Does Locker Alliance Network Improve Last Mile Delivery Efficiency?

An Analysis using Prize-collecting Traveling Salesman Model

Quanmeng Wang, Guodong Lyu, Long He, Chung-Piaw Teo

Institute of Operations Research and Analytics, National University of Singapore
School of Business and Management, Hong Kong University of Science and Technology
NUS Business School, National University of Singapore
GW School of Business, George Washington University

The Locker Alliance Network (LAN) is a recent smart nation initiative introduced in Singapore for parcel pickup by customers, to improve the efficiency of last mile operation. This government facility is open to all logistic service providers (LSPs) operating in the country. With more parcels being shifted to locker stations, the number of visits to home locations could be drastically reduced, and the length of the delivery trips to homes will decrease. However, in the case of LAN, the carriers have to substitute these home deliveries with visits to the locker stations, on a separate delivery trip. The challenge is to determine the appropriate size of the LAN (number and location of locker stations), since having too many or too few of these stations may increase the total length of delivery trips instead. Furthermore, given the interoperable nature of the system, how should the government design the network of locker stations to serve all LSPs operating in the country?

In this paper, we develop a LAN model to address these questions. For a given delivery profile, say from an LSP, we first develop a model to jointly minimize the length of the two delivery trips (to home locations and to locker stations). We show that this can be formulated as a Prize-Collecting TSP problem, and reformulated as a mixed integer second-order cone problem under the logit choice model. We develop a heuristic policy with provable approximation guarantee based on its continuous relaxation, for this class of network design problem. Our analysis also shows that there is an optimal number of locker stations needed for efficient operations, beyond which the efficiency of the last mile operations will deteriorate.

More importantly, we can use the model to design the interoperable network for multiple LSPs, with possibly different delivery volumes, as long as they have similar footprint (i.e., identical probability density for delivery locations). We show that the network expands (almost) in a nested fashion in this case, i.e., the optimal networks for LSPs with smaller scale are (almost) contained in the optimal network for the larger LSPs. Therefore, the optimal interoperable network is very close to the optimal network for the largest LSP, and the optimal density of the locker network is dictated by the optimal density of the largest LSP operating in the country. Participation of the largest LSP is therefore crucial in any government-run interoperable system to increase the efficiency of last mile delivery operations.

Key words: Smart City; Last Mile Delivery; Parcel Locker; Traveling Salesman Problem; Sustainability
1. Introduction

In the last decade, we have witnessed a dramatic growth in the sales of e-commerce retailing, with global sales projected to increase from $4.3 trillion in a pandemic-fueled year 2020 to $7.4 trillion in 2025.\(^1\) Intensive e-commerce transactions drive a massive influx of parcel deliveries and greatly promote the development of the logistics industry. Even under the shadow of the coronavirus pandemic, the global parcel volume exceeded 131.2 billion in 2020, with around 4160 parcels being shipped per second.\(^2\) This in turn brings compounded challenges to the urban logistics landscape, especially to the last leg of parcel delivery process from sorting stations (or warehouses) to consumers.

To maintain a sustainable urban logistics ecosystem, one of the central issues is to improve the last mile delivery efficiency (Savelsbergh and Van Woensel 2016, Hasija et al. 2020). In the traditional model for parcel delivery, the logistics service providers (LSPs) employ a fleet of carriers for parcel shipment, and customers usually wait at home (or some pre-arranged locations) to receive their parcels. In recent years, many e-commerce giants such as Amazon, Wayfair, and Alibaba etc., have all strived to ameliorate the online purchase experiences of their customers by ensuring a timely and reliable parcel delivery service (e.g., same-day delivery or next-day delivery). LSPs therefore have to streamline their parcel delivery operations to manage stringently the deadlines of the parcel deliveries. While advanced techniques have been deployed to eliminate the information gap between parcel carriers and consumers, failed deliveries often arise due to unforeseen circumstances (e.g., consumer being away from home at time of delivery). In these cases, LSPs need to direct the parcels to some designated delivery points or reschedule another parcel delivery, further reducing the efficiency in the last mile operations.

The concept of parcel locker system has attracted a lot of attention from this community recently. Parcel locker refers to a collection of drop boxes with electronic verification function. Through the locker channel, the carrier first drops a parcel in one compartment of the locker station and gets a unique access code generated by the system. The customer then receives the code and picks up the parcel from the locker at her/his own convenience. The benefits of locker system is clear. This technology decouples the processes of parcel delivery and consumer pickup, effectively reducing the coordination cost within the last mile. The locker system could also consolidate the scattered door-to-door deliveries into a few locker stations, allowing LSPs to deliver in bulks to a common location. A variety of parcel locker systems have been implemented worldwide. For example, Amazon hub lockers are strategically installed near convenience stores and apartment buildings in more than 900


cities/towns across the United States.\(^3\) Hive-Box operates more than 150,000 parcel lockers in China, to support the real-time monitoring of more than 9 millions daily parcel deliveries.\(^4\) Deutsche Post DHL aims to expand the scale of the packstation network to 12,000 by 2023 in Germany.\(^5\) Along this line, the Singapore government recently launched the Locker Alliance Network (LAN) pilot program in two towns, in partnership with the largest LSP and the largest locker operator in the country (one for each town). The LAN is setup to explore ways to transform the last-mile delivery infrastructure of the nation, with the aim to “enhance the productivity and efficiency of Singapore’s urban logistics sector, ease the strain on manpower for LSPs, and enable more sustainable delivery options (with lower carbon emissions)”\(^6\). Due to the success in the pilot program (at least for the program run by the largest LSP in one of the towns), the government has formally rolled out a locker network of 1,000 stations in 2021, as part of the Singapore Smart Nation Initiatives to shape the future of urban logistics ecosystem.\(^7\) Interestingly, the locker network is deployed, owned and operated by Pick\! Network, a wholly owned subsidiary of the Infocomm Media Development Authority (IMDA), Singapore’s postal services regulator. The decision to remove the largest LSP as an operator is seen to be the best way to rope in other LSPs to join the program. However, this also set up the stage for the Pick\! network to compete directly with the proprietary locker systems run by some of the larger LSPs in the country.

With more parcels being shifted to locker stations, the number of visits to home locations could be drastically reduced, and hence the length of the delivery trip to homes will decrease. However, the couriers have to visit the locker stations on a separate delivery trip in the case of LAN. As a result, the total length of delivery trips may possibly increase instead. This gives rise to the first fundamental problem in our study:

**Does LAN improve the last mile delivery efficiency, by reducing the total length of parcel delivery trips?**

The answer in turn hinges on how we set up the locker network. To understand this, we provide an example to illustrate how the locker network affects the length of delivery trips for a given LSP.

**EXAMPLE 1.** On a typical working day (11 July 2018, Wednesday), a carrier from our partner LSP needs to deliver parcels to 176 (unique) delivery locations at Punggol, Singapore. If the locker stations are not installed, the length of home delivery tour is 17.69 km.\(^8\)

---

\(^3\) Amazon hub lockers. Retrieved from https://www.amazon.com/b?ie=UTF8&node=6442600011
\(^7\) Locker alliance pilot. Retrieved from https://www.lockeralliance.net/about-locker-alliance/pilot-trial
\(^8\) For the ease of illustration, we use Euclidean distance to measure the delivery distance between two locations.
Figure 1  Comparison of the delivery trips under different locker networks.

(a) Delivery Trips with Locker Network A  (b) Delivery Trips with Locker Network B

Notes. In both figures, the big-black circle indicates the locker station, while the small-orange dot presents the home location. The edges connecting two adjacent nodes represent the (Euclidean) traveling path.

For the sake of discussion, suppose customers will opt to pick up from lockers if there is one installed within 400 meters of her/his home location. If the lockers are installed at the four corners (cf. Figure 1(a)), 51 home locations would be covered by the locker network, and the length of delivery trip to the remaining home locations reduces to 12.79 km. However, the carrier needs to run a separate delivery trip (with length 6.58 km) to the four locker stations. As a result, the total length of delivery trips is unfortunately 9.50% higher than the single trip without any lockers. On the contrary, Figure 1(b) shows that, if four lockers are installed at the central areas, the locker network covers 103 home locations, and the length of delivery trip to the remaining home locations shrinks to 9.78 km. In this case, the length of trip to locker stations is 3.86 km, and the total length of two delivery trips is 13.64 km, which is 22.89% lower than the trip without lockers. Intuitively, the traveling distance reduces in case (b) because the locker system wipes out the majority of short but dense traveling paths, while inducing a reasonable detour across the lockers.

The LAN design problem is complicated by the fact that the network configuration not only affects customer parcel pickup behavior, but also has a direct impact on the parcel delivery trips to both locker stations and home locations (Peppel and Spinler 2022). As depicted in Example 1, a properly-configured locker network would reduce the volume of home delivery and hence reduce the length of delivery trip to the remaining home locations, at the expense of operating a separate trip to locker stations. If the increase in the length of delivery trip to lockers could be compensated by the reduction in the trip to home locations, the delivery efficiency of the last mile system would be improved. To design the locker network, we need to address the stochasticity issue in the last mile system, since the set of home delivery locations varies over time, and the consumer choice towards the locker could also be probabilistic. This leads to a non-trivial optimization problem.
From the view of parcel locker practitioners, a dense network provides convenient parcel pickup service (i.e., shortening the pickup distance), lower the parcels dwell time in the system, and free up more locker capacity due to the faster turnaround time.\(^9\) Therefore, as a rule of thumb, a denser locker network is preferred. However, our model and solution show that there is a notion of “optimal density”\(^9\) for the LAN, and adding more lockers beyond that will reduce the last mile delivery efficiency instead. Furthermore, we note that the volume of parcel deliveries has a direct impact on the design of the optimal locker network. In general, the size of the optimal locker network increases with parcel volume. As a nationwide infrastructure, the LAN aims to be interoperable for all LSPs in the country. However, each LSP may have a different preference towards the optimal design of locker network. The optimal network configured based on the delivery volume of one LSP may not be optimal for other LSPs. To address this interoperability issue, a natural approach is to set up the locker network by pooling the demand of all the LSPs. This raises the second crucial problem to be addressed:

**Given the interoperable nature of the LAN system, does pooling the demand of all LSPs lead to a denser network of locker stations for all players?**

Ideally, when we pool the demand of more LSPs together to build the locker network, the economies of scale should incentivize the operator to install more lockers, so that the density of locker network would be increasing if more LSPs join the interoperable system. Unfortunately this is not always the case, especially if the delivery footprint of the LSPs in the country is similar. We have obtained from our collaborators delivery information of six leading LSPs in the country, in the month of Feb in 2016. Figure 2 compares their delivery profile (to HDB, public housing blocks in Singapore) across different postal districts in the country. While their market shares differ slightly, the normalized delivery densities have similar geographical patterns. In fact, zooming down to one region such as the Punggol town, the Pearson correlation coefficients of their delivery densities are between 0.56 and 0.82, indicating strong positive correlations. While the LSPs differ in the volume of parcels carried, their footprint, in terms of normalized delivery densities, appears to be similar. This turns out to play an important role in the determination of the optimal density in any interoperable locker network.

For LSPs with similar footprints, we show that the optimal networks for LSPs with smaller delivery volumes are (almost) contained in the optimal networks for the larger LSPs. This result implies that the government can build the LAN based on the optimal network for the largest LSP, and other LSPs can configure their own (near) optimal locker networks using a subnetwork from this system. Therefore, the optimal density of the LAN in the country is essentially determined by the density of the largest LSP!

In terms of methodology, our key contributions are as follows:

- We investigate the last mile delivery efficiency from a locker network design perspective. For a given consumer locker choice model (to choose parcel pickup at the LAN or to opt for home delivery), we extend the Beardwood-Halton-Hammersley (BHH) theorem (Beardwood et al. 1959) to approximate the length of delivery trip to the remaining home locations in a stochastic environment. Along the way, we formulate the LAN design model as a prize-collecting TSP, using non-linear function to model the penalty for the prize-collecting cost component of the problem (i.e., to characterize the length of delivery trip to home locations) and using the TSP formulation for the remaining component (i.e., to quantify the length of delivery trip to locker stations).

- We use a set of locker usage data from the locker pilot program to calibrate a logit choice model. This model allows us to further reformulate the LAN design problem as a mixed integer second-order cone problem. We develop a heuristic algorithm with provable approximation guarantee based on its continuous relaxation so that we can solve fairly large scale (non-linear) prize-collecting TSP problems efficiently. Furthermore, we explicitly characterize the performance gap between the optimal solution and the continuous relaxation. Our analysis is built on the classic result for the prize-collecting TSP model with linear penalty cost (Bienstock et al. 1993, Goemans 2009). To our knowledge, this is the first bound for the non-linear penalty case.
• Given the interoperable nature of the LAN system, we show that pooling the demand from many LSPs together will not lead to substantial increase in the network footprint. To see this, we use well-known results in the field of submodular approximation and parametric optimization, to demonstrate that a nested solution to an approximate model could be near-optimal for the original LAN design problem. This result implies that the optimal network expands (almost) in a nested fashion with the increase of delivery volume, and hence the optimal density of the LAN is essentially determined by the LSP with the largest scale in the country.

• We use our model to examine whether the LAN can improve the last mile delivery efficiency, using a set of data from the pilot program. We show that the delivery efficiency actually dropped under the existing LAN in Punggol, i.e., the total length of delivery trips increased. Our analysis reveals that the impact of LAN on last mile delivery efficiency is influenced largely by both consumer choice behavior (preferences for lockers vis-a-viz home deliveries) and demand volume of the LSPs. Our numerical results provide further validation to the phenomenon that the optimal locker network is nested for LSPs with different volumes. This phenomenon is generally true, and even if the solution is not nested, a significant number of locker locations overlap with each other.

The rest of the paper is organized as follows. We review relevant literature in Section 2. In Section 3, we describe the LAN design problem. In Section 4, we analyze the LAN design problem under logit choice model and investigate the impact of the interoperable nature of the LAN system. We describe the Singapore LAN pilot program in details and evaluate the performance of our network design model in Section 5. Section 6 concludes the paper. The technical proofs and numerical experiments are relegated to Appendix A and B, respectively.

2. Literature Review

Rohmer and Gendron (2020) contained a thorough description on the design of a locker system, and also relevant research directions from an operations research perspective. We refer the interested readers to this work for details. Here, we restrict our discussion to three main streams of study in the relevant literature: (i) benefits of parcel lockers, (ii) design of parcel locker network, and (iii) traveling salesman problem.

Benefits of Parcel Lockers. The locker system is widely seen to be a solution to increase delivery efficiency and reduce potential carbon emissions in urban cities (Hasija et al. 2020). For example, Lemke et al. (2016) argued that the locker system is environment-friendly, and quantified the benefit of locker system in terms of traveling distance reduction. Iwan et al. (2016) showed that the locker system could reduce the number of visits to the city area (e.g., reducing failed delivery and parcel return). Ranieri et al. (2018) documented that the reduction in the externalities cost induced by last mile delivery activities, such as traffic congestion and pollution. These benefits are also documented...
in (Edwards et al. 2009, 2010) as well as (Peppel and Spinler 2022). Schnieder et al. (2021) provided a formal comparison between the emission models of parcel lockers and home delivery. On a different note, Vakulenko et al. (2018) explored the customer value created from the locker network so that the operators could better adjust their parcel delivery service.

**Design of Parcel Locker Network.** This involves two key decisions, including the number and the location of the locker stations. A variety of approaches have been applied to construct locker networks, to minimize the operating cost or maximize the demand coverage. For example, Wu et al. (2015) used a set of public transit data to estimate the crown pattern, and used the insights to locate the locker stations at dense crowded places that are convenient for consumers. Deutsch and Golany (2018) formulated the network design problem as the classic uncapacitated facility location problem. In particular, the model assumed that consumers always opt for the nearest locker to pick up parcel, while the proportion of demand diverted to lockers other than their home decreases in the traveling distance.

The locker network design hinges on the consumer perception/utility of using lockers. Notably, the logit choice model is a natural option to formulate the consumer parcel pickup choice between locker and home, and this model was adopted to design the locker network by Lin et al. (2020). Lin et al. (2022) further introduced a threshold luce model, which allows for 0 probability of using lockers and generalizes the logit choice model, in the locker location problem. Lyu and Teo (2022) also considered logit choice model in the locker network design problem and used a set of locker usage data to calibrate the choice model. In particular, Lyu and Teo (2022) addressed the issue of heterogeneity between past location strategy and demand observed, and developed a facility location model for locker network design, without knowing the underlying hidden states in consumer choice model. However, the aforementioned works focused on maximizing the utilization of the locker network, and did not consider its impact on the delivery trips.

**Travelling Salesman Problem.** This paper is also related to the TSP model since the objective function of our LAN design model is to minimize the length of parcel delivery trips. More concretely, our network design problem is philosophically similar to the location-routing problem in the sense that the components of locker facility location and parcel delivery process are coupled (e.g., Baldacci et al. 2011, Janjevic et al. 2019, Pan et al. 2021). The location-routing problem is in general hard to solve. Our problem is even more complicated since the design of locker network affects customer behavior, and hence affects the collection of home delivery locations. Therefore, instead of formulating the complicated parcel delivery processes directly, we integrate the logit choice model and the BHH theory (Beardwood et al. 1959) to characterize the length of delivery trip to the remaining home locations, and treat the delivery trip to locker stations as a classic TSP. This allows us to examine the impact of locker network on the length of delivery trips explicitly. Interestingly, the integrated model
turns out to be a price-collecting TSP with non-linear penalty cost. The classic price-collecting TSP model with linear penalty cost has been investigated in the literature (e.g., Bienstock et al. 1993, Goemans 2009). However, to our knowledge, few works shed light on the non-linear penalty case.

Finally, we remark that the BHH theory has been widely adopted to approximate the expected length (or traveling time) of a TSP tour when the delivery locations are sampled randomly from an absolutely continuous probability density function over the service region. In the urban logistics context, Carlsson and Song (2018) applied this approximation technique to formulate the drone delivery model, and Carlsson et al. (2018) showed that this approach can also be used to study the service region partition problem. Furthermore, Stroh et al. (2021) leveraged on this approximation approach to characterize the parcel dispatch time function and design tactical models for the same-day delivery problem. Different from the aforementioned works, this classic theorem cannot be directly applied to formulate our LAN design model since the density function of the delivery locations changes after the installation of parcel lockers. This motivates us to extend the classic BHH theorem to solve our LAN design problem.

3. Locker Alliance Network Design Model

We start by formally introducing the problem setting, in which the LSP carrier is responsible for the parcel deliveries in a connected compact planar region $\mathcal{A}$. We assume that, on a daily basis, a fixed number of $N$ customers require delivery service, but the location of each customer is independently and identically generated according to a continuous density function $f(\cdot)$. Let $\tilde{z}_i$ denote the (random) location $i$ for $i = 1, 2, \ldots, N$, and $\mathcal{Z} := \{\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_N\}$ denote the set of random delivery points. There are in total $K$ candidates of locker locations, and the candidate set is denoted by $\mathbf{V} := \{v_1, v_2, \ldots, v_K\}$. The binary decision variable $x_k = 1$ indicates that the locker is installed at location $v_k$, and 0 otherwise. We denote $\mathbf{x} := (x_1, x_2, \ldots, x_K)$ as the vector of decision variables. With a slight abuse of notation, we let $\mathbf{V}(\mathbf{x}) := \{v_k \mid x_k = 1, k = 1, 2, \ldots, K\}$ represent the locker network, i.e., the set of locations with lockers being installed. Based on the deployed network $\mathbf{V}(\mathbf{x})$ and the home location, each consumer randomly chooses between delivery-to-locker and delivery-to-home. The LSP then arranges two separate trips for all deliveries to lockers and homes, respectively. We seek to minimize the length of two trips through a proper locker network design. Three key components are involved in the network model, including (i) consumer choice between parcel pickup from locker and home; (ii) traveling distance across the installed locker stations; and (iii) traveling distance to deliver the parcels to the remaining home locations. We detail the model formulation in what follows.

The leading LSPs in Singapore exploit a zonal assignment mechanism for parcel delivery. Under this mechanism, each carrier is mainly responsible for one particular zone and designs a TSP tour for parcel delivery each day.
Consumer Choice Model. Given the installed locker network $V(x)$ and realized customer home location $z \in A$, the consumer opts to locker $v_k \in V(x)$ for parcel pickup with probability $g(z, v_k, x) \in [0,1]$. Note that our model allows the probability to be a generic function of the realized customer location $z$, the locker location $v_k$, and the installation decision $x$. We let $g(z, 0, x) \in [0,1]$ denote the probability that the consumer from location $z$ stays with home delivery service, indexed by 0. We have

$$g(z, 0, x) = 1 - \sum_{v_k \in V(x)} g(z, v_k, x).$$

We do not enforce any functional form for $g(\cdot)$ in this section. Instead, we assume that the probability of choosing home delivery is always positive and continuous in $z$ regardless of the network deployed.

**Assumption 1.** For any locker network $V(x)$, the function $g(z, 0, x) > 0$ is absolutely continuous in $z$ within $A$.

Delivery Trip to Locker Stations. We note that the volume of parcels diverted to the lockers is in general much larger than the number of locker stations, and hence we can naturally assume that the carrier needs to visit all the installed locker stations in the network $V(x)$. While there are other measures for the efficiency of this portion of the delivery process, we use the classical TSP tour length to model this problem.\footnote{There are of course other variants that could be implemented. Some LSPs use timing or rate for the deliveries to measure efficiency, others may want to use the actual physical tour length chalked up by the delivery agents to model this problem.} To this end, we apply the classic TSP model to minimize the delivery trip to the locker stations. For the ease of modeling convenience, we assume further that the “depot” locker, indexed by $q$, is installed in the network (i.e., $x_q = 1$).\footnote{This assumption is introduced to formulate the TSP model. In fact, we can sequentially choose one locker station from the candidate set $V$ to be fixed as the “depot”, and compare the performance of each solution (Bienstock et al. 1993). The minimum gives rise to the optimal decision.} and let this locker be the starting point of the TSP tour. We denote $d_e$ as the traveling distance of arc $e \in E$ between two locker stations, where $E$ indicates the set of arcs in the fully-connected undirected network. The binary decision $y_e = 1$ indicates that the arc $e$ is in the network; and 0 otherwise. More concretely, let $\text{TSP}(x)$ denote the minimal traveling distance across the set of installed lockers $V(x)$, and we can formulate the TSP as follows:

$$\text{TSP}(x) := \min \sum_{e \in E} d_e y_e$$

s.t.

$$\sum_{e \in \delta(j)} y_e = 2x_j, \ \forall j = 1, 2, \ldots, K$$

$$\sum_{e \in \delta(S)} y_e \geq 2x_j, \ \forall j \in S, \ S \subseteq \{1, 2, \ldots, K\} \setminus \{q\}$$

$$y_e \in \{0, 1\}, \ \forall e \in E$$

\footnote{There are of course other variants that could be implemented. Some LSPs use timing or rate for the deliveries to measure efficiency, others may want to use the actual physical tour length chalked up by the delivery agents to model this problem.}
where the first set of constraints ensures that a candidate locker \( j \) is visited if and only if it is installed in the network, and the second set of constraints eliminates all the possible subtours. The notions \( \delta(j) \) and \( \delta(S) \) represent respectively the set of arcs that are connected to the node \( j \) and set \( S \). Conventionally, the subtour elimination constraint is implemented as a lazy constraint and we follow the computational strategy by Pferschy and Staněk (2017) to solve the TSP.

**Delivery Trip to Home Locations.** Characterizing the delivery length to the remaining home locations is the most challenging part of the problem. Ideally, we would measure it by the shortest tour through all home delivery locations. However, this comes with two critical difficulties: (i) the number of home delivery is usually large, which results in large scale TSP; and (ii) when designing locker network, the future (daily) home delivery locations are random, which results in a stochastic TSP. These make the problem intractable. We opt to measure the efficiency of this process using the seminal BHH theorem, originally derived by Beardwood et al. (1959), which depicts the expected length of a TSP tour across a sequence of delivery locations, based on the distribution from which the locations are randomly generated.

**Theorem 1.** (Beardwood et al. 1959) Suppose that \( \tilde{Z} := \{\tilde{z}_1, \tilde{z}_2, \ldots\} \) is a sequence of random delivery locations independent and identically generated according to an absolutely continuous probability density function \( f(\cdot) \) defined on a compact planar region \( A \). Let \( \tilde{Z}_M \) denote the subset of the first \( M \) delivery locations. Then the length of the optimal TSP tour across the first \( M \) locations, denoted by \( TSP(\tilde{Z}_M) \), satisfies

\[
\lim_{M \to \infty} \frac{TSP(\tilde{Z}_M)}{\sqrt{M}} = \beta(A) \int_A \sqrt{f(z)} dz, \quad \text{almost surely,}
\]

where \( \beta(A) \) is a constant that depends on the shape of \( A \).

Note that the exact value of \( \beta(A) \) is unknown, but it falls in the range \([0.6250, 0.9204]\) and can be further calibrated from data (cf. Beardwood et al. 1959, Carlsson et al. 2018). The BHH theorem has been extensively investigated and applied to approximate the length of TSP tours in a variety of urban logistics problems (e.g., Carlsson and Song 2018, Carlsson et al. 2018, Liu et al. 2021, Stroh et al. 2021). However, this theorem cannot be directly applied to our LAN case because of the following two key differences: (i) the number of customers who choose home delivery becomes a random variable, which is different from the integer \( M \) as stated in Theorem 1; and (ii) the location density of customers opting for home delivery is different from original density \( f(\cdot) \), as customers are more likely to choose home delivery if all installed lockers are far away. Therefore, we have to modify BHH theorem to capture these two important features involved in the LAN design problem.

We view the random customer choice between locker and home as a stochastic filtering process. Given the original density function \( f(\cdot) \), and conditional probability of choosing home delivery
Bernoulli filtering, the reshaped density function 
\[ \hat{M} \sim \text{tour across these locations, denoted by TSP} \]
first \[ \tilde{\hat{M}} \] probability density function random delivery locations independent and identical ly generated according to an absolutely continuous process is equivalent to generating \[ N \] customers who choose home delivery. It is easy to see for any integer \[ N \]
the customer’s location and locker network \[ V \] density \[ f \]
Recall that the original process unfolds as follows: \( (i) \) the \( N \) customer locations realize according to density \( f(\cdot) \); \( (ii) \) the customer determines if s/he chooses home delivery or locker delivery based on the customer’s location and locker network \( V(\mathbf{x}) \); and \( (iii) \) the LSP makes a delivery trip across all customers who choose home delivery. It is easy to see for any integer \( N \) and locker network \( V(\mathbf{x}) \), the process is equivalent to generating \( N \) locations based on the density \( \hat{f}(\cdot, \mathbf{x}) \) and performing a delivery trip across the first \( \tilde{M} \sim \text{Bin}(N, p(\mathbf{x})) \) locations. We let \( \tilde{Z}_{\tilde{M}} \) denote the subset of first \( \tilde{M} \) points and \( TSP(\mathbf{x}, \tilde{Z}_{\tilde{M}}) \) denote the optimal TSP tour length across these points. We now state our first key result regarding this random TSP problem.

**Theorem 2.** Given the locker network \( V(\mathbf{x}) \), and suppose \( \tilde{Z} := \{\tilde{z}_1, \tilde{z}_2, \ldots\} \) is a sequence of random delivery locations independent and identically generated according to an absolutely continuous probability density function \( \hat{f}(\cdot, \mathbf{x}) \) defined on a compact planar region \( \mathcal{A} \). Let \( \tilde{Z}_{\tilde{M}} \) be the subset of the first \( \tilde{M} \) points, where \( \tilde{M} \sim \text{Bin}(N, p(\mathbf{x})) \) is independent of \( \tilde{Z} \). Then the length of the optimal TSP tour across these locations, denoted by \( TSP(\mathbf{x}, \tilde{Z}_{\tilde{M}}) \), satisfies

\[
\lim_{{N \to \infty}} \frac{TSP(\mathbf{x}, \tilde{Z}_{\tilde{M}})}{\sqrt{p(\mathbf{x})N}} = \beta(\mathcal{A}) \int_{\mathcal{A}} \sqrt{\hat{f}(z, \mathbf{x})}dz, \text{ almost surely,}
\]

where \( \beta(\mathcal{A}) \) is a constant that depends on the shape of \( \mathcal{A} \), while \( p(\mathbf{x}) \) and \( \hat{f}(\cdot, \mathbf{x}) \) are defined respectively by Equation (3) and (4).

As stated in Theorem 2, the summation of \( N \) i.i.d. Bernoulli random variables admits the concentration phenomenon, i.e., \( \tilde{M} \) concentrates around its mean \( p(\mathbf{x})N \) with high probability when \( N \) is large. This motivate us to apply Equation (2) in Theorem 1 by substituting \( \tilde{M} \) with its mean \( p(\mathbf{x})N \) and use the reshaped density function \( \hat{f}(z, \mathbf{x}) \) for the original one. We provide a numerical experiment to validate the effectiveness of Theorem 2 in Appendix B.1 and detail a formal proof of Theorem 2 in Appendix A.1.
When \( N \) is sufficiently large, Theorem 2 implies that we can fairly approximate the length of the optimal stochastic TSP tour by a deterministic function. Recall that we raised two concerns about using the original BHH formula before the development of our extension: (i) the number of customers who opt for parcel pickup from lockers is affected by the network configuration; and (ii) the location density of customers who remain for home delivery is also affected by the network configuration. We remark that the approximation incorporates the aforementioned two features into consideration, i.e., the formula involves both \( p(x) \) and \( \hat{f}(\cdot, x) \). Therefore, we apply Theorem 2 to characterize the length of delivery trip to home locations,

\[
TSP(x, \tilde{Z}^{M}) \approx \sqrt{N} \beta(A) \int_A \int \sqrt{f(z)g(z, 0, x)} \, dz.
\]

**Locker Network Design Model.** We have established how the network configuration \( V(x) \) affects the delivery trips to both locker stations and home locations. Given a candidate set of \( K \) potential lockers, we can formulate the LAN design problem as:

\[
\begin{align*}
\text{(P)} \quad & \min_{x \in \{0, 1\}^K} TSP(x) + \sqrt{N} \beta(A) \int_A \int \sqrt{f(z)g(z, 0, x)} \, dz,
\end{align*}
\]

where the objective function is derived from Equations (1) and (6) directly. In model (P), we implicitly assume that two separate delivery trips are equally weighted, while our model allows the decision maker to choose the weight function flexibly. Note that the locker is composed of different compartments, and it is flexible to adjust the locker capacity based on the demand captured. Hence, we do not enforce a capacity constraint in our model.

Under an appropriate consumer choice model, the second term in the objective of model (P) is non-increasing in the number of lockers being installed, while the first term is clearly non-decreasing. Therefore, the visit to one more locker will possibly increase the traveling cost across the locker stations, but might collect higher “prize” by diverting more parcels to the locker network. In this way, our LAN design model (P) is indeed a prize-collecting TSP, which contains nonlinear prize-collecting cost component (i.e., the second term) and a standard TSP formulation for the first term. The class of prize-collecting TSPs is generally NP-hard due to the TSP component. For the case of linear prize-collecting term, Bienstock et al. (1993) and Goemans (2009) designed provable approximation algorithms by using the continuous relaxation of the problem. Along this direction, the problem of interest is to investigate the structure of this non-linear model and provide an efficient solution approach, under certain choice function \( g(\cdot) \).

**Remark 1.** We remark that a “depot” locker \( q \) is fixed to formulate the TSP(\( x \)) component in model (P). With a slight abuse of notation, we let \( P(q) \) represent a variant of model (P), in which we enforce \( x_q = 1 \) as the input, and let \( H^*(q) \) denote the optimal value of this model. We
also consider the case when we only install the depot locker, i.e., without formulating the TSP tour. We set this case as the benchmark and normalize the traveling cost to a single locker as 0. Notably, this case usually happens when the attractiveness of locker network is substantially lower than home delivery, and hence the delivery cost is negligible. To obtain the optimal network solution to model (P), we sequentially solve each variant model (P(q)) and search the minimal value from \{H^*(1), H^*(2), \ldots, H^*(K)\}. In this way, the optimal value of model (P), denoted by \(H^*\), can be represented as \(H^* = \min_{q=1,2,\ldots,K} H^*(q)\).

4. Network Design with Logit Choice Model

Note that the substitution effect exists across the installed lockers (Lyu and Teo 2022), and also among other types of retailing location problems (e.g., Glaeser et al. 2019). In analog to Lyu and Teo (2022), we also adopt the logit choice model to formulate the consumer choice toward different parcel pickup options and capture the possible substitution effect raised in the locker network. In this section, we analyze the LAN design problem under the logit choice model.

For a particular consumer from location \(z \in A\), we denote the attraction of locker \(v_k \in V\) to this consumer as \(\theta(z, v_k)\), and the attraction of home delivery is denoted by \(\theta(z, 0)\), where \(\{0\}\) can be treated as the outside option. Given a set of potential locker candidates \(V\), the consideration set for parcel delivery is represented by \(\{0\} \cup V\). In this way, we can specify the probability that a customer from location \(z\) opts for the installed locker network or stays with home delivery, respectively, as follows:

\[
\begin{align*}
g(z, v_k, x) &= \frac{\theta(z, v_k)x_k}{\theta(z, 0) + \sum_{j=1}^{K} \theta(z, v_j)x_j}, \quad \forall v_k \in V \\
g(z, 0, x) &= \frac{\theta(z, 0)}{\theta(z, 0) + \sum_{j=1}^{K} \theta(z, v_j)x_j}.
\end{align*}
\]

(7)

Note that the attraction/utility function \(\theta(z, v_k)\) and \(\theta(z, v_0)\) can be calibrated from data. Clearly, as long as both terms are positive and continuous in \(z\) for \(\forall z \in A\), then Assumption 1 holds in the LAN design problem under the logit choice model. Plugging Equation (7) into model (P), we can derive the network model under logit choice as follows:

\[
(P\text{-Logit}) \quad \min_{x \in \{0,1\}^K} \text{TSP}(x) + \sqrt{N} \beta(A) \int_{A} \int_{x} f(z) \left\{ \frac{\theta(z, 0)}{\theta(z, 0) + \sum_{j=1}^{K} \theta(z, v_j)x_j} \right\} dz.
\]

The model \((P\text{-Logit})\) seems to be complicated, in part due to the non-linear “prize-collecting” component. Interestingly, in Section 4.1, we show that this model can be further re-formulated as a mixed-integer second-order cone problem (MI-SOCP) so that we can solve mild cases of the LAN design problem by using existing solvers such as Gurobi. For the problem with large scale, we consider the continuous relaxation of this MI-SOCP and develop a provable approximation algorithm in
Section 4.2. More importantly, we analyze the impact of $N$ on the optimal network design solution to (P-Logit), and discuss the managerial insights obtained from building an interoperable locker system in Section 4.3.

4.1. Exact Reformulation

To provide a clear reformulation, we introduce a set of auxiliary variables $w(z)$ and $\rho(z)$ for each $z \in \mathcal{A}$ and enforce the following inequality:

$$w(z) \geq \frac{1}{\rho(z)} \geq \sqrt{\frac{\theta(z,0)}{\theta(z,0) + \sum_{j=1}^{K} \theta(z,v_j)x_j}}. \quad (8)$$

In this way, we can replace the square root term by $w(z)$ in the objective, and further formulate the fractional term (i.e., the choice probability) into a second-order cone constraint. We formally describe the MI-SOCP reformulation in Proposition 1.

**Proposition 1.** The model (P-Logit) can be exactly reformulated as the following mixed-integer second-order cone problem (P-MI-SOCP):

$$\begin{align*}
\text{(P-MI-SOCP) } & \min \ TSP(\mathbf{x}) + \sqrt{N} \beta(\mathcal{A}) \int_{\mathcal{A}} \sqrt{f(z)} w(z) dz \\
& \text{s.t. } \begin{pmatrix} \theta(z,0) + \sum_{j=1}^{K} \theta(z,v_j)x_j & \rho(z) \\ \rho(z) & 1/\theta(z,0) \end{pmatrix} \succeq 0, \forall z \in \mathcal{A} \\
& \begin{pmatrix} w(z) \\ 1 \rho(z) \end{pmatrix} \succeq 0, \forall z \in \mathcal{A} \\
& \rho(z) \geq 1, 0 \leq w(z) \leq 1, \forall z \in \mathcal{A} \\
& x_k \in \{0,1\}, \forall k = 1,2,\ldots,K.
\end{align*}$$

Furthermore, the continuous relaxation of this problem is a convex programming problem.

In model (P-MI-SOCP), the first and second set of constraints together reshape the fractional choice probability term into a second order cone format. The third set of constraints comes from Equation (8).

Notably, this reformulated model is an infinite dimensional convex programming problem due to the double-integration of all the possible locations $z \in \mathcal{A}$. For the ease of computation, a natural way to solve the problem is to discretize the planar region $\mathcal{A}$ into a collection of $R$ patches with the same area $\Delta$ (cf. Carlsson et al. 2018). We assume that the consumers from each patch $r = 1,2,\ldots,R$ are homogeneous, and denote $f(r)$ as the delivery density of this patch. With a slight abuse of notation, we let $\theta(r,0)$ and $\theta(r,v_k)$ denote respectively the attraction of home delivery and picking up from
locker \( v_k \) for the consumers from patch \( r \). The auxiliary variables \( w(r) \) and \( \rho(r) \) are also updated accordingly. In this way, model (P-MI-SOCP) can be further simplified to a Discretized version:

\[
(P\text{-MI-SOCP-D}) \min \ TSP(\mathbf{x}) + \sqrt{N} \beta(\mathcal{A}) \Delta \sum_{r=1}^{R} \sqrt{f(r)} w(r)
\]

\[
s.t. \left( \begin{array}{c}
\theta(r,0) + \sum_{j=1}^{K} \theta(r,v_j)x_j \\
\rho(r)
\end{array} \right) \geq 0, \forall r = 1, 2, \ldots, R \\
\left( \begin{array}{c}
w(r) \\
\rho(r)
\end{array} \right) \geq 0, \forall r = 1, 2, \ldots, R \\
\rho(r) \geq 1, 0 \leq w(r) \leq 1, \forall r = 1, 2, \ldots, R \\
x_k \in \{0, 1\}, \forall k = 1, 2, \ldots, K.
\]

In total, there are \((2R)\) continuous decision variables and \(O(K + K^2)\) binary decision variables, with \(O(K^2)\) variables being suppressed in the TSP(\(\mathbf{x}\)) component. We can solve the model to optimality for mild-scale problems, and construct the optimal network structure.

4.2. Continuous Relaxation and Approximation Bound

For large scale problems, we leverage the structure of the continuous relaxation of the problem to construct an approximation algorithm to solve the problem more efficiently. To be specific, the continuous relaxation of model (P-MI-SOCP) is defined by relaxing all the binary decisions to continuous variables. We denote the optimal value of problem (P-MI-SOCP) as \(H^*\), and denote \(H_C\) as the optimal value of the continuous version.

In the Singapore LAN case, we note that the home delivery is still a dominant option, compared to picking up parcels from the locker network. Through the pilot study, the proportion of home delivery was consistently above 80% between December 2018 and October 2019. This observation facilitates us to further bound the choice probability \(g(z,0,\mathbf{x})\), which would be a key ingredient for us to analyze to derive the approximation algorithm.

**Assumption 2.** The percentage of parcels delivered to home locations, regardless of the network \(V(\mathbf{x})\) deployed, is always lower bounded by a constant \(\Gamma \in (0, 1)\), i.e.,

\[
g(z,0,\mathbf{x}) = \frac{\theta(z,0)}{\theta(z,0) + \sum_{v \in V(\mathbf{x})} \theta(z,v)} \geq \Gamma, \forall V(\mathbf{x}),
\]

or equivalently,

\[
\sum_{v \in V(\mathbf{x})} \theta(z,v) \leq \theta(z,0)(1 - \Gamma)/\Gamma, \forall V(\mathbf{x}).
\]

We first solve the continuously relaxed problem instance and derive the optimal solution \(\mathbf{x}^C \in [0, 1]^K\). Define the following threshold “rounding” mechanism, and let

\[
x_k^{A}(\gamma) = \begin{cases} 
1, & \text{if } x_k^C \geq \gamma \\
0, & \text{otherwise,}
\end{cases}
\]
where $x^A(\gamma)$ denotes the approximate solution under the rounding threshold $\gamma \in (0, 1]$. We let $H^A(\gamma)$ denote the objective value under the solution $x^A(\gamma)$. Now, we are ready to state the performance guarantee of this approximate solution.

**Theorem 3.** For a given instance of problem (P-MI-SOCP) and a rounding threshold $\gamma$, we have

$$H^C \leq H^* \leq H^A(\gamma) \leq \max \left\{ \sqrt{1 + \frac{1 - \Gamma}{\Gamma}}, \frac{3}{2\gamma} \right\} H^C.$$ 

Our approximation algorithm is built on the results from Bienstock et al. (1993) and Goemans (2009) for linear price-collecting TSP. We extend their result to a special class of non-linear prize-collecting TSP. Note that the first term $\sqrt{1 + \frac{1 - \Gamma}{\Gamma}}$ is increasing in $\gamma$, while the second term $\frac{3}{2\gamma}$ is clearly decreasing. Therefore, for a given $\Gamma$, there exists an optimal $\gamma^* \in (0, 1]$ that optimizes the approximation bound $\max \left\{ \sqrt{1 + \frac{1 - \Gamma}{\Gamma}}, \frac{3}{2\gamma} \right\}$. To see this, we vary the value of $\Gamma$ from 0 to 1, and numerically compute the best approximation bound by choosing an appropriate $\gamma^*$. We visualize the result in the following Figure 3.

![Figure 3 Approximation bound over different $\Gamma$'s.](image)

We observe that the approximation bound could be fairly loose when $\Gamma$ is extremely small, i.e., when home delivery nearly has no attraction to the customers, it is not easy to balance the TSP cost over the installed lockers and the saving from home delivery. However, when $\Gamma$ exceeds a threshold $\frac{4}{9}$, the approximation bound reduces to $\frac{3}{2}$, which hits the well-known $\frac{3}{2}$ approximation bound for the classic TSP problem (Christofides 1976, Wolsey 1980). Therefore, when the home delivery is the default option for the consumers, the performance gap of our approximation algorithm mainly comes from locker TSP part. We formalize this result in the following corollary.
Corollary 1. Suppose $\Gamma \geq \frac{4}{9}$. The optimal rounding threshold $\gamma^* = 1$ and our approximation algorithm is $\frac{3}{2}$-optimal, i.e.,

$$\frac{H^A(\gamma^*)}{H^*} \leq \frac{3}{2}.$$ 

While Theorem 3 and Corollary 1 provide theoretical guarantees for our approximation algorithm, the actual performance could be better than the performance bound. We provide synthetic numerical experiments to evaluate the actual performance of our approximation algorithm in Appendix B.2.

4.3. Interoperable Network for All LSPs

In the above, we have addressed the LAN design problem for a given LSP. In this section, we move forward to address the network design problem for multiple LSPs by considering the interoperable nature of the LAN system. Note that each LSP may have a different preference towards the optimal design of locker network. Even if all LSPs serve the same population, i.e., share the same delivery density $f(\cdot)$ and customer behavior $\theta(\cdot)$, they normally differ in the volume of deliveries $N$. As a result, the optimal network configured based on the delivery volume of one LSP may not be optimal for other LSPs, and the government needs to build the network of facilities given that $N$ may vary for different LSPs. To understand the interoperability of the LAN system, we study how does the optimal network change with respect to $N$.

We combine results in two diverse areas – submodular approximation and parametric optimization – to enhance our insights into this issue. We claim that, although problem (P-Logit) is not submodular, the objective function can be fairly approximated by a submodular function with decreasing difference property. As a result, we can construct a nested optimal solution by solving the approximated LAN design problem, and demonstrate that the nested solution is near-optimal for the original LAN design problem. This result implies that the optimal networks for LSPs with smaller scales are (almost) contained in the optimal networks for the larger LSPs. Therefore, the government can build the LAN based on the delivery volume of LSP with the largest scale $N$.

**Submodular Approximation.** The objective function of (P-Logit) consists of the TSP component and price collecting component. While both terms are not submodular in general, we show that the summation of two terms can be approximated by a submodular function with bounded gap. To be concrete, we first utilize the strict supermodularity of $\sqrt{g(z,0,x)}$ to compensate the non-submodular structure of TSP(·) function, and then show that the $\sqrt{g(z,0,x)}$ function can be approximated by a linear function. We formalize the first step in the following proposition.

**Proposition 2.** Suppose $\theta(z,v_k) > 0$ for any $z$ and $v_k$, then

1. $\sqrt{g(z,0,x)}$ is strictly supermodular in $x$ for any given $z$;
2. $\exists \lambda^* > 0$ such that $T(x) := TSP(x) - \lambda^* \int_{z \in A} \sqrt{f(z)} \sqrt{g(z,0,x)} dz$ is submoular in $x$. 

Electronic copy available at: https://ssrn.com/abstract=4273960
In this way, we can rewrite the objective function of (P-Logit) as follows:

\[
TSP(x) - \lambda^* \int_{z \in A} \sqrt{f(z)} g(z, 0, x) dz + \left( \lambda^* + \sqrt{N} \beta(A) \right) \int_{z \in A} \sqrt{f(z)} g(z, 0, x) dz,
\]

where the first component \( T(x) \) is now submodular in \( x \). Next, we focus on the second component of Equation (9). As stated in Assumption 2, we assume that the proportion of home delivery service is bounded below by \( \Gamma \), or \( \sum_{j=1}^{K} \theta(z, v_j) \leq \theta(z, 0)(1 - \Gamma) / \Gamma \). This motivates us to approximate \( \sqrt{g(z, 0, x)} \) by the following linear function:

\[
\sqrt{g(z, 0, x)} = \frac{\theta(z, 0)}{\theta(z, 0) + \sum_{j=1}^{K} \theta(z, v_j)x_j} \approx G(z, x) := 1 - \frac{\Gamma(1 - \sqrt{T})}{1 - \Gamma} \sum_{j=1}^{K} \theta(z, v_j)x_j.
\]

Next, we characterize the approximation gap in the following Lemma 1.

**Lemma 1.** The ratio of \( G(z, x) \) over \( \sqrt{g(z, 0, x)} \) is upper bounded by a parametric constant, i.e.,

\[
\sqrt{g(z, 0, x)} \leq G(z, x) \leq C(\Gamma) \sqrt{g(z, 0, x)},
\]

where \( C(\Gamma) = \frac{2\sqrt{3(1-\Gamma^2)}}{9(1-\Gamma)} \sqrt{\frac{1-\Gamma^2}{\Gamma(1-\sqrt{T})}} \).

The proof is deferred to Appendix A.5. To make the discussion clearer, we numerically evaluate the approximation gap. Consider \( \Gamma \in \{80\%, 70\%, 60\%, 50\%\} \), we have the approximation gap \( C(\Gamma) \in \{1.005, 1.012, 1.025, 1.046\} \), respectively, i.e., the gap is within 5% for \( \Gamma \geq 50\% \). A more detailed comparison between \( \sqrt{g(z, 0, x)} \) function and the proposed approximation \( G(z, x) \) is detailed in Appendix B.3.

**Parametric Submodular Optimization.** We reformulate the objective function as Equation (9) and approximate \( \sqrt{g(z, 0, x)} \) by \( G(z, x) \), then our locker network design problem (P-Logit) can be approximated by the following Parametric Optimization problem:

\[
(P-PO) \min_{x \in \{0,1\}^K} F(x, t) := T(x) + tG(x)
\]

where \( G(x) = \int_{z \in A} \sqrt{f(z)} G(z, x) dz \) and \( t = \lambda^* + \sqrt{N} \beta(A) \). Notice that \( G(z, x) \) is a decreasing linear function of \( x \) for all \( z \), and this property is preserved for \( G(x) \) after the integration over \( z \). It is straightforward to verify that (i) \( F(x, t) \) is submodular in \( x \) for any \( t \), and (ii) \( F(\cdot) \) has a decreasing difference, i.e., \( \forall x' \geq x, t' > t, \)

\[
F(x', t) - F(x, t) = t(G(x') - G(x)) \geq t'(G(x') - G(x)) = F(x', t') - F(x, t').
\]

Furthermore, we can leverage the two properties to demonstrate that the optimal locker network could (almost) expand in a nested way. We formalize the result in the following proposition.
Proposition 3. Suppose the delivery volumes of $I$ different LSPs are ordered by $N_1 \leq N_2 \leq \ldots \leq N_I$. Let $\lambda^*$ be defined in Proposition 2 and $C(\Gamma)$ be defined in Lemma 1. Then

1. There exists an optimal solution $x^*(t_i)$ to (P-PO) at time $t_i = \lambda^* + \sqrt{N_i\beta(A)}$ for $i = 1, 2, \ldots, I$ that is nested as $t_i$ increases, i.e., $x^*(t_i) \leq x^*(t_i') \forall t_i < t_i'$. 

2. The performance gap between $x^*(t_i)$ and the optimal solution to (P-Logit) for the $i$-th LSP is bounded by $\frac{\lambda^*(C(\Gamma)-1)}{\sqrt{N_i\beta(A)}} + C(\Gamma)$.

Proposition 3 shows that we can construct a nested solution by solving the approximation model (P-PO) and prove that the nested solution is close to be optimal for the original locker network design problem (P-Logit). Note that the performance gap depends on the value of $\lambda^*$, which is independent of the delivery volume $N$. Clearly, when $N$ is sufficiently large, the performance gap converges to $C(\Gamma)$. This result implies that the optimal locker network expands (almost) in a nested fashion with the increase of delivery volume. To this end, we conclude that the optimal density of the LAN is essentially determined by the density of the largest LSP, and pooling the demand from many LSPs together will not lead to substantial increase in the network footprint. In Section 5.3.3, we numerically show that the optimal locker network could possibly expand in a nested fashion. Even if the nested result does not hold strictly, a significant number of locker locations designed for different LSPs overlap with each other.

5. Case Study: The Pilot Program of the Singapore LAN

The Singapore government officially launched the pilot program of LAN from 7 December 2018 to 31 December 2019. The program was run by the Infocomm Media Development Authority (IMDA), which develops and regulates the converging infocomm and media sectors in Singapore. During the pilot, 62 lockers were installed at Housing & Development Board (HDB) blocks, and 8 nearby Mass Rapid Transit (MRT) stations. Two operators – SingPost and Blu – won the bidding to help IMDA run the parcel locker system in two selected towns. In particular, SingPost installed 39 HDB lockers at Punggol town and 4 MRT lockers within/nearby Punggol town, while Blu installed 23 HDB lockers at Bukit Panjang town and 4 MRT lockers within/nearby Bukit Panjang town. The locker networks were designed mainly based on a conventional coverage model so that every resident can find a locker station within a five-minute walk (approximately 250 meters). In addition, the locker networks were open to all the other LSPs, but they were required to pay a service fee depending on the size of the parcels to be served. However, consumers did not need to pay the service fee if they opted to pick up their parcels from the locker network. During the pilot at Punggol town, the majority of parcels

---

SingPost could access the lockers installed at Punggol town without paying any additional service fee to itself, but needed to pay the service fee to use the lockers at Bukit Panjang town.
delivered to the lockers came from Singpost, and hence we can fully investigate how the consumers shift their parcel pickup behaviors.

In what follows, we take the case of Punggol for analysis. The data collected for our case study is described in Section 5.1, and the consumer choice model is calibrated in Section 5.2. Finally, we construct the locker network and study its impact in Section 5.3.

5.1. Description of Data Set

Punggol is a young town located in the north-east region of Singapore. It has a total land area of 9.57 square kilometers and the population size is 174,450.\(^{14}\) We obtain two sources of delivery records: (i) the first data set contains seven months of Singpost’s delivery-to-home records, from January 2019 to July 2019, with deliveries to 537 locations at Punggol. For each parcel home delivery record, we obtain the (unique) transaction identification, (consumer) mobile number, delivery time, delivery destination (address, longitude, and latitude), carrier identification, and delivery status (delivered or failed); (ii) the second data set contains seven months of delivery-to-LAN records, also from January 2019 to July 2019, with deliveries to 39 HDB lockers and 1 MRT locker within Punggol town. For each locker usage record, we also obtain the (unique) transaction identification, (consumer) mobile number, delivery time, delivery destination (locker location, longitude, and latitude), and delivery status (delivered or failed). We plot the daily average parcel volume to home/LAN in Figure 4(a). It shows clearly that home delivery is still the preferred mode of delivery option, counting for 84.66% across the seven months.

Note that the second data set does not record the residential location if the consumer opted to pick up parcels from lockers. As a result, we cannot directly measure the pickup distance from the residential location and the selected locker station. Fortunately, the two sources of data set are linked by the mobile number so that we can select a subset of consumers who picked up their parcels from both home and lockers. In total, 89.43% of locker users also had home delivery experience. We randomly choose a subset of 1000 locker usage records and 5666 home delivery records that can be used to calibrate consumer choices between home delivery and picking up from lockers. We visualize the locker usage patterns in Figure 4(b). The residential locations are randomly distributed at Punggol, and hence the selected subset is representative enough for choice model calibration.

5.2. Calibration of Consumer Choice Model

We consider the classic Multinomial Logistic (MNL) choice model to calibrate consumers’ parcel pickup choice between their home locations and locker stations, i.e., to calibrate the set of parameter \(\theta(z,v), \forall v \in \{0\} \cup V\) as stated in Equation (7). This forms the basis of model (P-Logit).

As shown in Figure 4(b), we have $K = 40$ lockers installed at different locations. Consumers are more likely to use the lockers nearby their residential locations, while the locker installed at MRT station clearly attracts more consumers who might live far away from the MRT station. This motivates us to separate the set of lockers as $V_{MRT}$ and $V_{HDB}$. For each consumer from location $z_i$, there are three types of parcel pickup options, including home, MRT locker, and the other HDB lockers. We denote the set of pickup options as $PickupOption_i$. Furthermore, we let $Distance_{i,k}$ denote the traveling distance from the location $z_i$ to the locker station $v_k$, $\forall v_k \in \{0\} \cup V$. Here we have $Distance_{i,0} = 0$ if the consumer opts to home delivery. Besides of understanding the impact of traveling distance on the utility towards the locker system, we also study the locational feature of each feature like the number of supermarkets nearby the locker. The set of feature variables includes $NumCarpark_k$ (number of car parks), $NumSupermarket$ (number of supermarkets), and $NumRestaurant$ (number of restaurants) that are located within 0.5 kilometers vicinity of the locker $k$. We also consider other features such as the number of bus stops and shopping malls. However, they are strongly correlated with the selected three features.

Under the framework of MNL choice model, we formulate the utility $\theta(z_i, v_k)$ obtained from locker $v_k$ by the customer from location $z_i$ as:

$$\log(\theta(z_i, v_k)) = \beta_1 PickupOption_i + \beta_2 Distance_{i,k} + \beta_3 NumCarpark_k$$
$$+ \beta_4 NumSupermarket_k + \beta_5 NumRestaurant_k.$$  \hspace{1cm} (10)

where $\{\beta_1, \beta_2, \ldots, \beta_5\}$ are the coefficients to be estimated.

We apply R package mlogit\textsuperscript{15} to estimate the customer utility toward different pickup options. We set the home delivery as default pickup option, and the utility of home delivery is normalized to be

\textsuperscript{15}Package ‘mlogit’. Retrieved from \url{https://cran.r-project.org/web/packages/mlogit/mlogit.pdf}. 

Notes. In (b), the big-black circle indicates the locker station, while the small-orange dot presents the home location. The line connecting the circle with the dot visualizes the parcel pickup pattern of a locker user.
1.00, i.e., \( \log(\theta(z_i, 0)) = 0.00 \). The model calibrations are provided in the following Table 1. As shown in Model (d), all the dependent variables in Equation (10) are significantly correlated with the utility of different pickup options at a significance level of 0.01. Compared to home delivery, the option of picking up parcels from the MRT locker station is more attractive as the coefficient 0.97 is positive, while the option of HDB locker is less attractive as the coefficient −1.62 is negative. The coefficient for distance is negative, and hence, the utility of parcel pickup decreases with the traveling distance of parcel pickup. We also observe that the lockers nearby car parks and supermarkets are slightly more attractive (possibly due to the pickup convenience concern), while the one nearby restaurant is less attractive. To check the robustness of our observations, we remove the feature variables sequentially from Model (d), and develop three simplified models, say Model (a)-(c). Consistent effects can be observed in all the models.

In the numerical experiments, we take Model (d) as the calibrated MNL choice model, which can be specified as follows.

\[
\log(\theta(z_i, v_k)) = \begin{cases} 
0.00 & v_k \in \{0\} \\
0.97 - 10.17 \times Distance_{i,k} + 0.11 \times NumCarpark_k & v_k \in V_{MRT} \\
-1.62 - 10.17 \times Distance_{i,k} + 0.11 \times NumCarpark_k & v_k \in V_{HDB} \\
+ 0.10 \times NumSupermarket_k - 0.03 \times NumRestaurant_k & v_k \in V_{MRT} \\
+ 0.10 \times NumSupermarket_k - 0.03 \times NumRestaurant_k & v_k \in V_{HDB}
\end{cases}
\]

### 5.3. Impact of Parcel Locker Network

In the case study, we first investigate the impact of installed 40 parcel lockers on the last mile delivery efficiency during the pilot. Next, we set the existing locker stations as the candidate set to construct the optimal locker network, and examine how the consumer choice model affects the network design. For the ease of exposition, we use the three-month worth of delivery records (i.e., from May to July, 2019) in the numerical experiments. Note that parcel deliveries are scheduled between Monday and Saturday. In the selected data set, there are in total 76 periods, and the average volume of daily parcel delivery is around 400, including both delivery-to-home and delivery-to-locker records.

We apply the model (P-MI-SOCP-D) to configure the optimal locker network, under the existing 40 candidate locker stations as described in Figure 4(b). Furthermore, we discrete the delivery region into \( R = 146 \) grids with equivalent areas, and estimate the delivery density \( f(r) \) at each grid \( r = 1, 2, \ldots, R \). As shown in Figure 5, the parcel deliveries are densely distributed in the center region of the town, while the delivery locations are relatively sparse at the corners.
Table 1 Calibration of MNL choice models.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model (a)</th>
<th>Model (b)</th>
<th>Model (c)</th>
<th>Model (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDB Locker</td>
<td>$-1.381^{***}$</td>
<td>$-1.721^{***}$</td>
<td>$-1.982^{***}$</td>
<td>$-1.621^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.098)</td>
<td>(0.130)</td>
<td>(0.141)</td>
</tr>
<tr>
<td>MRT Locker</td>
<td>$0.969^{***}$</td>
<td>$0.558^{***}$</td>
<td>$0.396^{**}$</td>
<td>$0.971^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.142)</td>
<td>(0.171)</td>
<td>(0.179)</td>
<td>(0.202)</td>
</tr>
<tr>
<td>Distance</td>
<td>$-10.107^{***}$</td>
<td>$-10.108^{***}$</td>
<td>$-10.137^{***}$</td>
<td>$-10.170^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.259)</td>
<td>(0.259)</td>
<td>(0.260)</td>
<td>(0.261)</td>
</tr>
<tr>
<td>Number of Car Parks</td>
<td>$0.086^{***}$</td>
<td>$0.099^{***}$</td>
<td>$0.110^{***}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.020)</td>
<td>(0.021)</td>
<td></td>
</tr>
<tr>
<td>Number of Supermarkets</td>
<td></td>
<td>$0.098^{***}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Restaurants</td>
<td></td>
<td></td>
<td>$-0.033^{***}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.005)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>6,666</td>
<td>6,666</td>
<td>6,666</td>
<td>6,666</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>$-4,068.446$</td>
<td>$-4,059.025$</td>
<td>$-4,054.134$</td>
<td>$-4,029.116$</td>
</tr>
</tbody>
</table>

*Note:* $^{*}p<0.1;^{**}p<0.05;^{***}p<0.01$

Figure 5 Discretization of the delivery region.

5.3.1. Change in the Length of Delivery Trip We set the original case without any locker as the benchmark. In this case, the length of TSP trip comes from home delivery, and we denote the length of TSP trip as $L^0$. Given a network solution $\mathbf{x}$, we denote $L(\mathbf{x})$ and $U(\mathbf{x})$ respectively as the length of TSP trips across the home locations and installed locker stations. We assume that consumers follow the MNL choice model developed in Section 5.2 to select their parcel pickup options between home and lockers. Due to the stochasticity of the choice model, we generate $T = 50$ random instances to evaluate the expected length of the delivery trip after the locker network. Note that
the relative change in home delivery \( \frac{L_0^0 - L(x)}{L_0^0} \) is always non-negative for any \( x \), while the relative change in total delivery \( \frac{L_0^0 - (L(x) + U(x))}{L_0^0} \) might be strictly negative if the network solution \( x \) is not well configured. Furthermore, by Theorem 1 and 2, we have

\[
\frac{L_0^0 - L(x)}{L_0^0} \approx 1 - \left[ \frac{\sqrt{N} \beta(A) \Delta}{\sqrt{N} \beta(A) \Delta} \sum_{r=1}^{R} \sqrt{f(r) g(r, 0, x)} \right] = 1 - \frac{\sum_{r=1}^{R} \sqrt{f(r) g(r, 0, x)}}{\sum_{r=1}^{R} \sqrt{f(r)}} \leq 1 - \sqrt{\Gamma},
\]

where the last inequality comes from Assumption 2. Hence, we can exploit Equation (11) to approximate the relative change in the length of home delivery trip, and we further show that the relative change is upper bounded by a constant that only depends on the utility of home delivery. In the pilot program, we can estimate that the relative home delivery change is around 6.81%, and is bounded above by 26.00% (\( \Gamma = 0.55 \)).

We compare the length of delivery tours before and after the pilot locker network. For the ease of computation, we apply the 2-Opt heuristic algorithm (cf. Croes 1958) to solve the TSP problem and approximate the length of delivery trip across home delivery locations. As shown in Figure 6, the length of home delivery tour reduces between 3% and 15% over the planning period. However, the total length of delivery tours unfortunately increases from 20% to 40%, due to the additional delivery trip to the installed locker stations. This comparison implies that the LAN indeed reduces the last mile delivery efficiency.

Figure 6 Comparison between the length of delivery tours before and after the pilot locker network.

(a) Home Delivery

(b) Total Delivery

Notes. The shaded region indicates the interval from 25 to 75 percentile. In (a), the horizontal dashed line represents the estimated change in the length of home delivery according to Equation (11).

5.3.2. Optimal Locker Network Design We construct the optimal locker network under the consumer parcel pickup choice model as calibrated from the pilot program, for a given LSP. Due to the low attractiveness of parcel lockers, only a small proportion of parcels can be diverted from home to locker. As a result, the reduction in the length of home delivery trip cannot compensate
for the length of delivery trip to any two lockers. Therefore, our optimal solution indicates to install one single locker nearby the Punggol MRT. This result is intuitive since the MRT locker can divert the highest proportion of parcel deliveries from home to locker. In this case, the proportion of home delivery is around 0.99, and the total length of delivery trips drops by 0.78%.

We note that this is the common dilemma faced by the locker operators. On one hand, the locker operators target to install more lockers to provide convenient delivery service and change consumers’ parcel pickup behavior in the long term. On the other hand, they are aware that the locker system is not profitable when the utilization rate is low. Running a nationwide locker system requires a lot of patience in building up demand, and nudging change in behaviour to improve locker adoption rate, before the system can be scaled up.

Our model provides a quick assessment on the performance of a given locker network. More concretely, a locker network solution $x$ can improve the last mile delivery efficiency if the length of delivery trip to the installed locker is smaller than the reduced length of delivery trip to home locations, i.e.,

$$U(x) \leq L^0 - L(x) \approx \sqrt{N} \beta(A) \Delta \sum_{r=1}^{R} \left( \sqrt{f(r)} - \sqrt{f(r)g(r, 0, x)} \right).$$

Clearly, the attraction of home delivery $g(\cdot)$ and demand volume/scale $N$ of LSPs are the critical success factors to reduce the length of delivery trips. Next, we examine the impact of consumer choice and demand volume on the locker network design by varying $\theta(r, 0)$ and $N$, respectively.

We first fix $N = 400$, and reduce the value of $\theta(r, 0)$ from 0.18 to $10^{-4}$. As shown in Table 2, we summarize the number of lockers installed $|x|$, length of delivery trip to the installed lockers $\text{TSP}(x)$, proportion of home delivery, length of delivery trip to home $\text{TSP}(x, \tilde{Z}_\Delta)$, reduction in the length of home delivery, and reduction in the total length of delivery trips. We show that there is a fundamental limit to the optimal level of density of the LAN, regardless of the share of preferences between home delivery and locker pickup.

- When the dominant option for parcel pickup is home delivery, only the MRT locker is installed to avoid inducing additional traveling cost. More lockers would be installed when the attraction of home delivery drops to 0.17. In this case, the locker network reduces the traveling distance to home locations by 5.10 km, and brings an additional traveling distance of 4.04 km. While around 27% consumers opt to lockers for parcel collection, the total length of delivery trips only reduce 3.25% since the delivery locations are densely distributed in the town.

- When the attraction of home delivery decreases to 0.10, the proportion of delivery-to-locker increases to 0.47, with 24 lockers being installed under the optimal solution. The scale of locker network grows slowly if we further reduce the attraction of home delivery to 0.01. Our result shows
Table 2  Impact of consumer choice on the delivery efficiency (∙ N = 400)

| \( \theta(r,0) \) | \( |x| \)  | \( TSP(x) \) | \( p(x) \)  | \( TSP(x, Z_M) \) | Relative Change (%) |
|-----------------|-------|-------------|----------|---------------|---------------------|
| 0.18            | 1     | 0.00        | 0.95     | 31.59         | 3.05                |
| 0.17            | 15    | 4.04        | 0.73     | 27.48         | 15.65               |
| 0.16            | 17    | 4.78        | 0.69     | 26.54         | 18.55               |
| 0.15            | 22    | 6.02        | 0.61     | 25.05         | 23.10               |
| 0.14            | 22    | 6.02        | 0.65     | 24.75         | 24.04               |
| 0.13            | 22    | 6.02        | 0.58     | 24.42         | 25.06               |
| 0.12            | 24    | 6.56        | 0.65     | 23.49         | 27.91               |
| 0.11            | 24    | 6.56        | 0.53     | 23.06         | 29.24               |
| 0.10            | 24    | 6.56        | 0.53     | 22.57         | 30.72               |
| 0.09            | 24    | 6.56        | 0.52     | 22.03         | 32.38               |
| 0.08            | 26    | 7.00        | 0.43     | 20.98         | 35.62               |
| 0.07            | 27    | 7.19        | 0.42     | 20.06         | 38.44               |
| 0.06            | 27    | 7.19        | 0.37     | 19.20         | 41.05               |
| 0.05            | 27    | 7.19        | 0.34     | 18.20         | 44.14               |
| 0.01            | 27    | 7.19        | 0.12     | 10.28         | 68.44               |
| 0.005           | 24    | 6.54        | 0.09     | 8.31          | 74.50               |
| 0.001           | 19    | 5.70        | 0.04     | 4.71          | 85.54               |
| 0.0005          | 18    | 5.59        | 0.02     | 3.52          | 89.20               |
| 0.0001          | 12    | 5.35        | 0.01     | 2.01          | 93.83               |

that 27 lockers (out of 40 candidates) are sufficient enough to divert around 90% parcels from home delivery to the locker network.

- Interestingly, when the parcel locker becomes the dominant parcel pickup option (i.e., \( \theta(r,0) \) is small enough), the number of lockers decreases with \( \theta(r,0) \) to prevent increasing the length of delivery trip to the installed lockers. Therefore, there is no need to ambitiously expand the network when the parcel locker becomes the default option.

We show four representative locker networks in Figure 7. In case (a), since the MRT locker must be installed, all the installed lockers are concentrated around the MRT station so as to minimize the length of delivery trip to the locker network. We also observe that all the installed lockers are located on the right hand side of the MRT station due to the higher volume of parcel delivery at the central region of the town. When the parcel locker becomes more attractive, the locker network first stretches to the right in case (b), and then expands along the periphery of the town to the left in case (c). However, when the parcel locker is much preferred to the home delivery option in case (d), the lockers located at the central region are uninstalled, and the remaining locker stations span the town to cover as many delivery locations as possible.

5.3.3. Nested Structure in Network Expansion  Recall that the LAN is interoperable for all LSPs in Singapore. One aim of the LAN is to attract as many LSPs as possible to use this service, for a sustainable last mile delivery ecosystem in the long term. We note that the delivery densities of all the leading LSPs in Singapore are similar across different estates, and hence we can examine
the impact of delivery volume by varying the parameter \( N \). As shown in Table 3, the reduction in the length of delivery trips decreases with the \( N \). This is a direct implication of Equation (12). We also observe that the optimal network density increases monotonically with \( N \).

| \( N \) | Delivery Trip to Locker \( |\mathbf{x}| \) | \( TSP(\mathbf{x}) \) | Delivery Trip to Home \( p(\mathbf{x}) \) | \( TSP(\mathbf{x}, \tilde{Z}_{\tilde{M}}) \) | Relative Change (%) |
|---|---|---|---|---|---|
| 100 | 1 | 0.00 | 0.92 | 15.56 | 4.51 | 4.51 |
| 200 | 1 | 0.00 | 0.92 | 22.00 | 4.51 | 4.51 |
| 300 | 22 | 6.02 | 0.53 | 20.09 | 28.81 | 7.49 |
| 400 | 24 | 6.56 | 0.54 | 22.57 | 30.72 | 10.58 |
| 500 | 29 | 7.62 | 0.46 | 24.14 | 33.74 | 12.81 |
| 600 | 29 | 7.62 | 0.47 | 26.44 | 33.74 | 14.64 |
| 700 | 29 | 7.62 | 0.46 | 28.56 | 33.74 | 16.05 |
| 800 | 29 | 7.62 | 0.46 | 30.53 | 33.74 | 17.19 |
| 900 | 30 | 7.85 | 0.46 | 32.14 | 34.24 | 18.17 |

Electronic copy available at: https://ssrn.com/abstract=4273960
We validate that the network indeed grows in a nested fashion when \( N \) increases. To see this, we first consider three LSPs with demand volume \( N_1 = 300 \), \( N_2 = 400 \), and \( N_3 = 800 \) respectively. The LAN operator constructs the network based on the largest demand volume \( N_3 = 800 \). As shown in Figure 8, given the optimal network (c), the first two LSPs can personalize a sub-network (a) and (b) from the 29 installed locker stations to minimize the length of their own delivery trips. Next, the other LSP with \( N_4 = 900 \) joins the interoperable system. To serve this “larger” LSP, the operator only needs to expand the existing network from (c) to (d) by installing one more locker. Consistent with the analysis in Section 4.3, our numerical results on other instances suggest that this phenomenon is generally true, and even if the solution is not nested, a significant number of locker locations overlap with each other, indicating that the optimal network designed for the LSP with the largest volume already contains many locations that the smaller LSPs can use to build up their own optimal sub-network.

**Figure 8**  Impact of demand volume on the locker network design \((\theta(r,0) = 0.10)\).

(a) \( N = 300 \)  
(b) \( N = 400 \)

(c) \( N = 800 \)  
(d) \( N = 900 \)

*Notes.* The blue dots represent the locker stations that are added to the expanded network when \( N \) increases.
6. Concluding Remarks

The LAN is a recent smart nation initiative introduced in Singapore for parcel pickup, to improve the delivery efficiency in the last leg of the delivery operations. In the LAN pilot program, we focus on the design of locker network and determine the appropriate size of the network. The problem is complicated by the fact that the network model not only affects customer parcel pickup behavior, but also has a direct impact on the parcel delivery trips.

We use a set of locker usage data from the LAN pilot program to calibrate a logit choice model, and generalize the classic BHH theory to approximate the length of delivery trip to home locations in a stochastic environment. We develop a compact LAN design model to jointly minimize the length of two separate delivery trips to the locker stations and home locations. Our numerical experiments show that a well-constructed LAN would reduce the total length of delivery trips by 10% when half of consumers opt to picking up parcels from the LAN. However, the network is not profitable when the attraction of parcel locker is low. We highlight that one critical issue faced by LAN operator is how to nudge the consumers to shift their parcel pickup behavior from home delivery to the lockers.

Our analysis also highlights the important role played by the larger LSPs in the operations of the interoperable system. Asked in the Singapore Parliament why IMDA did not contract existing private-sector parcel-locker operators to operate the Pick! network, Senior Minister of State for Communications and Information Ms. Sim Ann replied that “the current private-sector driven market for parcel lockers is largely fragmented, with competing players deploying their proprietary networks which are accessible only by selected delivery service providers. This has given rise to deep inefficiencies for local operators. The lockers are duplicated at high-traffic commercial locations and e-commerce marketplaces and delivery service providers are unable to access all lockers which may be available.” She emphasized that having a “neutral” parcel locker operator is key to ensuring that the parcel-locker network can be efficient.\(^\text{16}\) Our analysis explains why there is so much duplication (i.e., nested) in the locker networks set up by competing players, but also underscore the challenges ahead – while the interoperable system were given access to build lockers in public places, otherwise not allowed for the commercial players, the network set up to maximize delivery efficiency for all the LSPs will necessarily be the one that suit the largest LSP the most, and efforts need to be made to enroll the larger commercial players into the program, to avoid the head-to-head competition between the government and these players.

The interoperable locker system, as a novel solution to the innovation of last mile delivery, has piqued a lot of interest in both academia and industry. While we focus on minimizing the length of delivery trips in the last mile segment, we show that a well-constructed network should incentivize

different LSPs, especially the larger ones, to join the LAN to improve their delivery service, and shape the future of Singapore’s smart last mile operations. From an ecosystem perspective, it will be interesting to understand how this solution concept can help to reduce carbon emissions. We leave these issues to future research.

References


Electronic copy available at: https://ssrn.com/abstract=4273960


Appendix

A. Technical Proof

A.1. Proof of Theorem 2

**Theorem 2.** Given the locker network \( V(x) \), and suppose \( \tilde{Z} := \{\tilde{z}_1, \tilde{z}_2, \ldots\} \) is a sequence of random delivery locations independent and identically generated according to an absolutely continuous probability density function \( \hat{f}(\cdot, x) \) defined on a compact planar region \( A \). Let \( \tilde{Z}_M \) be the subset of the first \( M \) points, where \( M \sim \text{Bin}(N, p(x)) \) is independent of \( \tilde{Z} \). Then the length of the optimal TSP tour across these locations, denoted by \( TSP(x, \tilde{Z}_M) \), satisfies

\[
\lim_{N \to \infty} \frac{TSP(x, \tilde{Z}_M)}{p(x)N} = \beta(A) \int_A \sqrt{\hat{f}(z, x)} \, dz, \quad \text{almost surely,}
\]

where \( \beta(A) \) is a constant that depends on the shape of \( A \), while \( p(x) \) and \( \hat{f}( \cdot, x ) \) are defined respectively by Equation (3) and (4).

**Proof of Theorem 2.** For any given network \( x \), the goal is to prove the following,

\[
P \left( \lim_{N \to \infty} \frac{TSP(x, \tilde{Z}_M)}{p(x)N} = \beta(A) \int_A \sqrt{\hat{f}(z, x)} \, dz \right) = 1.
\]

The key is to use concentration bound on the sum of a sequence of independent Bernoulli random variables. When \( N \) is sufficiently large, \( \tilde{M} \) shall concentrate around its mean. In particular, given a finite \( N \), we can apply the Hoeffding’s inequality to \( \tilde{M} \) with a deviation of \( p(x)N^{3\over 2} \),

\[
P \left( \left| \tilde{M} - p(x)N \right| > p(x)N^{3\over 2} \right) \leq 2 \exp \left( -2p(x)^2N^{3\over 2} \right).
\]

We next establish that, in the limit case, \( \tilde{M} \) must stay around its mean with probability one by the Borel-Cantelli lemma. Denote \( a_N = p(x)(N - N^{3\over 2}) \), \( b_N = p(x)(N + N^{3\over 2}) \), and consider the event

\[
E_N = \{\tilde{M} < a_N \text{ or } \tilde{M} > b_N\}.
\]

The sum of probability of those events can be represented by

\[
\sum_{N=1}^{\infty} P(E_N) \leq 2 \sum_{N=1}^{\infty} 2 \exp \left( -2p(x)^2N^{3\over 2} \right) \leq 2 \int_1^{\infty} \exp \left( -2p(x)^2t^{3\over 2} \right) \, dt < \infty.
\]

Hence, the probability that event \( E_N \) happens infinitely many often is 0. In other words, we could restrict our attention on sample paths where \( \tilde{M} \) is eventually bounded between \((a_N, b_N)\), i.e.,

\[
\{(c_N)_{N=1}^{\infty} \mid \exists n \text{ s.t. } \forall N > n, a_N \leq c_N \leq b_N\}.
\]

Conditioning on \( \tilde{M} \) is on such sample paths, we can claim the following for \( N \) being sufficiently large,

\[
TSP(x, \tilde{Z}_{a_N}) \leq TSP(x, \tilde{Z}_M) \leq TSP(x, \tilde{Z}_{b_N}).
\]
Since both $a_N$, $b_N$ diverge to infinity, we are now ready to invoke the original BHH theorem and claim
\[
\beta(\mathcal{A}) \int_A \sqrt{\hat{f}(z, x)} dz \sqrt{1 - N^{-\frac{\theta}{2}}} + \frac{o(\sqrt{N})}{\sqrt{p(x) N}} \leq \frac{\text{TSP}(x, \hat{Z}_N)}{\sqrt{p(x) N}} \leq \beta(\mathcal{A}) \int_A \sqrt{\hat{f}(z, x)} dz \sqrt{1 - N^{-\frac{\theta}{2}}} + \frac{o(\sqrt{N})}{\sqrt{p(x) N}},
\]
for a sufficiently large $N$. As we only throw away a set of sample paths that is of probability 0, the desired result follows when $N$ tends to infinity.

A.2. Proof of Proposition 1

**Proposition 1.** The model (P-Logit) can be exactly reformulated as the following mixed-integer second-order cone problem (P-MI-SOCP):

\[
\text{(P-MI-SOCP)} \quad \min \ TSP(x) + \sqrt{N} \beta(\mathcal{A}) \int_A \sqrt{\hat{f}(z, x)} dz
\]

\[
\text{s.t.} \quad \begin{cases} 
\theta(z, 0) + \sum_{j=1}^{K} \theta(z, v_j) x_j - \frac{\rho(z)}{1/\theta(z, 0)} & \geq 0, \forall z \in \mathcal{A} \\
\left( \begin{array}{c} w(z) \\ 1 \end{array} \right) \left( \begin{array}{c} 1 \\ \rho(z) \end{array} \right) & \geq 0, \forall z \in \mathcal{A} \\
\rho(z) & \geq 1, 0 \leq w(z) \leq 1, \forall z \in \mathcal{A} \\
x_k & \in \{0, 1\}, \forall k = 1, 2, \ldots, K.
\end{cases}
\]

Furthermore, the continuous relaxation of this problem is a convex programming problem.

**Proof of Proposition 1.** Denote the optimal solution to model (P-Logit) as $x^1$ and the optimal solution to model (P-MI-SOCP) as $(x^2, w^*(z), \rho^*(z))$. We suppress the discussion over $z \in \mathcal{A}$ when there is no confusion. Furthermore, we denote the objective of model (P-Logit) and (P-MI-SOCP) as $h_1(\cdot)$ and $h_2(\cdot)$, respectively. In what follows, we target to show that $h_1(x^1) = h_2(x^2, w^*(z), \rho^*(z))$.

(i) Given the optimal solution $x^1$ to model (P-Logit), we construct a feasible solution $(x^1, w(z), \rho(z))$ to model (P-MI-SOCP), where $w(z) = 1/\rho(z)$ and \[
\frac{1}{(\rho(z))^2} = \frac{\theta_0(z)}{w(z)} = \frac{\theta_0(z)}{w(z, v_k) x_k} \leq 1.
\] It is straightforward to verify that $\rho(z) \geq 1, 0 \leq w(z) \leq 1$, and hence $(x^1, w(z), \rho(z))$ is a feasible solution to model (P-MI-SOCP). Therefore, we have

\[
h_1(x^1) = h_2(x^1, w(z), \rho(z)) \geq h_2(x^2, w^*(z), \rho^*(z)). \tag{13}
\]

The first equality comes from the construction of solution $(x^1, w(z), \rho(z))$, and the second inequality is due to the optimality of solution $(x^2, w^*(z), \rho^*(z))$.

(ii) Given the optimal solution $(x^2, w^*(z), \rho^*(z))$ to model (P-MI-SOCP). The first and second sets of feasibility constraints jointly force that

\[
w^*(z) \geq \frac{1}{\rho^*(z)} \geq \sqrt{\frac{\theta_0(z)}{\theta_0(z) + \sum_{k=1}^{K} \theta(z, v_k) x_k^2}}. \tag{14}
\]

Therefore, we have

\[
h_2(x^2, w^*(z), \rho^*(z)) \geq h_1(x^2) \geq h_1(x^1). \tag{15}
\]
The first equality comes from Equation (15), and the second inequality is due to the optimality of solution $x^1$.

Combining Equation (13) and (14), we have $h_1(x^1) = h_2(x^2, w^*(z), \rho^*(z))$. This completes the proof. 

A.3. Proof of Theorem 3

**Theorem 3.** For a given instance of problem (P-MI-SOCP) and a rounding threshold $\gamma$, we have

$$H^C \leq H^* \leq H^A(\gamma) \leq \max \left\{ \sqrt{1 + \gamma \frac{1 - \Gamma}{\Gamma}}, \frac{3}{2\gamma} \right\} H^C.$$ 

**Proof of Theorem 3.** We extend the technique from Bienstock et al. (1993), Goemans (2009) to prove this result. Let $(x^c, y^c, w^c(z), \lambda^c(z))$ be an optimal solution to the continuous relaxation of model (P-MI-SOCP) when $x_q$ is the fixed depot. Consider the rounding mechanism $S(\gamma) = \{ k | x^c_k \geq \gamma \}$, and let $T_\gamma$ denote the length of the tour produced by Christofides’ algorithm on the locker set $S(\gamma)$. Goemans (2009) showed that,

$$T_\gamma \leq \frac{3}{2\gamma} \sum_{e \in E} d_{e} y^c_e.$$ 

The inequality gives an upper bound on the TSP component. Next, we analyze the relation between the price collected under the rounding solution and price collected based on the continuous relaxation. With a slight abuse of notation, we define $(x^a, y^a, w^a(z), \lambda^a(z))$ as the solution after rounding. By definition, we have

$$w^a(z) = \sqrt{\frac{\theta(z, 0)}{\theta(z, 0) + \sum_{k \in S(\gamma)} \theta(z, v_k)}}.$$ 

Furthermore, we can derive

$$\left( \frac{w^a(z)}{w^c(z)} \right)^2 = \frac{\theta(z, 0) + \sum_{k=1}^{K} \theta(z, v_k) x^c_k}{\theta(z, 0) + \sum_{k \in S(\gamma)} \theta(z, v_k)} \leq \frac{\theta(z, 0) + \sum_{k \in S(\gamma)} \theta(z, v_k) + \sum_{k \notin S(\gamma)} \theta(z, v_k) \gamma}{\theta(z, 0) + \sum_{k \in S(\gamma)} \theta(z, v_k)} \leq 1 + \gamma \frac{\sum_{k \in S(\gamma)} \theta(z, v_k)}{\theta(z, 0)} \leq 1 + \gamma \frac{1 - \Gamma}{\Gamma}. \quad (16)$$

In Equation (16), we round up the value of $x^c_k$ to 1 and $\gamma$, according to $\gamma \leq x^c_k \leq 1$ and $x^c_k < \gamma$, respectively. Equation (17) follows directly from our assumption 2. As the analysis is independent of $z$, it naturally goes for the integration over all $z \in \mathcal{A}$,

$$\int_{\mathcal{A}} \sqrt{f(z)} w^a(z) dz \leq \sqrt{1 + \gamma \frac{1 - \Gamma}{\Gamma}} \int_{\mathcal{A}} \sqrt{f(z)} w^c(z) dz.$$
Combining the TSP component and prize collecting component, we can claim
\[ H^A(x) = T + \sqrt{N} \beta(A) \int_A \sqrt{f(z)} w^A(z) dz \]
\[ \leq \frac{3}{2\gamma} \sum_{e \in E} d_e y_e^C + \sqrt{N} \beta(A) \int_A \sqrt{f(z)} w^C(z) dz \]
\[ \leq \max \left( \frac{3}{2\gamma} \sqrt{1 + \gamma \frac{1 - \Gamma}{\Gamma}} \right) \left( \sum_{e \in E} d_e y_e^C + \int_A \sqrt{f(z)} w^C(z) dz \right) \]
\[ = \max \left( \frac{3}{2\gamma} \sqrt{1 + \gamma \frac{1 - \Gamma}{\Gamma}} \right) H^C. \]

A.4. Proof of Proposition 2

**Proposition 2.** Suppose \( \theta(z, v_k) > 0 \) for any \( z \) and \( v_k \), then

1. \( \sqrt{g(z,0,x)} \) is strictly supermodular in \( x \) for any given \( z \);
2. \( \exists \lambda^* > 0 \) such that \( T(x) := TSP(x) - \lambda^* \int_{z \in A} \sqrt{f(z)} g(z,0,x) dz \) is submoular in \( x \).

**Proof of Proposition 2.** Under the logit choice model, the term \( \sqrt{g(z,0,x)} \) is essentially \( \sqrt{\frac{1}{1 + a(x)}} \), where \( a(x) \) is a linear function of \( x \). Pick any \( x^1, x^2 \subseteq \{0,1\}^K \), we have
\[ \sqrt{\frac{1}{1 + a(x^1 \cap x^2)}} - \sqrt{\frac{1}{1 + a(x^1)}} = \frac{\sqrt{1 + a(x^1)} - \sqrt{1 + a(x^1 \cap x^2)}}{\sqrt{1 + a(x^1)} \sqrt{1 + a(x^1 \cap x^2)}} \]
\[ \geq \frac{\sqrt{1 + a(x^1) - 1 + a(x^1 \cap x^2)}}{\sqrt{1 + a(x^1 \cap x^2)}} \]
\[ \geq \frac{\sqrt{1 + a(x^1 \cap x^2)) - 1 + a(x^2)}}{\sqrt{1 + a(x^1 \cap x^2)}} \]
\[ = \sqrt{\frac{1}{1 + a(x^1 \cap x^2)}} - \sqrt{\frac{1}{1 + a(x^1)}} \]

where the first inequality follows from \( (1 + a(x^1))(1 + a(x^1 \cap x^2)) \leq (1 + a(x^1 \cup x^2))(1 + a(x^2)) \) and the second inequality follows from \( (1 + a(x^1))(1 + a(x^2)) \geq (1 + a(x^1 \cup x^2))(1 + a(x^1 \cap x^2)) \). Clearly, if neither \( x^1 \) nor \( x^2 \) is a proper subset of the other, the inequality becomes strict. This completes the proof for the first claim.

Based on the first claim, we can easily show that \( \int_{z \in A} \sqrt{f(z)} g(z,0,x) dz \) is strictly supermodular in \( x \) since \( f(z) \) is a density function. Therefore, the existence of desired \( \lambda^* \) is straightforward: any fixed set function can be turned into a submodular one by adding a large enough strictly submodular function to it. Namely, for any \( x^1 \) and \( x^2 \), the parameter \( \lambda^* \) has to be large enough so that
\[ \lambda^* \int_{z \in A} \sqrt{f(z)} \left( \sqrt{g(z,0,x^1 \cap x^2)} + \sqrt{g(z,0,x^1 \cup x^2)} - \sqrt{g(z,0,x^1)} - \sqrt{g(z,0,x^2)} \right) dz \]
\[ \geq TSP(x^1 \cap x^2) + TSP(x^1 \cup x^2) - TSP(x^1) - TSP(x^2). \] (18)
Since both \( \sqrt{g(\cdot)} \) and \( TSP(\cdot) \) are given, and there only exists a finite number of choices for \( x^1 \) and \( x^2 \), the existence of desired \( \lambda^* \) is obvious. However, this argument alone may not be informative about how large \( \lambda^* \)
should be. Therefore, we focus on finding a sufficient condition that is easier for checking an upper bound on $\lambda^*$. We first simplify the right hand side through Parsimonious property of TSP($x$) derived by Goemans and Bertsimas (1993). Their result allows us to write the Euclidean TSP as the following problem,

$$
\text{TSP}(x) := \min \sum_{e \in E} d_e y_e
$$

s.t.

$$\sum_{e \in \delta(j)} y_e \geq 2x_j, \ \forall j = 1, 2, \ldots, K$$

$$\sum_{e \in \delta(S)} y_e \geq 2x_j, \ \forall j \in S \subseteq \{1, 2, \ldots, K\} \setminus \{q\}$$

$$y_e \in \{0, 1\}, \ \forall e \in E$$

Let $y^1$ and $y^2$ be the optimal solution to TSP($x^1$) and TSP($x^2$) respectively. Based on the above formulation, we can verify that $y^1 \cup y^2$ is a feasible solution to TSP($x^1 \cup x^2$), which yields the result $\text{TSP}(x^1) + \text{TSP}(x^2) \geq \text{TSP}(x^1 \cup x^2)$. Hence, we can simplify the condition of Inequality (18) as follows:

$$\lambda^* \int_{z \in A} \sqrt{f(z)} \left( \sqrt{g(z, 0, x^1 \cup x^2)} - \sqrt{g(z, 0, x^1)} - \sqrt{g(z, 0, x^2)} \right) dz \geq \text{TSP}(x^1 \cap x^2).$$

Let $D$ denote the diameter of region $A$, and recall that the total number of candidate lockers is denoted by $K$. Therefore, the term $\text{TSP}(x^1 \cap x^2)$ is bounded above by $KD$. When proving the supermodularity of $\sqrt{g(z, 0, x^2)}$, it has been shown that, for a fixed $z$, we have

$$\sqrt{g(z, 0, x^1 \cap x^2)} + \sqrt{g(z, 0, x^1 \cup x^2)} - \sqrt{g(z, 0, x^1)} - \sqrt{g(z, 0, x^2)} \geq \frac{\sqrt{1 + a(x^1)} + \sqrt{1 + a(x^2)} - \sqrt{1 + a(x^1 \cup x^2)} - \sqrt{1 + a(x^1 \cap x^2)}}{\sqrt{1 + a(x^1 \cup x^2)}} \geq \frac{\theta(z, 0)}{\theta(z, 0) + \sum \theta(z, v_k) - \theta(z, 0)} = b(z),$$

where $k^* = \arg \min_k \theta(z, v_k)$. By denoting $B := \int_{z \in A} \sqrt{f(z)} b(z) dz$, we can now specify a sufficiently large value of $\lambda^*$ as $\frac{KD}{B}$, which only depends on the problem-related parameters while is independent of $N$. 

A.5. Proof of Lemma 1

**Lemma 1.** The ratio of $G(z, x)$ over $\sqrt{g(z, 0, x)}$ is upper bounded by a parametric constant, i.e.,

$$\sqrt{g(z, 0, x)} \leq G(z, x) \leq C(\Gamma) \sqrt{g(z, 0, x)},$$

where $C(\Gamma) = \frac{2\sqrt{\pi (1 - \Gamma \sqrt{T})}}{g(1 - \Gamma^2)} \sqrt{\frac{1 - \Gamma \sqrt{T}}{1 - \Gamma^2}}$.

**Proof of Lemma 1.** Since Assumption 2 implies $0 \leq \sum_{k=1}^K \frac{\theta(z, v_k)}{\theta(z, 0)} x_k \leq 1 - \Gamma$, it suffices to study the function $\sqrt{\frac{1}{1 + a}}$ on the interval $[0, \frac{1 - \Gamma}{1 + a}]$. By the convexity of $\sqrt{\frac{1}{1 + a}}$, the linear function $G(z, x)$ constructed by connecting two points $(0, 1)$ and $(\frac{1 - \Gamma}{1 + a}, \sqrt{T})$ should be an upper bound of $\sqrt{\frac{1}{1 + a}}$ on the interval $[0, \frac{1 - \Gamma}{1 + a}]$. This proves the first inequality.
To prove the second inequality, we first denote 

\[ m := \frac{1 - ma}{2\sqrt{1 + a}}. \]

The problem simplifies to study 

\[ h(a) := \frac{1 - ma}{2\sqrt{1 + a}} - m\sqrt{1 + a}, \]

which implies the original \( h(\cdot) \) function is maximized at \( a^* = \frac{1 - 2m}{3m} \). The desired upper bound follows by evaluating \( h(\frac{1 - 2m}{3m}) \) and substituting \( m \) with the expression of \( \Gamma \).

\[ \square \]

### A.6. Proof of Proposition 3

**Proposition 3.** Suppose the delivery volumes of \( I \) different LSPs are ordered by \( N_1 \leq N_2 \leq \ldots \leq N_I \). Let \( \lambda^* \) be defined in Proposition 2 and \( C(\Gamma) \) be defined in Lemma 1. Then

1. There exists an optimal solution \( x^*(t_i) \) to (P-PO) at time \( t_i = \lambda^* + \sqrt{N_i}\beta(A) \) for \( i = 1, 2, \ldots, I \) that is nested as \( t_i \) increases, i.e., \( x^*(t_i) \leq x^*(t_{i'}) \) for all \( t_i < t_{i'} \).

2. The performance gap between \( x^*(t_i) \) and the optimal solution to (P-Logit) for the \( i \)-th LSP is bounded by \( \frac{\lambda^* C(\Gamma) - 1}{\sqrt{N_i}\beta(A)} + C(\Gamma) \).

**Proof of Proposition 3.** For two time indexes \( t < t' \), let \( x^*(t), x^*(t') \) be two optimal solutions to (P-PO) under parameter \( t, t' \) respectively. We first prove that \( x^*(t) \cup x^*(t') \) is an optimal solution at time \( t' \).

\[
F(x^*(t), t) \leq F(x^*(t) \cap x^*(t'), t) \quad \text{(by the optimality of } x^*(t) \text{ for } t) \\
\implies F(x^*(t) \cup x^*(t'), t) \leq F(x^*(t'), t) \quad \text{(by the submodularity of } F(\cdot, t) \text{)} \\
\implies F(x^*(t) \cup x^*(t'), t') \leq F(x^*(t'), t') \quad \text{(by the decreasing difference property)}
\]

The desired result follows by the optimality of \( x^*(t') \) at time \( t' \). Similarly, we can validate that the intersection \( x^*(t) \cap x^*(t') \) is an optimal solution at time \( t \).

\[
F(x^*(t'), t') \leq F(x^*(t) \cap x^*(t'), t') \quad \text{(by the optimality of } x^*(t') \text{ for } t') \\
\implies F(x^*(t'), t) \leq F(x^*(t) \cup x^*(t'), t) \quad \text{(by the decreasing difference property)} \\
\implies F(x^*(t) \cap x^*(t'), t) \leq F(x^*(t), t) \quad \text{(by the submodularity of } F(\cdot, t) \text{)}
\]

The optimality of \( x^*(t) \) at time \( t \) implies the desired result in the first claim.

To prove the second claim, let \( x^*(t_i), \tilde{x}^*(t_i) \) denote an optimal solution to (P-PO) and (P-Logit) at time \( t_i = \lambda^* + \sqrt{N_i}\beta(A) \), respectively. Based on the optimality of \( x^*(t_i) \), we have

\[
T(x^*(t_i)) + t_i G(x^*(t_i)) \leq T(\tilde{x}^*(t_i)) + t_i G(\tilde{x}^*(t_i)). \tag{19}
\]

Plug in \( T(\cdot), G(\cdot) \) and combine with the lower bound of \( G(z, \cdot) \) stated in Lemma 1, we can simplify the left-hand-side of Equation (19) as:

\[
T(x^*(t_i)) + t_i G(x^*(t_i)) \\
= TSP(x^*(t_i)) - \lambda^* \int_{z \in A} \sqrt{f(z)} g(z, 0, x^*(t_i))dz + (\lambda^* + \sqrt{N_i}\beta(A)) \int_{z \in A} \sqrt{f(z)} g(z, x^*(t_i))dz \\
\geq TSP(x^*(t_i)) - \lambda^* \int_{z \in A} \sqrt{f(z)} g(z, 0, x^*(t_i))dz + (\lambda^* + \sqrt{N_i}\beta(A)) \int_{z \in A} \sqrt{f(z)} g(z, 0, x^*(t_i))dz \\
= TSP(x^*(t_i)) + \sqrt{N_i}\beta(A) \int_{z \in A} \sqrt{f(z)} g(z, 0, x^*(t_i))dz, \tag{20}
\]
which is the objective function of (P-Logit) evaluated under solution $x^*(t_i)$. We apply a similar analysis to the right-hand-side of Equation (19), and utilize the upper bound of $G(z, \cdot)$ stated in Lemma 1 to conclude:

$$
T(\tilde{x}(t_i)) + t_i G(\tilde{x}(t_i)) \\
= \text{TSP}(\tilde{x}(t_i)) - \lambda^* \int_{z \in A} \sqrt{f(z)} \sqrt{g(z, 0, \tilde{x}(t_i))} dz + (\lambda^* + \sqrt{N_i \beta(A)}) \int_{z \in A} \sqrt{f(z)} G(z, \tilde{x}(t_i)) dz \\
\leq \text{TSP}(\tilde{x}(t_i)) - \lambda^* \int_{z \in A} \sqrt{f(z)} \sqrt{g(z, 0, \tilde{x}(t_i))} dz + C(\Gamma)(\lambda^* + \sqrt{N_i \beta(A)}) \int_{z \in A} \sqrt{f(z)} \sqrt{g(z, 0, \tilde{x}(t_i))} dz \\
= \text{TSP}(\tilde{x}(t_i)) + \left( \frac{\lambda^* (C(\Gamma) - 1)}{\sqrt{N_i \beta(A)}} + C(\Gamma) \right) \sqrt{N_i \beta(A)} \int_{z \in A} \sqrt{f(z)} \sqrt{g(z, 0, \tilde{x}(t_i))} dz \\
\leq \left( \frac{\lambda^* (C(\Gamma) - 1)}{\sqrt{N_i \beta(A)}} + C(\Gamma) \right) \left( \text{TSP}(\tilde{x}(t_i)) + \sqrt{N_i \beta(A)} \int_{z \in A} \sqrt{f(z)} g(z, 0, \tilde{x}(t_i)) dz \right). 
$$

Combining Equation (20) and (21), the desired result in the second claim follows directly. \[\blacksquare\]

B. Numerical Experiments

In the numerical experiments, we formulate the locker network design models using Python language and solve the related optimization problems using Gurobi version 9.5.1 on a 2.70-GHz i7-6820HQ Central Processing Unit Windows Personal Computer with 16 GB Random Access Memory.

B.1. Numerical Validation of Theorem 2

We provide a numerical validation of Theorem 2 by showing that the TSP approximation by Equation (6) is fairly accurate. Consider a unit planar $\mathcal{A}$, and the delivery density is $f(z) = 1, \forall z \in A$. Four parcel lockers are installed respectively at the coordinates $(0.25, 0.25)$, $(0.25, 0.75)$, $(0.75, 0.25)$, and $(0.75, 0.75)$. We assume that each consumer opts to locker or home for parcel delivery with equal probability if there is one locker installed within a (Euclidean) distance of 0.25 units; otherwise home delivery is the only option. A fixed number of $N \in \{100, 150, \ldots, 400\}$ consumer home locations are uniformly sampled on the planar. In each case, we generate $T = 50$ random instances. According to Theorem 1, we first calibrate the parameter $\beta(A)$ when the locker is not installed, and then examine the gap between the exact length of the optimal TSP tour to the remaining home locations and the approximated one under the same $\beta(A)$ when the four lockers are installed. We plot the ratio of the LHS of Equation (6) to the RHS in Figure 9. It shows clearly that the mean value of the ratio over 50 instances is close to 1 and variation reduces with the increase of the number of delivery locations.

B.2. Numerical Validation of the Continuous Relaxation Model

In this section, we construct synthetic experiments to implement our exact reformulation and investigate the numerical performance of the continuous relaxation model. For the ease of exposition, we consider a unit grid planar $\mathcal{A}$ with $R = 10 \times 10$ equivalent girds. The area of each grid is set as $\Delta = 1/100$ and the delivery density is $f(r) = 1, \forall r = 1, 2, \ldots, R$ in this case. The locker is allowed to be installed at any corner or cross point, i.e., at the coordinates $(\frac{i}{10}, \frac{j}{10})$, $\forall i = 1, \ldots, 11, j = 1, \ldots, 11$. In total, there are $11 \times 11$ locker candidate locations. A fixed number of $N$ delivery locations are uniformly sampled on the planar. The attraction of
home delivery to the consumers located at grid $r$ is denoted by $\theta(r,0) = \alpha$, while the attraction of locker $v_k$ is represented by $\theta(r,v_k) = e^{-d_{r,k}}$, where $d_{r,k}$ denotes the Euclidean distance from the center of grid $r$ to the locker $k$, and the coefficient $\alpha > 0$ controls the proportion of home deliveries. We first calibrate the parameter $\beta(A)$ when the locker is not installed, and implement this parameter to solve the locker network design problem. Given the parameter $\alpha$, we estimate the value of $\Gamma$ in Assumption 2 by assuming all the lockers are installed, and numerically search the value of $\gamma$ to achieve the tightest bound in Theorem 3. In addition, we fix the locker at the central point $(0.5,0.5)$ to be installed so that we can minimize the length of delivery tour to the installed locker stations directly.

We vary the values of $N$ and $\alpha$ to investigate their impacts on the approximate solution. Table 4 shows that the selected rounding threshold $\gamma$ increases with the utility of staying home for parcel pickup. As a result, the approximate ratio becomes tighter if the consumers are more likely to pick up parcels at their homes. In this case, there is no need to install too many parcel lockers, and hence the difference between the exact model and approximate model is marginal. We also observe that the approximate ratio in general increases with the number of delivery locations. This is intuitive since it becomes harder to balance the trade-offs between two separate delivery trips when the pool of consumers is larger. Notably, while the theoretical ratio from Theorem 3 is bounded below 1.5, the actual ratio seems to be much closer to 1.

Note that we solve a discretized version of the network design problem (P-MI-SOC-P-D). Next, we examine the impact of discretization on the performance of both exact model and approximate model by varying the number of girds. For a fair comparison, we fix the $11 \times 11$ locker candidate locations in all cases. Taking $N = 400$ for illustration, Table 5 shows that the impact of $R$ on the performance of both models are negligible.
Table 4: Comparison between the exact solution and approximate solution.

<table>
<thead>
<tr>
<th>N</th>
<th>α</th>
<th>Γ</th>
<th>γ</th>
<th>Total Length of TSP Tours</th>
<th>Exact Solution</th>
<th>Approximate Solution</th>
<th>Theoretical Ratio</th>
<th>Actual Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>10</td>
<td>0.11</td>
<td>0.62</td>
<td>9.05</td>
<td>9.63</td>
<td>2.44</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.27</td>
<td>0.83</td>
<td>10.66</td>
<td>10.66</td>
<td>1.80</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.38</td>
<td>0.95</td>
<td>10.71</td>
<td>10.71</td>
<td>1.59</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>10</td>
<td>0.11</td>
<td>0.62</td>
<td>11.82</td>
<td>12.62</td>
<td>2.44</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.27</td>
<td>0.83</td>
<td>14.67</td>
<td>15.21</td>
<td>1.80</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.38</td>
<td>0.95</td>
<td>15.14</td>
<td>15.14</td>
<td>1.59</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>600</td>
<td>10</td>
<td>0.11</td>
<td>0.62</td>
<td>13.78</td>
<td>14.99</td>
<td>2.44</td>
<td>1.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.27</td>
<td>0.83</td>
<td>17.47</td>
<td>18.08</td>
<td>1.80</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.38</td>
<td>0.95</td>
<td>18.52</td>
<td>18.97</td>
<td>1.59</td>
<td>1.02</td>
<td></td>
</tr>
<tr>
<td>800</td>
<td>10</td>
<td>0.11</td>
<td>0.62</td>
<td>15.34</td>
<td>16.52</td>
<td>2.44</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.27</td>
<td>0.83</td>
<td>19.71</td>
<td>20.89</td>
<td>1.80</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.38</td>
<td>0.95</td>
<td>21.14</td>
<td>21.69</td>
<td>1.59</td>
<td>1.03</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Impact of discretization on the locker network design.

<table>
<thead>
<tr>
<th>R</th>
<th>α</th>
<th>Γ</th>
<th>γ</th>
<th>Total Length of TSP Tours</th>
<th>Exact Solution</th>
<th>Approximate Solution</th>
<th>Theoretical Ratio</th>
<th>Actual Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 × 5</td>
<td>10</td>
<td>0.11</td>
<td>0.61</td>
<td>11.81</td>
<td>12.60</td>
<td>2.44</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.27</td>
<td>0.83</td>
<td>14.66</td>
<td>15.20</td>
<td>1.80</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.38</td>
<td>0.94</td>
<td>15.14</td>
<td>15.14</td>
<td>1.59</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>10 × 10</td>
<td>10</td>
<td>0.11</td>
<td>0.62</td>
<td>11.82</td>
<td>12.62</td>
<td>2.44</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.27</td>
<td>0.83</td>
<td>14.67</td>
<td>15.21</td>
<td>1.80</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.38</td>
<td>0.95</td>
<td>15.14</td>
<td>15.14</td>
<td>1.59</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>20 × 20</td>
<td>10</td>
<td>0.11</td>
<td>0.62</td>
<td>11.83</td>
<td>12.62</td>
<td>2.44</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.27</td>
<td>0.83</td>
<td>14.67</td>
<td>15.21</td>
<td>1.80</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.38</td>
<td>0.94</td>
<td>15.14</td>
<td>15.14</td>
<td>1.59</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>40 × 40</td>
<td>10</td>
<td>0.11</td>
<td>0.61</td>
<td>11.83</td>
<td>12.62</td>
<td>2.44</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.27</td>
<td>0.83</td>
<td>14.67</td>
<td>15.21</td>
<td>1.80</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.38</td>
<td>0.94</td>
<td>15.14</td>
<td>15.14</td>
<td>1.59</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

B.3. Numerical Validation of the Approximation Function $G(z, x)$

We numerically validate the approximation accuracy of our linear approximation function $G(z, x)$ for the original function $\sqrt{g(z, 0, x)}$. To approximate $\sqrt{g(z, 0, x)}$ under logit choice model, we first rewrite

$$\frac{\theta(z, 0)}{\sqrt{\theta(z, 0) + \sum_{k=1}^{K} \theta(z, v_k) x_k}} = \sqrt{\frac{1}{1 + \sum_{k=1}^{K} \frac{\theta(z, v_k)}{\theta(z, 0)} x_k}}.$$  \hspace{1cm} (22)

Notice that Assumption 2 implies $\sum_{k=1}^{K} \frac{\theta(z, v_k)}{\theta(z, 0)} x_k \leq \frac{1-\Gamma}{\Gamma}$. Therefore, by picking different values for $x$ and denoting $a = \sum_{j=1}^{K} \frac{\theta(z, v_j)}{\theta(z, 0)} x_j$, Equation (22) can be expressed as $\sqrt{\frac{1}{1+a}}$ for some $a \in [0, \frac{1-\Gamma}{\Gamma}]$. Since the function $\sqrt{\frac{1}{1+a}}$ is convex on the interval $[0, \frac{1-\Gamma}{\Gamma}]$, one way to bound the function is to evaluate the function value on $a = 0$ and $a = \frac{1-\Gamma}{\Gamma}$, and connect the two points $(0, 1)$ to $(\frac{1-\Gamma}{\Gamma}, \sqrt{\frac{1}{\Gamma}})$ to form a linear function. This gives rise to the approximation function $G(z, x) = 1 - \frac{\Gamma(1-\Gamma)}{1-\Gamma} \sum_{j=1}^{K} \frac{\theta(z, v_j)}{\theta(z, 0)} x_j$ as proposed in Section 4.3. We provide a numerical validation for this approximation in the following Figure 10 under different values of $\Gamma$. The approximation is reasonably accurate, especially for large value of $\Gamma$.
Figure 10  Approximation accuracy under different values of $\Gamma$.

(a) $\Gamma = 50\%$

(b) $\Gamma = 60\%$

(c) $\Gamma = 70\%$

(d) $\Gamma = 80\%$

Notes. For different values of $\Gamma$, we plot the two functions $\sqrt{\frac{1}{1+a}}$ and $1 - \frac{\Gamma(1-a)}{1-a}$ on the interval $[0, \frac{1}{1-\Gamma}]$. 

Electronic copy available at: https://ssrn.com/abstract=4273960