Title
Asymptotic Approximations for the Transportation LP and Other Scalable Network Problems

Authors
Daganzo, Carlos F
Smilowitz, Karen R

Publication Date
2000-12-17

Asymptotic Approximations for the Transportation LP and Other Scalable Network Problems

Carlos F. Daganzo and Karen R. Smilowitz

Institute of Transportation Studies

Department of Civil and Environmental Engineering

University of California, Berkeley CA 94720

Working paper

August 4, 2000

Revised December 17, 2000

Asymptotic Approximations for the Transportation LP and Other Scalable Network Problems

Carlos F. Daganzo and Karen R. Smilowitz
Institute of Transportation Studies and Department of Civil and Environmental Engineering
University of California, Berkeley CA 94720
(August 4, 2000)
(Revised December 17, 2000)

ABSTRACT

Network optimization problems with a "scalable" structure are examined in this report. Scalable networks are embedded in a normed space and must belong to a closed family under certain transformations of size (number of nodes) and scale (dimension of the norm.) The transportation problem of linear programming (TLP) with randomly distributed points and random demands, the earthwork minimization problem of highway design, and the distribution of currents in an electric grid are examples of scalable network problems. Asymptotic formulas for the optimum cost are developed for the case where one holds the scale parameter constant while increasing the size parameter, *N*.

As occurs in some applied probability problems such as the Ising model of statistical mechanics, and the first passage of time of a random walk, the nature of the solution of *linear* problems depends on the dimensionality of the space. In the linear case, we find that the cost per node is bounded from above in 3+-dimensions (3+-D), but not in 1- and 2-D. Curiously, zone shape has no effect (asymptotically) on the optimum cost per point in 2^+-D, but it has an effect in 1-D. Therefore, the 2-D case can be viewed as a transition case that shares some of the properties of 1-D (unbounded cost) and some of the properties of 3-D (shape-independence). A simple formula for the 2-D, Euclidean TLP is given. Asymptotic results are also developed for a class of non-linear network problems with link costs that are a concave power function of flow. It is found that if these functions are strictly concave then the solution in 2+-D is bounded.

Asymptotic Approximations for the Transportation LP and Other Scalable Network Problems

Carlos F. Daganzo and Karen R. Smilowitz

Institute of Transportation Studies and Department of Civil and Environmental Engineering

University of California, Berkeley CA 94720

(August 4, 2000)

(Revised December 17, 2000)

1. INTRODUCTION

Asymptotic formulas exist for the traveling salesman problem, or "TSP" (Eilon et al., 1971, Karp, 1977, Daganzo, 1984a), and for the vehicle routing problem, or "VRP" (Eilon et al., 1971, Daganzo, 1984b, Haimovich et al., 1985, Newell and Daganzo, 1986a and 1986b, and Newell, 1986). The results apply to problems where N points are randomly and homogeneously distributed on a region of a metric plane with area A, and density $\delta = N/A$. In all cases the distance traveled per point for the TSP, or the "detour" distance per point for the VRP,[1] tends to a fixed multiple of $\delta^{-1/2}$ as N and A are increased in a fixed ratio. Hence, this local structure of the TSP and VRP problems allows the use of continuum approximations of the type proposed in Newell (1973) for inhomogeneous problems. Extensions of this type can be used to design and configure many kinds of one-to-many logistics systems; see Daganzo (1999). Hoping to extend these ideas to many-to-many logistics systems, this paper develops similar formulas for a version of the transportation

[1] This is the distance traveled in excess of the lower bound. The lower bound is the product of the (round trip) distance between the depot and a point, and the fraction of a vehicle's capacity consumed by each point.

linear programming problem (TLP) where points lie on a region of a linear normed space. The paper investigates whether the optimal solution of the TLP exhibits a local structure, and also presents asymptotic results for more general network problems.

Unfortunately, the TLP is more difficult to analyze than the TSP because more data are required to define a problem instance, and because the TLP solutions must include some long trips to balance interregional flows. The necessity of long trips suggests that the optimum solution may not have a local structure. Interestingly, as occurs in some applied probability problems (e.g., the Ising model of statistical mechanics, and the first passage of time of a random walk), the nature of the solution depends on the dimensionality of the space. It will be shown that while the TLP is not local in one dimension (1-D), it is local in 3^+-D, and in a more limited sense in 2-D. Related results are derived for more general network problems.

In order to mitigate the above-mentioned difficulties the simplest possible problems are presented first, followed by incremental generalizations. Section 2 introduces terminology and some background on dimensional analysis. Section 3 develops exact formulae for the optimum TLP distance in the 1-D case. Section 4 presents an upper bound for K-dimensional homogeneous problems in a cubic region. It is shown that if one holds δ constant while increasing N and the size of the cube in a constant ratio, then the distance traveled per point is of order N^0 in 3^+-D, as occurs with the TSP and the VRP, of order $log(N)$ in 2-D and of order $N^{1/2}$ in 1-D. Section 5 introduces lower bounds and shows that the upper bounds are asymptotically exact approximations. This section also shows that the distance formulae hold asymptotically for 2^+-D regions with non-cubic shapes, but that this is not true in 1-D. Section 6 shows how the results can be modified for inhomogeneous problems. Finally, in Section 7, results are generalized to a class of 'scalable' network problems.

2. BACKGROUND

2.1 Definitions

2.1.1 The TLP and ATLP

In this paper the TLP is defined as follows. Given are N points, a set of real, non-negative inter-point distances (or costs), $\{d_{ij}, \forall i,j = 1,...N \text{ with } i \neq j\}$, satisfying the triangle inequality, and a set of real-valued net supplies, v_i, at each point, measured in units of "items". Positive v_i are interpreted as supplies and negative values as demands. The goal is to find a set of real, non-negative shipments, $\{v_{ij}, \forall i,j = 1,...N \textit{ with } i \neq j\}$, that minimizes the total distance traveled while satisfying flow balance constraints at each point. Here, a link-based network formulation is adopted.

$$\text{(TLP)} \qquad \min \qquad z = \sum_{ij} d_{ij} v_{ij} \tag{1a}$$

$$\text{s. t.:} \qquad \sum_j (v_{ij} - v_{ji}) \leq v_i \; ; \quad \forall i \tag{1b}$$

$$v_{ij} \geq 0 \qquad ; \quad \forall i, j \, . \tag{1c}$$

Equations (1b) specify flow-conservation at each point, ensuring that the net flow (number of items) emanating from a point i never exceeds the net supply at i.

Problem TLP is feasible only if $\sum_i v_i \geq 0$, as can be seen by summing (1b) across i. The problem is "balanced" and denoted TLP(B) if $\sum_i v_i = 0$. In a balanced problem, constraints (1b) are satisfied as pure equalities. For infeasible TLP problems, we define a feasible (and balanced) auxiliary problem, ATLP, that includes a fictitious source, $i = 0$, with positive net supply,

$v_0 = -\sum_{i=1}^{N} v_i$, and distances, $d_{i0}, d_{0j} = M >> \sup(d_{ij})$. These distances represent a fixed penalty for failing to ship an item. The optimal cost of the auxiliary problem, z^*, includes two parts: a distance component corresponding to the real points, and a penalty component corresponding to the fictitious source. Because M is large, the set $\{d_{ij}\}$, including the fictitious source, continues to satisfy the triangle inequality. Thus, there can be no flow into the fictitious source in an optimum solution. Since the auxiliary problem is balanced, the outflow from the fictitious source must be v_0 with a penalty component v_0M. The distance component is the least total distance that satisfies the most demand, d^*.

For balanced problems, $v_0=0$, TLP ≡ ATLP, and the optimal objective of TLP and ATLP, z^*, is also the least distance required to satisfy the maximum demand, i.e., $z^* = d^*$. In summary, the relationship between the optimum of ATLP and d^* is:

$$d^* = z^* \qquad \text{if the TLP is feasible, } \sum_{i=1}^{N} v_i \geq 0 \tag{2a}$$

$$d^* = z^* - v_0 M \qquad \text{if the TLP is infeasible, } \sum_{i=1}^{N} v_i = -v_0 < 0 . \tag{2b}$$

2.1.2. The DTLP

In the solution of TLP and ATLP it is assumed that if supply exceeds demand, the excess supply is left at the origins. In a variant of ATLP, excess supplies are carried to the extra point, or "depot". This version of the problem will be called "depot-TLP", or DTLP. The DTLP is an ordinary TLP, where the net supply at the depot precisely balances the problem. In the DTLP the depot distances do not have to be fixed or large but they must be non-negative, $(d_{0j}, d_{i0}) \geq 0$, $\forall i, j$, and must satisfy the triangle inequality. The minimum of the DTLP objective function, z_D, will be denoted $d_D{}^*$.

Proposition 1 (DTLP as an upper bound to TLP). For any TLP and its associated DTLP, $d^* \leq d_D{}^*$. Furthermore, if the TLP is balanced, then $d^* = d_D{}^*$ ■

Proof: Recall that if the TLP is infeasible, a fictitious source is introduced in order to obtain z^* and d^*. The ATLP has the same constraints as the DTLP. Therefore, a set of optimal shipments for the DTLP, denoted $\boldsymbol{v}_{\boldsymbol{D}}^*$, is a feasible solution of the ATLP. The associated ATLP distance, including penalties, is denoted $z(\boldsymbol{v}_{\boldsymbol{D}}^*)$. If the DTLP is now adjusted by adding M to all the depot distances, $d'_{0j} = d_{0j} + M$ and $d'_{i0} = d_{i0} + M$, the objective function would become $d_D{}^* + v_0 M$, since v_0 is the total flow to/from the depot in the optimum solution of DTLP. All the link distances in the ATLP are less than or equal to those of the adjusted DTLP, and $\boldsymbol{v}_{\boldsymbol{D}}^*$ is feasible in both cases; therefore, it follows that $z(\boldsymbol{v}_{\boldsymbol{D}}{}^*) \leq d_D{}^* + v_0 M$. Since $z^* \leq z(\boldsymbol{v}_{\boldsymbol{D}}^*)$, it is also true that $z^* \leq d_D{}^* + v_0 M$, and hence that $d_D{}^* \geq z^* - v_0 M = d^*$ for the infeasible case.

In the feasible case, the non-depot flows $(i \neq 0, j \neq 0)$ in the optimal solution of the DTLP incur a cost $d_0 \leq d_D{}^*$. These non-depot flows are a feasible solution of the original TLP since the depot cannot act as a transshipment point in the optimum DTLP solution. Therefore, $d_0 \geq \mathrm{z}^* = d^*$ and it follows that $d^* \leq d_D{}^*$. If the problem is balanced, then $v_0 = 0$ and the DTLP and TLP problems coincide ■

In what follows, we look for the average of d^* over a set of solutions (e.g., over an infinite number of days) when conditions vary. It is assumed that points are embedded in a K-dimensional normed linear space where each point i is identified by a set of Cartesian coordinates, $\boldsymbol{x_i}$, and that distances, d_{ij}, are given by the norm of the Cartesian separation between points, $\| \boldsymbol{x}_i - \boldsymbol{x}_j \|$.

When conditions vary randomly, the notation $\langle Y \rangle$ and $\rangle Y \langle$ will be used to denote, respectively, the mean and variance of a random variable Y across the ensemble of possibilities; e.g., across all days. Different versions of the random TLP arise depending on which data are allowed to vary. In the simplest version of the problem the net supplies vary but the points are fixed on a K-dimensional square lattice (grid) with Cartesian spacing, l. Hence, N is fixed. If the v_i are independent, identically distributed (i.i.d.) random variables with mean 0 and variance σ^2, the problem instances will be generally unbalanced. However, if the v_i have zero means, variances σ^2 and covariances $\langle v_i v_j \rangle = -\sigma^2/(N-1)$, the problem is balanced.[2] Unless otherwise noted, it will also be assumed that the $\{v_i\}$ are multinormal. The modifiers "G" for "grid", "U" for "unbalanced" and "B" for "balanced" will sometimes be used as shorthand to specify the characteristics of a particular problem; e.g., TLP(U,G) and TLP(B,G) will designate unbalanced and balanced versions of the TLP where points are on a grid.

2.2. Dimensional analysis

Dimensional analysis can greatly simplify the solution task for any version of the TLP that can be completely specified in terms of just three constants. For example, given a norm, the constants are σ, l and N for DTLP(U,G). For other, more general versions of the problem (i.e, variations in demand that are not normal, points that are not fixed in number or location, or general service region shapes) these constants could be: σ, δ, and A. In general we look for $\langle d^* \rangle$, or alternatively for the average distance per point, defined as $\langle p^* \rangle = \langle d^* \rangle / N = \langle d^* \rangle / \delta A$.

Consideration shows that only two independent dimensionless parameters can be formed with either set of constants and the solution value, $\langle p^* \rangle$. The two parameters are N and $\frac{\langle p^* \rangle}{\sigma l}$ if the constants are σ, l and N; and δA and $\langle p^* \rangle \frac{\delta^{1/k}}{\sigma}$ if the constants are σ, δ and A. It therefore follows that the exact solution for $\langle p^* \rangle$ must be of the form:

$$\langle p^* \rangle = \sigma l f(N) \qquad \text{(in the first case)} \tag{3a}$$

and

$$\langle p^* \rangle = \sigma \delta^{-1/K} f(\delta A) \qquad \text{(in the second case),} \tag{3b}$$

where f is the only unknown left to be determined. This function will generally depend on the type of problem, the norm and the dimensionality of the space. The subscript "D" will be used with f when it refers to a DTLP. In a system of units where σ = 1 and l = 1, $\langle p^* \rangle = f(N)$. Thus, f has the interpretation of a "dimensionless distance per point".

3. EXACT RESULTS FOR THE 1-DIMENSIONAL CASE IN R^1

The balanced 1-D problem[3] with distance function $d_{ij} = |x_i - x_j|$ is simple. If one plots a curve of cumulative supply vs. the x-coordinate, $v(x) = \sum_{x_i \le x} v_i$, as shown in Fig. 1a, then d^* is the absolute area between *v(x)* and the x-axis. Figure 1a depicts a problem with evenly spaced points, (although this is not required); the following is true.

---

[2] This is true because, with these covariances, both the mean and variance of Σv_i are zero.

Result 1. (Deterministic and balanced TLP).

$$d^* = \int_{-\infty}^{+\infty} |v(x)| dx = \sum_{i=1}^{N-1} |v(x_i)| |x_{i+1} - x_i| \quad \blacksquare \tag{4a}$$

<u>Proof:</u> For any point, x_p, such as the one in Fig. 1a, $v(x_p)$ is the net flow across x_p because the aggregate supply and demand on both sides of x_p must be satisfied. Thus, $|v(x_p)|dx$ is a lower bound to the distance traveled in any small interval, (x_p, $x_p + dx$) where $v(x)$ is constant. Clearly, the sum on the right side of (4a) is a lower bound for d^*.

Conversely, a feasible solution can be constructed by considering horizontal slices of *dv* items (as shown on the figure) and transporting these quantities from the points where the slice intersects a rising portion of curve $v(x)$ to the adjoining points where it intersects a falling portion. In the case of the figure, *dv* items would be carried from A to B and from C to D. Thus, the summation of all the slices for small dv (still given by (4a)) is the distance of a feasible solution and an upper bound to d^*. ■

If points are evenly spaced, l distance units apart, (4a) reduces to

$$d^* = l \sum_{i=1}^{N-1} |v(x_i)| \,. \tag{4b}$$

If the net supplies are multinormal with $\langle v_i \rangle = 0$, $\rangle v_i \langle = \sigma^2$ and $\langle v_i, v_i \rangle = -\sigma^2 /(N-1)$, as is required for a homogeneous balanced problem, then the sum of the first i net supplies, $v(x_i)$, is a zero-mean normal variable with variance

$$\rangle v(x_i) \langle \;=\; i\acute{o}^{2} \left[1 - \frac{i-1}{N-1} \right]. \tag{5}$$

---

[3] 1-D problems arise in connection with highway construction projects and the minimization of earthwork "haul".

Recall too that if X is a normal random variable with zero mean, then

$$\langle |X| \rangle = [2\langle X \rangle / \pi]^{1/2} \quad . \tag{6}$$

Equations (5 and 8), applied to the expectation of (4b), yield the following result.

Result 2. (Random demand and balanced TLP). For the 1-D, homogeneous, zero-mean TLP(B,G) with normal demand,

$$\langle d^* \rangle = \left(2/\pi\right)^{1/2} \sigma l \sum_{i=1}^{N-1} \left(i \left[1 - \frac{i-1}{N-1} \right] \right)^{1/2} . \tag{7a}$$

The limit of this expression for $N \to \infty$ is[4]

$$\langle d^* \rangle \to \left(\frac{\pi}{32} \right)^{\frac{1}{2}} \sigma l N^{\frac{3}{2}} \quad . \ \blacksquare \tag{7b}$$

Note from (7b) that, for a given density of points $1/l$, the average distance traveled per point for the (1-D) TLP(B), $\langle p^* \rangle = \langle d^* \rangle / N$, satisfies:

$$\langle p^* \rangle \to \sqrt{\frac{\pi}{32}} \sigma l \sqrt{N} \tag{8}$$

This function increases without limit with the number of points, unlike in the TSP and the VRP, where there is a limit. The $\sqrt{N}$ dependence is caused by the long-range interactions arising from the flow balancing requirements. Equations (7b and 8) are quite general. They hold if the v_i are not normal, but satisfy the conditions of the central limit theorem, and also if the point locations

[4] Equation (7b) is true because its right side tends to

$$\left(2/\pi\right)^{1/2} \sigma l \int_0^N \left[x \left(1 - \frac{x}{N} \right) \right]^{\frac{1}{2}} dx = \left(2/\pi\right)^{1/2} \sigma l N^{\frac{3}{2}} \int_0^1 \left[\left(\frac{x}{N} \right) \left(1 - \frac{x}{N} \right) \right]^{\frac{1}{2}} d\left(\frac{x}{N} \right) = \left(\frac{\pi}{32} \right)^{\frac{1}{2}} \sigma l N^{\frac{3}{2}} .$$

vary across days as a homogeneous Poisson process, since (5) and (6) continue to hold under these conditions if $N \to \infty$.[5]

An expression for the unbalanced TLP is more difficult to obtain, but the task is easy for the DTLP. First define the cumulative demand for the depot *v'(x)* as shown in Fig. 1b; i.e.,

$$v'(x) = -v_0 H(x - x_0)\,, \tag{9}$$

where H is the Heaviside unit step function and x_0 is the depot location. Then, the same arguments used with result 1 establish that the absolute area between curves v and v' is d^*_D:

Result 3. (Deterministic DTLP).

$$d^*_D = \int_{-\infty}^{+\infty} \left| v(x) - v'(x) \right| dx\ . \ \blacksquare \tag{10}$$

If the depot is centrally located, then similar manipulations to those leading to Eq. (7b) now yield:

Result 4. (Random demand DTLP).

$$\left\langle d^*_D \right\rangle \to \sqrt{\frac{4}{9\pi}}\sigma\, l N^{\frac{3}{2}} \qquad \text{and} \qquad \left\langle p^*_D \right\rangle \to \sqrt{\frac{4}{9\pi}}\sigma\, l N^{\frac{1}{2}}\ . \ \blacksquare \tag{11}$$

---

[5] Clearly, (6) holds asymptotically because *v(x)* is normal for $N \to \infty$ even if the v_i aren't. To see that the asymptotic expression for $\rangle v(x) \langle$ remains the same note that if *i(x)* is the number of points in [0, x], then the conditional random variable $(v(x)|i(x))$ has zero mean, and variance $\rangle v(x) \,|\, i(x) \langle = i(x)\sigma^2 \left(1 - \frac{(i(x)-1)}{(N-1)}\right)$. Therefore, the unconditional variance is the expectation of this expression; i.e.,

$$\rangle v(x) \langle \ = \sigma^2 \left[\frac{\langle i(x) \rangle N}{N-1} - \frac{\langle i(x)^2 \rangle}{N-1} \right] = \sigma^2 \left[\frac{N^2 x}{L(N-1)} - \frac{1}{N-1} \left[\left(\frac{Nx}{L} \right)^2 + N \left(\frac{x}{L} \right) \left(1 - \frac{x}{L} \right) \right] \right],$$

which has the same limit, $\sigma^2 N \left[\frac{x}{L} \left(1 - \frac{x}{L} \right) \right]$, as Eq. (5).

Note that $\left\langle d_D^* \right\rangle > \left\langle d^* \right\rangle$, as one might expect. In the 1-D case $f(N) \sim \sqrt{N}$; the following sections show that $f(N) = O(log(N))$ in 2-D, and $f(N) = O(N^0)$ in higher dimensions.

The following sections show that $f(N) = O(log(N))$ in 2-D, and $f(N) = O(N^0)$ in higher dimensions.

4. UPPER BOUNDS

This section develops upper bounds for the distance traveled in several versions of the TLP and DTLP. The bounds are based on a bilevel algorithm for the DTLP that is described below. It is assumed that the service region has been partitioned into a finite number of subregions, C_I with their own subdepots.

Bilevel algorithm. Step 1 (lower level): Solve a DLTP for each C_I, using its subdepot, and route the items accordingly. Step 2 (upper level): Using the main depot, route the regional over- or under-supplies of step 1, $v_I = \sum_{i \in C_I} v_i$, from/to each subdepot as per an optimum DTLP.■ [6]

Since a DTLP is solved in step 2, the net flow of every subdepot is zero; i.e., items just pass through these points. It is therefore easy to express the result of the algorithm in path form, by specifying the number of items v_{ijk} that share the k^{th} path from origin i to destination j, including any intermediate subdepots. Since each point is part of a DTLP in step 1, the flows in and out of all points satisfy the conservation equations of the original DTLP. Therefore, the sums of the path

[6] The subscripts i and j are reserved for the original points, including the main depot in the case of the DTLP, but excluding all the subdepots. Capital letters, I, J are used for subdepots.

flows for every origin-destination pair, including the main depot, $\boldsymbol{v} = \left\{ v_{ij} = \sum_k v_{ijk}, \forall i, j \right\}$, are a feasible solution of the original problem with distance $d_D(\boldsymbol{v}) = \sum_{ij} v_{ij} d_{ij}$. If we let the combined distance of both steps of the bilevel algorithm be denoted $d^{(b)}$, it is now easy to see that the following is true:

Proposition 2 (Bilevel upper bound to DTLP). $d^{(b)} \geq d_D^*$. ■

<u>Proof:</u> From above, we see that $d_D(\boldsymbol{v}) \geq d_D^*$. Since the triangle inequality ensures that d_{ij} is a lower bound to the length of every path from i to j, d_{ijk}, we have:

$$d^{(b)} = \sum_{ijk} v_{ijk} d_{ijk} \geq \sum_{ijk} v_{ijk} d_{ij} = \sum_{ij} v_{ij} d_{ij} = d_D(\boldsymbol{v}) \geq d_D^*. \blacksquare$$

<u>4.1 Homogeneous, unbalanced problems with independent normal demands on a K-D lattice</u>

Consider now a K-dimensional cubic lattice of N points with Cartesian spacing l. Let $\bar{l}$ and $\underline{l}$ denote, respectively, the largest and smallest of the K distances between a point and its nearest neighbors in the direction of each axis. (For the Euclidean norm, $\bar{l} = \underline{l} = l$.) Assume too that $N^{1/K}$ is an integer, the points form a cube with the depot at its center, and that the net supplies are zero-mean, independent normal random variables with variance σ^2.

Define now two positive integers n and m such that $n^K m^K = N$, and a partition of the cube into $I = 1, 2, \ldots m^K$ identical cubes, C_I, with n points to a side and centrally located subdepots; see Fig.2. Since $N = n^K m^K$, we see from (3a) that the optimum total expected cost is $\langle d_D^* \rangle = m^K n^K \sigma l f_D(n^K m^K)$. A similar expression is now developed for the bilevel cost, $\langle d^{(b)} \rangle$.

Scalability of the bilevel algorithm: Because the normal distribution is infinitely divisible the two subproblems of the bilevel algorithm have normal demand. Furthermore, since both subproblems pertain to cubic lattices with centrally located subdepots and the same norm, they are DTLP's of the same type as the original problem. Therefore, they obey (3a) with the same f_D as the original, but different data, and this allows the expected total cost to be expressed as follows:

$$\langle d^{(b)} \rangle = m^K n^K \sigma l f_D\left(n^K\right) + m^K \left(\sigma n^{K/2}\right)(nl) f_D\left(m^K\right), \qquad (12)$$

where the first term is the aggregate cost of the m^k lower-level problems and the second term the cost of the high-level problem. The two parenthetical factors of the second term are the standard deviation and the lattice spacing of the upper level problem.

Proposition 2 implies that $\langle d^{(b)} \rangle \geq \langle d_D^* \rangle$. Thus,

$$f_D\left(n^K\right) + n^{\left(1-K/2\right)} f_D\left(m^K\right) \geq f_D\left(m^K n^K\right); \qquad m, n = 1, 2, 3 \ldots \qquad (13)$$

It is now possible to establish the following.

Theorem 1 (Upper bound for DTLP). If there is an $N_0 \geq 1$ such that f_D is monotone for $N > N_0$, then $f_D(N) = O\left(\sqrt{N}\right)$ in 1-D, $O(log(N))$ in 2-D and $O(N^0)$ in 3^+-D. ■

<u>Proof.</u> Consider the following subset of (13), corresponding to $m = 2$ and $n = 2, 4, 8 \ldots$

$$f_D\left(n^K\right) + n^{\left(1-K/2\right)} f_D\left(2^K\right) \geq f_D\left((2n)^K\right), \quad n = 2^j \; (j = 1, 2, \ldots), \qquad (14a)$$

and the related set of equalities,

$$f_D\left(n^K\right) + n^{\left(1-K/2\right)} f_D\left(2^K\right) = f_D\left((2n)^K\right), \quad n = 2^j \; (j = 1, 2, \ldots). \qquad (14b)$$

We look for the highest possible function with domain D = $\left\{n^K : n = 2^j, j = 1, 2, \ldots\right\}$ that satisfies (14a) and matches f_D when $n^K = 2^K$. Since (14a and 14b) have a recursive structure such a function, $\tilde{f}_D$, can be constructed by iterating (14b) starting with the given initial value for $n^K = 2^K$. The result, given below, is an upper bound for f_D in D.

$$\tilde{f}_D\left(n^K\right) = f_D\left(2^K\right)\left[\frac{n^{(1-K/2)}-1}{2^{(1-K/2)}-1}\right] \qquad \text{if } K \neq 2, \qquad (15a)$$

$$= f_D\left(2^K\right)\log_2(n) \qquad \text{if } K = 2. \qquad (15b)$$

Upon changing N for n^K in (15) it becomes apparent that $\tilde{f}_D(N)$ is $O\left(\sqrt{N}\right)$ if K=1, $O(log(N))$ if K = 2, and $O(N^0)$ if $K \geq 3$. Since f_D is bounded by $\tilde{f}_D$ in D, f_D satisfies the conditions of the theorem in D. The monotonicity of f_D guarantees that these conditions are also satisfied over the set of natural numbers. ■[7]

In view of Proposition 1, it is easy to see that the bounds also apply to the TLP(U).

Corollary 1 (Upper bounds for TLP(U)). Under the conditions of Theorem 1 stated at the outset of this section, $f(N)$ is $O\left(\sqrt{N}\right)$ in 1-D, $O(log(N))$ in 2-D, and $O(N^0)$ in 3^+-D. ■

4.2 Balanced problems, random point locations and other extensions

4.2.1 Balanced problems

Consider now the balanced TLP (still on a grid), where $\langle v_i v_j \rangle = -\sigma^2/(N-1)$. Recall that the associated DTLP, also balanced, is now identical to the TLP. The bilevel algorithm is not scalable now because the upper level problem is a DTLP(B), like the original problem, but the lower level subproblems exhibit correlations without being balanced. To avoid this difficulty, bounds will be used for the lower level problem instead of the exact result. These bounds are tightest when n = 2; therefore, we will set n = 2 in (12).

The cost of a feasible solution of a subproblem can be bounded by assigning to every pick-up and delivery the distance from the centroid. Since this distance is bounded by $K(\bar{l}/2)$, and since the expected number of pick-ups and deliveries in a subregion is $n^K\langle|v_i|\rangle = n^K\sqrt{\frac{2}{\pi}}\sigma$, the total lower level distance cannot exceed: $n^K m^K\sqrt{\frac{2}{\pi}}\sigma\,\bar{l}\,K\!\big/_2$. This quantity will replace the first term on the right side of (12).

The variance of the upper level net supplies is

$$\langle v_I^2\rangle = \left\langle \sum_{i,j\in C_I} v_i v_j \right\rangle = \sum_{i,j\in C_I}\langle v_i v_j\rangle \le \sum_{i\in C_I}\langle v_i^2\rangle = \sigma^2 2^K,$$

where the inequality results from neglecting the (negative) covariance terms. Thus, $\sigma 2^{K/2}$ is an upper bound for the standard deviation of v_I, and the second term of (12), which uses $\sigma 2^{K/2}$ as the standard deviation, is now an upper bound for the total high-level distance. The subscript “B” is used to designate the optimal solution of the balanced DTLP.

Hence, instead of (12) we have:

$$\langle d_B^{(b)}\rangle \le 2^K m^K \sigma\,\bar{l}\left(\frac{K}{\sqrt{2\pi}}\right) + 2^K m^K \sigma\, l 2^{1-K/2} f_B(m^K), \qquad \textit{for } m = 1, 2, \ldots \tag{16}$$

and since $\langle d_B^{(b)}\rangle \ge \langle d_B^*\rangle = (2m)^K \sigma\, l f_B((2m)^K)$, it follows that:

$$\frac{(\bar{l}/l)K}{\sqrt{2\pi}} + 2^{1-K/2} f_B(m^K) \ge f_B((2m)^K), \qquad \textit{for } m = 1, 2, \ldots \tag{17}$$

The arguments of Theorem 1 can now be repeated to obtain a recursive relation for the supremum of f_B, with essentially the same result. Therefore, we state without proof the following.

---

[7] The monotonicity assumption is needed because (13) can be satisfied with arbitrarily large values of $f_D(N)$

Theorem 2 (Upper bound for balanced problems). If there is an N_0 such that f_B is monotone for $N > N_0 \geq 1$, then $f_B(N) = O(\sqrt{N})$ in 1-D, $O(log(N))$ in 2-D and $O(N^0)$ in 3^+-D. ■

Since the TLP(B) and its associated DTLP are equivalent, Theorem 2 applies to both.

<u>4.2.2 Random point locations</u>

In this subsection points are randomly distributed in a cube as a homogeneous Poisson process with density δ so that (3b) holds for the DTLP. The subscript "R" will be used where appropriate to stress the changed nature of the problem.

If the cube is partitioned into m^K subcubes, with their centroids arranged in a cubic lattice (as before), and the bilevel algorithm is applied, then the expected total cost for the lower level problem is an aggregation of the costs of m^K scaled-down, random DTLP problems. Equation (3b) applies to each one of these subproblems if A is replaced by A/m^k. Therefore, the expected total lower level cost is

$$\left\langle d_R^{(b)} \right\rangle_L = \delta A \sigma \delta^{-1/K} f_{DR}\left(\delta \frac{A}{m^K} \right), \tag{18a}$$

since the expected number of total points at the lower level is δA.

The high level problem is a fixed location (grid) problem with m^K points and lattice spacing $A^{1/K}/m$. The variance of the net supply at each point is that of the excess demand in one cube, which is:[8] $\delta A \sigma^2 / m^K$. Thus, (3a) now yields:

---

whenever $N^{1/K}$ is a prime number; e.g. let $f_D(N) = M \to \infty$ if $N^{1/K}$ is prime, and $f_D(N) = 0$, otherwise

[8] The conditional mean and variance of excess demand in one cube with P points are 0 and $P\sigma^2$ respectively. Therefore, the unconditional variance is the mean of the conditional variance; i.e., $\delta A \sigma^2 / m^K$.

$$\left\langle d_R^{(b)} \right\rangle_H = m^K \left(\delta A \sigma^2 / m^K\right)^{1/2} \left(\frac{A^{1/K}}{m} \right) f_D\left(m^K\right). \tag{18b}$$

Note as well that the original DTLP problem satisfies:

$$\left\langle d_{DR}^* \right\rangle = (\delta A) \sigma \delta^{-1/K} f_{DR}(\delta A). \tag{18c}$$

Since Proposition 2 holds for each instance of the problem, it also holds for the expectation. Thus, $\left\langle d_R^{(b)} \right\rangle_L + \left\langle d_R^{(b)} \right\rangle_H \geq \left\langle d_{DR}^* \right\rangle$. Substituting (18a– 18c) for these terms, and dividing both sides of the resulting inequality by $\delta A \sigma \delta^{-1/K}$, we find:

$$f_{DR}\left(\delta \frac{A}{m^K}\right) + \frac{m^{K/2 - 1}}{(\delta A)^{1/2 - 1/K}} f_D\left(m^K\right) \geq f_{DR}(\delta A), \quad \text{for } \delta A \geq 0,\ m = 2, 3, \ldots \tag{19}$$

This system of inequalities has the same structure as (13), as can be seen by the change of variable $\delta A = n^K m^K$ (where n is now real and non-negative). The result is:

$$f_{DR}\left(n^K\right) + n^{1-K/2} f_D\left(m^K\right) \geq f_{DR}\left(n^K m^K\right), \quad \text{for } n \geq 0,\ m = 1, 2, 3, \ldots \tag{20}$$

The proof of Theorem 1 can now be repeated step by step with the same conclusion. Thus, the following is true.

Theorem 3 (Upper bounds for random location problems). The function f_{DR} of the random DTLP(U) obeys Theorem 1. Furthermore, insofar as the DTLP(U) distance bounds from above the TLP(U) distance, the result also holds for the latter. ■

The results can be extended to different variants of the problem using similar logic; e.g. if the number of points is fixed but their location is random and also if the problem is balanced.

4.2.3 Non-normal demand and the assignment problem

The results can also be extended to the case of non-normal demand by letting $\delta A \to \infty$ and then decomposing the problem in two levels, each with many points; i.e., where *m* is such that $\left(m^K, \delta A / m^K\right) \to \infty$. Only the logic behind the formal arguments is outlined here. If the main cube is partitioned as before, then the lower level problems are scaled-down versions of the main problem. This is true for both lattice and random problems. Therefore, Eqs.(18a and 18c) continue to hold for both types of problems; i.e., we can write

$$\left\langle d^{(b)} \right\rangle_L = (\delta A)\sigma\delta^{-1/K} f_X\left(\frac{\delta A}{m^K}\right) \qquad \text{and} \qquad \left\langle d^* \right\rangle = \delta A \sigma \delta^{-1/K} f_X(\delta),$$

where the subscript "*X*" signifies non-normal demand.

Since the expected number of points in the lower level problem tends to ∞, the distribution of subcentroid supplies tends to the normal, and the high level problem should behave as in prior sections for large *n*. In other words, for any desired tolerance level there should be a value m_0 such that $\left\langle d^{(b)} \right\rangle_H$ is given by (18b) to within the prescribed tolerance for all $m \geq m_0$.

Thus, f_X satisfies the same functional relation as f_R , Eq. (18), albeit only for $m \geq m_0$ and within a tolerance. The arguments of Theorem 1 can then be modified slightly (using $m = m_0$ instead of $m = 2$ as the fixed value of *m*, and incorporating the effect of the tolerance) with the same final result. Therefore, f_X also obeys Theorem 1.

Note that in the special case where the net supplies are binary random variables, $v_i = +1$ or -1 with probability 1/2, the TLP becomes the "assignment" LP. Therefore, Theorems 1 and 2 also apply to the assignment problem.

5. ASYMPTOTIC FORMULAE

Here we develop asymptotic formulae and investigate the effects of zone shape in 1-, 2- and 3^+- D spaces.

5.1 A lower bound

A lower bound to $\langle p^* \rangle$ for balanced problems, including the DTLP(U), is the product of the expected distance from a nearest neighbor and the mean absolute value of the net supply from a point. In all the problems studied, this product is a multiple of $\delta^{-1/K}\sigma$. Thus,

$$\langle p^* \rangle \geq c\delta^{-1/K}\sigma , \qquad (21)$$

where c is a problem-specific constant. The same result holds asymptotically for the TLP(U), since the amount not shipped becomes negligible as $\delta \to \infty$.

We have also found that in $3^+ - \mathrm{D}$, $\langle p^* \rangle = \sigma\delta^{-1/K} f(\delta A)$ where $f(\delta A)$ is bounded above by some constant, *C*, for all δA exceeding a certain value which we denote N_1. Equation (21) indicates that $f(\delta A)$ is also bounded from below by a positive number, *c*, so that if *f* is monotone in the sense of Theorem 1, then for all the problems studied there is a problem-specific positive constant c_0 such that:

$$\lim_{\delta A \to \infty} \langle p^* \rangle = c_0 \delta^{-1/K}\sigma , \qquad \text{if } K \geq 3 . \qquad (22)$$

5.2 Asymptotic behavior of the bilevel algorithm

It is shown in this section that the bilevel algorithm is asymptotically optimal in all cases where comparisons can be made with known asymptotic solutions. Therefore, it may be conjectured that it is also asymptotically optimal in the remaining cases discussed in this paper. This conjecture is tested in Section 5.3.

Assume that the bilevel algorithm is applied to a DTLP(U) problem whose dimensionless distance per point *f* satisfies the monotonicity condition, so that either Theorem 1 or Theorem 3 holds, and let $g\left(n^K, m^K\right)$ be the dimensionless distance per point obtained with the algorithm.[9] (Note that g is just an abbreviation for the left-hand side of either (13) or (20).) Then, the following is true.

Lemma 1. For the DTLP(U) in $3^+ - D$, $\left(g\left(n^K, m^K\right) - f\left(n^K\right)\right) / f\left(n^K\right) \to 0$ as $n \to \infty$, uniformly for all values of m^k. ■

Proof: Equations (13) and (20) reveal that the ratio in question is $n^{1-K/2} f_D\left(m^K\right) / f\left(n^K\right)$. Since $f_D(m^K) < C$ for $m^K > N_1$, and $f(n^K) < c$ (see Sec. 5.1), the ratio is bounded by $n^{1-K/2} C / c$ if $m^K > N_1$. Since $1 - K/2 < 0$, this function tends to zero as $n \to \infty$, and the lemma is proven. ■

Theorem 4. In R^1 and $3^+ - D$, the relative expected error of the bilevel algorithm for DTLP(U) tends to zero as $m, n \to \infty$.■

Proof: It follows from the monotonicity property and from (13), (20) that:

$$0 \le f\left(n^K\right) \le f\left(n^K m^K\right) \le g\left(n^K, m^K\right) \qquad \text{if } n^K \ge N_0 . \tag{23}$$

[9] Recall that n^K is the average number of points per subzone ($n^K = \delta A / m^K$ in Sec. 3.2.2) if points are random.

Consider now the $3^+ - D$ case. The relative expected error, defined as $\left(\langle d^{(b)}\rangle - \langle d_D^*\rangle\right)/\langle d_D^*\rangle$, is $\left(g\left(n^K, m^K\right) - f\left(n^K m^K\right)\right)/ f\left(n^K m^K\right)$. Equation (23) guarantees that it satisfies

$$0 \le \left(g\left(n^K, m^K\right) - f\left(n^K m^K\right)\right)/ f\left(n^K m^K\right) \le \left(g\left(n^K, m^K\right) - f\left(n^K\right)\right)/ f\left(n^K\right).$$

Since the right side of the above expression tends to zero as $m, n \to \infty$ by virtue of Lemma 1, so does the relative expected error.

Consider now the 1-D case where the second inequality of (23) has the form $f(nm) \le f(n) + n^{1/2} f(m)$. Recall from (8) that $f(N) = (\pi N/32)^{1/2}$ in R^1. Thus, the inequality becomes $\left(\frac{\pi n}{32}\right)^{1/2} m^{1/2} \le \left(\frac{\pi n}{32}\right)^{1/2}\left[1 + m^{1/2}\right]$, and the relative error is $m^{-1/2}$, which also tends to 0 as $m, n \to \infty$. ■

5.3 Approximate formulae for 2-D problems

Theorem 4 suggests that the bilevel algorithm with $n = m$ is also asymptotically optimal in 2-D, and therefore that for large values of n and m, $f\left(n^2 m^2\right) \approx f\left(n^2\right) + f\left(m^2\right)$. If this is true, f cannot be bounded from above. Instead, it would increase logarithmically. Therefore, we propose the following.

Conjecture. If N is large, an approximation for the dimensionless distance per point in 2-D is:

$$\langle p^*\rangle \delta^{1/2}/\sigma = f(N) \approx K_1 + K_2 log(N), \tag{24}$$

where K_1 and K_2 are dependent on the version of the problem. ■

To test this conjecture, a battery of Euclidean TLP(B) problems with random point locations and fixed N were solved. The data for these problems are described in the appendix.

Figure 3 displays the results on a diagram of $\langle p^* \rangle \delta^{1/2} / \sigma$ vs. *log*(*N*). The results speak for themselves. The equation of the line is:

$$f(N) = 0.42 + 0.031\, log_2(N)\,, \qquad \text{for } N \in [25{,}5000] \tag{25}$$

The deviations from the line are consistent with the standard errors estimated from the simulation.

Since the average number of items supplied per point is $\langle (v_i)^+ \rangle = \sigma / \sqrt{2\pi}$ in the case of normal demands, we see that the average distance traveled per item in the Euclidean case is estimated to be $f(N)\sqrt{2\pi}\delta^{-1/2}$. That is,

$$< distance \;\; per \;\; item > \;\approx \delta^{-1/2}(1 + 0.078\, log_2(N))\,.$$

As a point of reference, this distance is about twice as long as for the Euclidean TSP, for values of *N* one is likely to encounter in actual logistics problems.

<u>5.4 Size and shape effects</u>

This subsection explores the effect of zone size and shape. It shows that if f_S is the dimensionless distance per point function in a region of a specific size and shape, then $\lim_{N \to \infty} f_S(N) / f(N) = 1$ in 2^+-D, but not in 1-D. Therefore, the 2-D case can be viewed as a transition case that shares some of the properties of 1-D (unbounded *f*(*N*)) and some of the properties of 3-D (shape-independence).

The analysis is based on the conjecture that the DTLP and TLP, with either regular or random point locations, satisfy the following:

Strong monotonicity conjecture. If S is a region formed by a non-overlapping assembly of Q volume-A cubes, it is conjectured that there is a number N_{crit} for which the dimensionless distance per point for the region, $f_S(\delta AQ)$, satisfies:

$$f_S(\delta AQ) \geq f(\delta A), \qquad \forall \ \delta A \geq N_{crit} \ \blacksquare \tag{26}$$

This conjecture states that the dimensionless distance per point is larger in a region than in any component cube, if one holds the density of points constant. When the arrangement is itself a cube, as in Figs. 4a and 4b, (26) simply restates the monotonicity of f. The conjecture is also true for arrangements, such as those of Figs. 4c and 4d, whose points can be put in a 1:1 correspondence with those of a cube of identical volume by means of a mapping that preserves volume without increasing the distance between any pair of points.[10]

The following lemma relates the normalized distance functions for the DTLP, f_D and f_{DS}, and the diameter of the region, $\phi = sup(\|\boldsymbol{x_1} - \boldsymbol{x_2}\|)$.

Lemma 2. For some positive constant c, the expression $f_D(\delta A) + \phi c \delta^{1/K} (\delta A)^{-1/2}$ is an upper bound to the normalized distance function of the DTLP, $f_{DS}(\delta AQ)$.$\blacksquare$

<u>Proof.</u> If the bilevel algorithm is applied to the assembly with individual cubes as subregions, Proposition 2 ensures that $\langle d^{(b)} \rangle \geq \langle d_D^* \rangle$. The lower level cost of $\langle d^{(b)} \rangle_L$ is $Q\sigma\delta^{-1/K}(\delta A) f_D(\delta A)$. (The function f_D depends on the statistical distribution of the net supply). The upper level cost is

[10] To see this note that for regions with this weak contraction property, the optimum solution of a TLP or DTLP has a mapped equivalent in the cubic arrangement with lower or equal cost. Thus, $f_S(\delta AQ) \geq f(\delta AQ)$. Eq. (26) follows per the monotonicity of f.

bounded above by the product of the diameter of the region, ϕ, and the expected number of items shipped from the subcentroids, which is $Qc\left(\sigma^2 \delta A\right)^{1/2}$ for some constant *c*. (This constant would depend of the particulars of the problem; e.g., on the net supply distribution.) Therefore,

$$Q\sigma\delta^{-1/K}(\delta A) f_D(\delta A) + \phi Q c\sigma(\delta A)^{1/2} \geq \left\langle d^{(b)} \right\rangle \geq \left\langle d_D^* \right\rangle = \sigma\delta^{-1/K}(\delta AQ) f_{DS}(\delta AQ),$$

which implies

$$f_D(\delta A) + \phi c\delta^{1/K}(\delta A)^{-1/2} \geq f_{DS}(\delta AQ). \blacksquare \tag{27}$$

It is now possible to prove the following theorem.

Theorem 5 (Asymptotic shape independence). If (26) holds then $f_{DS}(\delta AQ)/ f_D(\delta A) \to 1$ as $\delta \to \infty$ (with *A* and *Q* constant) in $2^+ - D$, but not in R^1. ■

<u>Proof.</u> Consider the $2^+ - D$ case first. Then Eqs. (26) and (27) imply that

$$1 \leq \frac{f_{DS}(\delta AQ)}{f_D(\delta A)} \leq 1 + c_1 \frac{(\delta A)^{\frac{1}{K} - \frac{1}{2}}}{f_D(\delta A)}, \qquad \text{for } \delta A \geq N_{crit}, \tag{28}$$

where $c_1 = cA^{-1/K}\phi$.

Since $f_D(\delta A)$ increases without bound and $(\delta A)^{\frac{1}{K} - \frac{1}{2}}$ is bounded from above for $K \geq 2$, we see that the last member of (28) tends to 1 as $\delta \to \infty$. Thus, the theorem is proven for 2^+-D.

To see that the shape independence result does not hold in R^1 consider a DTLP for two segments of length *A* that are separated end to end by a distance *L* with the depot in the middle; see Fig.4d. Consideration of the logic behind Fig.1b and Eq.(10) shows that the separation of the two segments does not change the optimum v_{ij} but increases the total distance traveled by an amount *L*/2 for each item shipped out of a zone and a like amount for each item received extra-zonally.

(Items transshipped from one zone to the other pay L extra distance units, but items shipped from the depot only pay $L/2$). Thus, the expected distance is $\langle|v(A)|\rangle L$ and we have:

$$\langle d_S^* \rangle = \langle d_D^* \rangle + L\langle|v(A)|\rangle \to \langle d_D^* \rangle + L(2/\pi)^{1/2}\sigma(\delta A)^{1/2}.$$

The last expression follows from the central limit theorem and the formula for $\langle|v(A)|\rangle$ in the case of normal independent demand; see (6).

Then, since $\langle d_D^* \rangle = (4/9\pi)^{1/2}\sigma\delta^{-1}(\delta A)^{3/2} = (4/9\pi)^{1/2}\sigma A(\delta A)^{1/2}$ (see Eq. (7)), we find:

$$\left(\langle d_S^* \rangle - \langle d_D^* \rangle\right) \Big/ \langle d_S^* \rangle = 1 + L\left(\frac{3/\sqrt{2}}{A}\right) \neq 1. \blacksquare$$

Theorem 5 implies size and shape independence in 2^+-D (for a given density of points) because the limiting ratio is 1 for any Q and S. Note as well that, by letting $Q \to \infty$, regions with smooth boundaries can be considered. Therefore, the result is quite general. Similar results can be developed for the TLP.

A second order approximation to f_S in Euclidean K-space is proposed below, based on the example of Fig. 4c. If $K \geq 2$ and $L >> A^{1/K}$ then at optimality the items sent across cubes should be minimized, as occurred in the 1-D case. Therefore, the solution in each cube continues to be independent of L, and we can write:

$$\langle d_S^* \rangle = \text{(term independent of } L\text{)} + Lc_2\sigma(\delta A)^{1/2},$$

where $c_2 \to \sqrt{2/\pi}$ for the DTLP. It can also be shown that c_2 takes the same value for the TLP(B). Note that in 2-D L is proportional to the semi-perimeter of the region. Therefore, one may speculate that if K = 2, then a second order approximation for zones of irregular shape may be $\left\langle d_S^* \right\rangle - \left\langle d^* \right\rangle \approx \; \Delta c_2 \sigma (\delta A)^{1/2}$, where Δ is the change in semi-perimeter; i.e.,

$$\left\langle d_S^* \right\rangle \approx \sigma \delta^{-1/2} (\delta A) \left\{ f(\delta A) + c_2 \left(\frac{\Delta}{A^{1/2}} \right) \right\}.$$

Thus, for the random location TLP(B) with normal demands and Euclidean distances, the following is proposed:

$$f_S(\delta A) \approx 0.42 + \sqrt{\frac{2}{\pi}} \left(\frac{\Delta}{A^{1/2}} \right) + 0.31 \, log_2 (\delta A), \qquad (29)$$

where (25) has been used instead of $f(\delta A)$. For a rectangle with an aspect ratio of 2, the correction term is approximately 0.1; i.e. about 15% of the original value for problems of the size one is likely to encounter in logistics applications.

6. INHOMOGENEOUS PROBLEMS

In actual problems point locations may be fixed but irregular, and the supply data may be inhomogeneous, i.e., one may know statistics such as $\left\langle v_i \right\rangle = m_i \neq 0$, $\rangle v_i \langle = \sigma_i^2$ and $\left\langle v_i v_j \right\rangle = m_i m_j + \sigma_{ij}^2$. Section 6.1 presents an approximation for situations where only a few parameters summarizing the general distribution of the m_i, σ_i^2 and σ_{ij}^2 are known, and Sec. 6.2, develops bounds for problems where the detailed statistics are available.

<u>6.1 Confidence interval approximations based on summary data</u>

Assume first that the problem has independent demands $\left(\sigma_{ij}^2 = 0 \right)$ and is homogeneous. Assume too that the data, $\times = \left\{ \times_i = \left(m_i, \sigma_i{}^2 \right); i = 1, 2, \ldots N \right\}$, can be visualized as the realization of a

K-D homogeneous compound Poisson process over the region A with independent batch sizes (m_i, σ_i^2). Thus, the net supplies become conditional random variables, $\{v_i \,|\times\}$. Only the moments of $\boldsymbol{X}$ are known, i.e., $\langle (m_i, \sigma_i^2) \rangle = (0, \bar{\sigma}^2)$, and $\rangle (m_i, \sigma_i^2) \langle \;=\; (S^2, Z)$.

This set up is quite flexible. We have not said anything for example about the distribution of $\{m_i\}$, which can be asymmetric and therefore capture problems with many more origins and destinations (or vice versa). Deterministic supply problems are also included, if one just sets $\bar{\sigma}^2 = Z = 0$. We desire a formula for $\langle d^* \,|\times \rangle$ where the variations come from the different net supplies $\{v_i \,|\times\}$ that would be observed on each realization. It will turn out that Z is not needed for the approximation.

Consider first the unconditional random variable, d^*, whose variations come from those of the unconditional v_i. These are obtained by first choosing $\boldsymbol{X}$ and then $\{v_i \,|\times\}$. The unconditional mean distance $\langle d^* \rangle$ is given by the random-location results of Sec. 4.2.2 because the unconditional v_i are independent random variables with zero means $\langle v_i \rangle = \langle m_i \rangle = 0$, and identical variances, σ^2. The latter are $\sigma^2 \;=\; \rangle v_i \langle \;=\; \rangle m_i \langle + \langle \sigma_i^2 \rangle = S^2 + \bar{\sigma}^2$.

Since $\langle d^* \rangle$ is the mean of $\langle d^* \,|\times \rangle$, a confidence interval for $\langle d^* \,|\times \rangle$ is $\langle d^* \rangle \pm \kappa \left(\rangle\, [\langle d^* | \boldsymbol{X} \rangle\,]\, \langle \right)^{1/2}$, where κ is an appropriate number of standard deviations (e.g., 2 or 3) for a desired confidence level. Although the variance of $\langle d^* \,|\times \rangle$ in the 2nd term of this expression is unknown, this variance is bounded by the variance of d^*. This is true because $\rangle d^* \langle \;=\; \langle\, [\rangle d^* | \boldsymbol{X} \langle\,]\, \rangle \;+\; \rangle\, [\langle d^* | \boldsymbol{X} \rangle\,]\, \langle \;\geq\; \rangle\, [\langle d^* | \boldsymbol{X} \rangle\,]\, \langle$. Thus, an approximation for $\langle d^* | \boldsymbol{X} \rangle$ is:

$$\langle d^*|\boldsymbol{X}\rangle \approx \langle d^*\rangle \pm \kappa\left(\rangle d^*\langle\right)^{1/2}.$$

Obviously, a similar expression holds for the normalized distance per point:

$$\langle p^*|\boldsymbol{X}\rangle \approx \langle p^*\rangle \pm \kappa\left(\rangle p^*\langle\right)^{1/2}.$$

It is reasonable to expect $\rangle p^*\langle$ to be bounded or perhaps even decline with problem size, δA, because larger problems include more data. This was confirmed by the simulations, where it was estimated that

$$\left(\rangle p^*\langle\right)^{1/2} \le 0.15\sigma\delta^{-1/2}; \tag{30}$$

see Fig.3b. Since $\langle p^*\rangle$ declines with δ more slowly than $\left(\rangle p^*\langle\right)^{1/2}$ (except in 1-D), the error committed by using aggregate information goes to zero as the problem size increases; i.e.,

$$\underset{\delta\to\infty}{plim}\ \langle p^*|\boldsymbol{X}\rangle \big/ \langle p^*\rangle = 1. \tag{31}$$

Similar results can be developed for balanced problems and problems where only a subset of $\boldsymbol{X}$ is fixed. Note in particular that the perfectly deterministic problem, where the v_i are fixed, obeys the same asymptotic formulae. In this case one would put $\langle\sigma^2\rangle = \overline{\sigma}^2 = 0$.

6.2 Bounds for the detailed problem

It is assumed here that $\boldsymbol{X}$ is known. The results are based on the following self-evident property of Eqs. (1).

Superposition property of TLP: If $\boldsymbol{u} = \{u_{ij}\}$ and $\boldsymbol{y} = \{y_{ij}\}$ are feasible solutions of two TLP's with $u = \{u_i\}$ and $y = \{y_i\}$ as data, then $\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{y}$ is a feasible solution of the TLP with $v = u + y$ as data.■

Since $\boldsymbol{X}$ is known, let us decompose the net supplies as follows, $v_i = m_i + u_i$, where m_i is fixed and known and u_i varies across realizations. Define too, the "perturbation" function, $d^*(v)$, which returns d^* as a function of the data, $v = \{v_{ij}\}$. This function is convex in our case, because our problems always involve the minimization of a linear function over a convex set. It is now possible to show that the following is true.

Theorem 6 (Bounds for random demand and detailed data).

$$d^*(m) \leq \langle d^* \rangle \leq d^*(m) + \langle d^*(u) \rangle = d^*(m) + \bar{\sigma}\delta^{-1/K}(\delta A) f_S(\delta A), \tag{32}$$

where $\bar{\sigma}^2$ is the average of $\rangle u_i \langle$ across all points.■

<u>Proof:</u> Since d^* is a convex function, Jensen's inequality guarantees that $d^*(\langle v \rangle) \leq \langle d^*(v) \rangle$; i.e., that $d^*(m) \leq \langle d^* \rangle$. Therefore, the lower bound holds.

For any realization, the superposition of the optimum (deterministic) solution with m as data (and distance $d^*(m)$), and the optimum (random) solution with u as data (and distance $d^*(u)$) is a feasible solution of the real problem with v as data. Therefore, $d^*(v) \leq d^*(m) + d^*(u)$ for every realization. Clearly, the inequality must also hold for the averages across realizations and hence, $\langle d^* \rangle \leq d^*(m) + \langle d^*(u) \rangle = d^*(m) + \bar{\sigma}\delta^{-1/K}(\delta A) f_S(\delta A)$, which proves the theorem. (The last equality is based on the results of Sec. 5.1, since the u_i are net supplies with zero means.) ■

Whenever the second term of (32) is small relative to the first term, deterministic approximations are reasonable. Otherwise, since the dependence of $\langle p^* \rangle$ on δ is very weak in 2-D (and non-existent in 3-D for $\delta \to \infty$) the exact cost of the problem may perhaps still be approximated by a formula where the dependence is ignored. One may then be able to use continuum approximations for inhomogeneous problems where the spatial data change slowly with location. These practical matters, however, are left to future work.

7. SCALABLE NETWORKS

Here, the supply points are nodes i with net supplies v_i in a mixed graph with edges e. Assume first that the graph is directed, and is characterized by input and output sets, $I(i)$ and $O(i)$, that identify the edge pointing in and out of each node i. It is assumed that points are identified by Cartesian coordinates and that the non-negative edge distances d_e are given by a norm. The collection of all this information without the net supplies, {edges, nodes, depot, coordinates, norm and p}, will be called a "network". For any given set of net supplies and edge distances, we solve for the edge flows $v_{(e)}$ for the following network problem (NP), where $p \geq 0$.

$$\text{(NP)} \qquad \min \quad z = \sum_e d_e (v_{(e)})^p \tag{33a}$$

$$\text{s. t.:} \quad \sum_{e \in O(i)} v_{(e)} - \sum_{e \in I(i)} v_{(e)} \leq v_i \ , \qquad \forall i \tag{33b}$$

$$v_{(e)} \geq 0 \qquad , \qquad \forall e . \tag{33c}$$

Note the similarity to (1). As in that case, one can define feasible, auxiliary (ANP) versions of the problem, and DNP versions where one of the nodes is designated as a "depot".

The results in prior sections for fixed-point locations extend to this version of problem NP if the network is "scalable"; i.e., a network belonging to a family whose members are fully characterized by a scale parameter l and a size parameter $N = m^K n^K$, with the following two properties. First, the graph can be partitioned into m^K identical subgraphs that define subnetworks in the family, with scale parameter l and size parameter n^K. Second, paths can be extracted from the original graph to connect the subdepots and form a graph that belongs to the original family, with scale parameter nl and size parameter m^K. This definition of scalability is more general than the definition presented in Section 4.1 because depots are no longer required to be centrally located; see the 1-D example in Figure 5. The networks in this figure belong to a family of equally spaced nodes, which are connected from the left to the nearest neighbor and from the right by the second nearest neighbor. Note that the subnetworks are tiles that can be joined to fill the space and make larger networks. This is also true in 2 and 3 dimensions.

For these types of problems, dimensional analysis yields the following general solution:

$$\langle d^* \rangle = m^K n^K l \sigma^p f_N(N), \tag{34}$$

where the subscript "N" indicates that the dimensionless distance per point pertains to a network problem with a specific structure. The bilevel algorithm for the DNP yields flows $\boldsymbol{v_L}$ and $\boldsymbol{v_H}$ at each level that are feasible solutions to (33). Scalability implies that the resulting average total costs can be expressed as follows:

$$\left\langle d^{(b)} \right\rangle_L = \left\langle z(\boldsymbol{v_L}) \right\rangle = m^K n^K l \sigma^p f_N\left(n^K\right), \tag{35a}$$

$$\left\langle d^{(b)} \right\rangle_H = \left\langle z(\boldsymbol{v_H}) \right\rangle = m^K \left(\sigma n^{K/2} \right)^p (nl) f_N\left(m^K\right). \tag{35b}$$

If $p \leq 1$, we can write $\left\langle z(\boldsymbol{v_L}) \right\rangle + \left\langle z(\boldsymbol{v_H}) \right\rangle \geq \left\langle z(\boldsymbol{v_L} + \boldsymbol{v_H}) \right\rangle$, and since $(\boldsymbol{v_L} + \boldsymbol{v_H})$ is a feasible solution of the original problem, $\left\langle z(\boldsymbol{v_L} + \boldsymbol{v_H}) \right\rangle \geq \left\langle z^* \right\rangle$. Thus, if $p \leq 1$, then the sum of (35) is an upper bound for (34), and the following inequality results:

$$f_N\left(n^K\right) + n^{[1-K(1-p/2)]} f_N\left(m^K\right) \geq f_N\left(m^K n^K\right); \qquad m, n = 1, 2, 3 \ldots \tag{36}$$

This is the same as (13) when NP is linear (p = 1). Thus, Theorem 1 applies to linear NPs. Consideration shows that Theorems 2 and 4 also apply to the linear case.

Recall from the proof of Theorem 1 that the solution was bounded from above when the coefficient of f in the second term of the left side of (13) declined with n. This is also true now. Thus, in the nonlinear case the solution is bounded if $1-K(1-p/2)<0$. Note in particular that if $p<0$ (diseconomies of scale), the 2-D problem is bounded.

ACKNOWLEDGEMENTS

Research supported in part by the University of California Transportation Center.

REFERENCES

Daganzo, C.F. (1984a) "The length of tours in zones of different shapes", Trans. Res. B 18B, 135-146.

Daganzo, C.F. (1984b) "The distance traveled to visit N points with a maximum of C stops per vehicle: An analytic model and an application", Trans. Sci. 18 (4), 331-350.

Daganzo, C.F. (1999) *Logistics Systems Analysis* (3rd edition), Springer, Heidelberg, Germany.

Eilon, S., Watson-Gandy, C.D.T. and Christofides, N. (1971) Distribution Management: Mathematical Modelling and Practical Analysis, Hafner, New York, N.Y.

Haimovich, M. A., A. H. G. Rinnooy-Kan and L. Stougie (1985) “Bounds and heuristics for capacitated routing problems” Math.. Oper. Res. 10, 527-542.

Karp, R.M. (1977) "Probabilistic analysis of partitioning algorithms for the traveling salesman problem", Math. Opns. Res. 2, 209-224.

Newell, G.F. (1973) "Scheduling, location, transportation and continuum mechanics: some simple approximations to optimization problems", SIAM J. Appl. Math. 25(3), 346-360.

Newell, G.F. (1986) "Design of multiple vehicle delivery tours--III: Valuable goods", Trans. Res. B 20B(5), 377-390.

Newell, G.F. and Daganzo, C.F. (1986) "Design of multiple vehicle delivery tours--I: A ring-radial network", Trans. Res. B 20B(5), 345-364.

Newell, G.F. and Daganzo, C.F. (1986a) "Design of multiple vehicle delivery tours--II: Other metrics", Trans. Res. B 20B(5), 365-376.

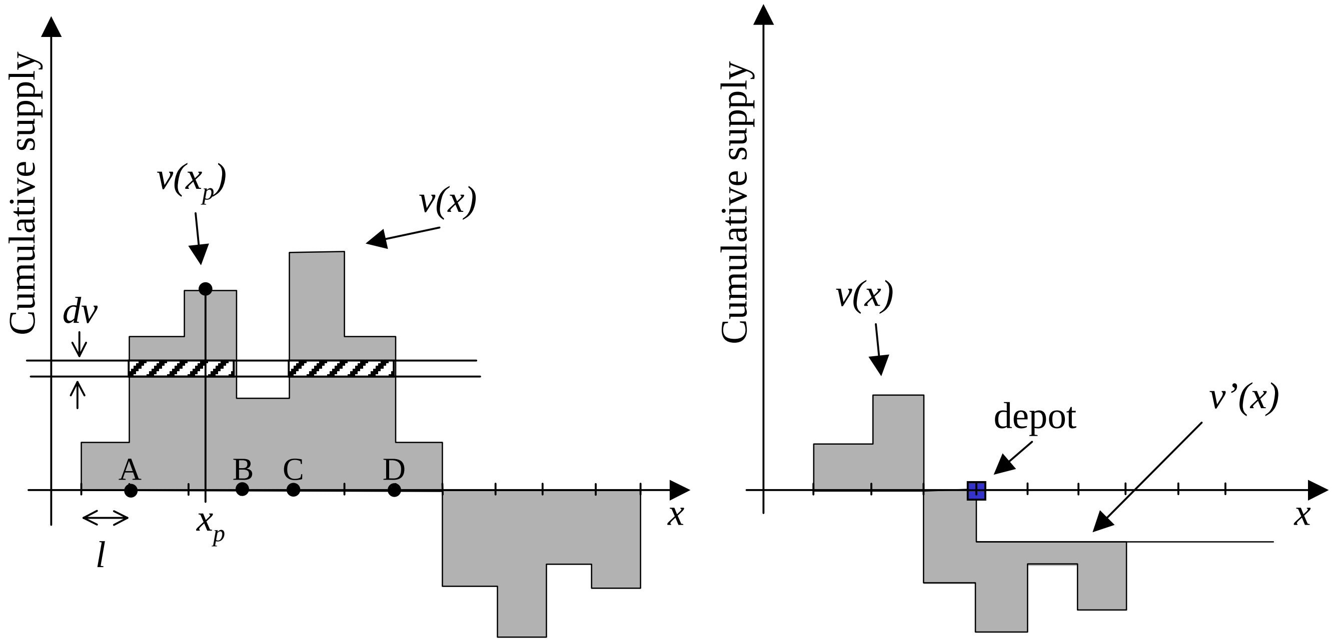

Figure 1. Graphical solutions of 1-D problems
(a) TLP(B); (b) DTLP(U)

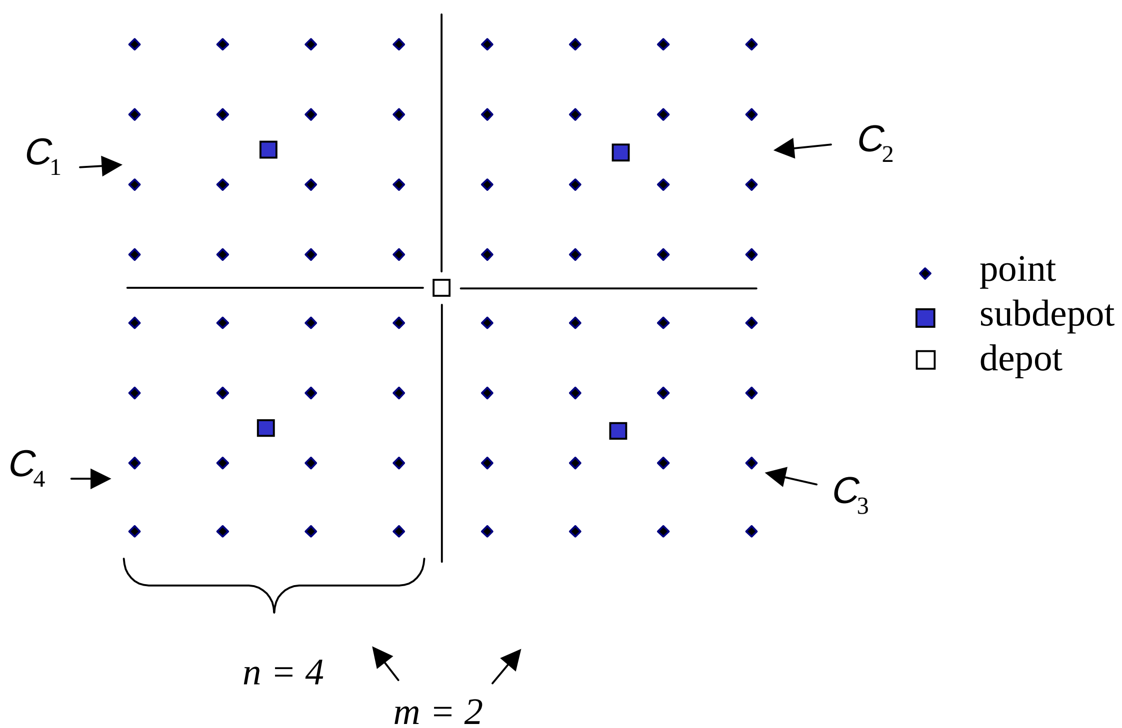

Figure 2. Partition of a 2-D lattice with N = 64 points

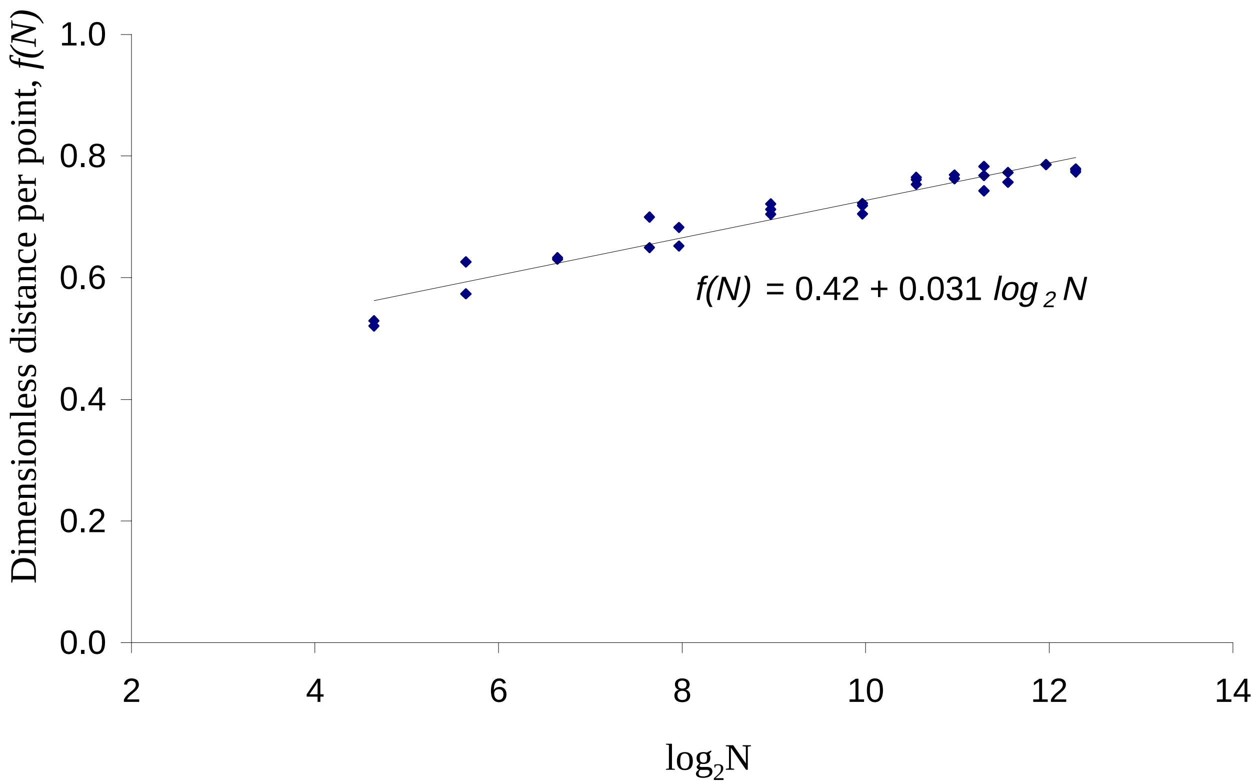

Figure3a. Dimensionless distance per point v. $\log_2 N$

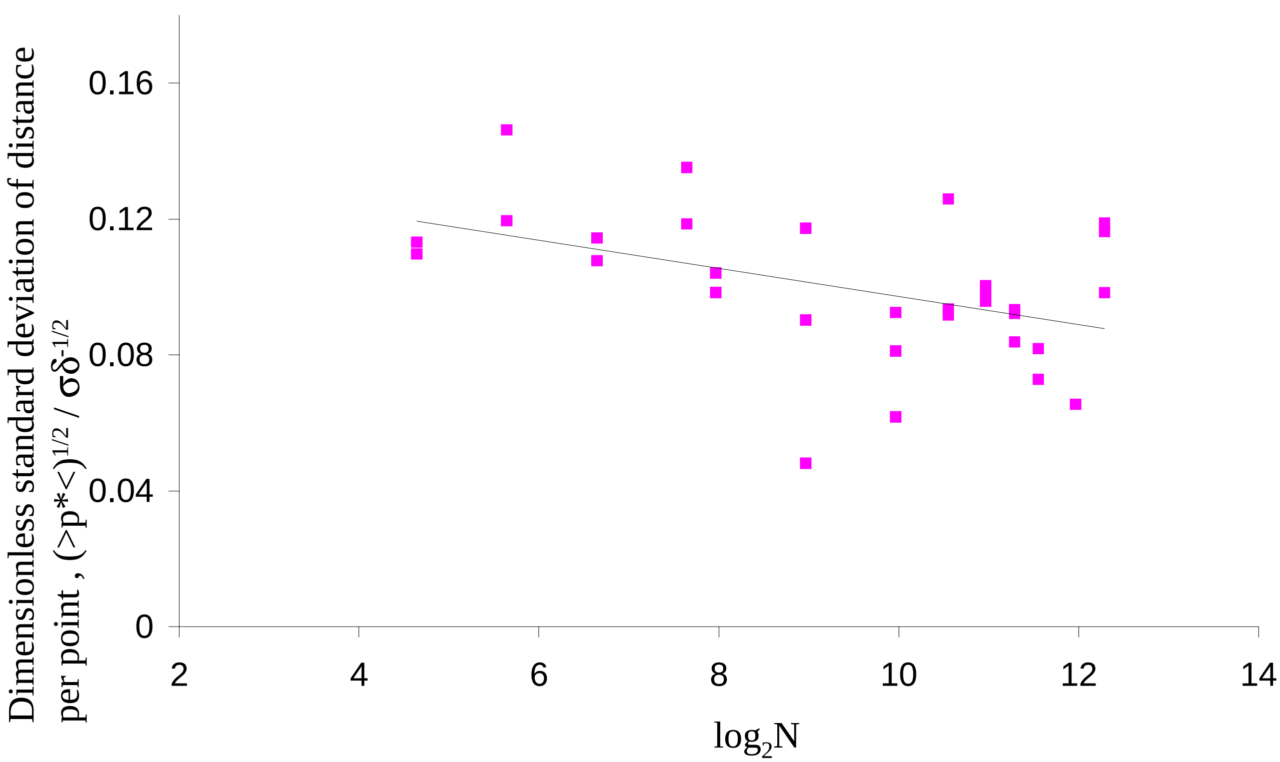

Figure3b. Dimensionless standard deviation of distance per point v. $\log_2 N$

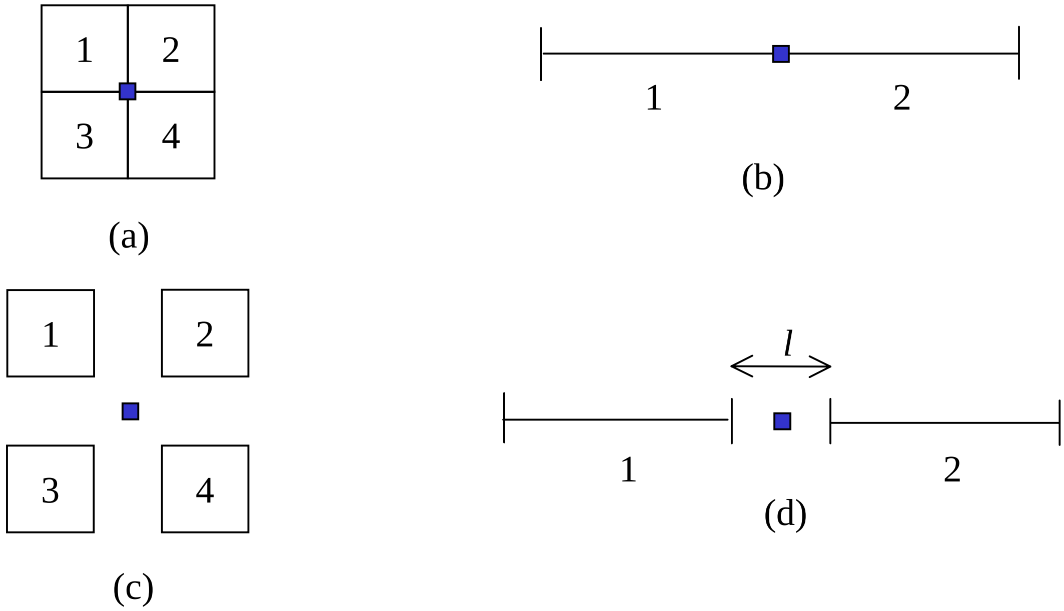

Figure 4. Shape effects

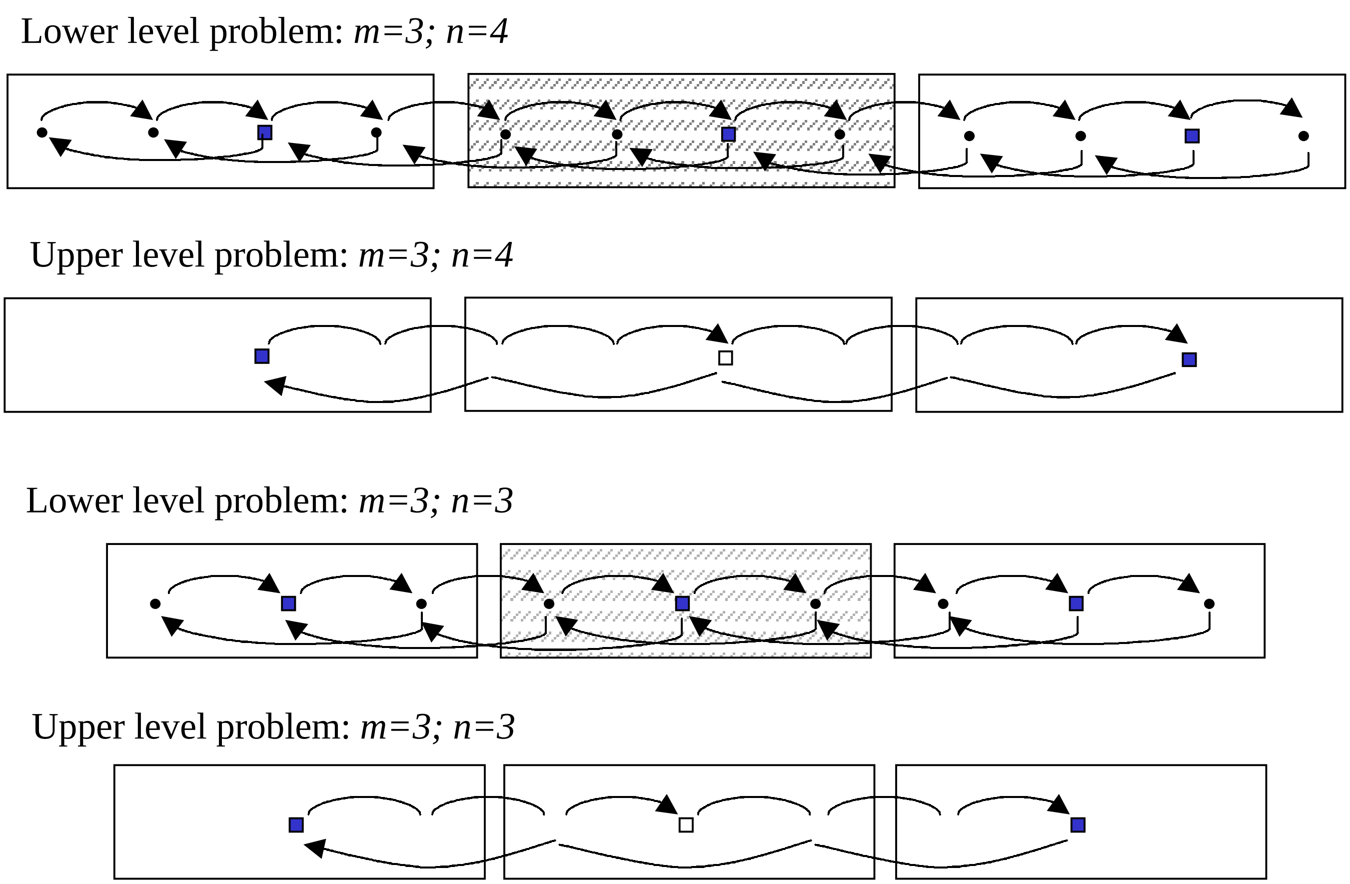

Figure 5. Scalable networks: Linear tile directed graph

APPENDIX

The total distance per point traveled was estimated using a Monte Carlo simulation of a single commodity transportation problem. It is assumed first that the service region is square and node coordinates were drawn from a uniform distribution over the service area. The net supply of all items at each point is normally distributed with mean 0 and standard deviation σ. The problem is assumed to be balanced, i.e., total supply equals total demand over the service area. A Euclidean metric was used to calculate distance.

Twenty-five test runs at various levels[1] of the following three parameters were performed: number of nodes (N), service area (A), and standard deviation in items (σ). In total, 769 simulations were run, with area ranging from 4000 to 90,000 area-units, 25 to 5,000 nodes, and standard deviation from 4.9 to 12.6.

[1] For N=4000, only 19 runs were performed

Simulation results

				distance per point	
Test set	N	A	σ	sample mean	sample standard deviation
1	25	5000	8.9	65.9	13.9
2	25	5000	6.3	47.3	10.1
3	50	5000	8.9	51.3	13.1
4	50	5000	6.3	39.6	7.6
5	100	4000	6.3	25.2	4.6
6	100	10000	8.9	56.6	9.6
7	200	10000	6.3	31.3	6.0
8	200	10000	8.9	41.1	7.5
9	250	5000	6.3	19.3	2.9
10	250	5000	4.9	14.3	2.2
11	500	75000	6.3	55.2	7.0
12	500	40000	8.9	56.4	3.8
13	500	40000	6.9	44.7	7.3
14	1000	40000	5.7	25.2	2.2
15	1000	50000	8.9	45.7	5.1
16	1000	50000	12.6	64.3	8.3
17	1500	50000	8.2	36.1	4.4
18	1500	50000	6.3	27.8	3.3
19	1500	50000	5.8	25.1	4.2
20	2000	50000	8.9	34.2	4.3
21	2000	75000	8.9	42.1	5.5
22	2000	75000	6.3	29.5	3.8
23	2500	75000	5.7	23.8	2.9
24	2500	90000	6.9	32.6	3.8
25	2500	75000	8.2	33.5	3.8
26	3000	75000	5.8	22.3	2.1
27	3000	75000	8.2	30.9	3.3
28	4000	90000	7.9	29.5	2.5
29	5000	75000	6.3	19.1	2.9
30	5000	75000	8.9	26.9	3.4
31	5000	90000	7.1	23.2	3.5