1. Introduction

The rapid growth of e-commerce in recent years has led to the popularity of same-day delivery, in which orders placed by a cutoff time are fulfilled by the end of the same service day. The thin margins associated with last-mile logistics necessitate careful planning and design of same-day delivery systems, of which service region partitioning is a key component.

However, the significant inherent unpredictability (in order times, customer locations, travel times, etc.) associated with these systems pose difficulties when analyzing overall system behavior using traditional discrete operations research models. Approximating discrete and stochastic aspects of a logistics system using deterministic continuous functions facilitates tactical design and planning. The central result in the use of continuous approximations for logistics is known as the BHH Theorem (Beardwood, Halton, and Hammersley 1959): the length of the optimal traveling salesman tour over \( n \) uniformly distributed points in a compact region approaches \( \gamma\sqrt{An} \) as \( n \) grows, where \( A \) is the area of the region and \( \gamma \) is a constant dependent on the underlying metric.

Recent work by Stroh, Erera, and Toriello (2019) uses a continuous approximation approach to determine optimal same-day dispatching policies with respect to total vehicle routing cost for single-vehicle and infinite-vehicle fleets. Specifically, orders arrive continuously over a service region at a rate \( \lambda \) from time \( t = 0 \) until an order cutoff time \( N \). Vehicles dispatch from a centrally located depot to serve orders, and the routing time for a dispatch is approximated with a continuous BHH-type function. All orders must be served and all vehicles must return to the depot by a fixed time \( T > N \). Figure 1 illustrates an example vehicle dispatching policy where two vehicles each complete one dispatch.

![Figure 1](image-url)  
**Figure 1**  
Vehicle dispatching example in same-day delivery setting
This recent work assumes every vehicle is responsible for orders in the entire service region. In our work, we use a similar approach to determine a partitioning and associated fleet size when each vehicle is responsible for a separate zone within the region.

2. Problem Definition and Approach

Suppose we are given a service region $R$ with a single depot. Our initial goal is to determine the minimum number of vehicles required to service a region if the region is partitioned into zones such that each vehicle is to independently serve a single zone. To this end, we first consider the related problem of independently maximizing the area of such a zone. Specifically, we determine the maximum area of a compact zone, whose centroid is $\rho$ time units of travel from the depot, such that the vehicle assigned to the zone can feasibly serve all orders which accumulate in the zone by making exactly $D$ dispatches. We assume that orders arrive continuously and uniformly over the service region at a rate $\lambda$ per unit area per unit time over the time interval $[0, N]$ and that all vehicles must return to the depot by a fixed time $T > N$. We assume for conciseness that $\lambda$ has been normalized to 1, and the remaining parameters have been scaled accordingly.

Each dispatch is to leave with all accumulated orders at its time of dispatch. By BHH, we model the time taken to serve the orders which accumulate in a zone over a time period of length $t$ to be $f(t) = 2\rho + \gamma A \sqrt{t}$, where $A$ is the area of the zone; this includes the linehaul travel time to and from the zone plus the travel time within the zone. Let $\omega_d$ represent the fraction of orders assigned to the $d$th dispatch. For a fixed number of dispatches $D$ and a given $\rho$, we formulate the following model to find the maximum area of a zone whose centroid is $\rho$ away from the depot:

$$\max \omega \quad A_D$$

$$\text{s.t.} \quad N \sum_{i=1}^{d} \omega_i + 2(D - d + 1)\rho + \sum_{j=d}^{D} \gamma A_D \sqrt{N\omega_j} \leq T \quad \forall d \in [D]$$

$$\sum_{d=1}^{D} \omega_d = 1, \quad \omega \geq 0 \quad \text{(M.1)}$$

For each $d \in [D]$, constraints (M.1) ensure that the total routing time of the $d$th dispatch and any subsequent dispatches does not exceed the difference between the $d$th dispatch time and $T$. For all feasible $\rho$, let the function $A_D(\rho)$ represent the optimal value of the model.

From a planning perspective, we are primarily interested in the cases where $D$ is small. While the model does not readily admit a closed-form optimal solution in general, in our work we derive relatively simple numerically solvable expressions for $D = 1, 2,$ and $3$.

When $D = 1$, it is necessarily true that $\omega_1 = 1$. Thus,

$$A_1(\rho) = \frac{T - N - 2\rho}{\gamma \sqrt{N}}$$

(1)
When $D = 2$, (M.2) allows us to set $\omega_2 = 1 - \omega_1$ and optimize over $\omega_1$. From (M.1), we are able to show that

$$A_2(\rho) = \frac{1}{\gamma \sqrt{N}} \sup_{\omega_1} \left\{ \min \left\{ \frac{T - \omega_1 N - 4\rho}{\sqrt{1 - \omega_1}}, \frac{T - N - 2\rho}{\sqrt{1 - \omega_1}} \right\} \right\}.$$  \hspace{1cm} (2)

Both expressions inside the inner braces are continuous with respect to $\omega_1$, allowing us to solve for $A_2(\rho)$ using a simple numerical one-dimensional minimizer (such as fminbound).

Similarly, for $D = 3$, we show that

$$A_3(\rho) = \frac{1}{\gamma \sqrt{N}} \sup_{\omega_1, \omega_2} \left\{ \min \left\{ \frac{T - \omega_1 N - 6\rho}{\sqrt{1 - \omega_1}} + \frac{T - (\omega_1 + \omega_2) N - 4\rho}{\sqrt{1 - \omega_1 - \omega_2}} + \frac{T - N - 2\rho}{\sqrt{1 - \omega_1 - \omega_2}} \right\} \right\}.$$  \hspace{1cm} (3)

In this case, it is necessary to optimize over both $\omega_1$ and $\omega_2$. However, a one-dimensional minimizer can still be used to find $A_3(\rho)$ by defining a function which fixes $\omega_1$ and optimizes over $\omega_2$, then optimizing this function over $\omega_1$. These area expressions and minimizing procedures can be directly extended for higher values of $D$. Additionally, for any $D$, a closed-form lower bound for $A_D(\rho)$ can be derived by setting $\omega_1 = \omega_2 = \ldots = \omega_D = \frac{1}{D}$.

It is not always necessarily true that $A_1(\rho) < A_2(\rho) < A_3(\rho) < \ldots$ or that a particular number of dispatches is feasible for a given $\rho$. Suppose we are allowed to use at most $D^*$ dispatches. Define

$$\bar{A}_{D^*}(\rho) = \max_{D \in \{D^*, \ldots \}} \{A_D(\rho) \mid D \text{ dispatches are feasible} \}.$$ 

Then, for a given $\rho$, the number of vehicles required per unit area is the reciprocal of $\bar{A}_{D^*}(\rho)$. Integrating over the service region $R$ gives an approximation for the minimum number of vehicles $V(R)$ required to serve the region (Erera 2000):

$$V(R) \approx \int_R \frac{1}{\bar{A}_{D^*}(\rho)} d\rho.$$  \hspace{1cm} (4)

3. Results

![Figure 2](image.png)  \hspace{1cm} $A_1(\rho), A_2(\rho),$ and $A_3(\rho)$ for $N = 75$, $T = 100$, and $\gamma = 1$

In numerical tests for small $D^*$, we observe that $\bar{A}_{D^*}(\rho)$ is continuous and smooth or piecewise smooth; the exact behavior depends on $D^*$ and the ratio between $N$ and $T$. Figure 2 plots $A_1(\rho)$,
$A_2(\rho)$, and $A_3(\rho)$ for $N = 75$ and $T = 100$ time units with $\gamma = 1$. In this case, note that $\bar{A}_3(\rho) = A_3(\rho)$. However, for smaller values of $N$ and some values of $\rho$, $\bar{A}_3(\rho) = A_2(\rho)$ or $\bar{A}_3(\rho) = A_1(\rho)$.

For these parameters, if $R$ is a disk of radius 6 time units with a centrally located depot, integrating the curves in Figure 2 as per (4) gives $V(R) \approx 59, 38$, and 34 vehicles for $D^* = 1, 2$, and 3, respectively. Such empirical evidence suggests rapidly diminishing savings in the number of vehicles required as $D^*$ increases. Analytically quantifying and bounding the rate of these savings will allow for a clearer understanding of optimal solution behavior, especially for larger $D^*$.

For $D = 2$ and $D = 3$, we observe, empirically, that exactly one of the following two alternatives hold at the optimal $\omega$ and $A_D$:

- the constraints (M.1) hold at equality for all $d \in [D]$, or
- for at least one $d \in [D]$, $\omega_d = 0$, and $A_D \leq A_{D-1}$.

In the second case, at least one dispatch has zero quantity, so the vehicle travels to the centroid of the zone and back without serving any orders. It follows that any such dispatches can be removed, implying that the maximum zone area can be achieved with fewer than $D$ dispatches.

4. Conclusions and Future Work

In this work, we propose a methodology for fleet sizing in same-day delivery systems to be used in service region partitioning. Using continuous approximations on order arrivals and vehicle routing times, we propose an optimization model to determine maximum zone areas as a function of a zone’s distance from the depot, and we show how these areas can be used to determine the number of vehicles required to serve the entire region. Continued work is focused in three directions:

(i) Developing closed-form expressions and bounds, such as for the values of (2) and (3).

(ii) Translating continuous approximation fleet sizing results into a feasible partitioning of a service region, as in Ouyang and Daganzo (2006), and validating theoretical results using real order data and road networks.

(iii) Relaxing the assumption that each zone within the service region is served by a single vehicle.

To this end, we are studying the routing cost minimization problem for general finite fleets and strengthening cost bounds derived from two distinct families of dispatching policies.

References


