On Planning and Design of Logistics Systems for Uncertain Environments

Carlos F. Daganzo

Department of Civil and Environmental Engineering
Institute of Transportation Studies
University of California, Berkeley

Alan L. Erera

Department of Industrial Engineering and Operations Research
University of California, Berkeley

January 5, 1999

Abstract

1. This paper addresses some issues that arise in the planning and design of logistics systems when the environment in which they are to be operated cannot be modeled accurately with certainty. The paper describes the analytical difficulties introduced by explicitly considering uncertainty, and suggests possible modeling steps that may result in more efficient, uncertainty-friendly plans.

1 Introduction

2. The two main goals of this paper are: (i) to describe the difficulties introduced by uncertainty in the planning and design of logistics systems, and (ii) to suggest approximate methods to systematically analyze the effects of uncertainty. The ideas are illustrated by means of two examples.

3. The effectiveness of conventional mathematical analysis methods, e.g. numerical optimization and optimization-based heuristics, for solving large-scale transportation/logistics problems involving deterministic data is well known. Example applications include vehicle routing, as indicated by the extensive literature on the "VRP" problem (see Fisher [10] and Bramel and Simchi-Levi [2] for recent reviews), and network problems such as the airline fleet assignment problem (see, e.g. Rushmeier and Kontogiorgis [14] or Hane *et al.* [13]) and the crew pairing problem (see, e.g. Vance *et al.* [15]).

4. Unfortunately, the standard methodologies are difficult to apply when uncertainty is a significant issue (i.e. for planning and design problems) and the solution effectiveness is notably reduced. In traditional *stochastic programming* approaches, approximate deterministic formulations are employed where uncertain values are replaced by expected values or by percentiles, but this is only appropriate for some problems and cannot always be done accurately and realistically. *Stochastic optimal control theory* and *dynamic programming* offer better ways to incorporate randomness into the optimization of systems that evolve over time (or another single dimension) but the scope of the problems that can be solved in this way is extremely narrow. The extensive literature that exists on the relatively simple problem of determining optimal inventory re-order policies from a single store [12] is an indication of the difficulties introduced by randomness. Thus, it should not be surprising to see that to solve large-scale problems involving uncertainty analysts invariably resort to heuristic formulations; e.g. of the "rolling horizon" type.

5. It should be clear that if one cannot anticipate when extra resources will be needed for a given task, a logistics system must be redundant; e.g. by maintaining larger inventories, using larger vehicle fleets, or by some other means. The challenge is to determine the most cost-effective form of redundancy required, and an operating/control strategy that will be able to exploit it. The goal of an analysis should be to explore the broadest possible space of system designs using an objective function that properly captures uncertainty. Since carefully idealized systems often can be examined accurately in generality, it is suggested in this paper that the possible forms

of redundancy should always be explored systematically with idealized models before embarking on a detailed numerical analysis.

6. Using two deterministic examples, Section 2 examines the issues introduced by uncertainty. Section 3 then describes its conventional treatment, the simplifications that are usually made, and suggests possible remedies; the examples of Section 2 are analyzed as proposed. Section 4 provides some closing comments.

2 Deterministic Analysis and Uncertainty

2.1 The static vehicle routing problem (VRP)

7. The vehicle routing problem has many variants and we consider here the problem of minimizing the transportation cost required to deliver (or collect) lots of small but varying sizes from a set of scattered customers with vehicles of fixed capacity, V. Transportation costs are assumed to be a linear function of the fleet size and the total vehicle distance traveled.

8. Suppose now that the problem involves many customers and many vehicle tours, and that a customer's demand can be split between vehicles. An efficient strategy in this case is to divide the service region into non-overlapping delivery zones containing V units of demand, elongating these zones toward the depot with a width that depends on the local density of customers, δ, as shown in Figure 1(a) and explained in [5], and then to route a vehicle within each zone with an "up and down" strip strategy [6]. If the delivery lot of the last customer in a tour does not fit in the vehicle, then that customer should also be visited by the following tour. If the delivery zones dove-tail reasonably well, then the distance of the VRP can be approximated by the integral over the service region of the following expression [5]:

$$2r\delta/C + 0.57\delta^{1/2} \tag{1}$$

which represents the delivery distance per unit area. In this expression r is distance from a point in the delivery zone to the depot and C is the average number of stops made by a vehicle; i.e. the ratio of V to the average delivery lot size, v. We assume for clarity of exposition that C is independent of location but δ may vary. Let us now examine the effect of using overlapping zones (redundancy).

9. If the stops in an area $\mathcal{A}_3$ were to be allocated to two different tours, as shown in Figure 1(b), the calculation would be different. One would have to calculate the distances for tours 1

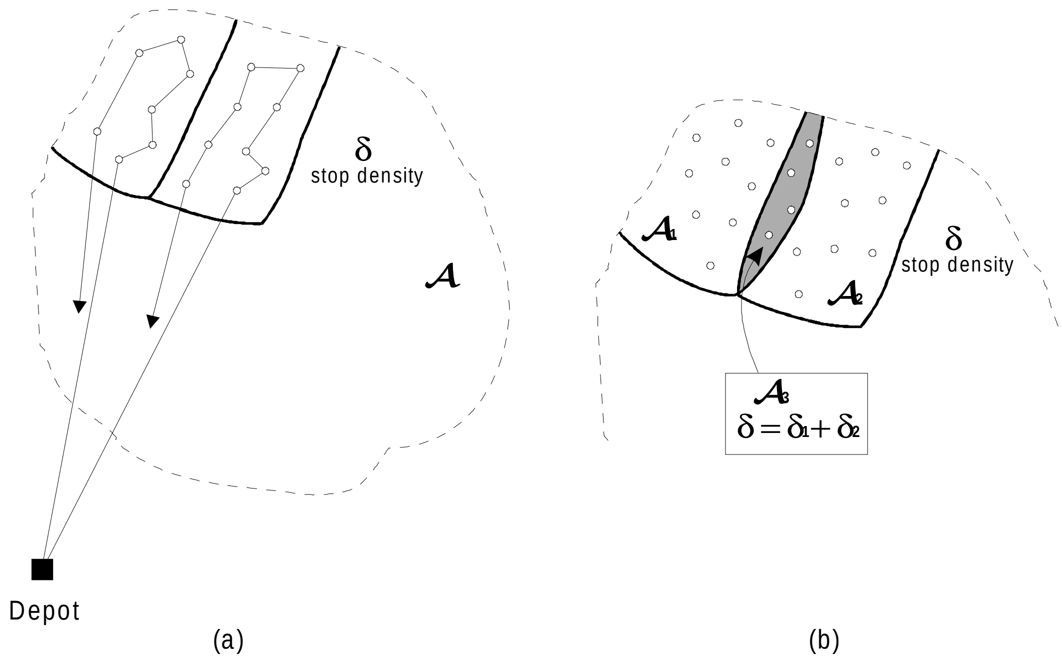

Figure 1: Non-overlapping (a) and Overlapping (b) Vehicle routing zones

and 2 separately by integrating (1) over the two zones $\mathcal{A}_1 \cup \mathcal{A}_3$ and $\mathcal{A}_2 \cup \mathcal{A}_3$, using in each case the proper customer density within $\mathcal{A}_3$. Suppose the customer densities for tours 1 and 2 in $\mathcal{A}_3$ are $\delta_1 > 0$ and $\delta_2 > 0$ ($\delta_1 + \delta_2 = \delta$). Consideration shows that if this is done then the total distance always exceeds that of the non-overlapping case by an amount:

$$\Delta = \int_{\mathcal{A}_3} 0.57[(\delta_1^{1/2} + \delta_2^{1/2}) - (\delta_1 + \delta_2)^{1/2}]da \geq 0 \tag{2}$$

where da is the differential of area. Note that Δ can never be negative, independent of our choices for $\mathcal{A}_3$, δ_1 and δ_2, because the square root function is subadditive. This result suggests that geographical areas should be served (non-redundantly) by single vehicles, but assumes that tours can be built with perfect a-priori information regarding lot sizes.

10. If customer locations and/or lot sizes are uncertain when planning, the fixed-zone strategy may be impractical, since the demand of some zones may exceed vehicle capacity. The desirability of alternative schemes then will depend on how and when lot size information becomes available and the degree of control that a dispatcher can exert over en-route vehicles. Researchers have attempted to address the problem when customer lot size information becomes known only after the arrival of the vehicle. Unfortunately, all of the mathematical algorithms that have been

proposed to date are based either on operating conditions that are unlikely to be feasible in practice as occurs for TSP partitioning heuristics [1], or on feasible forms of operation that are too restrictive to be appealing in practice. More discussion of these issues can be found in [9].

11. Demand that is uncertain prior to vehicle arrival may be managed for example by designing delivery zones as if the vehicle capacity were smaller ($V^- < V$) to ensure that few tours would overflow, and then serving the overflow customers with a set of secondary "sweeper" tours [7]. Gendreau *et al.* [11] optimizes such a scheme, but assumes quite restrictively that each sweeper tour can serve only the customers left behind by a single primary tour. More appealing ways of introducing redundancy exist but they are difficult to optimize with numerical methods. For example, redundancy can be introduced by: (i) eliminating the single tour restriction, (ii) designing overlapping routes as in Figure 1(b) to allow vehicles to cover for one another, and (iii) instructing vehicles with remaining capacity after their last delivery to stay where they are (or even reposition to strategic locations) so that they can more efficiently "sweep" the overflow. A mixed strategy combining elements of (i), (ii), and (iii) also may be desirable. Section 4 will show how strategy (i) can be designed using idealized models as an evaluation tool.

2.2 The warehouse location-inventory-routing problem (WLIRP)

12. The second example involves determining the number and location of warehouses to be supplied from a factory, and the vehicle routes and delivery schedules from the warehouses that are needed to serve a set of customers with time-dependent demands. The objective is to minimize the sum of the transportation, warehousing and customer inventory costs. This planning problem is very common; it arises for example in companies such as Clorox (consumer goods) and Safeway (grocery stores). As explained in [8], efficient designs for this type of problem do not require geographical redundancy when demands are known. Furthermore, detailed designs can be obtained via numerical optimization, as explained below.

13. Let x_{ij} be the distance from warehouse i to customer j. If the transportation cost c_{ij} of delivering d_{ij} items to j from i can be expressed as $c_{ij} = A_j + B_i d_{ij} x_{ij}$, independently of how many items are delivered to other customers, and if the transportation costs from the factory, o, to warehouse i are proportional to the item-Kms sent, $d_{io} x_{io}$ (so that cost $= B_i' d_{io} x_{io}$), then it is relatively easy to find efficient system designs as explained in paragraph 14. The two cost expressions just introduced are good approximations for many forms of transportation, although this may not always be apparent. For example, if deliveries from every warehouse occur with

VRP tours under the conditions described in paragraph 8, then the proposed expression for c_{ij} holds with $A_j = 0.57\delta_j^{-1/2}$ and $B_i = 2/V_i$. (The subscripts j and i have been used with δ and V to stress that the former parameter may vary across customers and the latter may vary across warehouses.)

14. The purpose of this paragraph is to establish the "easy" nature of the deterministic problem. The paragraph may be skipped without loss of continuity. For ease of exposition, it is assumed that all warehouses dispatch vehicles simultaneously at times $\{\tau_k\}$, and that each customer is served instantaneously with each dispatch*. If the cumulative customer demands as a function of time $D_j(t)$ are known then, conditional on two consecutive warehouse dispatch times τ_{k-1} and τ_k, one can calculate customer inventory costs for the intervening interval independent of the location of the warehouses.† The best dispatch schedule with a given number of dispatch intervals, K, and the resulting inventory cost, $z^*(K)$, can then be found with dynamic programming. Conversely, and quite fortunately, transportation costs depend on the schedule only through K. To see this, define an indicator decision variable, $\gamma_{ij}^{(k)}$, which is 1 if customer j is served from warehouse i in period k and 0 otherwise, and let $d_j^{(k)} = D_j(\tau_k) - D_j(\tau_{k-1})$ denote the demand of j in the kth interval. The transportation cost for customer j in this interval is then:

$$\sum_i \gamma_{ij}^{(k)}(A_j + B_i d_j^{(k)} x_{ij} + B_i' d_j^{(k)} x_{io}) \qquad \text{for } j \text{ fixed.} \tag{3}$$

The sum of (3) across all j and k is the total transportation cost. It should now be clear from the functional form of (3) that for any fixed set of x's (warehouse locations) and d's (dispatch schedules) the total transportation cost is minimized by setting $\gamma_{ij}^{(k)} = 1$ for the warehouse i that minimizes $B_i x_{ij} + B_i' x_{io}$. Because these terms are independent of $d_j^{(k)}$, the optimum allocation is the same for all dispatch intervals. Therefore, we can replace $\gamma_{ij}^{(k)}$ with γ_{ij} in the formulation. On recognizing that $\sum_k d_j^{(k)} = D_j(\tau_{\text{end}})$ is a constant, we can simplify the expression for the total transportation cost for all customers across all time periods to read as follows:

$$\sum_j \sum_i \gamma_{ij}(KA_j + B_i D_j(\tau_{\text{end}})x_{ij} + B_i' D_j(\tau_{\text{end}})x_{io}) \tag{4}$$

which can be further simplified to:

$$\sum_j \sum_i \gamma_{ij}(B_i D_j(\tau_{\text{end}})x_{ij} + B_i' D_j(\tau_{\text{end}})x_{io}) + K\sum_j A_j \tag{5}$$

*These assumptions can be relaxed, but doing this is beyond the scope of this paper.

†Warehouse inventories can be neglected because, given advance knowledge of demand, inbound shipments can be planned to arrive "just-in-time" for dispatch; this is the "cross-docking" role of warehouses.

since $\sum_i \gamma_{ij} = 1$ for all j. Since the number of warehouses is a variable, to solve the design problem we should add to (5) a term representing the fixed costs of opening warehouses at different locations i. For a fixed K the last term of (5) can be ignored and the remaining part of the objective function has the standard form of a location-allocation problem with a variable number of warehouses. This problem is "easy" to solve, and the resulting cost is denoted c^*. Hence, it is a simple matter to find the minimum of $z^*(K) + K \sum_j A_j + c^*$ over K, which gives the complete solution.

15. Uncertainty in customer demands, and the way in which uncertain demand becomes known as control decisions are made, complicates matters considerably. In addition to the decision variables considered in paragraph 14, one needs to determine appropriate "safety stock" inventory levels at the warehouses which can be used to absorb demand fluctuations during the orders' lead times. The status of the inventory stocks at any given time can also be used to decide if and how to adjust the basic ordering scheme and warehouse-customer allocations in real time. Unfortunately, determining optimal or near-optimal ways of doing so remains an unsolved problem.

16. One simple approach to this problem assumes that the warehouse-customer allocation is fixed (denoted $\mathcal{L}$), and allows the warehouse stocks to be replenished dynamically by varying the ordering frequency or the order size in response to changing demand. As suggested in [8] for the deterministic problem, and shown in Figure 2, one possible system design carves the service region into influence areas with centrally located warehouses, and all customers within an influence area are allocated to its warehouse. This method does not utilize "geographic redundancy" in the form of influence area overlap, and the warehouses can be controlled/operated independently. Individual warehouse safety stocks provide the buffer against uncertainty. Guidelines for the design and evaluation of this configuration can be derived easily (e.g., see [7]).

17. A more general but more complicated approach (suggested in Cheung and Powell [3]) would treat customer-warehouse allocations as control variables that depend on the inventory positions of the warehouses at the time of dispatch. By allowing customer shipments to come from more than one warehouse in this dynamic fashion, it should be clear that safety stocks can be reduced at the expense of higher transportation costs. Unfortunately, the formulation in [3] is unrealistic because the system's final state is not required to be equal to its initial state, and thus it ignores important future costs. Because these are hard to quantify, no way has yet been found of formulating this problem in detail without introducing a (heuristic) "rolling horizon" fix. This problem will be examined in a different way in Section 3.

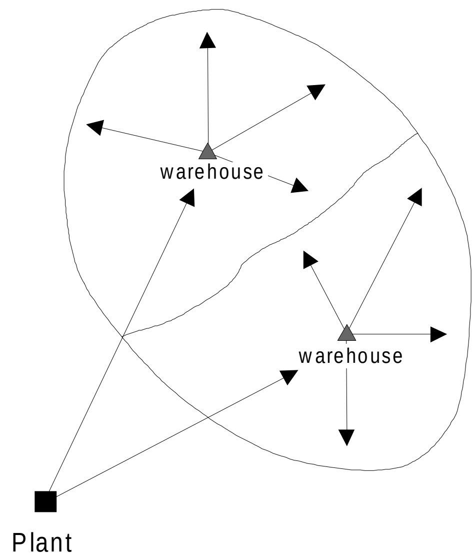

Figure 2: Influence areas with centrally-located warehouses

18. Section 2.1 (paragraphs 10 and 11 in particular) illustrated the difficulties introduced by randomness, and paragraphs 15 and 16 described the added difficulties introduced when one wants to design dynamic strategies over long time horizons; i.e. strategies that can be revised over time as information becomes available. Space considerations preclude us from discussing more complicated systems, such as many-to-many airline networks with supply uncertainty, but it should be clear that the same difficulties should arise in those cases; an expanded discussion of these issues can be found in [9]. The technical nature of the problem and the help that can be derived from simplified analyses are explained in the next section.

3 Treatment of Uncertainty

3.1 Conventional approach

19. Figure 3 contains a flowchart with the various components of a logistics problem. Decision variables are classified as being either of a "design" or "control" type. Design variables $\boldsymbol{D}$, such as the location and number of warehouses in the problem of Section 2.2, are chosen at the beginning of the study and have a lasting influence. Control variables $\boldsymbol{U}$, such as the dispatching times and requested amounts, are chosen dynamically by means of a strategy $\boldsymbol{S}$ while the system is in operation, assuming full information of the system history at each particular decision point.

Optimization tools such as mathematical programming or stochastic optimal control theory can be applied to solve the control problem for a given system design.

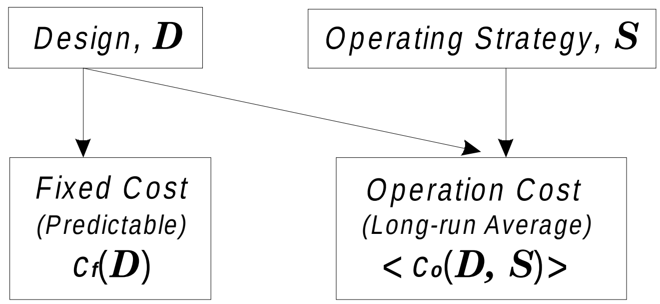

Figure 3: Logistics Problem Components

20. When successful, these tools find an algorithm (or strategy) $\boldsymbol{S}^*(\boldsymbol{D})$ that identifies the best possible set of dynamic controls—in the sense that the expected cost of operating the system with any other strategy $\boldsymbol{S}$, $< c_o(\boldsymbol{D}, \boldsymbol{S}) >$, always exceeds or equals the expected cost of operating it with $\boldsymbol{S}^*(\boldsymbol{D})$. This minimum expected cost is denoted $R(\boldsymbol{D})$ and, by analogy to stochastic programming, will be called the (design) recourse function.

21. The figure also illustrates that: (i) there are fixed design costs $c_f(\boldsymbol{D})$, (ii) the objective of the problem is to find the best design/control combination, and (iii) this may be achieved with a two-step process. The inner loop of this process identifies $\boldsymbol{S}^*(\boldsymbol{D})$ and $R(\boldsymbol{D})$.

22. If the set of allowable control strategies is very broad and the control problem is solved optimally, then experience shows that the design recourse function is usually: (i) very difficult to obtain, and (ii) of an unfavorable form for the outer optimization loop with respect to $\boldsymbol{D}$.

23. In view of this, it makes sense to simplify the control problem by limiting the search to a carefully chosen subset of all possible control strategies. It is particularly useful if the elements of this restricted set can be described in terms of numerical parameters $\boldsymbol{P}$ because then one can replace the mapping $< c_o(\boldsymbol{D}, \boldsymbol{S}) >$ with an ordinary function, $< C_o(\boldsymbol{D}, \boldsymbol{P}) >$. An example of this parameterization occurs in inventory control theory where the family of so-called (s, S)-reorder strategies is used as a proxy for all possible strategies[‡]. Of course, we should make sure that our subset of possible control strategies includes efficient near-optimal strategies, and that the function $< C_o(\boldsymbol{D}, \boldsymbol{P}) >$ is of a favorable form for optimization. Simplifications that achieve these goals may not be easy to find.

[‡] (s,S) policies are described by two parameters: the reorder trigger point and the fixed reorder quantity.

24. Therefore, one may want to simplify the design problem while ensuring that reasonable forms of redundancy are retained in the formulation. One good method consists in considering an idealized problem with symmetries that may reduce the number of decision variables in the combined design/control problem by several orders of magnitude. The idealized problem, which can be solved exactly, can then be a realistic testbed for design alternatives provided that the simplifications do not eliminate the phenomena of interest. Choosing a proper idealization is an art more than a science, but it is critically important. The right simplifications can help us eliminate from further consideration redundancies that are clearly inappropriate for a given case, and in this way narrow the scope of the non-idealized design problem to a manageable level. The next two subsections describe two idealized models that can be used to think about the problems described in Sections 2.1 and 2.2, and how the insights gained may help define design guidelines for the non-idealized problem.

3.2 The static VRP with uncertain demand

25. We show here how certain simplifications can be used to investigate designs for the static VRP with uncertain demand, VRP(UD). Of the three forms of design redundancy discussed in paragraph 11, we choose to evaluate (i); see Figure 4. Determining the primary delivery zones, $\mathcal{A}$, is the design problem, and choosing the routes of the secondary vehicles is the control problem. Construction of the primary vehicle routes is part of the design problem if the customer locations are known, and part of the control problem otherwise. Here we assume that the locations are known, but the methodology changes little if they are not. The main issue is selecting the size of the delivery region $A = |\mathcal{A}|$ because this entails a tradeoff between primary and secondary delivery costs. We show below how a simplified analysis of a continuum model can help generate a design.

26. In addition to the notation of Section 2.1, let μ be the coefficient of variation of the (uncertain) customer lot size. If the distribution of lot sizes is one where the central limit theorem holds approximately (e.g. if there are more than a few stops per tour), then the number of uncollected items in one zone is the non-negative part of a normal random variable, as in the well known "newsboy problem". For our problem, it is not difficult to show that the fraction of items overflowing, f, only depends on two parameters, α and β, which are:

$$\alpha^2 = \mu^2/(\delta A) \qquad \text{and} \qquad \beta = (V/v)/(\delta A) \tag{6}$$

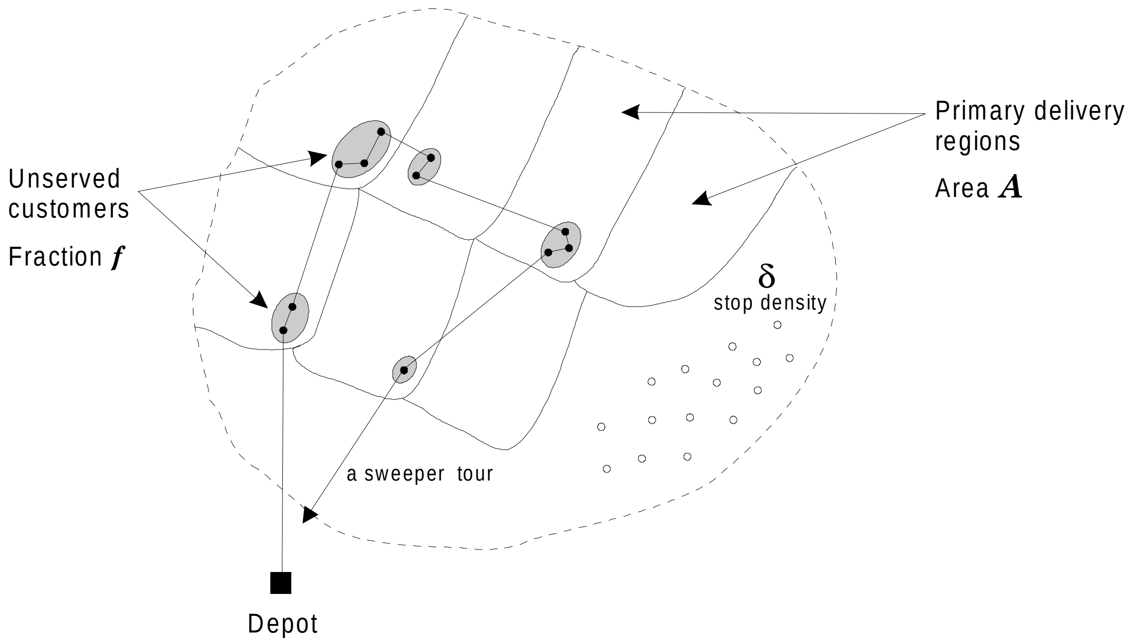

Figure 4: Primary/Secondary Operating Strategy for VRP(UD)

The first relation is the ratio of the coefficient of variation squared and the number of stops available in the zone; the second relation is the ratio of the average number of stops the vehicle can make and those available. The fraction f can be shown to be:

$$f = \alpha\Psi((\beta - 1)/\alpha) \tag{7}$$

where $\Psi(z)$ is the integral of the standard normal c.d.f. (cumulative distribution function), Φ, from $-\infty$ to $-z$. As shown in Figure 5(a), this function decreases toward zero; it can be expressed in terms of the standard normal density $\phi(z)$ and c.d.f.: $\Psi(z) = \phi(z) - z\Phi(-z)$. If customer lot sizes are mutually independent and small relative to the vehicle size, then the overflow fraction f is also approximately the fraction of customers that remain unserved; therefore, $f\delta$ is the density of customers for the secondary tours. We note that (6) and (7) imply a relation $f = F(A)$ between the overflow and our decision variable, and that this relation has an inverse $A = G(f)$; see Figure 5(b). Therefore, we can use f instead of A as the decision variable in the manipulations below.

27. If we imagine that the secondary stops are uniformly distributed, rather than clustered around corners of overflowing delivery regions (see Figure 4), we can write an expression for the total distance traveled per unit area for both the primary and secondary tours, using equation (1). (Consideration shows that the effect of clustering is so minor that it can be ignored in this type of analysis.) We may also want to add a level-of-service penalty k for every customer served

with a secondary tour; i.e., a term of the form $kf\delta$ for every unit area. The resulting distance per unit area is:

$$2r/G(f) + 2r(\delta/C)f + 0.57\delta^{1/2}[(1-f)^{1/2} + f^{1/2}] + kf\delta \qquad (8)$$

The first two terms represent the line-haul distance traveled by primary and secondary vehicles, and the third term the combined local delivery distance. The four components of (8) are plotted on Figure 5(c). As one may expect intuitively, the main trade-off occurs between the primary

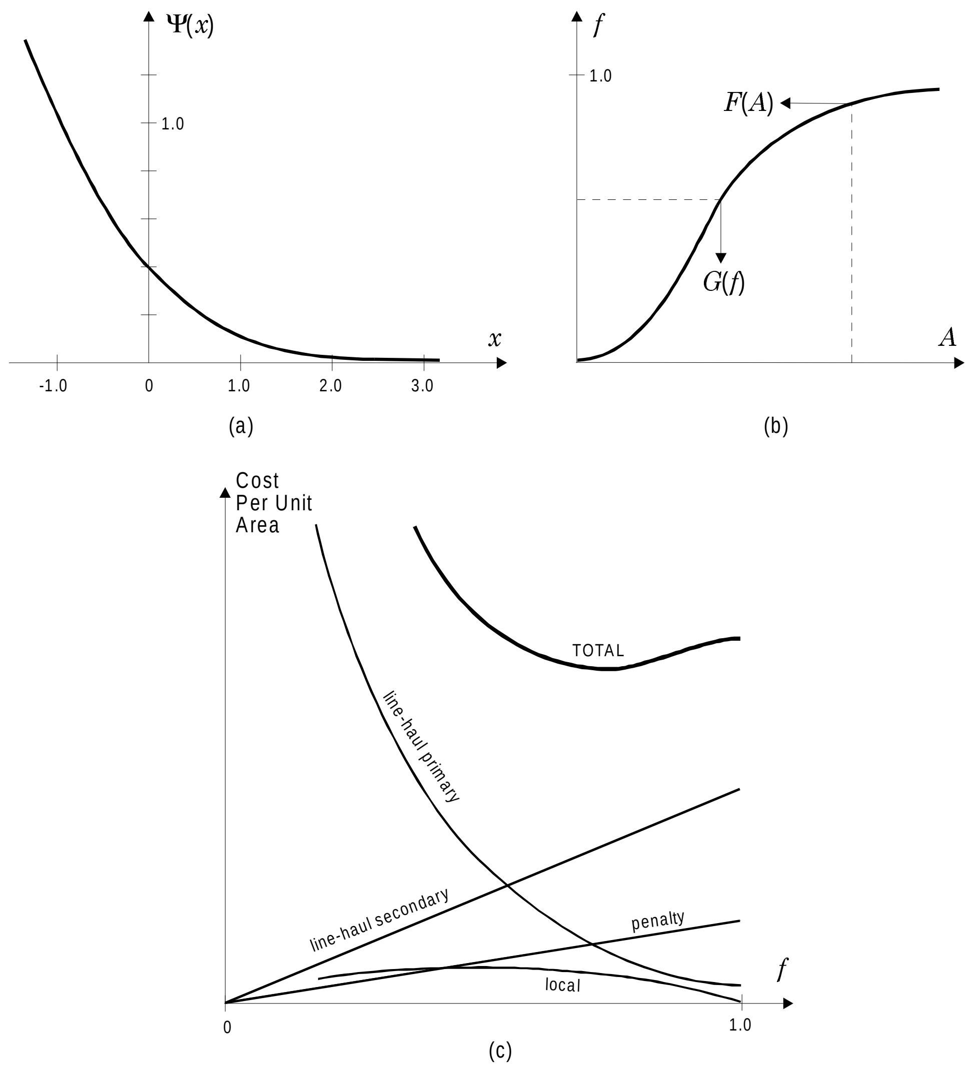

Figure 5: Analysis of the VRP(UD): (a) $\Psi(x)$, (b) $F(A)$ and $G(f)$, (c) Cost Per Unit Area as Function of f

and secondary line-haul costs. Examination of (8) reveals at a glance how the optimum value of f (and therefore A) depends on the parameters of the problem. For example, we see clearly that as r increases the last 3 components of (8) become relatively smaller, and therefore that the optimum f increases as we move away from the depot. Thus, we may want to use smaller primary zones near the depot.

28. One can also explore how the solution to our problem depends on δ, v, μ, etc., if these parameters change geographically. Then the optimum solution of (8) for suitable values of the parameters would indicate the desirable zone size that should be used in various geographical subregions of the service region. This information could then be used to generate a design, e.g., as done by Clarens and Hurdle [4] for a related problem and further discussed in [7].

29. Given this design, each primary vehicle follows the TSP tour constructed between the depot and the customers in its zone, returning to the depot when its capacity is reached and possibly leaving some customers unserved. The control problem is then to determine secondary vehicle tours through these skipped customers, and this is an ordinary deterministic VRP. Thus, the proposed methodology leads to practicable solutions of the combined design/control VRP(UD) problem. Analysis shows that these solutions are more efficient than those requiring sweeper tours to serve customers within a single zone, even if the latter problem can be configured closer to optimality (e.g. as proposed in Gendreau *et al.* ([11]).

30. The modeling approach of paragraphs 26 and 27 is quite useful. It has been proposed for the inventory-routing problem with uncertain demands [7] and can also be applied to other possible strategies for the VRP(UD); e.g., those that allow for overlapping delivery zones and for tours that do not return immediately to the depot as in strategy (iii) of paragraph 11. We are currently investigating these strategies and plan to conduct numerical tests to evaluate performance.

3.3 The warehouse-location-inventory-routing problem with uncertain demand, WLIRP(UD)

31. The complications introduced by uncertain demands in the WLIRP were mentioned in paragraphs 15-17. They are foreboding due to the multi-stage nature of the problem. As a result, no "exact" algorithm has been found for this problem, even for drastically simplified versions of it.

32. To reduce these difficulties to a manageable level while retaining sufficient flexibility to reduce safety stocks, we propose partitioning the set of warehouses into fixed subsets of size n to which customers are statically allocated. Warehouse subsets would "share" a safety stock

chosen to ensure that customer demand is met with very high probability. To prevent stockouts at individual warehouses, customers would be dynamically allocated within their subset in a transportation-efficient way. We would expect the reduction in safety stock to increase with n, but to be bounded from above, and the transportation costs also to increase with n albeit in a different way. We do not know the precise form of the latter relation but believe that it increases rather rapidly with n (see paragraph 38), and that there is a small $n = n^*$ which optimally balances the inventory savings with the transportation penalty. As an illustration of the modeling approach, we examine below the costs for the special case with $n = 2$ in some detail. (It is shown that with $n = 2$ the benefits of dynamic allocation almost always outweigh the drawbacks; i.e. that $n^* \geq 2$ in most cases.) Results for large n are also given without a derivation. They suggest that n^* should not be large.

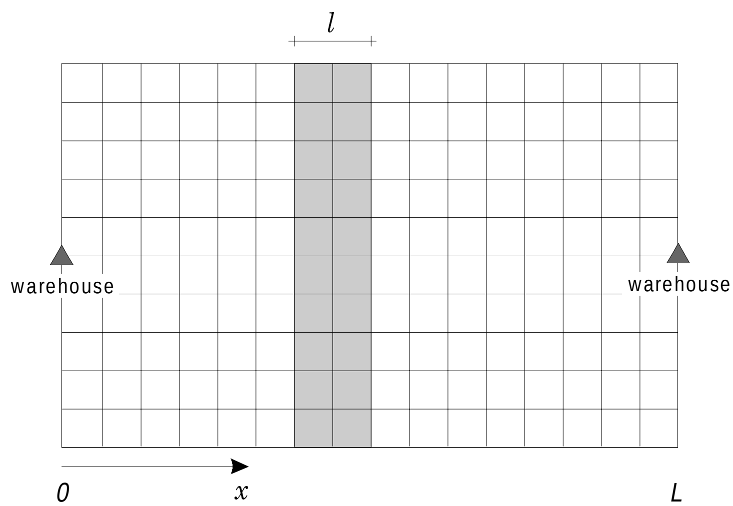

Figure 6: Idealized System for WLIRP(UD) with $n = 2$

33. We consider now the simplest possible example (Figure 6) which exhibits the aforementioned issues. It includes two warehouses centered on opposite sides of a rectangular service region, with base length L distance units. Travel on this region is permitted vertically and horizontally (L_1 metric)[§]. The demand in a vertical slice of the region ranging from abscissa x to $x + \ell$ during time $t, t + \tau$ is given by $D(x, x + \ell, t, t + \tau)$. Changing unpredictably with time, this demand is assumed to be a stationary process in x and t, with independent increments; accordingly, $D'\ell\tau$ will denote the expectation of $D(x, x+\ell, t, t+\tau)$ and $\gamma D'\ell\tau$ its variance, where γ is the process' index of dispersion. Assume that warehouses serve customers instantaneously (the latter do not carry a safety stock) and order from the factory regularly an amount equal

[§]Note that the shortest paths from either warehouse to a given customer require the same vertical distance, hence only horizontal distances are considered in this discussion to that depleted since the previous order (periodic review system). Units of time are chosen so that the time between warehouse reorders is 1 and this unit is referred to as a "day." Finally, the "lead time" between a warehouse order and the arrival of goods equals T reorder times.

34. Two cases will be compared: (a) *Static allocation*: customers with $x < L/2$ are allocated to the warehouse at $x = 0$ and the others to the one at $x = L$; and the safety stock at each warehouse is chosen to be three standard deviations of its total customer demand during one lead time, $3(\gamma D'(L/2)T)^{1/2}$, so as to ensure that the probability of a stockout is low. (b) *Dynamic allocation*: customers are dynamically assigned to a warehouse each period; and the *combined* safety stock is chosen to be three standard deviations of the lead-time demand in the complete service region, $3(\gamma D'LT)^{1/2}$. The dynamic allocation method is described in more detail below; it ensures that a customer goes unserved only if both warehouses are empty, and achieves this goal with the least possible item-Kms of travel between the warehouses and the customers.

35. *Static allocation evaluation.* The total system safety stock for this strategy is:

$$(18\gamma D'LT)^{1/2} \tag{9}$$

and the average item-Kms of travel in any given day are:

$$D'L^2/4 \tag{10}$$

36. *Dynamic allocation evaluation.* The total system safety stock is only:

$$(9\gamma D'LT)^{1/2} \tag{11}$$

If the inventory positions at the two warehouses at the beginning of a "day" are I_1 and I_2, and the cumulative demand for the "day" as a function of position $d(x)$ is also known, e.g. as shown by the curve in Figure 7, then the best allocation can be obtained graphically as depicted.[¶] We look for a point x^* that defines the influence areas for the day. Note that the item-Kms of travel are given by the shaded areas of the figure. If the demand can be satisfied, i.e. $d(L) < I_1 + I_2$, we first find x_1, x_2 such that $d(x_1) = I_1$, $d(L) - d(x_2) = I_2$, and then choose $x^* = \text{middle}(x_1, x_2, L/2)$, as shown in Figure 7. If the demand is not satisfied, which is rare, then customers in $(0, x_1)$ are served from 0, customers in (x_2, L) from L and those in (x_1, x_2) are lost. If we assume for the purpose of calculating the item-Kms that inventories are at their average positions at the

[¶]In many cases $d(x)$ may not be known sufficiently in advance for us to be able to achieve the best allocation; thus, our derivations are somewhat tilted in favor of this strategy.

beginning of the day, $I_1 = I_2 = D'L/2 + \frac{1}{2}(9\gamma D'LT)^{1/2}$, which also tilts the calculations slightly in favor of this strategy, then an approximate formula for the shaded area is (see appendix):

$$D'L^2/4 + (L/100)(\gamma D'L)^{1/2} \tag{12}$$

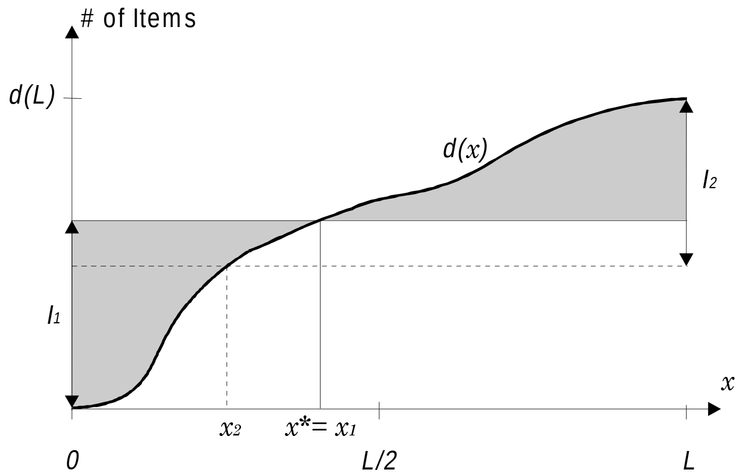

Figure 7: Dynamic warehouse allocation

37. We see from (9) and (11) that the dynamic strategy saves

$$1.24T^{1/2}(\gamma D'L)^{1/2} \tag{13}$$

items in inventory, but also see from (10) and (12) that it induces approximately

$$(L/100)(\gamma D'L)^{1/2} \tag{14}$$

extra item-Kms of travel every “day”. From the ratio of these quantities we see that for every truckload-“day” of inventory saved by the dynamics, a truck has to be driven $(L/124)/T^{1/2}$ Kms. A truckload of inventory for many goods such as automobiles costs on the order of \$30 per day, and also \$30 per “day” if we assume that 1 “day” = 1 day. (This number can be much higher for certain goods, such as jewelry, computer equipment, etc., but such goods may not be transported as described here.) Driving a truck costs on the order of \$1 per Km. Therefore, dynamic allocation will be attractive if $(L/T^{1/2}) < 3720$ Km. This should be the case even if $T = 1$ and the goods are much cheaper.

38. Asymptotic results for large n: We present here generalizations of (13) and (14) that include n as a parameter without a detailed derivation. The results show that n^* should not be large. First note that the rationale that led to (11) and (13) now yields $(1-(1/n)^{1/2})(9\gamma D'LT)^{1/2}$

for the inventory savings per terminal for large values of n, where L is the separation between terminals. The dynamic transportation costs can also be approximated analytically if the warehouse subsets are arranged one-dimensionally, and one uses a simple "greedy" allocation strategy. (Considered sequentially, e.g. from left to right, each warehouse would serve, starting with the last customer served by the previous warehouse, as many customers as it inventory position would allow without encroaching on the territory of the warehouse that follows.) This strategy is sub-optimal but easy to analyze. It is mathematically analogous to a Brownian queuing problem for which formulas exist. We find that the extra transportation cost per terminal increases linearly with $n^{1/2}$ for any given γ, D', L and T, according to the asymptotic formula: $kn^{1/2}(\gamma D' L^3/T)^{1/2}$ where k is a dimensionless coefficient which is $k = 1/6$ if the inventory positions are equal and $k = 1/3$ if they are random. An optimal strategy would treat customers on both sides symmetrically, and this would reduce k by more than a factor of 2. We believe that for an optimum strategy k would be somewhere between 1/10 and 1/25. The extra distance formula is more difficult to derive for other (non-one-dimensional) warehouse groupings but its rapid increase with n, and other qualitative behavior should not change much.

36. The results in paragraph 35 suggest that the optimum n^* is small and that it can be found with the help of simple idealized models. In order to design a system one would have to minimize an approximate "logistic cost function" in which the warehouse influence area diameter (L) and the size of the dynamic subset (n) would appear as decision variables. The dynamic allocation algorithm (control problem) would be relatively easy to solve since it decomposes by warehouse subset and n is small. A discussion of this issue, however, is beyond the scope of this paper.

4 Conclusion

37. As the examples in this paper have illustrated, uncertainty usually requires that redundancies be introduced into a system design. The design game is to determine which kinds of redundancies offer the most benefit for the least cost. If this is difficult to do with detailed models (which is usually the case) an approximate analysis with idealized models may yield the desired insights. Idealized models allow many more forms of redundancy to be evaluated without the ad hoc assumptions of detailed models, which are often limiting and hard to understand. Idealized models can identify efficient strategies that are simple enough to be implemented; i.e. strategies

that allow detailed designs for the original, non-idealized problem to be developed and the control subproblem to be solved, as occurred in the examples of this paper.

38. If prediction accuracy is important one can simulate the chosen design/control configuration (and perturbations to it) to obtain accurate cost estimates; these can be compared with the idealized predictions. In this respect, the most useful optimization methods would seem to be case-specific "meta-heuristics" that would allow us to sort through these perturbations while retaining the flavor of the basic design.

39. If closed form solutions can be developed, the expressions reveal at a glance how the solution depends on the input data. This is useful when proposals have to be made to management. For example, the analytic solution may indicate which data influence costs and which are irrelevant. The former may even suggest alternative problems that management should consider.

40. In closing, we recognize that the methodology proposed in this paper is more an art than a science but also note that once mastered it can be effectively applied quite broadly. We believe that the results of the approach can be very fruitful and hope that this paper will stimulate others to pursue similar avenues of thought in the future.

References

[1] D.J. Bertsimas. A vehicle routing problem with stochastic demand. *Operations Research*, 40:574–585, 1992.

[2] J. Bramel and D. Simchi-Levi. *The Logic of Logistics: Theory, Algorithms, and Applications for Logistics Management*. Springer-Verlag, New York, 1997.

[3] R.K.M. Cheung and W.B Powell. Models and algorithms for distribution problems with uncertain demands. *Transportation Science*, 30:43–59, 1996.

[4] G. Clarens and V.F. Hurdle. An operation strategy for a commuter bus system. *Transportation Science*, 9:1–20, 1975.

[5] C.F. Daganzo. The distance traveled to visit n points with a maximum of c stops per vehicle: an analytic model and an application. *Transportation Science*, 18:331–350, 1984.

[6] C.F. Daganzo. The length of tours in zones of different shapes. *Transportation Research, Part B*, 18B:135–146, 1984.

[7] C.F. Daganzo. *Logistics Systems Analysis*. Springer-Verlag, Berlin, 1991.

[8] C.F. Daganzo and G.F. Newell. Configuration of physical distribution networks. *Networks*, 16:113–132, 1986.

[9] A.L. Erera. Logistics systems and uncertainty. Working paper, Institute of Transportation Studies, University of California, Berkeley, 1998.

[10] M. Fisher. Vehicle routing. In M. Ball, T. Magnanti, C. Monma, and G. Nemhauser, editors, *Network Routing*, volume 8 of *Handbooks in Operations Research and the Management Sciences*, pages 1–33. Elsevier Science, Amsterdam, 1995.

[11] M. Gendreau, G. Laporte, and R. Séguin. An exact algorithm for the vehicle routing problem with stochastic demands and customers. *Transportation Science*, 29:143–155, 1995.

[12] S.C. Graves, A.H.G. Rinnooy Kan, and P.H. Zipkin, editors. *Logistics of Production and Inventory*, volume 4 of *Handbooks in Operations Research and the Management Sciences.* Elsevier Science, Amsterdam, 1993.

[13] C.A. Hane, C. Barnhart, E.L. Johnson, R. Marsten, G.L. Nemhauser, and G. Sigismondi. The fleet assignment problem: Solving a large-scale integer program. *Mathematical Programming*, 70:211–232, 1995.

[14] R.A. Rushmeier and S.A. Kontogiorgis. Advances in the optimization of airline fleet assignment. *Transportation Science*, 31:159–169, 1997.

[15] P.H Vance, C. Barnhart, E.L. Johnson, and G.L. Nemhauser. Airline crew scheduling: a new formulation and decomposition algorithm. *Operations Research*, 45:188–200, 1997.

Appendix

A1. Since stockouts are rare, we evaluate the item-Kms only for the case where $x_2 < x_1$. Three possibilities exist: (a) $x_1 < L/2$; (b) $x_2 > L/2$; (c) $x_2 < L/2 < x_1$. The average item-Kms traveled in case (c) will be $D'L/4$, and the average item-Kms for cases (a) and (b) will be larger. Insofar as $I_1 = I_2 = I$, the latter two averages should be equal to each other, by symmetry. Thus, the derivations below focus on case (a).

A2. Consider $d(x)$ now as a stochastic process. We know from the first passage time formulas for processes with independent, positive increments that x_1 is approximately normal with $E[x_1] = I/D'$ and $\text{var}(x_1) = (I/D')(\gamma/D')$. We also know from symmetry considerations that the expectation of the shaded area conditional on x_1, $A(x_1)$, is equal to the area of the two right triangles in Figure 8. (To see this note that for every realization of the process $d(x)$ we can define a dual realization $d'(x)$ by setting: $d'(x) = I - d(x_1 - x)$ if $x < x_1$, and $d'(x) = I + d(L) - d(L - x)$ if $x > x_1$. Our statement is true because dual pairs of realizations partition the sample space and because every dual pair has the same combined area: twice the shaded area.) Therefore,

$$A(x_1) = \frac{1}{2}[Ix_1 + D'(L - x_1)^2] \tag{A1}$$

If we let $\epsilon = L/2 - x_1$ and assume $\epsilon \ll L$ (not many extra miles) the above expression can be simplified:

$$A(\epsilon) = \frac{1}{2}[I(L/2 - \epsilon) + D'(L/2 + \epsilon)^2] = D'L^2/8 + IL/4 + (D'L - I)\epsilon/2 + D'\epsilon^2/2$$

$$\approx D'L^2/4 + (I - D'L/2)L/4 + (D'L - I)\epsilon/2$$

Thus, the expected added miles due to $x_1 < L/2$ are:

$$(I - D'L/2)(L/4)p(\epsilon) + \frac{1}{2}(D'L - I)E[\max(0, \epsilon)] \tag{A2}$$

where $p(\epsilon) = \Pr\{\epsilon > 0\} = \Pr\{d(L/2) > I\} = \Phi([D'L/2 - I]/(\gamma D'L/2)^{1/2})$. Recall that $[I - D'L/2]$ is the safety stock at $x = 0$ which is $\frac{1}{2}(9\gamma D'LT)^{1/2}$ as per (11). Thus, $p(\epsilon) = \Phi(-(9T/2)^{1/2})$, which is on the order of 0.01 or less, and the first term of (A2) becomes:

$$\frac{1}{2}(9\gamma D'LT)^{1/2}(L/4)\Phi(-(9T/2)^{1/2})$$

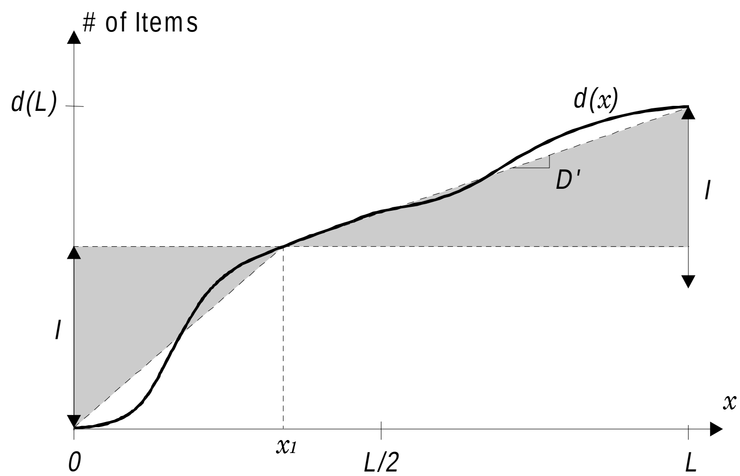

Figure 8: Calculation of Expected Item-Kms of Travel

If we use $I \approx D'L/2$ as an approximation in the expression for the variance of ϵ, the expectation of $\max(0, \epsilon)$ reduces to: $(\gamma L/2D')^{1/2}\Psi((9T/2)^{1/2})$, where Ψ is the previously defined integral of the standard normal c.d.f. Thus, the second term is:

$$\frac{1}{2}(D'L/2 - \frac{1}{2}(9\gamma D'LT)^{1/2})(\gamma L/2D')^{1/2}\Psi((9T/2)^{1/2})$$

and the total miles added become:

$$\frac{1}{2}(9\gamma D'LT)^{1/2}(L/4)\Phi(-(9T/2)^{1/2}) + (L/4)((\gamma D'L/2)^{1/2} - \gamma(9T/2)^{1/2})\Psi((9T/2)^{1/2})$$

Letting $\alpha = (\gamma D'L/2)^{1/2}$, $\beta = L/4$, and $f(T) = (9T/2)^{1/2}$, this expression simplifies to:

$$\alpha\beta\phi(f(T)) - \gamma\beta f(T)\Psi(f(T))$$

When T is close to 1, this expression is closely approximated by:

$$\approx (L/100)(\gamma D'L)^{1/2} \tag{A3}$$

as claimed in the text.