The Online Connected Facility Location Problem

Mário César San Felice*[1], David P. Willamson[2], and Orlando Lee[1]

[1] Unicamp, Institute of Computing, Campinas SP 13083-852, Brazil
felice@ic.unicamp.br, lee@ic.unicamp.br
[2] Cornell University, School of Operations Research and Information Engineering, Ithaca NY 14853-3801, USA
dpw@cs.cornell.edu

Abstract. In this paper we propose the Online Connected Facility Location problem (OCFL), which is an online version of the Connected Facility Location problem (CFL). The CFL is a combination of the Uncapacitated Facility Location problem (FL) and the Steiner Tree problem (ST). We give a randomized $\mathrm{O}(\log^2 n)$-competitive algorithm for the OCFL via the sample-and-augment framework of Gupta, Kumar, Pál, and Roughgarden and previous algorithms for Online Facility Location (OFL) and Online Steiner Tree (OST). Also, we show that the same algorithm is a deterministic $\mathrm{O}(\log n)$-competitive algorithm for the special case of the OCFL with $M = 1$.

Keywords: Online Algorithms, Competitive Analysis, Connected Facility Location, Steiner Tree, Approximation Algorithms, Randomized Algorithms.

1 Introduction

We start by presenting several problems that are relevant to this work.

In the Facility Location (FL) problem, we have a set of clients and a set of facilities in a metric space. Each facility has a cost associated with opening the facility. The cost of assigning a client to a facility is the distance between the two points. The goal of the problem is to select a set of facilities to open and to assign clients to open facilities so that the total cost of opening the facilities plus the cost of connecting clients to their assigned facilities is minimized. FL is an NP-complete problem that has been well studied; several constant ratio approximation algorithms are known for it [1–4]. It is particularly interesting that several different design techniques, such as LP rounding, primal-dual and local search, are successful at achieving good approximation ratios for this problem.

The online version of FL is the Online Facility Location problem (OFL), in which the clients are revealed one at a time and each one needs to be connected to an open facility before the next one arrives. As time progresses, no connection can be changed or opened facility can be closed. Algorithms for online problems are analyzed via competitive analysis [5]. An α-competitive algorithm returns a

* grant No. 2012/06728-3, São Paulo Research Foundation (FAPESP).

solution whose cost is within a factor of α of the cost of an optimal solution to the corresponding offline problem; α is called the competitive ratio of the algorithm. There are randomized and deterministic $O(\log n)$-competitive algorithms known for the OFL [6–10], where n is the number of clients. Also, the lower bound for the competitive ratio of an algorithm for OFL is $\Omega(\frac{\log n}{\log \log n})$ [7].

The Steiner Tree problem (ST) is a network design problem defined in a graph with edge costs. Its input is a set of terminals that need to be connected to each other. A solution for ST is a tree that contains all terminal nodes and that can contain other nodes, called Steiner nodes. The goal is to minimize the total cost of edges in the tree. The ST is also a well-studied NP-complete problem for which several different constant ratio approximation algorithms are known [11, 12], such as greedy, primal-dual, and randomized rounding algorithms.

The online version of ST is the Online Steiner Tree problem (OST), in which the terminals are revealed one at a time and each one needs to be connected to the current tree before the next one arrives. Also, no edge in the tree can be removed in the future. There are $O(\log n)$-competitive algorithms known for OST [13, 14], where n is the number of terminals. Also, the lower bound for the competitive ratio of an algorithm for OST is $\Omega(\log n)$ [13].

The Connected Facility Location problem (CFL) is a network design problem with two layers; it is motivated by the necessity of building networks in which the end users are connected to servers, with less expensive lower bandwidth connections, and the servers are connected to each other, through more expensive higher bandwidth connections. The input to the CFL is the same as the FL, except that there is a facility that is designated the root that represents the connection of the network to the outside world, and a parameter $M \geq 1$, which is a cost scaling factor. A solution for CFL is a set of open facilities (including the root), an assignment of clients to open facilities, and a tree spanning the open facilities. The goal of the problem is to minimize the total cost of opening facilities plus the total cost of connecting clients to their assigned facilities plus M times the cost of the edges in the tree spanning the open facilities. The CFL is an NP-complete problem; it has randomized and deterministic constant ratio approximation algorithms [15–20] that use techniques such as sample-and-augment, LP rounding and primal-dual. The CFL can be seen as a combination of FL with ST, using the cost scaling factor M.

Our Contributions. In this paper we propose the Online Connected Facility Location problem (OCFL) that is the online version of CFL. In the OCFL the clients are revealed one at a time and each one needs to be connected with a facility before the next one arrives. If a new facility is opened, it needs to be connected to the tree spanning the other opened facilities immediately. Also, no connection can be changed, opened facility can be closed or edge used in the tree can be removed in the future. We can also view the OCFL as the combination of OFL and OST, using a cost scaling factor M. Since the OCFL can be reduced to the OST, there is a lower bound of $\Omega(\log n)$ on the competitive ratio of any algorithm for the OCFL.

Our main result is a randomized $O(\log^2 n)$-competitive algorithm for the OCFL that uses the sample-and-augment technique of Gupta, Kumar, Pál, and Roughgarden [15], where n is the number of clients. We also show that the same algorithm is a deterministic $O(\log n)$-competitive algorithm for the special case of the OCFL with $M = 1$.

The sample-and-augment technique was developed by Gupta et al. [15] for a number of different problems; we illustrate it here with the single-source rent-or-buy problem. In this problem, we are given as input a graph with edge costs, a set of terminals, a root node, and a cost scaling factor $M \geq 1$. We must connect each terminal to the root. We can either buy an edge (at M times its cost), or rent an edge (but then each terminal using the edge must pay its cost). The sample-and-augment technique samples each terminal with probability $\frac{1}{M}$; it then uses an approximation algorithm for ST on the sampled terminals and buys the edges in the resulting tree; then it rents edges as needed to connect the nonsampled terminals to the tree of bought edges. The key idea of this technique is that by sampling with probability $\frac{1}{M}$ the algorithm balances the costs of the edges that should be bought and the cost of the edges that should be rented. We will use it in a similar way.

2 Problem Definitions

In this section we formally define the Online Connected Facility Location problem (OCFL). First we define its offline version, the Connected Facility Location problem (CFL), and then we describe what changes in the online version.

As mentioned previously the CFL combines the Facility Location problem (FL) with the Steiner Tree problem (ST). The input to the problem is a complete graph $G = (V, E)$, distances $d(i, j) : E \to R^+$ that respects the triangle inequality, facility opening costs $f(i) : V \to R^+$, clients $D \subseteq V$, possible facilities $F \subseteq V$, a root $r \in V$ and a parameter $M \geq 0$, that we call the cost scaling factor.

The goal is to serve the clients with the minimum cost. To serve the clients one must open a subset of facilities $F' \subseteq F$, connect each client in D to an opened facility in F' or to the root r, and give a tree T connecting the facilities in F' and r to each other. The cost to be minimized is the sum of the cost of the open facilities, the distance of each client to its assigned facility, and M times the cost of the tree T that connects $\{r\} \cup F'$, namely:

$$\sum_{i \in F'} f(i) + \sum_{j \in D} d(j, a(j)) + M \sum_{e=(i,j) \in T} d(i, j) \ ,$$

where $a : D \to V$ is a function that assigns each client in D to the root r or to a facility in F'.

The OCFL is the online version of CFL, so it combines the Online Facility Location problem (OFL) with the Online Steiner Tree problem (OST), just as CFL does with FL and ST. In the OCFL the clients in D arrive one at a time and the one that just arrived must be served before the next one arrives; in particular, it must be assigned to an open facility. The algorithm can open a facility for this purpose, but then the opened facility must be connected to the other opened facilities in the tree T. All decisions of the algorithm are irrevocable. In this case, this means that the algorithm cannot decide to remove from the current solution any facility previously opened, change to which facility a client is connected, even if a closer facility was opened, or remove an edge from the tree T.

3 Notation and Definitions

In what follows,

- $n = |D|$ is the number of clients. Notice that for CFL and OCFL the number of open facilities is upper bounded by n.
- compFL is the c_{OFL}-competitive primal-dual algorithm for the OFL from Fotakis [8] and Nagarajan and Williamson [9] papers, with $c_{\text{OFL}} \leq 4 \log n$,
- compST is the c_{OST}-competitive greedy algorithm for the OST from Imase and Waxman [13] paper, with $c_{\text{OST}} \leq \log n$,
- $i = a(j)$ means that j is connected to i by the online algorithm we are analyzing,
- $i = a^*(j)$ means that j is connected to i in the offline optimal solution with which we are comparing,
- $\text{path}(j, S)$ is the edge (j, v) with v being the closest node to j in S,
- $\text{ALG}_{\text{OCFL}}(D) = O + C + S$ is the cost of the Online CFL algorithm when serving the clients D, where O is the facility opening cost, C is the client connection cost and S is the Steiner tree cost,
- $\text{OPT}_{\text{CFL}}(D) = O^* + C^* + S^*$ is the cost of the CFL offline optimal solution with which we compare the cost of the online algorithm. It is also divided in facility opening, client connection and Steiner tree cost.

4 The Online CFL Algorithm

In this section we present a sample-and-augment algorithm for the OCFL that is based on the algorithm for the CFL presented in the Eisenbrand et al. [20] paper. Our algorithm uses the algorithm compFL as a subroutine when deciding which facilities to open and how to connect the clients. Also, it simulates the behavior of the compST algorithm when creating the tree that connects the open facilities.

The algorithm keeps a virtual solution that is competitive for the OFL. This solution serves all the clients that arrive and may have more open facilities, called virtual facilities, than the algorithm's actual solution. When a client j arrives it is served by compFL and connected to a virtual facility i. Also, the client j is sampled with probability $\frac{1}{M}$. A virtual facility is actually opened by the algorithm only when a client that was connected to it is sampled. If the client j is sampled i is opened and j is connected to it. Otherwise j is connected to the closest actually opened facility. Notice that i could already be open due to some previous sampled client.

The algorithm builds the tree T that connects the facilities as follows. When a facility i is actually opened, due to a client j that was sampled, the algorithm connects the client j to the tree T, using the shortest edge, and augments T to connect i to j. Although connecting the facility i directly to the tree T seems more intuitive, this behavior of the algorithm is useful during the analysis.

Data: $G = (V, E)$, d, f, F, root r and M
$D \leftarrow \emptyset$; $F' \leftarrow \emptyset$; $T \leftarrow \emptyset$;
$f(r) \leftarrow 0$;
send r to compFL;
$F' \leftarrow F' \cup \{r\}$; $V(T) \leftarrow V(T) \cup \{r\}$;
while *a new client j arrives* **do**
    send j to compFL;
    sample j with probability $p = \frac{1}{M}$;
    if *j was sampled and connected to a facility i that wasn't open* **then**
        $F' \leftarrow F' \cup \{i\}$;
        $T \leftarrow T \cup \{(i, j)\} \cup \{path(j, V(T))\}$;
    end
    let i be the closest open facility to j;
    $D \leftarrow D \cup \{j\}$; $a(j) \leftarrow i$;
end
return $(F' \setminus \{r\}, T, a)$;

Algorithm 1: The Online CFL algorithm.

4.1 Analysis of the Online CFL Algorithm

During this analysis we let $D' \subseteq D$ denote the clients sampled by the Online CFL algorithm. Note that D' is a random set.

First we bound the facility opening cost of the algorithm.

Lemma 1. *$O \leq c_{\text{OFL}}(O^* + C^*)$.*

Proof. Let O_{compFL} be the facility opening cost paid by compFL to serve $\{r\} \cup D$. Once our algorithm opens a subset of the facilities opened by compFL to serve $\{r\} \cup D$ we have that:

$$O \leq O_{\text{compFL}} \leq c_{\text{OFL}} \text{OPT}_{\text{FL}}(\{r\} \cup D) \leq c_{\text{OFL}}(O^* + C^*) \ , \quad (1)$$

where the last inequality follows since the optimal solution for CFL is a feasible solution for the OFL. □

Now we bound the expected cost of the Steiner tree T that connects the root and the opened facilities to each other.

Lemma 2. $E[S] \le c_{\mathrm{OST}}(S^* + C^*) + c_{\mathrm{OFL}}(O^* + C^*)$.

Proof. Define D'' as the set of sampled clients that were responsible for opening the facilities opened by the algorithm. The Online CFL algorithm builds a tree T by connecting the root r to each client in D'' using the compST algorithm. Then it augments T connecting each client $j \in D''$ to the facility i that was opened by it. So, we have that:

$$\begin{aligned} S &\le M\mathrm{compST}(\{r\} \cup D'') + M \sum_{j \in \{r\} \cup D''} d(j, a(j)) \\ &\le M c_{\mathrm{OST}} \mathrm{OPT}_{\mathrm{ST}}(\{r\} \cup D'') + M \sum_{j \in D''} d(j, a(j)) \\ &\le M c_{\mathrm{OST}} \mathrm{OPT}_{\mathrm{ST}}(\{r\} \cup D') + M \sum_{j \in D'} d(j, a(j)) \ , \end{aligned} \tag{2}$$

where the second inequality follows because $d(r, a(r)) = 0$ and the last inequality follows because $D'' \subseteq D'$.

We bound the expected cost of $\mathrm{OPT}_{\mathrm{ST}}$ when serving $\{r\} \cup D'$ as follows:

$$\begin{aligned} E[\mathrm{OPT}_{\mathrm{ST}}(\{r\} \cup D')] &\le E\left[\frac{S^*}{M}\right] + E\left[\sum_{j \in D'} d(j, a^*(j))\right] \\ &\le \frac{S^*}{M} + \sum_{j \in D} \frac{1}{M} d(j, a^*(j)) \le \frac{S^*}{M} + \frac{C^*}{M} \ , \end{aligned} \tag{3}$$

where the first inequality follows because the union of the optimal Steiner tree for CFL, T^*, and a connection from each client in D' to its facility in the optimal solution contains a tree that spans $\{r\} \cup D'$, since $r \in V(T^*)$; the third inequality follows because the probability that a client is sampled is $\frac{1}{M}$.

Let $a_{\mathrm{compFL}}(j)$ be the facility to which compFL connected j and C_{compFL} be the client connection cost of compFL when serving $\{r\} \cup D$. We bound the expected cost of $\sum_{j \in D'} d(j, a(j))$ as follows:

$$\begin{aligned} E\left[\sum_{j \in D'} d(j, a(j))\right] &= E\left[\sum_{j \in D'} d(j, a_{\mathrm{compFL}}(j))\right] \\ &= \sum_{j \in D} \frac{1}{M} d(j, a_{\mathrm{compFL}}(j)) \\ &= \frac{C_{\mathrm{compFL}}}{M} \\ &\le \frac{c_{\mathrm{OFL}}}{M} \mathrm{OPT}_{\mathrm{FL}}(\{r\} \cup D) \le \frac{c_{\mathrm{OFL}}}{M}(O^* + C^*) \ , \end{aligned} \tag{4}$$

where the first equality follows because $a(j) = a_{\mathrm{compFL}}(j)$ for $j \in D'$ and the last equality follows because $d(r, a_{\mathrm{compFL}}(r)) = 0$. The last inequality follows since the optimal solution for CFL is a feasible solution for the OFL.

Using the last three inequalities we have:

$$
\begin{aligned}
E[S] &\le E\left[Mc_{\mathrm{OST}}\mathrm{OPT}_{\mathrm{ST}}(\{r\} \cup D')\right] + E\left[M \sum_{j \in D'} d(j, a(j))\right] \\
&\le Mc_{\mathrm{OST}}\left(\frac{S^*}{M} + \frac{C^*}{M}\right) + M\left(\frac{c_{\mathrm{OFL}}}{M}(O^* + C^*)\right) \\
&\le c_{\mathrm{OST}}(S^* + C^*) + c_{\mathrm{OFL}}(O^* + C^*) \ , \qquad (5)
\end{aligned}
$$

which concludes the lemma. □

Using the two previous lemmas we can bound the expectation of the facility opening cost O and of the Steiner tree cost S of the Online CFL algorithm. Now we will bound the client connection cost.

Lemma 3. $E[C] \le c_{\mathrm{OFL}}(O^* + C^*) + c_{\mathrm{OST}}(S^* + C^* + c_{\mathrm{OFL}}(O^* + C^*))$.

Proof. First we define cost shares that divide the expected cost of compFL($\{r\} \cup D'$) and compST(F') between the clients. For each client j we call its cost share the buying cost b_j. We also divide the client connection cost C between the clients and, for each client j we call its share the renting cost r_j.

Then we analyze the algorithms compFL and compST to show that, when each client j arrives, the expected renting cost of j is at most its expected buying cost. Finally, using the linearity of expectation and summing over all the clients in D, we conclude the lemma.

Remember that D' is the set of sampled clients, $a(j)$ is the open facility to which the Online CFL algorithm connected j and $a_{\mathrm{compFL}}(j)$ is the open facility to which compFL connected j. Also, let $n(j)$ be the position of client j in the sequence of clients and $F'_{n(j)}$ be the set of facilities that were opened after the first $n(j)$ clients were served by the algorithm. Let $F'_{n(j)-1}$ be the facilities opened in the time step prior to the arrival of j.

Now we define the buying cost of a client j as:

$$
\begin{aligned}
&b_j = Md(j, a_{\mathrm{compFL}}(j)) + Md(a_{\mathrm{compFL}}(j), F'_{n(j)-1}) && \text{if } j \in D' \ , \\
&b_j = 0 && \text{if } j \text{ was not sampled} \ .
\end{aligned}
$$

Summing over all the clients in D we have:

$$
\begin{aligned}
\sum_{j \in D} b_j &= \sum_{j \in D'} b_j \\
&= \sum_{j \in D'} Md(j, a(j)) + \sum_{j \in D'} M(d(a(j), F'_{n(j)-1})) \\
&= M \sum_{j \in D'} d(j, a(j)) + M\mathrm{compST}(\{r\} \cup F') \ . \qquad (6)
\end{aligned}
$$

Let C_{compFL} be the client connection cost of compFL when serving $\{r\} \cup D$. We have that:

$$E\left[\sum_{j\in D'} d(j, a(j))\right] = \sum_{j\in D} \frac{1}{M} d(j, a_{\text{compFL}}(j)) = \frac{C_{\text{compFL}}}{M} \quad . \tag{7}$$

Similarly:

$$E\left[\sum_{j\in D'} d(j, a^*(j))\right] = \sum_{j\in D} \frac{1}{M} d(j, a^*(j)) = \frac{C^*}{M} \quad . \tag{8}$$

Bounding the expected cost of compST when it is serving $\{r\} \cup F'$ we have:

$$\begin{aligned} E[\text{compST}(\{r\} \cup F')] &\le c_{\text{OST}} E[\text{OPT}_{\text{ST}}(\{r\} \cup F')] \\ &\le c_{\text{OST}} E\left[\frac{S^*}{M} + \sum_{j\in D'} d(j, a^*(j)) + \sum_{j\in D'} d(j, a(j))\right] \\ &\le c_{\text{OST}} \left(\frac{S^*}{M} + \frac{C^*}{M} + \frac{C_{\text{compFL}}}{M}\right) \quad , \end{aligned} \tag{9}$$

where the second inequality follows because the union of the optimal Steiner tree for CFL, T^*, along with each client in D' connected to its facility in the optimal solution and its facility in the online solution, contains a tree that spans $\{r\} \cup F'$, since $r \in V(T^*)$.

Using the previous inequalities we bound the expected value of the total buying cost by:

$$\begin{aligned} E\left[\sum_{j\in D} b_j\right] &\le M \cdot E\left[\sum_{j\in D'} d(j, a(j))\right] + M \cdot E[\text{compST}(\{r\} \cup F')] \\ &= C_{\text{compFL}} + c_{\text{OST}}(S^* + C^* + C_{\text{compFL}}) \\ &\le c_{\text{OFL}} \text{OPT}_{\text{FL}}(\{r\} \cup D) + c_{\text{OST}}(S^* + C^* + c_{\text{OFL}} \text{OPT}_{\text{FL}}(\{r\} \cup D)) \\ &\le c_{\text{OFL}}(O^* + C^*) + c_{\text{OST}}(S^* + C^* + c_{\text{OFL}}(O^* + C^*)) \quad . \end{aligned} \tag{10}$$

We define the renting cost of a client j as follows:

$$r_j = d(j, F'_{n(j)}) \quad .$$

Summing over all the clients in D we have:

$$\sum_{j\in D} r_j = \sum_{j\in D} d(j, F'_{n(j)}) = C \quad , \tag{11}$$

where the last equality holds because the Online CFL algorithm connects each new client to the closest facility that was open at that moment.

Now we upper bound the expected renting cost of a client j using the expected buying cost of j. First we analyze the expectation of buying and of renting conditioned on the result of the first $n(j)-1$ coin tosses. We denote these respectively as $E[b_j|n(j)-1]$ and $E[r_j|n(j)-1]$.

Recalling that a client j has probability $\frac{1}{M}$ of being sampled and probability $\frac{M-1}{M}$ of not, we have:

$$\begin{aligned} E[r_j|n(j)-1] &= \frac{M-1}{M} d(j, F'_{n(j)-1}) + \frac{1}{M} d(j, F'_{n(j)}) \\ &\leq d(j, F'_{n(j)-1}) \\ &\leq \frac{1}{M}\Big(Md(j, a_{\text{compFL}}(j)) + Md(a_{\text{compFL}}(j), F'_{n(j)-1})\Big) \\ &= E[b_j|n(j)-1] \ , \end{aligned} \tag{12}$$

where the first equality holds because $F'_{n(j)} = F'_{n(j)-1}$ when the client j is not sampled and because $F'_{n(j)-1} \subseteq F'_n(j)$; the second inequality follows by the triangle inequality and the fact that the distance from j to the closest facility to $a_{\text{compFL}}(j)$ in $F'_{n(j)-1}$ has to be at least the distance from j to its closest facility in $F'_{n(j)-1}$.

Since this is true for all possible outcomes of the $n(j)-1$ first coin tosses, the result is true unconditionally, i.e.:

$$E[r_j] \leq E[b_j] \ . \tag{13}$$

We conclude the lemma using (10), (11) and (13), relying on the linearity of expectation as follows:

$$\begin{aligned} E[C] &= \sum_{j\in D} E[r_j] \leq \sum_{j \in D} E[b_j] \\ &\leq c_{\text{OFL}}(O^* + C^*) + c_{\text{OST}}(S^* + C^* + c_{\text{OFL}}(O^* + C^*)) \ . \end{aligned} \tag{14}$$

□

Using the three previous lemmas and that the competitive ratio of compFL and compST is $\mathrm{O}(\log n)$, we prove our main result in the next theorem.

Theorem 1. $E[\text{ALG}_{\text{OCFL}}(D)] \in \mathrm{O}(\log n^2 \text{OPT}_{\text{CFL}}(D))$.

Proof.

$$\begin{aligned} E[\text{ALG}_{\text{OCFL}}(D)] &= E[O + S + C] \\ &\leq c_{\text{OFL}}(O^* + C^*) + (c_{\text{OST}}(S^* + C^*) + c_{\text{OFL}}(O^* + C^*)) \\ &\quad + (c_{\text{OFL}}(O^* + C^*) + c_{\text{OST}}(S^* + C^* + c_{\text{OFL}}(O^* + C^*))) \\ &\leq 14 \log n \text{OPT}_{\text{CFL}}(D) + 4 \log^2 n \text{OPT}_{\text{CFL}}(D) \\ &= \mathrm{O}(\log^2 n) \text{OPT}_{\text{CFL}}(D) \ , \end{aligned} \tag{15}$$

where the last inequality follows because $c_{\text{OFL}} \leq 4 \log n$ and $c_{\text{OST}} \leq \log n$. □

4.2 Analysis of the Special Case of the Online CFL Problem with $M = 1$

Here we analyze the algorithm Online CFL in the special case when the Online Connected Facilty Location problem has $M = 1$.

In this analysis we let $D' \subseteq D$ denote the clients sampled by the Online CFL algorithm. In this special case, since $M = 1$, all clients are sampled by the Online CFL algorithm, so that $D' = D$. In fact, this means that the Online CFL algorithm is deterministic in this case.

First we bound the facility opening cost O and the client connection cost C.

Lemma 4. *$O + C \leq c_{\mathrm{OFL}}(O^* + C^*)$.*

Proof. Let O_{compFL} be the facility opening cost paid by compFL to serve $\{r\} \cup D$, and let C_{compFL} be the client connection cost of compFL when serving $\{r\} \cup D$. Since the OCFL algorithm opens exactly the facilities opened by compFL to serve $\{r\} \cup D$ and connects each client to the closest facility opened when that client was served, we have that:

$$\begin{aligned} O + C &= O_{\mathrm{compFL}} + C_{\mathrm{compFL}} \\ &\leq c_{\mathrm{OFL}} \mathrm{OPT}_{\mathrm{FL}}(\{r\} \cup D) \leq c_{\mathrm{OFL}}(O^* + C^*) \ , \end{aligned} \tag{16}$$

where the last inequality follows since the optimal solution for CFL is a feasible solution for the OFL. □

Now we bound the cost of the Steiner tree that connects the root and the opened facilities.

Lemma 5. *$S \leq c_{\mathrm{OST}}(S^* + C^*) + c_{\mathrm{OFL}}(O^* + C^*)$.*

Proof. Define the set D'' to be the set of sampled clients that were responsible for opening a facility. The Online CFL algorithm builds a tree T connecting the root r to each client in D'' using the compST algorithm, and then connects each client $j \in D''$ with the facility i that was opened by it. So we have that:

$$\begin{aligned} S &\leq M \mathrm{compST}(\{r\} \cup D'') + M \sum_{j \in \{r\} \cup D''} d(j, a(j)) \\ &\leq c_{\mathrm{OST}} \mathrm{OPT}_{\mathrm{ST}}(\{r\} \cup D) + \sum_{j \in D} d(j, a(j)) \\ &\leq c_{\mathrm{OST}} \mathrm{OPT}_{\mathrm{ST}}(\{r\} \cup D) + C \ , \end{aligned} \tag{17}$$

where the second inequality follows because $M = 1$, $D'' \subseteq D$ and $d(r, a(r)) = 0$.

We bound the cost of $\mathrm{OPT}_{\mathrm{ST}}$ when serving $\{r\} \cup D$ as follows:

$$\mathrm{OPT}_{\mathrm{ST}}(\{r\} \cup D) \leq S^* + C^* \ , \tag{18}$$

where the inequality follows because the union of the optimal Steiner tree for CFL, T^*, together with a connection from each client in D to its optimal facility, contains a tree that spans $\{r\} \cup D$, since $r \in V(T^*)$.

The cost of C can be bounded using the previous lemma:

$$C \leq c_{\mathrm{OFL}}(O^* + C^*) \ . \tag{19}$$

Using the last three inequalities we have:

$$\begin{aligned} S &\leq c_{\mathrm{OST}}\mathrm{OPT}_{\mathrm{ST}}(\{r\} \cup D) + C \\ &\leq c_{\mathrm{OST}}(S^* + C^*) + c_{\mathrm{OFL}}(O^* + C^*) \ , \end{aligned} \tag{20}$$

what concludes the lemma. □

Using the two previous lemmas and the fact that the competitive ratio of compFL and compST is $\mathrm{O}(\log n)$, we prove the next theorem.

Theorem 2. *When* $M = 1$ *we have that* $\mathrm{ALG}_{\mathrm{OCFL}}(D) \in \mathrm{O}(\log n \mathrm{OPT}_{\mathrm{CFL}}(D))$.

Proof.

$$\begin{aligned} \mathrm{ALG}_{\mathrm{OCFL}}(D) &= E[O + C + S] \\ &\leq c_{\mathrm{OFL}}(O^* + C^*) + (c_{\mathrm{OST}}(S^* + C^*) + c_{\mathrm{OFL}}(O^* + C^*)) \\ &\leq 9 \log n \mathrm{OPT}_{\mathrm{CFL}}(D) \\ &= \mathrm{O}(\log n)\mathrm{OPT}_{\mathrm{CFL}}(D) \ , \end{aligned} \tag{21}$$

where the last inequality follows because $c_{\mathrm{OFL}} \leq 4 \log n$ and $c_{\mathrm{OST}} \leq \log n$. □

It is worth noticing that an algorithm that does not sample the clients, but instead solves the OFL part of the problem and connects each open facility in an online Steiner tree, can be shown to be $\mathrm{O}(M \log n)$-competitive by an analysis very similar to the previous one.

5 Conclusion and Future Work

In this paper we proposed the Online Connected Facility Location problem and presented a randomized $\mathrm{O}(\log^2 n)$-competitive algorithm for it. The algorithm uses the sample-and-augment technique. We also showed that this algorithm is a deterministic $\mathrm{O}(\log n)$-competitive algorithm for the special case of the OCFL with $M = 1$, which is the best possible.

Two natural questions arise: first, is our algorithm $\mathrm{O}(\log n)$-competitive for the general case of the OCFL? Second, is the lower bound to the competitive ratio of OCFL algorithms greater than $\Omega(\log n)$?

Also, some directions for future work are to analyze our algorithm using the known distribution model from stochastic analysis, instead of the worst case model from competitive analysis, and to find a deterministic algorithm for the OCFL problem.

References

1. Shmoys, D.B.: Approximation algorithms for facility location problems. In: Approximation Algorithms for Combinatorial Optimization, Third International Workshop, APPROX 2000. Number 1913 in Lecture Notes in Computer Science, Berlin, Germany, Springer-Verlag (2000) 27–33
2. Mahdian, M., Ye, Y., Zhang, J.: Approximation algorithms for metric facility location problems. SIAM Journal on Computing **36** (2006) 411–432
3. Byrka, J., Aardal, K.: An optimal bifactor approximation algorithm for the metric facility location problem. SIAM Journal on Computing **39** (2010) 2212–2231
4. Li, S.: A 1.488 approximation algorithm for the uncapacitated facility location problem. Information and Computation **222** (2013) 45–58
5. Borodin, A., El-Yaniv, R.: Online Computation and Competitive Analysis. Press Syndicate of the University of Cambridge (1998)
6. Meyerson, A.: Online facility location. In: Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science. (2001) 426–431
7. Fotakis, D.: On the competitive ratio for online facility location. Algorithmica **50** (2008) 1–57
8. Fotakis, D.: A primal-dual algorithm for online non-uniform facility location. Journal of Discrete Algorithms **5**(1) (2007) 141–148
9. Nagarajan, C., Williamson, D.P.: Offline and online facility leasing. In: Integer Programming and Combinatorial Optimization, 13th International Conference, IPCO 2008. Number 5035 in Lecture Notes in Computer Science, Berlin, Germany, Springer-Verlag (2008) 303–315
10. Fotakis, D.: Online and incremental algorithms for facility location. SIGACT News **42**(1) (March 2011) 97–131
11. Vazirani, V.: Approximation Algorithms. Springer (2003)
12. Williamson, D.P., Shmoys, D.B.: The Design of Approximation Algorithms. Cambridge University Press (2011)
13. Imase, M., Waxman, B.M.: Dynamic Steiner tree problem. SIAM Journal on Discrete Mathematics **4**(3) (1991) 369–384
14. Buchbinder, N., Naor, J.S.: The design of competitive online algorithms via a primal-dual approach. Foundations and Trends in Theoretical Computer Science **3** (2009) 93–263
15. Gupta, A., Kumar, A., Pál, M., Roughgarden, T.: Approximation via cost sharing: Simpler and better approximation algorithms for network design. Journal of the ACM **54**(3) (2007) Article 11.
16. Gupta, A., Srinivasan, A., Tardos, É.: Cost-sharing mechanisms for network design. Algorithmica **50** (2008) 98–119
17. Swamy, C., Kumar, A.: Primal-dual algorithms for connected facility location problems. Algorithmica **40**(4) (2004) 245–269
18. Hasan, M.K., Jung, H., Chwa, K.Y.: Approximation algorithms for connected facility location. Journal of Combinatorial Optimization **16** (2008) 155–172
19. Jung, H., Hasan, M.K., Chwa, K.Y.: A 6.55 factor primal-dual approximation algorithm for the connected facility location problem. Journal of Combinatorial Optimization **18** (2009) 258–271
20. Eisenbrand, F., Grandoni, F., Rothvoß, T., Schäfer, G.: Connected facility location via random facility sampling and core detouring. Journal of Computer and System Sciences **76**(8) (2010) 709–726