A GIS-BASED DECISION SUPPORT SYSTEM FOR ANALYSIS OF ROUTE CHOICE IN CONGESTED URBAN ROAD NETWORKS

Yi-Hwa Wu[1]
Harvey J. Miller[2 *]
Ming-Chih Hung[1]

[1]Digitally Integrated Geographic Information Technologies (DIGIT) Laboratory,
University of Utah
[2]Department of Geography, University of Utah

260 S. Central Campus Dr. Room 270
Salt Lake City, UT 84112-9155
Email: {wu.yi-hwa, harvey.miller, mhung}@geog.utah.edu

[*]Author for correspondence

Acknowledgements: A grant from the Military Traffic Management Command - Transportation Engineering Agency through the University of Tennessee Transportation Center partially supported this research. The Digitally Integrated Geographic Information Technologies (DIGIT) Laboratory, University of Utah, provided additional support and computing facilities.

Abstract: Urban congestion patterns have become ubiquitous and complex. Traditional, “static” approaches are no longer adequate for analyzing network flows and conducting minimum cost routing. This paper reports on a GIS-based decision support tool for modeling dynamic network congestion and conducting minimum cost routing. The system predicts network flows at a detailed level of temporal resolution, capturing dynamic congestion propagation effects. A *Route Planner* module solves for the combined departure time and minimum cost routing required for a trip to reach its destination by a given deadline. The GIS provides effective decision support through its database management capabilities, graphical user interfaces and cartographic visualization. This supports analyses of "what-if?" scenarios for strategic planning and tactical management subject to unplanned network disturbances.

1. Introduction

There is little question that traffic congestion is a severe and growing problem in many urban areas across the globe. This unfortunate trend is likely to continue through the foreseeable future. In addition to its increasing incidence, congestion is also exhibiting pronounced spatio-temporal complexity due to changing usage of urban space and the daily time cycle for employment and social activities (Cervero 1986; Hanson 1995). These changes are confounding the traditional tools used for public sector transportation and land use planning as well as private sector logistical analysis.

Most of the transportation analysis tools used in public sector planning are based on a "static" view of network flows and travel costs. For example, most transportation and land use forecasting tools assume implicitly or explicitly that network flows reach a steady-state equilibrium; this is explicit in the commonly-used Wardrop's "user optimal" principle for network assignment (Wardrop 1952). While viable historically, this is an increasingly unrealistic treatment of contemporary urban transportation networks. As Robert Dial eloquently states, "…a static model of congestion is an oxymoron" (Dial 1995). While intelligent transportation systems (ITS) offer some promise in capturing realistic depictions of urban network flows, these systems can only report on traffic conditions in (near) real-time and do not forecast future flows resulting from changes in land use, infrastructure, transportation policy or demographics.

Similarly, private sector analysts solving tactical logistical problems typically conduct routing and scheduling using static network travel costs. The inaccuracy of the resulting solutions is exacerbated by the increasing reliance on just-in-time (JIT) procurement, production and distribution systems. An unplanned network disruption such as construction or accidents can propagate delays through the system and interfere with the intricate shipment timings required for JIT systems.

Clearly, new planning and logistical tools are required that capture dynamic congestion patterns in urban transportation networks. This paper reports on a GIS-based dynamic traffic assignment (DTA) decision support system, GIS-DTA, for dynamic congestion modeling and minimum cost routing. The system uses a discrete-time dynamic network assignment procedure to predict network flow at a detailed level of temporal resolution. Given an arrival deadline and destination, a *Route Planner* module estimates the combined departure time and minimum cost route based on the predicted dynamic traffic flows. The system can also assess the impact of unplanned network disruptions on network congestion, routing and estimated arrival times, supporting "what-if?" scenario modeling.

The next section of this paper reviews properties of urban network congestion and methods for network flow modeling, emphasizing the need for dynamic congestion predictions. Section 3 describes the methodology and system configuration used for developing the dynamic congestion module within GIS-DTA. Section 4 shows some preliminary results. Section 5 concludes the paper with some summary comments and directions for continued system development.

2. Literature Review

2.1. Urban Transportation Congestion

Traffic congestion has become a very undesirable aspect of urban living throughout the world. The increasing level of traffic congestion in urban areas is creating substantial negative externalities to firms and individuals. Firms suffer losses in productivity due to lengthy commute times for employees and difficulties in operating intricately-timed logistical operations such as "just-in-time" (JIT) procurement, distribution and production systems. In addition to degrading the quality-of-life for individuals, traffic congestion restricts accessibility to recreation, shopping, employment and other opportunities in urban areas by reducing the amount of time available for these activities. At the extreme, urban traffic congestion can cause serious harm to property and lives since response times for emergency vehicles can be highly uncertain due to unforeseen traffic congestion incidences.

Unfortunately, urban traffic congestion will likely worsen over the foreseeable future due to structural trends in both developed and developing regions of the world. Although population growth is tapering in many developed regions of the world, urban congestion is continuing to worsen due to continued suburbanization and increasing intensive use of the automobile. In developing regions, these trends combine with rapid population growth and rural-to-urban migration to create transportation systems that operate at near-standstill for substantial portions of the day. Consequently, in most urban areas, increases in transportation system capacity cannot match the corresponding increases in the volume of vehicles (Cervero 1986; Hanson 1995).

Traffic congestion patterns have become spatially and temporally complex. In the past, avoiding congestion was a simple matter of avoiding travel from the urban periphery to the core during the morning commute and the reverse during the afternoon commute. However, this simple spatio-temporal congestion pattern is being supplanted by less intuitive patterns. Congestion is becoming more geographically dispersed, even ubiquitous, due to declines in manufacturing employment relative to more geographically dispersed service sector employment. Service sector employment is also more dispersed temporally, with the traditional daily business cycle being replaced by "round-the-clock" employment cycles. Another complicating factor is a rapid increase in non-work trips relative to work trips (Cervero 1986; Hanson, 1995). These trends are combining to create complex congestion patterns being spread to all parts of an urban area as well as to other times of the day beyond the traditional morning and evening peak periods.

Traffic congestion is a dynamic phenomenon. Congestion does not occur everywhere, all-at-once. Instead, congestion occurs in specific locations and a propagates through network over time as congested conditions on a link spread to nearby links. In addition, since many urban transportation networks are operating at near capacity, they are especially vulnerable to congestion occurring as the result of unplanned incidents such as accidents and infrastructure failures (e.g., bridge closings, construction). These incidents result in congestion patterns that propagate from a localized incident through many portions of the network, potentially resulting in serious flow disruption.

The increasing ubiquity and complexity of urban congestion combined with its severe negative impacts suggests the need for new tools to analyze and predict congestion patterns. These tools are important both for tactical operations and strategic planning. Tactical operations such as JIT systems require tools for optimal routing and timing given expected traffic patterns. Also important is the ability to perform "what if?" scenario modeling to assess the robustness of tactical plans subject to unplanned incidences (e.g., accidents, bridge closings) and their effects on congestion patterns. Technologies such as intelligent transportation systems (ITS) that attempt to improve transportation efficiency also require methods for predicting future congestion patterns if intelligent travel recommendations are to be disseminated to travelers. Strategic transportation and land-use planning also requires methods for assessing the congestion patterns that result from new transportation system components or land use activities.

2.2. Equilibrium Analysis of Congested Networks

An effective approach to modeling transportation network congestion is through *equilibrium analysis*. The equilibrium approach captures the relationship between users' travel decisions and network performance. In congested networks, each user's route choice influences the travel cost of other users due to induced delays. Adjustments occur between users' decisions and network performance until a balanced pattern is achieved. This pattern is a network equilibrium in the sense that each traveler has no future incentive to change their route choices unless some external disturbance occurs (Sheffi, 1985). Empirical evidence, albeit limited in scoped, supports the existence of network equilibrium (Florian and Nguyen, 1976). Although not necessarily accurate with respect to predicting individual behavior, equilibrium methods can reproduce aggregate system properties and structural trends in a manner sufficient for tactical and strategic decision making.

The fundamental model of congested network equilibria is the *user optimal* (UO) principle, originally due to Wardrop (1952). UO requires that, at network equilibrium, no traveler can reduce his or her travel costs by unilaterally changing routes. An equivalent statement of this condition is that all used routes between an origin-destination (O-D) pair have the same, minimal cost and no unused route has a lower cost. Effective algorithms exist that solve for the network flows that correspond to the UO pattern; these include the *convex combinations method*, originally due to Frank and Wolfe (1956) (see Sheffi (1985) for an excellent introduction to this method).

The standard, "static" UO approach is most appropriate for long-term infrastructure planning and policy. It assumes a "steady-state" equilibrium and therefore cannot capture detailed temporal and spatial dynamics. However, these dynamics are important for capturing congestion patterns and properties plague many urban areas (Ben-Akiva 1985).

2.3. Dynamic Network Flow Modeling

The weakness of the steady state assumption in the UO condition has been long recognized by transportation researchers. Initial work on developing dynamic flow models began in the late seventies with Merchant and Nemhauser's (1978a, 1978b) pioneering work. Little progress was achieved during the subsequent decade, probably due to the problem's computational complexity and a lack of applications for the methods. This changed a decade later with the advent of the intelligent vehicle highway system (IVHS) program in the United States, now officially known as the ITS program. The power of computational platforms have also increased substantially, allowing researchers to solve dynamic flow patterns in increasingly realistic networks. Several approaches to the dynamic network flow problem have emerged, including: i) simulation-based approaches; ii) optimal control theory; iii) variational inequality; iv) dynamic systems approaches; and, v) mathematical optimization.

Simulation-based dynamic network flow modeling uses numerical methods to predict dynamic flow and network performance, often within a traditional iterative flow assignment procedure such as incremental assignment. Several systems have been developed, including DynaMIT (Massachusetts Institute of Technology 1996) and DYNASMART (Mahmassani and Peeta, 1993, 1995; Jayakrishnan et al. 1994; Hu and Mahmassani, 1995). Simulation models can approximate traffic dynamics accurately through the use of appropriate traffic flow functional relationships. Simulation-based approaches can also handle large, urban-scale transportation networks (Ben-Akiva, *et al*., 1997). Nevertheless, they do not have the attractive properties of the analytical models. In particular, solution quality is highly variable because they are not required to converge on a solution with given properties. Thus, the convergence characteristics and the solution quality cannot be determined.

The objective of conventional optimal control theory is to determine control strategies that cause a process to satisfy the physical constraints while at the same time minimize or maximize performance criterion. Wie (1989) first applied optimal control theory to the problem of dynamic network flows using a generalized UO criterion applied to a network with one origin-destination pair connected with parallel links. Several efforts had been made to extend this approach to continuous time, multiple origin-destination formulations (Friese et al., 1989; Wie et al., 1990; Ran et al, 1993; Boyce et al., 1995; Wie et al., 1994). While powerful, optimal control formulations can generate flow propagation anomalies such as congested traversal times being less than free flow times.

An alternative but related approach is to describe the behavior of traffic network as a variational inequality (VI) problem with exact flow propagation constraints. Friesz et al. (1993) developed a VI-based formulation based on a dynamic generalization of the static UO conditions. The model can be solved for a general network with endogenous departure times and route choices. Ran and Boyce (1996) presented some mathematical formulation and solution algorithms for dynamic transportation models which mostly are optimal control theory and variational inequalities. Chen (1999) provides a detailed discussion of variational inequality formulations for dynamic travel choice models. Because of its comprehensive capability to formulate transportation problems, VI has received in increasing attention from transportation network researchers (Wu et al., 1998; Xu et al., 1999). Although optimal control theory and VI-based approaches have improved our theoretical understanding of dynamic travel demands and network flows, their computational complexity and data requirements restricts their applications to modest network sizes for many application contexts.

The dynamic system approach generally takes the form of a system of differential equations that describe a trajectory of disequilibrium states tending towards an equilibrium (Cascetta and Cantarella, 1991). Depending on the assumptions made and the behavior captured, this equilibrium may be either a conventional steady state (static) UO equilibrium or a dynamic (moving) UO equilibrium (Friesz et al, 1996). Cantarella and Cascetta (1995) and Friesz et al. (1994) provide a detailed discussion of this approach. The dynamic systems approach has potential to provide a unified framework for both within-day and day-to-day (i.e., long-term learning effects) network flow dynamics. However, a practical solution algorithm has not been developed to date.

Optimization approaches directly generalize the static UO approach with suitable temporal modifications. In this approach, time-dependent network flows are defined by a set of link performance functions and an equilibrium condition that extends the Wardrop's UO principle. Merchant and Nemhauser (1978a, 1978b) provide an example of this strategy applied to a simple network with multiple origins but only one destination. Since their model is nonlinear and non-convex, it is difficult to apply to a general network. Ho (1980) formulated a piecewise linear version of the Merchant-Nemhauser model. Ho (1990) develops a nested decomposition algorithm that finds the global optimum through examining the set of optimal solutions of an embedded linear program. Carey (1986, 1987) expands this approach and reformulates the model as a convex, non-linear program.

Recently, Janson greatly improved the applicability of dynamic network flow modeling to real-world problems by developing a tractable discrete-time *dynamic user optimal* (DUO) approach (Janson 1991a, 1991b) The DUO conditions are a direct extension of Wardrop's UO conditions:

> (DUO) At network equilibrium, no traveler who departed (arrived) during the same time interval can reduce his or her travel costs by unilaterally changing routes. *Alternatively*: All used routes between an O-D pair have the same, minimal cost and no unused route has a lower cost for travelers that departed (arrived) during the same time interval.

Since travel times are variable, we cannot "fix" both departure and arrival times within the equilibrium conditions. Therefore, the DUO conditions assume either a known (fixed) departure or arrival time interval for flows and require equivalent minimal travel costs for all flows "scheduled" to depart or arrive during each interval.

Janson (1991a, 1991b) formulates a DUO equivalent optimization problem with a convex objective function and non-linear dynamic flow constraints. The DUO problem can be solved for realistic, urban-scale network with reasonable computational times (Robles and Janson 1995; Boyce et al. 1997). The DUO data requirements are also quite modest, with the major additional data item relative to the static UO requirements being a temporal origin-destination network at the level of resolution dictated by the discrete time intervals.

Janson and Robles (1995) extend the basic DUO formulation and develop a *quasi-continuous* DUO. Since it approximates better continuous-time trajectories, the quasi-continuous DUO approach handle dynamic flow propagation and spillback effects in a more realistic manner. However, its is not as computationally tractable as the standard DUO approach.

Due to its computational efficiency, particularly for modest, desktop platforms, we have chosen the Janson (1991a, 1991b) DUO model for our GIS-linked dynamic network flow modeling software system. In the next section of this paper, we discuss the software system design and the DUO approach in detail.

3. Methodology

The dynamic network congestion modeling software system is a decision support tool that interfaces the Janson (1991a, 1991b) DUO module with a geographic information system (GIS), specifically, Arc/Info®. The DUO module solves for the dynamic network flows while the GIS provides effective decision support through its database management capabilities, graphical user interfaces and cartographic visualization of complex spatial and temporal congestion patterns and the time-dependent minimum cost routes. The GIS interface also allows the analyst to change the road network to reflect failure or capacity reduction due to planned or unplanned disruption.

In this section, we discuss the methodology for the dynamic network congestion modeling software system. We first discuss the DUO model formulation and its solution algorithms. We then briefly discuss its data requirements. Finally, we provide an overview of the software system and discuss some critical database design issues.

3.1.DUO Model Formulation

The DUO equilibrium condition is that all used routes between an origin-destination (O-D) pair have same minimal cost and no unused route has a lower cost for travelers that departed during the same time interval. The DUO model assumes a known temporal O-D matrix, with each time slice corresponding to a discrete time interval over the study horizon. Based on this exogenous data, the DUO minimization problem, when solved, determines the dynamic flow patterns that satisfies the DUO principle while meeting the O-D flow constraints imposed by the matrices. The DUO problem is:

$$MIN \quad \sum_{k\in L}\sum_{t\in T}\int_0^{x_k^t} f_k^t(w)dw \tag{ 1}$$

Subject to

$$x_k^t = \sum_{p\in P}\sum_{d\in T} v_p^d \alpha_{pk}^{dt} \qquad \text{for all } k\in L\,,\ t\in T \tag{ 2}$$

$$q_{rs}^d = \sum_{p\in P_{rs}} v_p^d \qquad \text{for all } r\in Z\,,\ s\in Z\,,\ d\in T \tag{ 3}$$

$$v_p^d \geq 0 \qquad \text{for all } p\in P\,,\ d\in T \tag{ 4}$$

$$\alpha_{pk}^{dt} \in \{0,1\} \qquad \text{for all } p\in P\,,\ k\in K_p\,,\ d\in T\,,\ t\in T \tag{ 5}$$

$$\sum_{t\in T}\alpha_{pk}^{dt} = 1 \qquad \text{for all } p\in P\,,\ k\in K_p\,,\ d\in T\,,\ t\in T \tag{ 6}$$

$$b_{pn}^d = \sum_{t\in T}\sum_{k\in K_{pn}} f_k^t(x_k^t)\alpha_{pk}^{dt} \qquad \text{for all } p\in P\,,\ n\in N\,,\ d\in T\,,\ t\in T \tag{ 7}$$

$$[b_{pn}^d - t\Delta t]\alpha_{pk}^{dt} \leq 0 \qquad \text{for all } p\in P\,,\ n\in N\,,\ k\in L_n\,,\ d\in T\,,\ t\in T \tag{ 8}$$

$$[b_{pn}^d - (t-1)\Delta t]\alpha_{pk}^{dt} \geq 0 \qquad \text{for all } p\in P\,,\ n\in N\,,\ k\in L_n\,,\ d\in T\,,\ t\in T \tag{ 9}$$

where

N = set of all nodes;
Z = set of all origin-destination zone (trip begin/end nodes);
L = set of all links (directed arcs);
L_n = set of all links incident form node *n*.
P = set of all routes between all zone pairs.
P_{rs} = set of all routes from zone *r* to zone *s*;
K_p = set of all links on route *p*;
K_{pn} = set of all links on route *p* prior to node *n*;
Δt = duration of each time interval (same for all *t*);
T = set of all time intervals in the full analysis period;
x_k^t = amount of traffic flow between all zone pairs assigned to link *k* in time interval *t*;
v_p^d = amount of traffic flow departing in time interval *d* assigned to route *p* ;

$f_k^t(x_k^t)$ = travel impedance (travel time) on link k in time interval t;

q_{rs}^d = amount of traffic flow from zone r to zone s departing in time interval d via any route;

α_{pk}^{dt} = 0-1 variable indicating whether trips departing in time interval d and assigned to route p use link k in time interval t (0 = no, 1 = yes) ;

b_{pn}^d = travel time of route p from its origin to node n for trips departing in time interval d ;

The DUO formulation implies a directed space-time network $G(N, L, T)$, where N is the set of nodes, L is the set of directed links and T is study time horizon (i.e., the number of discrete time intervals). The temporal incidence variable α_{pk}^{dt} indicates topology of the time-expanded network. Equation (5) defines α_{pk}^{dt} to be a zero-one variable that indicates whether trips assigned to route p departing in time interval d using link k in time interval t. Equation (2) requires total flow on link k in time interval t to be the sum of flows departing in any time interval on any route that uses link k in time interval t. Flows on route p departing in time interval d will only be assigned to link k in time interval t as designated by the α_{pk}^{dt}. Equation (3) requires route flows to sum to the proper trip departure totals in each time interval between each zone pair. Equation (4) requires the flows on all routes to be nonnegative. Equation (6) ensures that the trip of any route p departing in a given time interval d can be assigned to a particular link k in only one time interval t. This requires any link k to only be used in the time interval t in which trips assigned to route p departing in time interval d.

Equations (1-5) collapse to a static UO equilibrium problem for a single time interval: in this case, the incidence α_{pk}^{dt} could be pre-specified as either zero or one based on the input network's topology. In the time-expanded network implied by the DUO, topology is determined by physical connectivity and space-time "reachability" dictated by the network travel times. Travel times are determined by the link loadings in the current solution. Therefore, the α_{pk}^{dt} variables are endogenous and fluid, changing with the network flows in the current solution.

Since the node time intervals are endogenous, the DUO problem has nonlinear flow conservation constraints and requires additional conditions to insure temporally continuous flows (equation (7-9)). Equation (7) is added to sum the travel times to each node *n* along links in each route *p*, denoted by the link set K_{pn}. Equations (8-9) then force each route to use its links, given by the link set K_p, in the time intervals that are compatible with the travel time to each node.

3.2.DUO Solution Algorithms

Janson (1991a, 1991b) provides two solution approaches are available for solving the DUO problem. The *dynamic traffic assignment* (DTA) procedure is a heuristic approach. Janson's DTA procedure assigns link flows based on current flow levels, ratios of future (not yet assigned) travel demands and flows assigned in previous intervals (Janson, 1991a). It is not a convergent solution for DUO equilibrium, but instead produces traffic flows that tend to satisfy the DUO conditions. The *convergent dynamic algorithm* (CDA) solves the DUO problem exactly by decomposing the main DUO problem into two subproblems. The first subproblem is equivalent to the static UO assignment problem. It can be solved using the convex combination methods, such as. Frank-Wolfe (F-W) method of linear combination (Frank and Wolfe, 1956). The second subproblem updates the temporal incidence variables and enforces conditions for temporally continuous flows. The procedure iterates between the static UO and the temporal flow subproblems until computational convergence (Janson 1991b).

Due to the decomposition procedure, CDA requires 3-4 times as much computational effort for a simple network than Janson's DTA (Janson, 1991b). Since we are developing our software system for common desktop platforms, our current implementation uses Janson's DTA procedure. In the future system development, the CDA procedure will also be implemented for the quasi-continuous DUO formulation (Janson and Robles, 1995) to allow users to opt for the slower but more accurate solution procedure for smaller problems or if extended computation time is available.

A key element of Janson's DTA procedure is the tracking of vehicle trip flows through a network in a link-by-link or node-by-node basis in successive time intervals.

Flows must be tracked across the network by the time interval and by the destination to which they are headed. In other words, the procedure makes route decisions at the time of trip departure on the basis of projected link travel times that account for changes in travel demand over future time intervals. Link travel times in each time interval are computed based in each link's volume: these in turn determine the routes to which vehicles will be assigned.

Figure 1 provides a detailed flowchart for Janson's DTA procedure. In the initialization stage, two issues must be addressed: i) appropriate duration for the discrete time intervals; and ii) network link flow in "previous" (prior to study horizon) time intervals. The appropriate interval duration is crucial for estimating the dynamic traffic flow using Janosn's DTA procedure. If the time duration is too long, then the dynamic flows will fall into a static equilibrium; if it is too short, unrealistic flow propagation effects can occur. Janson (1991a) suggests choosing an interval that is approximately four or five times the average link travel times in the study network. This minimizes flow estimate variation between intervals.

Since Janosn's DTA procedure finds and loads shortest path trees based on projected link volume in current and future time intervals, it requires the traffic flow data in previous time intervals to estimate link volumes into the initial intervals of the study time horizon. A larger number of pre-study horizon time intervals improves accuracy of the initial interval within the study horizon but requires additional data.

The main computational effort of Janson's DTA procedure is to find a shortest path tree from each origin to all other destinations based on the projected link travel times in the current and future time intervals. The shortest path algorithm used is the well-known Dijkstra algorithm (1959). The search routine finds shortest paths based on link travel times as projected during the time intervals in which they will be traversed. To minimize memory array space allocation, we calculate the projected link volumes and link travel times as needed in the shortest path routine. The main computational increase of this routine over a "static" shortest path routine is that the time interval in which each tree uses each link must be recorded.

For each link, the current and projected link volumes are estimated as weighted combinations of the final volume assigned to that link in the previous interval (t - 1) and the volume assigned thus far in the current interval *t*. The weighting is by ratios of total trip departures form all origins in intervals (*t* - 1), *t*, and (*t* + *n*), where *n* is the number of time intervals in the study horizon.

$$y_{ij}^{t+n} = \theta^t \omega_{t-1}^{t+n} x_{ij}^{t-1} + \left(1-\theta^t\right)\omega_t^{t+n} x_{ij}^t \tag{11}$$

for all, $ij \in K$, $t+n \in T$, $n \geq 0$

where :

x_{ij}^t = assigned volume on link *ij* in the current interval *t* after trips departing in time intervals 1 through (*t* - 1) have been assigned, and while trips departing in time interval *t* are being assigned.

x_{ij}^{t-1} = assigned volume on link *ij* in time interval (*t* - 1) (i.e. just prior to the current time interval) after all trips departing in time intervals 1 through (*t* - 1) have been assigned.

y_{ij}^{t+n} = projected volume on link *ij* in interval *t* + *n* (where $n \geq 0$ and $t + n \in D$) after trips departing in time intervals through (*t* - 1) have been assigned, and while trips departing in interval t are being assigned.

$f_{ij}^{t+n}(y_{ij}^{t+n})$ = projected travel impedance (travel time) of link *ij* in the current or future time interval $(t+n)$ computed directly as a function of the projected volume y_{ij}^{t+n}, where $n \geq 0$ and $t+n \in T$.

Q^t = total number of trips departing from all zones in time interval *t* (i.e. total inflow to the network in time interval *t*).

θ^t = proportion of Q^t that has not yet been assigned to the network in time interval *t*.

$\omega_{t-1}^{t+n} = Q^{t+n}/Q^{t-1}$,a measure of systemwide travel demand and projected traffic volumes in time interval *t* + *n* relative to *t* - 1.

The standard “Bureau of Public Roads” link performance function, which relate capacity, current volume and free-flow traversal time to estimate travel time as a function of current flow, are used in current implementation for measuring link impedance (see Ortúzar and Willumsen 1994).

Janson’s DTA procedure assigns trips from origins in random order so as to minimize ordering effects due to the sequential loading of flows on the network links. This can create variability in flow volumes among time intervals not related to systematic changes in travel demand. In order to reduce this variability, Janson’s DTA procedure finds NTREES shortest paths from each origin during a given time interval, where NTREES is some positive integer. The procedure evenly divides the flows departing from the origin during that time interval among the NTREES shortest path trees, i.e., each tree receives 1/NTREES of the departing flows during the time interval. A larger NTREES reduces the random variability at the expense of greater computational burden.

3.3. Data Requirements

Data requirements for the DUO model include link performance functions and a temporal origin-destination (O-D) matrix at the same level of temporal resolution as the specified discrete time intervals. Ideally, the temporal O-D matrix should be constructed from flow data “tagged” with the time of day when each trip occurred. These data can be aggregated to the discrete time intervals of the DUO model. The critical “time stamp” is the departure or arrival time of each trip.

If a temporal O-D matrix cannot be obtained directly from primary data it must be estimated. As a member of ITS American, Utah Transportation Laboratory uses ITS technique and real-time traffic data to obtain the observed link traffic counts. Furthermore, Janson and Southworth (1992) discuss a method that uses the dynamic traffic assignment procedure to estimate departure times from observed link traffic counts; these data are often readily available. Another, less sophisticated, option is to temporally disaggregate a daily O-D flow matrix. The simplest method is to divide the O-D matrix equally into the *n* daily intervals implied by the specified time duration. However, since O-D flows typically exhibit irregular peaks rather than an even daily distribution, this approach is crude. Daily O-D flows could be distributed over the time period of interest by using daily peak profile curves; this would provide more realistic estimates of the time-dependent O-D flows

3.4. System Design

The current prototype system is a “loosely-coupled” integration between Arc/Info$^{®}$ (version 7) GIS software and a custom DTA module written in C++. The DTA module is a stand-alone system that runs separately from the DOS command prompt. Although running as a separate program, the program directly reads and writes Arc/Info$^{®}$ INFO files, allowing the GIS software to manage the input data and visualize model results.

The procedure for running the prototype system is as follows. The user first develops the required data as Arc/Info$^{®}$ coverages. An ARC coverage represents the transportation network in the study area while a POINT coverage maintains the origin and destination locations and the O-D matrix (i.e., each point in the coverage maintains data on its flow to all other destinations, in other words, its "row" of the O-D matrix). The DTA procedure reads the INFO files for these coverages and writes new INFO files, one file for each time interval modeled. These can be visualized and queried within the cartographic context of the network coverage using Arc/Info.

In addition to the standard visualization and spatial data querying techniques available within Arc/Info$^{®}$, we have built a *Route Planner* interface that allows the user to conduct minimum cost routing subject to the estimated dynamic flows and congestion within the network. One procedure allows the user to find a minimum cost route between an O-D pair based on a specified departure interval. The procedure reads the INFO file generated by the DTA procedure and returns the minimum cost route and an estimated arrival time at the destination. Another procedure allows the user to input an O-D pair, the earliest possible departure time from the origin and the latest possible arrival time at the destination. The procedure searches all the possible minimum cost routes across departure time intervals and generates a set of departure times and minimum cost routes that meets the specified time window.

3.5. Database Design

A critical issue when coupling the GIS software with the DTA module is the spatio-temporal database model. The DTA procedure generates a large volume of spatio-temporal data. Ineffective data management will result in a computational bottleneck that can seriously degrade the performance of the software system. Also, instead of developing a completely customized database design, it is worthwhile to take advantage of the data management capabilities of existing software system, particularly Arc/Info® and possibly other commercial database management systems in future system development. These software systems offer other data input, data management, analytical and geographic visualization capabilities that can provide effective decision support.

Figure 2 provides the conceptual design of the spatio-temporal database to support the GIS-DTA coupling. This design treats time as a complementary and independent facet of spatial (the georeferenced network) and thematic (network attributes, travel demand) data domains. Successive time-stamped *versions* of database entities are maintained in a temporal data structure. A new version of a database entity is defined when some change occurs between successive time intervals. New versions of the flow on each link are recorded for each DTA time interval since this attribute changes frequently. Conversely, new versions of the network and non-flow link attributes will be recorded infrequently since these entities do not change often (at least with respect to the modeling domain). This design minimizes database storage requirements since data redundancy is at a minimum.

The database design allows a very straightforward topology to support querying and data retrieval. Each DTA projects has a header control file which contains pointers that connect the spatial, thematic, and temporal data. Spatial and thematic data are static and therefore only stored once. The temporal flow pattern (e.g. flow volume, travel time) is recorded for each time interval. The unique spatial object IDs from spatial database connects the three data domains.

The separation of themes, time and space in the spatio-temporal database design preserves compatibility with the data model and data structures currently used in GIS packages such as Arc/Info®. The three domains are separated into distinct structures but are united by the spatial object IDs and temporal indices that form the intersections between domains. The link-node network topology can be maintained in the standard vector spatial data format supported by most GIS packages. The static-state attributes of street network, such as capacity, lanes, speed limit, etc., can be maintained in either a commercial database management system (DBMS) or within the GIS software.

3.6. Decision Support Tool Development

The current GIS-DTA prototype system is complete with respect to its basic database design, dynamic flow computation and minimum cost route querying capabilities. At present, we are working on developing user-friendly graphical user interfaces and decision support tools. This will allow a non-technically oriented planner or logistical analyst to use the system with a minimum of guidance. This section describes some of these current system enhancements. The first enhancements concern the *Route Planning* module for conducting time-critical logistics while the second set concerns visualization enhancements to support public sector planning.

Multi-attribute route planning. The current decision support tools for dynamic routing are based only on travel time. Other attributes that could affect route preferences are currently not considered. For example, some drivers may prefer highway rather than arterial streets to avoid traffic signals. Similarly, we may wish to restrict drivers' routes to familiar rather than unknown areas. When shipping hazardous materials, we also may wish to incorporate geographic risk information through the GIS software and include this when planning for routes.

Incorporating non-temporal attributes into route selection requires decision support tools for multiattribute decision making. We are currently developing this functionality based on a powerful graphical technique developed by Coutinho-Rodriques et al. (1996). The *Best AGainst Least* (BAGAL) technique displays the properties of a given solution simultaneously along several specified attribute dimensions (e.g., cost, risk) normalized and with a common origin. Also plotted are user-specified limits for each dimension, allowing the decision maker to easily assess the solution sperformance along each dimension. The combination of attribute-space visualization with the geographical space visualization provided through the GIS software will provide effective decision support for multi-attribute route planning subject to time constraints.

*Animation Tools***.** Most of the dynamic network flow research in the literature emphasizes the convergence properties of the algorithms or the temporal resolution of results. However, the increasing temporal resolution of DTA solution has the potential to overwhelm the analyst attempting to assess the impacts of a change in infrastructure or traffic management control policy on congestion patterns. The development of visualization tools for dynamic traffic congestion modeling can greatly enhance the usability of the dynamic flow results.

Recent advances in dynamic visualization have occurred both in the transportation realm (e.g., Ganter and Cashwell 1994) and for dynamic data in general (e.g., Acevedo and Masuoka 1997). These techniques greatly enhance interpretation by supporting exploratory analysis of the extensive dynamic information generated from dynamic modeling and simulation systems. This can also provide decision support by structuring the search for good solutions and allowing easy comparisons among competing solutions.

In a previous project, we developed a "temporal interpolation" technique to create computer animations from satellite images that simulate gradual landscape change. We are currently modifying this technique to create computer animation to simulate network congestion dynamics. In the animations, the congestion levels of roads are gradually changed from one time point to the next, represented by changes of color. This technique is designed to overcome the visual gap created from displaying network dynamics by only showing the congestion level at selected time points.

4. Results

In this section, we illustrate some of the functionality of the GIS-DTA system using examples from a real world transportation network. The network comprises northeast Salt Lake City, Utah. It consists of 7812 directed links, 2328 nodes and 331 O-D zones. The time interval for model runs is five minutes. A three hours time horizon results in thirty-six consecutive time “slices” of dynamic traffic patterns. We derived a daily O-D matrix from a travel survey conducted by the University of Utah during spring 1994 and a local daily peak profile curve.

Figure 3 and Figure 4 represents the congestion pattern of the “Sugarhouse” area of Salt Lake City (the southeast quadrant of the study area) in different time intervals. For display purposes, Figures 2 and 3 classify links into one of two categories, namely, “normal traffic” (flow is less than 80% of design capacity) and “very congested” (flow is greater than 80% of capacity). Note that a pair of directed arcs represents two-way streets: these are offset slightly for display. These figures illustrate the detailed temporal and spatial resolution of flow patterns provided by the system. Figure 3 shows the initial period of study horizon (first five minutes) when the traffic congestion is still light while Figure 4 shows a middle interval of study horizon (one hour and 40 minutes later) when the whole area has become extremely congested. Note that, as expected, congestion exhibits a "contagious" pattern, that is, congestion tends to occur in proximal and adjacent links in the network. This realistic congestion property is not captured well by static network flow models.

Figure 5 and Figure 6 illustrate the *Route Planner* functionality as well as some of the complexities involved in minimum cost routing subject to dynamic congestion effects. Figure 5 provides the minimum cost (least travel time) routes for opposite travel directions between the same O-D zones in the same time interval. Different travel directions between the same locations face entirely different congestion patterns and hence require different minimum cost routes. Figure 6 illustrates the temporal complexity of congestion pattern. In the successive time periods, (intervals 34 and 35) the congestion pattern changes sufficiently that the minimum cost route is completely different from a geometric perspective. Despite this large geometric difference, the total travel times between the two routes are only slightly different.

The *Route Planner* allows a wide range of dynamic routing queries based on travel times and the time window available. Table 1 illustrates the results of a querying the possible departure times and minimum cost routes from Downtown Salt Lake City to the Sugarhouse area constrained by departing no earlier than interval 8 (40 minutes clock time) and arriving before the end of interval 21 (one hour and 45 minutes clock time). Table 1 summarizes the six combined departure times/routes that meet these temporal constraints. These routes can also be visualized using the *Route Planner* (although this is difficult to convey in this paper without color illustrations).

Table 1: Querying for departure times and routes within a specified time window

Departure Time Interval	Travel Time (minutes)	Arrival Time Interval	
8	32.90	14	Shortest travel time
9	35.70	16	
10	40.70	18	
11	43.52	19	
12	42.20	20	
13	40.68	21	Latest departure time

5. Conclusion

This paper discusses a prototype of a decision support tool for analyzing dynamic congestion patterns and conducting routing and logistical analysis subject to these flows. A dynamic congestion module captures complex spatio-temporal congestion patterns and dynamic propagation of congestion through the network from localized incidences. The dynamic traffic assignment procedure at the core of the module has reasonable data and computational requirements. The GIS software provides effective management of the detailed spatial data and the ability to query and visualize model results. A sophisticated spatio-temporal database design links the GIS with the dynamic flow module and enhances their capabilities by providing efficient data storage and retrieval of model results.

Our current efforts are directed towards continued development of decision support for the route planning and "what-if?" scenario modeling. The *Route Planner* module provides the optimal routing and timing to support tactical operation of time-critical logistical operations such as “just-in-time” systems. The current prototype provides decision support for routing based only on travel time. We are in the process of developing tools for multi-objective routing decisions to explore solutions simultaneously along several specified attribute dimensions (e.g., cost, risk). The animation techniques also under development will enhance interpretation by supporting exploratory analysis of the extensive dynamic information generated from the modeling system.

As stated previously, our current software is a "loosely-coupled" prototype system. We are continuing software development along several avenues. Our first priority is a "tightly-coupled" system with seamless integration among the dynamic congestion module, GIS software and decision support tools. We are also developing a more sophisticated dynamic congestion module based on the quasi-continuous DUO (Janson and Robles, 1995), which can handle dynamic flow propagation and spillback effects in a more realistic manner. This will allow the user to trade-off computation speed for accuracy when modeling dynamic flows.

Literature Cited

Acevedo, W. and Masuoka, P. (1997) "Time-series animation techniques for visualizing urban growth," *Computers and Geosciences*, 23, 423-435.

Ben-Akiva, M. (1985) “Dynamic network equilibrium research,” *Transportation Research A*, 19A, 429-431.

Ben-Akiva, M. E., Koutsopoulos, H. N. Mishalano, R. G. and Yang, Q. (1997) “Simulation laboratory evaluating dynamic traffic management systems,” *Journal of Transportation Engineering*, ACSE, pp. 283-289.

Boyce, D. E., Ran, B. and LeBlanc, L. J. (1995) “Solving an instantaneous dynamic user-optimal route choice model,” *Transportation Science*, 29(2), 128-142.

Boyce, D. E., Lee, D-H, Janson, B. N. and Berka, S. (1997) “Dynamic route choice model of large-scale traffic network,” *Journal of Transportation Engineering*, ASCE, 123(4), 276-282.

Cantarella, G. E. and Cascetta, E. (1995) “Dynamic processes and equilibrium in transportation networks: towards a unifying theory,” *Transportation Science*, 29(4), 305-329.

Carey, M. (1986) “A constraint qualification for a dynamic traffic assignment model,” *Transportation Science*, 20, 55-58.

Carey, M. (1987) “Optimal time-vary flows on congested networks,” *Operations Research*, 35, 58-69.

Cascetta, E. and Cantarella, G. E. (1991) “A day-to-day and within-day dynamic stochastic assignment model,” *Transportation Research A*, 25A, 277-291.

Cervero, R. (1986) *Suburban Gridlock*, New Brunswick, N.J.: Center for Urban Policy Research.

Chen, H.-K. (1999) *Dynamic Travel Choice Models: A Variational Inequality Approach*, Springer-Verlag.

Coutinho-Rodriques, J., Current, J., Climaco, J. and Ratick, S. (1996) “An interactive spatial decision support system for multiobjective hazmat location-routing problems,” Working Paper WPS 95-39, Max M. Fisher College of Business, The Ohio State University

Dial, R. B. (1995) “T2: Another multipath probabilistic traffic assignment model that obviates path enumeration,” working paper, Volpe National Transportation Systems Center, Cambridge, MA.

Dijkstra, E. W. (1959) “A note on two problems in connexion with graphs,” *Numerische Mathematik*, 1, 269-271.

Florian, M and Nguyen, S. (1976) “An application and validation of equilibrium trip assignment models,” *Transportation Science*, 10, 374-390.

Frank, M. and Wolfe, P. (1956) “An algorithm for quadratic programming,” *Naval Research Logistics Quarterly*, 3, 95-110.

Friesz, T. L., Luque, F. J., Tobin, R. L. and Wie, B.-W. (1989) “Dynamic network traffic assignment considered as a continuous time optimal control problem,” *Operations Research*, 37, 893-901.

Friesz, T. L., Berstein, D., Smith, T. E., Tobin, R. L. and Wie, B.-W. (1993) “A variational inequality formulation of the dynamic network user equilibrium problem,” *Operations Research*, 41, 179-191.

Friesz, T. L., Berstein, D. H., Mehta, N. J., Tobin, R. L. and Ganjalizadeh, S. (1994) “Day-to-day dynamic network disequilibrium and idealized driver information systems,” *Operations Research*, 42, 1120-1136.

Friesz, T. L., Bernstein, D. and Stough, R. (1996) “Dynamical systems, variational inequalities and control theoretic models for predicting time-varying network flows,” *Transportation Science*, 30, 14-32.

Ganter, J. H. and Cashwell, J. W. (1994) "Display techniques for dynamic network data in transportation GIS," *GIS-T '94 Proceedings*, 42-53.

Hanson, S. (1995) “Getting there: Urban transportation in context,” in S. Hanson (ed.) *The Geography of Urban Transportation*, 2ed., New York: The Guilford Press, 305-341.

Ho, J. K. (1980) “A successive linear optimization approach to the dynamic traffic assignment problem,” *Transportation Science*, 14, 295-305.

Ho, J. K. (1990) “Solving the dynamic traffic assignment problem on a hypercube multicomputer,” *Transportation Research B*, 24B, 443-451.

Hu, T. Y. and Mahmassani, H. S. (1995) “Evolution of network flows under real-time information: Day-to-day dynamic simulation assignment framework.” *Transportation Research Record*, 1493, 46-56.

Janson, B. N. (1991a). Dynamic traffic assignment for urban road networks, *Transportation Research B*, 25B: 143-161.

Janson, B. N. (1991b) “Convergent algorithm for dynamic traffic assignment,” *Transportation Research Record*, 1328, 69-80.

Janson, B. N. and Southworth, F. (1992) “Estimating departure times from traffic counts using dynamic assignment,” *Transportation Research B*, 26B 3-16.

Janson, B. N. and Robles, J. (1995) “Quasi-continuous dynamic traffic assignment model,” *Transportation Research Record*, 1493, pp. 199-206.

Jayakrishnan, R., Mahmassani, H. S. and Hu, T. Y. (1994) “An evaluation tool for advanced traffic information and management systems in urban networks,” *Transportation Research C*, 2C, 129-147.

Mahmassani, H. S. and Peeta, S. (1993) “Network performance under system optimal and user equilibrium dynamic assignment: Implications for advanced traveler information systems,” *Transportation Research Record*, 1408, 83-93.

Mahmassai, H. S. and Peeta, S. (1995) “System optimal dynamic assignment for electronic route guidance in a congested traffic network,” In N. H. Gartner and G.

Improta (eds.) *Urban Traffic Networks—Dynamic Flow Modeling and Control*, Springer-Verlag, pp. 3-37.

Massachusetts Institute of Technology (1996) *Development of a Deployable Real-Time Dynamic Traffic Assignment System, Task C Interim Report: Conceptual Design and Framework*, Intelligent Transportation Systems Program.

Merchant, D. K. and Nemhauser, G. L. (1978a) "A model and an algorithm for the dynamic traffic assignment problem," *Transportation Science*, 12, 183-199.

Merchant, D. K. and Nemhauser, G. L. (1978b) "Optimality conditions for a dynamic traffic assignment model," *Transportation Science*, 12, 200-207.

Ortúzar, J. D. and Willumsen, L. G. (1994) *Modelling Transport*, 2nd edition, New York: John Wiley.

Ran, B., Boyce, D. E. and LeBlanc, L. J. (1993) "A new class of instantaneous dynamic user-optimal traffic assignment models," *Operations Research*, 41, 192-202.

Ran, B. and Boyce, D. E. (1996) *Modeling Dynamic Transportation Networks: An Intelligent Transportation System Oriented Approach*, Second Revised Edition, Springer-Verlag.

Robles, J. and Janson, B. N. (1995) "Dynamic traffic modeling of the I-25/HOC corridor southeast of Denver," *Transportation Research Record*, 1516, 48-60.

Sheffi, Y. (1985) *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Englewood Cliffs, NJ: Prentice-Hall.

Wardrop, J. G. (1952) "Some theoretical aspects of road traffic research," *Proceedings of the Institution of Civil Engineers*, *Part II*, 1, 325-378.

Wie, B.-W. (1989) "An application of optimal control theory to dynamic user-equilibrium traffic assignment," *Transportation Research Record*, 1251, 66-73.

Wie, B.-W. and Friesz, T. L. and Tobin, R. L. (1990) "Dynamic user optimal traffic assignment on congested multidestination networks," *Transportation Research B*, 24B, 431-442.

Wie, B.-W., Tobin, R. and Friesz, T. (1994) "The augmented Lagrangian method for solving dynamic network traffic assignment models in discrete time," *Transportation Science*, 28, 204-220.

Wu, J. H., Chen, Y. and Florian, M. (1998) "The continuous dynamic network loading problem: A mathematical formulation and solution method," *Transportation Research B*, 32B, 173-187.

Xu, Y. W., Wu, J. H., Florian, M., Marcotte, P. amd Zhu, D. I. (1999) "Advances in the continuous dynamic network loading problem", *Transportation Science*, 33, 341-353.

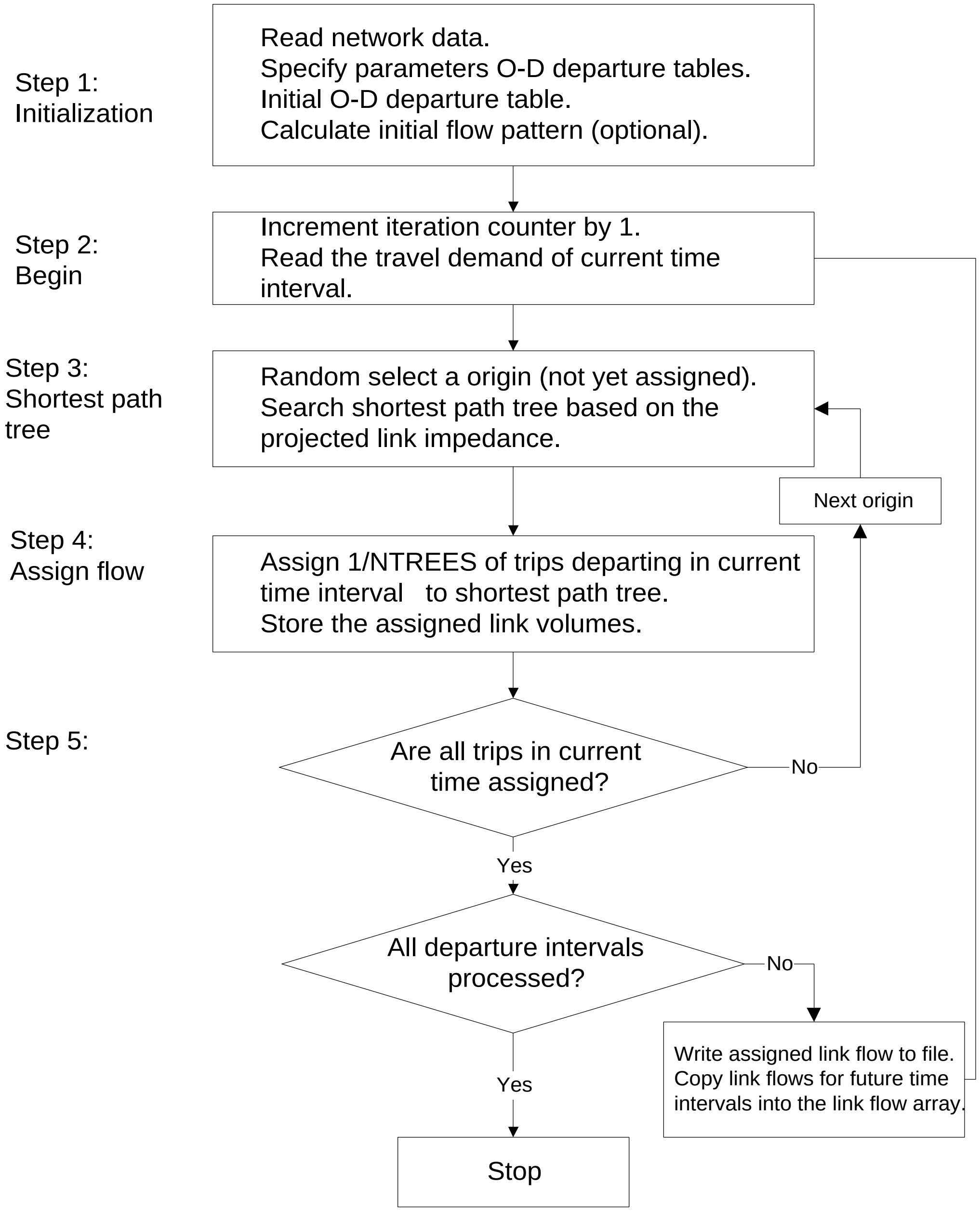

Figure 1 DTA computation procedure

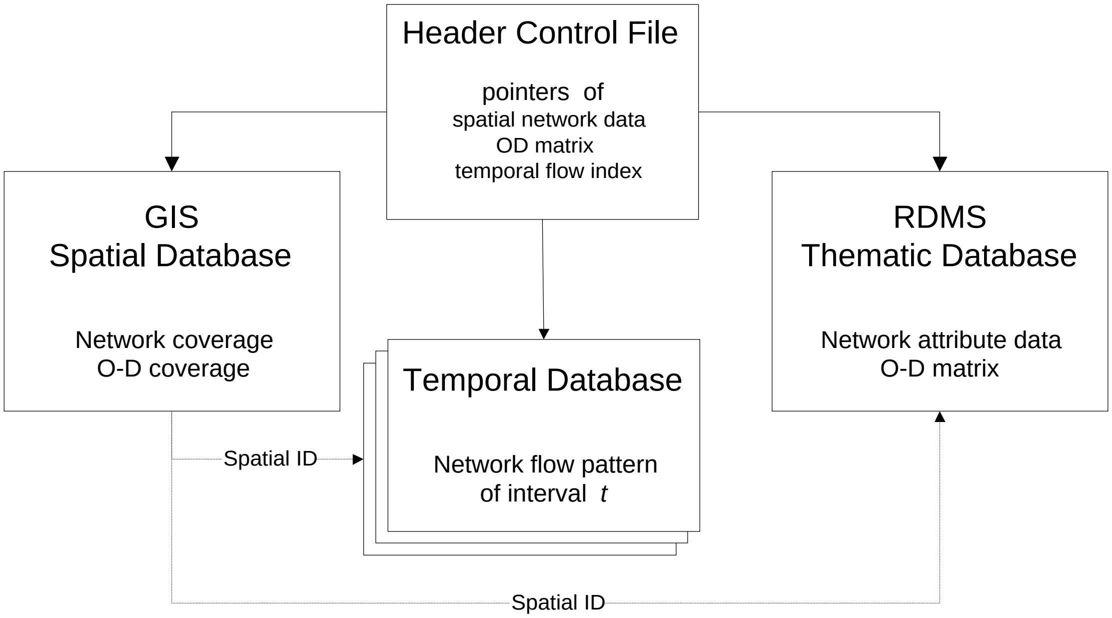

Figure 2 Conceptual database design

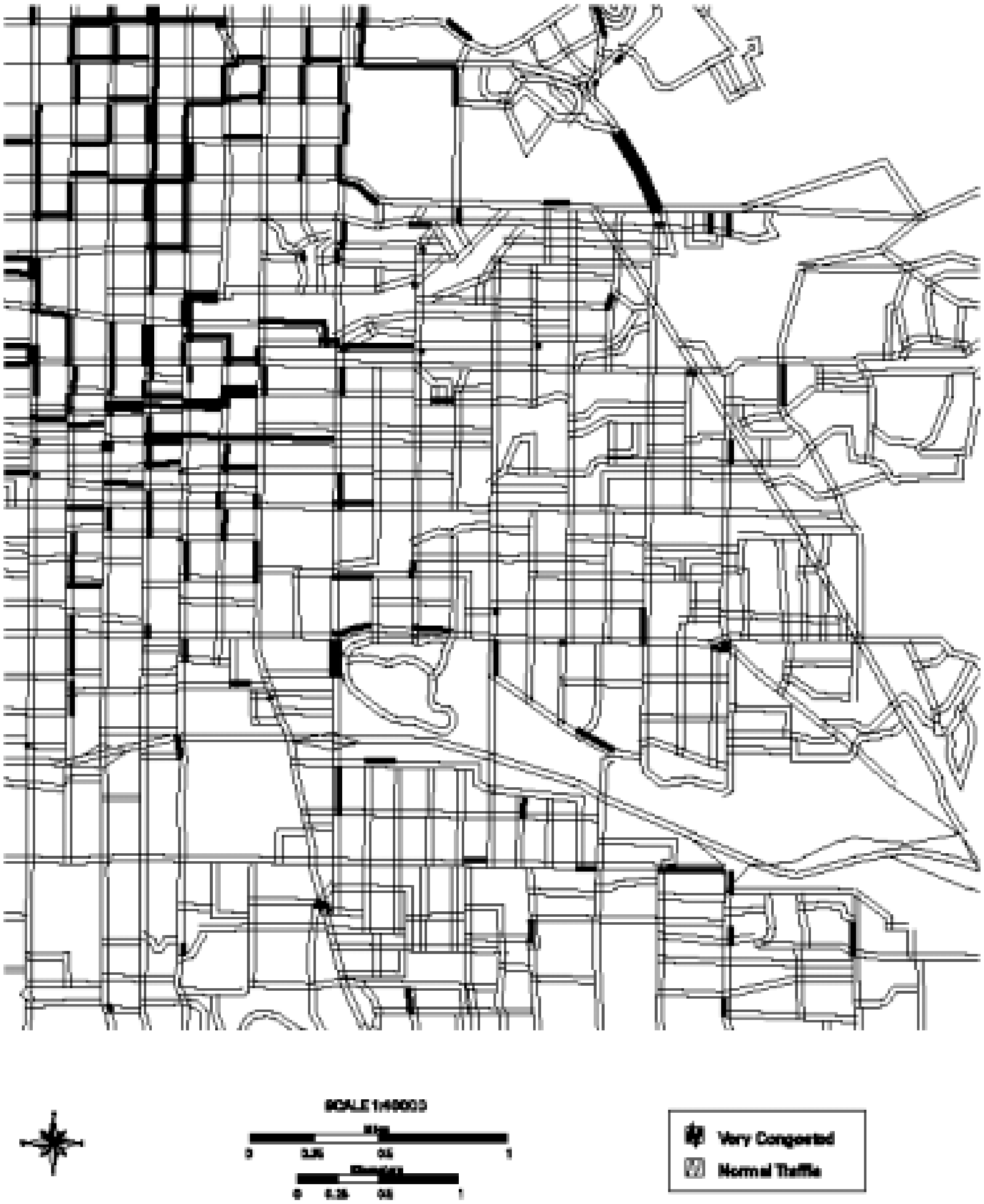

Figure 3: Salt Lake City network congestion pattern, time interval 1

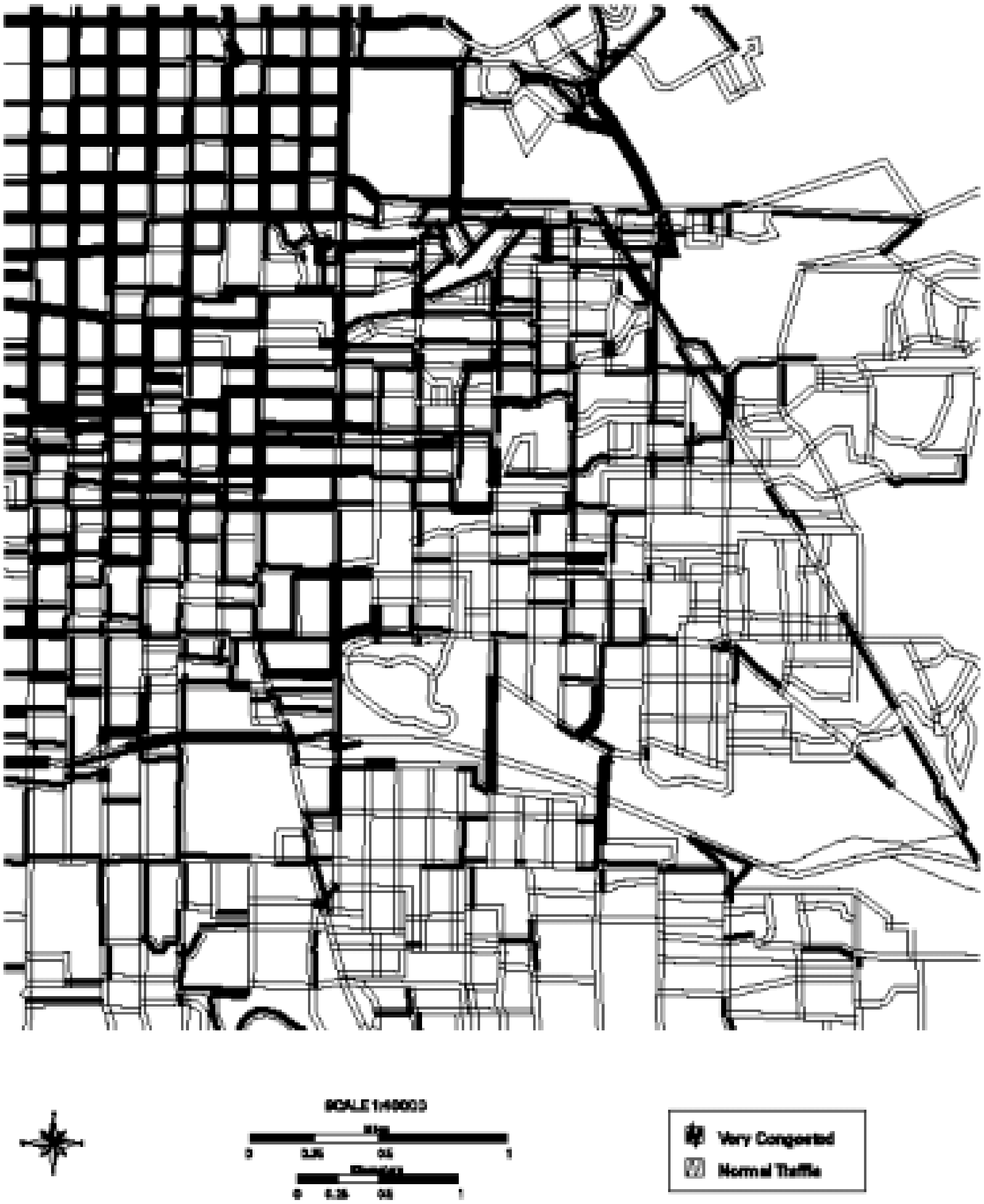

Figure 4: Salt Lake City network congestion pattern, time interval 20

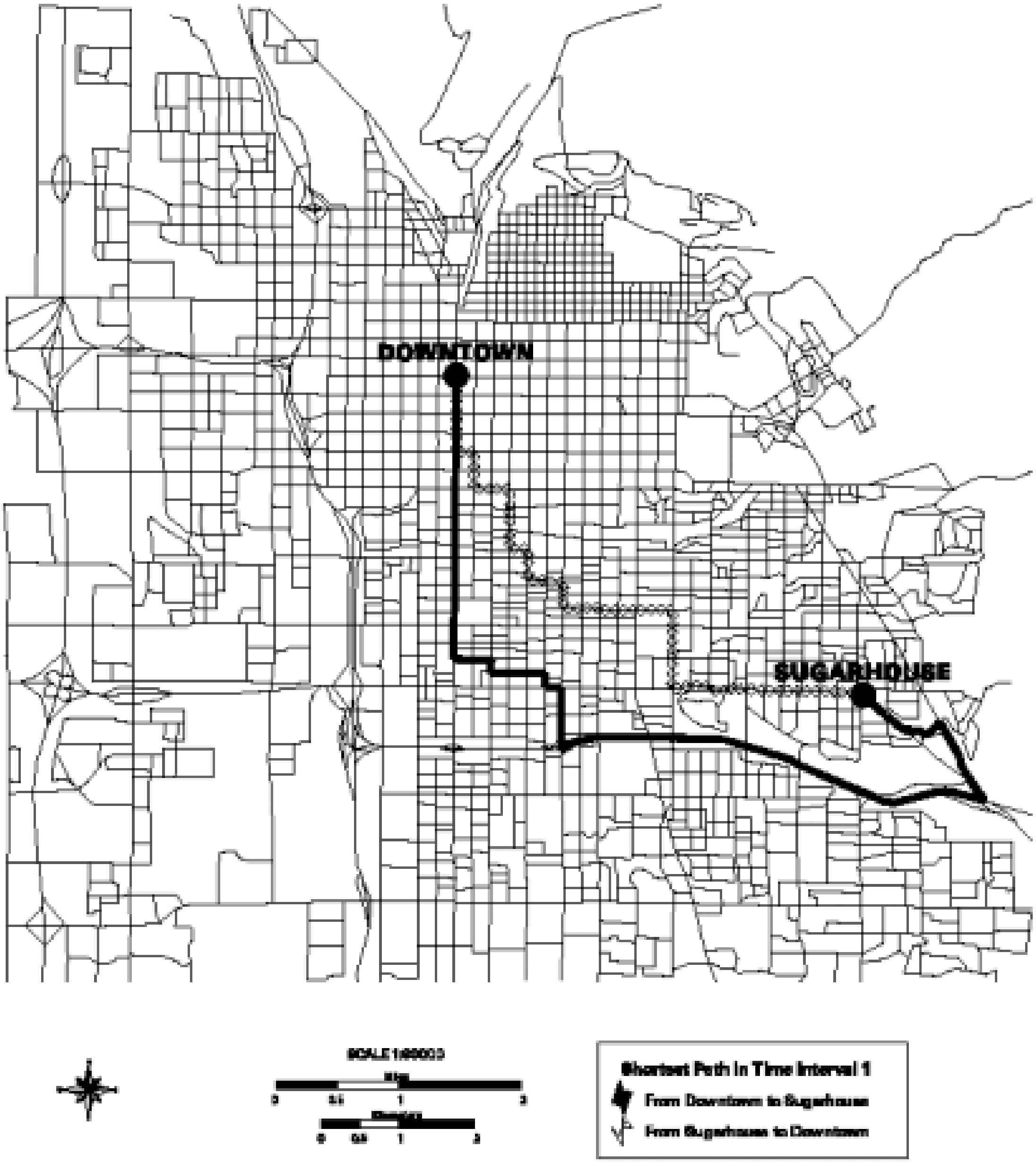

Figure 5: Minimum cost routes for opposite travel directions, time interval 1

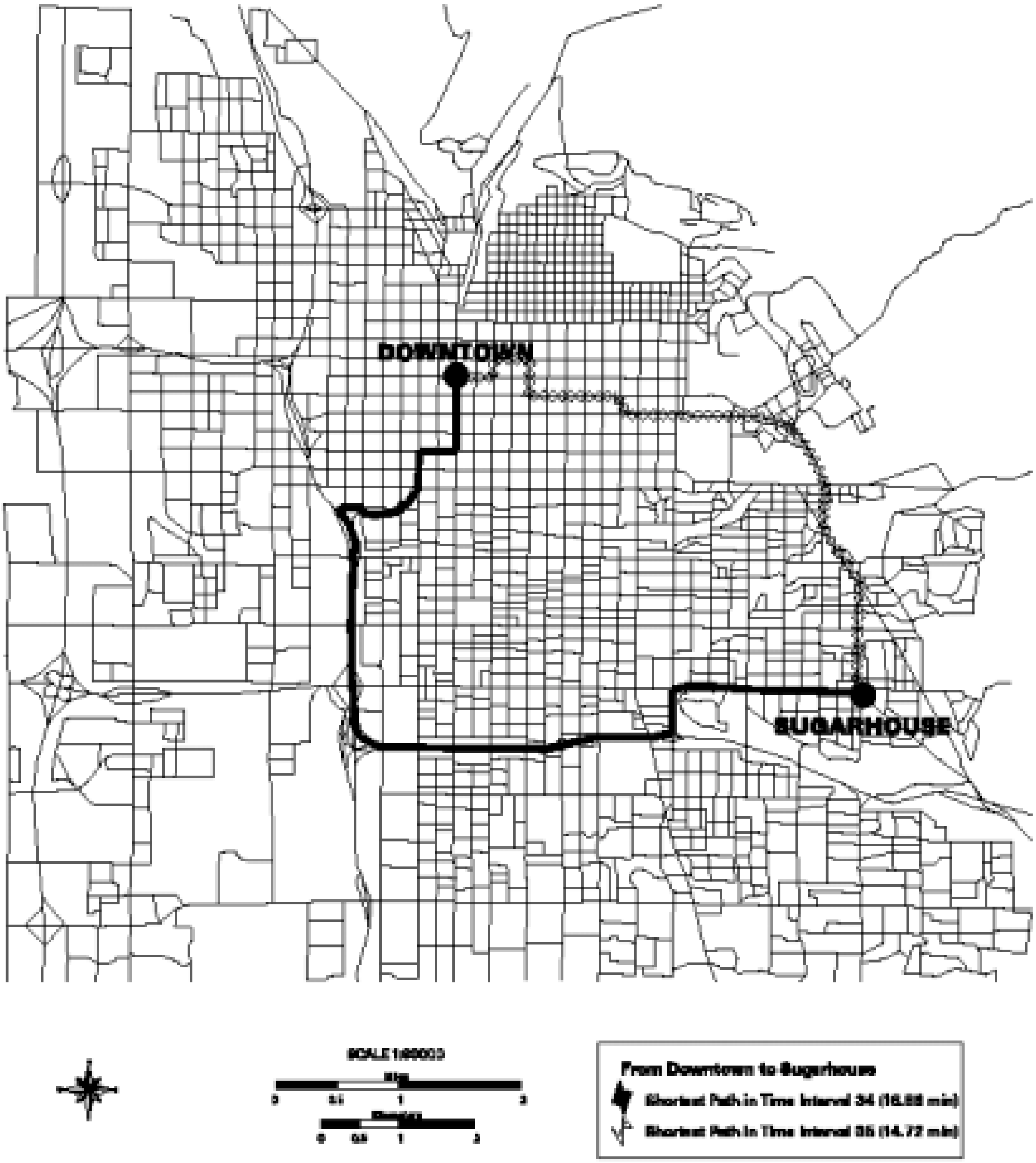

Figure 6: Minimum cost routes in successive time intervals