Report in Wirtschaftsmathematik
Nr. 81/2002

The continuous stop location problem in public transportation networks

A. Schöbel[*] H.W. Hamacher[†] A. Liebers[‡] D. Wagner[§]

March 29, 2002

Abstract

In this paper we consider the location of stops along the edges of an already existing public transportation network. This can be the introduction of bus stops along some given bus routes, or of railway stations along the tracks in a railway network. The positive effect of new stops is given by the better access of the potential customers to their closest station, while the increasement of travel time caused by the additional stopping activities of the trains leads to a negative effect. The goal is to *cover* all given demand points with a minimal amount of additional traveling time, where covering may be defined with respect to an arbitrary norm (or even a gauge). Unfortunately, this problem is NP-hard, even if only the Euclidean distance is used. In this paper, we give a reduction to a finite candidate set leading to a discrete set covering problem. Moreover, we identify network structures in which the coefficient matrix of the resulting set covering problem is totally unimodular, and use this result to derive efficient solution approaches. Various extensions of the problem are also discussed.

[*]University of Kaiserslautern and Fraunhofer ITWM
[†]University of Kaiserslautern and Fraunhofer ITWM
[‡]SIEMENS AG, Konstanz
[§]University of Konstanz

1 Introduction

Establishing stops within a transportation network is fundamemtal for offering public transportation service, see e.g., [DAL82].
Stops (or stations) may have the following conflicting effects on the public transportation network. On the one hand, many stops are advantageous, since they can cover many of the (potential) customers. In bus transportation, often the covering radius is assumed to be 400 m, meaning that a customer will think about using a bus, only if the next bus stop is within a distance of at most 400 m. In rail transportation, the covering radius is larger, and is usually assumed to be 2 km. On the other hand, each additional stop will increase the transportation time (e.g. by 2 minutes), which makes the service unattractive to the customers. Moreover, stops are costly. Among the many possible different objective functions taking both mentioned effects into account, one frequently discussed goal is to establish as few stops as possible in such a way, that (nearly) all customers are covered. This has been done in [MDSF98, Mur01a, Mur01b]. In these papers, the public transportation network in Brisbane, Australia was analyzed in detail and it turned out that 84.5 % of the stops are not necessary in terms of covering demand points within an Euclidean distance of 400 m, i.e. closing them would not decrease the actual number of covered customers. The stop location problem was treated in a discrete setting in these papers, i.e., the authors either consider only the actual stops, or they assume that a finite candidate set of new stops is given. This leads to an unweighted set covering problem (like the one tackled in [TSRB71]). In the context of stop location this problem has been solved by [Mur01a] using the Lagrangean-based set covering heuristic of [CFT99]. Very recently, another discrete stop location model has been developed by Laporte et al. [LMO02]. They investigate which candidate stops along one given line in Sevilla should be opened, taking into account constraints on the interstation space. The main part of their paper deals with the determination of the coverage of a new stop using a gravitation model. Finally, the problem is solved by a longest path algorithm in an acyclic graph.

As part of a project with the largest German rail company (DB), another objective function has been considered, see [HLS$^+$01]. In this project the main goal was the minimization of the overall traveling time over all customers, which is given by the access time of the customers to their (Euclidean) closest station, their travel time within the vehicle and their time to go from the final station to their destination. Another difference to the approaches mentioned above is, that for the new stations all possible locations along the tracks were allowed in [HLS$^+$01]. It turned out that due to the savings in the access times, it is possible to decrease the overall travel times by establishing (many) new stations. This result was obtained by a genetic algorithm, using the real data of DB. A covering model similar to that in [HLS$^+$01] has been considered and solved in [KPS$^+$02].

In both models, covering was defined with respect to the Euclidean distance, i.e. the distance from a new stop to a demand point was measured in the Euclidean way. This may be unrealistic, since in most cities, distances are rather network-based, which often can be approximated by polyhedral norms or even gauges (e.g. by the rectangular distance in Manhattan).

In the current paper we combine both approaches and extend them to a more generalized concept of covering. This leads to our *continuous stop location problem* (CSLP). The main characteristics of (CSLP) are the following.

- We do not assume a given finite candidate set, but allow a continuous set of possible locations for the stops, given by the current bus routes or railway tracks, which are assumed to be a piecewise linear set in the plane.

- The goal is to cover a prespecified set of demand points (which might represent e.g. settlements, shopping centers, industrial areas, or recreational points) with minimal additional travel time. For this we assume that the traffic load, i.e. the number of customers along each bus route or railway track is given. Assuming equal weights for all tracks the goal reduces to minimizing the number of stops.

- Furthermore, we allow a more general concept of covering in the following sense. For each demand point we allow its own distance measure which we assume to be a norm or a gauge, and which can be chosen according to the local structure of the streets or the environment of the repective demand point.

The remainder of the paper is structured as follows. In the next section we present a mathematical model for the (CSLP) and show that it can be used in different practical settings (referring to the opening and/or closing of stops).

Furthermore we prove the NP-hardness of the problem. Section 3 establishes a *finite dominating set* result by showing that the continuous problem can be reduced to one where only a finite candidate set needs to be considered. This allows to apply results of (discrete) set covering problems. Using the underlying information given by the bus routes or railway tracks, we develop, however, much more efficient solution approaches. To this end, in Section 4 we analyze the special case of (CSLP) where only a linear piece of the bus routes or railway tracks is considered. We show that in this case, the covering matrix is totally unimodular and give extensions to general networks. The model is further extended in the remaining sections. In Section 5 we do not require covering of all demand points but allow only to cover a fixed percentage of the population, or try to maximize the covered population with a given number of stops. In Section 6 we develop a finite dominating set which is independent of the given definition of covering, i.e. which can be applied for any covering radius. Finally, we conclude the paper by briefly pointing out further extensions under research.

2 The continuous stop location problem

Let $\mathcal{P} \subseteq \mathbb{R}^2$ be the given finite set of demand points, each of them represented by a point p in the plane, and let $G = (V, E)$ be a planar connected graph with straight-line embedding in the plane. (In real-world data sets, the nodes represent either existing stations or important breakpoints.)

$$T = \cup_{e \in E} e = \{x \in \mathbb{R}^2 : x \in e \text{ for some } e \in E\} \subseteq \mathbb{R}^2$$

denotes the set of points on edges of G, where each edge $e \in E$ is given by a line segment in the plane. Furthermore, let c_e be the traffic load along edge $e \in E$, and let c_v be the traffic load through station $v \in V$, respectively, measured, for example, in customers per day. The goal is to establish stops (or stations), which are represented by points in T. To calculate the additional travel time induced by some set of stations $S \subseteq T$ we take the number of customers which are affected by the additional stopping activities and multiply them by the time t_{stop} for an additional stop. This is specified in the following notation:

Definition 1 *Given $s \in T$ let*

$$g(s) := \begin{cases} s & \text{if } s \in V \\ e & \text{if } s \in e, s \notin V \end{cases} .$$

Furthermore, given a finite set $S \subseteq T$ we define

$$f(S) = \sum_{s \in S} t_{\text{stop}} c_{g(s)}.$$

For an infinite set S we define $f(S) = \infty$.

Since t_{stop} is a constant, it can be neglected for the optimization process. Furthermore, note that $f(S) = |S|$ if all weights $c_i = 1$. Hence, we will refer to the *unweighted problem* if we deal with this special case of minimizing the number of stations.

To deal with the accessibility of the potential customers, we introduce the notation of *covering* with respect to a distance measure d as follows.

Definition 2 *Given $r > 0$, a demand point $p \in \mathcal{P}$ is covered by a point $s \in T$ if $d_p(p, s) \leq r$. Furthermore, given a set $\mathcal{P}$ and a set $S \subseteq T$,*

$$\text{cover}_{\mathcal{P}}(S) = \{p \in \mathcal{P} : p \textit{ is covered by some } s \in S\}.$$

If it is clear to which set $\mathcal{P}$ we refer, we just write cover*(S). Furthermore, for $s \in S$ we use* cover(s) *for* cover$(\{s\})$.

Note that d_p may be any distance measure (dependent on the demand point p), assigning some nonnegative distance to each pair of points in the plane. One choice is to imagine d as the Euclidean distance, but we will also allow other norms or gauges.

The following two problems are of interest.

(CSLP) New planning of all stations: Given $\mathcal{P}$ and T, find a set $S^* \subseteq T$ such that $\text{cover}_{\mathcal{P}}(S^*) = \mathcal{P}$ and $f(S^*)$ is minimal.

(CSLP') Opening additional stations: Given $\mathcal{P}'$, T', and a set of already existing stations $S' \subseteq T'$, find a set $S^* \subseteq T'$ such that $\text{cover}_{\mathcal{P}'}(S^* \cup S') = \mathcal{P}'$ and $f(S^*)$ is minimal.

Lemma 1 *Both problems are equivalent.*

Proof: To come from (CSLP) to (CSLP'), just define $S' := \emptyset$. For the reduction from (CSLP') to (CSLP) we need to change

$$\mathcal{P} := \mathcal{P}' \setminus \text{cover}_{\mathcal{P}'}(S'),$$

leaving the set of track $T = T'$ as it is.

QED

Lemma 1 shows that we only need to discuss one of the mentioned problems, and this will be the *continuous stop location problem* (CSLP) since it simplifies the notation. We first make the following observations.

Lemma 2 *Given an instance of (CSLP), the following properties hold.*

1. *(CSLP) has a feasible solution if and only if* $\text{cover}(T) = \mathcal{P}$.
2. *If (CSLP) has a feasible solution, then it also has a finite solution with cardinality less than or equal to* $|\mathcal{P}|$.

Proof: The first part is obvious. For the second part, if (CSLP) has a feasible solution, we know that $\text{cover}(T) = \mathcal{P}$, so for all $p \in \mathcal{P}$

$$\min_{s \in T} d_p(p, s) \leq r.$$

Choosing

$$\begin{aligned} s(p) &:= \text{argmin}_{s \in T} d_p(p, s) \quad \text{and} \\ S &:= \{s(p) : p \in \mathcal{P}\} \end{aligned}$$

yields a finite feasible solution with $|S| \leq |\mathcal{P}| < \infty$.

QED

For most other results in this paper we have to make precise what we mean by the distance measure d. Here we use an approach, originally introduced by Minkowski ([Min67]), in which distance measures are defined by their unit balls. In our practical setting such a unit ball about a given demand point p may be interpreted as the set of points B_p around p which can be reached within one unit of time, maybe within maximal 10 minutes. Since bus connections, or streets may be of different quality, this needs not be a Euclidean ball but can have other shapes. We next explain how to derive a distance measure d if such a unit ball B is given.

Let B be a compact convex set in $\mathbb{R}^2$ containing the origin in its interior. The *gauge* of a point $x \in \mathbb{R}^2$ with respect to B is then defined as

$$\gamma_B(x) := \inf\{\lambda > 0 : x \in \lambda B\}.$$

B is called the unit ball of the gauge γ_B. If, furthermore, B is symmetric with respect to the origin, the resulting function γ_B defines not only a gauge, but a norm. On the other hand, given a gauge γ, its associated unit ball is uniquely defined by

$$B_\gamma := \{x \in \mathbb{R}^2 : \gamma(x) \leq 1\}.$$

Note that any gauge γ obtains the following properties:

- $\gamma(x) \geq 0$ for all $x \in \mathbb{R}^2$
- $\gamma(x) = 0$ if and only if $x = \underline{0}$ (where $\underline{0}$ denotes the origin in $\mathbb{R}^2$)
- $\lambda\gamma(x) = \gamma(\lambda x)$ for all $x \in \mathbb{R}^2$ and for all $\lambda \geq 0$
- $\gamma(x + y) \leq \gamma(x) + \gamma(y)$ for all $x, y \in \mathbb{R}^2$ (triangle inequality),

while for norms we additionally have $\gamma(x) = \gamma(-x)$ which is not necessarily true for gauges.

Given a gauge γ we define the distance from $x \in \mathbb{R}^2$ to $y \in \mathbb{R}^2$ w.r.t. γ in the usual way by

$$d(x, y) := \gamma(x, y) := \gamma(y - x).$$

In the following we assume that some covering radius r is given and fixed (this assumption will be relaxed in Section 6), and that for each $p \in \mathcal{P}$ we furthermore have given a gauge γ_p with unit ball B_p, i.e. the distance from p to some point $x \in \mathbb{R}^2$ is given by

$$\gamma_p(x - p) = \inf\{\lambda > 0 : x \in p + \lambda B_p\}.$$

Using the denotation

$$B_p^r := p + rB_p,$$

we get

$$\gamma_p(p, x) \leq r \text{ if and only if } x \in B_p^r.$$

Next, we discuss the complexity status of (CSLP). It turns out that even the unweighted version of (CSLP) is NP-hard in the special case that γ_p is the Euclidean distance for all $p \in \mathcal{P}$.

Theorem 1 *(CSLP) is NP-hard.*

Proof: We show the NP-hardness of the following special case of (CSLP), in which we do not use weights and take the Euclidean distance as distance measure for all given demand points.

(Euclidean unweighted CSLP-Decision version) Given a finite set $\mathcal{P} \subseteq \mathbb{R}^2$, a set of points $T = \cup_{e \in E} e \subseteq \mathbb{R}^2$ of a planar connected graph with edge set E, and a positive integer $K \leq |\mathcal{P}|$ does there exist a collection of K discs $\mathcal{C}$ with radius r and center points in T, such that each $p \in \mathcal{P}$ lies in at least one of the discs $C \in \mathcal{C}$?

The proof is a reduction of Geometric Covering by Discs to the above decision version of (CSLP). Geometric Covering by Discs has been shown to be NP-complete [Joh82] and can be stated as follows:

(Geometric Covering by Discs) Given a finite set $\mathcal{P}$ of integer-coordinate points in the plane and positive integers d and $K \leq |\mathcal{P}|$, can the points of $\mathcal{P}$ be covered by at most K discs of diameter d?

Now take an instance of Geometric Covering by Discs. To construct an instance of the unweighted (CSLP) we

- use the same set $\mathcal{P}$ of points and
- the same numbers $r := \frac{1}{2}d$ and K and
- define the set of edges E of a connected planar graph as follows: For each unordered pair of points p_1 and p_2 from $\mathcal{P}$,
  - (i) add the line segment from p_1 to p_2 as an edge and also
  - (ii) add a sufficiently large piece of the bisector of p_1 and p_2 to E.

Claim: $\mathcal{P}$ can be covered by at most K discs of diameter d, if and only if $\mathcal{P}$ can be covered by at most K discs of radius r which all have their center points in $T = \bigcup E$.

To see this, first assume that $\mathcal{P}$ can be covered by some collection $\mathcal{C}$, consisting of at most K discs of diameter d. Then, for each disc $C \in \mathcal{C}$:

Case 1: If C contains no points of $\mathcal{P}$, then disregard C.

Case 2: If C contains only one point p of $\mathcal{P}$, then replace C by the disc with center point p (p is in T, since all line segments from p to any other point from $\mathcal{P}$ are in T) and radius r.

Case 3: If C contains a set of points $A \subseteq \mathcal{P}$ with $|A| \geq 2$ (i.e., C contains more than one point of $\mathcal{P}$), then replace C by a disc with center point q and radius r, where q is the center point of the smallest enclosing circle of A.

Since C covers A, the diameter of the smallest enclosing circle is smaller or equal to d, and hence the disc with radius r and center point q also covers A. Note that finding q is a well-known problem of location theory and can be done in linear time [Meg83].

Moreover it is known that q always lies on at least one bisector of points in A [EH72], such that q satisfies $q \in T$.

In summary, $\mathcal{P}$ can be covered by at most K discs of diameter d, all with center points in T.
The other direction is trivial.

QED

As a consequence, the unweighted and the weighted optimization versions of (CSLP) are NP-hard problems. In the next section we discuss the relation of these problems to set covering problems by transforming them to a discrete stop location problem.

3 A finite dominating set

The goal of this section is to reduce the (CSLP) to a set covering problem by determining a finite dominating set (FDS), i.e. a finite set of candidates $\mathcal{S}_{\text{cand}} \subseteq T$, for which we know that it contains an optimal solution S^*, if the problem is feasible at all. Throughout this section, let us hence assume that (CSLP) has a feasible solution, which is established easily by checking if

$$B_p^r \cap T \neq \emptyset \text{ for all } p \in \mathcal{P}$$

(see Lemma 2). We define

$$\mathcal{S} := \{s \in T : \gamma_p(p, s) = r \text{ for some } p \in \mathcal{P}\},$$

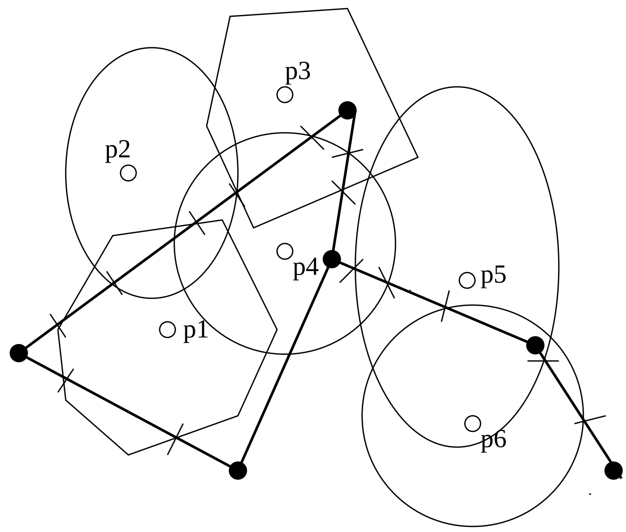

Figure 1: The set of candidates along the tracks.

i.e. $\mathcal{S}$ is given by the intersection points of T with the boundaries of the balls B_p^r, which are the unit balls B_p of radius r about the demand points p (see Figure 1).
To analyze $\mathcal{S}$ we consider the situation along each edge $e \in E$ separately. On an edge $e \in E$ with endpoints $v_1^e, v_2^e \in V$ we can talk about an ordering and use the notation of intervals as follows.

- Let $\leq_e$ denote the natural ordering along the line segment e, induced by $v_1^e \leq_e v_2^e$.
- For $s_1, s_2 \in e$ let $[s_1, s_2]_e = \{s \in e : s_1 \leq_e s \leq_e s_2\}$ denote the set of points on e between s_1 and s_2, and $]s_1, s_2[_e = \{s \in e : s_1 <_e s <_e s_2\}$ denote the set of points strictly between s_1 and s_2.

Consequently, $\mathcal{S}$ can be ordered along each edge $e \in E$. Unfortunately, it may happen, that $|\mathcal{S} \cap e| = \infty$, i.e. the boundary of some B_p^r contains a linear piece which coincides with e in infinitely many points, see $p_2 + rB_2$ in Figure 2. To overcome this problem, note that for *all* $p \in \mathcal{P}$ either $B_p^r \cap e = \emptyset$, or

$$B_p^r \cap e = [s_{p,1}^e, s_{p,2}^e] \text{ for some } s_{p,1}^e, s_{p,2}^e \in e.$$

In the following, we will show that only the endpoints $s^e_{p,1}$ and $s^e_{p,2}$ of the interval need to be considered, also in the case of $|\mathcal{S} \cap e| = \infty$. We define

$$\mathcal{S}^e_{\text{cand}} := \{v^e_1, v^e_2\} \cup \begin{cases} \bigcup_{p \in \mathcal{P}} \{s^e_{p,1}, s^e_{p,2}\} & \text{if } \bigcup_{p \in \mathcal{P}} \{s^e_{p,1}, s^e_{p,2}\} \text{ contains at least one} \\ & \text{point of the interiour of } e \\ \{\frac{v^e_1 + v^e_2}{2}\} & \text{otherwise} \end{cases},$$

where instead of $\frac{v^e_1 + v^e_2}{2}$ any point in the interior of the edge e may be used. This case makes sure that at least one point of edge e is included in $\mathcal{S}^e_{\text{cand}}$ covering the (unlikely) case that the whole edge e lies in the interior of at least one B^r_p, does not intersect any of the other $B^r_{p'}$, and satisfies $c_e \leq \min\{c_{v^e_1}, c_{v^e_2}\}$, i.e. has smaller cost than both of its endpoints.
If $c_{v^e_1} \leq c_e$, or $c_{v^e_2} \leq c_e$ it may happen that one of the endpoints of e is included in the optimal solution, hence both v^e_1 and v^e_2 must be in $\mathcal{S}^e_{\text{cand}}$.
The result is a finite set of points

$$\mathcal{S}^e_{\text{cand}} =: \{s_0, s_1, \ldots, s_{N_e+1}\}$$

for which we assume $s_0 := v^e_1 \leq_e s_1 \ldots \leq_e s_{N_e} \leq s_{N_e+1} := v^e_2$. (The situation is depicted in Figure 2.)

Lemma 3 *Let e be an edge of E, and let $s \in]s_j, s_{j+1}[_e$ for some $j \in \{0, 1, \ldots, N_e\}$. Then*

$$\text{cover}(s) \subseteq \text{cover}(s_j) \cap \text{cover}(s_{j+1}).$$

Proof: Suppose, e.g., $\text{cover}(s) \not\subseteq \text{cover}(s_j)$, i.e. there exists $p \in \mathcal{P}$ such that $\gamma_p(s, p) \leq r$ and $\gamma_p(s_j, p) > r$. Note that for any gauge γ_p the distance $\gamma_p(p, s)$ from s to p is continuous if we fix p and move s along a line. Hence, the intermediate value theorem yields a point $\tilde{s} \in]s_j, s[_e$ with $\gamma_p(\tilde{s}, p) = r$, i.e. $\tilde{s} \in \mathcal{S}^e$. This contradicts the definition of s_j and s_{j+1} as consecutive points in $\mathcal{S}^e_{\text{cand}}$.

QED

Now we are able to prove that

$$\mathcal{S}_{\text{cand}} := \bigcup_{e \in E} \mathcal{S}^e_{\text{cand}}$$

is, indeed, a finite dominating set.

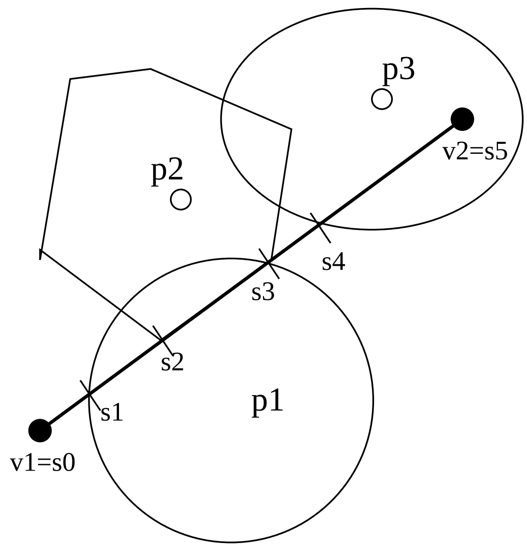

Figure 2: The set of candidates along one edge $e \in E$.

Theorem 2 *Either (CSLP) is infeasible, or there exists an optimal solution* $S^* \subseteq \mathcal{S}_{\text{cand}}$.

Proof: Let $S^* \subseteq T$ be optimal, but $S^* \not\subseteq \mathcal{S}_{\text{cand}}$. The goal is to replace each $s \in S^* \setminus \mathcal{S}_{\text{cand}}$ by a point in $\mathcal{S}_{\text{cand}}$ without loosing feasibility or optimality. To this end, take some $s \in S^* \setminus \mathcal{S}_{\text{cand}}$. Note that $s \notin V$, i.e. $g(s) = e \in E$, especially let

$$s \in]s_j, s_{j+1}[_e \text{ for some } j \in \{0, 1, \ldots, C_e\}.$$

Since $\mathcal{S}_{\text{cand}}$ contains at least one point of the interior of e, we can assume without loss of generality that $s_j \notin V$. Then Lemma 3 yields that $\text{cover}(s) \subseteq \text{cover}(s_j)$ such that

$$S' := S^* \setminus \{s\} \cup \{s_j\}$$

still covers $\mathcal{P}$. Moreover, since $g(s) = g(s_j) = e$ both points have the same costs c_e and we finally get $f(S') = f(S^*)$, which completes the proof.

QED

Note that the unweighted (CSLP) either is infeasible, or there exists an optimal solution

$$S^* \subseteq \bigcup_{p \in \mathcal{P}, e \in E} \{s^e_{p,1}, s^e_{p,2}\}$$

if $\bigcup_{p\in\mathcal{P},e\in E}\{s^e_{p,1}, s^e_{p,2}\} \neq \emptyset$. Furthermore, if γ_p is a strictly convex norm for all $p \in \mathcal{P}$ (e.g. if all γ_p are the Euclidean distance), we have that the original set $\mathcal{S}$ is already a finite set, i.e.

$$\mathcal{S} \subseteq \bigcup_{p\in\mathcal{P},e\in E} \{s^e_{p,1}, s^e_{p,2}\}.$$

Together, in the unweighted Euclidean case, $\mathcal{S}_{\text{cand}} := \mathcal{S}$ suffices as finite dominating set, if $\mathcal{S} \neq \emptyset$.

Using Theorem 2, (CSLP) can be formulated as a set covering problem. For all $s \in \mathcal{S}_{\text{cand}}$ let x_s be a variable with the following meaning:

$$x_s = \begin{cases} 1 & \text{if } s \in S^* \\ 0 & \text{otherwise} \end{cases}.$$

Define $A^{\text{cov}} = (a_{ps})_{p\in\mathcal{P},s\in\mathcal{S}_{\text{cand}}}$ with elements

$$a_{ps} = \begin{cases} 1 & \text{if } \gamma_p(p,s) \leq r \\ 0 & \text{otherwise} \end{cases}$$

as the matrix containing the covering information. Then (CSLP) is equivalent to a set covering problem, where the rows of the covering matrix A^{cov} correspond to the demand points, while the columns represent the possible candidates.

$$\begin{array}{rrcl} \min & cx & & \\ s.t. & A^{\text{cov}}x & \geq & \underline{1} \\ & x & \in & \{0,1\}^{|\mathcal{S}_{\text{cand}}|}, \end{array} \tag{1}$$

The cost vector $c \in \mathbb{R}^{|\mathcal{S}_{\text{cand}}|}$ in (1) is given by

$$c_s = c_{g(s)} \text{ for all } s \in \mathcal{S}_{\text{cand}},$$

and $\underline{1} \in \mathbb{R}^{|\mathcal{P}|}$ is the vector consisting of a 1 in each component.

As a consequence, to solve (CSLP) any set covering approach can be used. The set covering problem is, however, NP-hard. Thus, we will develop a more efficient approach in the next section by taking advantage of the special structure of the covering matrix A^{cov}.

4 Decomposition approach

In this section, we first consider (CSLP) along a single edge $e \subseteq T$ and show that in this case, the problem can be solved efficiently. To this end, note that a $(0,1)$- matrix A has the *consecutive ones property* (e.g., [GJ79]) if the columns of A can be orderd such that for all rows i of A the following holds.

$$a_{ik} = 1, a_{il} = 1, \text{ and } k \leq l \Longrightarrow a_{ij} = 1 \text{ for all } k \leq j \leq l.$$

Lemma 4 *Consider (CSLP) in the special case that $T = e$ consists of a single edge. Then the covering matrix A^{cov} of (CSLP) has the consecutive ones property.*

Proof: Each column of A^{cov} represents a candidate $s \in \mathcal{S}_{\text{cand}}$. Order the columns of A^{cov} according to the order of the candidates induced by $\leq_e$ (see the notation on page 10). We show that according to this order, A^{cov} has the consecutive ones property. To this end, take a line of A^{cov} (belonging to a demand point $p \in \mathcal{P}$) such that

$$a_{ps_k} = 1 \text{ and } a_{ps_l} = 1 \text{ for some } s_k \leq_e s_l.$$

This means

$$\begin{aligned} \gamma_p(p, s_k) &\leq r \text{ and} \\ \gamma_p(p, s_l) &\leq r, \end{aligned}$$

and equivalently,

$$s_k \in B_p^r \text{ and } s_l \in B_p^r.$$

Since B_p^r is a convex set, the whole segment $[s_k, s_l]$ lies in B_p^r, i.e. all s with $s_k \leq_e s \leq_e s_l$ satisfy $s \in B_p^r$, yielding $a_{ps} = 1$.

QED

Theorem 3 *(CSLP) can be solved by linear programming in the special case that $T = e$ is a single edge.*

Proof: From Lemma 4 we already know that the coefficient matrix A^{cov} of the integer programming formulation (1) has the consecutive ones property. This means, A^{cov} is a totally unimodular matrix (see e.g. Chapter 3.1. of [NW88]) and hence the IP-relaxation of (1) solves the integer program.

QED

More efficient procedures for solving set covering problems, where the covering matrix has the consecutive ones property are based on the theory of network matrices, or use a transformation to a shortest path problem, see [Sch01].

We now turn back to (CSLP) in a general network. The result of Theorem 3 can be extended, if the input data $\mathcal{P}$ and T satisfies the condition of the next result.

Lemma 5 *Given $\mathcal{P}, T$ for (CSLP), suppose that each demand point in $\mathcal{P}$ can be covered by exactly one edge $e \subseteq T$, i.e.*

$$|\{e \in E : \gamma_p(p, e) \leq r\}| = 1 \text{ for all } p \in \mathcal{P}.$$

Then A^{cov} has the consecutive ones property. Furthermore, (CSLP) can be decomposed into $|E|$ smaller problems, each of them containing only one edge of G as set of tracks, i.e. with input data $cover(e), T = e$.

Proof: To order the columns of A, use for each candidate s as first component its edge $g(s)$ (in arbitrary order), and as second component the ordering of the candidates induced by $\leq_e$. Choose an arbitrary ordering of the edges $e \in E$ and order the columns lexicographically according to their two components. Since each demand point can be covered from at most one edge, and for each edge e_i the covering matrix A_i^{cov} has the consecuctive ones property, we get that the covering matrix

$$A^{\text{cov}} = \begin{pmatrix} A_1^{\text{cov}} & & & \\ & A_2^{\text{cov}} & & \\ & & \cdots & \\ & & & A_{|E|}^{\text{cov}} \end{pmatrix}$$

for the complete problem, also has the consecutive ones property. Moreover, since A^{cov} has no coupling constraints, the whole problem can be separated into its blocks.

QED

An instance of the stop location problem with a coefficient matrix without the consecutive ones property is given in Figure 3. Numbering the candidates

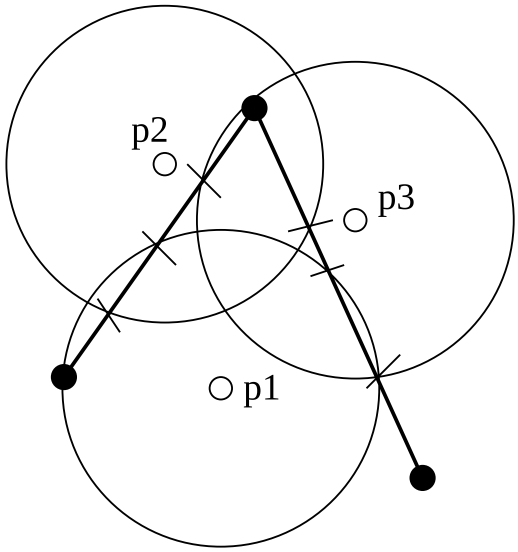

Figure 3: An instance of (CSLP) without consecutive ones property.

in $\mathcal{S}$ (without V which is sufficient in the unweighted case) from left to right, A^{cov} is given by

$$A^{\text{cov}} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix},$$

which cannot be reordered to satisfy the consecutive ones property.

We remark, however, that in practice the condition of Lemma 5 is often satisfied, since in practice the problem to be considered is (CSLP') referring to the establishment of additional stations. In the corresponding practical data the existing stations V are usually the endpoints of the edges. Now consider a node $v \in V$ with incident edges e_i and e_j. Since there exists a station in v, a large part of

$$\text{cover}(e_i) \cap \text{cover}(e_j)$$

is already covered. Hence, it is very unlikely that there exists an uncovered demand point within

$$(\text{cover}(e_i) \cap \text{cover}(e_j)) \setminus \text{cover}(v).$$

5 Extending the basic model

Among many useful extensions of the basic continuous stop location problem we consider the following variant. In a practical setting, one might not want to cover all demand points $\mathcal{P}$ but only a given percentage of the population. For this, we assume that for each demand point, we have given a weight w_p representing the number of customers who would like to use public transportation, if the next station is closer than r. For a subset $S \subseteq T$ define

$$w(S) = \sum_{p \in \text{cover}(S)} w_p$$

as the number of (potential) customers living in S. Then it makes sense to consider the following variants of (CSLP):

(CSLP-1) Given $\mathcal{P}$, T, and $Q_1 \in \mathbb{R}$, find a set $S^* \subseteq T$ such that $w(S^*) \geq Q_1$ and $f(S^*)$ is minimal.

(CSLP-2) Given $\mathcal{P}$, T, and $Q_2 \in \mathbb{R}$, find a set $S^* \subseteq T$ such that $f(S^*) \leq Q_2$ and $w(S^*)$ is maximal.

Theorem 4 $\mathcal{S}_{\text{cand}}$ *is a finite dominating set for (CSLP-1) and for (CSLP-2).*

Proof: We use exactly the proof of Theorem 2, where $\text{cover}(S^*) \subseteq \text{cover}(S')$ yields that

$$\begin{array}{rcl} w(S^*) & \leq & w(S') \text{ and} \\ f(S^*) & \geq & f(S'). \end{array}$$

This shows the theorem.

QED

Using Theorem 4, both problems can be formulated as integer programs. To keep track of the population covered by the new stops, we have to know, which demand points are covered and which not. In addition to the variables x_s defined for (CSLP), we therefore define another set of binary variables

$$y_p = \begin{cases} 1 & \text{if demand point } p \text{ is covered} \\ 0 & \text{otherwise} \end{cases},$$

and let $w = (w_1, w_2, \ldots, w_{|\mathcal{P}|})$. The IP model for (CSLP-1) is

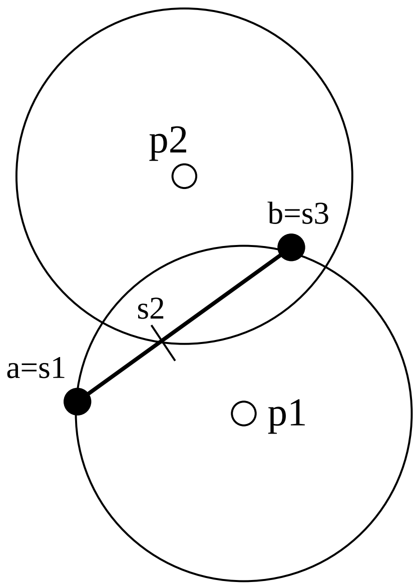

Figure 4: The coefficient matrix of (CSLP-1) is not totally unimodular.

$$\begin{array}{rrcl} \min & cx & & \\ \text{s.t.} & A^{\text{cov}}x - y & \geq & \underline{0} \\ & wy & \geq & Q_1 \\ & x & \in & \{0,1\}^{|\mathcal{S}_{\text{cand}}|} \\ & y & \in & \{0,1\}^{|\mathcal{P}|}, \end{array} \tag{2}$$

and (CSLP-2) is given by

$$\begin{array}{rrcl} \max & wy & & \\ \text{s.t.} & A^{\text{cov}}x - y & \geq & \underline{0} \\ & cx & \leq & Q_2 \\ & x & \in & \{0,1\}^{|\mathcal{S}_{\text{cand}}|} \\ & y & \in & \{0,1\}^{|\mathcal{P}|}. \end{array} \tag{3}$$

Note that even for a single edge $T = e$ and weights $w_p \in \{0,1\}$ the coefficient matrix of problem (CSLP-1) is **not** totally unimodular, an example with two demand points is given in Figure 4. For this example we get the following coefficient matrix of (CSLP-1)

$$\begin{pmatrix} 1 & 1 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix},$$

which is not totally unimodular.

On the other hand, using the same example for (CSLP-2) with $c_s = 1$, the coefficient matrix is given by

$$\begin{pmatrix} 1 & 1 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 1 & 1 & 1 & 0 & 0 \end{pmatrix},$$

which is a totally unimodular matrix (see Lemma 6 below).
To get rid of the coupling constraint $wy \geq Q_1$ in CSLP-1 one can consider the Lagrange relaxation with respect to this constraint.

$$\begin{array}{lrcl} \min & cx + \lambda(Q_1 - wy) & & \\ \text{s.t.} & A^{\text{cov}}x - y & \geq & \underline{0} \\ & x & \in & \{0,1\}^C \\ & y & \in & \{0,1\}^{|\mathcal{P}|}, \end{array} \tag{4}$$

where $\lambda > 0$.

Lemma 6 *In the case of one edge, weights $w_p = 1$, and costs $c_s = 1$*

1. *the Lagrange-relaxation of (CSLP-1) (with fixed multiplier λ) can be solved by linear programming.*
2. *(CSLP-2) can be solved by linear programming.*

Proof: Since A^{cov} is totally unimodular it is well known that the coefficient matrix of (4) given by

$$\begin{pmatrix} A^{\text{cov}} & -I \end{pmatrix},$$

also is totally unimodular, which shows both parts of the lemma.

QED

Efficient approaches for solving (CSLP-2) are given in [KPS+02] by using dynamic programming or the equivalence to the p-center problem on the line. In [Sch01] a transformation to a cardinality constrained shortest path problem in an acyclic digraph is suggested.

6 Varying the radius

The finite dominating set $\mathcal{S}_{\text{cand}}$ is dependent on the given radius r. If more than one value for r should be considered it would be advantageous to have a finite dominating set which is **not** dependent on the specific value of r. Such a set will be derived in this section, for the special case that $c_v \leq c_e$ for all $v \in V$, and incident edges e.
Using the denotation of $\mathcal{S}^e_{\text{cand}}$ (see page 11) we define

$$\begin{aligned}\mathcal{S}^*_{\text{cand}} \; := \; & \{x \in T : \text{ there exist } p_1, p_2 \in \mathcal{P} \text{ such that } \gamma_{p1}(p_1, x) = \gamma_{p2}(p_2, x)\} \\ & \cup \bigcup_{p \in \mathcal{P}, e \in E} \operatorname{argmin}_{x \in e} \gamma_p(p, x) \cup \bigcup_{e \in E} \{\frac{v_1^e + v_2^e}{2}\} \cup V\end{aligned}$$

The first set of candidates can be derived by intersecting the bisector between each pair of demand points with T, the second set describes the projection points from all demand points onto all edges, and the third set makes sure that V and one point of the interior of each edge is included in $\mathcal{S}^*_{\text{cand}}$.

Theorem 5 *$\mathcal{S}^*_{\text{cand}}$ is a dominating set for (CSLP), (CSLP-1), and (CSLP-2) for any given radius $r > 0$, if $c_v \leq c_e$ for all $v \in V$, $e \in E$ with v, e are incident.*

Proof: Take $r > 0$. Using Theorems 2 and 4 we can assume that there exists an optimal solution $S^* \subseteq \mathcal{S}_{\text{cand}}$, where $\mathcal{S}_{\text{cand}}$ is the specific candidate set with respect to r. Take $s \in S^*$ with $s \notin \mathcal{S}^*_{\text{cand}}$. This means, $s \notin V$ and $s \neq \frac{v_1^e + v_2^e}{2}$ for any of the edges $e \in E$. Consequently, s lies in the interior of some edge e, and there exists exactly one $p \in \mathcal{P}$ such that

$$\gamma_p(p, s) = r,$$

(otherwise, $s \in \mathcal{S}^*_{\text{cand}}$). Move s along e within B^r_p (i.e. getting closer to p) to a new point, until one of the following conditions is satisfied for the first time.

- s reaches some point in $\mathcal{S}^*_{\text{cand}}$, let this point be denoted by s^*.
- s leaves some $B^r_{p'}$: Denote this point by s'. It is important that $p \neq p'$, since otherwise, e.g., the projection point from p onto e would have been met before. Then we have

$$\begin{aligned} \gamma_{p'}(s') = r \quad & \text{and} \quad \gamma_p(s') \leq r \\ \gamma_{p'}(s) \leq r \quad & \text{and} \quad \gamma_p(s) = r \end{aligned}$$

meaning that $h(x) = \gamma_{p'}(x) - \gamma_p(x)$ satisfies $h(s') \geq 0$ and $h(s) \leq 0$ such that the intermediate value theorem yields a point s^* with $h(s^*) = 0$, i.e. $s^* \in \mathcal{S}^*_{\text{cand}}$.

Due to the construction, in both cases we have found a point $s^* \in \mathcal{S}^*_{\text{cand}}$ such that

$$\begin{aligned} \text{cover}(\{s\}) &\subseteq \text{cover}(\{s^*\}), \text{ and} \\ c_{g(s)} &\geq c_{g(s^*)}, \end{aligned}$$

the latter holding since either $g(s) = g(s^*) = e$, or $g(s) = e$ and $g(s^*) = v$ with v and e incident to each other, and hence satisfying $c_{g(s)} = c_e \geq c_v = c_{g(s^*)}$. Finally we get that $S' := S^* \setminus \{s\} \cup \{s^*\}$ satisfies

$$\begin{aligned} w(S^*) &\leq w(S') \\ f(S^*) &\geq f(S'), \end{aligned}$$

which completes the proof.

QED

We remark that, dependent on the structure of the bisectors, $\mathcal{S}^*_{\text{cand}}$ needs not be a finite set (for extensive discussion about determining bisectors for gauge-distance functions, see [Wei99]), but can be made finite along the lines of Section 3. We further point out, that the size of $\mathcal{S}^*_{\text{cand}}$ is of $\mathrm{O}(|\mathcal{P}|^2 + |\mathcal{P}||E|)$ which is much larger than the size of $\mathcal{S}_{\text{cand}}$, which is of order $\mathrm{O}(|\mathcal{P}|)$.

While the assumption of Theorem 5 is in practice usualy satified for all nodes with degree at most two, it may happen that it does not hold for nodes with more than two incident edges. Unfortunately, the next result yields that no finite dominating set is possible without this assumption.

Lemma 7 *In the general case, no finite dominating set which can be used for each value of r exists for (CSLP), (CSLP-1), and for (CSLP-2).*

Proof: Consider the star-shaped graph with four nodes and three edges, all edges incident with node v_0 (see Figure 5). Let one demand point p be given, such that $\text{argmin}_{x \in e_1} \gamma_p(p, x) = v_0$, and consider, e.g., the Euclidean distance l_2.

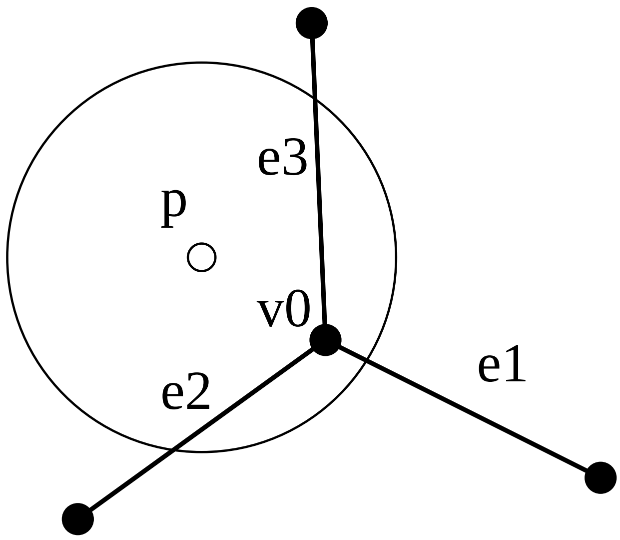

Figure 5: No finite dominating set is possible independent of r.

Let $w_{e_1} = 2, w_{e_2} = w_{e_3} = 100$, and $w_{v_0} = 102$ be (realistic) traffic loads. Now take any finite candidate set, and let s be the candidate of e_1 which is closest to v_0. Note that $l_2(p, v_0) < l_2(p, s)$. Then let r such that

$$l_2(p, v_0) < r < l_2(p, s)$$

Then the optimal solution is a point on e_1 between v_0 and s, which is not contained in the finite candidates set.

QED

7 Other extensions and further research

One other straightforward extension is to allow demand regions instead of demand points. This can be done as long as for each region $p \subseteq \mathbb{R}^2$ we get that

$$e \cap \{x \in \mathbb{R}^2 : x \text{ covers } p\}$$

is an interval, which, e.g., is the case if

$$\{x \in \mathbb{R}^2 : x \text{ covers } p\} \subseteq \mathbb{R}^2$$

is a convex set. Also, various definitions of *covering* may be used in this context.

Acknowledgement

We want to thank Frank Plastria for suggesting the topic of Section 6.

References

[CFT99] A. Caprara, M. Fischetti, and P. Toth. A heuristic method for the set covering problem. *Operations Research*, 47(5):730–743, 1999.

[DAL82] M. J. Demetsky, M. Asce, and B. B.-M. Lin. Bus stop location and design. *Transportation Engineering Journal*, 108:313–327, 1982.

[EH72] J. Elzinga and D.W. Hearn. Geometrical solutions for some minimax location problems. *Transportation Science*, 4:379–394, 1972.

[GJ79] M.R. Garey and D.S. Johnson. *Computers and Intractability — A Guide to the Theory of NP-Completeness*. Freeman, San Francisco, 1979.

[HLS+01] H.W. Hamacher, A. Liebers, A. Schöbel, D. Wagner, and F. Wagner. Locating new stops in a railway network. *Electronic Notes in Theoretical Computer Science*, 50(1), 2001.

[Joh82] D. S. Johnson. The NP-completeness column: An ongoing guide. *Journal of Algorithms*, 3:182–195, 1982.

[KPS+02] E. Kranakis, P. Penna, K. Schlude, D.S. Taylor, and P. Widmayer. Improving customer proximity to railway stations. Technical report, ETH Zürich, 2002.

[LMO02] G. Laporte, J.A. Mesa, and F.A. Ortega. Locating stations on rapid transit lines. *Computers and Operations Research*, 29:741–759, 2002.

[MDSF98] A. Murray, R. Davis, R.J. Stimson, and L. Ferreira. Public transportation access. *Transportation Research D*, 3(5):319–328, 1998.

[Meg83] N. Megiddo. Linear time algorithms for linear programming in $\mathbb{R}^3$ and related problems. *SIAM Journal on Computing*, 12:759–776, 1983.

[Min67] H. Minkowski. *Gesammelte Abhandlungen, Band 2*. Chelsea Publishing Company, New York, 1967.

[Mur01a] A. Murray. Coverage models for improving public transit system accessibility and expanding access. Technical report, Department of Geography, Ohio State University, 2001.

[Mur01b] A. Murray. Strategic analysis of public transport coverage. *Socio-Economic Planning Sciences*, 35:175–188, 2001.

[NW88] G. L. Nemhauser and L. A. Wolsey. *Integer and Combinatorial Optimization*. Wiley, 1988.

[Sch01] A. Schöbel. Set covering problems with consecutive ones property. Technical report, Universität Kaiserslautern, 2001.

[TSRB71] C. Toregas, R. Swain, C. ReVelle, and L. Bergman. The location of emergency facilities. *Operations Research*, 19:1363–1373, 1971.

[Wei99] A. Weissler. *General Bisectors and their Application in Planar Location Theory*. Shaker, 1999.