimhan G, Algorithms for facility location problems with outliers, *Proceedings of SODA*, 2001, 642–651.

[14] Xu G and Xu J, An LP-rounding algorithm for approximating uncapacitated facility location problem with penalties, *Inform. Process. Lett.*, 2005, **94**: 119–123.

[15] Xu G and Xu J, An improved approximation algorithm for uncapacitated facility location problem with penalties, *J. Comb. Optim.*, 2008, **17**: 424–436.

[16] Hayrapetyan A, Swamy C, and Tardos É, Network design for information networks, *Proceedings of SODA*, 2005, 933–942.

[17] Chudak F A and Nagano K, Efficient solutions to relaxations of combinatorial problems with submodular penalties via the Lovász extension and non-smooth convex optimization, *Proceedings of SODA*, 2007, 79–88.

[18] Ravi R and Sinha A, Multicommodity facility location, *Proceedings of SODA*, 2004, 342–349.

[19] Mahdian M, Facility location and the analysis of algorithms through factor-revealing problems, Ph.D.'s degree thesis, Massachusetts Institute of Technology, Cambridge, MA, 2004.

[20] Li G, Wang Z, and Wu C, Approximation algorithms for the stochastic priority facility location problem, *Optimization*, 2013, **62**(7): 919–928.

[21] Jain K and Vazirani V V, Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and Lagrangian relaxation, *J. ACM*, 2001, **48**: 274–296.