The Multi-Stop Station Location Problem

Felix J. L. Willamowski[1a], Miriam Ganz[a,b], and Erik Mühmer[a]

[a] *RWTH Aachen University, Lehrstuhl für Operations Research, Kackertstr. 7, D-52072 Aachen, Germany*
[b] *TomTom Location Technology Germany GmbH, An den Treptowers 1, D-12435 Berlin, Germany*

Abstract

We introduce the (directed) multi-stop station location problem. The goal is to install stations such that ordered (multi-)sets of stops can be traversed with respect to range restrictions that are reset whenever a station is visited. Applications arise in telecommunications and transportation, e.g., charging station placement problems. The problem generalizes several network optimization problems such as (directed) Steiner tree problems. We show strong intractability results of the directed and undirected version under different complexity assumptions; that is, there are no constant factor approximation algorithms, unless P = NP and there are no polylogarithmic approximation algorithms for the directed version, unless NP $\subseteq$ DTIME$(n^{\text{polylog}(n)})$. By a transformation from the directed version to shortest path problems we obtain a linear approximation algorithm.

1. Introduction

The *(directed) multi-stop station location* problem ((D)MSLP) is given by a (directed) graph $G = (F \cup V, E)$ with edge costs $c^E : E \to \mathbb{Q}_{\geq 0}$, edge lengths $\ell : E \to \mathbb{Q}_{\geq 0}$, and station installation costs $c^F : F \to \mathbb{Q}_{\geq 0}$. Moreover, there is a set of trips T. Each trip $t \in T$ is represented by an ordered (multi-)set of stops $\mathcal{S}_t \in V^{m_t}$ for $m_t \in \mathbb{Z}_{\geq 2}$ and is equipped with a length bound $r_t \in \mathbb{Q}_{\geq 0}$. A solution to the (D)MSLP consists of a subset of stations $F^* \subseteq F$ and a node-path $\mathcal{P}_t$ in $G[F^* \cup \mathcal{S}_t]$ for each trip $t \in T$ containing exactly its ordered stops ($\mathcal{S}_t = \mathcal{P}_t \setminus F^*$). Each path $\mathcal{P}_t$ has to obey the length bound r_t, i.e., the lengths of the subpaths between the first stop and the first used station, any two consecutive stations, the last station

[1] Corresponding author; e-mail: willamowski@or.rwth-aachen.de

and the last stop of the original path or between the first and the last stop, if $\mathcal{P}_t = \mathcal{S}_t$, may not exceed the length bound r_t. The objective is to minimize the costs for installing stations and the length of the paths:

$$\sum_{f \in F^*} c^F(f) + \sum_{t \in T} c^E(\mathcal{P}_t) \quad .$$

For the purposes of distinguishability, whenever we consider the directed version of the problem, we denote the directed graph with $D = (V, A)$, the arc costs with $c^A : A \to \mathbb{Q}_{\geq 0}$, and the arc lengths with $\ell : A \to \mathbb{Q}_{\geq 0}$.

The MSLP and DMSLP have many applications in practice such as charging station placement problems for (long-haul) e-mobility services. In the same way, the expansion of a hydrogen filling infrastructure for fuel cell powered cars can also be depicted by it. Besides its applications in the transport and mobility sector, it can also be used for planning telecommunications networks. The (directed) network design problem with relays (NDPR) [1, 2] is a related problem that arises in the context of telecommunications and logistics. A (directed) graph with edge costs, edge lengths, and costs for relays as well as a distance bound and a set of origin-destination node pairs is considered. The goal is to place the relays on the nodes in such a way that all origin-destination node pairs can be served and the costs for installing the relays and edge costs are minimized. In contrast to our problem, only the origin and destination nodes are fixed and each selected edge must only be paid once. Another related problem is the regenerator location problem [3] that deals with the design of optical networks and is a special case of the NDPR. As the quality of optical signals deteriorates with the traveled distance, the aim is to place regenerators at nodes such that all nodes are able to communicate with each other while minimizing the installation cost. In the context of electric vehicle charging, Zheng and Peeta [4] address a variant of the NDPR in which relay nodes have capacities and the number of relay nodes in a path, which connects origin and destination, is restricted.

The MSLP not only has many applications in practice, but is also strongly related to many graph and network optimization problems in theory. Consider for example the *generalized connectivity* problem (GCP), also known as *group Steiner forest* problem, where we are given an edge-weighted graph $G = (V, E)$ and a collection of distinct demands $\mathcal{D} = \{(S_1, T_1), \ldots, (S_k, T_k)\}$; each demand (S_i, T_i) is a pair of disjoint node subsets. The goal is to find a minimum weight subgraph connecting every demand, i.e., there is a path in the subgraph between a node of S_i and a node of T_i. The MSLP generalizes the GCP as follows. We create for every edge $e \in E$ a station f_e at the same cost as the edge and for every node $v \in V$ a station f_v with

cost 0. We introduce for every $i \in [k]$ nodes s_i and t_i, a trip (s_i, t_i) with length bound $r_i = 0$, and stations $f^v_{S_i}$ for $v \in S_i$ and $f^v_{T_i}$ for $v \in T_i$ with cost 0. Furthermore, for each $i \in [k]$ we set the lengths of the edges $\{s_i, s\}$ for $s \in S_i$, $\{t_i, t\}$ for $t \in T_i$, and $\{v, f_v\}$ for $v \in S_i \cup T_i$ to 0. For every edge $e = \{u, v\} \in E$ we set the length of the edges $\{f_u, f_e\}$ and $\{f_v, f_e\}$ to be 0. Finally, we set all not considered edge lengths to 1 and all edge costs to 0. It is easy to see that there is a correspondence between solutions of the both instances and that these solutions have the same cost. Therefore, the MSLP generalizes problems such as shortest path, minimum spanning tree, Steiner tree, Steiner forest, non-metric facility location, tree multicast, and group Steiner tree [5]. Moreover, since the previous problems are a special cases of the MSLP, lower bounds carry over. In particular, there is no $\mathcal{O}(\log^{2-\varepsilon}(n))$ approximation for group Steiner tree [6], unless NP $\subseteq$ ZTIME$(n^{\text{polylog}(n)})$ and no $(1-\varepsilon)\ln(n)$-approximation for non-metric facility location, unless P = NP [7]. Despite of the proof sketch we have stated, we will perform a reduction from the set cover problem to MSLP in Section 2 to obtain properties of the MSLP instances more relevant in practice.

The DMSLP is additionally strongly related to some other problems. Consider for example the *directed Steiner forest* problem (DSFP), also known as *directed Steiner network* problem, where we are given a directed graph $D = (V, A)$, arc costs $c : A \to \mathbb{Q}_{\geq 0}$, and a set of (ordered) demands $\mathcal{D} = \{(s_1, s_2), \ldots, (s_k, t_k)\}$; each demand (s_i, t_i) is a pair of distinct vertices. The goal is to find a minimum weight subgraph connecting the demands, i.e., there is a directed (s, t)-path in the subgraph for every demand pair $(s, t) \in \mathcal{D}$. The DMSLP generalizes the DSFP as follows. We create for every arc $a \in A$ a station f_a with the same cost as the arc and for every vertex $v \in V$ a station f_v with cost 0. We introduce for every $i \in [k]$ vertices s_i and t_i, and a trip (s_i, t_i) with length bound $r_i = 0$ and we set the lengths of the arcs (s_i, f_{s_i}) and (f_{t_i}, t_i) to be 0. For every arc $a = (u, v) \in A$ we set the lengths of the arcs (f_u, f_a) and (f_a, f_v) to be 0. Finally, we set all not considered arc lengths to be 1 and all arc costs to be 0. Again, it is easy to see that there is a correspondence between solutions of the both instances and that these solutions have the same cost. Therefore, the DMSLP generalizes problems such as directed shortest path, directed minimum spanning tree, and directed Steiner tree. Moreover, since the previous problems are a special cases of the DMSLP, lower bounds carry over. In particular, for any $\varepsilon > 0$ there is no $\mathcal{O}(2^{\log^{1-\varepsilon}(n)})$-approximation for DSFP [8], unless NP $\subseteq$ DTIME$(n^{\text{polylog}(n)})$. Despite the proof sketch we have stated, we will perform a reduction from the label cover problem to DMSLP in Section 2 to obtain properties of the DMSLP instances more relevant in practice. Al-

though MSLP and DMSLP capture many graph and network optimization problems, there are some properties arising in such problems that are not covered, e.g., node- or edge-disjointness.

In this paper our contribution is not only limited to formally proving the NP-hardness of both the MSLP and DMSLP, but we also show that there are no constant factor approximation algorithms, unless P = NP, even when restricting to metric instances. For DMSLP we extend this result to (non-metric) instances with only one trip and prove that there are no polylogarithmic approximations, unless NP $\subseteq$ DTIME$(n^{\text{polylog}(n)})$. Many other network design problems allow trivial approximation algorithms, but this seems not to be the case for our problems. We obtain a linear approximation algorithm by stating a non-trivial transformation from DMSLP to shortest path problems.

2. Computational Complexity

In this section, we show the intractability of MSLP and DMSLP under different complexity assumptions. The intractability results also hold if we are allowed to violate the range constraints up to a constant factor.

Definition 1. *An* (α, β)-approximation algorithm *for the (D)MSLP is a polynomial-time algorithm that computes a solution* $(F^*, \{\mathcal{P}_t\}_{t\in T})$ *not exceeding the length bound* r_t *up to a factor of* α *for each* $t \in T$ *and having cost of at most* $\beta \cdot \mathrm{OPT}$*, where* OPT *denotes the minimum cost not exceeding the original length bound* r_t *for each* $t \in T$*.*

In the following theorem, we show that MSLP is hard to approximate within any constant factor, unless NP = P, even when restricting to metric instances. Furthermore, the result carries over if algorithms are allowed to violate the length bounds by a factor of $(2 - \varepsilon)$ for every $\varepsilon > 0$. We prove this results by a reduction from the unweighted set cover problem. The *set cover* problem is given by a finite ground set of elements U, and a finite collection of subsets $\mathcal{S}$ where each $S \subseteq U$ for $S \in \mathcal{S}$. The goal is to find a subcollection $\mathcal{S}^* \subseteq \mathcal{S}$ minimizing the cardinality $|\mathcal{S}^*|$ such that each element of the ground set is covered, i.e., $\cup_{S\in\mathcal{S}^*} S = U$.

Theorem 2. *For any* $\varepsilon > 0$*, there is no* $(2 - \varepsilon, (1 - \varepsilon)\ln(\frac{\sum_{t\in T} m_t}{2}))$*-approximation algorithm for MSLP, unless* NP = P*, even if restricting to instances with* $c^E = \ell$ *is a metric.*

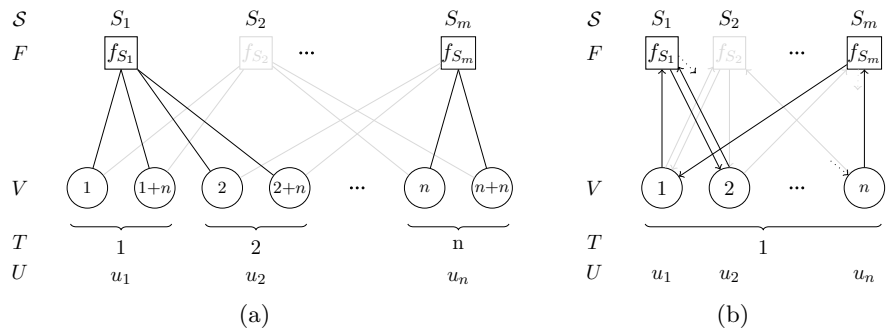

Figure 1: Schematic representation of the instances used in the set cover reductions of Theorem 2 and Theorem 3; (a) MSLP (b) DMSLP. All edges/arcs with length within the length bounds are presented in gray and feasible solutions in black.

Proof. We reduce from set cover, which is known to be $(1-\varepsilon)\ln(|U|)$-hard to approximate for every $\varepsilon > 0$, unless NP = P [7].

Let $U = \{u_1, u_2, \ldots, u_n\}$ be the ground set and $\mathcal{S}$ be the collection of subsets of the set cover instance. We construct an instance of MSLP as follows (cf. Fig. 1a). We create a node set $V = \{1, 2, \ldots, 2n\}$ and for each subset $S \in \mathcal{S}$ a station location f_S, i.e., the station locations are given by $F = \{f_S \mid S \in \mathcal{S}\}$ with costs $c^F \equiv M \in \mathbb{Z}_{\geq 0}$. We will determine the value of M later. The graph of the multi-stop station location instance is the complete graph on $V \cup F$. We set the costs and lengths of the edges $\{\{i, f_S\} \mid i \in [n], S \in \mathcal{S}, i \in S\}$ and $\{\{i+n, f_S\} \mid i \in [n], S \in \mathcal{S}, i \in S\}$ to be 1 and all other edge costs and lengths to be 2. For each $t \in T := [n]$, we define a trip with stops $\mathcal{S}_t = (t, t+n)$ and length bound $r_t = 1$ representing the element $u_t \in U$ of the ground set of the set cover instance.

Given a feasible cover of the set cover instance with subcollection $\mathcal{S}'$, we create a feasible solution to the MSLP instance as follows. For each set $S \in \mathcal{S}'$ we add the station $f_S \in F$ to the solution. We know that for each element $u_t \in U$ in the set cover instance there is some set $S_t \in \mathcal{S}'$ such that $u_t \in S_t$; we then add the path $\mathcal{P}_t = (t, f_{S_t}, t+n)$ to the solution. The resulting solution to the MSLP instance has cost of at most $M \cdot |\mathcal{S}'| + 2n$. Since this holds for every feasible solution of the set cover instance, we have $\mathrm{OPT}_{\text{MSLP}} \leq M \cdot \mathrm{OPT}_{\text{SC}} + 2n$, where OPT_{SC} and $\mathrm{OPT}_{\text{MSLP}}$ denote the optimal values of the set cover and the MSLP instance, respectively.

Let us assume that there is a $(2-\varepsilon, (1-\varepsilon)\ln(\frac{\sum_{t\in T} m_t}{2}))$-approximation algorithm for MSLP for some $\varepsilon > 0$. We then get a solution $(F^*, \{\mathcal{P}_t\}_{t\in T})$ by only using edges with length 1. Therefore, for each $t \in [n]$ the path $\mathcal{P}_t$

has to use an edge $\{t, f_{S_t}\}$ for some set $S_t \in \mathcal{S}$ with $u_t \in S_t$, since all other edges incident to t have length 2. Furthermore, there has to be a path from f_{S_t} to $t+n$ for $t \in [n]$ ending with an edge $\{f_{S_{t+n}}, t+n\}$ with $t \in S_{t+n}$, since otherwise the edge would have length 2. Thus, removing this path and using the edge $\{f_{S_t}, t+n\}$ does not increase the cost. So, we can and do assume that for each trip $t \in T$ the path $\mathcal{P}_t$ has exactly the form $(t, f_{S_t}, t+n)$ for some $S_t \in \mathcal{S}$ with $t \in S$. Thus, F^* induces a solution to the set cover instance as follows. For $t \in [n]$ we add S_t to the set cover solution. Note that the total number of sets used is exactly equal to the number of stations $|F^*|$ and that the cost of the MSLP solution is $M \cdot |F^*| + 2n$ which is bounded from above by $(1-\varepsilon)\ln(\frac{\sum_{t\in T} m_t}{2}) \cdot \mathrm{OPT}_{\mathrm{MSLP}}$. Then, the total number of sets in the constructed set cover solution is

$$\begin{aligned} |F^*| &\leq \frac{(1-\varepsilon)\ln(\frac{\sum_{t\in T} m_t}{2}) \cdot \mathrm{OPT}_{\mathrm{MSLP}} - 2n}{M} \\ &\leq (1-\varepsilon)\ln(n) \cdot \mathrm{OPT}_{\mathrm{SC}} + \frac{2n \cdot (\ln(n) - 1)}{M}, \end{aligned}$$

bounded from above by $(1-\varepsilon)\ln(n) \cdot \mathrm{OPT}_{\mathrm{SC}}$ for $M = 2n^2$ and all sufficiently large values of n. Thus, there is a $(1-\varepsilon)\ln(n)$-approximation algorithm for set cover, implying P = NP. □

For DMSLP we improve the hardness result of Theorem 2 and show that the result also holds in case of having only one trip.

Theorem 3. *For any $\varepsilon > 0$, there is no $(1-\varepsilon)\ln(\sum_{t\in T} m_t - 1)$-approximation algorithm for the DMSLP, even if restricting to instances with $|T| = 1$, unless* P = NP.

Proof. The proof is similar to the proof of Theorem 2, except that we can reduce the number of stops, combine them into one trip, and change some arc lengths (cf. Fig. 1b). We create a DMSLP instance with vertices $V = [n]$ and station locations $F = \{f_S \mid S \in \mathcal{S}\}$ with $c^F \equiv 1$. We set the lengths of the arcs $\{(i, f_S) \mid S \in \mathcal{S}, i \in S\}$, $\{(f_S, i+1) \mid S \in \mathcal{S}, i \in S\}$, and $\{(f_S, 1) \mid S \in \mathcal{S}, n \in S\}$ to be 1 and all other arc lengths to be 3. For the ease of presentation, we set $c^E \equiv 0$. We define one trip with stops $\mathcal{S}_1 = (1, 2, \ldots, n, 1)$ and length bound $r_1 = 2$. Note, that each vertex $i \in V$ is associated with an element $u_i \in U$ and all outgoing arcs with length at most 2 are the arcs $\{(i, f_S) \mid S \in \mathcal{S}, i \in S\}$. Following the reasoning of the proof of Theorem 2, we have that each feasible solution of the set cover instance transforms into a feasible solution of the DMSLP instance and vice

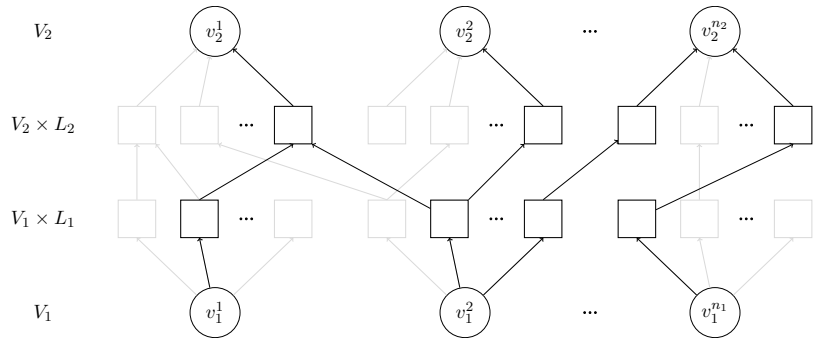

Figure 2: Schematic representation of the directed multi-stop station location instance used in the label cover reduction of Theorem 4, with edges $\{v_1^1, v_2^1\}$, $\{v_1^2, v_2^1\}$, $\{v_1^2, v_2^2\}$, $\{v_1^2, v_2^{n_2}\}$, and $\{v_1^{n_1}, v_2^{n_2}\}$. All arcs with length at most the length bound are presented in gray. A Feasible solution is depicted in black.

versa. Moreover, since $c^F \equiv 1$ and $c^A \equiv 0$ the cost of the DMSLP solution equals the total number of sets used in the set cover solution. We conclude that if there is a $(1-\varepsilon)\ln(\sum_{t\in T} m_t - 1)$-approximation algorithm for some $\varepsilon > 0$ for DMSLP with only one trip, then there is a $(1-\varepsilon)\ln(n)$-approximation algorithm for set cover, implying P = NP. □

We strengthen the inapproximability result of Theorem 3 to a hardness of $2^{\log^{1-\varepsilon}(\sum_{t\in T} m_t)}$ for every $\varepsilon > 0$, under the stronger complexity assumption NP $\not\subseteq$ DTIME$(n^{\text{polylog}(n)})$ by a reduction from a minimization version of label cover [9, Chapter 16.4]. This *label cover* problem is given by a bipartite graph $G = (V_1 \cup V_2, E)$, possible labels L_1 for V_1 and possible labels L_2 for V_2, and a non-empty relation $\emptyset \neq R_{v_1,v_2} \subseteq L_1 \times L_2$ of acceptable labels for each edge $\{v_1, v_2\} \in E$ with $v_1 \in V_1$ and $v_2 \in V_2$. The goal is to find sets of labels L_{v_1} and L_{v_2} for each node $v_1 \in V_1$ and $v_2 \in V_2$ such that for every edge $\{v_1, v_2\} \in E$ with $v_1 \in V_1$ and $v_2 \in V_2$ there is a label $\ell_1 \in L_{v_1}$ and a label $\ell_2 \in L_{v_2}$ with $(\ell_1, \ell_2) \in R_{v_1,v_2}$, while minimizing the total number of labels $\sum_{v_1\in V_1} |L_{v_1}| + \sum_{v_2\in V_2} |L_{v_2}|$.

Theorem 4. *For any $\varepsilon > 0$, there is no $2^{\log^{1-\varepsilon}(\frac{\sum_{t\in T} m_t}{2})}$-approximation algorithm for DMSLP, unless* NP $\subseteq$ *DTIME$(n^{\text{polylog}(n)})$, even if restricting to instances with $c^E = \ell$ obeys the triangle inequality.*

Proof. We reduce from label cover, which is known to be $2^{\log^{1-\varepsilon}|E|}$-hard to approximate for every $\varepsilon > 0$, unless NP $\subseteq$ DTIME$(n^{\text{polylog}(n)})$ [9, Theorem 16.37]. Let $(V_1 \cup V_2, E)$ be the input graph of the label cover instance, L_1

and L_2 be the labels for V_1 and V_2 respectively, and R_{v_1,v_2} be the relation for $\{v_1, v_2\} \in E$ with $v_1 \in V_1$ and $v_2 \in V_2$. We create an instance of DMSLP as follows (cf. Fig. 2). We create a vertex set $V = V_1 \cup V_2$ and station locations $F = (V_1 \times L_1) \cup (V_2 \times L_2)$ and let the costs of the station locations be M. We will determine the value of M later. We denote the station locations of $V_1 \times L_1$ and $V_2 \times L_2$ by pairs such as (v_1, ℓ_1) and (v_2, ℓ_2), respectively. We set the lengths of the following arcs to 1:

$$\begin{aligned} A_1 &= \{(v_1, (v_1, \ell_1)) \mid v_1 \in V_1, \ell_1 \in L_1\}, \\ A_{1,2} &= \{((v_1, \ell_1), (v_2, \ell_2)) \mid v_1 \in V_1, v_2 \in V_2, (\ell_1, \ell_2) \in R_{v_1,v_2}\}\}, \text{ and} \\ A_2 &= \{\{(v_2, \ell_2), v_2\} \mid v_2 \in V_2, \ell_2 \in L_2\}. \end{aligned}$$

We set all other arc lengths to 2; all values larger than 1 would also be sufficient. For each edge $\{v_1, v_2\} \in E$ with $v_1 \in V_1$ and $v_2 \in V_2$ we define a trip $t = (v_1, v_2)$ with stops $\mathcal{S}_t = (v_1, v_2)$ and length bound $r_t = 1$.

Given a feasible labeling for the label cover instance with labels $L_{v_1} \subseteq L_1$ for each $v_1 \in V_1$ and labels $L_{v_2} \subseteq L_2$ for each $v_2 \in V_2$, we create a feasible solution to the DMSLP instance as follows. For each node $v \in V_1 \cup V_2$ and label $\ell \in L_v$ we add the station (v, ℓ) to the solution. We know that for each edge $\{v_1, v_2\} \in E$ with $v_1 \in V_1$ and $v_2 \in V_2$ in the label cover solution there is some label $\ell_1 \in L_{v_1}$ and $\ell_2 \in L_{v_2}$ such that $(\ell_1, \ell_2) \in R_{v_1,v_2}$; we then add the path $\mathcal{P}_{(v_1,v_2)} = (v_1, (v_1, \ell_1), (v_2, \ell_2), v_2)$ to the solution. The resulting solution to the DMSLP instance has cost of at most $M \cdot (\sum_{v_1 \in V_1} |L_{v_1}| + \sum_{v_2 \in V_2} |L_{v_2}|) + 3|E|$. Since this holds for every feasible solution of the label cover instance, we have that the optimal value of the DMSLP instance ($\mathrm{OPT}_{\mathrm{DMSLP}}$) is bounded from above by $M \cdot \mathrm{OPT}_{\mathrm{LC}} + 3|E|$, where $\mathrm{OPT}_{\mathrm{LC}}$ denotes the optimal value of the label cover instance.

Let us assume that there is a $(2 - \varepsilon, 2^{\log^{1-\varepsilon} \frac{\sum_{t \in T} m_t}{2}})$-approximation algorithm for DMSLP for some $\varepsilon > 0$. We then get a solution $(F^*, \{\mathcal{P}_t\}_{t \in T})$ only using arcs in $A_1 \cup A_{1,2} \cup A_2$, since $r_t < 2$ for every $t \in T$. Note that in $A_1 \cup A_{1,2} \cup A_2$ the only directed path connecting vertices $v_1 \in V_1$ and $v_2 \in V_2$ (if possible) uses the arcs $(v_1, (v_1, \ell_1))$, $((v_1, \ell_1), (v_2, \ell_2))$, and $((v_2, \ell_2), v_2)$ for some labels $\ell_1 \in L_1$ and $\ell_2 \in L_2$. By construction $(\ell_1, \ell_2) \in R_{v_1,v_2}$, since otherwise the arc $((v_1, \ell_1), (v_2, \ell_2))$ would have length 2. Thus, the stations F^* induce a label cover solution as follows. For each path $\mathcal{P}_{(v_1,v_2)} = (v_1, (v_1, \ell_1), (v_2, \ell_2), v_2)$ we label v_1 with ℓ_1 and v_2 with ℓ_2. Note that the total number of labels used is exactly equal to the number of stations $|F^*|$. The cost of the DMSLP is $M \cdot |F^*| + 3|E|$ which is bounded from above by $2^{\log^{1-\varepsilon} \frac{\sum_{t \in T} m_t}{2}}$. Then, the total number of labels in the label cover solution

is

$$
\begin{aligned}
|F^*| &\leq \frac{2^{\log^{1-\varepsilon} \frac{\sum_{t\in T} m_t}{2}} \cdot \mathrm{OPT}_{\text{DMSLP}} - 3|E|}{M} \\
&\leq 2^{\log^{1-\varepsilon}|E|} \cdot \mathrm{OPT}_{\text{LC}} + \frac{3|T| \cdot 2^{\log^{1-\varepsilon}|T|} - 3|T|}{M}
\end{aligned}
$$

bounded from above by $2^{\log^{1-\varepsilon}|E|} \cdot \mathrm{OPT}_{\text{LC}}$ for $M = |T|^2$ and all sufficiently large values of $|E| = |T|$. Thus, there is a $2^{\log^{1-\varepsilon}|E|}$-approximation algorithm for label cover, implying NP $\subseteq$ DTIME$(n^{\text{polylog}(n)})$. □

Remark 5. *For all values of $|T|$, the total number of stops is at least as large as the number of nodes $\sum_{t\in T} m_t \geq |V|$. Therefore, Theorem 2, Theorem 3, and Theorem 4 carry over to hardness results in terms of $|V|$.*

3. Approximation Algorithm

In the previous section, we investigated the intractability of MSLP and DMSLP and have seen that it is unlikely in terms of computational complexity that there are approximation algorithms with small factors. In this section we provide a linear approximation algorithm. For the following theorem it is worth mentioning that we can assume that $\sum_{t\in T} m_t \geq 2|T|$.

Theorem 6. *There is a $(\sum_{t\in T} m_t - |T|)$-approximation for (D)MSLP.*

In order to prove this theorem, we first investigate the case $|T| = 1$ with the following lemma. Then, Theorem 6 follows immediately. For the case $|T| = 1$ we transform instances of DMSLP to instance of the directed (s,t)-shortest path problem, which is given by a directed graph and arc weights; the goal is to find a weight-minimum directed subgraph that admits a directed path from s and t.

Lemma 7. *There is a $(\sum_{t\in T} m_t - |T|)$-approximation algorithm for DMSLP instances having only one trip ($|T| = 1$).*

Proof. Let $(F, V, c^F, c^A, \ell, \mathcal{S}_1, r_1)$ be a DMSLP instance $|T| = 1$ and with stops $\mathcal{S}_1 = (v^1, v^2, \ldots, v^{m_1})$. We create a shortest path instance as follows. For each $i \in [m_1 - 1]$ we create two copies of the station locations $F_1^i = F$ and $F_2^i = F$ and refer to the copies as $(f,i)_1$ and $(f,i)_2$, respectively. Furthermore, we create two vertices s and t. Then, the vertex set of the shortest path instance is given by $\{v^1\} \cup (\cup_{i\in[m_1-1]} F_1^i) \cup (\cup_{i\in[m_1-1]} F_2^i) \cup \{v^{m_1}\}$ with

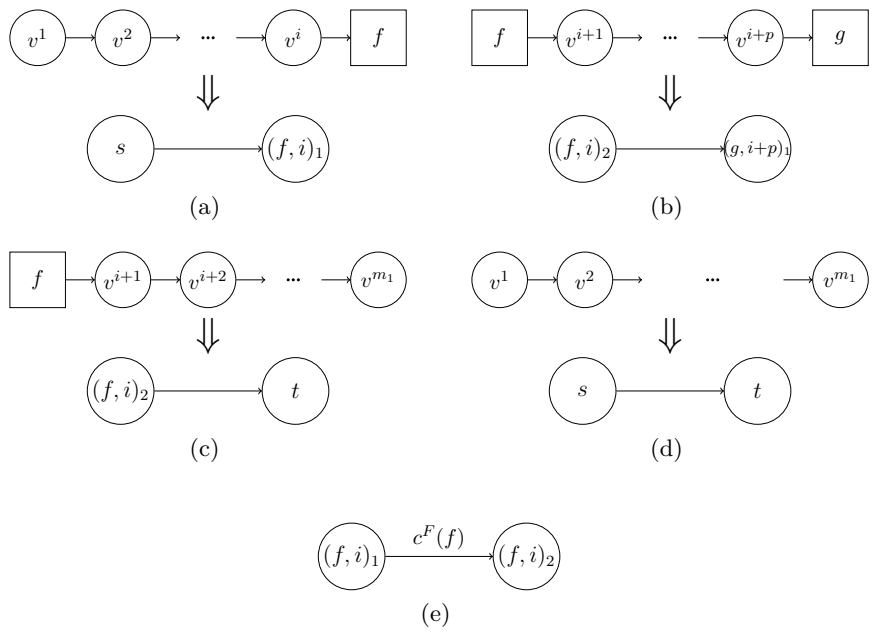

Figure 3: Schematic construction of the shortest path instance in Lemma 7.

$s = v^1$ and $t = v^{m_1}$. We create the arc set of the shortest path instance, representing paths in the DMSLP instance obeying the length bound, as follows. Note that we can and do assume that a feasible solution of the DMSLP instance contains each station at most once between two consecutive stops, because otherwise we could remove the directed cycle without increasing the cost. For $i \in [m_1 - 1]$, we first create arcs from s to $(f,i)_1 \in F_1^i$ if the length of the path $(v^1, v^2, \ldots, v^i, f)$ is within the length bound r_1 and set the weight to $c^A(v^1, v^2, \ldots, v^i, f)$ (cf. Fig. 3a); second, arcs from $(f,i)_2 \in F_2^i$ to $(g, i+p)_1 \in F_1^{i+p}$ for $p \in \{0, 1, \ldots, m_1 - i - 1\}$ if the length of the path $(f, v^{i+1}, v^{i+2}, \ldots, v^{i+p}, g)$ is within the length bound r_1 and set the weight to $c^A(f, v^{i+1}, v^{i+2}, \ldots, v^{i+p}, g)$ (cf. Fig. 3b); third, arcs from $(f,i)_2 \in F_2^i$ to t if the length of the path $(f, v^{i+1}, v^{i+2}, \ldots, v^{m_1})$ is within the length bound r_1 and set the weight to $c^A(f, v^{i+1}, v^{i+2}, \ldots, v^{m_1})$ (cf. Fig. 3c), and fourth, an arc from s to t if the length of the path $(v^1, v^2, \ldots, v^{m_1})$ is within the length bound r_1 and set the weight to $c^A(v^1, v^2, \ldots, v_1^m)$ (cf. Fig. 3d). Finally, we create arcs from $(f,i)_1 \in F_1^i$ to $(f,i)_2 \in F_2^i$ with weight $c^F(f)$ (cf. Fig. 3e).

Let $(F^*, \mathcal{P}_1)$ be a feasible solution of the DMSLP instance with the path given by $\mathcal{P}_1 = (v^1, f_1^1, f_2^1, \ldots, f_{k_1}^1, v^2, f_1^2, f_2^2, \ldots, f_{k_{m_1-1}}^{m_1-1}, v^{m_1})$. Then, a feasi-

ble solution to the shortest path instance is provided by

$$\mathcal{P} = (s, (f_1^1, 1)_1, (f_1^1, 1)_2, (f_2^1, 1)_1, (f_2^1, 1)_2, \ldots, \\ (f_{k_1}^1, 1)_1, (f_{k_1}^1, 1)_2, (f_1^2, 2)_1, (f_1^2, 2)_2, (f_2^2, 2)_1, \ldots, (f_{k_{m_1-1}}^{m_1-1}, m_1 - 1)_2, t).$$

This path has cost of at most $(m_1 - 1) \cdot c^F(F^*) + c^A(\mathcal{P}_1)$, since we use at most all stations F^* between two consecutive stops and the arc cost carry over. Furthermore, this holds for every feasible solution of the DMSLP instance and thus, we have $\mathrm{OPT}_{\mathrm{sp}} \leq (m_1 - 1) \cdot \mathrm{OPT}_{\mathrm{DMSLP}(1)}$, where $\mathrm{OPT}_{\mathrm{SP}}$ and $\mathrm{OPT}_{\mathrm{DMSLP}(1)}$ denote the optimal values of the shortest path instance and the DMSLP instance with only one trip, respectively.

By construction, every solution of the shortest path instance can be transformed into a solution of the DMSLP instance with at most the cost of the shortest path solution. Again, this holds for every solution and, thus, we have $\mathrm{OPT}_{\mathrm{DMSLP}(1)} \leq \mathrm{OPT}_{\mathrm{SP}}$. Since we can solve the directed (s, t)-shortest path problem to optimality, it follows that we get a solution with cost of at most $(m_1 - 1) \cdot \mathrm{OPT}_{\mathrm{DMSLP}(1)}$. □

Proof of Theorem 6. This is easy to see, since the optimal cost of each trip, if solved solely, is at most $\mathrm{OPT}_{\mathrm{DMSLP}}$. By using Lemma 7 and summation over all trips we get the desired bound. Finally, we can use the algorithm for the directed multi-stop station location problem for the undirected case. □

We have seen that we can solve each trip independently of the other trips to achieve the approximation factor of Theorem 6. However, we can improve the introduced algorithm by solving the trips iteratively. After each trip, we set the station costs of all stations already visited by a previously solved trip to zero. As a result, the cost of the computed shortest path is closer to the actual cost caused by the solution with respect to the DMSLP instance (if at least one station is visited that is also used by a previous trip). Thus, for a given problem instance, the iterative approximation algorithm constructs a solution that is at least as good as the solution constructed by the approximation algorithm that processes the trips independently.

4. Conclusion

We introduced the (directed) multi-stop station location problem, which has many applications in practice such as telecommunications and logistics, for example in charging station placement problems for long-haul bus services. For the two problems we showed a strong relationship to graph and

network design problems. For the MSLP we proved that for any $\varepsilon > 0$ there is no $(2-\varepsilon, (1-\varepsilon)\ln(\frac{\sum_{t\in T} m_t}{2}))$-approximation algorithm, unless NP = P, even when restricting to instances with $c^E = \ell$ is a metric. Moreover, we showed that for any $\varepsilon > 0$ there is no $\mathcal{O}(\log^{2-\varepsilon}(\frac{\sum_{t\in T} m_t}{2}))$-approximation, unless NP $\subseteq$ ZTIME$(n^{\text{polylog}(n)})$. For the DMSLP we proved that for any $\varepsilon > 0$, there is no $(1-\varepsilon)\ln(\sum_{t\in T} m_t - 1)$-approximation algorithm even when restricting to instances with $|T| = 1$, unless P = NP. Furthermore, we proved that for any $\varepsilon > 0$, there is no $2^{\log^{1-\varepsilon}(\frac{\sum_{t\in T} m_t}{2})}$-approximation algorithm under the stronger complexity assumption NP $\not\subseteq$ DTIME$(n^{\text{polylog}(n)})$. We developed an approximation algorithm for our problems based on shortest path computations. We derived a linear approximation factor, $(\sum_{t\in T} m_t) - T$.

Although, we proved strong lower bounds on the approximation guarantees, there is a gap between lower and upper bound for the MSLP as well as for the DMSLP. Therefore, it would be interesting whether the MSLP admits a polylogarithmic approximation algorithm or such a hardness result. For the DMSLP it would be nice to know whether there is a sublinear approximation algorithm or if such an algorithm does not exist.

Moreover, the practical aspects of the work with applications of MSLP and DMSLP in the context of electric vehicle charging station placement for long-haul e-mobility services will be considered in a future paper. Along with this, we propose (exact) approaches to solve the MSLP and DMSLP, which we evaluate with respect to practically orientated instances.

References

[1] E. A. Cabral, E. Erkut, G. Laporte, R. A. Patterson, The network design problem with relays, European Journal of Operational Research 180 (2007) 834–844.

[2] M. Leitner, I. Ljubić, M. Riedler, M. Ruthmair, Exact approaches for the directed network design problem with relays, Omega 91 (2020) 102005.

[3] S. Chen, I. Ljubić, S. Raghavan, The regenerator location problem, Networks 55 (2010) 205–220.

[4] H. Zheng, S. Peeta, Routing and charging locations for electric vehicles for intercity trips, Transportation Planning and Technology 40 (2017) 393–419.

[5] N. Alon, B. Awerbuch, Y. Azar, N. Buchbinder, J. S. Naor, A general approach to online network optimization problems, ACM Trans. Algorithms 2 (2006) 640–660.

[6] E. Halperin, R. Krauthgamer, Polylogarithmic inapproximability, in: Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing, STOC '03, Association for Computing Machinery, New York, NY, USA, 2003, pp. 585–594.

[7] I. Dinur, D. Steurer, Analytical approach to parallel repetition, in: Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing, STOC '14, Association for Computing Machinery, New York, NY, USA, 2014, pp. 624–633.

[8] Y. Dodis, S. Khanna, Design networks with bounded pairwise distance, in: Proceedings of the thirty-first annual ACM symposium on Theory of computing, STOC '99, Association for Computing Machinery, New York, NY, USA, 1999, pp. 750–759.

[9] D. P. Williamson, D. B. Shmoys, The Design of Approximation Algorithms, Cambridge University Press, USA, 1st edition, 2011.