Providing Reliable Route Guidance: A Case Study Using Chicago Data

Yu (Marco) Nie * and Xing Wu
Department of Civil and Environmental Engineering
Northwestern University, 2145 Sheridan Road, Evanston, IL 60208
Peter Nelson and John Dillenburg
University of Illinois, Chicago

July 7, 2009

Abstract

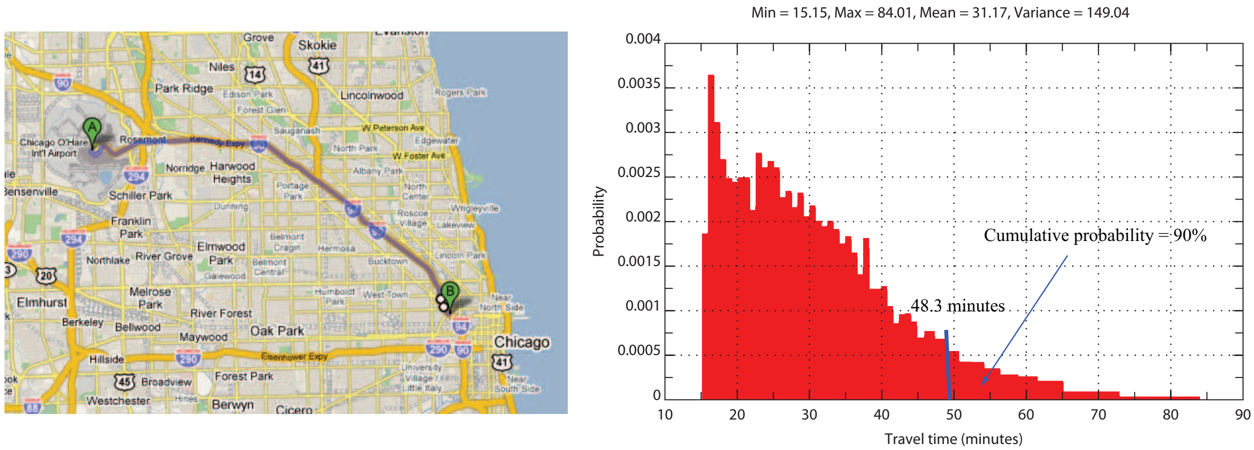
Reliable route guidance can be generated from solving the reliable *a priori* shortest path problem, which finds paths that maximize the probability of arriving on time. This paper aims to demonstrate the usefulness and feasibility of such route guidance using a case study. A hybrid discretization approach is first developed to improve the efficiency in computing convolution integral, which is an important and time-consuming component of the reliable routing algorithm. Methods to construct link travel time distributions are discussed and implemented with the data from the case study. Particularly, the travel time distributions on arterial streets are estimated from linear regression models calibrated from freeway data. Numerical experiments demonstrate that optimal paths are substantially affected by the reliability requirement in rush hours, and that reliable route guidance could generate up to 10 - 20 % of travel time savings. The study also verifies that existing algorithms can solve large-scale problems with modest computational resources.

Keywords: reliable *a priori* shortest path problem; route guidance; linear regression; case study; travel time distribution

1 Introduction

Motorists are becoming increasingly dependent on route guidance to plan unfamiliar trips. Just like a search engine can help internet users to locate useful information on web, a route guidance systems find best routes for motorists from a complex road system. As of today, many personal vehicles have built-in or adds-on GPS-based route guidance systems which can provide en-route guidance. Some of these equipments can even receive and make use of real-time traffic information. In the absence of such an in-vehicle unit, a priori driving directions (e.g. those provided by Internet-based map engines) are often used in trip planning.

*Corresponding author



(a) Interstate 94/90 from Chicago (Ohio St.) to Ohare International Airport (source: Google Map)

(b) Travel Time Distribution for that corridor during morning rush hour (6-10 AM)

Figure 1: An illustration of travel time variances in the Chicago area

Most existing route guidance systems assume that the road travel times are deterministic. When the stochastic nature of the system is acknowledged, selecting the route that is the fastest *on average* is often believed to be the right strategy. However, our day-to-day experience suggests just the opposite: not only is travel time random, but also a route with the least expected travel time is not always desirable due to, for instance, large variances. This is especially true in large metropolitan areas where random disruptions of various sorts consume a large portion of the total journey time. Figure 1 shows an example from the Chicago area. The right panel in the figure displays the empirical distribution of travel times observed in weekday morning rush hours on a stretch of freeway that connects Chicago downtown to O'hare International Airport (shown in the left panel), the second busiest airport in the US. Note that the travel times vary from as short as 15 minutes to as long as 80 minutes in this period of weekdays. In light of the magnitude of the variance, it is not surprising that the travel time estimated by existing rout guidance systems often turns out to be wildly inaccurate. Moreover, the figure also shows that if a traveler wish to arrive at the airport on time with a 90% chance, 48 minutes has to be budgeted for travel, which is more than 50% more than the mean travel time (31 minutes). Existing route guidance systems do not allow users to incorporate *reliability* into route choice. Nor are they able to inform users with reliability information of their recommended routes. Consequently, these systems essentially leave it to motorists to choose between running the risk of being late or budgeting a large buffer time, of which much is likely to be wasted. *Reliable route guidance* studied in this research addresses precisely these issues.

This paper studies reliable route guidance from an application point of view. The theoretical aspects of the reliable routing problem, including formulations and solution algorithms, have been developed in the literature (Frank 1969, Miller-Hooks 1997, Fan, Kalaba & Moore 2005*a*, Nie & Wu 2009*c*) and therefore are not the present focus. Nevertheless, a critical algorithmic issue, namely the discretization scheme used for evaluating convolution integrals, is further developed in this study. In particular, a hybrid approach is proposed which combines the advantages of the existing schemes (see Section 3 for details). Three key issues pertinent to application and deployment are addressed, using the Chicago metropolitan region as a case study. We first discuss how the necessary inputs, particularly the construction of empirical travel time distributions on both freeways and arterial streets, can be obtained from various sources of traffic data, such as loop detectors and toll transponders. Due to the lack of observations on arterial streets, linear regression models have to be used to estimate travel time distributions on them. The case study is focused on examining the benefits of reliable routing. We find that reliable route guidance could generate up to 10 - 20 % of travel time savings for motorists who travel during rush hours and seek high reliability. Another noteworthy finding is that highly reliable routes often tend to prefer major arterial to freeways and highways in rush hours. Last but not least, our experiments indicate that producing reliable route guidance is computationally viable even on very large regional networks, despite the fact the underlying optimization problem has a non-deterministic polynomial complexity.

For the remainder, Section 2 reviews the literature of reliability routing guidance. Section 3 presents the formulation and solution algorithms, and describes the newly developed hybrid discretization scheme. The case study is presented in Section 4, as well as the preparation of input data including travel time distribution. Section 5 reports and discusses experiment results, and Section 6 concludes the study with a summary of finding.

2 Literature review

Route guidance algorithms direct vehicles from an origin to a destination along a path that ~~are~~ considered "optimal" one way or another. Depending on whether or not the guidance is coordinated by a central control unit, the algorithms can be classified as "centralized" or "decentralized". They can also be labeled as "adaptive" or "a prior", according to whether or not en-route re-routing is allowed. Two other factors that are often used in classification are dynamics (i.e., if travel time varies over time of day) and uncertainties (i.e., if travel time is random). This research considers decentralized, *a prior* route guidance for stochastic and static networks [1]. The focus is to incorporate travel reliability as an integrated objective of route guidance. By *static*, we mean that the travel time distributions remain constant within each routing process. We have to restrict to the static case not because of methodological limitations [2], but rather due to data availability. The static label does not exclude, however, the possibility of changing travel time distributions according to time-of-day from one routing process to another. In the case study, reliable routes generated for morning rush hour are like to be different from those for evening peak period.

When uncertainty is concerned, "optimal" routing has many different meanings. A classic definition considers a routing strategy optimal if it incurs the least expected travel time (LET). The definition of optimality in this paper has to do with reliability, recognizing that a LET route (or policy) may be subject to high risks and therefore is not desirable to a *risk averse* traveler. ~~%subsectionLET problems~~

Reliability-based stochastic routing has been studied extensively, with the majority of the literature focused on *a priori* path problems. As to "optimality", different researchers give different definitions. In his seminal work, Frank (1969) defines the optimal path as the one that maximizes the probability of realizing a travel time equal to or less than a given threshold. An exact method is provided to compute the continuous probability distribution for the travel time on shortest paths. Mirchandani (1976) presents a recursive algorithm to solve a discrete version of Frank's problem. However, both methods are suitable only for small instances since they require

[1]Corresponding "adaptive" problems are related to "a prior" counterparts, and are usually simpler to solve.

[2]Note that both Miller-Hooks (1997) and Nie & Wu (2009*c*) deal with "dynamic" version of the problem.

enumerating all paths. Sigal, Alan, Pritsker & Solberg (1980) suggests using the probability of being the shortest path as an optimality index. Analytical formulas are given to evaluate such an index (assuming that all paths are enumerated) which involves the calculation of multiple integrals. The expected utility theory of von Neumann & Morgenstern (1967) has also been used to define path optimality. It has been shown (Loui 1983, Eiger, Mirchandani & Soroush 1985) that the Bellman's principle of optimality can be used to find maximum-utility paths when affine or exponential functions are used. For quadratic utility functions and/or special distributions that are uniquely determined by the first two moments, Loui (1983) showed that the maximum expected-utility problem is reduced to a class of bi-criteria shortest path problems that trade off the mean and variance of path travel times. These bi-criteria problems can be formulated using generalized dynamic programming (DP) (see, e.g., Carraway, Morin & Moskowitz 1990) based on the non-dominance relationship. More general nonlinear utility functions may be approximated by piecewise linear functions (see Murthy & Sarkar 1996, Murthy & Sarkar 1998). The mean-variance tradeoff can be treated in other ways. For example, Sivakumar & Batta (1994) adds an extra constraint into the shortest path problem to ensure that the identified LET paths have a variance smaller than a benchmark. In Sen, Pillai, Joshi & Rathi (2001), the objective function of stochastic routing becomes a parametric linear combination of mean and variance. In either case, DP cannot be applied. Instead, nonlinear or integer programming solution techniques must be used. Stochastic routing has also been discussed in the context of robust optimization, that is, a path is optimal if its worst-case travel time is the minimum (Yu & Yang 1998, Montemanni & Gambardella 2004). However, such robust routing problems are NP-hard even under restrictive assumptions (Yu & Yang 1998). Miller-Hooks & Mahmassani (1998*a*) defines the optimal path in a stochastic and time-varying network as the one that realizes the least possible travel time. Miller-Hooks (1997) and Miller-Hooks & Mahmassani (2003) explore other definitions of optimality based on first-order stochastic dominance (FSD) and *definite stochastic dominance*. Label-correcting algorithms are proposed (Miller-Hooks 1997, Miller-Hooks & Mahmassani 1998*c*, Miller-Hooks & Mahmassani 1998*b*) to find non-dominated paths under the stochastic dominance rules. Recognizing that the exact algorithm does not have a polynomial bound, heuristics are considered (Miller-Hooks 1997) which attempt to limit the size of the retained non-dominated paths by a predetermined number. As noted in Miller-Hooks (1997) (Chapter 5), however, these heuristics may not identify any non-dominated paths. Reliability has also been defined using the concept of connectivity (Chen, Bell, Wang & Bogenberger 2006, Kaparias, Bell, Chen & Bogenberger 2007). This approach models reliability as the probability that the travel time on a link is greater than a threshold. Accordingly, the reliability on a path is the product of link reliability (assuming independent distributions). A software tool known as ICNavS was developed based on this approach (Kaparias et al. 2007).

This research is built upon the prior work of Nie and Wu (2009*c*, 2009*b*, 2009*a*), which defines the objective of routing as maximizing the probability of arriving on-time and formulate the problem using general dynamic programming. This definition of optimality is identical to that of Frank (1969) and closely related to the concept of first-order stochastic dominance (e.g. Hadar & Russell 1969, Miller-Hooks & Mahmassani 2003). While this definition seems to intuitively address the perception of travel reliability, solving the resulting routing problem on real networks has been considered impractical because it requires path enumeration. Nevertheless, this research is built on the premises that recent advances in algorithmic development warrants a fresh look at this once "intractable" problem. We shall first present the formulation of the problem and its solution algorithms in what follows.

Table 1: Notation

s	the destination of routing
c_{ij}	travel times on link ij which is a random variable
$p_{ij}(\cdot)$	probability density function of c_{ij}
$\pi_{ij}(\cdot)$	cumulative distribution function (CDF) of c_{ij}
k^{rs}	path k that connects origin-destination path $r-s$.
K^{rs}	a set of all paths that connect r and s
$u_k^{rs}(b)$	the maximum probability of arriving at s through path k^{rs} on-time or earlier, departing from r with a time budget b.
$u^{rs}(b)$	the maximum probability of arriving at s through any path $k^{rs} \in K^{rs}$ on-time or earlier, departing from r with a time budget b.
Γ^{rs}	FSD-admissible paths between the OD pair rs
Ω^{rs}	FSD-optimal paths between the OD pair rs

3 Problem formulation and a solution algorithm

Consider a directed and connected network $G(\mathcal{N}, \mathcal{A}, \mathcal{P})$ consisting of a set of nodes $\mathcal{N}$ ($|\mathcal{N}| = n$), a set of links $\mathcal{A}$ ($|\mathcal{A}| = m$), and a probability distribution $\mathcal{P}$ describing the statistics of link travel times. Table 1 lists other notation to be used frequently. This paper does not consider the correlations among different c_{ij}, since these are difficult to establish from existing data. In addition, although p_{ij} does vary from one period of time to another, the time resolution of the data is only fine enough to allow a "static" routing decision for each period (although it is changed between periods). Technically, thus, the reliable shortest problem considered herein is a static version of those studied in (Nie & Wu 2009*c*) and (Nie & Wu 2009*b*).

The problem of providing reliable route guidance can be formulated as the so-called reliable a priori shortest path (RASP) problem. To present the formulation, we first need to define the first-order stochastic dominance(FSD) and the associated admissibility.

Definition 1 (First-order stochastic dominance (FSD) $\succ_1$) *Path k^{rs} dominates path l^{rs} in the first order, denoted as $k^{rs} \succ_1 l^{rs}$, if the CDF of π_k^{rs} never lies below that of π_l^{rs} and the inequality holds strictly at least at one point.*

Definition 2 (FSD-admissible path) *A path l^{rs} is FSD-admissible if there is no path k^{rs} such that $k^{rs} \succ_1 l^{rs}$.*

The RASP problem equals the problem of identifying all FSD-admissible paths between $(i, s), \forall i \neq s$ (Nie & Wu 2009*c*). However, it is possible that an FSD-admissible path is not shortest for any on-time arrival probability. To clarify this point, we define FSD optimality in the following.

Definition 3 (FSD-optimal path) *A path k^{rs} is FSD-optimal if 1) it is FSD-admissible and 2) it provides the highest on-time arrival probability from node r to node s for some time budget b.*

We shall denote the set of FSD-admissible and FSD-optimal paths between the OD pair (r, s) with Γ^{rs} and Ω^{rs}, respectively. Note that Ω^{rs} is the subset of Γ^{rs} by definition. At any node $i \in \mathcal{N}$, define $u^{is}(b) \equiv \max\{u_k^{is}(b), \forall k^{is} \in \Omega^{is}, \forall b\}$. The function $u^{is}(\cdot)$ is called *Pareto frontier function* at node i, which constitutes optimal solutions of the RASP problem.

The problem of finding all FSD-admissible paths can be solved using the following label correcting algorithm (see Nie & Wu (2009*c*) for the proof of convergence and a complexity analysis):

Algorithm FSD-LC

Step 0 Initialization. Let 0^{ss} be a dummy path from the destination to itself. Initialize the scan list $Q = \{0^{ss}\}$. set $\pi_0^{ss}(b) = 1, \forall b$.

Step 1 Select the first path from Q, denoted as l^{js}, and delete it from Q.

Step 2 For any predecessor node i of j, create a new path k^{is} by extending l^{js} along link ij.

step 2.1 Calculate the distribution of π_k^{is} from the distribution of π_l^{js} by convolution integral (Details are give below).

step 2.2 Compare the new path k^{is} with current *Pareto frontier*. If the frontier is dominated by k^{is}, update the frontier with the distribution of π_k^{is}, drop all existing FSD-admissible paths at node i, and set $\Gamma^{is} = \{k^{is}\}, \Omega^{is} = \{k^{is}\}$. Otherwise, further compare the distribution of the new paths and all existing FSD-admissible paths to check FSD admissibility. If any of the existing path dominates k_{is}, drop k_{is} and go back to Step 2; otherwise, delete all paths that are dominated by k_{is} from Γ^{is}, set $\Gamma^{is} \cup \{k_{is}\}$, and update $Q = Q \cup \{k^{is}\}$.

Step 3 If Q is empty, retrieve Ω^{is} from Γ^{is}, and evaluate the Pareto frontier function $u^{is}(\cdot), \forall i$, stop; otherwise go to Step 1.

If random link travel times follow a continuous probability density function p_{ij}, the distribution of path travel time π_k^{rs} can be calculated recursively from the following convolution integral

$$u_k^{is}(b) = \int_0^b u_k^{js}(b-w)p_{ij}(w)dw, \ \forall b \in [0, T] \tag{1}$$

where T is the largest possible travel time between any O-D pair. Typically, the convolution integral has to be evaluated using numerical methods which involves discretization. The simplest discretization scheme is to divide $[0, T]$ evenly into L intervals of length ϕ. The corresponding probability mass function P_{ij} reads

$$P_{ij}(b) = \begin{cases} \int_b^{b+\phi} p_{ij}(w)dw & b = 0, \phi, ..., (L-1)\phi \\ \int_b^{\infty} p_{ij}(w)dw & b = L\phi \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Accordingly, the evaluation of convolution integral in Equation (1) is replaced with a finite sum as follows:

$$u_k^{is}(b) = \sum_0^b u_k^{js}(b-\phi)P_{ij}(\phi), \ \forall b = 0, \phi, \cdots, L\phi \tag{3}$$

The above method has two shortcomings. First, the cumulative function has to be evaluated up to the predetermined upper bound $T = L\phi$ in the convolution. Note that T should equal the longest possible travel time between any O-D pair so that the computed distribution functions can cover the entire domain. However, it is hard to estimate T a priori, and to bypass the difficulty, T has to be set as an arbitrarily large number, which often turns out to be wasteful, and sometimes prohibitively expensive. Second, the uniformly spaced discrete points are not effective in representing heterogenous probability mass concentration. They frequently overly represent the flat portions, while at times fail to capture the hot spots where rapid changes take place. To address these issues, (Nie & Wu 2009*a*) suggested discretizing T such that each interval has the same probability mass $\epsilon < 1$, which is predetermined. The scheme discretizes the domain to $L = [1/\epsilon]_-$ intervals, where $[a]_-$ denotes the largest integer smaller than a. Since the domain is not uniformly discretized, however, the formula (3) is no longer applicable. Instead an alternative numerical method was proposed in (Nie & Wu 2009*a*) to perform the convolution. Although this method addresses the problem of undetermined T, it still poorly responds to the ireregular concentration of probability mass, essentially because identical probability mass is required in each interval.

In light of the above limitation, this study adopts a hybrid approach to allow both variable length of discrete intervals and different probability mass in each interval. The hybrid approach starts from a set of L uniform intervals (whose length may vary from one random variable to another), and computes the probability mass functions with Equation (2). However, a *consolidation* procedure is incorporated to merge consecutive intervals together such that no interval has a probability mass smaller than $1/L$. "Merging" two consecutive discrete intervals means removing the boundary between them and assigning the sum of their probability masses to the new interval. The consolidation produces a set of effective support intervals (ESI), whose size is often much smaller than L according to our experience. Consequently, the hybrid approach brings about significant computational benefits, without compromising the accuracy and stability of numerical convolution.

We now show how the convolution can be performed using the hybrid discretization scheme. Consider two random variables X and Y, whose domain is represented (after discretization and consolidation), respectively, by the following sets of break points: $[x_0, x_1, ..., x_{L(X)}]$, $[y_0, y_1, ..., y_{L(Y)}]$, where $L(X)$ and $L(Y)$ are numbers of ESI. Accordingly, the discrete support points are defined as

$$S^X = [\frac{x_0 + x_1}{2}, \cdots, \frac{x_{L(X)-1} + x_{L(X)}}{2}], S^Y = [\frac{y_0 + y_1}{2}, \cdots, \frac{y_{L(Y)-1} + y_{L(Y)}}{2}],$$

and the probability mass functions are

$$P^X = [P^X_1, \cdots, P^X_{L(X)}], P^T = [P^Y_1, \cdots, P^Y_{L(Y)}].$$

The following method can be used to compute P^Z for $Z = X \oplus Y$,$\oplus$ denotes convolution integral.

Convolution based on the hybrid discretization approach

Step 0 Set $z_{\min} = x_0 + y_0, z_{\max} = x_{L(X)} + Y_{L(Y)}$. Divide $[z_{\min}, z_{\max}]$ into L intervals of uniform length, and compute $\phi = (z_{\max} - z_{\min})/L$. Intialize $P^Z_l = 0, \forall l = 1, \cdots, L$.

Step 1 for $i = 1, 2, \cdots, L(X)$,

for $j = 1, 2, \cdots, L(Y)$

Calculate $ts = S^X_i + S^Y_j$ and $tp = P^X_i \times P^Y_j$. Define $l = \left[\frac{ts - z_{\min}}{\phi}\right]_-$, and set $P^Z_l = P^Z_l + tp$

end for

end for

Step 2 Consolidate P^Z to get effective support intervals and the associated probability mass functions.

Finally, we note that the determination of FSD-admissibility relies on comparing CDFs. In the hybrid discretzation scheme, the CDF of any consolidated distribution can be evaluated at any feasible point using linear interpolation. Thus, FSD-admissibility can be examined at any desired resolution and is not restricted by the discrete points used to represent distributions.

4 Input data for reliable route guidance

This section discusses issues associated with input data necessary to providing reliable route guidance. The focus is on how to construct link travel time distributions, which, to the best of our knowledge, are not readily available in most places. We first describe the case study and the available data sources.

Our case study considers the Chicago metropolitan area, which is the third largest metropolitan area in the US, and not coincidentally, one of the most congested cities too. According to the latest mobility report (Schrank & Lomax 2007), an average commuter in Chicago area wasted 46 hours due to traffic congestion in 2007. Perhaps more important, the travel time in the Chicago area seems more unreliable than any other major metropolitan areas in the US. The same mobility report indicates that an average commuter in the Chicago area has to budget 2.07 times of the free flow travel time for an important trip (which requires 95% probability of on-time arrival), the highest index in the country. On the other hand, Chicago has archived a rich set of traffic data in both public and private sectors. In particular, the GCM (Gary-Chicago-Milwaukee corridor) traveler information system (www.gcmtravel.com) broadcasts ~~and archives~~ real-time traffic data collected from loop detectors, toll transponders and other devices operated by departments of transportation in Illinois, Wisconsin and Indiana. This system also provides (point-to-point) real-time travel time for most toll roads in the region (known as I-PASS data).
This research uses GCM data~~base~~ as the the primary source for traffic data on freeways and toll roads. The GCM data come from two main sources: loop detectors and electronic toll transponders (IPASS), which cover freeways and toll roads respectively. Figure 2 shows a topology of the Chicago network used in the case study, as well as the location of loop detectors and IPASS toll booths. The network data are from the latest travel planning model prepared by the Chicago Metropolitan Agency for Planning (CMAP). As revealed in Figure 2, a major problem is the lack of data on arterial and local streets, which constitute the majority of links in the network. Recognizing that this is a universal problem that has not yet overcome in the current paradigm of data collection, we estimate travel time distributions on these streets, as to be detailed in the following.

4.1 Data for freeway and toll roads

In the GCM database, loop detectors record speed, occupancy and flow rate approximately every 5 minutes. Likewise, travel times on toll roads between two I-PASS toll booths are obtained from in-vehicle transponders, and subsequently aggregated and written into database every 5 minutes. About 825 loop detectors and 174 I-PASS detectors (each I-PASS detector corresponds to an origin-destination pair of toll booths) from GCM database are used in this study (see Figure 2). Specifically, the loop detector data collected from 2004 10/10 to 2008 10/11, and the I-PASS detector data from 2004 10/9 to 2008 7/3, are employed.

We first need to identify links that are "covered" by either I-PASS detector, loop detector, ~~or both~~. To determine which link in the CMAP network is associated with a loop detector, the

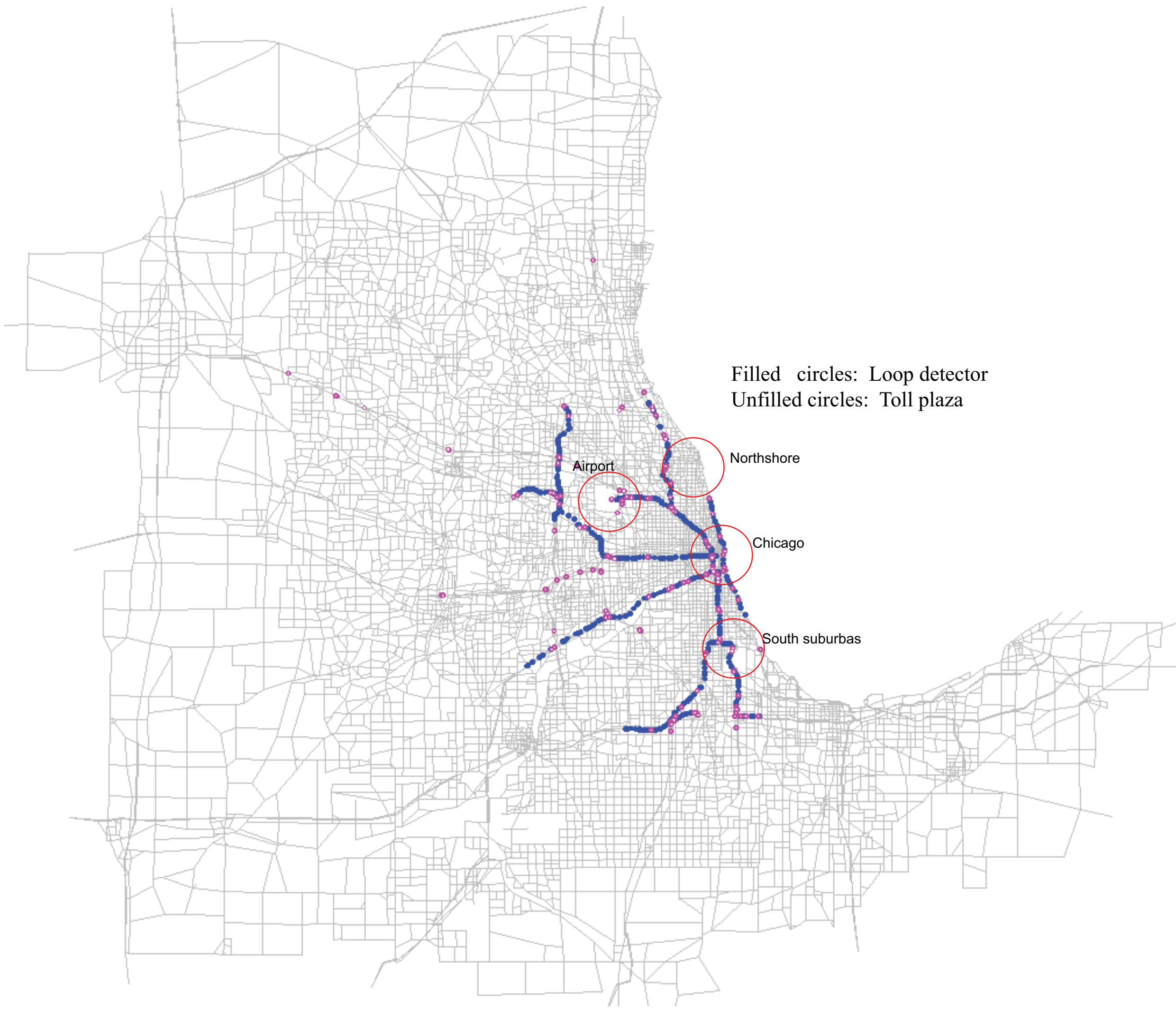


Figure 2: The Chicago network from Chicago Metropolitan Agency for Planning

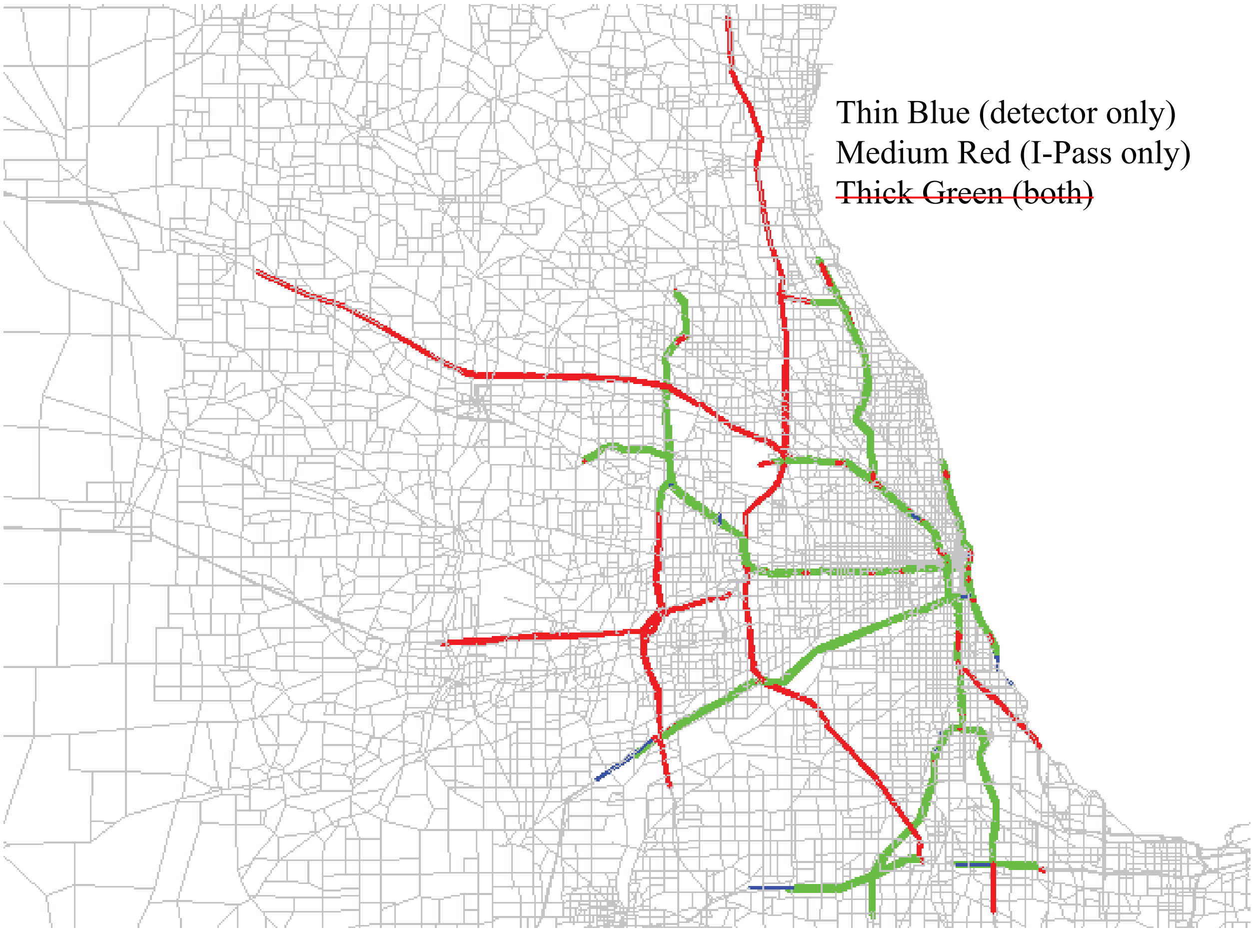


Figure 3: Covered links in the CMAP network (total covered links = 765)

coordinates (longitude and latitude) of the detector are used to find the closest freeway link. Finding the I-PASS covered links requires more work, which consists of two major steps. In the first, the starting and ending points of the I-PASS detector are located in the CMAP network. The second step identifies and marks all links used by the fastest (not shortest) path connecting the two points. Thus, vehicles that pass two I-PASS toll booths in sequence are assumed to always stay on the toll roads, which in most cases constitute the fastest alternative. Note that a link in the CMAP network may be covered by more than one loop detectors~~, or by both loop detectors and I-PASS detectors~~. In total, 765 out of 44331 links are covered in one way or another, as shown in Figure 3.

For links covered by loop detector(s), the recorded speed in a 5-minute interval is used to estimate link travel time for the corresponding interval, i.e.,

$$\tau_a^d(t) = l_a / v_a^d(t) \tag{4}$$

where $\tau_a^d(t)$ and $v_a^d(t)$ are travel time and speed on link a recorded by detector d for time interval t, and l_a is the link length. If for an interval, a link contains more than one recorded travel time, the arithmetic average of calculated travel time values is taken as the nominal link travel time, that is,

$$\tau_a^d(t) = \frac{\sum_{d \in D(a)} l_a / v_a^d(t)}{|D(a)|} \tag{5}$$

where $D(a)$ is the set of loop detectors associated with link a at a given time. As for I-PASS detectors, we need to estimate link travel times on covered link based on the recorded path travel times for a given time period. This is a difficult exercise for two reasons. First, how the travel delays (if there is any) experienced on a path may be *spatially* distributed is unknown. Second, an I-PASS record tagged by one time interval might contribute link travel times at other time intervals [3]. It is hard to solve either problem unless further information is available, such as supplementary loop detector data. For simplicity, we assume that path travel times are distributed to links proportional to their lengths, that is

$$\tau_a^i(t) = \frac{l_a}{\sum_{a \in k^{rs}} l_a} c_k^{rs}(t) \tag{6}$$

where k^{rs} denotes the shortest path connecting nodes r (the starting node of the link associated with the origin toll booth) and s (the ending node of the link associated with the destination toll booth), and $c_k^{rs}(t)$ is the recorded travel time on the path for time interval t. While this simplification would certainly introduce errors, we note that the magnitude of errors may be alleviated when multiple I-PASS records are available for the same stretch of toll roads. Equation (6) also implies that the travel time on a path at one time interval contributes to its covered links for the same interval. This shortcoming is not as serious as it sounds, since eventually the travel time data will be aggregated on a period of a coupe of hours. That is to say, as long as the misplaced link travel times do not go into a wrong period (which is certainly possible but is much less likely), they will not seriously distort the aggregated distributions. Finally, in light of the above problems, this study only uses I-PASS data where no loop detector data are available. To be precise, the travel time on link a at time t is calculated as follows

$$\tau_a(t) = \begin{cases} \tau_a^d(t) & \text{if loop data are available} \\ \tau_a^i(t) & \text{if } \cancel{\text{only}} \text{ I-PASS data are available} \end{cases} \tag{7}$$

Once link travel times are obtained, the empirical distributions can be constructed using the following procedure.

Construct Empirical Distribution

Step 1 Find $L_a = \min\{\tau_a(t), \forall t \in \Lambda\}, U_a = \min\{10 l_a / v_a^0, \max\{\tau_a(t), \forall t\}\}$, where Λ is a set of valid time intervals, and v_a^0 is free flow speed (or speed limit) on link a.

Step 2 Divide $[L_a, U_a]$ into M intervals, and let $\delta_a = (U_a - L_a)/M$. Find the set $D_m = \{\tau_a(t) | \forall t \in \Lambda, (m-1)\delta a \leq \tau_a(t) < m\delta\}, \forall m = 1, ..., M$

Step 3 Obtain the probability mass for each interval m using $P_m = \frac{|D_m|}{|\Lambda|}$.

It is noted that link travel time distribution may be affected by various factors, such as time-of-day and seasonal effects. Conceiably, one should consider a different reliable routing decision for rush hour and off-peak period. To address this issue, the travel time data are disaggregated

[3]Note that the time interval that identifers an I-PASS record must be tied to either the entry (origin) or exit (designation), since the travel times between most I-PASS toll booths are longer than 5 minutes.

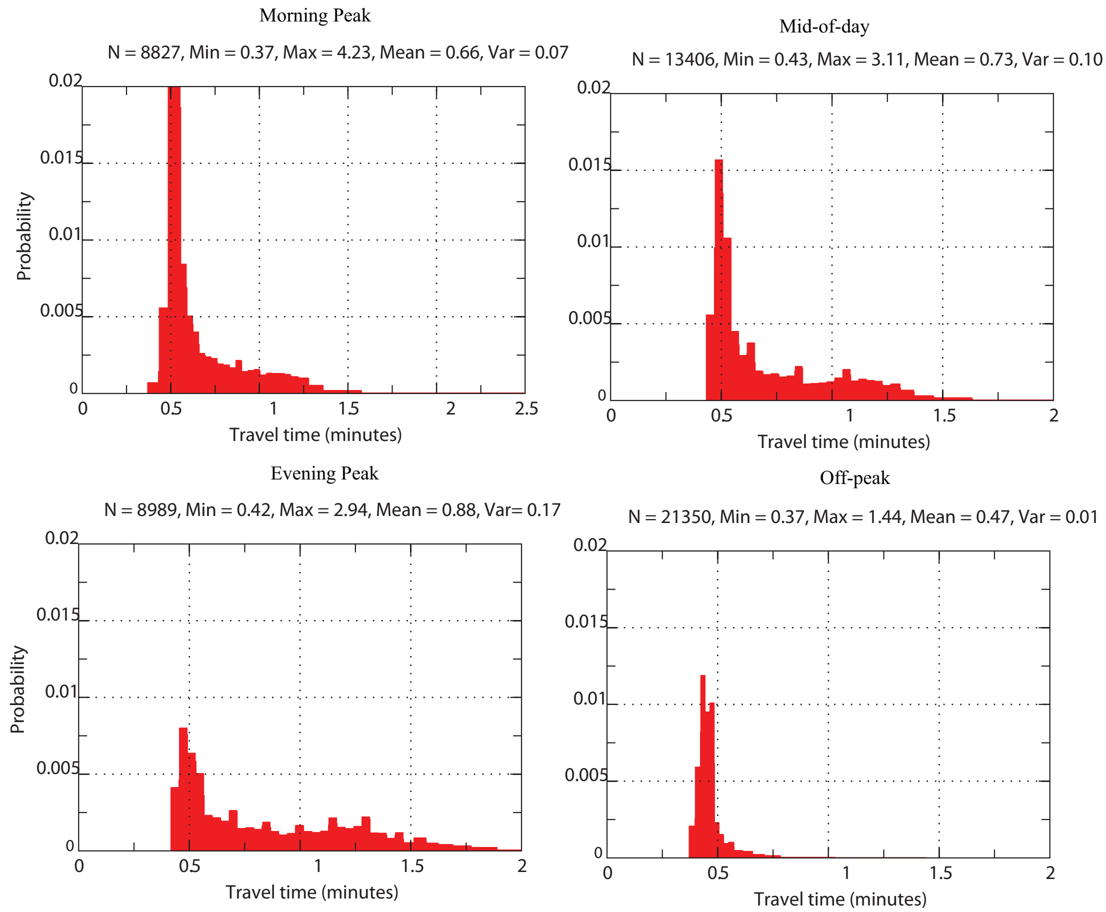


Figure 4: Comparison of link travel time distribution at different time-of-day periods (Link 19815, Spring, Weekday)

according to three key factors: time-of-day, day-of-week and season. Specifically, each day is divided into four periods, namely, morning peak period (6 am - 10 am), mid-of-day period (10 am - ~~15~~ pm), evening peak period (~~15~~ pm - ~~20~~ pm) and off-peak period (20 pm - 6 am). Days in a week are first grouped into weekends and weekdays. In addition, Friday, Saturday and Sunday form individual groups because they have somewhat different travel patterns. Finally, a year is grouped into Spring (months of March, April and May), Summer (months of June, July and August), Fall (months of September, October, November) and Winter (months of December, January, February). For each of the three factor, an additional group is added to address the case of no-segmentation. For instance, ~~T~~he segmentation for time-of-day contains 5 instead of 4 groups: morning peak, mid-of-day, evening peak, off-peak and whole-day (no segmentation for time-of-day). Therefore, in total, there are $5 \times 6 \times 5 = 150$ possible combinations. Accordingly, we generate 150 different distributions for all the 765 covered links.

A comparison of travel time distributions on a sample link for the four time-of-day periods is given in Figure 4. The results suggest that the most congested period is evening peak, followed by mid-of-day, morning peak and off-peak. Also, the variance of the distributions increases as the road becomes more congested. For example, the standard deviation for off-peak and evening peak is 0.1 and 0.4 minutes, respectively.

4.2 Data for arterial and local streets

No observations are available for arterial and local streets in the CMAP network. Consequently, the travel time distributions on these links have to be estimated indirectly. The estimation process involves two main steps: select an appropriate functional form, and estimate mean and variance.

Travel time on freeway and arterial is known to closely follow a Gamma distribution (e.g. Polus 1979). Figure 4 provides further confirmation. Gamma distributions have also been adopted in various studies of stochastic routing problems (e.g. Fan et al. 2005*a*, Nie & Wu 2009*c*). Therefore, this study adopts Gamma distribution to describe the travel time distribution on arterial and local streets. The probability density function of a Gamma distribution is

$$f(x) = \frac{1}{\theta^{\kappa}\Gamma(\kappa)}(x-\mu)^{\kappa-1}e^{-(x-\mu)/\theta}; x \geq \mu, \theta, \kappa \geq 0 \tag{8}$$

where θ is the *scale parameter*; κ is the *shape parameter*; μ is the *location parameter*; and $\Gamma(\cdot)$ is the Gamma function which takes the following form

$$\Gamma(z) = \int_0^{\infty} t^{z-1}e^{-t}dt$$

Having selected the functional form, we proceed to show how to estimate the three parameters required in a Gamma distribution. We first note that the mean and variance of a Gamma distribution are $\kappa\theta$ and $\kappa\theta^2$, respectively. Thus, if we know mean (denoted as u), variance (denoted as σ^2) and μ, then κ and θ can be obtained by

$$\theta = \frac{\sigma^2}{u-\mu}, \kappa = (\frac{u-\mu}{\sigma})^2 \tag{9}$$

The CMAP travel demand model (CMAP 2006) is used to estimate congested travel times on arterial streets. Note that the planning model is designed to capture average traffic conditions in the network for the designated time period on a typical weekday. The CMAP travel demand model represents a classical four-step process of *trip generation*, *trip distribution*, *mode choice*, and *traffic assignment*, with considerable modifications used to enhance the distribution and mode choice procedures. The original CMAP model divides a day into eight periods: off-peak (8 PM - 6 AM), pre-morning-peak (6-7 AM), morning-peak (7 - 9 AM), post-morning-peak (9-10 AM), mid-of-day (10 AM - 2 PM), pre-evening-peak (2 - 4 PM), evening-peak (4 - 6 PM), and post-evening-peak (6 - 8 PM). Note that in GCM data peak periods combine the pre and post periods defined in the CMAP model. For simplification, the assignment results for the peak periods (morning and evening) in the CMAP model are used to represent those from pre to post peak periods. Specifically, each link obtains from the CMAP model a mean travel time for each of the four predetermined GCM periods: morning peak, mid-of-day, evening peak and off-peak.

To estimate the mean (u), variance (σ^2) and the location parameter (μ) is less straightforward, since they are not readily available from the travel demand model. We postulate that the mean and variance of travel times on a link are related to its free flow travel time τ^0 and the level of congestion $\rho = \tau - \tau^0$, where τ is travel time from traffic assignment (note that the subscript a is suppressed for simplicity). This relationship may be estimated from freeway data using statistical models. The simplest linear regression model reads

$$u = a_1\tau^0 + b_1\rho + c_1 \tag{10}$$

$$\sigma = a_2\tau^0 + b_2\zeta\rho + c_2 \tag{11}$$

Table 2: Results of the linear regression for the mean-variance model (Equations 10-11) and the location model (Equation 12)

time-of-day	Variance Model				Mean Model				Location Model		
periods	a_1	b_1	c_1	R^2	a_2	b_2	c_2	R^2	a	b	R^2
AM PEAK	0.309	0.870	0.580	0.444	1.127	0.546	-2.056	0.910	0.843	-4.106	0.958
PM PEAK	0.368	0.685	2.967	0.400	1.143	0.563	0.336	0.872	0.860	-3.533	0.964
MIDDAY	0.283	1.076	2.040	0.346	1.100	0.630	-1.145	0.889	0.857	-3.608	0.956
OFF PEAK	0.178	0	-1.031	0.907	0.1778	0.0000	-0.017	0.516	0.831	-0.0876	0.937

where a_1, b_1, c_1, a_2, b_2 and c_2 are coefficients to be estimated from linear regression. ζ is a pre-determined parameter to account for the fact the the existence of signal control may increase variances. $\zeta = 1$ if no signal exists on the link; otherwise ζ is taken from a uniform distribution between $[1.1, 1.3]$. ζ is an artificial instrument. For all 765 links covered by GCM data, u and σ can be obtained from the empirical distribution and ρ is known from the CMAP travel demand model. Thus, a linear regression can be performed to determine the coefficients, which in turn are employed to estimate u and σ for arterial streets. We note that a linear model is needed for each of the four time-of-day periods.

A similar linear model can be constructed to estimate the location parameter μ, which delineates the smallest possible travel time on a link. We note that μ is likely to be smaller than τ_0 because motorists may drive well beyond the speed limit or the nominal "free-flow travel speed". Moreover, it seems reasonable to assume that the level of congestion does not affect μ. Thus, the linear model used to estimate μ is given by

$$\mu = a\tau^0 + b \tag{12}$$

The linear regression results for each of the four time-of-day periods are given in Table 2. As shown, the mean model and the location model (Equation 12) fit the data rather well (high R^2). However, the fitness of the variance model is not impressive. We tried, without much success, to introduce various forms of non-linearity into the model, such as using the variance (σ^2) instead of the stand deviation on the left hand side of Equation (11), or consider $(\tau - \tau_0)^2$ on the right hand side. Apparently, the travel time variances are affected by many other factors not included in the simple linear model. Finally, we note that the coefficient of the congestion index is near zero in both mean and variance models for off-peak period because congestion is negligible in that period. Figure 5 reports distributions on a sample link (ID=18548, located on the northbound Columbus Dr. in the Chicago downtown) The results show that the street is slightly more congested during the mid-of-day and evening peak periods.

5 Experimental results and discussions

Numerical experiments are presented in this section to show the usefulness of reliable route guidance and the feasibility of the existing algorithm in solving the real-size problems. For the first, we consider two real-world routing examples: one is between Chicago downtown and the O'hare international airport (ORD); another is from the northshore to Chicago south suburbs.

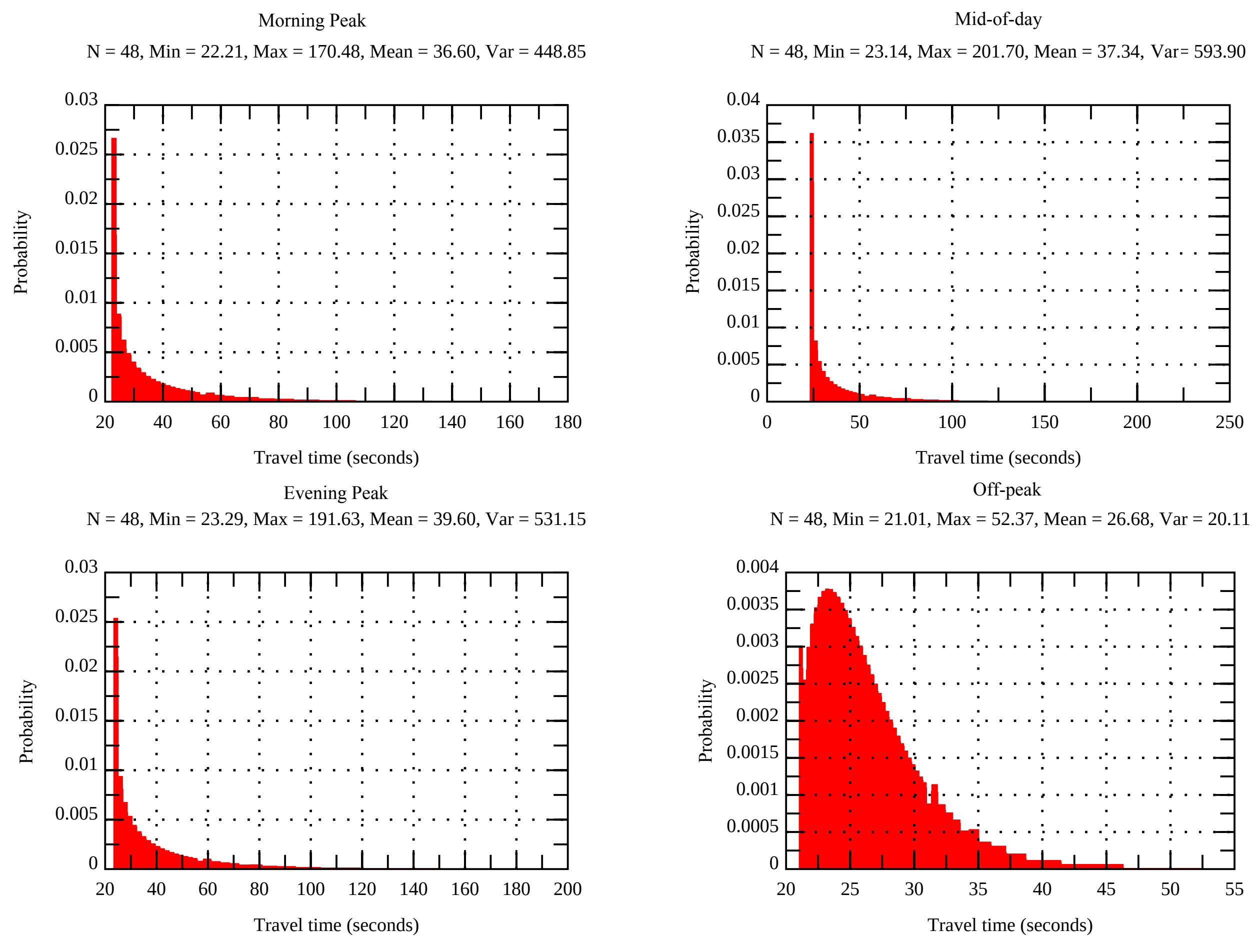
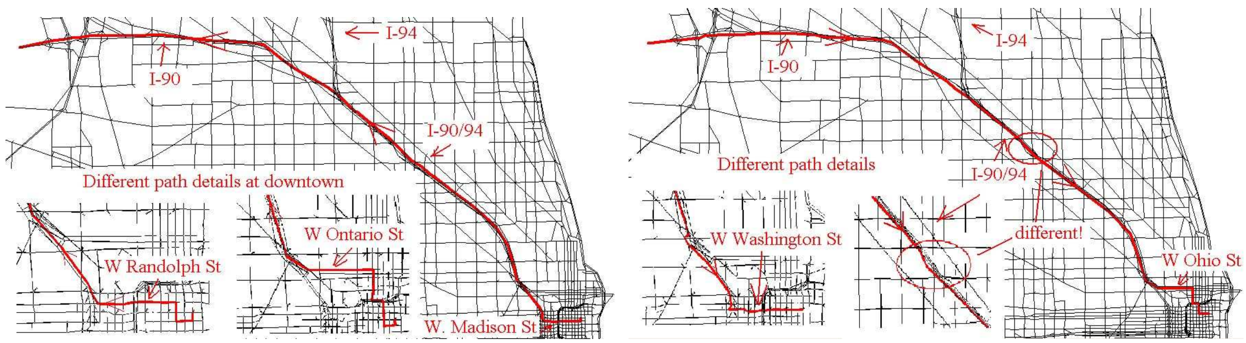


Figure 5: Comparison of distributions of link travel time at different time-of-day periods (Link 18548)



(a) Typical path from downtown to ORD

(b) Typical path from ORD to downtown

Figure 6: FSD-admissible paths between Chicago downtown and O'hare international airport (ORD) during the mid-of-day period of weekdays

Only the three non-off-peak periods in weekdays are considered. Finally, the algorithm FSD-LC was coded using MS-C++ and tested on a Windows XP x64 Workstation with two Xeno 3.0GHz CPUs and 8GB RAM.

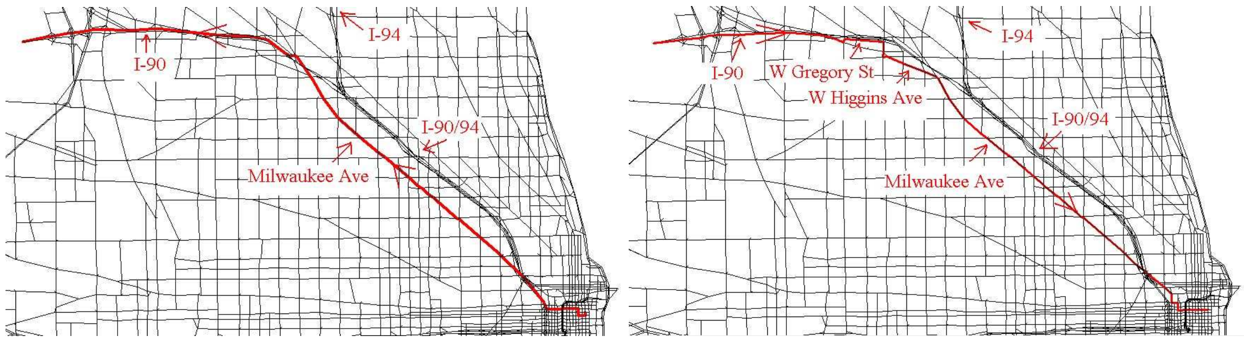
5.1 From Chicago downtown to O'hare international airport (ORD)

In this experiment, the intersection of Wabash St. and Washington St. is selected to represent the Chicago downtown, and the airport is represented by the end of I-190, the highway that serves the terminals. The most obvious choice for this routing problem, which is also suggested by both Google Map and Yahoo Maps, is to use freeway I-90/94 and I-90. Our results agree with this popular routing policy in general but have interesting discrepancies. For the mid-of-day period, most FSD-admissible paths always heavily use I-90 and I-90/94, and the differences between these paths are trivial, see Figure 6.

For the morning peak period, however, the reliable route guidance suggests that motorists should avoid I-90/94 and use an arterial street (N. Milwaukee Ave.) instead, if they wish to have an on-time arrival probability higher than 44% (to airport) or 62%(from airport) on-time arrival probability. To arrive the airport with 95% probability, for example, the path in Figure 7(a) requires a time budget of 33 minuets 57 seconds while the path mostly using I-90 and I-90/94 needs 37 minutes and 18 seconds. The reliable route guidance thus leads to a 10% saving in travel budget. In the evening peak of weekdays, motorists who drive from the airport to the city are recommended to avoid I-90 until they pass the diverge of I-90 and I-94 (see Figure 8(a)). For 95% on-time arrival probability, the path in Figure 8(a) requires a time budget of 34 minutes and 30 seconds, while the path in Figure 8(b) needs 39 minutes and 38 seconds. In both periods, our results suggest that avoiding the entire or part of the popular freeways will help motorists budget less time for better reliability.

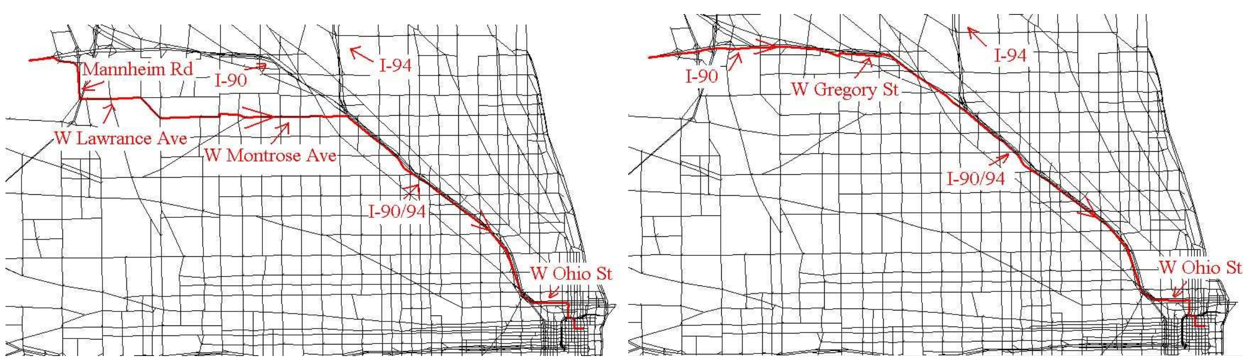
We note that the path given in Figure 8(b) is actually the best for 50% on-time arrival probability (i.e. the average performance). At this probability, a motorisit using the path only need to budget 34 minutes 17 seconds for travel. As a comparison, the more reliable path in Figure 8(a) needs a slightly higher budget (34 minutes 34 seconds) for 50% probability.

Figure 9 reports the CDFs of travel times on the FSD-admissible paths. Note that the CDFs during the mid-of-day period are very close to each other. A close look reveals that all these admissible paths heavily use I-90/94 and I-90, with negligible topological differences. On the



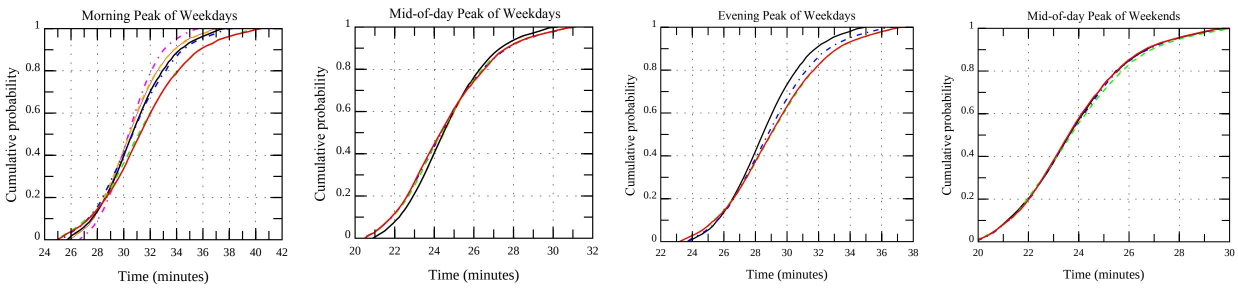
(a) from downtown to ORD (b) from ORD to downtown

Figure 7: Shortest paths between Chicago downtown and O'hare international airport (ORD) during the morning peak of weekdays, given the desired on-time arrival probability is 95%



(a) 95% on-time arrival probability (b) 50% on-time arrival probability

Figure 8: Shortest paths from Chicago O'hare international airport (ORD) to downtown during the evening peak of weekdays, given the desired on-time arrival probabilities are 95% and 50%, respectively



(a) (b) (c) (d)

Figure 9: Cumulative density functions (CDFs) of travel times of FSD-admissible paths for trips from Chicago downtown (Node 2057) to ORD (node 13051) during different time periods

other hands, the CDFs for the morning and evening peak periods are quite different, which are related to the large topological difference of the admissible paths.

5.2 From the northshore to Chicago south suburbs

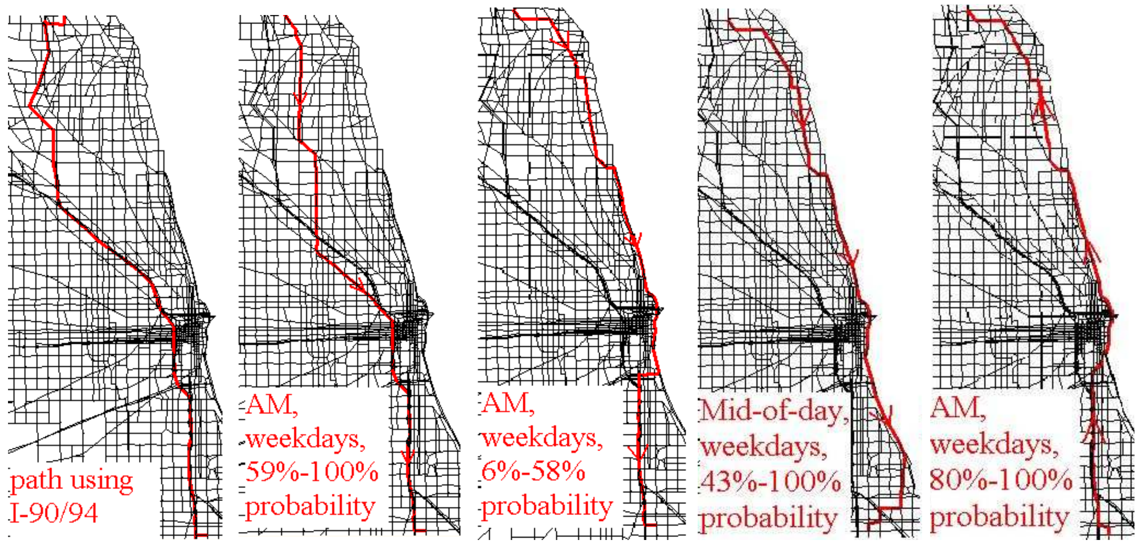
The challenge of this routing problem (see Figure 2) is how to travel through Chicago downtown and its peripheral area. Typically, motorists have two options: I-90/94 and Lake Shore Dr. The origin and the destination are deliberately selected so that both options could become attractive. Specifically, the intersection of Illinois Rd. and Locust Rd. on the northshore, and the intersection of E.79 St. and S Martin Luther King Dr. in the south suburbs are selected.

The reliable route guidance generally suggests that for weekdays motorists should stay away from I-90/94, especially for the portion north of Chicago downtown if driving from north to south. Our results indicate that the popular choice shown in Figure 10(a) is shortest only when the on-time arrival probability is very low (6% for morning peak, and 18% for mid-of-day). For the evening peak, this path is not even FSD-admissible. For the mid-of-day and the evening peak periods, Lake Shore Dr. is more reliable. Figure 10(d) shows that the shortest path during the mid-of-day period uses the Lake Shore Dr, which guarantees an on-time arrival probability higher than 43%. The FSD-admissible paths for the evening peak are similar and not reported separately. For the morning peak, however, Lake Shore Dr. is preferred only if a motorist wants to arrive on time with a probability lower than or equal to 59% (see Figure 10(c)). For higher reliability motorists need to use various arterial streets until they are close to downtown, and then switch I-90/94 (see Figures 10(b)).

Driving from south to north during weekdays is a different story. Experiments show that the path in Figure 10(a) (inverse direction) represents most of FSD-admissible path with the exception of the morning peak. For that period, Lake Shore Dr. is always recommended and any path using I-90/94 is not FSD-admissible.

5.3 Computational performance

We proceed in this section to examine the computational performance of the reliable routing algorithms. Table 3 reports the consumed CPU times for solving each of the two routing problems (in both directions). As shown, our algorithm solves the RASP problem within 30 seconds in most cases. Considering the complexity of the problem and the size of the network (15037 nodes and 44331 links), the performance is more than acceptable, even from a practical point of view. Note



(a) (b) (c) (d) (e)

Figure 10: Shortest paths from the north suburbs (node 4983) to south suburbs (node 12393) or in the inverse direction during some periods

that the CPU times reported in Table 3 were used to find FSD-admissible paths from all origins to the the specified destination. Therefore, once the computation is done, little effort is needed to obtain admissible paths for another origin to the same destination. Moreover, it may be possible to expedite the computation if only one O-D pair is of interest. This possibility, however, is not attempted in the current implementation. Another observation from the table is that the level of congestion seems to have significant impacts on the CPU time. It is relatively easier to solve the weekend models, likely because they are subject to less congestion and hence travel time variances. Also, the weekend routing models generate less admissible paths in general. In fact the AM and PM periods in weekends have only one admissible path in most cases, which makes sense since these periods are unlikely to be congested and therefore the most obvious choice usually prevails.

6 Summary

The overarching goal of this research is to demonstrate that incorporating reliability measures into routing decisions is both useful and feasible. To these ends, we present a proof-of-concept case study of reliable route guidance on a large regional network from the Chicago area.

Our experiments indicate that best paths do vary substantially with the reliability requirement, measured in this paper by the probability of arriving on-time or earlier. For motorists who travel during rush hours and seek high reliability, reliable route guidance could generate up to 10 - 20 % of travel time savings. Interestingly, highly reliable routes often tend to prefer major arterial to freeways and highways in rush hours. We recognize that this phenomenon could well have been caused by the underestimation of travel time variances on arterial streets. Recall that the distributions on arterial streets were estimated using linear models calibrated from freeway data, which were not fitted very well for the variances. Nevertheless, staying away from congested freeways during rush hours, particularly when you have important appointments, does

Table 3: Performance of the algorithm and results of the RASP problems for routes between Chicago downtown and ORD, and between Chicago north suburbs and south suburbs

	Weekdays			Weekends		
	AM Peak	Mid-of-day	PM Peak	AM Peak	Mid-of-day	PM Peak
From downtown (node 2057) to ORD (node 13051):						
CPU time	29.58	18.69	16.58	12.25	19.14	8.50
95%	33.95	28.27	32.85	19.32	27.76	29.37
50%	30.21	24.30	28.43	17.53	23.58	24.95
Same path?	Yes	No	Yes	Yes	Yes	Yes
# paths	7	5	4	1	5	1
From ORD (node 13051) to downtown (node 2057):						
CPU time	29.58	23.70	14.58	15.69	15.36	28.02
95%	34.50	30.70	39.82	20.00	27.05	28.32
50%	31.02	25.83	34.28	17.87	21.87	23.58
Same path?	No	Yes	No	Yes	No	No
# paths	6	2	2	1	2	4
From north suburbs (node 4983) to south suburbs (node 12393):						
CPU time	65.88	74.39	20.42	15.52	46.53	33.74
95%	50.37	48.97	54.15	32.43	46.02	47.17
50%	46.12	44.55	48.73	32.43	38.87	40.28
Same path?	No	No	Yes	No	Yes	No
# paths	7	10	2	2	1	4
From south suburbs (node 12393) to north suburbs (node 4983):						
CPU time	60.83	39.00	33.74	14.19	36.25	12.08
95%	50.50	46.46	49.85	33.58	43.55	41.70
50%	45.22	40.18	42.88	30.45	36.98	36.15
Same path?	No	No	Yes	No	Yes	No
# paths	10	6	6	1	3	1

Note: 1) CPU time is measured in seconds; 2) ”95%” and ”50%” are two on-time arrival probabilities respectively, and it gives the minimum travel time (in minutes) under 95% or 50% on-time arrival probability; 3) ”Same path?” is to show whether the shortest paths under 95% and 50% on-time arrival probability are exactly the same; 4) ”# paths” means the number of FSD-admissible paths sound a conventional wisdom. Such advice appeals to many motorists probably because freeways are less amenable to effective recourses when uncertainty strikes. The results from this case study may have provided a piece of empirical evidence to support this perception.

The paper presents and implements methods for constructing link travel time distributions, which are key inputs to providing reliable route guidance. Since these methods primarily rely on traffic and travel planning data that are available in many (if not most) large metropolitan areas in the US, they can be transferred to other regions upon further validation. A major problem tackled in this research is the data availability on arterial and local streets. Hitherto little traffic information has been archived for these streets, even on those equipped with intersection-related sensors. The method proposed in this research is indeed a compromise instead of a resolution. On the one hand, that the level of congestion affects travel time variances is widely noticed, and indeed supported by our data. Therefore, predicting variances from a congestion index seems a reasonable idea. On the other hand, it is likely that travel time variances depend on many other factors in a complex manner, and therefore the simple linear model may not work very well. Above all, using freeway data to project distributions on arterial streets may turn out to be ineffective, since these links are quite different in nature. It will be more desirable to calibrate the models using travel time data directly collected on arterial streets. Currently, the research team is seeking to use AVL (automatic vehicle location) data collected by buses. We will report those results in a subsequent work.

The study also verifies the capability of existing algorithms in solving large-scale reliable routing problems within reasonable amount of time. As noted, in most cases the algorithm found all-to-one reliable paths (FSD-admissible paths) within half a minute on an up-to-date workstation. Considering the non-deterministic polynomial complexity of the problem and the sheer size of the network, the reported performance is deemed satisfactory. To achieve real-time reliable route guidance (e.g., through an en-vehicle navigation system), further improvements are still needed. Possible strategies include but are not limited to: imposing higher-order stochastic dominance to reduce the number of non-dominant paths; adopting approximation methods; and exploiting the special properties of a one-to-one (instead of all-to-one) shortest path problem. We leave these further developments also to the future research.

Finally, it is worth noting that the methodologies presented herein can be used in applications other than reliable route guidance. They can help, for instance, transportation planning agency understand the level of service in heavily traveled corridors in terms of reliability, or construct aggregated travel reliability indexes for the entire or part of the highway network of interest.

Acknowledgement

We want to thank Mr. Kermit Wies from Chicago Metropolitan Agency for Planning, who kindly provides CMAP network and travel demand data used in this project.

References

Andreatta, G. & Romeo, L. (1988), 'Stochastic shortest paths with recourse', *Networks* **18**(3), 193–204.

Carraway, R. L., Morin, T. L. & Moskowitz, H. (1990), 'Generalized dynamic programming for multicriteria optimization', *European Journal of Operational Research* **44**(1), 95–104.

Chen, Y., Bell, M. G. H., Wang, D. & Bogenberger, K. (2006), 'Risk-averse time-dependent route guidance by constrained dynamic a* search in decentralized system architecture', *Transportation Research Record* **1944**, 51–57.

Cheung, R. K. (1998), 'Iterative methods for dynamic stochastic shortest path problems', *Naval Research Logistics* **45**(8), 769–789.

CMAP (2006), Travel demand modeling for the conformity process in northeastern illinois, available at http://www.cmap.illinois.gov /uploadedfiles/publications/other_publications /pm25_conformity_analysis_b.pdf, last accessed on 6/10/2009, Chicago Metropolitan Agency for Planning.

Croucher, J. (1978), 'A note on the stochastic shortest route problem', *Naval Research Logistics* **25**(4), 729–732.

Eiger, A., Mirchandani, P. B. & Soroush, H. (1985), 'Path preferences and optimal paths in probabilistic networks', *Transportation Science* **19**(1), 75–84.

Fan, Y., Kalaba, R. & Moore, J. (2005*a*), 'Arriving on time', *Journal of Optimization Theory and Applications* **127**(3), 497–513.

Fan, Y., Kalaba, R. & Moore, J. (2005*b*), 'Shortest paths in stochastic networks with correlated link costs', *Computers and Mathematics with Applications* **49**(9-10), 1549–1564.

Frank, H. (1969), 'Shortest paths in probabilistic graphs', *Operations Research* **17**(4), 583–599.

Fu, L. (2001), 'An adaptive routing algorithm for in-vehicle route guidance systems with real-time information', *Transportation Research Part B* **35**(8), 749–765.

Fu, L. & Rilett, L. R. (1998), 'Expected shortest paths in dynamic and stochastic traffic networks', *Transportation Research Part B* **32**(7), 499–516.

Gao, S. & Chabini, I. (2006), 'Optimal routing policy problems in stochastic time-dependent networks', *Transportation Research Part B* **40**(2), 93–122.

Hadar, J. & Russell, W. (1969), 'Rules for ordering uncertain prospects', *American Economic Review* **59**, 25–34.

Hall, R. W. (1986), 'The fastest path through a network with random time-dependent travel time', *Transportation Science* **20**(3), 182–188.

Hart, E. P., Nilsson, N. J. & Raphael, B. (1968), 'A formal basis for the heuristic determination of minimum cost paths', *IEEE Transactions on Systems, Science, and CyberneticsPart C: Applications and Reviews* **4**, 100–107.

Kaparias, I., Bell, M., Chen, Y. & Bogenberger, K. (2007), 'Icnavs: a tool for reliable dynamic route guidance', *IET Intellegent Transportation Systems* **1**(4), 225–253.

Loui, R. P. (1983), 'Optimal paths in graphs with stochastic or multidimensional weights', *Communications of the ACM* **26**(9), 670–676.

Miller-Hooks, E. (1997), Optimal Routing in Time-Varying, Stochastic Networks: Algorithms and Implementations, PhD thesis, Department of Civil Engineering, University of Texas at Austin.

Miller-Hooks, E. D. (2001), 'Adaptive least-expected time paths in stochastic, time-varying transportation and data networks', *Networks* **37**(1), 35–52.

Miller-Hooks, E. D. & Mahmassani, H. S. (1998*a*), 'Least possible time paths in stochastic, time-varying networks', *Computers and Operations Research* **25**(2), 1107–1125.

Miller-Hooks, E. D. & Mahmassani, H. S. (2000), 'Least expected time paths in stochastic, time-varying transportation networks', *Transportation Science* **34**(2), 198–215.

Miller-Hooks, E. D. & Mahmassani, H. S. (2003), 'Path comparisons for a priori and time-adaptive decisions in stochastic, time-varying networks', *European Journal of Operational Research* **146**(2), 67–82.

Miller-Hooks, E. & Mahmassani, H. (1998*b*), On the generation of nondominated paths in stochastic, time-varying networks, *in* 'Proceedings of TRISTAN III (Triennial Symposium on Transportation Analysis)', San Juan.

Miller-Hooks, E. & Mahmassani, H. (1998*c*), 'Optimal routing of hazardous materials in stochastic, time-varying transportation networks', *Transportation Research Record* **1645**, 143–151.

Mirchandani, P. B. (1976), 'Shortest distance and reliability of probabilistic networks', *Computers and Operations Research* **3**(4), 347–355.

Montemanni, R. & Gambardella, L. (2004), 'An exact algorithm for the robust shortest path problem with interval data', *Computers and Operations Research* **31**(10), 1667–1680.

Murthy, I. & Sarkar, S. (1996), 'A relaxation-based pruning technique for a class of stochastic shortest path problems', *Transportation Science* **30**(3), 220–236.

Murthy, I. & Sarkar, S. (1998), 'Stochastic shortest path problems with piecewise linear concave linear functions', *Management Science* **44**(11), 125–136.

Nie, Y. & Wu, X. (2009*a*), 'Implementation issues in approximate algorithms for reliable a priori shortest path problem', *Journal of the Transportation Research Board* **Forthcoming**.

Nie, Y. & Wu, X. (2009*b*), 'Reliable a priori shortest path problem with limited spatial and temporal dependencies', *In Proceedings of the 18th International Symposium on Transportation and Traffic Theory, accepted* .

Nie, Y. & Wu, X. (2009*c*), 'Shortest path problem considering on-time arrival probability', *Transportation Research Part B* **43**, 597–613.

Nie, Y., Wu, X. & de Mello, T. H. (2009), 'Optimal path problems with second-order stochastic dominance constraints', *Transportation Research Part B* **under review**.

Polus, A. (1979), 'A study of travel time and reliability on arterial routes', *Transportation* **8**(2), 141–151.

Polychronopoulos, G. H. & Tsitsiklis, J. N. (1996), ‘Stochastic shortest path problems with recourse’, *Networks* **27**(2), 133–143.

Provan, J. S. (2003), ‘A polynomial-time algorithm to find shortest paths with recourse’, *Networks* **41**(2), 115–125.

Schrank, D. & Lomax, T. (2007), 2007 urban mobility study, Technical report, Texas Transportation Institute, Texas A&M University, College Station, TX.

Sen, S., Pillai, R., Joshi, S. & Rathi, A. (2001), ‘A mean-variance model for route guidance in advanced traveler information systems’, *Transportation Science* **35**(1), 37–49.

Sigal, C. E., Alan, A., Pritsker, B. & Solberg, J. J. (1980), ‘The stochastic shortest route problem’, *Operations Research* **28**(5), 1122–1129.

Sivakumar, R. & Batta, R. (1994), ‘The variance-constrained shortest path problem’, *Transportation Science* **28**(4), 309–316.

von Neumann, J. & Morgenstern, O. (1967), *Theory of Games and Economic Behavior*, Wiley, New York.

Waller, S. T. & Ziliaskopoulos, A. K. (2002), ‘On the online shortest path problem with limited arc cost dependencies’, *Networks* **40**(4), 216–227.

Yu, G. & Yang, J. (1998), ‘On the robust shortest path problem’, *Computers and Operations Research* **25**(6), 457–468.