Designing Meet-in-the-Middle Service Systems

Clarence Worrell* Jourdain Lamperski*† Lisa Maillart*

Department of Industrial Engineering, University of Pittsburgh

Abstract

A service system is ideally both *accessible* and *operationally efficient*; no *client* should have to travel too far from their client location to obtain service, and no *agent* should have to travel too far from a service *facility* to administer service. Accessibility and operational efficiency are conflicting objectives that system planners balance through *meet-in-the-middle sites*, i.e., intermediate locations between facilities and client locations at which clients and agents can meet. Consider bus stops in transportation systems, garbage bins in waste management systems, and outreach sites in healthcare systems. In this work, we consider locating service facilities and meet-in-the-middle sites amongst a given set of client locations with the aim of optimally balancing accessibility and operational efficiency. While a number of studies consider (re)locating meet-in-the-middle sites in service systems with existing facilities, we consider more fully (re)designing service systems and investigating the associated benefits. Locating facilities and meet-in-the-middle sites is naturally more challenging because *optimal* meet-in-the-middle site locations depend on facility locations. We measure accessibility and operational efficiency as the maximum cost that a client and agent, respectively, incur for travel. The goal is then to minimize the maximum cost that an agent or client incurs for travel, subject to the condition that at most k facilities can be built. We establish first principles, derive mixed-integer linear programming formulations, develop a polynomial-time approximation algorithm, and present a numerical study that investigates the empirical performance of the proposed methods. A vaccine campaign design case study demonstrates that locating facilities and meet-in-the-middle sites (in comparison with locating meet-in-the-middle sites given facilities) provides an average 8% reduction and an up to 37% reduction in travel distance across 37 districts in Ghana.

1 Introduction

When designing a service system, planners aim to ensure that it is both *accessible* and *operationally efficient*. No *client* should have to travel too far from their client location to *retrieve* service (i.e., the system is accessible), and no service *agent* should have to travel too far from a service *facility* to administer service (i.e., the system operates efficiently).

*Emails: {clw117, lamperski, maillart}@pitt.edu

†Corresponding author. Address: 1025 Benedum Hall, Pittsburgh, PA 15261, USA.

Naturally, there is a trade-off between accessibility and operational efficiency. If an agent is sent from a facility to provide service to a client, then the system is not as efficient, but it is more accessible. On the other hand, if a client is required to travel to a facility, then the system is more efficient, but it is not as accessible. Urban planners balance accessibility and operational efficiency with *meet-in-the-middle sites*, i.e., intermediate locations between facilities and client locations at which clients and agents can meet. (A meet-in-the-middle site is a concept that we term.)

Below we discuss examples of meet-in-the-middle sites in transportation, waste management, and healthcare systems.

- **Bus stops in transportation systems.** Consider a bus transportation system. It would be highly inaccessible to require that all commuters travel to a bus depot (i.e., facility) in order to catch a bus, and it would be highly inefficient to send a bus to the location of each commuter in its service region. Instead, commuters and buses meet-in-the-middle at bus stops.

- **Garbage bins in waste management systems.** Consider a waste management system in an urban area. Residents of urban areas typically dispose of their waste in garbage bins (aka. dumpsters). A garbage truck then delivers the waste to a waste management facility.

  In regions with ample resources, garbage bins can be located within close proximity to each residential complex. In regions with more limited resources, bins must be further spread out, and residents consequently may need to travel a non-trivial distance to dispose of their waste. Parrot et al. [29] and Zia and Devadas [45] report that bin locations in Yaoundé and Kanpur, respectively, favor operational efficiency at the cost of accessibility.

- **Outreach sites in healthcare systems.** Consider a healthcare system that has *outreach sites*. An outreach site is a designated location that people who would not otherwise have access to healthcare can go to obtain care. Health workers travel from health facilities to administer care at outreach sites.

  In 2002, the World Health Organization and its partners introduced the "reaching every district" (RED) strategy [39] with the goal of improving vaccination rates in areas with low coverage (i.e., making service more accessible). Outreach trips made from medical facilities to remote locations are an important component of the RED strategy [42].

A number of system-specific studies investigate (re)locating meet-in-the-middle sites in service systems that have existing facilities with the aim of improving accessibility, operational efficiency, or both. Transportation studies investigate (re)locating bus stops [4, 11, 23, 24, 28], and waste management studies investigate (re)locating waste bins [10, 17, 29, 30, 45]. In regions with no or limited facilities, however, there is still an opportunity/need to more fully design service systems, i.e., jointly locate facilities and meet-in-the-middle sites. For instance, there is a growing body of work that explores locating medical clinics and outreach sites in developing regions [13, 14, 21, 43, 44]. Existing systems can also potentially be redesigned (analogous to relocating meet-in-the-middle sites in systems with existing facilities) if the benefits of redesign outweigh the costs.

In this work we consider locating facilities and meet-in-the-middle sites amongst a given set of client locations with the aim of optimally balancing accessibility and operational efficiency. We measure operational efficiency as the maximum cost that an agent incurs for traveling to administer service, and we measure accessibility as the maximum cost that a client incurs for traveling to retrieve service. The goal is to locate facilities and meet-in-the-middle sites at given client locations in order to minimize the maximum cost that an agent or client incurs for travel, subject to the condition that at most k facilities can be built. No budget is placed on the number of meet-in-the-middle sites given that they are presumably cheap to erect. The problem can be thought of as an analog of the k-center problem in which agents and clients can meet-in-the-middle; the goal in the k-center problem is to locate k facilities that minimize the maximum distance between a client and a nearest facility. Thus, we refer to the problem as the *meet-in-the-middle k-center problem*.

Naturally, the meet-in-the-middle k-center problem does not fully account for certain, potential, practical considerations (e.g., locating facilities at locations other than client locations, population size/density, and routing). That said, the meet-in-the-middle k-center problem is elegant as well as interpretable and consequently can provide useful insights for system planners. Furthermore, our study investigates (re)locating facilities and meet-in-the-middle sites (instead of meet-in-the-middle sites given facilities). Locating facilities and meet-in-the-middle sites is a naturally more challenging because *optimal* meet-in-the-middle site locations depend on facility locations.

Simultaneous versus sequential location. To better inform the (re)design of service systems, we are particularly interested in understanding the potential benefit of *simultaneously* locating facilities and meet-in-the-middle sites (our work) versus *sequentially* locating meet-in-the-middle sites given a set of facility locations (existing studies).

As a motivating example, suppose that a system planner has the budget to build one facility at any of the client locations depicted in Plot (a) of Figure 1. Further suppose that the planner aims to minimize the maximum taxicab distance that an agent or client travels (i.e., travel costs are distances). If the planner does not account for meet-in-the-middle sites when locating facilities, then the planner aims to minimize the maximum distance from the facility to a client (and faces an instance of the k-center problem with $k = 1$). It is readily verified that building the facility depicted in Plot (b) minimizes the maximum distance, so suppose the planner builds this facility. Agent versus client travel are arbitrarily specified in Plot (b). If there are no meet-in-the-middle sites, then the maximum distance that an agent or client travels equals 5 under this facility. However, if there are meet-in-the-middle sites, and they are subsequently (optimally) located given the facility from Plot (b), as depicted in Plot (c), then the maximum travel distance equals 4. Plot (d) demonstrates that if the planner simultaneously locates the facility and meet-in-the-middle sites (and hence builds the facility depicted in this plot), then the maximum travel distance equals 3.

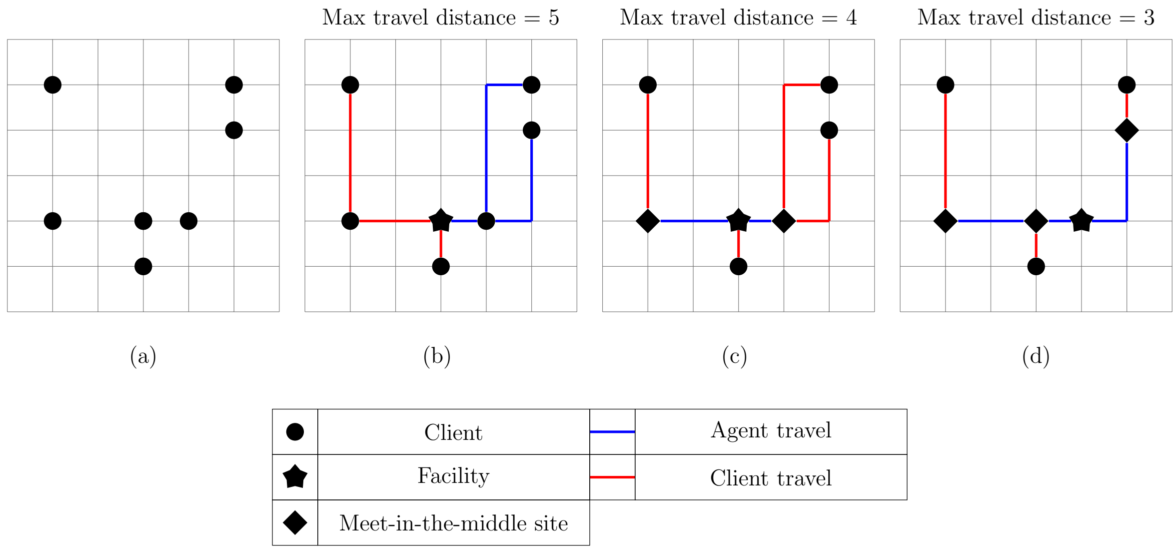


Figure 1: Plot (a) illustrates the set of client locations. Plot (b) depicts a facility that minimizes the maximum distance to a client; agent versus client travel are arbitrarily specified. Plot (c) illustrates sequentially locating meet-in-the-middle sites given the facility in Plot (b). Finally, Plot (d) depicts simultaneously locating facilities and meet-in-the-middle sites.

1.1 Summary of contributions

The meet-in-the-middle k-center problem. We present a formal description of the the meet-in-the-middle k-center problem in Subsection 2.1. We study the problem under a variety of different cost regimes (i.e., assumptions about the client and agent travel costs) that we discuss in Subsection 2.2. Finally, we position the problem within the existing literature in Subsection 2.3.

First principles. In Section 3, we establish the following first principles that play a fundamental role throughout our work.

- **Minimum balancing costs.** In Subsection 3.1, we identify the concept of *minimum balancing costs*. A minimum balancing cost is the minimum cost to the system (i.e., the maximum receiving or retrieving cost incurred) at which a given client can obtain service from a given facility (possibly by meeting at a meet-in-the-middle site). In our methodological developments, we will think of a minimum balancing cost for a given client and facility as the distance between the client and facility; although we show that minimum balancing costs do not define a metric in any of the cost regimes specified in Subsection 2.2.

- **Locating meet-in-the-middle sites.** In Subsection 3.2, we develop a polynomial-time algorithm for the subproblem of locating meet-in-the-middle sites given facilities; see Algorithm 1. As a byproduct of our analysis of the algorithm, we see that the meet-in-the-middle k-center problem reduces to the k-center problem with minimum balancing costs as distances. In other words, the meet-in-the-middle k-center problem is in some sense no harder than the k-center problem! This statement, however, is somewhat misleading: even if the receiving and retrieving costs are equal and induce a metric, the meet-in-the-middle k-center problem does not reduce to a metric k-center problem, so there is some nuance in this reduction.

Meet-in-the-middle location methods. In Section 4, we develop the following meet-in-the-middle location methods.

- **MILP formulations.** We propose two mixed-integer linear programming (MILP) formulations in Subsection 4.1. The first MILP formulation readily follows from the description of the meet-in-the-middle k-center problem, while the second formulation follows from the reduction established in Subsection 3.2. In computational experiments that we present in Section 5, we observe that a commercial solver (Gurobi) solves the latter formulation approximately one order of magnitude faster.
- **Greedy approximation algorithm.** In Section 4.2, we develop a greedy approximation algorithm. Motivated by the reduction presented in Section 3.2, we consider an extension of a greedy approximation algorithm for the k-center problem. We establish theoretical approximation guarantees for the algorithm in different cost regimes and show that all of the approximation guarantees are tight. The approximation analysis is non-trivial because the meet-in-the-middle k-center problem does not reduce to the metric k-center problem in any of the cost regimes.

Computational study. In Section 5, we present a computational study. We investigate the empirical performance of the meet-in-the-middle location methods on synthetic instances. We also present a vaccine campaign design case study across districts in Ghana and quantify the benefits of simultaneous location.

2 Problem statement and related work

We present a description of the meet-in-the-middle k-center problem in Subsection 2.1, discuss the travel cost regimes that we consider in Subsection 2.2, and review related work in Subsection 2.3.

2.1 The meet-in-the-middle k-center problem

Suppose that there are n clients indexed by the set $[n] := \{1, \ldots, n\}$. We must decide (i) at which of the n client locations to build k facilities and (ii) which client locations to designate as meet-in-the-middle sites. If a client's location has a facility or is designated a meet-in-the-middle site, then the client receives service at the location. Otherwise, the client retrieves service from a facility or a meet-in-the-middle site. That is, we assume that agents do not travel from facility locations to client locations unless the client locations are designated as meet-in-the-middle sites. The assumption is without loss of generality because we do not assume there is a cost to erect meet-in-the-middle sites or a budget on the number of meet-in-the-middle sites.

There is a cost of $c_{ij} \in \mathbb{R}_{\geq 0}$ for an agent to travel from a facility at the location of client $j \in [n]$ to administer service at the location of client $i \in [n]$. There is a cost of w_{ij} for client

i to travel from their location to retrieve service from a facility or meet-in-the-middle site at the location of client j. We assume that $c_{ii} = 0$ and $w_{ii} := 0$ for each $i \in [n]$. We collect the agent and client travel costs into the $n \times n$ matrices c and w, respectively. That is, c_{ij} and w_{ij} are the ij-th entries of the matrices c and w, respectively.

Suppose that we locate facilities and meet-in-the-middle sites at the locations of clients $F \subseteq [n]$ and $M \subseteq [n] \setminus F$, respectively. To be succinct, we refer to F and M as a set of facilities and meet-in-the-middle sites, respectively. Note that $R = [n] \setminus (F \cup M)$ is the set of clients who retrieve service. The operational efficiency under facilities F and meet-in-the-middle sites M, as measured by the maximum cost that an agent incurs for travel, is given by

$$r_1(F, M) := \max_{i \in M} \min_{f \in F} c_{if}.$$

Note that for a given meet-in-the-middle site, we assume an agent travels to the site from a facility that yields the smallest cost. The accessibility under facilities F and meet-in-the-middle sites M, as measured by the maximum cost that a client incurs for travel, equals

$$r_2(F, M, R) := \max_{i \in R} \min_{f \in F \cup M} w_{if}.$$

Note that we assume that a retrieving client retrieves service from a facility or meet-in-the-middle site that brings them the smallest travel cost. Finally, define

$$f(F, M, R) := \max \{r_1(F, M), r_2(F, M, R)\} \tag{1}$$

to be maximum cost that a client or agent incurs for travel, which captures the balance of operational efficiency and accessibility.

The goal in the meet-in-the-middle k-center problem is to optimally balance operational efficiency and accessibility, subject to the condition that we can locate at most k facilities. Formally, the meet-in-the-middle k center problem is given by the optimization problem:

$$\begin{aligned} \min_{F,M,R} \quad & f(F, M, R) \\ \text{s.t.} \quad & |F| = k \\ & F \subseteq [n] \\ & M \subseteq [n] \setminus F \\ & R = [n] \setminus (F \cup M). \end{aligned} \tag{2}$$

Relation to the k-center problem and NP-hardness. It is not difficult to see that the meet-in-the-middle k-center problem captures the metric k-center problem (and more general k-center problem with distances that do not induce a metric). Indeed, we can encode an instance of the metric k-center problem with metric d (or distances d that do not induce a metric) as an instance of the meet-in-the-middle k-center problem with $c = d$ and $w_{ij} = \max_{p,q \in [n]} d_{pq}$ for all $i \neq j \in [n]$. (Here the definition of the client costs w ensures that it is always optimal to locate a meet-in-the-middle site at a client location that does not have a facility.) So, because the metric k-center problem is NP-hard [18], the meet-in-the-middle k-center problem (2) is NP-hard as well.

2.2 Travel cost regimes

We study the meet-in-the-middle k-center problem under the travel cost regimes listed below. All meet-in-the-middle location methods that we develop in Section 4 are applicable in general (i.e., under no assumptions on the agent and client travel costs); the cost regimes are mainly considered to study the theoretical and empirical performance of the meet-in-the-middle location methods in different settings.

Metric costs. In the metric costs regime, agent and client travel costs individually define metrics on $[n]$. That is, the costs are symmetric

$$c_{ij} = c_{ji} \quad \text{and} \quad w_{ij} = w_{ji}, \qquad i, j \in [n], \tag{3}$$

and the costs satisfy the triangle inequality

$$c_{ij} \leq c_{ik} + c_{kj} \quad \text{and} \quad w_{ij} \leq w_{ik} + w_{kj}, \qquad i, j, k \in [n]. \tag{4}$$

Recall that, by definition, $c_{ij}, w_{ij} \in \mathbb{R}_{\geq 0}$ and $c_{ii} = w_{ii} = 0$ for $i, j \in [n]$, so we do not explicitly state this metric property above. The metric costs regime is natural to consider, especially given that the k-center problem has been extensively studied under the assumption that distances define a metric on $[n]$; see Subsection 2.3.

Related metric costs. In the related metric costs regime, conditions (3) and (4) of the metric costs regime hold, and

$$c_{ij} \leq \min\{c_{kj} + w_{ik}, c_{ki} + w_{jk}\} \text{ for all } i, j, k \in [n] \text{ such that } i \neq j \neq k. \tag{5}$$

Condition (5) states that the cost for an agent to travel from the location of client $j \in [n]$ to the location of client $i \in [n]$ is less than (i) the total system cost for client i to meet-in-the-middle at the location of client k with an agent from a facility at the location of client j, and (ii) the total system cost for client j to meet-in-the-middle at the location of client k with an agent from a facility at the location of client i.

Note that, under conditions (3) and (4), if $c_{ij} \leq w_{ij}$ for all $i, j \in [n]$, then condition (5) holds. Accordingly, if the agent as well as client travel costs individually define metrics, and the client costs are higher the agent costs, then the conditions of the related metric costs regime are met. Client travel costs are typically higher than agent costs in developing regions given that residents must walk to travel.

Equal metric costs. In the equal metric costs regime, conditions (3) and (4) of the metric costs regime hold, and the agent and client travel costs are equal:

$$c_{ij} = w_{ij}, \qquad i, j \in [n], \tag{6}$$

The equal metric costs regime captures the setting in which it is of interest to minimize the maximum distance that a facility agent or client travels, as opposed to the corresponding costs that they incur. The motivating example from Section 1 depicted in Figure 1 thus falls under the equal metric costs regime.

Finally, note that the metric costs regime is the most general regime, and the equal metric costs regime is the most restrictive regime. That is, if the conditions of the equal metric cost regime are met, then (it is straightforward to verify) the conditions of the related metric costs regime are met, and if the conditions of the related metric costs regime are met, then the conditions of the metric costs regime are met.

2.3 Related work

First, we position our study within the broad facility location literature. Next, we review service-specific studies that consider locating meet-in-the-middle sites. Finally, we summarize related work on the k-center problem (and its variations).

Facility location. We direct the reader to the recent survey [5] as well as the surveys [26, 34, 35] for overviews of the facility location literature. Classically, facility location problems are studied for their application in the manufacturing industry. Now, they are also (if not more) studied for their application in the service industry; see [5, 7] for recent service facility location surveys. Some studies consider general service systems (as in our work), while others focus on particular types of services systems (e.g., transportation, waste management, and healthcare systems).

The meet-in-the-middle k-center problem has its own peculiarities. Regarding the model itself, two different of types of locations must be established, namely facility locations and meet-in-the-middle sites. The task of establishing different types of locations also arises in hierarchical facility location [25, 36]. In hierarchical facility location, each client must be assigned to a facility at the bottom of the hierarchy, but in our problem, client demand can be satisfied by either a facility or meet-in-the-middle site. Regarding the objective of the meet-in-the-middle k-center problem, maximum-based objectives are on one hand commonly considered in the service facility location literature [5, 7]; they ensure that no client or agent incurs an excessively large cost at the expense of other clients or agents. On the other hand, our objectives uniquely captures the balance between two conflicting quantities, operational efficiency and accessibility. In contrast, location studies typically tackle conflicting objectives with multi-objective optimization [8].

Locating meet-in-the-middle sites. Below we review system-specific studies that investigate locating meet-in-the-middle sites. From a methodological standpoint, all of these studies propose MILP formulations or heuristics.

- **Bus stops in transportation systems.** There is a recognized need to account for accessibility and operational efficiency when locating bus stops [37, 41]. There are two different location approaches. In the first approach, stops are first located, and then buses are routed through the stops [4, 28, 23]. Stops are located to ensure accessibility (e.g., by solving a set cover problem that ensures each commuter is within a specified walking distance to a stop [4]), and then routes are chosen to maximize operational efficiency. In the second approach, bus routes are first selected, and then stops are located on the routes [11, 24]. Gleason [11] and Murray [24] both consider minimizing the number of stops subject to the condition that all commuters are within a given walking distance of a stop. (This is also a set covering problem). Aside from accounting for routing, both of these lines of work are different from ours in two ways. First, in locating stops, they account for accessibility but not operational efficiency. (Operational efficiency is only accounted for when routing). Second, they assume that that bus depots are already located.

- **Garbage bins in waste management systems.** Purkayastha et al. [30] survey work on locating garbage bins in urban areas. In developing regions, bins are not as accessible [29, 45], so corresponding location studies focus on maximizing accessibility. For instance, Kao and Lin [17] consider minimizing the sum of walking distances to bins. There are also studies that consider locating bins in developed regions [10]. Ghiani et al. [10] account for both accessibility and operational efficiency. To the best of our knowledge, there is not a study that investigates simultaneously locating waste management facilities and bins

- **Outreach sites in healthcare systems.** Healthcare facility location problems have received a considerable amount of research attention; see Daskin and Dean [6] for a review of single-level models, Rahman et al. [31] for a review specific to developing countries, and [1, 32] for more recent reviews. Studies [14, 13, 21, 43, 44] investigate simultaneously locating clinics and outreach sites. All of these studies account for accessibility and operational efficiency as well. In contrast to our work, these studies consider set cover models. They also only develop MILP formulations.

The k-center problem. The k-center problem has been extensively studied under the assumption that distances between locations define a metric. Under this assumption, the problem is referred to as the *metric* k-center problem. The metric k-center problem is NP-hard [18]. Hochbaum [15] further shows that for $\alpha < 2$, it is NP-hard to solve the metric k-center problem to within an approximation factor of α. There are a number of 2-approximation algorithms for the metric k-center problem, including the simple greedy algorithm of Gonzalez [12]. Often this algorithm is referred to as the Gon algorithm. In this work we develop an extension of the Gon algorithm for the meet-in-the-middle k-center problem (2); see Algorithm 2. Other 2-approximation algorithms include the Sh algorithm [38] and the HS algorithm [16]. The CDS algorithm proposed in [9] is yet another approximation algorithm, but it has an approximation factor of 3. Nonetheless, it tends to achieve better empirical performance than the Gon, Sh, and HS algorithms.

Fewer works study the setting in which distances do not define a metric. A few studies consider the *asymmetric* k-center problem, namely the k-center problem under the assumption that the metric does not satisfy the symmetry property, but still satisfies the triangle inequality [3, 27, 40]. In particular, [27, 40] propose $O(\log^*(m))$-approximation algorithms, and [3] proposes approximation algorithms with improved approximation factors of $O(\log^*(k))$, where $\log^*$ is the iterated logarithm. Finally, [20] investigates the case in which distances are induced by smooth and strongly convex objectives. In Section 3.2 we show that the meet-in-the-middle k-center problem reduces to a k-center problem with distances that are not necessarily symmetric nor satisfy the triangle inequality.

Many variations of the k-center problem have been considered in the literature; we discuss the most relevant. A few studies consider the possibility of routing through locations to serve clients. Study [19] develops an approximation algorithm for the edge-dilation k-center problem. The goal in the edge-dilation k-center problem is to choose k facility locations to minimize the maximum ratio of geometric to graph distance between a client and their nearest facility. Another related problem is the k-next center problem [2, 22]. The goal in the k-next center problem is to choose k facility locations that minimize the maximum distance between clients and their nearest facility *plus* the distance from this nearest facility to its nearest facility (the backup facility).

3 First principles

We introduce as well as study the concept of minimum balancing costs in Subsection 3.1, and we consider the subproblem of locating meet-in-the-middle sites in Subsection 3.2.

3.1 Minimum balancing costs

First, we define the quantities

$$r_{ij} := \begin{cases} \min_{k\in[n]\setminus\{i,j\}} \max\{c_{kj}, w_{ik}\} & i \neq j \\ 0 & i = j, \end{cases} \qquad i, j \in [n]. \tag{7}$$

That is, r_{ij} is the smallest value (with respect to the maximum agent and client travel cost incurred) at which client i can meet-in-the-middle with an agent from the location of client j. We define the *minimum balancing costs* by

$$d_{ij} := \min\{c_{ij}, w_{ij}, r_{ij}\}, \qquad i, j \in [n]. \tag{8}$$

The minimum balancing cost d_{ij} is the smallest possible value (with respect to maximum agent and client travel cost incurred) at which client i can obtain service from a facility at the location of client j (possibly through meeting-in-the-middle at the location of some other client).

Regarding the notation, we will think of d_{ij} as the distance between client i and client j in our methodological developments; in particular, we show in Subsection 3.2 that the meet-in-the-middle k-center problem reduces to a k-center problem with distances given by minimum balancing costs. As we demonstrate below, however, minimum balancing costs do not necessarily define a metric in any of the cost regimes described in Subsection 2.2. Finally, we collect the minimum balancing costs into the $n \times n$ matrix d. That is, d_{ij} is the ij-th entry of d.

Minimum balancing costs in different cost regimes. Clearly $d_{ij} \in \mathbb{R}_{\geq 0}$ and $d_{ii} = 0$, so we restrict our attention to the symmetry and triangle inequality conditions.

Example 3.1 below shows that minimum balancing costs are not necessarily symmetric and do not necessarily satisfy the triangle inequality in the metric costs regime. Furthermore, the example shows that both symmetry and the triangle inequality can be severely violated.

Example 3.1. Suppose that there are $n = 3$ clients, and suppose the agent and client travel costs are given by

$$c = \begin{bmatrix} 0 & \epsilon & M+\epsilon \\ \epsilon & 0 & M \\ M+\epsilon & M & 0 \end{bmatrix} \quad \text{and} \quad w = \begin{bmatrix} 0 & M & M+\epsilon \\ M & 0 & \epsilon \\ M+\epsilon & \epsilon & 0 \end{bmatrix},$$

respectively, where $\epsilon, M \in \mathbb{Z}_{\geq 0}$ such that $\epsilon \ll M$. It is easy to check that these costs individually define metrics (i.e., satisfy (3) and (4)), so the conditions of the metric costs regime are met. Observe that

$$d_{13} = \min\{c_{13}, w_{13}, \min\{\max\{c_{23}, w_{12}\}\}\} = \min\{M+\epsilon, M+\epsilon, M\} = M,$$

$$
\begin{aligned}
d_{31} &= \min\{c_{31}, w_{31}, \min\{\max\{c_{21}, w_{32}\}\}\} = \min\{M+\epsilon, M+\epsilon, \epsilon\} = \epsilon, \\
d_{12} &= \min\{c_{12}, w_{12}, \min\{\max\{c_{32}, w_{13}\}\}\} = \min\{\epsilon, M, M\} = \epsilon, \\
d_{23} &= \min\{c_{23}, w_{23}, \min\{\max\{c_{13}, w_{21}\}\}\} = \min\{M, \epsilon, M\} = \epsilon.
\end{aligned}
$$

We see that the minimum balancing costs do not satisfy the symmetry condition because $d_{13} = M \gg \epsilon = d_{31}$, and they do not satisfy the triangle inequality because $d_{13} = M \gg 2\epsilon = d_{12} + d_{23}$. □

Example 3.2 below shows that minimum balancing costs are not necessarily symmetric in the related metric costs regime. (Example 3.3 further below implies that the symmetry condition does not necessarily hold as well.)

Example 3.2. Suppose that there are $n = 3$ locations, and suppose that the agent and client travel costs are given by

$$
c = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix} \quad \text{and} \quad w = \begin{bmatrix} 0 & 1 & 3 \\ 1 & 0 & 2 \\ 3 & 2 & 0 \end{bmatrix}.
$$

It is readily verified that the conditions (3), (4), and (5) of related metric costs regime are met. Observe that

$$
\begin{aligned}
d_{13} &= \min\{c_{13}, w_{13}, \min\{\max\{c_{23}, w_{12}\}\}\} = \min\{2, 3, 1\} = 1 \\
d_{31} &= \min\{c_{31}, w_{31}, \min\{\max\{c_{12}, w_{32}\}\}\} = \min\{2, 3, 2\} = 2.
\end{aligned}
$$

Hence the symmetry condition does not hold because $d_{13} = 1 \neq 2 = d_{31}$.

Example 3.3 below shows that minimum balancing costs do not necessarily satisfy the triangle inequality in the equal metric costs regime, implying that the minimum balancing costs do not necessarily satisfy the triangle inequality in the related metric costs regime as well. It is straightforward to verify that minimum balancing costs are always symmetric in the equal metric costs regime.

Example 3.3. Suppose that there are $n = 5$ clients. Let $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 4$, and $x_5 = 6$. Also let $c_{ij} = w_{ij} = |x_i - x_j|$ for $i, j \in [5]$, so the agent and client travel costs are equal. It is readily verified that these costs also individually define metrics, so the conditions of the equal metric costs regime are met. We have that

$$
\begin{aligned}
d_{15} &= \min\{c_{15}, \min\{\max\{c_{12}, c_{25}\}, \max\{c_{13}, c_{35}\}, \max\{c_{14}, c_{45}\}\}\} \\
&= \min\{6, \min\{5, 4, 4\}\} = 4, \\
d_{13} &= \min\{c_{13}, \min\{\max\{c_{12}, c_{23}\}, \max\{c_{14}, c_{43}\}, \max\{c_{15}, c_{53}\}\}\} \\
&= \min\{2, \min\{1, 4, 6\}\} = 1, \\
d_{35} &= \min\{c_{35}, \min\{\max\{c_{31}, c_{15}\}, \max\{c_{32}, c_{25}\}, \max\{c_{34}, c_{45}\}\}\} \\
&= \min\{4, \min\{6, 5, 2\}\} = 2.
\end{aligned}
$$

Hence the triangle inequality does not hold because $d_{15} = 4 \not\leq 3 = d_{13} + d_{35}$.

Example 3.2 and 3.3, in contrast to Example 3.1, do not demonstrate that symmetry and the triangle inequality can be severely violated. Proposition 3.1 below shows that symmetry and the triangle inequality can only be violated up to small multiplicative constants in the related metric costs regime. It follows that the triangle inequality can only be violated up to a small multiplicative constant in the equal metric costs regime as well. The proof of the proposition does not provide any additional insight, so we postpone it to Appendix A.1

Proposition 3.1. *In the related metric costs regime, namely when conditions* (3), (4), *and* (5) *are met, it holds that*

$$d_{ij} \leq 2d_{ji}, \qquad i, j \in [n],$$

as well as

$$d_{ij} \leq 2(d_{ik} + d_{kj}), \qquad i, j, k \in [n].$$

Proof. See Appendix A.1. □

Remark 3.1. Example 3.2 establishes that the analysis of Proposition 3.1 is tight with regards to symmetry. However, Example 3.3 does not establish that the analysis is tight with respect to the triangle inequality. We do not use Proposition 3.1 to establish any of the main results in this work, so we do not concern ourselves with further investigating whether or not the analysis is tight.

3.2 Locating meet-in-the-middle sites given facilities

Given facilities $F \subseteq [n]$, we consider the subproblem of locating meet-in-the-middle sites. That is, we consider the following subproblem of (2):

$$\min_{M,R} \{f(F, M, R) : M \subseteq [n] \setminus F \text{ and } R = [n] \setminus (F \cup M)\}. \tag{9}$$

We present a description of the algorithm that we propose for solving (9) in Algorithm 1 below. We show that the algorithm returns an optimal solution to the subproblem under no assumptions on the agent or client travel costs; see Theorem 3.1 below. Furthermore, as a byproduct of this analysis, we show that the the meet-in-the-middle k-center problem reduces in polynomial time to the k-center problem (also under no assumption on the agent or client travel costs).

Algorithm 1 takes facilities F, agent travel costs c, and client travel costs w as input. The first few steps of algorithm are dedicated to initialization. Steps 1 and 2 compute the quantity r_{if} and the minimum balancing cost d_{if} as defined in (7) and (8), respectively, for each client-facility pair $(i, f) \in ([n] \setminus F) \times F$. Step 3 computes a facility f_i that is closest to each client $i \in [n] \setminus F$ without a facility. Step 4 initializes the set M of meet-in-the-middle sites.

Steps 5-14 iterate through the clients $[n] \setminus F$ without facilities and construct the set M of meet-in-the-middle sites. At each iteration, the algorithm adds index $i \in [n] \setminus F$ to the set M if a most cost-effective way for client i to obtain service from facility f_i (with respect maximum agent and client travel cost incurred) is to send an agent from facility f_i to client i; see Steps 6-7. If it is not a most cost-effective way, then the algorithm checks in Step 8 whether it is most cost-effective for client i to meet-in-the-middle somewhere with an agent from facility f_i. If this is the case, then the algorithm selects the client j in Step 9 whose location should serve as the meet-in-the-middle site. If $j \notin F$, then the algorithm adds j to M in Step 11. This constitutes an iteration of the algorithm. Note that the algorithm does not add any indices to M in an iteration if it is (strictly) most cost-effective for client i to retrieve service from facility f_i.

Algorithm 1 Locating meet-in-the-middle sites

Input: Facilities $F \subseteq [n]$, agent travel costs $c \in \mathbb{R}^{n\times n}_{\geq 0}$, and client travel costs $w \in \mathbb{R}^{n\times n}_{\geq 0}$

1: $r_{if} \leftarrow \min_{k\in[n]\setminus\{i,f\}} \max\{c_{kf}, w_{ik}\}$ for each $(i,f) \in ([n] \setminus F) \times F$
2: $d_{if} \leftarrow \min\{c_{if}, w_{if}, r_{if}\}$ for each $(i,f) \in ([n] \setminus F) \times F$
3: $f_i \leftarrow \operatorname{argmin}_{f\in F} d_{if}$ for each $i \in [n] \setminus F$
4: Initialize meet-in-the-middle sites $M \leftarrow \emptyset$
5: **for** $i \in [n] \setminus F$ **do**
6:     **if** $d_{if_i} = c_{if_i}$ **then**
7:         $M \leftarrow M \cup \{i\}$
8:     **else if** $d_{if_i} = r_{if_i}$ **then**
9:         $j \leftarrow k' \in \operatorname{argmin}_{k\in[n]\setminus\{i,f_i\}} \max\{c_{kf_i}, w_{ik}\}$
10:         **if** $j \notin F$ **then**
11:             $M \leftarrow M \cup \{j\}$
12:         **end if**
13:     **end if**
14: **end for**
15: Set retrieving clients $R \leftarrow [n] \setminus (F \cup M)$
16: Return (M, R)

Computational complexity. Step 1 of Algorithm 1 requires $O(n^3)$ operations, and Step 2 requires $O(n^2)$ operations. Assuming $|F| = k$, Step 3 requires $O((n-k)k)$ operations, and Steps 4-14 require $O((n-k)n)$ operations because Step 9 requires $O(n)$ operations. If the quantities in Step 9 are precomputed, which can be done in Step 1 at no additional expense to the big-O computational complexity, then Steps 4-14 only require $O(n-k)$ operations. Therefore, the computational complexity of Algorithm 1 is $O(n^3)$. Consequently, Algorithm 1 runs in polynomial time.

Analysis of Algorithm 1. For the remainder of this section, we mainly set out to show that Algorithm 1 returns an optimal solution to (9).

Let OPT_F denote the optimal value of (9). We present a lower bound on OPT_F in Lemma 3.1 below. The lower bound holds because d_{if} is the smallest possible value (with respect to the maximum agent and client travel cost incurred) at which client $i \in [n] \setminus F$ can obtain service (possibly through a meet-in-the-middle site) from facility $f \in F$. We present a formal proof that unpacks the definition of the minimum balancing costs d in Appendix A.2.

Lemma 3.1. $\max_{i\in[n]\setminus F} \min_{f\in F} d_{if} \leq \mathrm{OPT}_F$.

Proof. See Appendix A.2 $\square$

Now consider Lemma 3.2 below. The lemma states that Algorithm 1 outputs a solution to subproblem (9) whose objective value provides a lower bound on the lower bound presented in Lemma 3.1. Because subproblem (9) is a maximization problem, Theorem 3.1 then immediately follows. We can also conclude that the lower bound presented Lemma 3.1 is tight; we formally make note of this observation in Theorem 3.2.

Intuition for Lemma 3.2 follows from the the discussion provided above about Steps 5-14 of Algorithm 1. That is, indices are added to the set M of meet-in-the-middle sites so that clients $[n] \setminus F$ can obtain service in the most cost-effective way. We present a formal proof of Lemma 3.2 that follows this logic in Appendix A.3.

Lemma 3.2. *The output* (M, R) *of Algorithm 1 satisfies*

$$f(F, M, R) \leq \max_{i \in [n] \setminus F} \min_{f \in F} d_{if}.$$

Proof. See Appendix A.3. □

Theorem 3.1. *The output* (M, R) *of Algorithm 1 is optimal for subproblem* (9)*.*

Proof. The theorem immediately follows from Lemmas 3.1 and 3.2. □

Theorem 3.2. $\max_{i \in [n] \setminus F} \min_{f \in F} d_{if} = \mathrm{OPT}_F$.

Proof. The theorem immediately follows from Lemmas 3.1 and 3.2. □

Reduction to the k-center problem. Consider the following k-center problem that uses minimum balancing costs (8) as distances:

$$\min_{F \subseteq [n] : |F| = k} \max_{i \in [n] \setminus F} \min_{f \in F} d_{if}. \tag{10}$$

From Theorem 3.2, it follows that the optimal value of (10) equals the optimal value of the meet-in-the-middle k-center problem (2). Furthermore, we see from Theorem 3.1 that given an optimal solution F^* to (10), we can run Algorithm 1 with input F^* to obtain (M^*, R^*) such that (F^*, M^*, R^*) is optimal for the meet-in-the-middle k-center problem (2). Thus, the meet-in-the-middle k-center problem reduces in polynomial time to the k-center problem. Note, however, that because our minimum balancing costs do not necessarily define a metric in any of the cost regimes (see Section 2.2), the reduction does not provide a reduction to the metric k-center problem.

4 Meet-in-the-middle location methods

In Subsection 4.1, we present MILP formulations, and in Subsection 4.2, we present and study a greedy approximation algorithm. Both the formulations and greedy method are applicable under no assumptions on the agent or client travel costs.

4.1 MILP formulations

First, we introduce an *assignment formulation* that assigns each meet-in-the-middle site to a facility and each retrieving client to a facility or meet-in-the-middle site. Next, we introduce a *k-center formulation* based on the reduction provided at the end of Subsection 3.2.

Assignment formulation. For each $i \in [n]$, we introduce a binary variable $x_i \in \{0,1\}$ that indicates if we build a facility at the location of client i. For each $i \neq j \in [n]$, we introduce a binary variable $y_{ij} \in \{0,1\}$ that indicates if an agent travels from a facility at the location of client j to a meet-in-the-middle site at the location of client i, and we introduce the binary variable $z_{ij} \in \{0,1\}$ that indicate if client i retrieves goods from the location of client j.

Consider the following assignment MILP formulation of (2):

$$
\begin{aligned}
\min_{\eta,x,y,z} \quad & \eta \\
\text{s.t.} \quad & \eta \geq c_{ij} y_{ij} && i \neq j \in [n] && \text{(11a)} \\
& \eta \geq w_{ij} z_{ij} && i \neq j \in [n] && \text{(11b)} \\
& y_{ij} \leq x_j && i \neq j \in [n] && \text{(11c)} \\
& z_{ij} \leq x_j + \sum_{\ell \in [n]: \ell \neq j} y_{j\ell} && i \neq j \in [n] && \text{(11d)} \\
& x_i + \sum_{j \in [n]: j \neq i} y_{ij} + \sum_{j \in [n]: j \neq i} z_{ij} = 1 && i \in [n] && \text{(11e)} \\
& \sum_{i=1}^{n} x_i = k && i \in [n] && \text{(11f)} \\
& \eta \in \mathbb{R}, \ x \in \{0,1\}^n \\
& y_{ij}, z_{ij} \in \{0,1\} && i \neq j \in [n].
\end{aligned}
$$

Constraints (11a) and (11b) ensure that the objective equals the maximum (agent or client) travel cost incurred. Constraints (11c) enforce that agents can only be sent to meet-in-the-middle sites from locations with facilities. Constraints (11d) ensure that clients can only retrieve service from a location with a facility or meet-in-the-middle site. Constraints (11e) ensure that each client is serviced. Finally, constraints (11f) ensure that there are k facilities.

Note that if (x^*, y^*, z^*, η^*) is optimal for formulation (11), then (F^*, M^*, R^*) defined by

$$
\begin{aligned}
F^* &= \{i \in [n] : x_i^* = 1\}, \\
M^* &= \{i \in [n] : y_{ij}^* = 1 \text{ for some } j \neq i \in [n]\} \\
R^* &= \{i \in [n] : z_{ij}^* = 1 \text{ for some } j \neq i \in [n]\}
\end{aligned}
$$

is an optimal solution for the meet-in-the-middle k-center problem (2).

k-center formulation. As in the assignment formulation, we introduce a binary variable $x_i \in \{0,1\}$ for each $i \in [n]$ that indicates whether we build a facility at the location of client i. For each $i, j \in [n]$, we introduce a binary variable $y_{ij} \in \{0,1\}$ that indicates if client i is "assigned" to a facility at the location of client j. More precisely, $y_{ij} = 1$ if the facility at the location of client j is a facility that is closet to location i (breaking ties arbitrarily) under distances that are given by minimum balancing costs (8).

We formulate the k-center problem (10) as the MILP:

$$
\begin{aligned}
\min_{x,y,\eta} \quad & \eta \\
\text{s.t.} \quad & \eta \geq \sum_{j \in [n]} d_{ij} y_{ij} && i \in [n] && \text{(12a)} \\
& y_{ij} \leq x_j && i, j \in [n] && \text{(12b)} \\
& \sum_{j \in [n]} y_{ij} = 1 && i \in [n] && \text{(12c)} \\
& \sum_{i \in [n]} x_i = k && i \in [n] && \text{(12d)} \\
& x \in \{0,1\}^n, \ y \in \{0,1\}^{n \times n}, \ \eta \in \mathbb{R}.
\end{aligned}
$$

Constraints (12a) enforce that the objective equals the maximum distance from a location without a facility to a location with a facility. Constraints (12b) ensure that a client can only be assigned to a client location that has a facility. Constraints (12c) enforce that each client is assigned to a facility. Finally, constraints (12d) ensure that there are k facilities.

Proposition 4.1 shows how to construct an optimal solution to the meet-in-the-middle k-center problem (2) from an optimal solution to the MILP (12). The proposition immediately follows from the discussion at the end of Subsection 3.2.

Proposition 4.1. *Suppose that* (x^*, y^*, η^*) *is an optimal solution to* (12)*. Then* (F^*, M^*, R^*) *is an optimal solution to* (2)*, where* (M^*, R^*) *is the output of Algorithm 1 run with input* $F^* = \{i \in [n] : x_i^* = 1\}$.

4.2 Greedy algorithm

We present a description of the greedy algorithm that we propose in Algorithm 2 below. Steps 1 and 2 of the algorithm initialize the quantity r_{ij} and the minimum balancing cost d_{ij} as defined in (7) and (8), respectively, for each $i, j \in [n]$. Steps 3-7 apply the greedy algorithm of [12] for the k-center problem, using minimum balancing costs (8) as distances. More specifically, in Step 3, our algorithm initializes the set of facilities F with an arbitrarily chosen client $i \in [n]$. In Steps 4-7, the algorithm adds new clients to F (according to the rule described in Step 5, which chooses a client that is farthest from the facilities F) until F contains k indices. In Step 8, the algorithm applies Algorithm 1 with input F to obtain meet-in-the-middle sites M and retrieving clients R. Finally, the algorithm returns (F, M, R) as a solution to the meet-in-the-middle k-center problem (2).

Algorithm 2 Greedy algorithm

Input: Agent travel costs $c \in \mathbb{R}^{n\times n}_{\geq 0}$ and client travel costs $w \in \mathbb{R}^{n\times n}_{\geq 0}$

1: $r_{ij} \leftarrow \min_{k\in[n]\setminus\{i,j\}} \max\{c_{kj}, w_{ik}\}$ for each $i, j \in [n]$
2: $d_{ij} \leftarrow \min\{c_{ij}, w_{ij}, r_{ij}\}$ for each $i, j \in [n]$
3: Initialize facility locations $F \leftarrow \{i\}$, where $i \in [n]$
4: **while** $|F| < k$ **do**
5: $\quad i \leftarrow j' \in \operatorname{argmax}_{j\in[n]\setminus F} \min_{f\in F} d_{jf}$
6: $\quad F \leftarrow F \cup \{i\}$
7: **end while**
8: $M, R \leftarrow$ Algorithm 1 with input F, c, and w
9: Return (F, M, R)

Computational complexity. Running Steps 1 and 2 of Algorithm 2 requires $O(n^3)$ operations. Step 5 of Algorithm 2 requires $O(n\ell)$ operations at the ℓ-th iteration. It follows that, as stated, Steps 4-7 use $O(nk^2)$ operations in total. By using $O(nk)$ memory, this total can be reduced to $O(nk)$ operations. Recall from Section 3.2 that Algorithm 1 has $O(n^3)$ computational complexity. The total computational complexity of Algorithm 2 is therefore given by $O(n^3)$.

Approximation guarantees. The greedy algorithm developed in [12] is a 2-approximation algorithm for the *metric* k-center problem, but it is not immediately clear to what extent we can establish a similar guarantee for Algorithm 2. The main challenge is that distances in our setting (i.e., minimum balancing costs) do not necessarily define a metric in any of the cost regimes of interest (see Subsection 2.2). Note that we could utilize the fact that minimum balancing costs approximately define a metric in some of the cost regimes (see Proposition 3.1), but an alternative analysis actually provides better (in fact tight) approximation guarantees.

We present our main results in Theorem 4.1 below. The theorem states that Algorithm 2 provides a 3-approximation in the related metric costs regime and a 2-approximation in the more restrictive equal metric costs regime. Accordingly, the 2-approximation algorithm of [12] extends in the equal metric costs regime but not in the more general related metric costs regime.

Theorem 4.1. *The greedy algorithm (Algorithm 2) is a* 3*-approximation algorithm for the meet-in-the-middle k-center problem* (2) *in the related metric costs regime (i.e., when conditions* (3)*,* (4)*, and* (5) *hold) and a* 2*-approximation algorithm in the equal metric costs regime (i.e., when conditions* (3)*,* (4)*, and* (6) *hold).*

We present our proof of Theorem 4.1 here. Let (F^*, M^*, R^*) be an optimal solution to the meet-in-the-middle k-center problem (2). Define

$$\mathrm{OPT} := f(F^*, M^*, R^*)$$

to be the optimal value of (2). For each $f \in F^*$, define the sets:

$$M^*_f := \left\{ i \in M^* : f \in \operatorname*{argmin}_{f\in F^*} c_{if} \right\},$$

$$R_f^* := \left\{ i \in R^* : \left[\underset{j \in \{f\} \cup M_f^*}{\operatorname{argmin}} \; w_{ij} \right] \cap \left[\underset{j \in F^* \cup M^*}{\operatorname{argmin}} \; w_{ij} \right] \neq \emptyset \right\},$$
$$C_f := \{f\} \cup M_f^* \cup R_f^*.$$

If $i \in M_f^* \subset M^*$, then it is optimal for the meet-in-the-middle site at the location of client i to be serviced by an agent from facility f. If $i \in R_f^* \subset R^*$, then it is optimal for client i to retrieve goods from facility f or one of the meet-in-the-middle sites M_f^*. We refer to the set C_f as an *optimal cluster*. Note that the optimal clusters cover $[n]$ (i.e., $[n] = \cup_{f \in F^*} C_f$), but two different optimal clusters do not necessarily have a nonempty intersection.

Before proving Theorem 4.1, we make note of Lemma 4.1 below. The lemma states that if two indices are in the same optimal cluster, then the distance (i.e., minimum balancing cost) between the two corresponding clients is less than a constant multiple of OPT, where the constant depends on the cost regime. The proof of the lemma requires a significant amount of casework, so we present the proof in Appendix A.4.

Lemma 4.1. *Let $f \in F^*$ and $i, j \in C_f$. Then*

(i) $d_{ji} \leq 3\mathrm{OPT}$ *in the related metric costs regime, and*

(ii) $d_{ji} \leq 2\mathrm{OPT}$ *in the equal cost metric costs regime.*

Proof. See Appendix A.4 □

We are now prepared to prove Theorem 4.1. Our proof follows in a similar spirit to the proof of [12]. In contrast, however, our proof exploits Lemma 4.1 instead of metric properties (which are not present in our setting).

Proof of Theorem 4.1. Let (F, M, R) denote the output of Algorithm 2. From Theorem 3.2, it holds that $f(F, M, R) = \max_{j \in [n] \setminus F} \min_{f \in F} d_{jf}$. Let $j \in [n] \setminus F$. It is then sufficient to show that

$$\min_{f \in F} d_{jf} \leq \alpha \mathrm{OPT}, \tag{13}$$

where $\alpha = 3$ or $\alpha = 2$, depending on the cost regime.

Because the optimal clusters cover $[n]$, it holds that $j \in C_f$ for some $f \in F^*$. We consider two cases:

Case 1. Suppose that no two indices in F belong to the same optimal cluster. Then there is an index $i \in F \cap C_f$. Observe that

$$\min_{f \in F} d_{jf} \leq d_{ji} \leq \alpha \mathrm{OPT},$$

where the last inequality follows from Lemma 4.1. Thus (13) holds.

Case 2. Suppose that two indices in F are in the same optimal cluster, say $k, \ell \in \mathcal{C}_{\hat{f}}$ for some $\hat{f} \in F^*$. Without loss of generality, suppose that Algorithm 2 added k to F before ℓ, and let $\hat{F}$ denote the set of indices that were added to F before ℓ. Because $\hat{F} \subseteq F$, we have that

$$\min_{f \in F} d_{jf} \leq \min_{f \in \hat{F}} d_{jf} \leq d_{\ell k} \leq \alpha \mathrm{OPT},$$

where the second inequality follows from Step 5 of Algorithm 2, and the last inequality from Lemma 4.1. □

We conclude this subsection by noting that the factors of 2 and 3 stated in Theorem 4.1 are tight. For the sake of space, we postpone a proof of this fact to Appendix A.5

Proposition 4.2. *There exist instances of the meet-in-the-middle k-center problem* (2) *with related and equal metric costs per which Algorithm 2 provides a* 3*-approximation and* 2*-approximation, respectively.*

Proof. See Appendix A.5. □

5 Computational experiments

In Subsection 5.1, we apply the methods of Section 4 to synthetic meet-in-the-middle k-center instances. We (i) compare the solve times of the assignment MILP formulation (11) and k-center MILP formulation (12), (ii) investigate the empirical approximation performance of Algorithm 2, and (iii) explore the benefits of simultaneously locating facilities and meet-in-the-middle sites in comparison with sequentially locating facilities and meet-in-the-middle sites. In Subsection 5.2, we apply our methods in a vaccine campaign design case study.

Hardware and software. Our computational study is performed on an x86_64 Linux server with eight Intel(R) Xeon(R) E5-2697 v4 2.30GHz CPUs and 16 Gb memory. The MILP formulations are solved using Gurobi Optimizer 11.02, and our algorithms are coded using Python 3.8.

5.1 Synthetic instances

Instance test bed. We generate instances of the meet-in-the-middle k-center problem as follows. First, we draw n client locations from the uniform distribution on the unit square $[0,1]^2$. Then, we set the agent travel cost c_{ij} for each $i, j \in [n]$ to be the Euclidean distance between the location of client i and j. Next, we set the client travel costs. To this end, we consider two approaches, yielding two different problem instances. In the first approach, we set $w_{ij} = c_{ij}$ for each $i, j \in [n]$, which provides an equal metric cost regime instance. In the second approach, we set $w_{ij} = 2c_{ij}$ for each $i, j \in [n]$, which provides a related metric cost regime instance (see Subsection 2.2).

For each $n \in \{10, 50, 100, 200, \ldots, 800\}$, we apply this generation procedure 10 times, yielding $10 \times 10 \times 2 = 200$ instances. To fully define the instances, we must also specify k. For each of the 200 instances, we consider 5 values of k, yielding $200 \times 5 = 1000$ instances in total. Specifically, we take k such that $k/n \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. We report averaged results (along with confidence intervals) across the 10 instances for each pair (n, k) and instance type (i.e., equal or related metric costs).

MILP solve times. For each instance in the test bed, we apply Gurobi to the corresponding assignment MILP (11) and k-center MILP (12) for 3 minutes. We report average solve times in seconds in Figure 2. If Gurobi does not terminate in 3 minutes, then we report 3 minutes

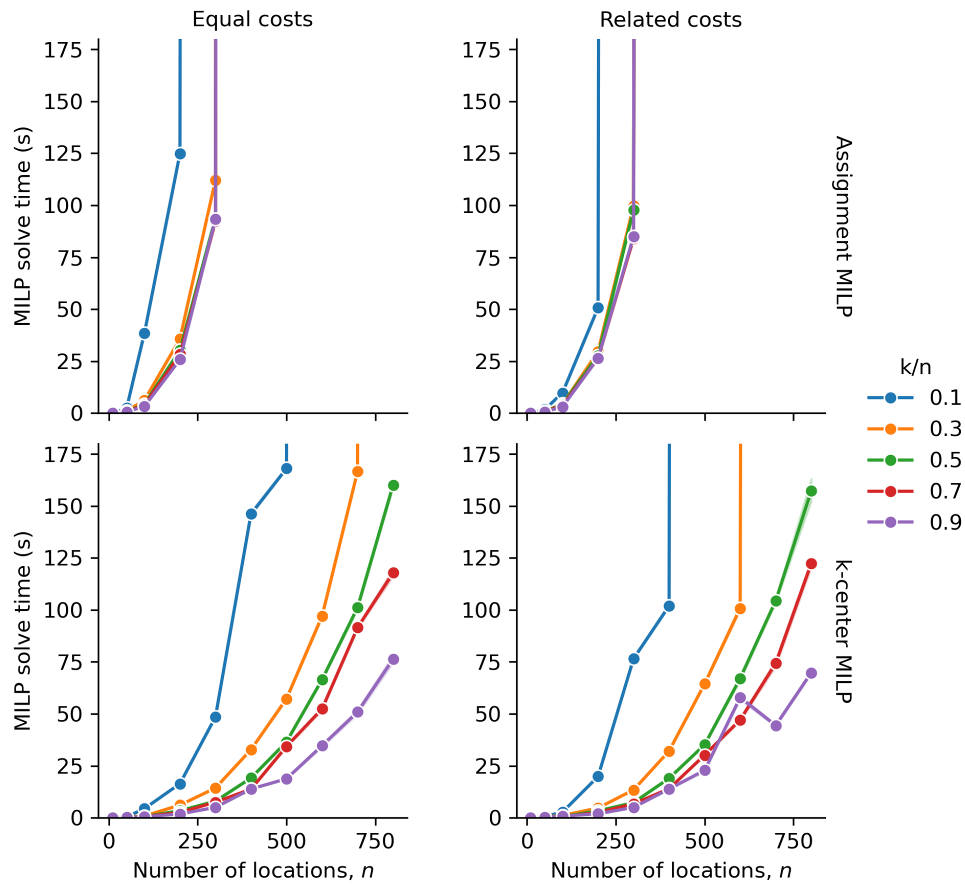


Figure 2: Average solve time in seconds of assignment MILP (11) and k-center MILP (12) against number n of locations on test bed instances. If Gurobi does not terminate in 3 minutes, then solve time is reported as 3 minutes. (Solve times of 3 minutes are truncated from the plots.)

as the solve time. (Recall that we must apply Algorithm 1 after solving the k-center MILP to compute meet-in-the-middle sites and retrieving clients. We do not report the runtime of Algorithm 1 because it is negligible compared to the MILP runtime for moderately large values of n, as reported.)

Not surprisingly, solve times grow with the number of client locations n. On a more interesting note, more time is required to solves instances with a smaller proportion k/n of facilities. Also, we see that the k-center MILP solves an order of magnitude faster than the assignment MILP. Finally, equal metric cost instances require slightly more time to solve, most likely due to the symmetry in their agent and client travel cost structure.

Approximation performance of Algorithm 2. We apply Algorithm 2 to instances in the test bed with $n \in \{10, 50, 100, 200, \dots, 400\}$ and record the empirical approximation ratio, namely $f(F, M, R)/f(F^*, M^*, R^*)$, where (F^*, M^*, R^*) is an optimal solution found with Gurobi, and (F, M, R) is the output of Algorithm 2. Taking $n \in \{10, 50, 100, 200, 300, 400\}$ ensures that Gurobi finds an optimal solution (to the k-center MILP formulation); see Figure 2. We report the average empirical approximation ratio (along with 95% confidence intervals) in Figure 3.

Recall Theorem 4.1, which states that Algorithm 2 is a 2-approximation algorithm in the equal metric costs regime and a 3-approximation algorithm in the related metric costs regime. We see that the approximation guarantees indeed hold on the test bed, and the empirical approximation ratio is higher on related cost instances. In particular, as the proportion k/n of facility locations increases, the empirical approximation ratio improves more on equal costs instances than on related costs instances. While the empirical approximation ratio improves as the proportion of facilities increases, there is no distinguishable dependence on the number n of locations.

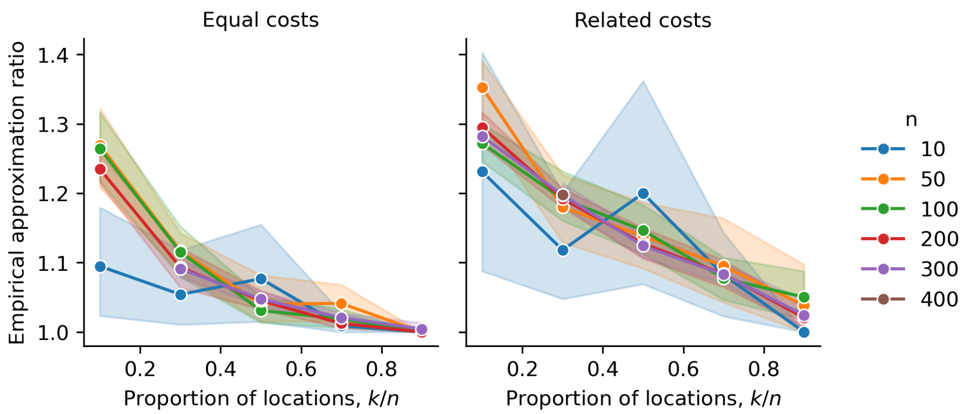


Figure 3: Average empirical approximation ratio of Algorithm 2 (with 95% confidence intervals) against proportion of locations k/n on test bed instances with $n \in \{10, 50, 100, 200, 300, 400\}$.

Simultaneous versus sequential location. We compare simultaneously versus sequentially locating facilities and meet-in-the-middle sites. To this end, we consider a similar setup to the example illustrated in Figure 1 from Section 1.

For each instance in the test bed with $n = 100$, we perform the following. Analogous to Plot (b) of Figure 1, in which there are no meet-in-the-middle sites, we compute an optimal solution $\hat{F} \subseteq [n]$ to the k-center MILP (12) with $d = c$ (instead of taking d to be the minimum balancing costs). We record the maximum travel cost $f(\hat{F}, \emptyset, [n] \setminus \hat{F})$ incurred when there are no meet-in-the-middle sites. We also record the percent reduction to this maximum travel cost provided by simultaneous location:

$$\frac{f(\hat{F}, \emptyset, [n] \setminus \hat{F}) - f(F^*, M^*, R^*)}{f(\hat{F}, \emptyset, [n] \setminus \hat{F})},$$

where (F^*, M^*, R^*) is an optimal solution to the meet-in-the-middle k-center problem (found with Gurobi). Analogous to Plot (c) of Figure 1, in which meet-in-the-middle sites are sequentially located, we apply Algorithm 1 with $\hat{F}$ as input to obtain $(\hat{F}, \hat{M}, \hat{R})$. We record the maximum travel cost $f(\hat{F}, \hat{M}, \hat{R})$ incurred under sequential location. We also record the percent reduction to this maximum travel cost provided by simultaneous location:

$$\frac{f(\hat{F}, \hat{M}, \hat{R}) - f(F^*, M^*, R^*)}{f(\hat{F}, \hat{M}, \hat{R})}.$$

We report the average maximum travel costs and percentage reductions (along with 95% confidence intervals) against the proportion k/n of facility locations in Figures 4 and 5, respectively.

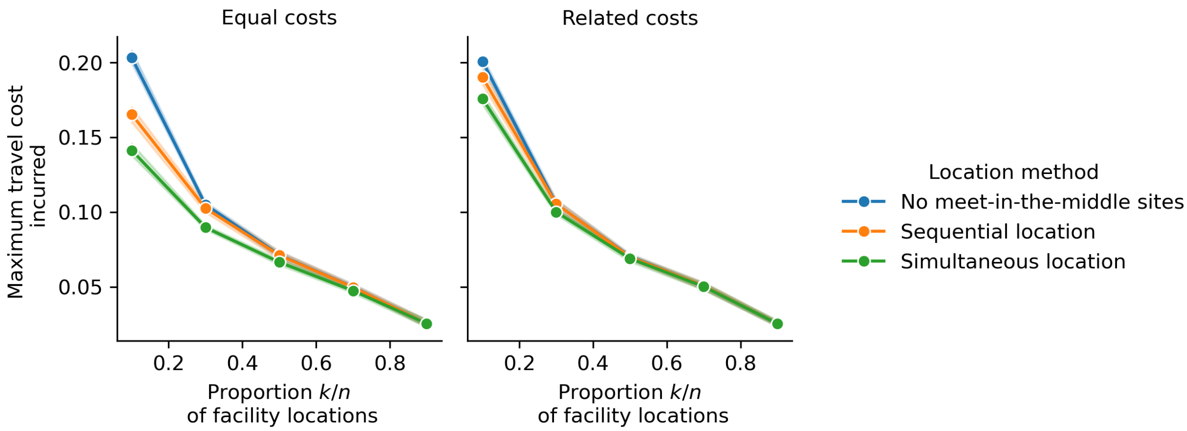


Figure 4: Average maximum travel cost (with 95% confidence intervals) against the proportion k/n of facilities locations on test bed instances with $n = 100$.

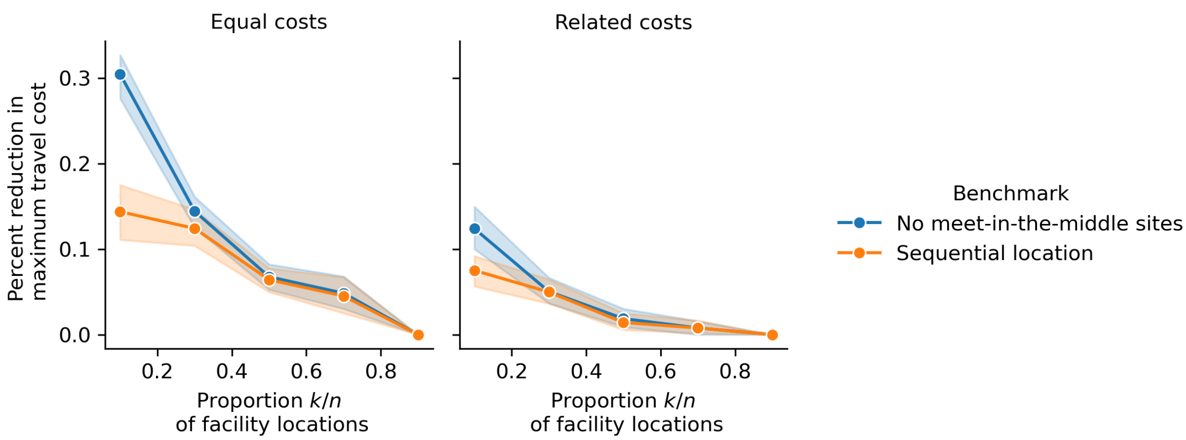


Figure 5: Average percent reduction in maximum travel cost (with 95% confidence intervals) provided by simultaneous location against the proportion k/n of facilities locations on test bed instances with $n = 100$. Percentage reductions are presented for benchmarks that do not locate meet-in-the-middle sites and that perform sequential location.

We focus our discussion on Figure 5. Naturally, the percent reduction for no meet-in-the-middle sites is greater than the percentage reduction for sequential location. Although, the benefit of meeting-in-the-middle (i.e., difference between blue and orange curve) diminishes as the proportion k/n of facility locations increases because facilities are more densely located. Both percent reductions decrease as the proportion k/n of facility locations increases. Interestingly, the percent reductions are larger on equal cost instances; on related cost instances, it is more important to build meet-in-the-middle sites closer to clients, as client travel costs contribute more to the objective, so facility locations are not as significant. When there are a small number of facilities, we see that simultaneous location provides an approximate, average 15% and 8% reduction relative to sequential location for equal and related costs instances, respectively; we explore the extent to which this holds on real-world data as well in the Subsection 5.2.

5.2 Ghana vaccine campaign design case study

Ghana is divided into 16 regions that are further divided into 261 districts. In this subsection we consider a district-level vaccine campaign design case study. We use the Health Facilities dataset from the Ghana Open Data Initiative [33]. The dataset contains the locations of health establishments across 10 regions and 171 districts.

We consider locating vaccine hubs (facilities) and outreach (meet-in-the-middle) sites at existing health establishments in each district. We assume that the vaccine hubs have storage vessels for the vaccines as well as on-site health workers who are trained to administer the vaccines. The on-site workers can travel with vaccines to outreach sites to administer vaccines. Patients can travel to one of the vaccine hubs or outreach sites to be vaccinated.

We measure the distance between two health establishments (candidate locations for vaccine hubs and outreach sites) by the geospatial distance between their latitude and longitude coordinates in terms of miles. We restrict our attention to the 37 districts that have between 25 to 50 health establishments. For each of the 37 districts and $k \in \{1, 2, 3\}$, we record the percentage reduction in the maximum travel distance provided by simultaneous location, in the same manner as Subsection 5.1; specifically, see Figure 5 and the corresponding discussion. We report histograms for the percentage reduction in maximum travel distance across the 37 districts in Figure 6, based on benchmarks that do not locate any meet-in-the-middle sites and that perform sequential location.

Naturally, percent reduction of the maximum travel distance under no meet-in-the-middle sites is greater than under sequential location. We restrict our discussion here to percent reduction for sequential location. As the number k of facilities increase, it appears that more districts benefit from simultaneous location, i.e., more districts have a non-zero percent reduction. More precisely, for $k = 1, 2, 3$, there is a nonzero percent reduction in $12, 22, 26$ of the districts, respectively. For $k = 3$ facilities, there is an average percent reduction of 8% and a maximum percent reduction of 37% (for the district of Jomoro). Accordingly, simultaneous location is more effective for some districts than others.

To gain insight into the factors that make simultaneous location particularly effective, we plot the facilities and meet-in-the-middle sites from simultaneous and sequential location for the district of Jomoro in Figure 7. Simultaneous location serves the bottom right region with one facility, while sequential location serves the same region with two facilities. This in turn frees up one facility that can then be located at the somewhat remote location of Elubo. Accordingly, simultaneous location is potentially particularly effective for regions that have remote locations. Knowing that coverage can be extended through meet-in-the-middle sites, simultaneous location can locate less facilities in heavily populated areas and allocate more facilities to the outskirts of a region.

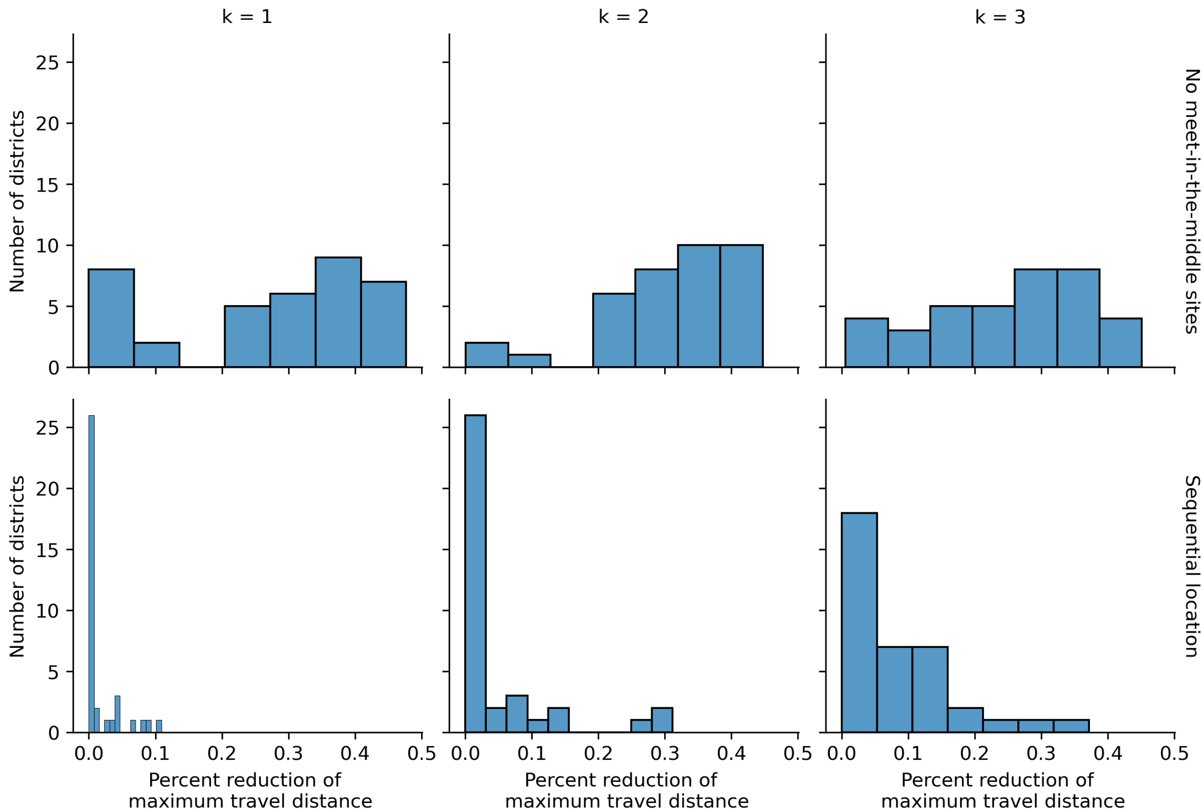
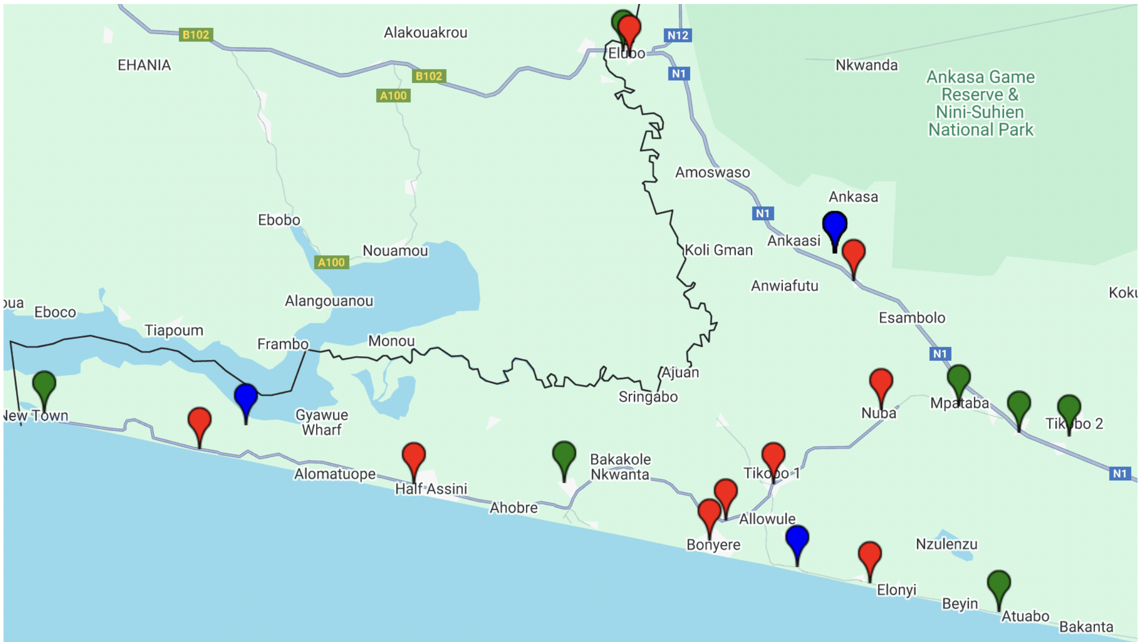


Figure 6: Histograms of the percentage reduction in maximum travel distance provided by simultaneous location across 37 districts in Ghana for $k = 1, 2, 3$ facilities. Percentage reductions are presented for benchmarks that do not locate meet-in-the-middle sites that perform sequential location.

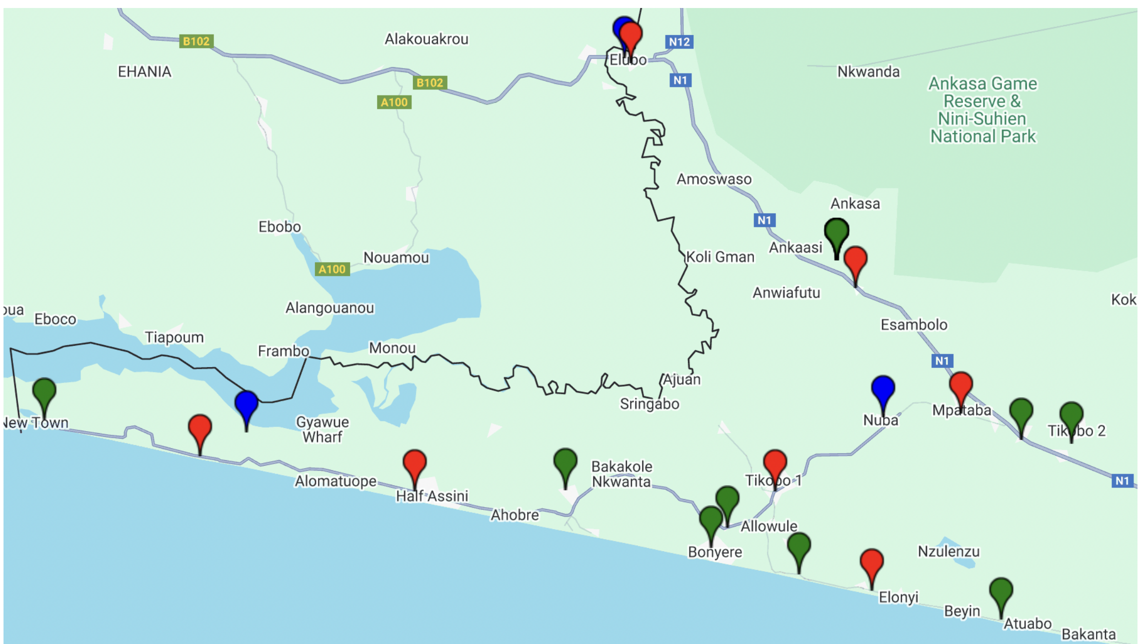
6 Discussion

We considered locating service facilities and meet-in-the-middle sites with the aim of optimally balancing accessibility and operational efficiency. To this end, we proposed and studied the meet-in-the-middle k-center problem. In Section 3 we established first principles, i.e., minimum balancing costs and a polynomial-time algorithm for locating meet-in-the-middle sites given facilities. From these first principles, we reduced the meet-in-the-middle k-center problem to the k-center problem, which in turn led to the development of the k-center MILP formulation (12) as well as a greedy approximation algorithm (i.e., Algorithm 2). Finally, we tested the methods on synthetic and real-world (inspired) instances in Section 5.

Our study provides a few interesting insights. Even if agent and client travel costs both define metrics, minimum balancing costs do not necessarily define a metric; see Subsection 3.1. Consequently, our reduction to the k-center problem cannot be trivially exploited to develop an approximation algorithm. The meet-in-the-middle k-center problem appears to be more challenging to approximate than the k-center problem; Algorithm 2 is a 3-approximation in the related costs regime (see Theorem 4.1), rather than a 2-approximation (like for the k-center problem). Furthermore, an approximation guarantee remains to be es-



(a) Sequential location of facilities and meet-in-the-middle sites



(b) Simultaneous location of facilities and meet-in-the-middle sites

Facility	Meet-in-the-middle site	Client

Figure 7: Sequential location (a) and simultaneous location (b) of $k = 3$ facilities and meet-in-the-middle sides for the Jomoro district of Ghana. Note that a facility is located near Elubo under simultaneous location.

tablished for the regime in which agent and client travel costs individually define metrics but are unrelated. Our computational experiments demonstrate that Gurobi solves the k-center MILP (12) an order of magnitude faster than the assignment MILP (11). Consequently, our reduction to the k-center problem proves practically useful. Finally, both our synthetic and real-world (inspired) computational experiments demonstrate that simultaneously locating facilities and meet-in-the-middle sites can (but does not always) significantly help to balance accessibility and operational efficiency. In particular, simultaneous location is potentially particularly effective in regions that have remote locations.

There are a few interesting directions for future research:

Modeling considerations. Naturally, it is of interest to account for other modeling considerations. It would be particularly interesting to investigate the extent to which the results presented herein extend under a budget constraint on the number of meet-in-the-middle sites. It is also of interest to account for fixed costs, supplies, demands, and capacities. Furthermore, one could investigate other objectives, such as the total travel distance or demand-weighted travel distance.

Approximation performance. Recall that we established that approximation factor of 3 for Algorithm 2 (see Theorem 4.1) is tight in related metric cost regime (see Proposition 4.2). Can an algorithm with a superior approximation guarantee be developed? Also, recall that we only established approximation guarantees for the equal and related metric costs regimes. Can an approximation algorithm be developed for the regime in which the agent and client travel costs define *unrelated* metrics? Finally, can one develop an approximation algorithm for the case in which there is a budget on the number of meet-in-the-middle sites?

References

[1] A. Ahmadi-Javid, P. Seyedi, and S. S. Syam, *A survey of healthcare facility location*, Computers & Operations Research, 79 (2017), pp. 223–263.

[2] M. Albareda-Sambola, Y. Hinojosa, A. Marín, and J. Puerto, *When centers can fail: A close second opportunity*, Computers & Operations Research, 62 (2015), pp. 145–156.

[3] A. Archer, *Two O(log*k)-approximation algorithms for the asymmetric k-center problem*, in Integer Programming and Combinatorial Optimization, K. Aardal and B. Gerards, eds., Heidelberg, 2001, Springer Berlin, pp. 1–14.

[4] D. Bertsimas, A. Delarue, and S. Martin, *Optimizing schools' start time and bus routes*, Proceedings of the National Academy of Sciences, 116 (2019), pp. 5943–5948.

[5] D. Celik Turkoglu and M. Erol Genevois, *A comparative survey of service facility location problems*, Annals of Operations Research, 292 (2020), pp. 399–468.

[6] M. Daskin and L. Dean, *Location of health care facilities*, in Operations Research and Health Care, Springer, 2004, pp. 43–76.

[7] R. Z. Farahani, S. Fallah, R. Ruiz, S. Hosseini, and N. Asgari, *OR models in urban service facility location: A critical review of applications and future developments*, European Journal of Operational Research, 276 (2019), pp. 1–27.

[8] R. Z. Farahani, M. SteadieSeifi, and N. Asgari, *Multiple criteria facility location problems: A survey*, Applied mathematical modelling, 34 (2010), pp. 1689–1709.

[9] J. Garcia-Diaz, J. Sanchez-Hernandez, R. Menchaca-Mendez, and R. Menchaca-Mendez, *When a worse approximation factor gives better performance: A 3-approximation algorithm for the vertex k-center problem*, Journal of Heuristics, 23 (2017), pp. 349–366.

[10] G. Ghiani, D. Laganà, E. Manni, and C. Triki, *Capacitated location of collection sites in an urban waste management system*, Waste management, 32 (2012), pp. 1291–1296.

[11] J. M. Gleason, *A set covering approach to bus stop location*, Omega, 3 (1975), pp. 605–608.

[12] T. F. Gonzalez, *Clustering to minimize the maximum intercluster distance*, Theoretical Computer Science, 38 (1985), pp. 293–306.

[13] M. Hasanzadeh Mofrad, *Optimizing vaccine clinic operations in low and middle income countries*, PhD thesis, University of Pittsburgh, 2016.

[14] Y. Hashimoto and H. Ribeiro Carretti, *Development and application of an immunization network design optimization model for UNICEF*, (2020).

[15] D. S. Hochbaum, *When are NP-hard location problems easy?*, Annals of Operations Research, 1 (1984), pp. 201–214.

[16] D. S. Hochbaum and D. B. Shmoys, *A best possible heuristic for the k-center problem*, Mathematics of Operations Research, 10 (1985), pp. 180–184.

[17] J.-J. Kao and T.-I. Lin, *Shortest service location model for planning waste pickup locations*, Journal of the Air & Waste Management Association, 52 (2002), pp. 585–592.

[18] O. Kariv and S. L. Hakimi, *An algorithmic approach to network location problems. I: The p-centers*, SIAM Journal on Applied Mathematics, 37 (1979), pp. 513–538.

[19] J. Könemann, Y. Li, O. Parekh, and A. Sinha, *An approximation algorithm for the edge-dilation k-center problem*, Operations Research Letters, 32 (2004), pp. 491–495.

[20] J. Lamperski, O. A. Prokopyev, and L. G. Wrabetz, *Min-max-min optimization with smooth and strongly convex objectives*, SIAM Journal on Optimization, 33 (2023), pp. 2435–2456.

[21] J. Lim, E. Claypool, B. Norman, and R. J., *Coverage models to determine outreach vaccination center locations in low and middle income countries*, Operations Research for Health Care, (2016).

[22] A. D. López-Sánchez, J. Sánchez-Oro, and A. G. Hernández-Díaz, *GRASP and VNS for solving the p-next center problem*, Computers & Operations Research, 104 (2019), pp. 295–303.

[23] J. L. Moura, B. Alonso, Á. Ibeas, and F. J. Ruisánchez, *A two-stage urban bus stop location model*, Networks and Spatial Economics, 12 (2012), pp. 403–420.

[24] A. T. Murray, *A coverage model for improving public transit system accessibility and expanding access*, Annals of Operations Research, 123 (2003), pp. 143–156.

[25] C. Ortiz-Astorquiza, I. Contreras, and G. Laporte, *Multi-level facility location problems*, European Journal of Operational Research, 267 (2018), pp. 791–805.

[26] S. H. Owen and M. S. Daskin, *Strategic facility location: A review*, European journal of operational research, 111 (1998), pp. 423–447.

[27] R. Panigrahy and S. Vishwanathan, *An O(log*n) approximation algorithm for the asymmetric p-center problem*, Journal of Algorithms, 27 (1998), pp. 259–268.

[28] J. Park and B.-I. Kim, *The school bus routing problem: A review*, European Journal of operational research, 202 (2010), pp. 311–319.

[29] L. Parrot, J. Sotamenou, and B. K. Dia, *Municipal solid waste management in africa: Strategies and livelihoods in yaoundé, cameroon*, Waste management, 29 (2009), pp. 986–995.

[30] D. Purkayastha, M. Majumder, and S. Chakrabarti, *Collection and recycle bin location-allocation problem in solid waste management: A review*, Pollution, 1 (2015), pp. 175–191.

[31] S. Rahman and D. Smith, *Use of location-allocation models in health service development planning in developing nations*, European Journal of Operational Research, 123 (2000), pp. 437–452.

[32] A. Rais and A. Viana, *Operations research in healthcare: A survey*, International Transactions in Operational Research, 18 (2011), pp. 1–31.

[33] Ghana Open Data Initiative, *Health facilities*. `https://data.gov.gh/dataset/health-facilities`. Accessed: 08-20-2024.

[34] C. S. ReVelle and H. A. Eiselt, *Location analysis: A synthesis and survey*, European Journal of Operational Research, 165 (2005), pp. 1–19.

[35] C. S. Revelle, H. A. Eiselt, and M. S. Daskin, *A bibliography for some fundamental problem categories in discrete location science*, European Journal of Operational Research, 184 (2008), pp. 817–848.

[36] G. Şahin and H. Süral, *A review of hierarchical facility location models*, Computers & Operations Research, 34 (2007), pp. 2310–2331.

[37] A. A. Saka, *Model for determining optimum bus-stop spacing in urban areas*, Journal of Transportation Engineering, 127 (2001), pp. 195–199.

[38] D. B. Shmoys, *Computing near-optimal solutions to combinatorial optimization problems*, in Combinatorial Optimization, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, AMS Publications, 1995, pp. 355–397.

[39] J. Vandelaer, J. Bilous, and D. Nshimirimana, *Reaching every district (RED) approach: A way to improve immunization performance*, Bulletin of the World Health Organization, 86 (2008), pp. A–B.

[40] S. Vishwanathan, *An O(log* n) approximation algorithm for the asymmetric p-center problem*, in Proceedings of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 1996, pp. 1–5.

[41] R. Wei, X. Liu, Y. Mu, L. Wang, A. Golub, and S. Farber, *Evaluating public transit services for operational efficiency and access equity*, Journal of Transport Geography, 65 (2017), pp. 70–79.

[42] World Health Organization and others, *In-depth evaluation of the reaching every district approach in the African region.* `https://cdn.who.int/media/docs/default-source/immunization/reaching-every-district-red/1-afro-1-red-evaluation-report-2007-final.pdf`, 2007. Online; accessed June 13, 2023.

[43] Y. Yang and J. Rajgopal, *Learning combined set covering and traveling salesman problem*, preprint, arXiv:2007.03203, (2020).

[44] Y. Yang and J. Rajgopal, *Outreach strategies for vaccine distribution: A multi-period stochastic modeling approach*, Operations Research Forum, 2 (2021).

[45] H. Zia and V. Devadas, *Urban solid waste management in kanpur: Opportunities and perspectives*, Habitat international, 32 (2008), pp. 58–73.

A Omitted proofs

A.1 Proof of Proposition 3.1

Proof of Proposition 3.1. First we show that

$$\min\{c_{ij}, w_{ij}\} \leq 2d_{ij}, \qquad i, j \in [n]. \tag{14}$$

For each $\ell \in [k]$, it follows from (5) that

$$c_{ij} \leq \min\{c_{\ell j} + w_{i\ell}, c_{\ell i} + w_{j\ell}\} \leq c_{\ell j} + w_{i\ell} \leq 2\max\{c_{\ell j}, w_{i\ell}\},$$

so $\min_{\ell\in[n]\setminus\{i,j\}} \max\{c_{\ell j}, w_{i\ell}\} \geq \frac{1}{2}c_{ij}$, and hence

$$d_{ij} = \min\left\{c_{ij}, w_{ij}, \min_{\ell\in[n]\setminus\{i,j\}} \max\{c_{\ell j}, w_{i\ell}\}\right\} \geq \min\left\{c_{ij}, w_{ij}, \frac{1}{2}c_{ij}\right\} \geq \frac{1}{2}\min\{c_{ij}, w_{ij}\}.$$

Multiplying the above inequality by 2 gives inequality (14).

Regarding the first statement of the proposition,

$$d_{ij} = \min\{c_{ij}, w_{ij}, r_{ij}\} \leq \min\{c_{ij}, w_{ij}\} = \min\{c_{ji}, w_{ji}\} \leq 2d_{ji},$$

where the second equality follows from (3) and the second inequality from (14). Moving on to the second statement of the proposition,

$$\begin{aligned}
d_{ij} &= \min\{c_{ij}, w_{ij}, r_{ij}\} \\
&\leq \min\{c_{ij}, w_{ij}\} \\
&\leq \min\{c_{ik} + c_{kj}, c_{kj} + w_{ik}, c_{ki} + w_{jk}, w_{ik} + w_{kj}\} \\
&= \min\{c_{ik} + c_{kj}, w_{ik} + c_{kj}, c_{ik} + w_{kj}, w_{ik} + w_{kj}\} \\
&\leq \min\{c_{ik}, w_{ik}\} + \min\{c_{kj}, w_{kj}\} \\
&\leq 2(d_{ik} + d_{kj}),
\end{aligned}$$

where the second inequality follows from (4) as well as 5, the second equality from (3), and the last inequality from (14). $\square$

A.2 Proof of Lemma 3.1

Proof of Lemma 3.1. Suppose that (M^*, R^*) is optimal for (9); that is, $f(F, M^*, R^*) = \mathrm{OPT}_F$. Taking $j \in [n] \setminus F = M \cup R$, it is sufficient to show that $\min_{f\in F} d_{jf} \leq \mathrm{OPT}_F$. We consider two cases:

Case 1. Suppose that $j \in M^*$. Then

$$\begin{aligned}
\min_{f\in F} d_{jf} = \min_{f\in F} \min\{c_{jf}, w_{jf}, r_{jf}\} \leq \min_{f\in F} c_{jf} \leq \max_{i\in M^*} \min_{f\in F} c_{if} \\
= r_1(F, M^*) \leq \max\{r_1(F, M^*), r_2(F, M^*, R^*)\} = \mathrm{OPT}_F.
\end{aligned}$$

Case 2. Suppose that $j \in R^*$. Let $\ell \in \mathrm{argmin}_{k\in F\cup M^*} w_{jk}$. We consider two subcases:

Case 2.1. Suppose that $\ell \in F$. Then

$$\begin{aligned}\min_{f\in F} d_{jf} &= \min_{f\in F}\min\{c_{jf}, w_{jf}, r_{jf}\} \le \min_{f\in F} w_{jf} = w_{j\ell} = \min_{k\in F\cup M^*} w_{jk} \le \max_{i\in R^*}\min_{k\in F\cup M^*} w_{ik}\\ &= r_2(F, M^*, R^*) \le \max\{r_1(F, M^*), r_2(F, M^*, R^*)\} = \mathrm{OPT}_F.\end{aligned}$$

Case 2.2. Suppose that $\ell \in M^*$. Then we have that

$$\begin{aligned}\min_{f\in F} d_{jf} &= \min_{f\in F}\min\{c_{jf}, w_{jf}, r_{jf}\} \le \min_{f\in F} r_{jf} = \min_{f\in F}\min_{k\in[n]\setminus\{j,f\}}\max\{c_{kf}, w_{jk}\}\\ &\le \min_{f\in F}\max\{c_{\ell f}, w_{j\ell}\} = \max\left\{\min_{f\in F} c_{\ell f}, w_{j\ell}\right\} \le \max\left\{\max_{i\in M^*}\min_{f\in F} c_{if}, \max_{i\in R^*}\min_{k\in F\cup M^*} w_{ik}\right\}\\ &= \max\left\{r_1(F, M^*), r_2(F, M^*, R^*)\right\} = \mathrm{OPT}_F,\end{aligned}$$

where the last inequality follows from $\ell \in M^*$, $j \in R^*$, and $\ell \in \mathrm{argmin}_{k\in f\cup M^*} w_{jk}$. □

A.3 Proof of Lemma 3.2

Proof of Lemma 3.2. We must show that

$$f(F, M, R) = \max\{r_1(F, M), r_2(F, R)\} \le \max_{i\in[n]\setminus F}\min_{f\in F} d_{if}. \tag{15}$$

First we show that

$$r_1(F, M) \le \max_{i\in[n]\setminus F}\min_{f\in F} d_{if}. \tag{16}$$

Let $j \in M$ and $f_j \in \mathrm{argmin}_{f\in F} d_{jf}$. Algorithm 1 either adds j to M in Step 7 or in Step 11. In the former case, since $f_j \in F$,

$$\min_{f\in F} c_{jf} \le c_{jf_j} = d_{jf_j} = \min_{f\in F} d_{jf}, \tag{17}$$

where the first equality follows from Step 6. Now let us consider the latter case. From Steps 8-13, there is an index $i \in [n] \setminus F$ such that for $f_i \in \mathrm{argmin}_{f\in F} d_{if}$, it holds that $d_{if_i} = r_{if_i}$ and $j \in \mathrm{argmin}_{k\in[n]\setminus\{i,f_i\}}\max\{c_{kf_i}, w_{ik}\}$. Because $f_i \in F$,

$$\min_{f\in F} c_{jf} \le c_{jf_i} \le \max\{c_{jf_i}, w_{ij}\} = r_{if_i} = d_{if_i} = \min_{f\in F} d_{if}, \tag{18}$$

where the first equality follows from the definition of r_{if_i}. So, from (17) and (18), we see that $r_1(F, M) = \max_{i\in M}\min_{f\in F} c_{if} \le \max_{i\in M}\min_{f\in F} d_{if}$, and consequently (16) holds.

Next we show that

$$r_2(F, M, R) \le \max_{i\in[n]\setminus F}\min_{f\in F} d_{if}. \tag{19}$$

Let $j \in R$ and $f_j \in \mathrm{argmin}_{f\in F} d_{jf}$. It follows from Steps 4-15 that either $d_{jf_j} = w_{jf_j}$ or $d_{jf_j} = r_{jf_j}$. In the former case, since $f_j \in F$,

$$\min_{k\in F\cup M} w_{jk} \le w_{jf_j} = d_{jf_j} = \min_{f\in F} d_{jf}. \tag{20}$$

Now let us consider the latter case, namely $d_{jf_j} = r_{jf_j}$. From Steps 8-13, there is an index $i \in [n] \setminus F$ such that $i \in \text{argmin}_{k \in [n] \setminus \{j, f_j\}} \max\{c_{kf_j}, w_{jk}\}$, and by Steps 10-12 specifically, either $i \in F$ or $i \in M$. Hence, $i \in F \cup M$, and

$$\min_{k \in F \cup M} w_{jk} \leq w_{ji} \leq \max\{c_{jf_j}, w_{ji}\} = r_{jf_j} = d_{jf_j} = \min_{f \in f} d_{jf}. \tag{21}$$

So, from (20) and (21), it holds that

$$r_2(F, M, R) = \max_{i \in R} \min_{j \in F \cup R} w_{ij} \leq \max_{i \in [n] \setminus F} \min_{f \in F} d_{if},$$

implying that (19) holds.

The desired result (15) follows from (16) and (19). □

A.4 Proof of Lemma 4.1

Proof of Lemma 4.1. If $i = j$, then the desired result follows immediately because then $d_{ji} = d_{jj} = 0$, so assume that $i \neq j$. Either $i = f$, $i \in M_f^*$, or $i \in R_f^*$; we consider these three cases below.

Case 1. Suppose that $i = f$. If $j \in M_f^*$, then

$$d_{ji} = d_{jf} \leq c_{jf} \leq \text{OPT}.$$

If $j \in R_f^*$, then it is either optimal for client j to retrieve from client f's location or the location of some client $k \in M_f^*$. In the former case,

$$d_{ji} = d_{jf} \leq w_{jf} \leq \text{OPT},$$

and in the latter case,

$$d_{ji} \leq r_{ji} \leq \max\{c_{ki}, w_{jk}\} \leq \text{OPT}.$$

Case 2. Suppose that $i \in M_f^*$. Either (*i*) $j = f$, (*ii*) $j \in M_f^*$, or (*iii*) $j \in R_f^*$; we consider these three cases below as subcases.

Case 2.1. Suppose that $j = f$. Then

$$d_{ji} = d_{fi} \leq c_{fi} = c_{if} \leq \text{OPT},$$

where the last inequality follows from the fact that $i \in M_f^*$.

Case 2.2. Suppose that $j \in M_f^*$. Then

$$d_{ji} \leq c_{ji} \leq c_{jf} + c_{fi} = c_{jf} + c_{if} \leq 2\text{OPT},$$

where the last inequality follows from the fact that $i, j \in M_f^*$.

Case 2.3. Suppose that $j \in R_f^*$. It is either optimal for client j to retrieve from the location of client f, client i, or some client $k \neq i \in M_f^*$. In the first case, we see that

$$d_{ji} \leq r_{ji} \leq \max\{c_{fi}, w_{jf}\} = \max\{c_{if}, w_{jf}\} \leq \text{OPT},$$

where the last inequality follows from $i \in M_f^*$ and the assumption that it is optimal for client j to retrieve from location of client f. In the second case, it holds that

$$d_{ji} \leq w_{ji} \leq \mathrm{OPT}.$$

And in the last case, we have that

$$\begin{aligned}d_{ji} &\leq r_{ji} \leq \max\{c_{ki}, w_{jk}\} \leq \max\{c_{kf} + c_{fi}, w_{jk}\} = \max\{c_{kf} + c_{if}, w_{jk}\} \\ &\leq \max\{2\mathrm{OPT}, \mathrm{OPT}\} = 2\mathrm{OPT},\end{aligned}$$

where the last inequality follows from the fact that $i, k \in M_f^*$ and the assumption that it is optimal for client j to retrieve from location of client k.

Case 3. Suppose that $i \in R_f^*$. Either (i) $j = f$, (ii) $j \in M_f^*$, or (iii) $j \in R_f^*$; we consider these three cases below.

Case 3.1. Suppose that $j = f$. It is either optimal for client i to retrieve from the location of client $j = f$ or some client $k \in M_f^*$. In the former case,

$$d_{ji} = d_{fi} \leq w_{fi} = w_{if} \leq \mathrm{OPT},$$

and in the latter case,

$$d_{ji} = d_{fi} \leq c_{fi} \leq c_{kf} + w_{ik} \leq 2\mathrm{OPT},$$

where the last inequality follows from the fact that $k \in M_f^*$ and it is optimal for client i to retrieve from the location of client k.

Case 3.2. Suppose that $j \in M_f^*$. It is either optimal for client i to retrieve from the location of client f, client j, or some client $k \neq j \in M_f^*$. In the first case, we have that

$$d_{ji} \leq c_{ji} \leq c_{fj} + w_{if} = c_{jf} + w_{if} \leq 2\mathrm{OPT},$$

where the last inequality follows from the fact $j \in M_f^*$ and it is optimal for client i to retrieve from the location of client f. In the second case, we see that

$$d_{ji} \leq w_{ji} = w_{ij} \leq \mathrm{OPT}.$$

And in the final case, we obtain a different approximation, depending on the cost regime. In the related metric costs regime,

$$d_{ji} \leq c_{ji} \leq c_{jf} + c_{fi} \leq c_{jf} + c_{kf} + w_{ik} \leq 3\mathrm{OPT},$$

where the last inequality follows from the fact that $j, k \in M_f^*$ and it is optimal for client i to retrieve from location of client k. And in the equal costs regime,

$$\begin{aligned}d_{ji} &\leq r_{ji} \leq \max\{c_{fi}, c_{jf}\} \leq \max\{c_{fk} + c_{ki}, c_{jf}\} = \max\{c_{kf} + c_{ik}, c_{jf}\} \\ &\leq \max\{2\mathrm{OPT}, \mathrm{OPT}\} = 2\mathrm{OPT},\end{aligned}$$

where the last inequality follows from the fact that $j, k \in M_f^*$ and it is optimal for client i to retrieve from the location of client k.

Case 3.3. Suppose that $j \in R_f^*$. It is either optimal for client i to retrieve from the location of client f or some client $k \in M_f^*$. First let us consider the case in which it is optimal for location i to retrieve from client f's location. In this case, it is either optimal for client j to retrieve from the location of client f as well or some client $\ell \in M_f^*$. In the former case,

$$d_{ji} \leq w_{ji} \leq w_{jf} + w_{fi} = w_{jf} + w_{if} \leq 2\mathrm{OPT},$$

and in the latter case,

$$\begin{aligned} d_{ji} \leq r_{ji} \leq \max\{c_{\ell i}, w_{j\ell}\} &\leq \max\{c_{\ell f} + w_{fi}, w_{j\ell}\} = \max\{c_{\ell f} + w_{if}, w_{j\ell}\} \\ &\leq \max\{2\mathrm{OPT}, \mathrm{OPT}\} = 2\mathrm{OPT}, \end{aligned}$$

where the last inequality follows from the fact that $\ell \in M_f^*$ and it is optimal for client i and j to retrieve from location of client f and ℓ, respectively.

Next we consider the case in which it is optimal for client i to retrieve from the location of client $k \in M_f^*$. It is either optimal for client j to retrieve from the location of client k or some client $\ell \in M_f^*$. In the former case,

$$d_{ji} \leq w_{ji} \leq w_{jk} + w_{ki} = w_{jk} + w_{ik} \leq 2\mathrm{OPT}.$$

Now let us consider the latter case in which it is optimal for client j to retrieve from the location of client ℓ. We obtain a different approximation guarantee, depending on the cost regime. In the related metric costs regime,

$$\begin{aligned} d_{ji} \leq r_{ji} \leq \max\{c_{\ell i}, w_{j\ell}\} &\leq \max\{c_{\ell f} + c_{fi}, w_{j\ell}\} \leq \max\{c_{f\ell} + c_{kf} + w_{ik}, w_{j\ell}\} \\ &\leq \max\{3\mathrm{OPT}, \mathrm{OPT}\} = 3\mathrm{OPT}, \end{aligned}$$

where the last inequality follows from the fact that $k, \ell \in M_f^*$ and that it is optimal for client i and j to retrieve from the location of client k and ℓ, respectively. Now, in the equal cost regime,

$$\begin{aligned} d_{ji} \leq r_{ji} \leq \max\{c_{fi}, c_{jf}\} &\leq \max\{c_{fk} + c_{ki}, c_{j\ell} + c_{\ell f}\} = \max\{c_{kf} + c_{ik}, c_{j\ell} + c_{\ell f}\} \\ &\leq \max\{2\mathrm{OPT}, 2\mathrm{OPT}\} = 2\mathrm{OPT}, \end{aligned}$$

where the last inequality follows from the fact that $k, \ell \in M_f^*$ and that it is optimal for clients i and j to retrieve from the locations of clients k and ℓ, respectively. □

A.5 Proof of Proposition 4.2

To establish Proposition 4.2, we consider simple instances of (2) (instead of an infinite family of instances) to provide clear insights into the properties that make an instance challenging for Algorithm 2.

Proof of Proposition 4.2. First we consider the related metric costs regime case. Consider the meet-in-the-middle 1-center instance with the following agent and client travel costs, respectively.

$$c = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{bmatrix} \quad \text{and} \quad w = \begin{bmatrix} 0 & 1 & 2 & 5 \\ 1 & 0 & 1 & 4 \\ 2 & 1 & 0 & 3 \\ 5 & 4 & 3 & 0 \end{bmatrix}.$$

It is readily verified that the costs individually define metrics (i.e., conditions (3) and (4) hold). Furthermore, condition (5) holds because condition (4) holds and $c \leq w$. Thus, the assumptions of the related metric costs regime are satisfied.

Suppose that we initialize Algorithm 2 in Step 3 with location $i = 1$. Because $k = 1$, the algorithm then proceeds to Step 8. It is straightforward to verify that, in Step 8, Algorithm 1 returns receiving locations $M = \{2, 4\}$ and retrieving locations $R = \{3\}$. Thus Algorithm 2 returns the solution $F = \{1\}$, $M = \{2, 4\}$, $R = \{3\}$. Because $c_{21} = 1$, $r_{31} = 1$, and $c_{41} = 3$, it holds that the objective value of this solution equals 3. Now consider the solution $F^* = \{3\}$, $M = \{2, 4\}$, and $R = \{1\}$. Because $c_{23} = 1$, $c_{43} = 1$, and $r_{13} = 1$, it holds that the objective value of this is solution is 1. Therefore, the approximation ratio of the solution of Algorithm 2 is 3.

Now we consider the equal metric costs regime case. Consider a meet-in-the-middle 2-center instance with

$$c_{ij} = w_{ij} = |i - j|, \quad i, j \in \{1, \ldots, 7\}.$$

Clearly the assumptions of the equal metric cost regime are satisfied. Suppose that we initialize Algorithm 2 in Step 3 with $i = 1$. In Steps 4-6, the algorithm either adds index 6 to F or index 7 because $d_{21} = 1$, $d_{31} = 1$, $d_{41} = 2$, $d_{51} = 2$, $d_{61} = 3$, and $d_{71} = 3$. Suppose that the algorithm adds index 7 to F. It is straightforward to verify that, in Step 8, Algorithm 1 can return receiving locations $M = \{2, 3, 5, 6\}$ and retrieving locations $R = \{4\}$. Thus Algorithm 2 can return the solution $F = \{1, 7\}$, $M = \{2, 3, 5, 6\}$, and $R = \{4\}$. The objective value of this solution equals 2. Now consider the solution $F^* = \{3, 5\}$, $M^* = \{2, 4, 6\}$, and $R^* = \{1, 7\}$. The objective value of this solution is 1. Therefore, the approximation ratio of the solution of Algorithm 2 is 2. $\square$