Business-to-Consumer E-Commerce: Home Delivery in Megacities

Yixiao Huang[a], Martin Savelsbergh[b], Lei Zhao[a]

[a]*Department of Industrial Engineering, Tsinghua University, Beijing, 100084, China*
[b]*H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, 30332*

Abstract

To deliver to consumers in densely populated urban areas, companies often employ a two-echelon logistics system. In a two-echelon logistics system, the entry point for goods to be delivered in the urban area is a city distribution center (CDC). From a CDC the goods are transported to an intermediate facility, from where the goods are delivered to the consumers. By restricting the set of potential vehicle routes employed in one or both of the echelons, it is possible to significantly reduce the complexity of the delivery operations, which is a common practice in real-life environments. We study the impact on delivery costs of such strategies and demonstrate that when the number of orders to be delivered is large and the location density of delivery addresses is high, such strategies can have near-optimal performance. To more easily accommodate delivery volume growth and to more effectively handle daily delivery volume variations, we introduce a simple aggregation concept, which leads to further quality improvements. We provide further insight by means of a worst-case analysis for a specific geographic topology.

Keywords: megacities, business-to-consumer e-commerce, city logistics, two-echelon vehicle routing

1. Introduction

In the United Nations report "World Urbanization Prospects – The 2014 Revision" (https://esa.un.org/unpd/wup/), it is observed that, at the time of the writing of the report, 54% of the world's population is living in urban areas and that that proportion will continue to increase. The authors of the report also predict a continuing growth in the number of megacities, i.e., cities with a population of more than 10 million, from 28 in 2014 to 41 in 2030. See Figure 1 for urbanization and megacities trends – reproduced from the report.

At the same time, business-to-consumer e-commerce is booming worldwide. For example, in Beijing, a daily average of 327.2 items per square kilometer were delivered to consumers in 2016 (source: State Post Bureau of P.R. China, http://www.spb.gov.cn/xw/dtxx_15079/201701/t20170114_959038.html, in Chinese, last accessed on 9 May, 2017). It is therefore

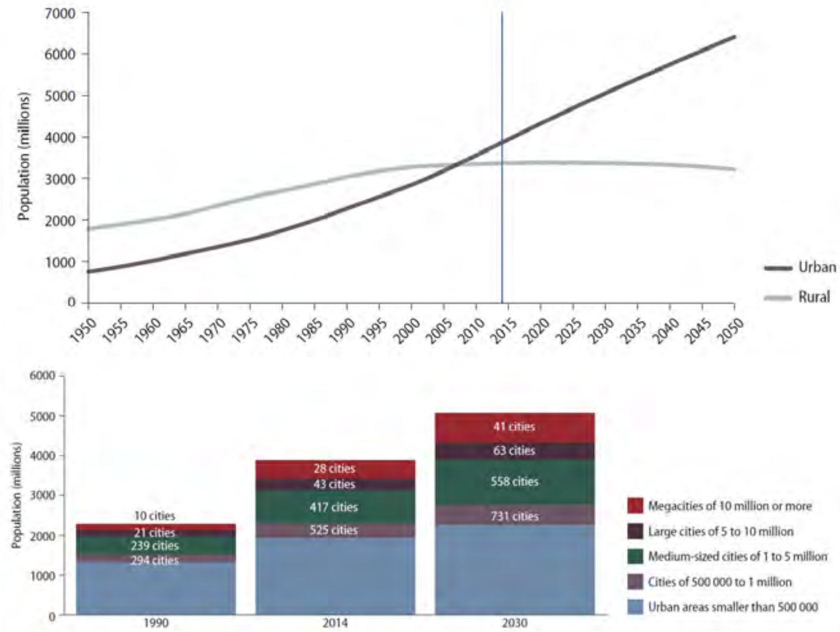


Figure 1: Urbanization and megacities trends.

critical that industry and government find sustainable and cost-effective ways to accommodate the continuing urbanization and the growth of business-to-consumer e-commerce, and to limit any negative impact on the environment and the quality of life of those who live in urban areas.

To efficiently deliver to consumers living in densely populated urban areas, companies often employ a two-echelon logistics system. In a two-echelon logistics system, the entry point for goods to be delivered in the urban area is a city distribution center (CDC). From there the goods are transported to an intermediate facility, from where the goods are delivered to the consumers. Typically, large vehicles are used for the transportation of goods from a CDC to an intermediate facility, and small vehicles are used for the transportation of goods from an intermediate facility to a consumer. More and more, the vehicles used to make last-mile deliveries are environmentally friendly, such as electric cars or electric tricycles. Such two-echelon logistics systems control inventory holding costs through pooling and transportation costs through consolidation (in the first echelon).

The academic community as well has started to study two-echelon delivery systems, which has resulted in a growing, but at the moment still small, body of literature on the so-called two-echelon vehicle routing problem (2E-VRP), e.g., Crainic et al. (2009) and Hemmelmayr et al. (2012), with a focus on developing effective and efficient algorithms. Unfortunately, the size of the instances that can be handled by these algorithms is orders of magnitudes smaller than the sizes encountered in real-life settings (in terms of the number of orders that needs to be delivered and the location density of the delivery addresses).

Our focus is somewhat different. We seek to assess the impact on delivery costs of employing a two-echelon logistics system in which a restricted set of vehicle routes is used in the second echelon. Our study is inspired by the two-echelon logistics system employed by companies providing home delivery services in Beijing (e.g., Amazon China, JD.com, SF Express). More specifically, the two-echelon logistics system investigated has a single CDC, divides the coverage area into regions, each with an intermediate facility, and divides each region into cells, each served by a single driver. The drivers themselves decide the (single) route that they will use to deliver the orders in their cell. Note that once the regions and the cells have been defined, managing the delivery system is straightforward. All that needs to be done is determining how to supply the intermediate facilities, which corresponds to solving a small – since the number of intermediate facilities is usually small – split delivery vehicle routing problem; once the goods to be delivered arrive at an intermediate facility, they are simply handed to the driver responsible for the cells containing the delivery locations. We conjecture that when the number of orders that needs to be delivered is large and the location density of the delivery addresses is high, then this type of two-echelon logistics system has near-optimal performance (as long as the individual drivers use a reasonable route to deliver their assigned orders). The research presented in this paper supports that conjecture.

To more easily accommodate growth, i.e., an increase in the number of daily deliveries, and to more effectively handle daily variations, i.e., day-to-day differences in the number of deliveries in a cell, we propose a delivery strategy using *blocks*. A block is an aggregation of cells to be served by a small number of drivers (up to three). The introduction of blocks adds very little operational complexity as all that needs to be done is dividing the deliveries in a block over the drivers assigned to the block, which can be done, for example, by a simple sweep-like procedure that balances the workload of the drivers assigned to the block.

In summary, our research contributes to the existing literature in the following ways:

- We analyze the performance of urban delivery systems designed for operational simplicity, empirically and analytically, in order to demonstrate that such systems are not only practical, but also highly effective and therefore an excellent choice for home delivery in (mega)cities.

- We conduct an extensive computational study that demonstrates that urban delivery systems designed for operational simplicity have near-optimal performance when the number of orders that needs to be delivered is sufficiently large and the location density of the delivery addresses is sufficiently high.

- We show that the introduction of blocks leads to a significantly more robust urban delivery system capable of accommodating (moderate) growth and cost-effectively handling day-to-day differences in delivery volumes.

- We deliver further insight by providing a worst-case analysis for a specific two-echelon geographic topology.

The remainder of the paper is organized as follows. We review relevant literature in Section 2. We formally define the two-echelon logistics system being studied and the strategies for operating such a system in Section 3. We introduce and investigate the block design (sub)problem in Section 4. We present the results of an extensive computational study comparing different strategies for operating a two-echelon logistics system in Section 5. We present analytical results, in the form of a worst-case analysis, in Section 6. We conclude with final remarks in Section 7.

2. Literature review

We review relevant literature in this section, namely the two-echelon vehicle routing problem (2E-VRP), the territory-based vehicle routing problem, and the territory design problem. Interested readers may refer to Cuda et al. (2015) for a comprehensive review of 2E-VRPs. Note that the territory-based vehicle routing problem and the territory design problem in the literature consider single-echelon logistics systems (with one depot or multiple depots) and have not yet explored multi-echelon logistics systems.

The majority of 2E-VRP studies focus on the algorithmic design. Various integer programming solution techniques are applied to develop exact methods. Baldacci et al. (2013) propose a new formulation for the 2E-VRP. They develop an exact method via decomposing the problem into a limited set of multi-depot vehicle routing problems with side constraints. Jepsen et al. (2013) implement a branch-and-cut algorithm to solve the symmetric 2E-VRP with a special branching scheme, which is able to solve 47 out of a set of 93 instances to optimality. Santos et al. (2015) propose a branch-and-cut-and-price method based on a strengthened reformulation overcoming symmetry issues. Among these exact algorithms, the largest instances solved consist of 100 consumers (Baldacci et al., 2013).

On the heuristic and metaheuristic side, Perboli et al. (2011) propose math-based heuristics to decide the assignment between consumers and intermediate facilities based on the information obtained by solving the linear relaxation of a flow-based model. Hemmelmayr et al. (2012) develop an adaptive large neighborhood search (ALNS) algorithm with a variety of repair and destroy operators. Based on extensive numerical experiments, they show that the ALNS algorithm outperforms other existing algorithms in instances of the 2E-VRP and location-routing problem. Wang et al. (2015) propose a hybrid metaheuristic to decide the assignment between consumers and intermediate facilities. Breunig et al. (2016) propose a hybrid metaheuristic, combining local search with destroy-and-repair principles and tailored operators of selecting intermediate facilities. Among the heuristics and metaheuristics, the largest instances solved consist of 200 consumers (Hemmelmayr et al., 2012; Breunig et al., 2016).

Crainic et al. (2010) explicitly consider the impact of important instance parameters on the cost of 2E-VRP, e.g., consumer distribution, CDC location, number of intermediate facilities. They also compare the 2E-VRP to the (single-echelon) capacitated VRP and conclude that 2E-VRP outperforms the capacitated VRP in most cases, in particular when the CDC is located externally to the coverage area and the intermediate facilities are located between the CDC and the consumers.

There are a few emerging papers considering more practical variants of 2E-VRPs, which are summarized in Table 1. Crainic et al. (2009) comprehensively analyze the city logistics practice and model the city logistics as a time-dependent 2E-VRP with satellite synchronization and time windows. Soysal et al. (2015) take the fuel consumption and time-dependency in the second echelon into consideration and study a practical 2E-VRP inspired by a supermarket chain in the Netherlands. Grangier et al. (2016) consider the multi-trip 2E-VRP with satellite synchronization and time windows. Song et al. (2016) consider a variant of 2E-VRP allowing direct deliveries from CDCs to consumers. Dellaert et al. (2016) propose a branch-and-price method to solve the 2E-VRP with time windows. Li et al. (2016a,b) both consider the 2E-VRPs in linehaul delivery systems, where Li et al. (2016a) further consider the carbon emission in the objective function. Liu et al. (2017) consider the 2E-VRP with stochastic demands and propose a simulation-based tabu search method. Among the above studies, only Li et al. (2016a,b) assume predefined assignments between consumers and intermediate facilities, which is similar to the setting of regional division in this paper.

Table 1: Characteristics of papers studying 2E-VRP variants.

	TD	TW	MWT	SS	MT	FC	DD	SD
Crainic et al. (2009)	✓	✓		✓				
Soysal et al. (2015)	✓					✓		
Grangier et al. (2016)		✓		✓	✓			
Song et al. (2016)							✓	
Dellaert et al. (2016)		✓						
Li et al. (2016a)		✓	✓	✓		✓		
Li et al. (2016b)		✓	✓	✓				
Liu et al. (2017)								✓

TD: Time-dependency. MWT: Maximum working time. TW: Time windows. SS: Satellite synchronization.
MT: Multi-trip. FC: Fuel consumption. DD: Directly delivery. SD: Stochastic demands.

The routing based on fixed area is called the territory-based VRP in the literature, usually in a single-depot setting. Wong and Beasley (1984) consider to partition the service area into subareas, each assigned with one vehicle. They propose a heuristic algorithm based on first constructing an initial customer allocation and then improving it via interchange operations. Zhong et al. (2007) consider a different problem. Territories are classified to the core area, which is served by the same driver every day, and the flex zone, where the customers can be reassigned daily. Additionally, they consider the learning curve of each driver to characterize drivers' familiarities of each area. They use a tabu search method to strategically design the territory. Following the setting of core areas and flex zones in Zhong et al. (2007), Schneider et al. (2015) incorporate the time window and propose a modular territory routing approach.

There is another stream of literature focusing on the territory design, which tactically decides service regions for multiple depots, usually based on the measure of the expected routing distance. The expected routing distance is approximated in order to avoid solving complex optimization problems, where readers may refer to Langevin et al. (1996) for a review of approximation methods in freight transportation. Carlsson (2012) consider the territory planning for the vehicle routing problem with stochastic customer demand. Carlsson and Delage (2013) consider the robust territory partition to balance the workload of vehicles among the regions. Lei et al. (2015) consider a joint dynamic districting and multiple traveling salesmen problem with multi-periods. The studies above approximate the expected routing distance via Beardwood-Halton-Hammersley Theorem (Beardwood et al., 1959). Alternatively, there are papers approximating the expected routing distance via Continuous Approximation (Daganzo, 1984). Ouyang (2007) proposes a method that partitions an area into vehicle routing zones based on continuous approximation. Franceschetti et al. (2017) consider the region partition and fleet planning simultaneously for city logistics. Qi et al. (2016) design the service regions of a shared mobility system for last-mile delivery.

3. A two-echelon city logistics system

The two-echelon city logistics system considered serves consumers in an urban area and can be defined on a directed graph $G = (N, A)$, where N is the set of nodes and A is the set of directed arcs. The node set N consists of three subsets: the (single-node) set N_0 representing a city distribution center (CDC) located at the boundary of the urban area, set N_S representing a set of intermediate facilities strategically located in the urban area, and set N_C representing a set of consumers that have to be served and whose home locations are scattered across the urban area. Each consumer $i \in N_C$ requires a delivery of size q_i. The arc set A consists of two subsets: the first-echelon arc set $A_1 = \{(N_0 \cup N_S) \times (N_0 \cup N_S)\}$ representing links connecting the CDC with the intermediate facilities and connecting the intermediate facilities themselves, and the second-echelon arc set $A_2 = \{(N_S \cup N_C) \times (N_S \cup N_C)\}$ representing the links connecting the intermediate facilities with the consumers and connecting the consumers themselves. Each arc $(i, j) \in A$ has an associated distance d_{ij}. Note that direct delivery from the CDC to a consumer is not an option.

The logistics company that operates the two-echelon city logistics system has a fleet of K_1 homogeneous vehicles to serve the first echelon, each with capacity Q_1, and a fleet of K_2 homogeneous vehicles to serve the second echelon, each with capacity $Q_2 < Q_1$. Note that deliveries from the CDC to the intermediate facilities can be split, i.e., an intermediate facility may receive deliveries from multiple vehicles. The objective of the company is to minimize the total delivery cost, which is a function of the total distance traveled. We let c_1 denote the unit cost of travel in the first echelon and let c_2 denote the unit cost of travel in the second echelon.

Note that the setting introduced above assumes known deterministic demand. This is representative of situations where a logistics company operates a delivery policy, in which orders placed in a specific period are guaranteed to be delivered in a specific later period. The simplest such policy guarantees that all orders placed today will be delivered tomorrow. A more sophisticated policy may guarantee that all orders placed during an 8-hour window will be delivered during the subsequent 8-hour window. The logistics company can employ different strategies to serve the demand: an *unconstrained delivery strategy*, in which there is complete flexibility in constructing the delivery routes for the vehicles in both echelons, and a *constrained delivery strategy*, in which there is limited flexibility in constructing the delivery routes in one or both of the echelons.

3.1. The unconstrained delivery strategy

The *unconstrained delivery strategy* does not impose any restrictions on the delivery routes in the two echelons. That is, no restrictions are placed on which intermediate facility serves a particular consumer (which implies that it is possible not to use an intermediate facility at all). When the unconstrained delivery strategy is employed, a delivery plan minimizing the total cost is obtained by solving a two-echelon vehicle routing problem.

The two-echelon vehicle routing problem can be viewed as a combination of a split delivery vehicle routing problem (SDVRP) in the first echelon and a multi-depot vehicle routing problem in the second echelon. However, these two problems are coupled because the demand of an intermediate facility is the sum of the consumer demands served from that intermediate facility, which makes the two-echelon vehicle routing problem computationally challenging.

We denote the cost of a delivery plan constructed under the unconstrained delivery strategy as Z^U.

3.2. A constrained delivery strategy

In practice, to avoid the computational challenges associated with the unconstrained delivery strategy and to reduce the complexity of the delivery operation itself, many logistics companies employ a constrained delivery strategy, in which the urban area is partitioned into a set of regions R, each containing one of the intermediate facilities. A constrained delivery strategy using regions imposes that consumers in a region are served from the intermediate facility in that region. We denote the cost of a delivery plan constructed under a constrained delivery strategy using regions as Z^R. Clearly, we have $Z^R \geq Z^U$.

Note that when employing a constrained delivery strategy, the delivery problem is decomposed into a SDVRP in the first echelon, and several capacitated vehicle routing problems (CVRPs), one for each region, in the second echelon. Note too that the SDVRP and CVRPs are decoupled due to the predefined assignment of consumers to intermediate facilities, which makes obtaining a minimum cost delivery plan much easier.

To reduce the complexity of the delivery operation even further, each region $r \in R$ is partitioned in a set of cells C_r. All consumers in a cell $c \in C_r$ are served by one driver, starting the delivery route from the intermediate facility in the region. Note that this implies that rather than solving a CVRP, a number of traveling salesman problems (TSPs) have to be solved. In practice, rather than solving a TSP for the consumers assigned to a driver, the driver is allowed to serve his assigned consumers in the order that he believes is the best, based on his intimate knowledge of the small area of the city he is responsible for.

We denote the cost of a delivery plan constructed under a constrained delivery strategy using cells as Z^C. Clearly, we have $Z^C \geq Z^R \geq Z^U$.

3.3. Blocks

Rather than using cells, we propose to use blocks, where a block is an aggregation of cells. That is, each region $r \in R$ is first partitioned into a set of cells C_r, where a cell represents a geographical area that, on average, can be served by a single driver, and then a block $b \in B_r$ is defined as the union of one or more cells $c \in C_r$, where blocks satisfy a "connectivity" condition, e.g., for each cell in a block with more than one cell, there is at least one adjacent cell in the block.

Introducing blocks creates flexibility to deal with growth in demand as well as with variability in demand. When the average demand in a cell is less than the capacity of a vehicle, assigning a vehicle to multiple cells may result in a cost reduction. Similarly, when the average demand in a cell is more than the capacity of a vehicle, assigning a few cells to a small number of vehicles (fewer cells than vehicles) may avoid infeasibility and may result in a cost reduction. This flexibility is especially important when new markets (new geographical areas) are opened and there may be rapid growth, and when demand shows significant day-to-day variability.

We denote the cost of a delivery plan constructed under a constrained delivery strategy using blocks as Z^B. Clearly, we have $Z^C \geq Z^B \geq Z^R \geq Z^U$.

On a day-to-day basis, the consumers in a block have to be partitioned among the drivers. For the remainder, we assume that a simple sweep heuristic is used to assign consumers to drivers in a block. That is, we set the intermediate facility of the region as the origin. We sort the consumers in a block in a non-descending order of their angles (i.e., locations are represented in a polar coordinate system). Starting with the consumer with the smallest angle, we assign consumers to a vehicle until the capacity is reached, and repeat.

4. The block design problem

We call two cells i and j adjacent if they share a common boundary segment, not just a point (Brun, 2002). With each region $r \in R$, we associate a graph $G_r = \{N_r, E_r\}$ with node set N_r representing each cell $c \in C_r$ and arc set E_r representing pairs of cells $c, c' \in C_r$ that are adjacent. Let $\bar{Q}$ be the target block capacity. The block design problem for a region $r \in R$ seeks to minimize the number of blocks required to cover the total demand in region r such that the total demand in a block does not exceed the target block capacity and a block consists of a set of pairwise adjacent cells. Introducing the following decision variables

$$y_b = \begin{cases} 1, & \text{if cell } b \text{ represents block } b, \\ 0, & \text{otherwise,} \end{cases}$$

$$x_{ib} = \begin{cases} 1, & \text{if cell } i \text{ belongs to block } b, \\ 0, & \text{otherwise,} \end{cases}$$

the generic block design problem can be formulated as follows:

$$
\begin{aligned}
\min_{x_{ib}, y_b} \quad & \sum_{b \in N_r} y_b && \text{(1a)}\\
\text{s.t.} \quad & \sum_{i \in N_r} d_i x_{ib} \le \bar{Q} y_b, \ \forall b \in N_r, && \text{(1b)}\\
& \sum_{b \in N_r} x_{ib} = 1, \ \forall i \in N_r, && \text{(1c)}\\
& x_{bb} = y_b, \ \forall b \in N_r, && \text{(1d)}\\
& x_{ib} = 1 \Rightarrow \text{cell } i \text{ and cell } b \text{ are connected}, && \text{(1e)}\\
& y_b \in \{0,1\}, \ \forall b \in N_r,\\
& x_{ib} \in \{0,1\}, \ \forall i, b \in N_r.
\end{aligned}
$$

This is a generalized assignment problem with additional connectivity constraints (1e). Constraints (1b) ensure that the total demand of a block does not exceed the target block capacity $\bar{Q}$. Constraints (1c) guarantee that each cell belongs to a block and constraints (1d) enforce that the cell chosen to represent block b belongs to block b. Constraints (1e) are to model the connectivity requirements and will be discussed in detail below. The objective function (1a) is to minimize the number of blocks.

We consider two formulations of the connectivity constraints (1e), both inspired by multi-commodity flow concepts.

The first formulation requires that for each cell i assigned to block b there is a unit flow in G_r from cell b to cell i, where b is the cell representing block b. Therefore, we introduce the flow variables of f_{jk}^{ib} representing a flow from cell b to cell i ($i \neq b$), for any arc $(j,k) \in E_r$. Then, constraints (1e) can be formulated as follows:

$$
\begin{aligned}
\sum_{(j,k) \in E_r} f_{jk}^{ib} - \sum_{(k,j) \in E_r} f_{kj}^{ib} &= \begin{cases} x_{ib}, & j = b, \\ -x_{ib}, & j = i, \\ 0, & \text{otherwise}, \end{cases} && \forall j, i, b \in N_r, i \neq b, && \text{(2a)}\\
f_{jk}^{ib} &\le x_{ib}, && \forall (j,k) \in E_r, \forall i, b \in N_r, i \neq b, && \text{(2b)}\\
f_{jk}^{ib} &\le x_{jb}, && \forall (j,k) \in E_r, \forall i, b \in N_r, i \neq b, && \text{(2c)}\\
f_{jk}^{ib} &\le x_{kb}, && \forall (j,k) \in E_r, \forall i, b \in N_r, i \neq b, && \text{(2d)}\\
f_{jk}^{ib} &\ge 0, && \forall (j,k) \in E_r, \forall i, b \in N_r, i \neq b.
\end{aligned}
$$

Constraints (2a) ensure flow conservation. Constraints (2b)–(2d) enforce that a flow only traverses cells in the same block.

Because enforcing connectivity in this way introduces a huge number of additional continuous variables ($|E_r| \cdot |N_r|^2$), we introduce the following variation. Consider a block b and all the cells that belong to block b, we will enforce that there is a flow from cell b to all other cells and that each of these other cells "absorbs" a unit flow. In this case, the connectivity is also guaranteed. Therefore, we introduce the flow variables of f_{jk}^{b} representing the flow in block b, for any arc $(j,k) \in E_r$. Notice that we have $f^b_{jk} = \sum_{i \in C_r, i \neq b} f^{ib}_{jk}$. Then, constraints (1e) can be formulated as follows:

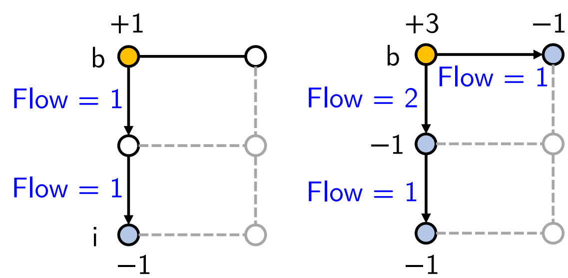
$$\sum_{(j,k)\in E_r} f^b_{jk} - \sum_{(k,j)\in E_r} f^b_{kj} = \begin{cases} \sum_{i\neq b} x_{ib}, & j = b, \\ -x_{jb}, & \text{otherwise.} \end{cases} \qquad \forall j, b \in N_r, \tag{3a}$$

$$f^b_{jk} \leq (|N_r| - 1)x_{jb}, \qquad \forall (j,k) \in E_r,\ b \in N_r, \tag{3b}$$

$$f^b_{jk} \leq (|N_r| - 1)x_{kb}, \qquad \forall (j,k) \in E_r,\ b \in N_r, \tag{3c}$$

$$f^b_{jk} \geq 0, \qquad \forall (j,k) \in E_r,\ b \in N_r.$$

Constraints (3a)–(3c) are similar to constraints (2a)–(2d). The major difference is that flows are aggregated; the outflow at the source (cell b) is the size of block minus 1 (cell b itself is excluded), and the inflow at each other cell of the block is 1. Notice that we can further improve the big-M value (currently set to $|N_r|-1$) in constraints (3b) and (3c), which is equal to the largest possible number of cells in a block, i.e., $M = \max\{|\mathcal{C}| : \mathcal{C} \subset C_r, \sum_{i\in\mathcal{C}} d_i \leq \bar{Q}\}$. In Figure 2, we illustrate how the two network flow formulations model connectivity.



(a) f^{ib}_{jk} (b) f^b_{jk}

Figure 2: Illustration of the two network flow formulations to model connectivity (orange nodes are source nodes and gray nodes are sink nodes).

Because it is likely that there are many optimal solutions to formulation (1), we employ a hierarchical optimization approach, where we first minimize the number of blocks and subsequently minimize a function of the distance from the intermediate facility (denoted as S_r) to the cell representing a block and the distance from the cells of a block to the cell representing the block in order to induce a "desirable" shape of a block, i.e.,

$$\min_{x_{ib}, y_b} \quad \sum_{b\in C_r} d_{S_r b} y_b + \sum_{i,b\in C_r} d_{ib} x_{ib}. \tag{4}$$

Note that this secondary objective (4) can be viewed as an approximation of the distance required by the vehicles serving the block.

5. A computational study

Our computational study seeks to generate insights into the creation and use of block designs as well as into the effects of restricting route choices in the second echelon in order to simplify the delivery operations. We start, however, by describing our experimental setup.

5.1. Experimental setup

We use the urban area of Beijing as the basis of our investigation. Specifically, we use the publicly available information from the Beijing Bureau of Urban Planning to define four regions and 87 cells, see regions 01-04 in Figure 3. We assume that the intermediate facility in a region is located at the region's centroid. Each cell is approximated by a simple polygon. The city distribution center (CDC) is located at the southwest corner of the coverage area. We also assume that capacity can simply be measured in terms of the number of orders, i.e.,

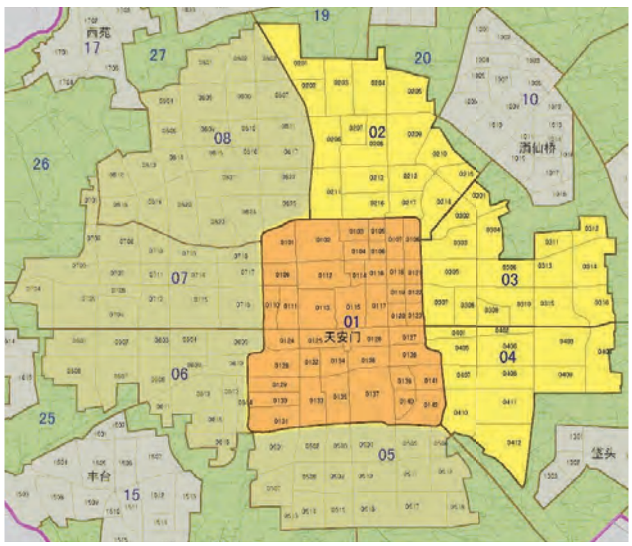


Figure 3: Network structure based on a part of Beijing urban area

deliveries corresponding to small packages. The capacity of first-echelon vehicles is 1000 and the capacity of tricycles used in the second echelon is 40. Furthermore, we assume that the unit cost $c_1 = c_2 = 1$, i.e., the objective is to minimize the total distance traveled (measured by the Euclidean distance) in the two-echelon system.

To produce a specific instance, i.e., a realization of demand, we generate $n_c = \lfloor \rho A_c \rfloor$ consumers uniformly in each cell c, where ρ is a density parameter and A_c is the area of cell c. We use an "accept-reject" method to generated consumer location. Specifically, we construct a rectangle enclosing the cell and uniformly generate a point in the rectangle. If the point is in cell c, we accept it as a consumer location; otherwise the point is rejected. We terminate the process when n_c consumer locations have been accepted. To decide whether a point is in the cell we have to solve a "point in polygon" problem, which we do using the ray casting algorithm (Huang and Shih, 1997). We generate five instances for each density $\rho = 2, 4, \ldots, 14$ consumers per square kilometer.

To obtain a delivery plan for an instance when the unconstrained delivery strategy is employed, we use the Adaptive Large Neighborhood Search (ALNS) algorithm for solving instances of the 2E-VRP developed by Hemmelmayr et al. (2012). To obtain a delivery plan for an instance when a constrained delivery strategy is employed, we use the branch-and-cut algorithm developed by Archetti et al. (2014) for solving instances of the SDVRP (i.e., to get a delivery plan for the first echelon), and the record-by-record algorithm developed by (Groër et al., 2010) for solving instances of the CVRP (i.e., to get a delivery plan for the second echelon). We implement the instance generation and the block design formulations via Python 2.7.8 calling commercial solver Gurobi 6.5.0.

Due to the complexity of the 2E-VRP, we solve the 2E-VRP instances on a server with a processor E5-2650 at 2.3 GHz and 256 GB memory. The rest experiments are run on a laptop with a processor i5 at 2.2 GHz and 4 GB memory.

5.2. Computational experiments

The main focus of our computational study is analyzing the impact of restricting route choices in order to simplify delivery operations, but we start our computational study with an investigation of the block design problem.

5.2.1. The block design problem

We first compare the two block design formulations, i.e., the disaggregated flow formulation (2) and the aggregated flow formulation (3). Somewhat surprisingly, the aggregated flow formulation consistently ourperforms the disaggregated flow formulation. To illustrate the difference in performance of the two formulations, we choose one instance and summarize the Gurobi logs for the solution of the two block design formulations in Table 2.

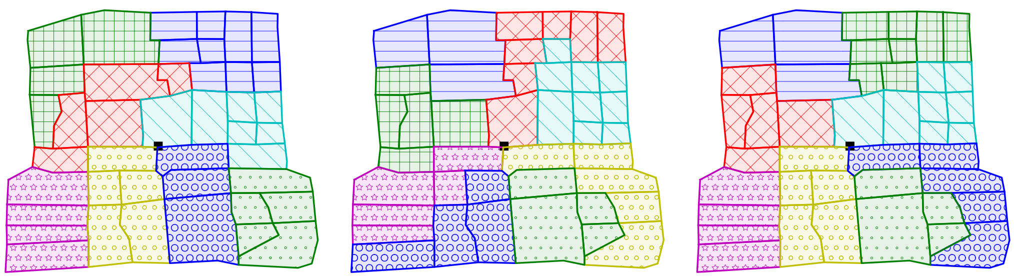
Table 2: Gurobi logs of the two block design formulations (Region 1, $\rho = 2/\text{km}^2$).

	Disaggregated flow	Aggregated flow
# continuous var.	269.89×10^3	6.43×10^3
# binary var.	1806	1806
Presolve time (s)	20.03	0.45
Root solve time (s)	628.77	1.14
Total time (s)	1124.65	2.94
Root simplex iter.	172.50×10^3	2.58×10^3
Total simplex iter.	262.51×10^3	2.58×10^3
Obj. of solution(s) found by root node heuristics	34	34 and 3
# nodes explored	0	0
# cutting planes	64	0

We observe that the difference in the number of continuous variables has a significant impact. Presolving the model and solving the root relaxation take significantly longer with the disaggregated flow formulation. Maybe more importantly, the root node heuristics embedded in Gurobi are more effective with the aggregated flow formulation. These results are representative for the behavior of the two formulations. Therefore, in the following experiments, we use the aggregated flow formulation for the block design problems.

To gain insight into the benefits of using a hierarchical optimization approach to solving instances of the block design problem and into the impact of the different ways to model the connectivity requirement, we conduct the following experiments.

Let M_1 denote the variant in which we do not use hierarchical optimization and simply minimize the number of blocks, i.e., use objective (1a), and let M_2 denote the variant in which we do use hierarchical optimization and first minimize the number of blocks and then minimize the distance from the intermediate facility to the cells representing the blocks, i.e., use objective (4). For comparison purposes, we also introduce variants M_2^1 and M_2^2, in which the secondary objective consists of only the first term of (4) and only the second term of (4), respectively.



(a) M_2 (b) M_2^1 (c) M_2^2

Figure 4: Block designs produced by model variants (density = 10/km^2, target block capacity = $2Q_2$).

Figure 4 illustrates a few of the block designs obtained for Region 1. As expected, in the block design produced by M_2^1, each block has a cell representing the block that is adjacent or close to the intermediate facility, whereas in the design produced by M_2^2, the cells in a block are more "clustered". Although the objective of M^2 combines those of M_2^1 and M_2^2, the design produced by M_2 is more similar to that of M_2^2.

In Table 3, we report the maximum time required to solve an instance in Region 1 for the different model variants. Note that all variants of the hierarchical optimization approach require solving M_1 first. To highlight the time required to solve the second-level problem, we only show the time of solving the second level for the M_2 variants with a plus sign "+" to indicate that this represents the additional solution time (on top of the time required to solve M_1). Somewhat surprisingly, M_2^1 seems to be much harder to solve than M_2 or M_2^2.

Next, we compare the second-echelon distance of the associated delivery plans (produced by solving the TSPs for each block using the record-to-record algorithm – the TSPs that result when the orders in the block are divided over the drivers using the sweep heuristic). The results can be found in Table 4, where the reported distance is the average over the five instances and the best average distance is highlighted in bold font. As expected, the performance of the hierarchical optimization variants is better than the performance of M_1,

Table 3: Solution time (in seconds) in Region 1 ($\bar{Q} = 2Q_2$).

Density	M_1	M_2	M_2^1	M_2^2
2	0.84	+ 4.35	+ 2.29	+ 6.38
4	0.80	+ 5.14	+ 1.98	+ 4.07
6	0.99	+ 5.46	+ 10.52	+ 3.24
8	19.00	+ 2.91	+ 383.15	+ 2.64
10	154.11	+ 12.01	+ 142.34	+ 21.92
12	102.50	+ 3.43	+ 3600.00*	+ 9.50
14	3600.00*	+ 2.72	+ 3237.63	+ 8.18

*Time limit reached.

and among the hierarchical optimization variants M_2^2 slightly outperforms the others, but not by much.

Table 4: Second-echelon distance ($\bar{Q} = 2Q_2$).

	Region 1				Region 2			
Density	M_1	M_2	M_2^1	M_2^2	M_1	M_2	M_2^1	M_2^2
2	77.02	**76.86**	82.08	**76.86**	**51.10**	**51.10**	**51.10**	**51.10**
4	113.30	**112.03**	115.19	**112.03**	**78.09**	80.03	78.29	80.03
6	156.33	**152.77**	155.57	153.37	106.10	105.78	**101.63**	102.34
8	198.98	195.45	**195.00**	195.45	125.96	**124.42**	129.10	125.33
10	223.64	219.63	**212.56**	214.03	144.76	**140.78**	148.77	**140.78**
12	256.06	**250.39**	255.12	255.65	171.99	172.53	177.53	**169.14**
14	296.83	288.89	**287.67**	289.11	195.14	**191.12**	**191.12**	**191.12**
	Region 3				Region 4			
Density	M_1	M_2	M_2^1	M_2^2	M_1	M_2	M_2^1	M_2^2
2	**43.04**	**43.04**	**43.04**	**43.04**	**38.93**	**38.93**	**38.93**	**38.93**
4	66.19	**65.37**	68.17	**65.37**	60.48	53.66	**52.89**	53.66
6	91.90	**83.78**	84.97	**83.78**	75.31	**72.71**	73.02	73.62
8	108.10	107.36	**103.57**	105.91	89.93	**84.19**	91.35	**84.19**
10	121.21	**120.15**	120.64	121.21	100.21	**99.82**	**99.82**	**99.82**
12	148.74	**147.05**	147.47	147.10	115.76	**115.71**	**115.71**	**115.71**
14	172.18	**170.05**	**170.05**	171.27	145.12	144.21	144.33	**142.23**

As mentioned in Section 4, the introduction of blocks creates flexibility to handle growth in demand and variability in demand. In the following experiments, we seek to better understand the impact of the target block capacity.

In Figure 5, we show the number of cells assigned to a block for increasing consumer density (i.e., growing demand) and for varying target block capacities, i.e., $\bar{Q} = Q_2, \bar{Q} = 2Q_2$, and $\bar{Q} = 3Q_2$. As expected, we see that the number of cells assigned to a block decreases as the consumer density increases, that fewer cells are assigned to a block when the target

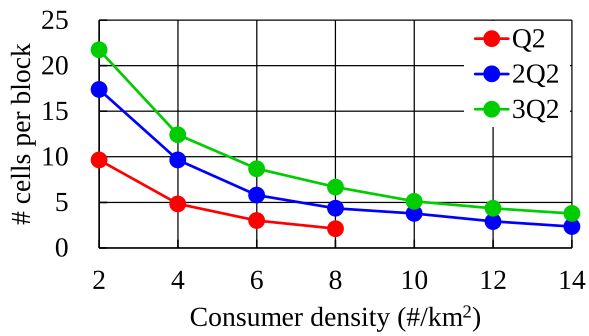
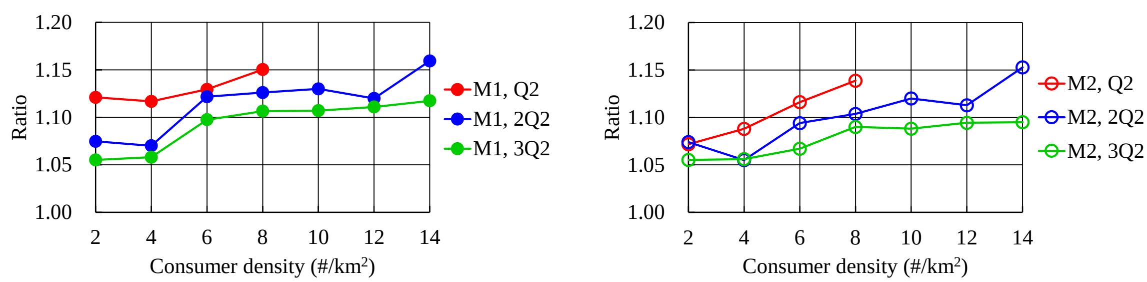


Figure 5: Number of cells per block.

block size is smaller, and that when the consumer density is larger than eight, the number of orders in a cell exceeds the vehicle capacity and setting the target block size to the vehicle capacity results in infeasibility.

More interestingly, in Figure 6, we compare the performance of M_1 and M_2 for varying target block capacities, i.e., $\bar{Q} = Q_2, \bar{Q} = 2Q_2$, and $\bar{Q} = 3Q_2$. We report the ratio of the second-echelon distance obtained when employing a constrained delivery strategy with a specific choice of target block size and the second-echelon distance obtained when employing the unconstrained delivery strategy. Again, as expected, we observe that introducing flexibility,



(a) M_1 (b) M_2

Figure 6: Second-echelon distance ratio for various choices of target block size.

i.e., using target block capacities of $2Q_2$ and $3Q_2$, leads to lower second-echelon distances, and, again, that the use of hierarchical optimization tends to be beneficial.

Overall, the larger block capacity increases the performance of block design, as vehicles have more flexibility within a block. Besides, considering the primary and secondary objective functions is beneficial, especially when the consumer density is small and the block capacity is small, which results in multiple solutions that minimizes the number of blocks.

5.2.2. Impact of restricting route choices

As mentioned above, we use the ALNS algorithm of Hemmelmayr et al. (2012) to produce a delivery plan for an instance when we use the unconstrained delivery strategy. Since the algorithm incorporates random choices, to obtain a high-quality solution in a given amount of time, there is a trade-off between the time given to a single solve (i.e., the maximum number of iterations) and the number of replications (i.e., the number of times an instance is solved). Our first computational experiment is geared towards identifying good choices for these parameters. We considered three combinations of maximum number of iterations and number of replications: 50,000 and 5, 50,000 and 10, and 100,000 and 5. The results for the five instances with density $\rho = 10$ can be found in Table 5. As we are interested in the best possible delivery plan, we focus on the minimum objective function value among the solutions for the different replications and highlight those in bold font. We also calculate the relative difference (column "Diff.") with respect to the results for 50,000 iterations and 5 replications. (Note that the column "Time" gives the average solution time per replication.) These results suggest that the benefits of more replications or more iterations is marginal, and, therefore, in the experiments to be discussed next, we have set the number of iterations to 50,000 and the number of replications to 5.

We conjectured that when the number of orders that needs to be delivered is large and the location density of the delivery addresses is high, then a two-echelon logistics system employing a constrained delivery strategy has near-optimal performance. The next computational experiment substantiates this conjecture.

In Figure 7, we show the ratio of the total distance of the constrained delivery strategy using cells and that of the unconstrained delivery strategy, the ratio of the total distance of the constrained delivery strategy using regions and that of the unconstrained delivery strategy, as well as the ratio of the total distance of the constrained delivery strategy using blocks (with target block capacity equal to $3Q_2$) and that of the unconstrained delivery strategy, for different consumer densities. Note that, when consumer densities are larger than $8/\text{km}^2$, there exist cells of which the demands are larger than the vehicle capacity. Under such a situation, additional vehicles are assigned to these cells to ensure the feasibility.

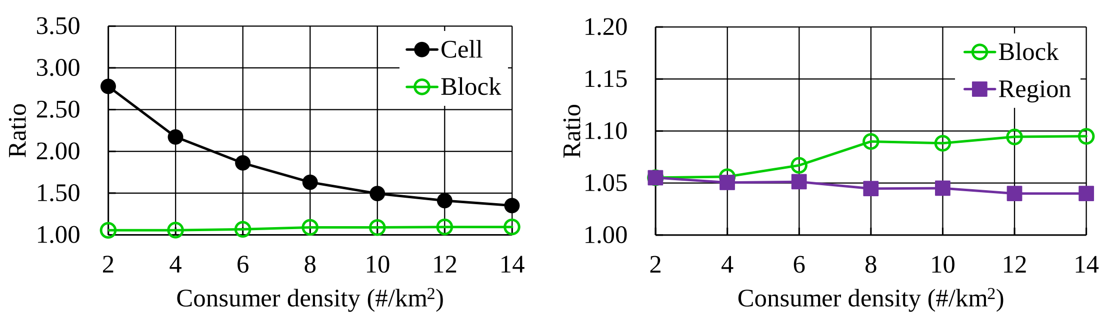


Figure 7: Total distance ratio (Note that scales of the y-axes are different).

We see that, as conjectured, when the consumer density increases, the ratio of the total distance of the constrained delivery strategy using cells and that of the unconstrained delivery strategy decreases. Furthermore, the ratio of the total distance of the delivery strategy using regions and that of the unconstrained delivery strategy shows a similar trend, but much less

Table 5: Results of ALNS with different parameters.

Ins.	50k iterations, 5 replications				50k iterations, 10 replications					100k iterations, 5 replications				
	Min	Ave.	Max	Time (s)	Min	Diff.	Ave.	Max	Time (s)	Min	Diff.	Ave.	Max	Time (s)
1	597.7	609.1	627.3	2115.4	598.0	0.1%	620.3	666.1	1604.2	**597.6**	0.0%	622.2	702.0	2976.6
2	597.6	610.1	629.5	1783.8	**593.3**	-0.7%	631.3	736.8	1698.2	599.0	0.2%	604.9	613.4	2460.8
3	**598.1**	606.5	603.7	1331.6	600.3	0.4%	615.3	641.3	1662.5	603.7	0.9%	627.7	647.1	2960.2
4	**591.5**	618.2	655.3	1576.2	597.3	1.0%	612.5	627.0	1656.1	603.2	1.9%	615.6	633.6	2195.4
5	586.2	615.1	653.4	1355.4	**584.6**	-0.3%	613.3	632.2	2087.3	596.3	1.7%	626.1	683.6	2962.2

pronounced. Note too that the latter ratio never exceeds 1.06 (even for very low consumer densities).

Importantly, using blocks, even with a relatively small target block capacity, improves the quality of the delivery plans significantly (compared to using cells). In fact, the ratio of the total distance of the constrained delivery strategy using blocks and that of the unconstrained delivery strategy is small, never more than 1.10 regardless of the consumer density. Comparing to the constrained delivery strategy using regions (which can be seen as using a single block of sufficiently large target block capacity), the extra distance traveled of the constrained delivery strategy using blocks is at most 5.4% (when consumer density is $14/\text{km}^2$).

These results have important practical implications. They demonstrate that employing delivery strategies aimed at reducing operational complexity, both in terms of planning and execution, does not have to lead to huge cost increases; employing a constrained delivery strategy using blocks leads to high-quality, low-cost delivery plans.

To obtain additional insights, we further decompose the total distance into the first-echelon distance and the second-echelon distance. The results can be found in Figures 8 and 9, respectively. (Note that the scales of the y-axes in Figures 8 and 9 are different.) As expected, we observe that the difference in first-echelon distance is small for all consumer densities, and that it is the second-echelon distance that shows differences, especially when the consumer densities are small.

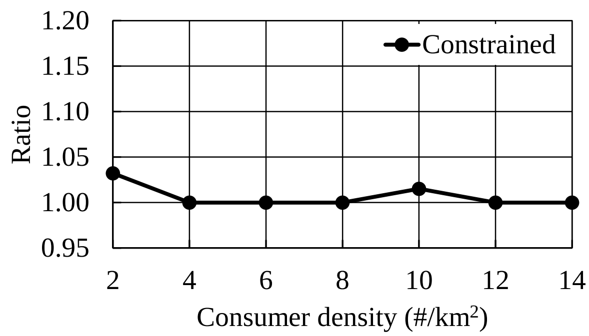


Figure 8: First-echelon distance ratio.

6. Analytical results

In this section, we provide analytical results in the form of a worst-case analysis of the ratio between the optimal value of the constrained delivery strategy using regions and the optimal value of the unconstrained delivery strategy, i.e., Z^R/Z^U, for specific geographic topologies. More specifically, we consider geographic topologies in which regions are either square or rectangular and distances that are either based on the Euclidean metric or the Manhattan metric.

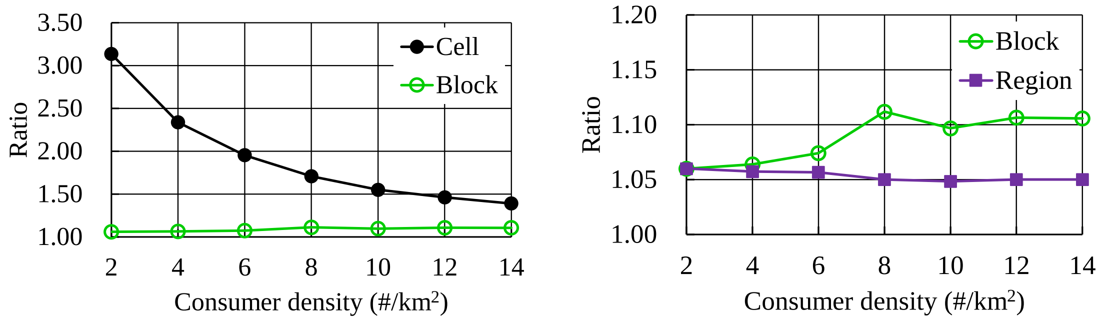


Figure 9: Second-echelon distance ratio (Note that scales of y-axes are different).

In this section, we only present the analysis for the case with square regions and Euclidean distances; the analysis for the other cases can be found in Appendix A. Therefore, we have a setting with the following characteristics:

1. The CDC is located either on the boundary of the urban area or within the urban area.
2. The urban area is partitioned into $m_a \times m_b$ regions where m_a and m_b are even numbers.
3. Each region is a square with unit length sides.
4. The intermediate facility is located at the center of each region.
5. Each region has only one cell, i.e., the region itself.
6. There is at least one consumer in each region.
7. The capacity of the vehicles used in both echelons is such that an optimal solution for the unconstrained delivery strategy uses a single delivery route in both echelons (i.e., the vehicle capacity is sufficiently large).
8. The unit cost of travel in the first echelon, c_1, and the unit cost of travel in the second echelon, c_2, satisfy $c_1 \geq c_2$.

We define Z_1^D and Z_2^D as the first-echelon and second-echelon delivery cost under delivery strategy D, where $D = R$ (the constrained delivery strategy using regions) or U (the unconstrained delivery strategy). We have that the total cost for delivery strategy D, Z^D, satisfies $Z^D = Z_1^D + Z_2^D$, for $D = R$ or U.

In the following, we first determine the worst-case ratio of Z_2^R/Z_2^U, i.e., the ratio of the second-echelon cost between the constrained delivery strategy using regions and the unconstrained delivery strategy. Then, we determine the worst-case ratio of Z^R/Z^U, i.e., the ratio of the total cost between the constrained delivery strategy using regions and the unconstrained delivery strategy.

6.1. Worst-case ratio of the second-echelon cost

We start by analyzing a few specific subnetworks. The results will form the building blocks for analyzing a general network. We first consider a two-region subnetwork, where we compare the cost of the constrained *routes* in these two regions and the unconstrained *path* passing through this subnetwork (which forms a segment of the unconstrained route).

Furthermore, we assume that the unconstrained path enters at a long boundary of the subnetwork and exits at the opposite long boundary of the subnetwork, as shown in Figure 10. Let $Z_2^{U,\mathcal{P}}$ denote the cost of the unconstrained path. We have the following result.

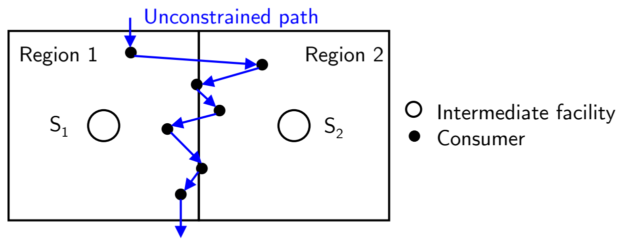


Figure 10: Illustration of the unconstrained path.

Theorem 1. *In a two-region subnetwork in which an unconstrained path enters at a long boundary and exits at the opposite long boundary, we have that* $\dfrac{Z_2^R}{Z_2^{U,\mathcal{P}}} \le 2\sqrt{2}+2$ *and this bound is tight.*

Proof. Define N_r as the set of consumers in region r ($r = 1$ or 2) and define $N = N_1 \bigcup N_2$ as the set of all consumers. Given the consumers and their corresponding locations, we can obtain an optimal *unconstrained path* $\mathcal{P}^{U,*} = \{O, n_{[1]}, n_{[2]}, \ldots, n_{[|N|]}, D\}$ with minimum cost $Z_2^{U,\mathcal{P}} = c_2 d(\mathcal{P}^{U,*})$, where $d(\mathcal{P}^{U,*})$ is the length of path $\mathcal{P}^{U,*}$ and where O and D represent points on the entering and exiting boundary, respectively. Using an optimal unconstrained path, we can construct associated constrained routes as follows:

1. Partition the consumers $n_{[1]}, n_{[2]}, \ldots, n_{[|N|]}$ to two sets based on the regions they belong to. Assume the two sets are $\{n^1_{[1]}, \ldots, n^1_{[|N_1|]}\}$ and $\{n^2_{[1]}, \ldots, n^2_{[|N_2|]}\}$.
2. Construct the constrained route in region r as $\mathcal{R}_r^{R,h} = \{S_r, n^r_{[1]}, n^r_{[2]}, \ldots, n^r_{[|N_r|]}, S_r\}$, $r = 1$ or 2.

Denote the cost of these constrained routes by $Z_2^{R,h}$ (superscript h indicates that this represents a heuristic solution). We have $Z_2^{R,h} = c_2(d(\mathcal{R}_1^{R,h}) + d(\mathcal{R}_2^{R,h}))$, where $d(\mathcal{R}_r^{R,h})$ is the length of route $\mathcal{R}_r^{R,h}$ ($r = 1$ or 2). Thus we have

$$\frac{Z_2^R}{Z_2^{U,\mathcal{P}}} \le \frac{Z_2^{R,h}}{Z_2^{U,\mathcal{P}}} = \frac{d(\mathcal{R}_1^{R,h}) + d(\mathcal{R}_2^{R,h})}{d(\mathcal{P}^{U,*})}.$$

We further decompose the length of the constrained route $\mathcal{R}_r$ ($r = 1$ or 2) into two terms, $d_r^s = d(S_r, n^r_{[1]}) + d(n^r_{[|N_k|]}, S_r)$, the distance from the intermediate facility S_r to the first consumer and from the last consumer back to the intermediate facility S_r (highlighted as pink lines in Figure 11), and $d_r^v = \sum_{i=1}^{|N_r|-1} d(n^r_{[i]}, n^r_{[i+1]})$, the distance traveled visiting consumers in region r (highlighted as red lines in Figure 11). Furthermore, we define the constrained path *inside* region r as $\mathcal{P}_r^{R,h} = \{n_{[1]}^r, n_{[2]}^r, \ldots, n_{[|N_r|]}^r\}$. Therefore we have

$$\frac{Z_2^R}{Z_2^{U,\mathcal{P}}} \leq \frac{d_1^s + d_2^s}{d(\mathcal{P}^{U,*})} + \frac{d_1^v + d_2^v}{d(\mathcal{P}^{U,*})} \tag{5}$$

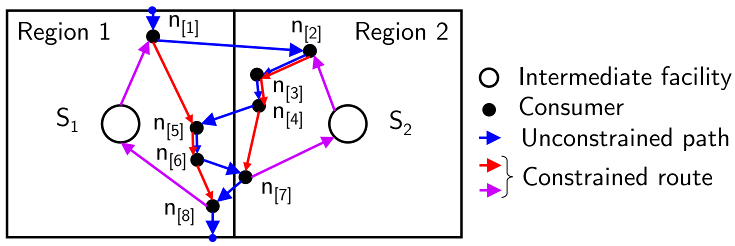


Figure 11: Illustration of the unconstrained path and the constrained routes.

In the following, we will analyze the two terms in the righthand side of inequality (5) separately.

The values of d_1^s, d_2^s, and $d(\mathcal{P}^{U,*})$ depend on the locations of consumers. We have that $\min d(\mathcal{P}^{U,*}) = 1$ which occurs when all consumer nodes are close to the shared boundary of the two regions. And we have that $\max d_r^s = \sqrt{2}$ which occurs when the first consumer $n_{[1]}^r$ and the last consumer $n_{[|N_r|]}^r$ are located at the corners of the square region. Therefore, we have

$$\frac{d_1^s + d_2^s}{d(\mathcal{P}^{U,*})} \leq \frac{\max d_1^s + \max d_2^s}{\min d(\mathcal{P}^{U,*})} = 2\sqrt{2}. \tag{6}$$

As illustrated in Figure 11, we have $d_1^v \leq d(\mathcal{P}^{U,*})$ and $d_2^v \leq d(\mathcal{P}^{U,*})$ because both constrained paths $\mathcal{P}_r^{R,h}$ ($r = 1$ or 2) take short cuts compared to the unconstrained path $\mathcal{P}^{U,*}$ (as the triangle inequality holds). Thus, we have

$$\frac{d_1^v + d_2^v}{d(\mathcal{P}^{U,*})} \leq 2. \tag{7}$$

Combining expressions (6) and (7), we obtain the bound that

$$\frac{Z_2^R}{Z_2^{U,\mathcal{P}}} \leq 2\sqrt{2} + 2.$$

The following worst-case example shows that the bound is tight. In Figure 12, each consumer node is close to a corner of the region (ϵ away from the two boundaries). Thus, we have

$$\frac{Z_2^R}{Z_2^{U,\mathcal{P}}} = \frac{2(\sqrt{2} - 2\sqrt{2}\epsilon) + 2(1 - 2\epsilon)}{1 + 2\epsilon},$$

and when $\epsilon \to 0$, indicating that the consumer nodes are getting infinitely close to corners, the ratio of the length of the constrained routes and the length of the unconstrained path gets infinitely close to $2\sqrt{2} + 2$, which proves that the bound is tight. □

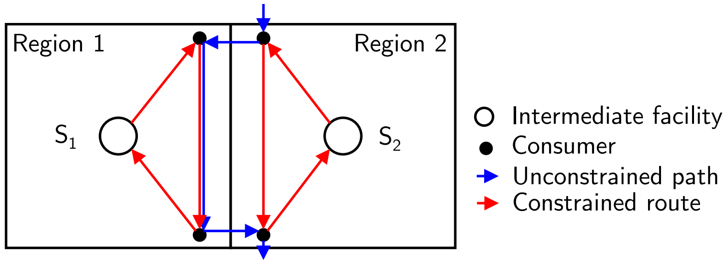


Figure 12: A two-region worst-case example that shows the bound is tight.

Next, we consider a subnetwork with 2×2 regions. If the unconstrained path enters and exits at opposite boundaries of the subnetwork, it is a straightforward extension of Theorem 1, and the worst-case ratio is $2\sqrt{2}+2$. Thus, we focus on the situation where the unconstrained path enters and exits at adjacent boundaries of the subnetwork. Using techniques similar to those used in the proof of Theorem 1, we are able to show the same worst-case ratio. A four-region (2×2 regions) worst-case example is shown in Figure 13, where a consumer node located at the corner of neighboring regions corresponds to multiple

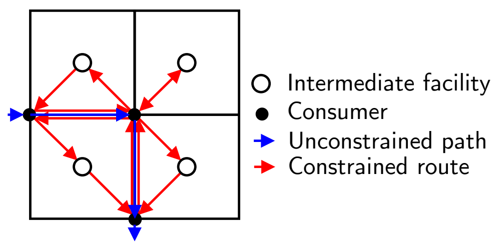


Figure 13: A four-region (2×2 regions) worst-case example that shows the bound is tight.

consumers in the corners of these regions (and infinitely close to the corner). More formally

Theorem 2. *In a subnetwork with 2×2 regions in which an unconstrained path enters and exits at its boundaries, we have that $\dfrac{Z_2^R}{Z_2^{U,\mathcal{P}}} \leq 2\sqrt{2}+2$ and this bound is tight.*

Proof. Similar to Theorem 1. □

Next, we consider a network with $m_a \times m_b$ regions. Let route $\mathcal{R}^{U,*}$ be an optimal second-echelon unconstrained route with minimum length. We denote the length of route $\mathcal{R}^{U,*}$ by $d(\mathcal{R}^{U,*})$. We start by deriving a lower bound on $d(\mathcal{R}^{U,*})$, which will be used in the proof of the worst-case ratio of Z_2^R/Z_2^U in a network of $m_a \times m_b$ regions.

Lemma 3. *A lower bound on $d(\mathcal{R}^{U,*})$ in a network with $m_a \times m_b$ regions is $m_a m_b/2 + (\sqrt{10}+\sqrt{2})/2 - 2$.*

Proof. Since we have assumed that m_a and m_b are even, we can view the network as consisting of $m_a/2 \times m_b/2$ subnetworks of 2×2 regions. To get a lower bound on $d(\mathcal{R}^{U,*})$ in the network, we assume that the unconstrained route only visits the consumers in the center of each of the 2×2 subnetworks (because we have assumed that each region has at least one consumer, each region has to be visited and visiting the center of a 2×2 subnetwork is the best possible). Therefore, the unconstrained route visits a grid of $m_a/2\times m_b/2$ consumers, where the length of an arc connecting two nearest consumers (referred to as adjacent consumers and an adjacent arc) is two. In the following, we discuss two cases based on the parity of $m_a/2$ and $m_b/2$.

Case 1. Either $m_a/2$ or $m_b/2$ is even. If we ignore the start and end at an intermediate facility and consider a route visiting only the $m_a/2\times m_b/2$ consumers, then Thompson (1977) has shown that if either $m_a/2$ or $m_b/2$ is even, an optimal route $\mathcal{R}'$ only uses adjacent arcs. As $\mathcal{R}'$ consists of $m_a m_b/4$ adjacent arcs, its length $d(\mathcal{R}')$ is $m_a m_b/4\times 2 = m_a m_b/2$.

Next, we modify the route $\mathcal{R}'$ so that it includes an intermediate facility. We take an intermediate facility S and the two adjacent consumers, say n_1 and n_2, that minimize $d(n_1,S)+d(S,n_2)$. Then, we replace the arc (n_1,n_2) with arcs (n_1,S) and (S,n_2), as shown in Figure 14, where $d(n_1,S)+d(S,n_2)=(\sqrt{10}+\sqrt{2})/2$ and $d(n_1,n_2)=2$. The length of the resulting route gives a lower bound on $d(\mathcal{R}^{U,*})$, i.e., $d(\mathcal{R}^{U,*})\geq m_a m_b/2+(\sqrt{10}+\sqrt{2})/2-2$.

Case 2. Both $m_a/2$ and $m_b/2$ are odd. In this case an optimal route uses at least one non-adjacent arc (Thompson, 1977). Thus, its length will be greater than or equal to that of route $\mathcal{R}'$ of Case 1, which completes the proof. □

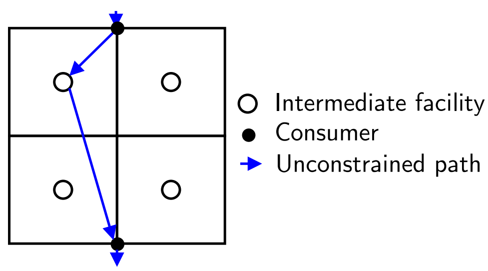


Figure 14: Illustrating the incorporation of an intermediate facility in a second-echelon unconstrained route.

Next, we show that the worst-case ratio of Z_2^R/Z_2^U in a network with $m_a\times m_b$ regions is $2\sqrt{2}+2$, the same as the worst-case ratio $Z_2^R/Z_2^{U,\mathcal{P}}$ for the subnetworks considered in Theorems 1 and 2.

Theorem 4. *Consider a network with $m_a\times m_b$ regions. We have* $\dfrac{Z_2^R}{Z_2^U}\leq 2\sqrt{2}+2$ *and this bound is tight.*

Proof. Consider the optimal unconstrained second-echelon route $\mathcal{R}^{U,*}$ starting and ending at intermediate facility $S^{U,*}$ and visiting all consumers in the sequence $\{S^{U,*},n_{[1]},\ldots,n_{[|N|]},S^{U,*}\}$. Using the visiting sequence of consumers in $\mathcal{R}^{U,*}$, we construct a constrained route for each region r, $\mathcal{R}_r^{R,h}=\{S_r,n^r_{[1]},\ldots,n^r_{[|N_r|]},S_r\}$, which visits the consumers in the region in the same sequence as in $\mathcal{R}^{U,*}$, and denote the combined length of these constrained routes as $Z_2^{R,h}$. We define $d^s = \sum_{r=1}^{R} \Big(d(S_r, n^r_{[1]}) + d(n^r_{[|N_r|]}, S_r) \Big)$ to be the summation of the first-leg and last-leg distance of the constrained routes over all regions and $d^v = \sum_{r=1}^{R} \sum_{i=1}^{|N_r|-1} d(n^r_{[i]}, n^r_{[i+1]})$ to be the summation of the lengths of the paths visiting consumers in the constrained routes over all regions. Clearly, we have $Z_2^{R,h} = c_2(d^s + d^v)$. This implies that we have

$$\frac{Z_2^R}{Z_2^U} \leq \frac{Z_2^{R,h}}{Z_2^U} = \frac{d^s}{d(\mathcal{R}^{U,*})} + \frac{d^v}{d(\mathcal{R}^{U,*})}.$$

Again, we analyze the terms $d^s/d(\mathcal{R}^{U,*})$ and $d^v/d(\mathcal{R}^{U,*})$ separately.

The worst-case ratio of the first term can be derived from Lemma 3. We have $\frac{d^s}{d(\mathcal{R}^{U,*})} \leq \frac{\max d^s}{\min d(\mathcal{R}^{U,*})} \leq \frac{m_a m_b \sqrt{2}}{m_a m_b/2 + (\sqrt{10} + \sqrt{2})/2 - 2} \leq 2\sqrt{2}$.

Next, we consider the second term $d^v/d(\mathcal{R}^{U,*})$. We define two arc sets A^* and $\bar{A}^*$ for the arcs in the constrained routes related to the term d^v. For any arc $(n^r_{[i]}, n^r_{[i+1]})$ $(i = 1, \ldots, |N_r| - 1)$ of the constrained route in region r $(r = 1, \ldots, R)$, if it belongs to $\mathcal{R}^{U,*}$, we add arc $(n^r_{[i]}, n^r_{[i+1]})$ to set A^*, otherwise we add arc $(n^r_{[i]}, n^r_{[i+1]})$ to set $\bar{A}^*$. That is, we partition the arcs in the constrained routes considered in the distance term d^v. Define distance $d(A^*)$ and $d(\bar{A}^*)$ as the summation of distance of all arcs in A^* and $\bar{A}^*$, respectively. Thus we have $d^v = d(A^*) + d(\bar{A}^*)$. We have $d(A^*) \leq d(\mathcal{R}^{U,*})$ as the arcs in A^* are a subset of arcs in $\mathcal{R}^{U,*}$.

Next, we discuss the bound of $d(\bar{A}^*)/d(\mathcal{R}^{U,*})$. It is obvious that, if the inequality $d(\bar{A}^*) \leq d(\mathcal{R}^{U,*})$ holds, we have $d^v/d(\mathcal{R}^{U,*}) \leq 2$ which would complete the proof. Unfortunately, this inequality does not always hold (see an example in Figure 15). For an arc $(i, j) \in \bar{A}^*$ in region r, let the corresponding unconstrained path in $\mathcal{R}^{U,*}$ starting at consumer i and ending at consumer j be $\mathcal{P}_{ij}^{U,*}$. It is obvious that

$$d(i, j) \leq d(\mathcal{P}_{ij}^{U,*}), \tag{8}$$

as the arc (i, j) is a short cut of the path $\mathcal{P}_{ij}^{U,*}$ due to the triangle inequality. Furthermore, when we sum inequality (8) over the arc set $\bar{A}^*$, we obtain

$$\begin{aligned} \sum_{(i,j)\in \bar{A}^*} d(i, j) &\leq \sum_{(i,j)\in \bar{A}^*} d(\mathcal{P}_{ij}^{U,*}) \\ d(\bar{A}^*) &\leq \sum_{(i,j)\in \bar{A}^*} d(\mathcal{P}_{ij}^{U,*}). \end{aligned} \tag{9}$$

Unfortunately, we may not have $\sum_{(i,j)\in \bar{A}^*} d(\mathcal{P}_{ij}^{U,*}) \leq d(\mathcal{R}^{U,*})$, because some of the unconstrained paths can overlap. An example is shown in Figure 15, where path $\mathcal{P}_{i_1 j_1}^{U,*}$ and $\mathcal{P}_{ij}^{U,*}$ overlap (more specifically $\mathcal{P}_{i_1 j_1}^{U,*}$ is a subpath of $\mathcal{P}_{ij}^{U,*}$). In such a situation, the sum on the right-hand side in inequality (9) counts certain segments of the unconstrained route more than once and can be larger than $d(\mathcal{R}^{\mathcal{U},*})$.

In the following, we identify a condition that guarantees $\sum_{(i,j)\in \bar{A}^*} d(\mathcal{P}_{ij}^{U,*}) \leq d(\mathcal{R}^{U,*})$.

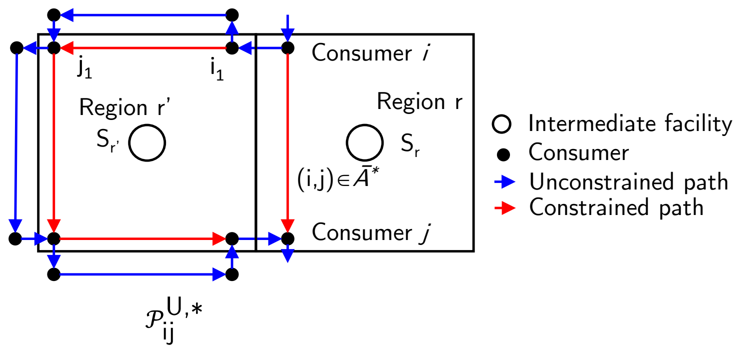


Figure 15: Example of $d(\bar{A}^*) > d(\mathcal{R}^{U,*})$

Condition 4.1. *For each arc $(i,j) \in \bar{A}^*$ in region r, the corresponding unconstrained path $\mathcal{P}_{ij}^{U,*} \subset \mathcal{R}^{U,*}$ starting at consumer i and ending at consumer j visits region r and only a single adjacent region.*

An example of when this condition is satisfied is illustrated in Figure 16.

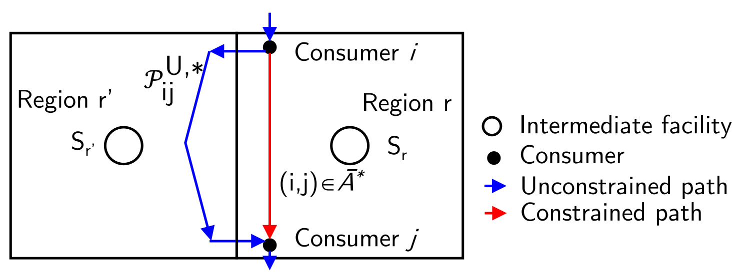


Figure 16: Illustration of the condition that guarantees $\sum_{(i,j)\in\bar{A}^*} d(\mathcal{P}_{ij}^{U,*}) \leq d(\mathcal{R}^{U,*})$.

When Condition 4.1 holds, there are no arcs (i',j') and $(i,j) \in \bar{A}^*$ such that their corresponding unconstrained paths $\mathcal{P}_{i'j'}^{U,*}$ and $\mathcal{P}_{ij}^{U,*}$ overlap (which will be proved later). That is, the paths $\mathcal{P}_{ij}^{U,*}$ for $(i,j) \in \bar{A}^*$ are disjunct. When the paths $\mathcal{P}_{ij}^{U,*}$ for $(i,j) \in \bar{A}^*$ are disjunct, it is easy to obtain $\sum_{(i,j)\in\bar{A}^*} d(\mathcal{P}_{ij}^{U,*}) \leq d(\mathcal{R}^{U,*})$. Combining inequality (9), we have

$$d(\bar{A}^*) \leq \sum_{(i,j)\in\bar{A}^*} d(\mathcal{P}_{ij}^{U,*}) \leq d(\mathcal{R}^{U,*}).$$

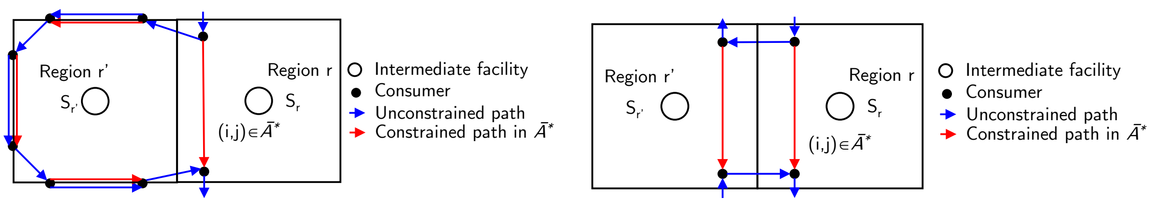
To see that Condition 4.1 guarantees that the paths $\mathcal{P}_{ij}^{U,*}$ for $(i,j) \in \bar{A}^*$ are disjunct, assume, to the contrary, that there exist (i,j) and $(i',j') \in \bar{A}^*$ such that $\mathcal{P}_{i'j'}^{U,*}$ overlaps with $\mathcal{P}_{ij}^{U,*}$. If $\mathcal{P}_{i'j'}^{U,*}$ is a segment of $\mathcal{P}_{ij}^{U,*}$ (as in Figure 15), then $\mathcal{P}_{ij}^{U,*}$ has to leave and enter region r', which violates Condition 4.1. If $\mathcal{P}_{i'j'}^{U,*}$ is not a segment of $\mathcal{P}_{ij}^{U,*}$, but the paths have a common arc, then that violates the optimality of the unconstrained route $\mathcal{R}^{U,*}$.

In the following, we show that even if Condition 4.1 is not satisfied, which implies that the inequality $d(\bar{A}^*) \leq d(\mathcal{R}^{U,*})$ may not hold, the ratio Z_2^R/Z_2^U is still no more than $2\sqrt{2}+2$.

We do so by investigating five cases based on the number of arcs in $\bar{A}^*$ in a region that violate Condition 4.1 as well as their position within the region. The cases are listed below:

- Case 1. There exists at most *one* arc $(i, j) \in \bar{A}^*$ in a region violating Condition 4.1 and nodes i and j are in adjacent quadrants of the region.
- Case 2. There exists at most *one* arc $(i, j) \in \bar{A}^*$ in a region violating Condition 4.1 and nodes i and j are in non-adjacent quadrants of the region.
- Case 3. There exist at most *two* arcs in $\bar{A}^*$ in a region violating Condition 4.1.
- Case 4. There exist at most *three* arcs in $\bar{A}^*$ in a region violating Condition 4.1.
- Case 5. There exist *more than three* arcs in $\bar{A}^*$ in a region violating Condition 4.1.

Case 1. There exists at most *one* arc $(i, j) \in \bar{A}^*$ in a region, say region r, violating Condition 4.1 where consumers i and j are in adjacent quadrants of region r. Let the region adjacent to the side defined by the two quadrants be region r'. Because arc (i, j) violates Condition 4.1, the unconstrained path $\mathcal{P}_{ij}^{U,*}$ visits more than two regions. We can further consider two situations based on the number of times that the unconstrained path enters and exits the neighboring region r'. Note that the number of times that the unconstrained path exits region r' equals the number of times it enters region r'. Because arc (i, j) violates Condition 4.1, the path enters and exits region r' at least twice. Therefore, we consider two subcases: the path enters and exits region r' more than twice (Case 1.1.) and exactly twice (Case 1.2.).



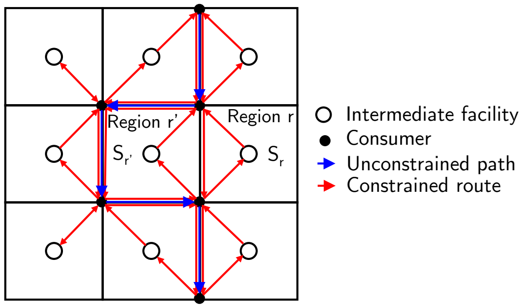
(a) Case 1.1. (b) Case 1.2.

Figure 17: Illustrations of Case 1.

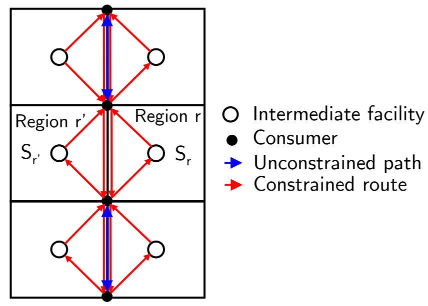
Case 1.1. The unconstrained path enters and exits region r' *more than twice*, as illustrated in Figure 17(a). First, assume that these constrained arcs in region r' satisfy Condition 4.1. That is, each corresponding unconstrained path visits region r' and one neighboring region. Then, the worst-case example (based on the neighboring regions visited) is shown in Figure 18(a), where the unconstrained path enters and exits region r' four times. However, we have that

$$\frac{d(\mathcal{R}^R)}{d(\mathcal{P}^{U,*})} \leq \frac{11 + 9\sqrt{2}}{5} < 2\sqrt{2} + 2.$$

The reason that the ratio is less than $2\sqrt{2} + 2$ is that the unconstrained path takes a detour around region r', and the length of the unconstrained route "moves away" from the lower



(a) Case 1.1.



(b) Case 1.2.

Figure 18: Instances with worst-case ratios of Case 1.

bound derived in Lemma 3. Next, assume that one of the three constrained arcs in region r' in Figure 18(a) violates Condition 4.1. Again, the worst-case ratio of $d(\mathcal{R}^R)/d(\mathcal{P}^{U,*})$ is less than $2\sqrt{2}+2$ as the length of the detours in the unconstrained route increases.

Case 1.2. The unconstrained path enters and exits region r' *exactly twice*, as illustrated in Figure 17(b). Similarly, as shown in Figure 18(b), the worst-case ratio in these regions is

$$\frac{d(\mathcal{R}^R)}{d(\mathcal{P}^{U,*})} \leq \frac{6+6\sqrt{2}}{4} < 2\sqrt{2}+2.$$

The reason is similar: the unconstrained path takes a detour and the length of the unconstrained route moves away from the lower bound derived in Lemma 3.

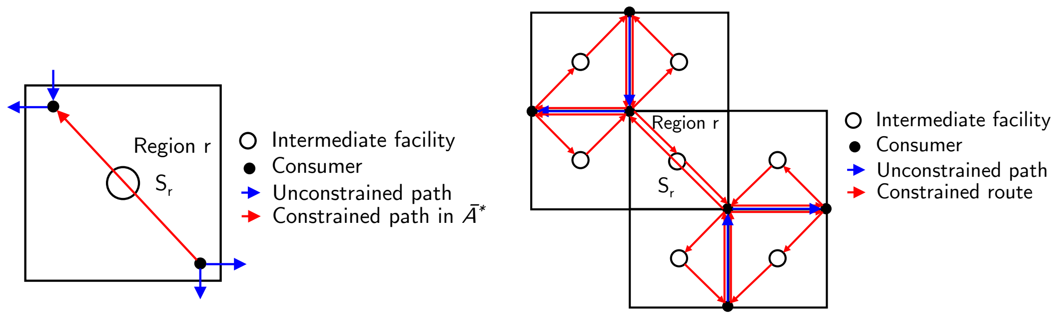
Case 2. There exists at most *one* arc $(i,j) \in \bar{A}^*$ in a region, say region r, violating Condition 4.1 where consumers i and j are in non-adjacent quadrants of the region (illustrated in Figure 19(a)). An instance resulting in the worst-case ratio is shown in Figure 19(b). The worst-case ratio of Case 2 is

$$\frac{d(\mathcal{R}^R)}{d(\mathcal{P}^{U,*})} \leq 2\sqrt{2}+2.$$

Case 3. There exist at most *two* arcs $(i_1,j_1),(i_2,j_2) \in \bar{A}^*$ in a region, say region r, violating Condition 4.1, as illustrated in Figure 20(a). Similar to Case 1, the worst-case ratio of Case 3 is less than $2\sqrt{2}+2$.

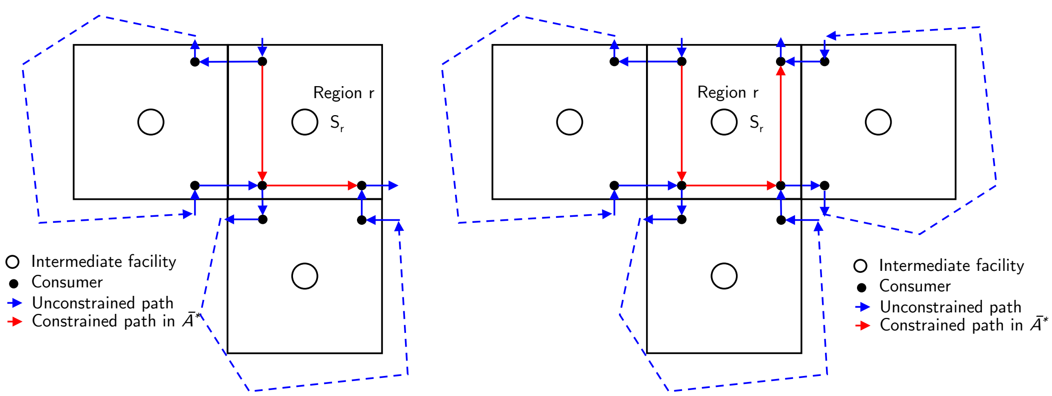
Case 4. There exist at most *three* arcs $(i_1,j_1),(i_2,j_2),(i_3,j_3) \in \bar{A}^*$ in a region, say region r, violating Condition 4.1, as illustrated in Figure 20(b). Similar to Case 1, the worst-case ratio of Case 4 is less than $2\sqrt{2}+2$.

Case 5. There exist *more than three* arcs in $\bar{A}^*$ in a region, say region r, violating Condition 4.1. Note that there cannot exist a region with four arcs in $\bar{A}^*$ connecting the four quadrants of region r, because an unconstrained route leading to such a situation cannot be optimal. Thus, there exist multiple arcs in $\bar{A}^*$ in adjacent quadrants of region r. Therefore, let us consider a situation in which there are two arcs in $\bar{A}^*$ in adjacent quadrants. Denote the two arcs as (i,i') and (j',j), respectively.



(a) Illustration of Case 2 (b) Instance with worst-case ratio

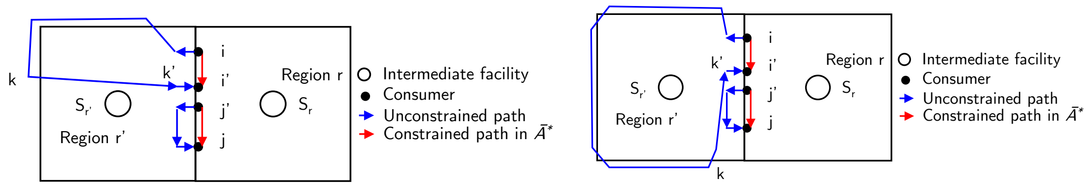
Figure 19: Illustration of Case 2 and instance with worst-case ratio of Case 2.



(a) Case 3. (b) Case 4.

Figure 20: Illustration of Cases 3 and 4.

If the corresponding unconstrained paths $\mathcal{P}^{U,*}_{i,i'}$ and $\mathcal{P}^{U,*}_{j',j}$ only visit region r and an adjacent region r', then Condition 4.1 is satisfied. Therefore, at least one of the unconstrained paths visits more than two regions, say path $\mathcal{P}^{U,*}_{i,i'}$. There are two possibilities: (1) a segment of



(a) First possibility of Case 5 (b) Second possibility of Case 5

Figure 21: Illustration of Case 5.

$\mathcal{P}^{U,*}_{i,i'}$ crosses the middle of a region, as illustrated in Figure 21(a), where section (k, k') crosses the middle of region r', and (2) all segments of $\mathcal{P}^{U,*}_{i,i'}$ are along the boundary of a region, as illustrated in Figure 21(b). The first possibility results in a smaller ratio than in Case 1.1 (Figure 17(a)). The second possibility, cannot occur, because an unconstrained route leading to such a situation cannot be optimal.

A worst-case example demonstrating that the bound is tight can be found in Figure 22, where the network contains $m_a \times m_b$ regions with $m_a = \infty$ and $m_b = 4$. (Note that this worst-case example "concatenates" the worst-case examples presented for Theorems 1 and 2. More specifically, it contains four (2×2)-region subnetworks and an infinite number of two-region subnetworks. □

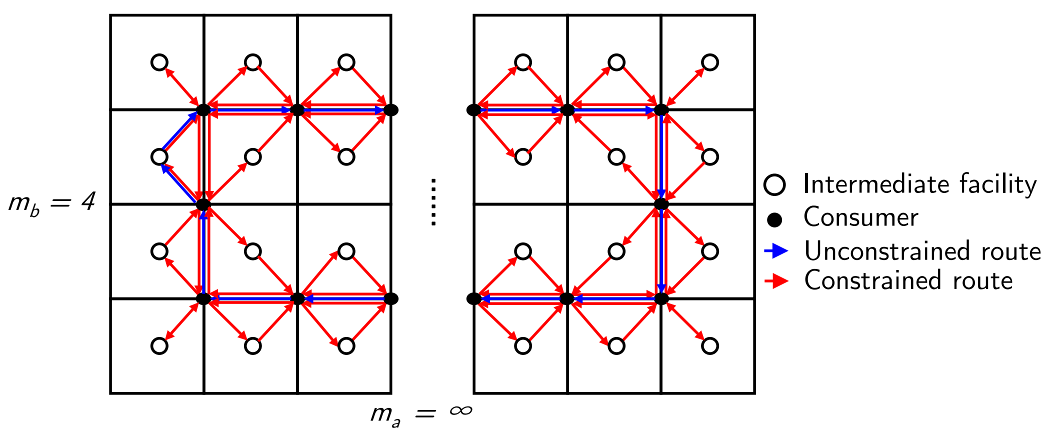


Figure 22: A worst-case example that shows the bound is tight.

6.2. Worst-case ratio of total cost

Based on the previous results, we are able to obtain the worst-case ratio of the total cost of the constrained delivery strategy using regions and the unconstrained delivery strategy.

Theorem 5. *Consider a network with $m_a \times m_b$ regions. We have that $\frac{Z^R}{Z^U} \le 2\sqrt{2}+2+2c_1/c_2$, where c_1 is the first-echelon unit cost and c_2 is second-echelon unit cost, and this bound is tight.*

Proof. We first analyze the first-echelon cost of both delivery strategies. The unconstrained first-echelon route is a round trip from the CDC to the nearest intermediate facility. We denote the length of this round-trip by $\Lambda^U(N_0, N_S) = 2d(\text{CDC}, S^*_{[1]})$, where $S^*_{[1]}$ is the nearest intermediate facility to the CDC. The constrained first-echelon route, which starts and ends at the CDC, visits the $m_a \times m_b$ intermediate facilities. Because the distance between two "neighboring" intermediate facilities is 1, the length of an optimal constrained first-echelon route is

$$Z_1^R = m_a m_b - 1 + \Lambda^R(N_0, N_S),$$

where $\Lambda^R(N_0, N_S)$ represents the distance from the CDC to the first intermediate facility visited plus the distance from the last intermediate facility visited to the CDC. We have $\Lambda^R(N_0, N_S) = d(\text{CDC}, S^*_{[1]}) + d(\text{CDC}, S^*_{[2]})$, where $S^*_{[2]}$ is the second nearest intermediate facility to the CDC. Note that $S^*_{[1]}$ and $S^*_{[2]}$ are "neighboring" intermediate facilities, i.e., $d(S^*_{[1]}, S^*_{[2]}) = 1$. Thus, we have $\Lambda^R(N_0, N_S) - 1 \le \Lambda^U(N_0, N_S)$ due to the triangle inequality.

Next, we decompose the ratio Z^R/Z^U into terms Z_1^R/Z^U and Z_2^R/Z^U and analyze them separately. We have

$$\begin{aligned}
\frac{Z_1^R}{Z^U} &= \frac{Z_1^R}{Z_1^U + Z_2^U} \\
&= \frac{c_1(m_a m_b + \Lambda^R(N_0, N_S) - 1)}{c_1\Lambda^U(N_0, N_S) + Z_2^U} \\
&\le \frac{c_1 m_a m_b + c_1(\Lambda^R(N_0, N_S) - 1)}{c_1\Lambda^U(N_0, N_S) + c_2(m_a m_b/2 + (\sqrt{10}+\sqrt{2})/2 - 2)} && (10) \\
&\le \frac{c_1 m_a m_b + c_1(\Lambda^R(N_0, N_S) - 1) - c_1\Lambda^U(N_0, N_S)}{c_1\Lambda^U(N_0, N_S) + c_2(m_a m_b/2 + (\sqrt{10}+\sqrt{2})/2 - 2) - c_1\Lambda^U(N_0, N_S)} && (11) \\
&\le \frac{c_1 m_a m_b}{c_2(m_a m_b/2 + (\sqrt{10}+\sqrt{2})/2 - 2)} \\
&< 2c_1/c_2. && (12)
\end{aligned}$$

Note that inequality (10) is derived from Lemma 3. Note also that inequality (11) is based on the fact that the righthand side of inequality (10) is greater than 1 (as $c_1 \ge c_2$) and that $\frac{a}{b} \le \frac{a-\Delta}{b-\Delta}$ for any a, b, and Δ satisfying $a \ge b > \Delta \ge 0$. Furthermore, we have

$$\begin{aligned}
\frac{Z_2^R}{Z^U} &= \frac{Z_2^R}{Z_1^U + Z_2^U} \\
&\le \frac{Z_2^R}{Z_2^U} \\
&\le 2\sqrt{2} + 2. && (13)
\end{aligned}$$

Combing inequalities (12) and (13) gives

$$\frac{Z^R}{Z^U} \le 2\sqrt{2} + 2 + 2c_1/c_2.$$

A worst-case example that demonstrates that the bound is tight can be found in Figure 23. For clarity, the first-echelon constrained route (i.e., the route visiting the intermediate facilities) is not shown in the figure. This completes the proof. □

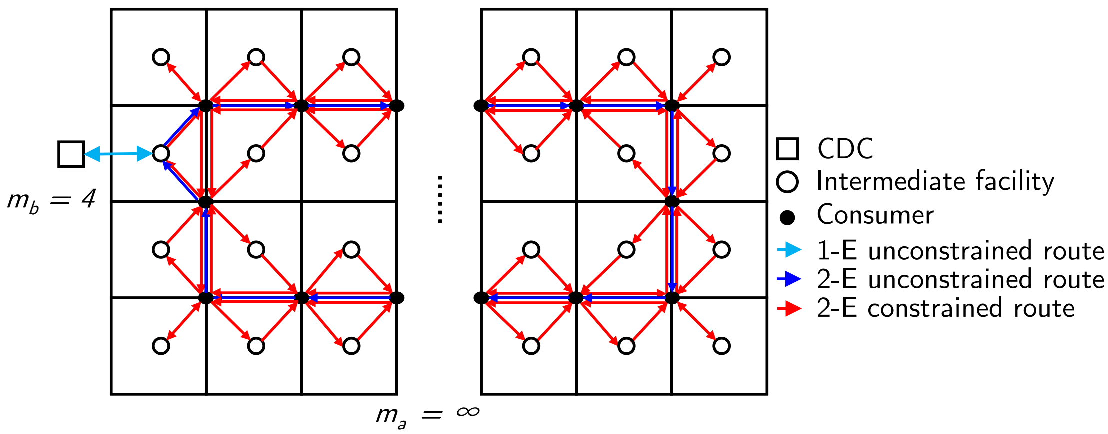


Figure 23: An instance with the worst-case ratio.

7. Conclusions

To deliver to consumers in densely populated urban areas, companies often employ a two-echelon logistics system. By restricting the set of potential vehicle routes employed in one or both of the echelons, it is possible to significantly reduce the complexity of the delivery operation, which is a common practice in real-life environments. We have shown that when consumer density is high the impact on delivery cost of such operational complexity reducing strategies is small. Furthermore, we have shown that simple enhancements, e.g., aggregation of delivery cells, can more easily accommodate growth and more effectively handle daily variations. In addition, we have provided related theoretical results in the form of a worst-case analysis for specific geographic topologies.

The major conclusion of our efforts is that to minimize home delivery costs in megacities the focus should be on the design of the delivery system rather than on the optimization of routing and scheduling. This is interesting, because the academic community has focused (and still focuses) primarily on the routing and scheduling aspects.

Acknowledgement

This research was jointly supported by the National Natural Science Foundation of China (NSFC) under Project No. 71361130017 and the Netherlands Organization for Scientific Research (NWO) under Project No. 629.001.012.

References

Archetti, C., Bianchessi, N., Speranza, M., 2014. Branch-and-cut algorithms for the split delivery vehicle routing problem. European Journal of Operational Research 238 (3), 685–698.

Baldacci, R., Mingozzi, A., Roberti, R., Calvo, R., 2013. An exact algorithm for the two-echelon capacitated vehicle routing problem. Operations Research 61 (2), 298–314.

Beardwood, J., Halton, J., Hammersley, J., 1959. The shortest path through many points. In: Mathematical Proceedings of the Cambridge Philosophical Society. Vol. 55. pp. 299–327.

Breunig, U., Schmid, V., Hartl, R., Vidal, T., 2016. A large neighbourhood based heuristic for two-echelon routing problems. Computers & Operations Research 76, 208–225.

Brun, Y., 2002. The four color theorem. MIT Undergraduate Journal of Mathematics, 21–28.

Carlsson, J., 2012. Dividing a territory among several vehicles. INFORMS Journal on Computing 24 (4), 565–577.

Carlsson, J., Delage, E., 2013. Robust partitioning for stochastic multivehicle routing. Operations Research 61 (3), 727–744.

Crainic, T., Perboli, G., Mancini, S., Tadei, R., 2010. Two-echelon vehicle routing problem: A satellite location analysis. Procedia: Social and Behavioral Sciences 2, 5944–5955.

Crainic, T., Ricciardi, N., Storchi, G., 2009. Models for evaluating and planning city logistics systems. Transportation Science 43 (4), 432–454.

Cuda, R., Guastaroba, G., Speranza, M., 2015. A survey on two-echelon routing problems. Computers & Operations Research 55, 185–199.

Daganzo, C., 1984. The distance traveled to visit n points with a maximum of c stops per vehicle: An analytic model and an application. Transportation Science 18 (4), 331–350.

Dellaert, N., Saridarq, F., Van Woensel, T., Crainic, T., 2016. Branch & price based algorithms for the two-echelon vehicle routing problem with time windows. Working paper, CIRRELT.

Franceschetti, A., Honhon, D., Laporte, G., Van Woensel, T., Fransoo, J., 2017. Strategic fleet planning for city logistics. Transportation Research Part B: Methodological 95, 19–40.

Grangier, P., Gendreau, M., Lehuede, F., Rousseau, L., 2016. An adaptive large neighborhood search for the two-echelon multiple-trip vehicle routing problem with satellite synchronization. European Journal of Operational Research 254, 80–91.

Groër, C., Golden, B., Wasil, E., 2010. A library of local search heuristics for the vehicle routing problem. Mathematical Programming Computation 2 (2), 79–101.

Hemmelmayr, V., Cordeau, J.-F., Crainic, T., 2012. An adaptive large neighborhood search heuristic for two-echelon vehicle routing problems arising in city logistics. Computers & Operations Research 39 (12), 3215–3228.

Huang, C., Shih, T., 1997. On the complexity of point-in-polygon algorithms. Computers & Geosciences 23 (1), 109–118.

Jepsen, M., Spoorendonk, S., Ropke, S., 2013. A branch-and-cut algorithm for the symmetric two-echelon capacitated vehicle routing problem. Transportation Science 47 (1), 23–37.

Langevin, A., Mbaraga, P., Campbell, J., 1996. Continuous approximation models in frieght distribution: An overveiew. Transportation Research Part B: Methodological 30 (3), 163–188.

Lei, H., Laporte, G., Liu, Y., Zhang, T., 2015. Dynamic design of sales territories. Computers & Operations Research 56, 84–92.

Li, H., Zhang, L., Lv, T., Chang, X., 2016a. The two-echelon time-constrained vehicle routing problem in linehaul-delivery ssystems considering carbon dioxide emissions. Transportation Research Part D: Transport and Environment 49, 231–245.

Li, H., Zhang, L., Lv, T., Chang, X., 2016b. The two-echelon time-constrained vehicle routing problem in linehaul-delivery systems. Transportation Research Part B: Methodological 94, 169–188.

Liu, R., Tao, Y., Hu, Q., Xie, X., 2017. Simulation-based optimisation approach for the stochastic two-echelon logistics problem. International Journal of Production Research 55 (1), 187–201.

Ouyang, Y., 2007. Design of vehicle routing zones for large-scale distribution systems. Transportation Research Part B: Methodological 41, 1079–1093.

Perboli, G., Tadei, R., Vigo, D., 2011. The two-echelon capacitated vehicle routing problem: Models and math-based heuristics. Transportation Science 45 (3), 364–380.

Qi, W., Li, L., Liu, S., Shen, Z., 2016. Shared mobility for last-mile delivery: Design, operational perscriptions and environmental impact. Working paper, University of California, Berkeley.

Santos, F., Mateus, G., da Cunha, A., 2015. A branch-and-cut-and-price algorithm for the two-echelon capacitated vehicle routing problem. Transportation Science 49 (2), 355–368.

Schneider, M., Stenger, A., Schwahn, F., Vigo, D., 2015. Territory-based vehicle routing in the presence of time-windows constraints. Transportation Science 49 (4), 732–751.

Song, L., Gu, H., Huang, H., 2016. A lower bound for the adaptive two-echelon capacitated vehicle routing problem. Journal of Combinatorial Optimization, 1–23.

Soysal, M., Bloemhof-Ruwaard, J., Bektas, T., 2015. The time-dependent two-echelon capacitated vehicle routing problem with environmental considerations. International Journal of Porduction Economics 164, 366–378.

Thompson, G., 1977. Hamiltonian tours and paths in rectangular lattice graphs. Mathematics Magazine 50 (3), 146–149.

Wang, Y., Ma, X., Xu, M., Liu, Y., Wang, Y., 2015. Two-echelon logistics distribution region partitioning problem based on a hybrid particle swarm optimization-genetic algorithm. Expert Systems with Applications 42, 5019–5031.

Wong, K., Beasley, J., 1984. Vehicle routing using fixed delivery areas. Omega 12 (6), 591–600.

Zhong, H., Hall, R., Dessouky, M., 2007. Territory planning and vehicle dispatching with driver learning. Transportation Science 41 (1), 74–89.

Appendix A. Further analytical results

The worst-case ratios with the Manhattan distance metric and square regions are shown below. The proofs are omitted as they are similar to those in Section 6.

Theorem 6. *Consider a network with* $m_a \times m_b$ *square regions and distances based on the Manhattan metric. We have* $\frac{Z_2^R}{Z_2^U} \leq 6$*, and this bound is tight.*

Theorem 7. *Consider a network with* $m_a \times m_b$ *square regions and distances based on the Manhattan metric. We have* $\frac{Z^R}{Z^U} \leq 6 + 2c_1/c_2$*, where* c_1 *is the first-echelon unit cost and* c_2 *is the second-echelon unit cost, and this bound is tight.*

We generalize the results above and consider the situation in which each region is a rectangle with width l_a and height l_b. Without loss of generality, we assume that $l_b > l_a$.

Theorem 8. *Consider a network with* $m_a \times m_b$ *rectangular regions. With the Manhattan distance metric, we have* $\frac{Z_2^R}{Z_2^U} \leq 4(l_a + l_b)/l_a + 2$*. With the Euclidean distance metric, we have* $\frac{Z_2^R}{Z_2^U} \leq 2\sqrt{l_a^2 + l_b^2}/l_a + 2$*. Both bounds are tight.*

Theorem 9. *Consider a network with* $m_a \times m_b$ *rectangular regions. With the Manhattan distance metric, we have* $\frac{Z^R}{Z^U} \leq 4(l_a + l_b)/l_a + 2 + 2c_1/c_2$*. With the Euclidean distance metric, we have* $\frac{Z^R}{Z^U} \leq 2\sqrt{l_a^2 + l_b^2}/l_a + 2 + 2c_1/c_2$*. Both bounds are tight.*