MICHIGAN OHIO UNIVERSITY TRANSPORTATION CENTER

Alternate energy and system mobility to stimulate economic development.

Report No: MIOH UTC SC2 p2-4 2010-Final
MDOT Report No: RC1545

ENABLING CONGESTION AVOIDANCE AND REDUCTION IN THE MICHIGAN-OHIO TRANSPORTATION NETWORK TO IMPROVE SUPPLY CHAIN EFFICIENCY: FREIGHT ATIS

FINAL REPORT

PROJECT TEAM

Dr. Ratna Babu Chinnam
Dr. Alper Murat
Industrial & Systems Engineering
Wayne State University
4815 Fourth Street
Detroit, Michigan 48202, USA

Dr. Gregory Ulferts
School of Business Administration
University of Detroit Mercy
4001 W. McNichols Rd
Detroit, Michigan 48221, USA

Report No: MIOH UTC SC2p2-4 2010-Final
SC2, Series, Projects 2, 3 and 4 October 2010
FINAL REPORT

Developed By: **Ratna Babu Chinnam, Assoc. Professor**
Principal Investigator, WSU
r_chinnam@wayne.edu
313-577-4846

Alper Murat, Asst. Professor
Investigator, WSU
amurat@wayne.edu
313-577-3872

With Contributions By: **Gregory Ulferts, Professor**
Co-Principal Investigator, UDM
ulfertgw@udmercy.edu
313-993-1219

SPONSORS

This is a Michigan Ohio University Transportation Center project supported by the U.S. Department of Transportation, the Michigan Department of Transportation, Wayne State University and the University of Detroit Mercy.

ACKNOWLEDGEMENT

The Project Team would like to acknowledge support from the Michigan Department of Transportation (MDOT), Ford MP&L, UPS and C.H. Robinson during the course of this research.

ABSTRACT

In just-in-time (JIT) manufacturing environments, on-time delivery is a key performance measure for dispatching and routing freight vehicles. Growing travel time delays and variability, attributable to increasing congestion in transportation networks, are greatly impacting the efficiency of JIT logistics operations. Recurrent and non-recurrent congestion are the two primary reasons for delivery delay and variability. Over 50 percent of all travel time delays are attributable to non-recurrent congestion sources such as incidents. Despite its importance, state-of-the-art dynamic routing algorithms assume away the effect of these incidents on travel time. In this study, we propose a stochastic dynamic programming formulation for dynamic routing of vehicles in non-stationary stochastic networks subject to both recurrent and non-recurrent congestion. We also propose alternative models to estimate incident induced delays that can be integrated with dynamic routing algorithms.

Proposed dynamic routing models exploit real-time traffic information regarding speeds and incidents from Intelligent Transportation System (ITS) sources to improve delivery performance. Results are very promising when the algorithms are tested in a simulated network of southeast Michigan freeways using historical data from the MITS Center and Traffic.com.

TABLE OF CONTENTS

LIST OF FIGURES

PAGE

LIST OF TABLES

PAGE

1. EXECUTIVE SUMMARY

The overall goal of this project is to develop effective static and dynamic routing algorithms for congestion avoidance and reduction for commercial cargo carriers given real-time information regarding recurring and non-recurring congestion by Advanced Traveler Information Systems (ATIS).

Just-in-time supply chains require reliable deliveries. However, travel times on road networks are unfortunately stochastic in nature. This randomness might stem from multiple sources. One of the most significant sources is the high volume of traffic due to commuting. This kind of traffic congestion is called *recurrent congestion* for it usually occurs at similar hours and days on a given network. The most used approach to deal with recurrent congestion is building 'buffer time' into the trip, e. starting the trip earlier to end the trip on time. However, these buffers significantly increase driver and equipment idle time (i.e., reduce utilization).

Intelligent Transportation Systems (ITS) that collect and provide real-time traffic data are now available in most urban areas and traffic monitoring systems are beginning to provide real-time information regarding incidents. In-vehicle communication technologies, both GPS and non-GPS based, are also enabling drivers' access to this information, facilitating vehicle routing and re-routing for congestion avoidance. We are proposed dynamic vehicle routing models that use ITS traffic information to avoid both recurrent and non-recurrent congestion in stochastic transportation networks.

We developed a dynamic vehicle routing model based on Markov decision process (MDP) formulation for the non-stationary stochastic shortest path problem. The state set of the MDP is based on the position of the vehicle, the time of the day, and the traffic congestion states of the roads. ITS data from southeast Michigan road network, collected in collaboration with M-DOT's Michigan Intelligent Transportation System Center (MITS) and Traffic.com, is used to illustrate the performance of the proposed models. Recurrent congestion states of the roads and their transition patterns are determined using historic and real-time traffic data from MITS Center. In particular, states are determined using Gaussian mixture model (GMM) based clustering. To address issues of 'curse of dimensionality' common to MDPs and the recognition that information from distant arcs are unreliable and less likely to influence 'optimal' path selection, we formulated the MDP state space such that only the roads/arcs that are in proximity to the vehicle affect local decisions.

Our dynamic routing models also account for non-recurring congestion stemming from incidents. Our incident models attempt to address two questions: 1) Estimate the affect of incident on travel time (incident-induced travel time delay) and 2) Estimate the incident clearance time (incident clearance time). We estimate incident-induced arc travel time delay using a decay function based on incident severity and duration parameters. Time required to clear the incident and restore the traffic is usually defined as incident clearance time and most of the delay due to incident is experienced during this period. We model the incident-clearance process using a Markov chain with an eventual absorbing state of incident clearance. Given that a road network may encounter both types of congestion concurrently, our dynamic routing models integrally account for both types of congestion and their interactions.

Our experiments clearly illustrate the superior performance of the SDP-derived dynamic routing policies over traditional static-path-based routing. The savings however depend on the network states as well as the time of day. The savings are higher during peak times and lower when traffic tends to be static (especially at nights). Experiments also show that explicit treatment of non-recurrent congestion stemming from incidents can yield significant savings.

2. ACTION PLAN FOR RESEARCH

We start with the data collection from multiple directions. On the *network structure* (network topology, design parameters, arc characteristics) side, we developed a network represents freeway and highways of the southeast Michigan. We test our dynamic routing decisions as well as incident (i.e. accidents, breakdowns) delay models on this network.

For southeast Michigan corridor ***arc velocity data***, we have collaborated with the MITS Center and signed a data-sharing agreement with Traffic.com. We have had multiple meetings with MITS Center to develop a better understanding for their traffic monitoring system (for southeast Michigan highways) and have also received data representing several months of traffic flow (such as velocity, occupancy) for the southeast Michigan highways. We have analyzed this data to improve the quality of the models being developed for dynamic vehicle routing decision support when operating with access to Advanced Traveler Information Systems (ATIS) information. For instance, through our analyses, we identified the need for representing each arc's congestion with a different number of states (and not force all arcs to be modeled with two states – i.e., congested and uncongested). Accordingly, we have refined the recurrent congestion state modeling by employing the Gaussian Mixture Model clustering method for automated detection of number of states and state velocity thresholds. In addition to MITS Center data, we now have access to Traffic.com's sensor database covering majority of highways in the southeast Michigan corridor. These datasets (and the networks resulting from them) are playing a critical role for evaluating and refining our dynamic routing algorithms. For ***incident data*** collection, we collaborated with the MITS Center. We received several months of incident data from Monroe Pendelton and Mark Burrows of MITS Center. This data set allowed us to initiate modeling of non-recurring congestion (incidents and special events) in routing applications. In addition to MITS Center data, Traffic.com also has an extensive archive of incident data which we are currently using to develop parametric incident delay models, models of particular interest to SEMCOG.

We have initially constructed a simple hypothetical road network simulator to build, test, and validate our algorithms in Matlab. This simulator allowed us to experiment with various network, velocity and incident scenarios. In the second year, we developed a southeast Michigan road network model that covers the sensors from both MITS Center and Traffic.com. We constructed these networks using archived historical traffic ITS data provided by our research partners, MITS Center and Traffic.com. One instance from this network encompasses main freeways and arterials extending from the intersection of I-94 and I-275 to the intersection of I-696 and I-75. In addition to network construction, we developed a data extraction and network configuration tool that allowed us to automate the loop sensor velocity and incident data extraction from the ITS databases. This tool takes in the origin-destination coordinates as inputs to identify and locate loop sensors.

Subsequently, this tool first extracts sensor velocity data and incident data and then configures routing models by determining such model inputs as arc travel time distributions by departure time of day. This tool encompasses data extraction, filtering, and cleaning procedures and is based on a Microsoft Access database with Matlab interface for efficient network configuration and algorithmic implementation.

We developed compact yet effective parametric incident duration and incident delay models. We extend and refine these models by calibrating according to the incident data obtained from the MITS Center and Traffic.com. In addition, we have also extended the algorithmic framework by incorporating more realistic "non-recurring congestion" modeling and exploitation logic into the algorithms.

Extensive evaluations of our SDP algorithms on hypothetical networks revealed significant reductions in trip completion times in comparison with deterministic algorithms and static stochastic algorithms that do not account for non-recurring congestion information. We first developed a parametric multiplicative incident model for the incident delay. This model accounts for the real-time traffic congestion, incident duration, incident severity, incident response. In the second year, we have further extended previous incident model by coupling the parametric delay model with an incident clearance Markov model. The incident clearance model is a non-stationary Markov chain model in which the incident clearance probability increases with the duration of the incident. Our incident model is integrated within the recurring congestion modeling and algorithmic framework.

We have developed the road-network model for the southeast Michigan region and identified some set of origin-destination pairs for major freight routes. On these routes, we have extracted the road-network recurring and non-recurring congestion data sets and calibrated these arcs accordingly. We implemented our static and dynamic models and algorithms in these major freight routes and compared the performance differences between typical baseline routing algorithms and our stochastic dynamic routing algorithms.

3. INTRODUCTION

Supply chains that rely on just-in-time (JIT) production and distribution require timely and reliable freight pickups and deliveries from the freight carriers in all stages of the supply chain. The requirements have even spread to the supply chains' service sectors with the adoption of cross docking, merge-in-transit, and e-fulfillment, especially in developed countries with keen concern in process improvement[1]. For example, in Osaka and Kobe, Japan, as early as 1997, 52 percent (by weight) of cargo deliveries and 45 percent of cargo pickups had designated time windows or specified arrival times [2]. These requirements have now become the norm in the U.S. as well. For example, many automotive final assembly plants in Southeast Michigan receive nearly 80 percent of all assembly parts on a JIT basis (involving five to six deliveries/day for each part with no more than three hours of inventory at the plant). However, road transportation networks are experiencing ever growing congestion, which greatly hinders all travel and certainly the freight delivery performance. The cost of this congestion is growing rapidly, reaching $78 billion by 2005 (from $20 billion in 1985) just in the U.S. large metropolitan areas alone [3]. This congestion is forcing logistics solution providers to add significant travel time buffers to improve on-time delivery performance, causing idle vehicles due to early arrivals.

Figure 1, for example, illustrates the magnitude of these buffers for 2003 in the automotive industry heavy Detroit Metro area, reaching over 70 percent during peak congestion periods of the day to achieve 95 percent on-time delivery performance [4]. Given that automotive plants are heavily relying on JIT deliveries, this is increasingly forcing the automotive original equipment manufacturers (OEMs) and others to carry increased levels of safety inventory to cope with the risk of late deliveries.

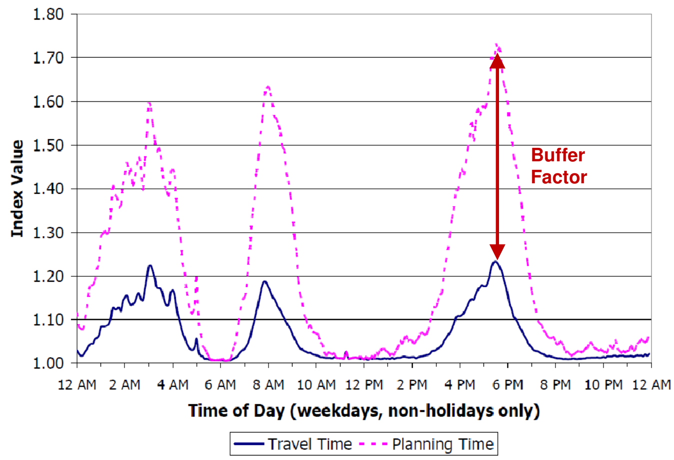

Figure 1. Extra Buffer Time Needed for On-Time Delivery with 95 Percent Confidence in Detroit [4]

The average trip travel time varies by the time of day. Travel time delays are mostly attributable to the so called ‘recurrent’ congestion that, for example, develops due to high volume of traffic seen during peak commuting hours. Incidents, such as accidents, vehicle breakdowns, bad weather, work zones, lane closures, special events, etc. are other important sources of traffic congestion. This type of congestion is labeled ‘non-recurrent’ congestion in that its location and severity is unpredictable. The Texas Transportation Institute [5] reports that over 50 percent of all travel time delays are attributable to the non-recurrent congestion. Despite its importance, current state-of-the-art dynamic routing algorithms assume away the effect of these incidents on travel time.

The standard approach to deal with congestion is to build additional ‘buffer time’ into the trip (i.e., starting the trip earlier so as to end the trip on time), as illustrated in Figure 1. Intelligent Traffic Systems (ITS), run by state agencies (e.g., the Michigan Intelligent Transportation Systems (MITS) Center in southeast Michigan) and/or the private sector (e.g., Traffic.com operating in many states), are providing real-time traffic data (e.g., lane speeds and volumes) in many urban areas. These traffic monitoring systems are also beginning to provide real-time information regarding traffic incidents and their severity. In-vehicle communication technologies, such as satellite navigation systems, are also enabling drivers access to this information en-route.

4. OBJECTIVE

The objective of our study is to develop methods for routing vehicles in stochastic road network environments representative of real-world conditions. In the literature, some aspects of this problem have been studied at some level but there does not exist any study that takes into account all aspects of our dynamic routing problem.

Specifically, our objectives are:

1) Methods for accurate and efficient representation of recurrent congestion, in particular, identification of multiple congestion states and their transition patterns.
2) Integrated modeling and treatment of recurrent and non-recurrent congestion for vehicle routing and demonstrating the need and value of such integration.

5. SCOPE

In this paper, we precisely consider JIT pickup/delivery service, and propose a dynamic vehicle routing model that exploits real-time ITS information to avoid both recurrent and non-recurrent congestion. We limit the scope to routing a vehicle from an origin point (say depot or warehouse) to a destination point.

Our problem setting is the non-stationary stochastic shortest path problem with both recurrent and non-recurrent congestion. We propose a dynamic vehicle routing model based on a Markov decision process (MDP) formulation. Stochastic dynamic programming is employed to derive the routing 'policy', as the static 'paths' are provably suboptimal for this problem. The MDP 'states' cover vehicle location, time of day, and network congestion state(s). Recurrent network congestion states and their transitions are estimated from the ITS historical data. The proposed framework employs Gaussian mixture model based clustering to identify the number of states and their transition rates, by time of day, for each arc of the traffic network. To prevent exponential growth of the state space, we also recommend limiting the network monitoring to a reasonable vicinity of the vehicle. As for non-recurrent congestion attributable to incidents, we estimate the incident-induced arc travel time delay using a stochastic queuing model.

6. LITERATURE SURVEY

In the classical deterministic shortest path (SP) problem, the cost of traversing an arc is deterministic and independent on the arrival time to the arc. The stochastic SP problem (S-SP) is a direct extension of this deterministic counterpart where the arc costs follow a known probability distribution. In S-SP, there are multiple potential objectives, and the two most common ones are the minimization of the total expected cost and maximization of the probability of being lowest cost [6]. To find the path with minimum total expected cost, Frank [7] suggested replacing arc costs with their expected values and subsequently solving as a deterministic SP. Loui [8] showed that this approach could lead to sub-optimal paths and proposed using utility functions instead of the expected arc costs. Eiger et al. [9] showed that Dijkstra's algorithm [10] can be used when the utility functions are linear or exponential.

Stochastic SP problems are referred as stochastic time-dependent shortest path problems (STD-SP) when arc costs are time-dependent. Hall [11] first studied the STD-SP problems and showed that the optimal solution has to be an 'adaptive decision policy' (ADP) rather than a single path.

In an ADP, the node to visit next depends on both the node and the time of arrival at that node, and therefore the standard SP algorithms cannot be used. Hall [11] employed the dynamic programming (DP) approach to derive the optimal policy. Bertsekas and Tsitsiklis [12] proved the existence of optimal policies for STD-SP. Later, Fu and Rilett [13] modified the method of Hall [11] for problems where arc costs as continuous random variables. They showed the computational intractability of the problem based on the mean-variance relationship between the travel time of a given path and the dynamic and stochastic travel times of the individual arcs.

They also proposed a heuristic in recognition of this intractability. Bander and White [14] modeled a heuristic search algorithm AO* for the problem and demonstrated significant computational advantages over DP, when there exists known strong lower bounds on the total expected travel cost between any node and the destination node. Fu [15] discussed real-time vehicle routing based on the estimation of immediate arc travel times and proposed a label-correcting algorithm as a treatment to the recursive relations in DP. Waller and Ziliaskopoulos [16] suggested polynomial algorithms to find optimal policies for stochastic shortest path problems with one-step arc and limited temporal dependencies. Gao and Chabini [17] designed an ADP algorithm and proposed efficient approximations to time and arc dependent stochastic networks. An alternative routing solution to the ADP is a single path satisfying an optimality criterion. For identifying paths with the least expected travel (LET) time, Miller-Hooks and Mahmassani [18] proposed a modified label-correcting algorithm. Miller-Hooks and Mahmassani [19] extends [18] by proposing algorithms that find the expected lower bound of LET paths and exact solutions by using hyperpaths.

All of the studies on STD-SP assume deterministic temporal dependence of arc costs, with the exception of Waller and Ziliaskopoulos [16] and Gao and Chabini [17]. In most urban transportation networks, however, the change in the cost of traversing an arc over-time is stochastic and there are very few studies addressing this issue. Most of these studies model this stochastic temporal dependence through Markov chain modeling and propose using the real-time information available through ITS systems for observing Markov states. In addition, all of these studies assume that recourse actions are possible such that the vehicle's path can be re-adjusted based on newly acquired congestion information. Accordingly, they identify optimal ADPs. Polychronopoulos and Tsitsiklis [20] is the first study to consider stochastic temporal dependence of arc costs and to suggest using online information en route. They considered an acyclic network where the cost of outgoing arcs of a node is a function of the environment state of that node and the state changes according to a Markovian process. They assumed that the arc’s state is learned only when the vehicle arrives at the source node and the state of nodes are independent. They also proposed a DP procedure to solve the problem. Polychronopoulos and Tsitsiklis [21] consider a problem when recourse is possible in a network with dependent undirected arcs and the arc costs are time independent. They proposed a DP algorithm to solve the problem and discussed some non-optimal but easily computable heuristics. Azaron and Kianfar [22] extended [20] by evolving the states of current node as well as its forward nodes with independent continuous-time semi-Markov processes for ship routing problem in a stochastic but time invariant network. Kim et al. [23] studied a similar problem as in [20] except that the information of all arcs are available real-time. They proposed a DP formulation where the state space includes states of all arcs, time, and the current node. They stated that the state space of the proposed formulation becomes quite large making the problem intractable.

They reported substantial cost savings from a computational study based on the southeast Michigan road network. To address the intractable state-space issue, Kim et al. [24] proposed state space reduction methods. A limitation of Kim et al. [24], is the modeling and partitioning of travel speeds for the determination of arc congestion states. They assume that the joint distribution of velocities from any two consecutive periods follows a single unimodal Gaussian distribution, which cannot adequately represent arc travel velocities for arcs that routinely experience multiple congestion states.

Moreover, they also employ a fixed velocity threshold *(50 mph)* for all arcs and for all times in partitioning the Gaussian distribution for estimation of state-transition probabilities (i.e., transitions between congested and uncongested states). As a result, the value of real-time information is compromised rendering the loss of performance of the dynamic routing policy. Our proposed approach addresses all of these limitations.

6.1. Non-recurrent Incidents and Incident Clearance

All of the shortest-path studies reviewed above consider stochastic arc costs that are mostly attributable to recurrent congestion. However, as stated earlier, over 50 percent of all traffic congestion is attributable to non-recurrent incidents and has to be accounted for dynamic routing. Incident-induced delay time estimation models are widely studied in the transportation literature. These models can be categorized into three groups based on their approaches: shockwave theory [25-27], queuing theory [28-33], and statistical (regression) models [34-36]. All of these modeling approaches have certain requirements such as loop-sensor data or assumptions regarding traffic/vehicle behavior. For instance, the shockwave theory based models require extensive loop sensor data for accurate positioning and progression of the shockwave. Both the queuing and shockwave theory based models require assumptions about the vehicle arrival process. Regression models, as empirical methods, cannot handle missing data without compromising on accuracy.

In all these three modeling methods, the delay due to incident is a function of incident duration. Thus, the correct estimation of incident duration is fundamental and there are various distributions suggested. Gaver [37] derived probability distributions of delay under flow stopping. Truck-involved incident duration is studied by Golob et al. [38] and employs lognormal distribution. Analysis of variance is examined by Giuliano [39] and a truncated regression model to estimate incident duration is proposed by Khattak et al. [40] for incident durations in Chicago area. Gamma and exponential distributions are also suggested as good representations of incident duration distribution [41]. Since the likelihood of ending an incident is related to how long it has lasted, hazard-based models are also suggested extensively. An overview of duration models applications is presented by Hensher and Mannering [42]. Nam and Mannering [43] applied hazard-based duration models to model distribution of detect/report, respond and clear durations of incidents. Using the empirical data of two years from the state of Washington, they showed that detect/report and respond times are Weibull distributed and the clearance duration is log-logistic distributed. Modeling incident delay in conjunction with vehicle routing is in its nascence. Ferris and Ruszczynski [44] present a problem in which arcs with incidents fail and become permanently unavailable. They model the problem as an infinite-horizon Markov decision process. Thomas and White [45] consider the incident clearance process and adopt the models in Kim et al. [23] for routing under non-recurrent congestion.

They model the incident delay using a multiplicative model and the incident clearance time as a non-stationary Markov chain, with transition probabilities following a Weibull distribution with an increasing instantaneous clearance rate. To model incident-induced delay, they multiply the incident arc's cost by a constant and time-invariant scalar. However, they do not account for recurrent congestion and assume arc costs are time-invariant and deterministic. In our approach, we address these limitations by joint consideration of recurrent and non-recurrent congestion as well as more appropriate representation of incident-induced delay and clearance.

7. METHODOLOGY: MODELING RECURRENT AND NON-RECURRENT CONGESTION

7.1. Recurrent Congestion Modeling

Let the graph $G=(N,A)$ denote the road network where N is the set of nodes (intersections) and $A \subseteq N \times N$ is the set of directed arcs between nodes. For every node pair, $n', n \in N$, there exists an arc $a \equiv (n,n') \in A$, if and only if, there is a road that permits traffic flow from node n to n'. Given an origin-destination (OD) node pair, the trip planner's problem is to decide which arc to choose at each decision node such that the expected total trip travel time is minimized. We denote the origin and destination nodes with n_0 and n_d, respectively. We formulate this problem as a finite horizon Markov decision process (MDP), where the travel time on each arc follows a non-stationary stochastic process.

An arc, $a \equiv (n,n') \in A$ is labeled as observed if its real-time traffic data (e.g., velocity) is available through the traffic information system. An observed arc's traffic congestion can be in $r+1 \in \mathbb{Z}^+$ different states at time t. These states represent arc's congestion level and are associated with the real-time traffic velocity on the arc. We begin with discussing how to determine an arc's congestion state given the real-time velocity information and defer the discussion on estimation of the congestion state parameters to Section 5. Let $c_a^{i-1}(t)$ and $c_a^i(t)$ for *i=1,2,...,r+1* denote the cutoff velocities used to determine the state of arc a given the velocity at time t on arc a, $v_a(t)$. We further define $s_a^i(t)$ as the i^{th} traffic congestion state of arc a at time t, i.e. $s_a^1(t)=\{1\}$ and $s_a^r(t)=\{\text{Congested at level r}\}=\{r\}$. For instance, if there are two congestion levels (e.g., *r+1=2*), then there will be one congested state and the other will be uncongested state, i.e., $s_a^0(t)=\{\text{Uncongested}\}=\{0\}$ and $s_a^1(t)=\{\text{Congested}\}=\{1\}$. Congestion state, $s_a^i(t)$ of the arc a at time t can then be determined as:

$$s_a(t)=\left\{i, \text{if } c_a^{i-1}(t) \le v_a(t) < c_a^i(t)\right\} \tag{1}$$

We assume the congestion state of an arc evolves according to a non-stationary Markov chain and the travel time is normally distributed at each state. In a network with all arcs observed, $S(t)$ denotes the traffic congestion state vector for the entire network, i.e., $S(t)=\{s_1(t),s_2(t),...,s_{|A|}(t)\}$ at time *t*. For presentation clarity, we will suppress (*t*) in the notation whenever time reference is obvious from the expression. Let the state realization of $S(t)$ be denoted by $s(t)$.

It is assumed that arc traffic congestion states are independent from each other and have the single-stage Markovian property. In order to estimate the state transitions for each arc, two consecutive periods' velocities are modeled jointly. Accordingly, the time-dependent single-period state transition probability from state $i=s_a(t)$ to state $j=s_a(t+1)$ is denoted with $P\{s_a(t+1)=j \mid s_a(t)=i\}=\alpha_a^{ij}(t)$. The transition probability for arc a, $\alpha_a^{ij}(t)$, is estimated from the joint velocity distribution as follows:

$$\alpha_a^{ij}(t)=\frac{\left|c_a^{i-1}(t)\le V_a(t)<c_a^i(t)\cap c_a^{j-1}(t+1)<V_a(t+1)<c_a^j(t+1)\right|}{\left|c_a^{i-1}(t)\le V_a(t)<c_a^i(t)\right|} \tag{2}$$

Let $T_a(t,t+1)$ denote the matrix of state transition probabilities from time t to time $t+1$, then we have $T_a(t,t+1)=\left[\alpha_a^{ij}(t)\right]_{ij}$. We further assume that arc *a*'s congestion state is independent of other arcs' states, i.e. $P\{s_a(t+1)\mid s_{a'}(t+1),s_a(t)\}=P\{s_a(t+1)\mid s_a(t)\}=\alpha_a^{ij}(t)$ for $\forall a'\in A$. Note that the single-stage Markovian assumption is not restrictive for our approach as we could extend our methods to the multi-stage case by expanding the state space [46]. Let network be in state $S(t)$ at time *t* and we want to find the probability of the network state $S(t+\delta)$, where δ is a positive integer number. Given the independence assumption of arcs' congestion states, this can be formulated as follows:

$$P\left(S(t+\delta)\mid S(t)\right)=\prod_{a=1}^{|A|}P\left(s_a(t+\delta)\mid s_a(t)\right) \tag{3}$$

Then the congestion state transition probability matrix for each arc in δ periods can be found by the Kolmogorov's equation [47]:

$$T_a(t,t+\delta)=\left[\alpha_a^{ij}(t)\right]_{ij}\times\left[\alpha_a^{ij}(t+1)\right]_{ij}\times...\times\left[\alpha_a^{ij}(t+\delta)\right]_{ij} \tag{4}$$

With the normal distribution assumption of velocities, the time to travel on an arc can be modeled as a non-stationary normal distribution. We further assume that the arc's travel time depends on the congestion state of the arc at the time of departure (equivalent to the arrival time whenever there is no waiting). It can be determined according to the corresponding normal distribution:

$$\delta(t,a,s_a)\sim N\left(\mu(t,a,s_a),\sigma^2(t,a,s_a)\right) \tag{5}$$

where $\delta(t,a,s_a)$ is the travel time on arc a at time t with congestion state $s_a(t)$; $\mu(t,a,s_a)$ and $\sigma(t,a,s_a)$ are the mean and standard deviation of the travel time on arc a at time t with congestion state $s_a(t)$. For the clarity of notation, we hereafter suppress the arc label from the parameter space wherever it is obvious, i.e. $\delta(t,a,s_a)$ will be referred as $\delta_a(t,s)$.

We assume that objective of dynamic routing is to minimize the expected travel time based on the real-time information. The nodes (intersections) of the network represent decision points where a routing decision can be made. Since our algorithm is also applicable for a network with incidents, in the next section we present our incident modeling approach, and then integrate the recurrent congestion and incident models.

7.2. Incident Modeling

In this section, we develop incident models which measure the incident clearance time and the delay experienced as a result of incident. In section 4, we integrate recurrent congestion and incident models with the dynamic routing model.

7.2.1 Estimating Incident Duration

The incident duration is defined as the total of detection/reporting, response, and clearance times. Due to the nature of most incident response mechanisms, the longer the incident has not been cleared, the more likely that it will be cleared in the next period. For example, the probability of an incident being cleared in the 15th minute, given that it has lasted 14 minutes, is greater than the probability of it being cleared in the 14th minute given that it has lasted 13 minutes. This is because it is more likely that someone has already reported the incident and an incident response team is either on the way or has already responded. Let t be the time to clear the incident. Then, we have the increasing hazard rate property, e.g., $\lambda(t+1)> \lambda(t)$, where $\lambda(t)=f(t)/(1-F(t))$ is the hazard rate of incident clearance in duration t, and $f(t)$ and $F(t)$ are the density and cumulative density functions of the clearance duration, respectively. We choose the Weibull distribution with increasing hazard rate to model the incident clearance duration.

Whenever there is an incident on an arc in the network, we assume that its starting time (t_{inc}^0), current status (i.e. cleared/not cleared), expected duration (μ), and standard deviation (σ) are available through ITS incident management and incident database systems. Hence, we can estimate the parameters of the Weibull distribution ($\varphi(a,b)$) of the incident clearance duration [47]. Furthermore, if an incident occurs en route, we may simply re-optimize the routing policy by assuming that the new origin node is the node that the driver is at or arrives next.

7.2.2 Estimating Incident-Induced Delay

Our incident delay model is based on [29]. Here incident-induced delay function, $\Theta(\cdot)$, is based on the incident duration φ, road nonincident capacity denoted with c (vehicle per hour, or *vph* in short), road capacity during the incident denoted with $\rho(vph)$ and arrival rate of vehicles to the incident arc denoted with $q(vph)$. Given these parameters for an incident started at t_{inc}^{0}, the vehicle arriving to the incident arc at time (t) experiences the following expected incident-induced delay:

$$E\left(\Theta(.)\right)=\left(\frac{c-\rho}{c}\right)\left(D_{12}-D_1P_3\right)+P_2d_m \tag{6}$$

where $D_{12}=\int_{D_1}^{D_2}x.\varphi(x)\,dx$, $D_1=\left(\frac{c-\rho}{c-q}\right)\left(t-t_{inc}^{0}\right)$, $D_2=\left(\frac{q}{\rho}\right)\left(t-t_{inc}^{0}\right)$, $d_m=\left(\frac{q-\rho}{\rho}\right)\left(t-t_{inc}^{0}\right)$, $P_1=\int_{0}^{D_2}x.\varphi(x)\,dx$, $P_2=\int_{D_2}^{\infty}x.\varphi(x)\,dx$, and $P_3=1-\left(P_1+P_2\right)$.

In order to track the amount of time that each arc has spent in the incident state, we define an incident duration vector defined over all the arcs, $I(t)$, i.e. $I(t)=\left\{i_1(t),i_2(t),...,i_{|A|}(t)\right\}$. Note that if an arc a is not an incident arc, then $i_a(t)$=0, otherwise $i_a(t)=t-t_{inc}^{0}(a)$ and $0<i_a(t)<\infty$, where $t_{inc}^{0}(a)$ is the incident onset time on arc a. For presentation clarity, we will hereafter omit the arc reference from the incident onset time, i.e. $t_{inc}^{0}=t_{inc}^{0}(a)$, whenever incident arc reference is obvious.

The incident delay model is an additive model, in that, $\Theta(\cdot)$ represents the delay time by which the arc travel time under same conditions (congestion state and the time) will be increased by a duration amounting to the incident induced delay. Specifically, given the arc travel time without the incident, $\delta_a(t,s,i=0)$, and the incident parameters, (φ,c,ρ,q,i), we can express the arc travel time with incident as:

$$\delta_a(t,s,i)=\delta_a(t,s,i=0)+\Theta_a\left(\varphi,c,\rho,q,i=t-t_{inc}^{0}\right) \tag{7}$$

We make the following assumptions for the incident delay function:

Assumption 1: Incident delay is only experienced on the incident arc (no propagation of the incident delay effect in the remainder of the network).

Assumption 2: Incident delay function is additive which amplifies the incumbent arc travel time.

Assumption 3: Incident delay function, $\Theta(\cdot)$, is such that the total delay associated by deciding to wait at a node (e.g., waiting time plus the incident delay), is not less than the case without waiting.

In practice, the incident effect propagates in the network in the form of a shockwave after a certain duration following the incident. Since our goal is to investigate the impact of incidents on the travel time, we choose to focus on the most important ingredient, namely the incident-induced delay on the incident arc. Hence, *Assumption 1* is acceptable under certain scenarios. One scenario is where the incident duration is not long enough that vehicles divert to alternative arcs or the capacity of alternative arcs is sufficiently large to accommodate the diversion without any change in their congestion state. The additive model assumption (*Assumption 2*) is appropriate since the travel time delay of a particular incident depends on both the incident characteristics and the incumbent travel time on the arc. *Assumption 3* is consistent with our network and travel time assumptions where we assume that waiting at a node (or on an arc) is not permitted and/or does not provide travel time savings (first-in-first-out property). The following lemma provides a requirement for the incident model parameters such that the *Assumption 3* holds.

Lemma 1. The incident-induced delay parameters (*c*,*q*), satisfying the following condition for the minimal waiting time of Δ (smallest discrete time interval), ensures that waiting at the incident node does not reduce the expected travel time.

$$\mu_a(t_k+\Delta,s)-\mu_a(t_k,s)\geq -\frac{q}{c}\Delta$$

Proof. Let $a\in A$ denote the incident arc with origin and destination nodes (n_k,n_{k+1}). Further, let $t_{k+1}=t_k+\delta_a(t_k,s,t_k-t_{inc}^0)$ represent the arrival time to the node n_{k+1} after departing from n_k at time t_k. Then the expected travel time from node n_k to the trip destination node (n_d) under an optimal policy is $E\left\{\delta_a\left(t_k,s,i=\left(t_k-t_{inc}^0\right)\right)+F^*\left(n_{k+1},t_k+\delta_a\left(t_k,s,t_k-t_{inc}^0\right),w\right)\right\}$, where the second term is the cost-to-go from node n_{k+1} at time t_{k+1} with congestion state vector w for future arcs at t_{k+1}. Let's denote the expected travel time from node n_k to the trip destination node (n_d) at time t_k and $t_k+\Delta$ with $D(t_k)$ and $D(t_k+\Delta)$, respectively.

$$D(t_k)=\delta_a\left(t_k,s,t_k-t_{inc}^0\right)+F^*\left(n_{k+1},t_k+\delta_a\left(t_k,s,t_k-t_{inc}^0\right),w\right)$$
$$D(t_k+\Delta)=\delta_a\left(t_k+\Delta,s,t_k+\Delta-t_{inc}^0\right)+F^*\left(n_{k+1},t_k+\delta_a\left(t_k+\Delta,s,t_k+\Delta-t_{inc}^0\right),w\right)$$

Assumption 3 states that at any node arrival time (t_k), waiting at the node does not lead to lower destination arrival time than without waiting. We write this condition for the minimal waiting time of Δ unit time (smallest discrete time interval),

$$E\{D(t+\Delta)\}-E\{D(t)\}\geq -\Delta .$$

We assume that cost-to-go functions alone satisfy this relationship as we assumed that link travel times (in both congestion states) and state transitions are such that waiting at a node does not provide travel time savings in the recurrent congestion (e.g., first-in-first-out property). For Δ waiting time this leads to the following relation for every t_k:

$$F^*\left(n_{k+1},t_k+\Delta+\delta_a\left(t_k+\Delta,s,t_k+\Delta-t_{inc}^0\right),w\right)-F^*\left(n_{k+1},t_k+\delta_a\left(t_k,s,t_k-t_{inc}^0\right),w\right)\geq -\Delta .$$

Hence, we have the following relation:

$$E\left\{\delta_a\left(t_k+\Delta, s, t_k+\Delta-t_{inc}^0\right)\right\}-E\left\{\delta_a\left(t_k, s, t_k-t_{inc}^0\right)\right\} \geq -\Delta,$$

where,

$$\begin{aligned} E\left\{\delta_a\left(t_k, s, t_k-t_{inc}^0\right)\right\} &= E\left\{\delta_a\left(t_k, s, i=0\right)+\Theta_a\left(\varphi, c, \rho, q, i=t_k-t_{inc}^0\right)\right\} \\ &= \mu_a\left(t_k, s\right)+E\left\{\Theta_a\left(\varphi, c, \rho, q, i=t_k-t_{inc}^0\right)\right\}, \end{aligned}$$

and, $\mu_a\left(t_k, s\right)$ is the mean travel time on arc a at time t_k with congestion state s. The expression $E\left\{\Theta_a\left(\varphi, c, \rho, q, i=t_k-t_{inc}^0\right)\right\}$ can be expressed in two alternative closed-form expressions. In the first case, we assume that the vehicle experiences the maximum delay (i.e. fixed-delay regime in Fu and Rilett [29]),

$$E\left\{\Theta_a\left(\varphi, c, \rho, q, i=t_k-t_{inc}^0\right)\right\}=\frac{q-\rho}{\rho}\left(t_k-t_{inc}^0\right).$$

The other alternative is the variable-delay regime in which the vehicle experiences a delay somewhere between the no-delay and the maximum delay [29].

$$E\left\{\Theta_a\left(\varphi, c, \rho, q, i=t_k-t_{inc}^0\right)\right\}=\frac{c-\rho}{c}\mu_{inc}-\frac{c-q}{c}\left(t_k-t_{inc}^0\right).$$

Note that the waiting decision at the incident node is reasonable only in the case of incident queue dissipation, i.e. either the incident is cleared but the queue is not fully dissipated or the incident is not cleared but the vehicle will exit the link before the clearance. This corresponds to the variable-delay regime and we will show that this holds true by comparing the conditions derived for each case. We first express the no node waiting condition under incident for variable-delay regime as:

$$\begin{aligned} & E\left\{\delta_a\left(t_k+\Delta, s, t_k+\Delta-t_{inc}^0\right)\right\}-E\left\{\delta_a\left(t_k, s, t_k-t_{inc}^0\right)\right\} \geq -\Delta \\ & \mu_a\left(t_k+\Delta, s\right)+E\left\{\Theta_a\left(\varphi, c, \rho, q, t_k+\Delta-t_{inc}^0\right)\right\}-\mu_a\left(t_k, s\right)-E\left\{\Theta_a\left(\varphi, c, \rho, q, t_k-t_{inc}^0\right)\right\} \geq -\Delta \\ & \mu_a\left(t_k+\Delta, s\right)-\mu_a\left(t_k, s\right)-\frac{c-q}{c}\left(t_k+\Delta-t_{inc}^0\right)+\frac{c-q}{c}\left(t_k-t_{inc}^0\right) \geq -\Delta \\ & \mu_a\left(t_k+\Delta, s\right)-\mu_a\left(t_k, s\right) \geq -\frac{q}{c}\Delta . \end{aligned}$$

When we take the limit $\Delta \rightarrow 0$, we have,

$$\frac{d\mu_a(t, s)}{dt}\bigg|_{t=t_k} \geq -\frac{q}{c}.$$

In the maximum delay case, the no node waiting condition can be expressed as:

$$E\left\{\delta_a\left(t_k+\Delta, s, t_k+\Delta-t_{inc}^0\right)\right\}-E\left\{\delta_a\left(t_k, s, t_k-t_{inc}^0\right)\right\} \geq -\Delta$$

$$\mu_a\left(t_k+\Delta, s\right)+E\left\{\Theta_a\left(\varphi, c, \rho, q, t_k+\Delta-t_{inc}^0\right)\right\}-\mu_a\left(t_k, s\right)-E\left\{\Theta_a\left(\varphi, c, \rho, q, t_k-t_{inc}^0\right)\right\} \geq -\Delta$$

$$\mu_a\left(t_k+\Delta, s\right)-\mu_a\left(t_k, s\right)+\frac{q-\rho}{\rho}\left(t_k+\Delta-t_{inc}^0\right)-\frac{q-\rho}{\rho}\left(t_k-t_{inc}^0\right) \geq -\Delta$$

$$\mu_a\left(t_k+\Delta, s\right)-\mu_a\left(t_k, s\right) \geq -\frac{q}{\rho}\Delta .$$

When we take the limit $\Delta \rightarrow 0$, we have,

$$\frac{d\mu_a(t,s)}{dt}\Big|_{t=t_k} \geq -\frac{q}{\rho}.$$

Note that since the capacity under incident is less than regular capacity, i.e. $c>\rho$, we have the condition for variable-delay regime more strict than the fixed-delay regime, i.e., $-q/c>q/\rho$. Hence, for arbitrary waiting time Δ, no node waiting condition under incident is:

$$\mu_a\left(t_k+\Delta, s\right)-\mu_a\left(t_k, s\right) \geq -\frac{q}{c}\Delta .$$

□

8. METHODOLOGY: DYNAMIC ROUTING MODEL WITH RECURRENT AND NON-RECURRENT CONGESTION

We assume that the objective of our dynamic routing model is to minimize the expected travel time based on real-time information where the trip originates at node n_0 and concludes at node n_d. Let's assume that there is a feasible path between (n_0, n_d) where a path $p=(n_0,..,n_k,..,n_{K-1})$ is defined as sequence of nodes such that $a_k \equiv (n_k, n_{k+1}) \in A$, $k=0,..,K-1$ and K is the number of nodes on the path. We define set $a_k \equiv (n_k, n_{k+1}) \in A$ as the *current arcs* set of node n_k, and denoted with $CrAS(n_k)$. That is, $CrAS(n_k) \equiv \{a_k : a_k \equiv (n_k, n_{k+1}) \in A\}$ is the set of arcs emanating from node n_k.

Each node on a path is a decision stage (or epoch) at which a routing decision (which node to select next) is to be made. Let $n_k \in N$ be the location of k^{th} decision stage, t_k is the time at k^{th} decision stage where $t_k \in \{1,...,T\}$, $T>t_{K-1}$. Note that we are discrediting the planning horizon.

We next define our look ahead policy for projecting the congestion states in the network. While optimal dynamic routing policy requires real-time consideration and projection of the traffic states of the complete network, this approach makes the state space prohibitively large. In fact, there is little value in projecting the congestion states well ahead of the current location. This is because the projected information is not different than the long run average steady state probabilities of the arc congestion states. Hence, an efficient but practical approach would tradeoff the degree of look ahead (e.g., number of arcs to monitor) with the resulting projection accuracy and routing performance. This has been very well illustrated in Kim et al. [24]. Thus we limit our look ahead to finite number of arcs that can vary by the vehicle location on the network. The selection of the arcs to monitor would depend on factors such as arc lengths, value of real-time information, and the arcs' congestion state transition characteristics. For ease of presentation and without loss of generality, we choose to monitor only two arcs ahead of the vehicle location and model the rest of the arcs' congestion states through their steady state probabilities. Accordingly, we define the following two sets for all arcs in the network. $ScAS(a_k)$, the successor arc set of arc a_k, $ScAS(a_k) \equiv \{a_{k+1} : a_{k+1} \equiv (n_{k+1}, n_{k+2}) \in A\}$, i.e., the set of outgoing arcs from the destination node (n_{k+1}) of arc a_k. $PScAS(a_k)$, the post-successor arc set of arc a_k, $PScAS(a_k) \equiv \{a_{k+2} : a_{k+2} \equiv (n_{k+2}, n_{k+3}) \in A\}$ i.e., the set of outgoing arcs from the destination node (n_{k+2}) of arc a_{k+1}.

Since the total trip travel time is an additive function of the individual arc travel times on the path plus a penalty function measuring earliness/tardiness of arrival time to the final destination, the dynamic route selection problem can be modeled as a dynamic programming model. The state of the system at k th decision stage is denoted by $\Omega\left(n_k, t_k, s_{a_{k+1} \cup a_{k+2}, k}, I_k\right)$. This state vector is composed of the state of the vehicle and network and thus characterized by the current node (n_k), the current node arrival time (t_k), and $s_{a_{k+1} \cup a_{k+2}, k}$ the congestion state of arcs $a_{k+1} \cup a_{k+2}$ where $\{a_{k+1} : a_{k+1} \in ScAS(a_k)\}$ and $\{a_{k+2} : a_{k+2} \in PScAS(a_k)\}$, and incident durations (I_k) of the network at stage k , i.e. $I_k \equiv I(t_k)$. The action space for the state $\Omega\left(n_k, t_k, s_{a_{k+1} \cup a_{k+2}, k}, I_k\right)$ is the set of current arcs of node n_k, denoted with $CrAS(n_k)$.

At every decision stage, the trip planner evaluates the alternative arcs from $CrAS(n_k)$ based on the remaining expected travel time. The expected travel time at a given node with the selection of an outgoing arc is the expected arc travel time on the arc chosen and the expected travel time of the next node. Let $\pi = \{\pi_0, \pi_1, ..., \pi_{K-1}\}$ be the policy of the trip and is composed of policies for each of the *K-1* decision stages. For a given state $\Omega_k(n, t, S, I)$, the policy $\pi_k(\Omega_k)$ is a deterministic Markov policy which chooses the outgoing arc from node n_k, i.e., $\pi_k(\Omega_k) = a \in CrAS(n_k)$. Therefore the expected travel cost for a given policy vector $\pi = \{\pi_0, \pi_1, ..., \pi_{K-1}\}$ is as follows:

$$F_0(n,t,S,I) = \underset{\delta_k}{E}\left\{ g_{K-1}(\Omega_{K-1}) + \sum_{k=0}^{K-2} g_k(\Omega_k, \pi_k(\Omega_k), \delta_k) \right\} \tag{8}$$

where (n_0, t_0, S_0, I_0) is the starting state of the system. δ_k is the random travel time at decision stage k, i.e., $\delta_k \equiv \delta(t_k, \pi_k(\Omega_k), s_a(t_k), i_a(t_k)) + \Theta(\varphi, c, \rho, q, i)$ and $\Theta(\varphi, c, \rho, q, i=0) = 0$, i.e. the incident delay of an arc without incident. $g_a(\Omega_k, \delta_k)$ is cost of travel on arc $a = \pi_k(\Omega_k) \in CrAS(n_k)$ at stage k, i.e., if travel cost is a function (ϕ) of the travel time, then $g(\Omega_k, \pi_k(\Omega_k), \delta_k) \equiv \phi(\delta_k)$. Then the minimum expected travel time can be found by minimizing $F_0(n,t,S,I)$ over the policy vector $\pi = \{\pi_0, \pi_1, ..., \pi_{K-1}\}$ as follows:

$$F_0^*(n,t,S,I) = \min_{\pi=\{\pi_0,\pi_1,...,\pi_{K-1}\}} F_0(n,t,S,I) \tag{9}$$

The corresponding optimal policy is then $\pi^* = \underset{\pi=\{\pi_0,\pi_1,...,\pi_{K-1}\}}{\arg\min} F_0(n,t,S,I)$. Hence, the Bellman's cost-to-go equation for the dynamic programming model can be expressed as follows [46]:

$$F_k^*(\Omega) = \min_{\pi_k} \underset{\delta_k}{E}\left\{ g_k(\Omega, \pi(\Omega), \delta) + F_{k+1}^*(\Omega) \right\} \tag{10}$$

For a given policy $\pi_k(\Omega_k) = a_k \in CrAS(n_k)$, we can re-express the cost-to-go function by writing the expectation in the following explicit form:

$$\begin{aligned} F_k(n,t,S,I \mid a) = \sum_{\delta_k} P_k(\delta \mid n,t,S,I,a)\big[g_k(\Omega,a,\delta) + \\ \sum_{s_{a_{k+1}}} P\left(s_{a_{k+1}}(t_k) \mid s_{a_{k+1}}(t_k+\delta_k)\right) \sum_{s_{a_{k+2}}} P\left(s_{a_{k+2}}(t)\right) \sum_{I_{k+1}} P\left(I(t_k) \mid I(t_k+\delta_k)\right) F_{k+1}\left(n,(t_k+\delta_k),S,I\right)\big] \end{aligned} \tag{11}$$

where $P_k(\delta \mid n,t,S,I,a)$ is the probability of travelling arc a_k in δ_k periods. $P\left(s_{a_{k+2}}(t_{k+1})\right)$ is the long run probability of arc $a_{k+2} : a_{k+2} \in PScAS(a_k)$ being in state $s_{a_{k+2},k+1}$ in stage $k+1$. This probability can be calculated from the historical frequency of a state for a given arc and time.

We use backward dynamic programming algorithm to solve for $F_k^*(\Omega)$, $k = K-1, K-2, .., 0$. In the backward induction, we initialize the final decision epoch such that, $\Omega_{K-1} = \Omega(n_{K-1}, t_{K-1})$, n_{K-1} is destination node, and $F_{K-1}(\Omega) = 0$ if $t_{K-1} \le T$. Accordingly, a penalty cost is accrued whenever there is delivery tardiness, e.g., $t_{K-1} > T$.

9. DISCUSSION OF RESULTS: EXPERIMENTAL STUDY

This section demonstrates the performance of the proposed algorithm on a network from southeast Michigan with real-time traffic data from the Michigan Intelligent Transportation Systems (MITS) Center. MITS center is the hub of ITS technology applications at the Michigan Department of Transportation (MDOT) and oversees a traffic monitoring system composed of 180 freeway miles instrumented with 180 Closed Circuit TV Cameras, Dynamic Message Signs, and 2260 Inductive Loops. The methods also utilize real-time and archived data from Traffic.com, a private company that provides traffic information services in several states and also operates additional sensors and traffic monitoring devices in Michigan. Traffic.com also provides information regarding incidents causing non-recurrent congestion (e.g., incident location, type, severity, and times of incident occurrence and clearance). We implemented all our algorithms and methods in Matlab 7 and executed on a Pentium IV machine (with 1.6 GHz speed processor and 1024 MB RAM) running Microsoft Windows XP operating system.

Our experimental study is outlined as follows: Section 5.1 introduces two road networks from southeast Michigan used for demonstrating the performance of the proposed algorithms along with a description of their general traffic conditions. Section 5.2 describes the process and the results from modeling of recurrent congestion for the networks. Section 5.3 reports savings from employing the proposed dynamic routing model under recurrent congestion for a network with multiple OD pairs. Section 5.4 presents the experimental setup that involves an incident and reports results and savings from employing the proposed dynamic routing model under both recurrent and non-recurrent congestion.

(a) (b)

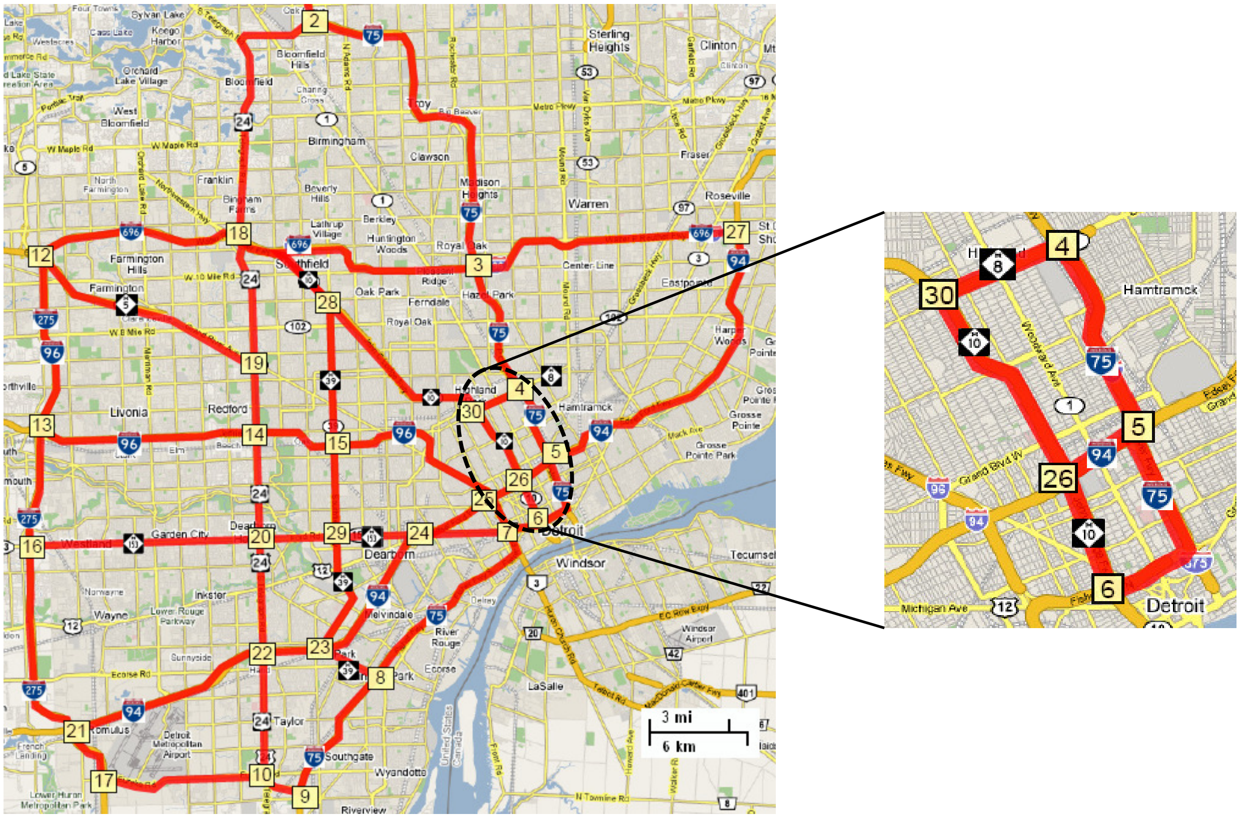

Figure 2. (a) South-East Michigan Road Network Considered for Experimental Study. (b) Sub-Network from Southeast Wayne County

9.1. Sample Networks and Traffic Data

This section introduces the road networks from southeast Michigan used for demonstrating the performance of the proposed algorithms along with a description of their general traffic conditions. As illustrated in Figure 2, the sample network covers southeast Michigan freeways and highways in and around the Detroit metropolitan area.

The network has 30 nodes and a total of 98 arcs with 43 observed arcs (with real-time ITS information from MITS Center) and 55 unobserved arcs. Real-time traffic data for the observed arcs is collected from MDOT Center for 23 weekdays from January 21, 2008 to February 20, 2008 for the full 24 hours of each day at a resolution of an observation every minute. The raw traffic speed data from the MITS Center is cleaned with a series of procedures from Texas Transportation Institute and Cambridge Systematics [4] to improve quality and reduce data errors.

A small part of our full network, labeled *sub-network*, is used here to better illustrate the methods and results (Figure 2b). The sub-network has five nodes and six observed arcs, with more details provided in Table 1.

Table 1. Information Regarding Sub-Network Nodes and Arcs

			FROM		TO	
Arc ID	Freeway	Length (miles)	Node #	Description (Exit #)	Node #	Description (Exit #)
1	I-94	1.32	5	216	26	215
2	M-8	1.75	4	56A (I-75)	30	7C (M-10)
3	I-75	3.13	4	56A	5	53B
4	I-75	2.81	5	53B	6	50
5	M-10	3.26	30	7C	26	4B
6	M-10	1.42	26	4B	6	2A

In the experiments based on the sub-network, node 4 is considered as the origin node and node six as the destination node of the trip. Given the OD pair, we present the speed data for the six different arcs of the sub-network in Figure 3. It can be seen clearly that the traffic speeds follow a stochastic non-stationary distribution that vary with the time of the day. The mean speeds and standard deviations for these same arcs are shown in Figure 4, clearly revealing the non-stationary nature of traffic.

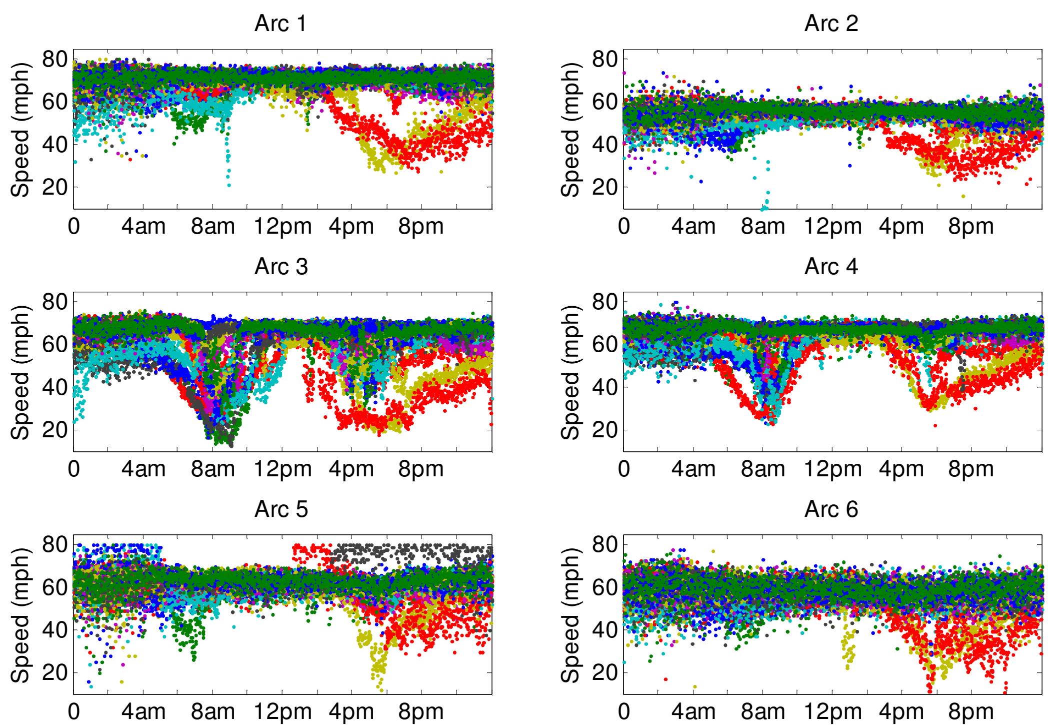

Figure 3. Raw Traffic Speeds for Arcs on Sub-Network (mph) at Different Times of the Day

(Data: Weekday traffic from January 21 to February 20. Each color represents a distinct day of 23 days)

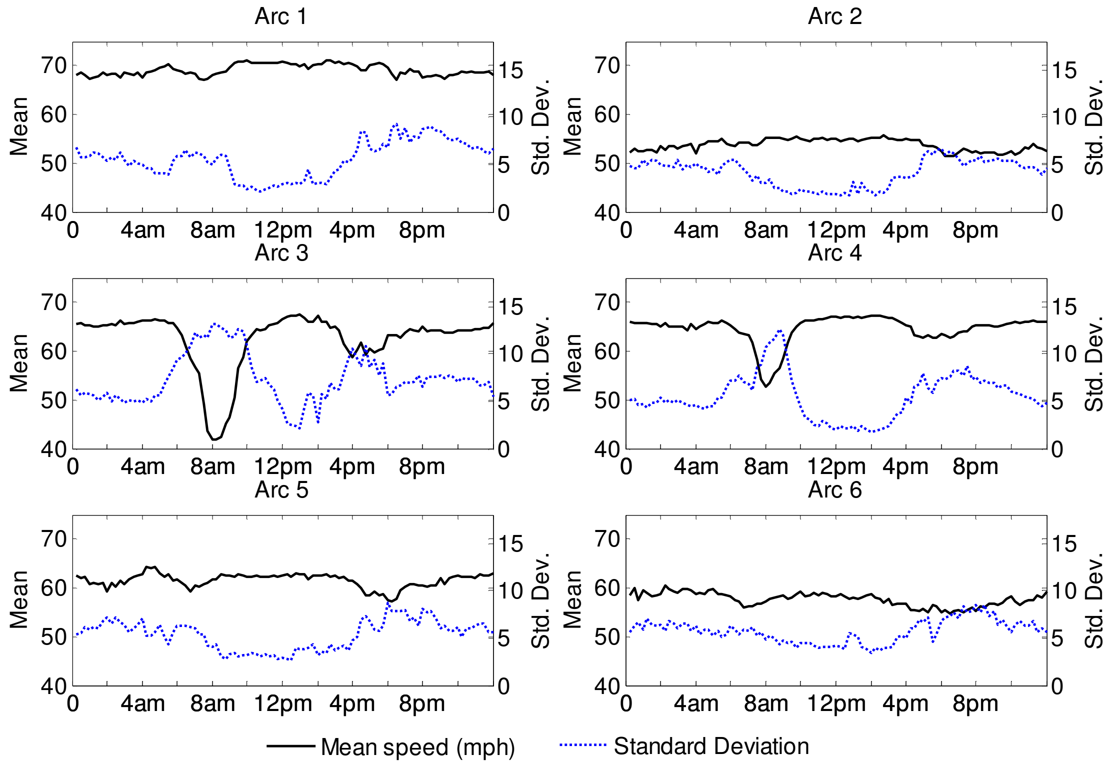

Figure 4. Traffic Mean Speeds (mph) and Standard Deviations by Time of the Day for Arcs on Sub-Network. (15 minute time interval resolution)

9.2 Recurrent Congestion Modeling

The proposed dynamic routing algorithm calls for identification of different congestion states and estimation of their state transition rates as well as arc traverse times by time of day. Given the traffic speed data from the MITS Center, we employed the Gaussian Mixture Model (GMM) clustering technique to determine the number of recurrent-congestion states for each arc by time of day. In particular, we employed the greedy learning GMM clustering method of Verbeek [48] for its computational efficiency and performance. To estimate the number of congestion states, traffic speed data from every pair of two consecutive time periods, t and t+1, are clustered and modeled using a bi-variate joint Gaussian distribution ($\boldsymbol{\mu}^i_{t,t+1};\boldsymbol{\Sigma}^i_{t,t+1}$), where i denotes the i^{th} cluster. The Gaussian distribution assumption has been employed by others in the literature (see Kim et al. [23]). The clusters are ordered by their means and the densities of their projections onto the two axes are employed to identify the congestion state speed intervals, as illustrated in Figure 5. Formally, the cut-off speed between congestion state-pair $(i,i+1)$ for arc a at time t is denoted by $c^i_a(t)$ and is calculated as follows: $c^i_a(t)=x, x: f_{i_t}(x)=f_{(i+1)_t}(x)$ where $f(\cdot)$ is the projected probability density function for state i. Unlike most clustering methods, the GMM clustering procedure employed does not call for specification of number of clusters (i.e., congestion states) in advance and can determine the optimal number of clusters based on the maximum likelihood and model complexity measures.

However, we did limit the number of clusters to two, considered quite adequate for modeling recurrent-congestion, and to limit estimation errors attributable to data scarcity. As expected, the GMM procedure generally yielded mostly two states, even without the constraint, as in Figure 5 (resulting in states denoted 'congested' and 'uncongested' states with c_1^1(8:30)= 64.9 mph), and rarely a single state during periods of low traffic (as in Figure 6). Following these observations, we have adopted two congestion states in representing arc congestion dynamics. Note that this does not compromise from the accuracy of congestion modeling, rather provides uniformity in the algorithmic data structures across all arcs in the network.

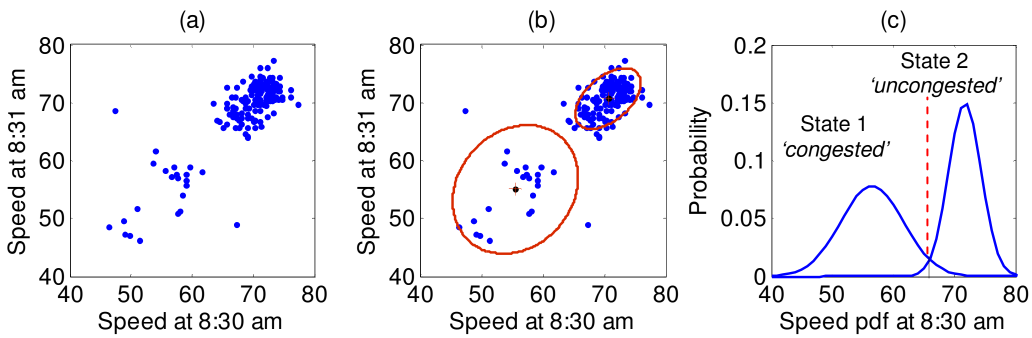

Figure 5. (a) Joint Plots of Traffic Speeds in Consecutive Periods for Modeling State-Transitions at 8:30 am, for Arc 1; (b) Cluster Joint Distributions of Speed at 8:30 am Generated by GMM; (c) Partitioned Traffic States Based on Projections

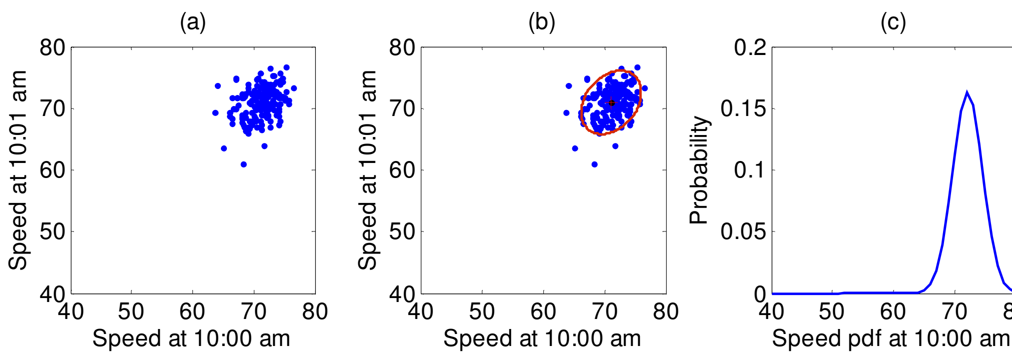

Figure 6. (a) Joint Plots of Traffic Speeds in Consecutive Periods for Modeling State-Transitions at 10:00 am, for Arc 1; (b) Single Cluster Joint Distribution of Speed at 10:00 am Generated by GMM; (c) Partitioned Traffic States Based on Projections

The parameters of the traffic state joint Gaussian distributions (i.e., $\boldsymbol{\mu}^i_{t,t+1};\boldsymbol{\Sigma}^i_{t,t+1}$) along with the computed cut-off speeds (if GMM yields more than one state) are employed to calculate travel time distribution parameters and the transition matrix elements as explained in section 3. In the event that two states are identified by GMM, α_t denotes the probability of state transition from congested state to congested state where as β_t denotes the probability of state transition from uncongested state to uncongested state.

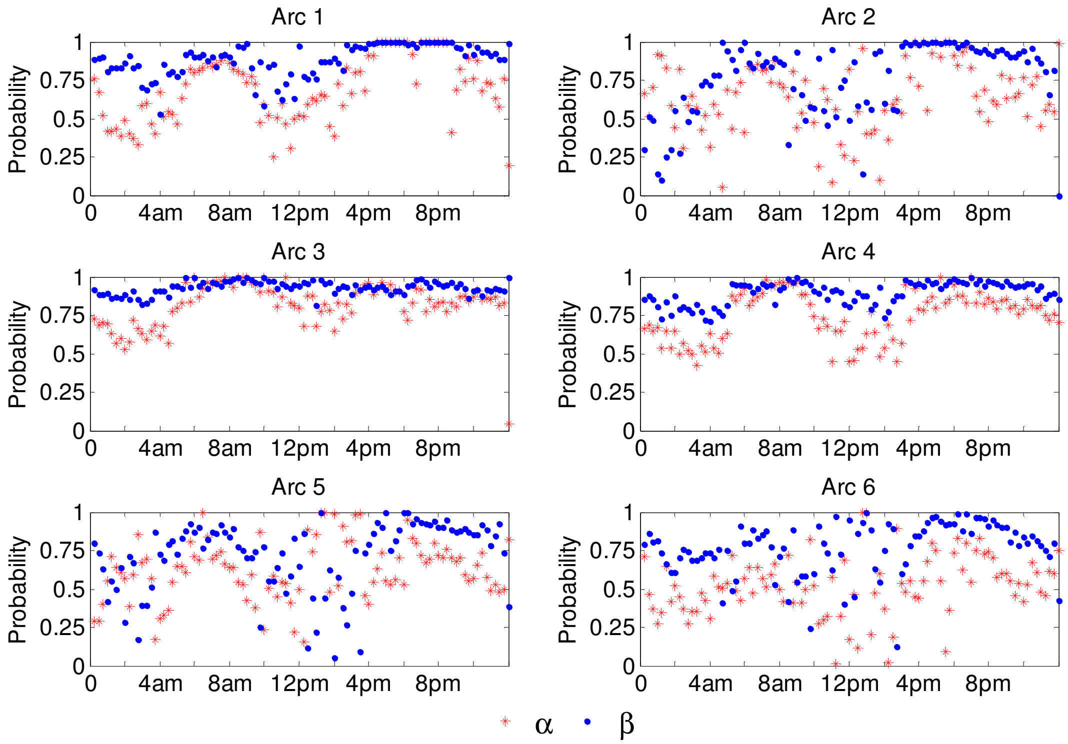

Figure 7. Recurrent Congestion State-Transition Probabilities for Arcs on Sub-Network. α: Congested to Congested Transition; β: Uncongested to Uncongested Transition Probability (Plotted with 15-minute time interval resolution)

Figure 7 plots these transition rates for the different arcs of the sub-network. Note that the state transitions to same states (i.e., congested to congested or uncongested to uncongested) are more likely during peak demand time periods, which increase the value of the congestion state information, and is the case in practice. For the sub-network, the mean and standard deviation of arc travel times are illustrated in Figure 8 and Figure 9, respectively, by traffic state and time of day.

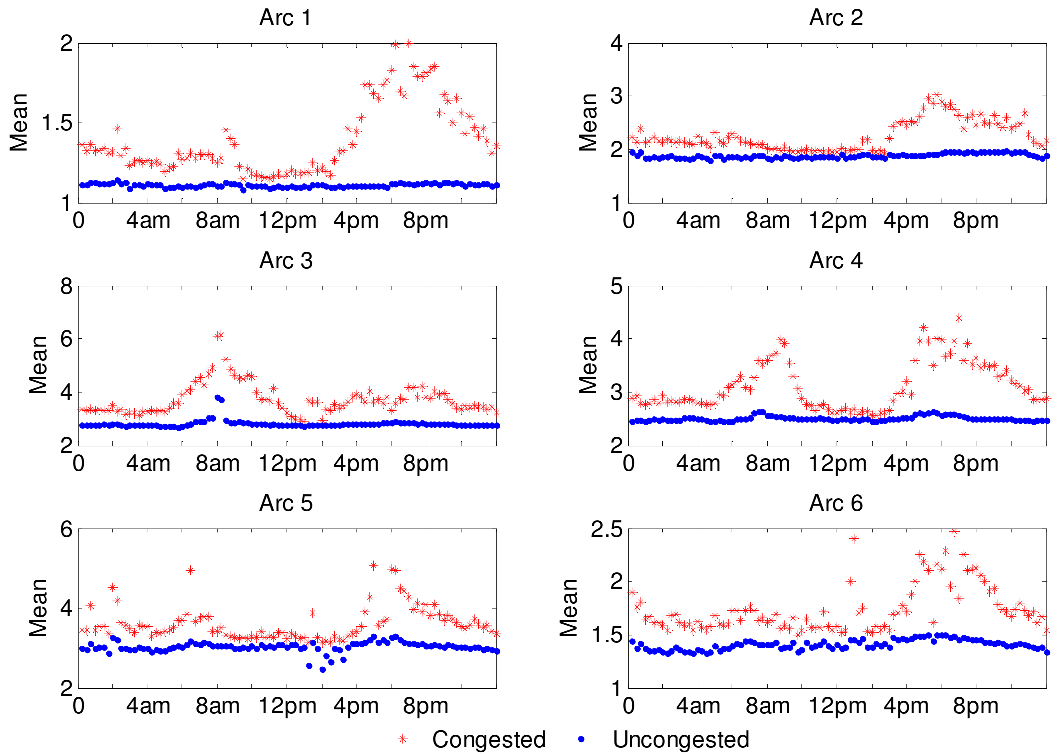

Figure 8. Sub-Network Arc Travel Time Means in Minutes (Plotted with 15-minute time interval resolution)

9.3 Results from Modeling Recurrent Congestion

This section highlights the potential savings from explicit modeling of recurrent congestion during dynamic vehicle routing. First, we discuss the results for routing on the *sub-network*. As stated earlier, we consider *node 4* as the origin node and *node 6* as the destination node of the trip. Three different path options exist (*path 1: 4-5-6; path 2: 4-5-26-6;* and *path 3: 4-30-26-6*). Note that our aim is not to identify an optimal path, rather, to identify the best policy based on the time of the day, location of the vehicle, and the traffic state of the network (for paths can be sub-optimal under non-stationary networks). However, in practice, almost all commercial logistics software aim to identify a robust (static) path that is best on the average. In this context, given the traffic flow histories for the arcs of the sub-network, *path1: 4-5-6* would be most robust, for it dominates other paths most of the day under all network states. Hence, we identify *path 1* as the *baseline path* and show the savings from using the proposed dynamic routing algorithm with regard to baseline path. Since we limit the traffic state look ahead to only successor and post-successor arcs, there are five arc states to be considered at the starting node of the trip. This implies that there are 2^5=32 starting network traffic state combinations.

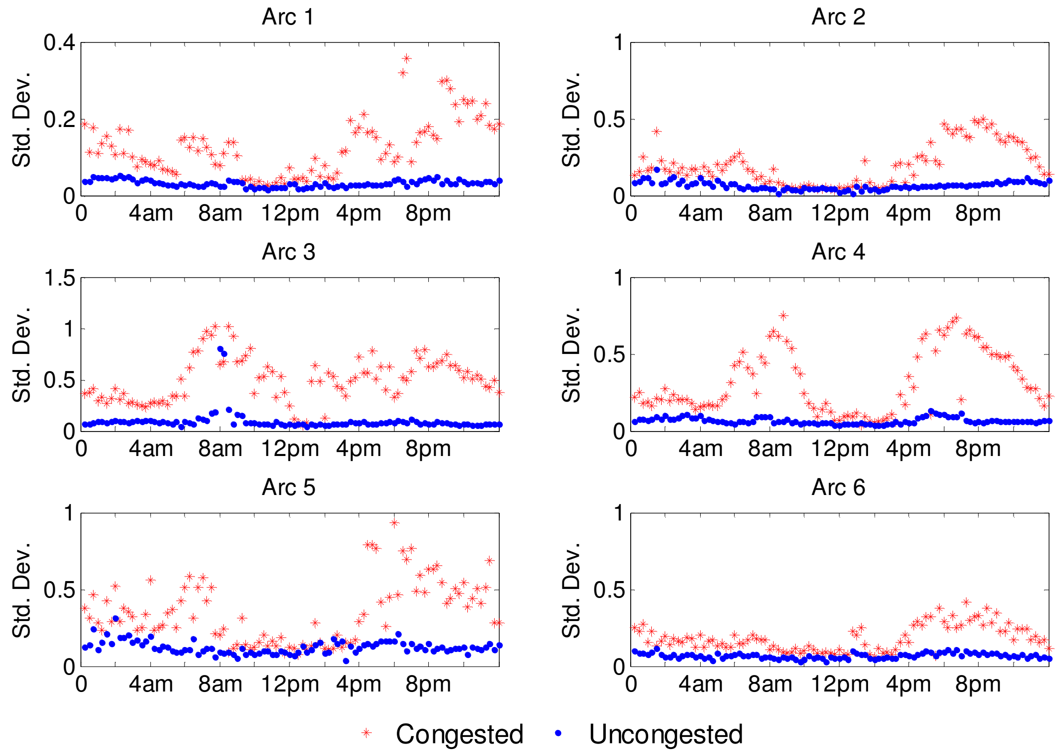

Figure 9. Sub-Network Arc Travel Time Standard Deviations in Minutes (Plotted with 15-minute time interval resolution)

We simulated the trip 10,000 times for each of these starting network traffic state combinations throughout the day for 15-minute interval starting times (yielding (24×60)/15=96 trip start times). Figure 10a plots the mean baseline path travel times over 10,000 simulation runs for every combination of the sub-network traffic state (all 32 of them) and Figure 10b plots the mean travel times for the dynamic policy.

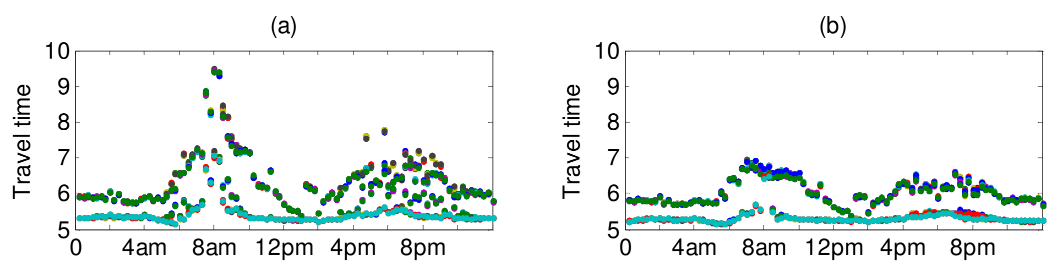

Figure 10. Mean travel times for all state combinations of the sub-network (each color represents a different state combination): (a) Baseline path. (b) Dynamic vehicle routing policy

Figure 11(a) plots the corresponding percentage savings from employing the dynamic vehicle routing policy over the baseline path for each network traffic state combination and Figure 11b shows the average savings (averaged across all network traffic states, treating them equally likely). It is clear that savings are higher and rather significant during peak traffic times and lower when there is not much congestion, as can be expected.

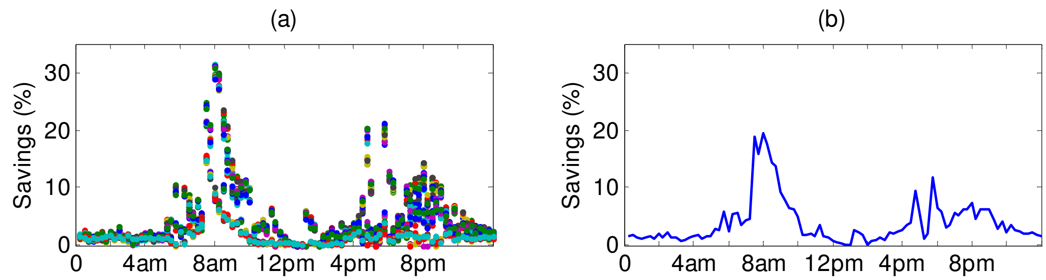

Figure 11. Savings from employing dynamic vehicle routing policy over baseline path: (a) Savings for each of the 32 network state combinations (b) Average savings across all state combinations

Besides the sub-network (Figure 2b), as listed in Table 2, we have also identified 5 other origin and destination (OD) pairs in the southeast Michigan road network (Figure 2a) to investigate the potential savings from using real-time traffic information under a dynamic routing policy. Unlike the *sub-network*, these OD pairs have both observed and unobserved arcs and each OD pair has several alternative paths from origin node to destination node.

Table 2 : Origin-Destination Pairs Selected from Southeast Michigan Road Network

	ORIGIN		DESTINATION	
OD Pair	Node #	Description (Intersection of)	Node #	Description (Intersection of)
1	2	I-75 & US-24	21	I-275 & I-94
2	12	I-96 & I-696	25	I-96 & I-94
3	19	M-5 & US-24	27	I-696 & I-94
4	23	I-94 & M-39	13	I-96 & I-275
5	3	I-75 & I-696	15	I-96 & M-39

Once again, we identify the *baseline path* for each OD pair (as explained for the case of routing on the *sub-network*) and show percentage savings in mean travel times (over 10,000 runs) over the baseline paths from using the dynamic routing policy. Figure 12 plots the percentage savings for each network traffic state combination and Figure 13 shows the average savings (averaged across all network traffic states, treating them equally likely). The savings are consistent with results from the *sub-network*, somewhat validating the sub-network results, with higher savings once again during peak traffic times.

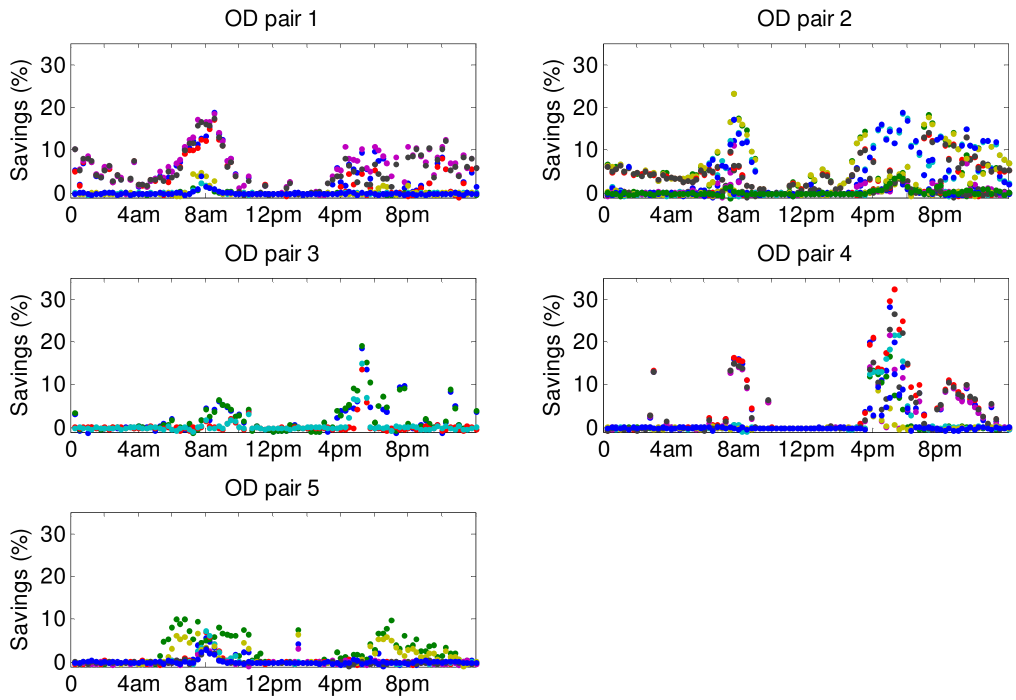

Figure 12. Savings of Dynamic Policy Over Baseline Path During the Day for All Starting States of Given OD Pairs (with 15-minute time interval resolution)

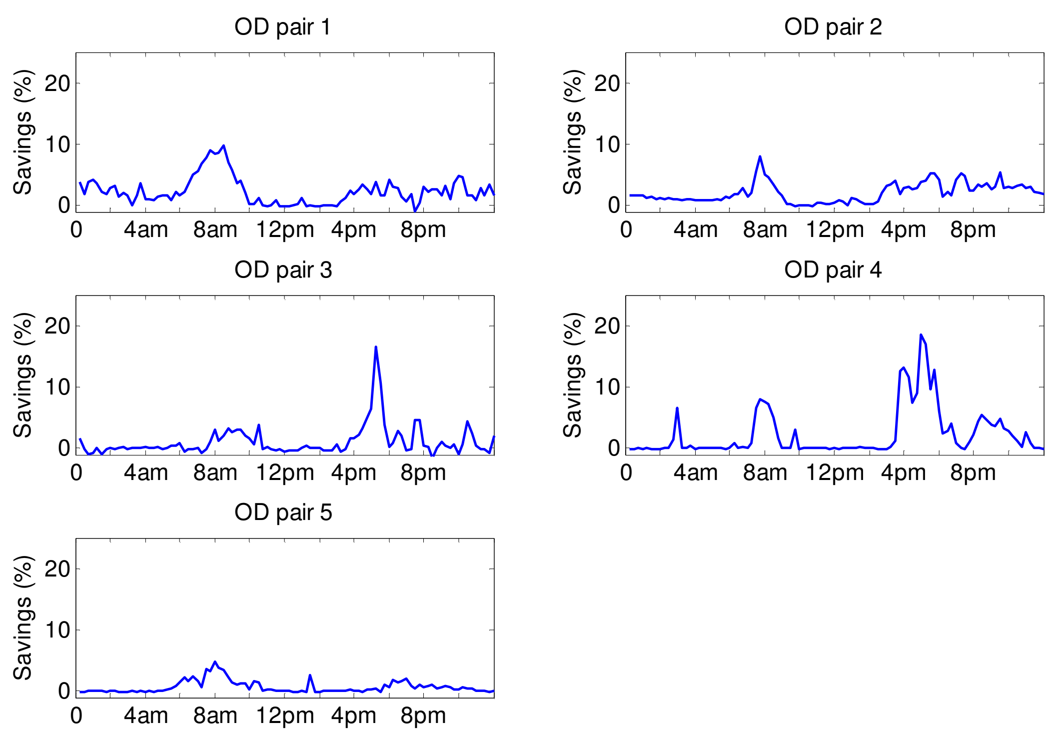

Figure 13. Average Savings of Dynamic Policy Over Baseline Path During the Day for All starting states of given OD pairs (with 15-minute time interval resolution)

9.4. Impact of Modeling Incidents

This section highlights the potential savings from explicit modeling of non-recurrent congestion along with modeling of recurrent congestion during dynamic vehicle routing. As for the setting, we focus on the sub-network (Figure 2b). We derive the dynamic routing policies in two ways. Initially, the dynamic policy does not account for non-recurrent congestion even though there is an incident in the network. Later, we allow the dynamic policy to explicitly account for non-recurrent congestion information to generate the optimal policy. We show the results for 6 starting times during the day (to study the impact of non-stationary traffic on savings): 6:30 am, 9:00 am, 10:30 am, 4:00 pm, 5:30 pm and 7:00 pm. To achieve a good comparison, we set all parameters of the incident to be the same for all starting times. We create an incident on either arc 3, or 4, or 6 with duration mean of 10 minutes and standard deviation of 5 minutes, following a Weibull distribution (scale parameter of 11 and a shape parameter of 2). We assume that all the arcs of the sub-network have a capacity of 1800 vehicles per hour (vph) under normal conditions and that the incident reduces their capacity to 1080 vph. Also, we assume in-flow traffic arrival rate for each arc to be 1500 vph during these operation times. We have also validated the assumption of no node waiting for incident arcs using the condition derived in Lemma 1.

The percentage savings from the explicit modeling of non-recurrent congesting along with recurrent congestion during dynamic vehicle routing are illustrated in Figure 14. The results are very compelling and pertain to three different scenarios. In the first scenario, the incident occurs 10 minutes before vehicle's departure from the starting node. In the second and third scenarios, the incident occurs 20 minutes and 30 minutes before vehicle's departure from the starting node, respectively. For example, if the vehicle departs the origin node at 6:30 am, incident is simulated to occur at 6:20 am or 6:10 am or 6:00 am, and incident has not yet been cleared in all three cases.

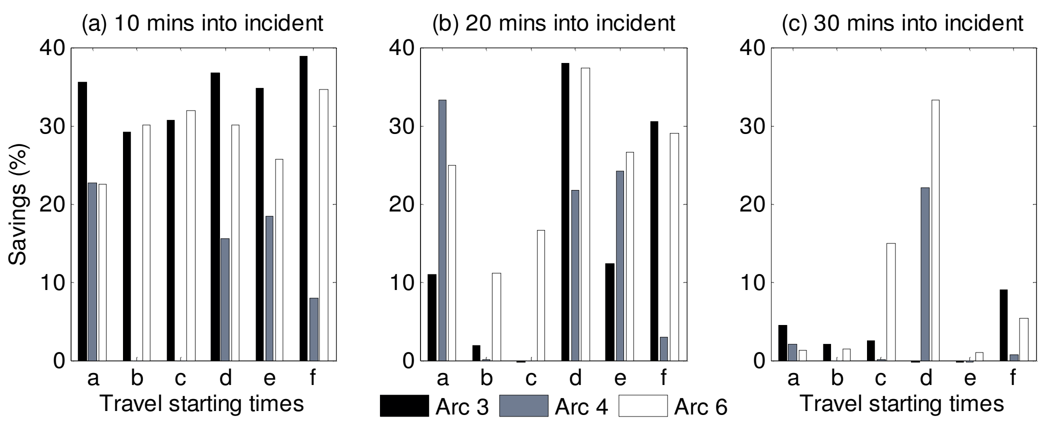

Figure 14. Savings realized by dynamic routing based on modeling both recurrent and non-recurrent congestion compared to the dynamic routing with only recurrent congestion modeling: a: 6:00, b: 7:30, c: 9:00, d: 16:00, e: 17:30, and f: 19:00. Incident is either on arc 3, or 4, or 6. Trip starts (a) 10 minutes (b) 20 minutes (c) 30 minutes after incident has occurred

The savings for the first scenario are presented in Figure 14a. Since arc 3 is close to the origin node, the effect of incident is generally high which leads to greater savings. Arc 4 is a downstream arc (i.e., it is not connected to the origin node), thus the incident is partially cleared by the time the vehicle reaches there. Subsequently, the impact of the incident on arc travel time and the savings are lesser. Arc 6 is also a downstream arc but the dynamic policy (without taking into account the non-recurrent congestion) sometimes chooses this arc, thus there are savings associated with explicit modeling of non-recurrent congestion. Due to space constraints, we are not presenting results from incidents on other arcs. The results for other arcs vary for similar reasons. The results for the second scenario (e.g., 20 minutes into the incident) are presented in Figure 14b. The savings for this scenario are less than the first scenario since the incident has partially or fully cleared by the time the vehicle reaches the incident arcs. Otherwise, we generally see consistency in savings with the first scenario. Figure 14c presents the results for the third scenario and savings for this scenario are mostly less than the other scenarios since the incident is more likely to be fully cleared by the time the vehicle reaches the incident arcs. To illustrate the results better, we also report the path distributions for the case where incident took place on arc 4 (because of space limits, we are not showing the other results). Figure 15a reports the path distribution of the dynamic policy in the absence of explicit modeling of non-recurrent congestion due to the incident that took place 10 minutes before trip start time. Figure 15b, c, and d report path distributions under explicit modeling of incidents and the resulting non-recurrent congestion, with trip start times of 10, 20, and 30 minutes into the incident, respectively. Since the incident is on path 1, there is no routing on path 1 for the case when trip starts just 10 minutes after the incident occurred (Figure 15b). As time passes, since the probability of incident clearance and no delay regime increases, dynamic routing policy starts to select this path as well (Figure 15d and d).

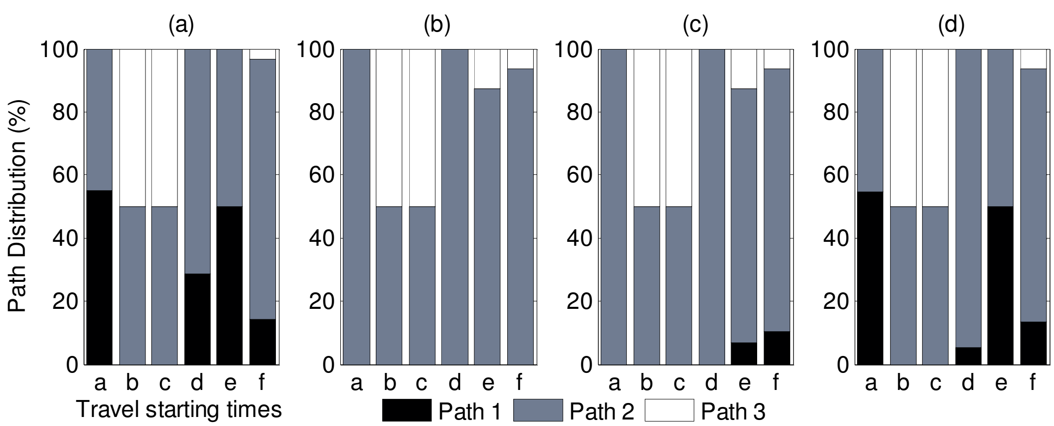

Figure 15. Path distribution from dynamic routing under an incident on arc 4 for different trip start times: a: 6:00, b: 7:30, c: 9:00, d: 16:00, e: 17:30, f: 19:00. (a) Results without modeling incident and trip starts 10 minutes into incident. (b), (c), and (d) report path distributions under explicit modeling of incidents, with trip start times of 10, 20, and 30 minutes into the incident, respectively

10. CONCLUSIONS

The paper proposes practical dynamic routing models that can effectively exploit real-time traffic information from Intelligent Transportation Systems (ITS) regarding recurrent congestion, and particularly, non-recurrent congestion stemming from incidents (e.g., accidents) in transportation networks. With the aid of this information and technologies, our models can help drivers avoid or mitigate trip delays by dynamically routing the vehicle from an origin to a destination in road networks. While non-recurrent congestion is known to be responsible for a major part of network congestion, extant literature mostly ignores this in proposing dynamic routing algorithms. We model the problem as a non-stationary stochastic shortest path problem with both recurrent and non-recurrent congestion. We propose effective data driven methods for accurate modeling and estimation of recurrent congestion states and their state transitions. A Markov decision process (MDP) formulation that generates a routing "policy" to select the best node to go next based on a "state" (vehicle location, time of day, and network congestion state) is proposed to solve the problem. While optimality is only guaranteed if we employ the full state of the transportation network to derive the policy, we recommend a limited look ahead approach to prevent exponential growth of the state space. The proposed model also estimates incident-induced arc travel time delay using a stochastic queuing model and uses that information for dynamic re-routing (rather than anticipate these low probability incidents).

ITS data from southeast Michigan road network, collected in collaboration with Michigan Intelligent Transportation System Center and Traffic.com, is used to illustrate the performance of the proposed models. Our experiments clearly illustrate the superior performance of the SDP derived dynamic routing policies when they accurately account for recurrent congestion (i.e., they differentiate between congested and uncongested traffic states) and non-recurrent congestion attributed to incidents. Experiments show that as the uncertainty (standard deviation) in the travel time information increases, the dynamic routing policy that takes real-time traffic information into account becomes increasingly superior to static path planning methods. The savings however depend on the network states as well as the time of day. The savings are higher during peak times and lower when traffic tends to be static (especially at night). Experiments also show that explicit treatment of non-recurrent congestion stemming from incidents can yield significant savings.

11. RECOMMENDATIONS FOR FURTHER RESEARCH

Further research will focus on developing dynamic routing algorithms for supporting 'milk-runs' where a vehicle departs from an origin to serve several destinations in a network with one or more of the following settings: 1) stochastic time-dependent network where vehicles may encounter recurrent and/or non-recurrent congestion during the trip, 2) vehicle must pickup/deliver within specific time-windows at customer locations, 3) stochastic dependencies and interactions between arcs' congestion states, and 4) anticipate and respond to the behavior of the rest of the traffic to the real-time ITS information.

12. RECOMMENDATIONS FOR IMPLEMENTATION

The research identified a number of recommendations for implementation to help leverage the full potential of dynamic routing of freight vehicles using real-time ITS information.

- Implementation mechanisms for ensuring data quality. Recommendations include collecting and sharing up-to-date sensor maintenance and placement information in the implementation network. This allows the users of the models and algorithms developed to revise their estimates of the travel times as well as traffic behavior under incident conditions. While the majority of the arteries do not have sensors, we found that some of the sensors in major highways are inactive. The absence of these sensors on the large segments of highways creates quality problems associated with distribution estimations for travel times as well as state transitions. In addition, the data collected from the some of the sensors are found to be inaccurate, e.g. inconsistent speed data, which may be attributable to weather, sensor's health state, and communication network inefficiencies. Recommendations for the missing sensor information or inaccurate data captures include benchmarking the traffic condition on the network segments devoid of sensors with those having sensors and reconciling and using a linear regression estimation of the speed data at any given time between adjacent sensors on a highway segment.

- The routing policies derived in this research are for point-to-point routing of the freight carrying vehicle. In most JIT systems this routing is performed in the form of milk run pickups and deliveries. Implementation recommendation for milk runs is to enumerate the potential order of customer visits and then using the historical congestion state probabilities for links emanating from each customer node visited.

- The off-line routing policy generation is impractical given the large number of links and incident state possibilities. Recommendation for implementation is to communicate the route actions (which road network link to select next) to the driver through a wireless connection (e.g., satellite) in real time. The identification of the real-time routing decisions is achieved through a centralized dynamic routing decision support system implementing the models and algorithms developed in this research. The decision support system is recommended to extract the real-time traffic congestion information from the ITS server. When the server is down or there are communication problems, the default operating mode for the decision support system is to assume the long-term congestion state probabilities. Figure 16 illustrates the recommended framework for data communication and decision support integration.

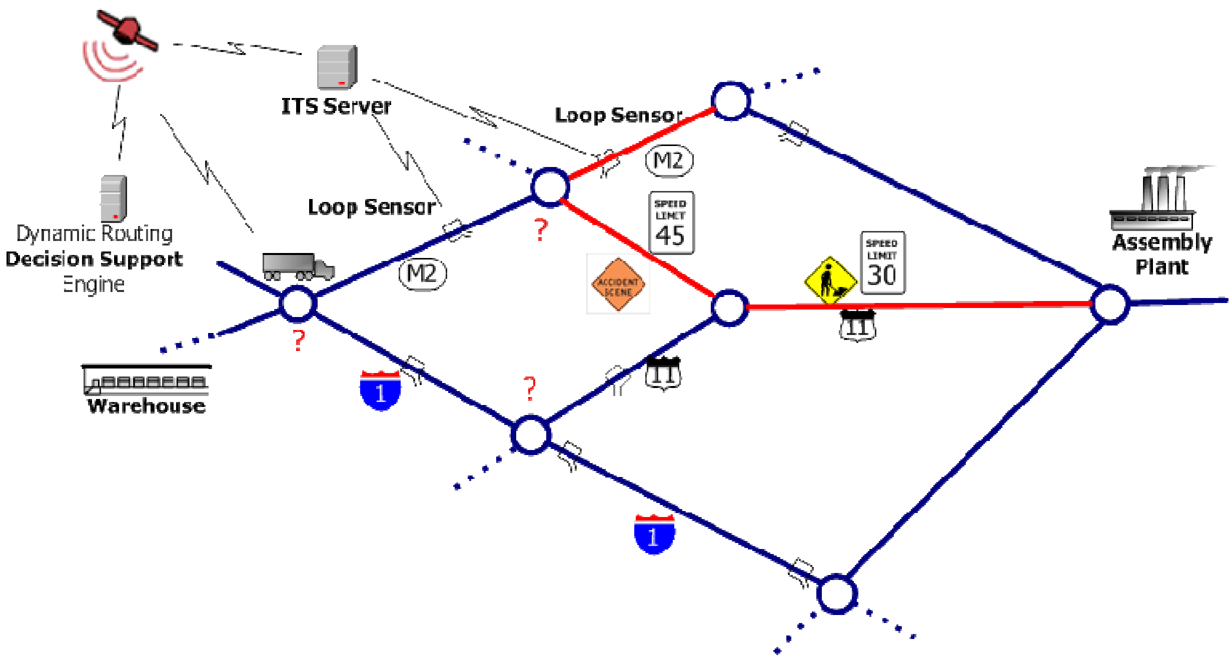

Figure 16. Recommended Framework for Data Communication and Decision Support Integration

13. BIBLIOGRAPHY

[1] Chang T-S, Wan Y-W, OOI WT. A stochastic dynamic traveling salesman problem with hard time windows. European Journal of Operational Research 2009;198(3):748-59.

[2] Taniguchi E, Thompson RG, Yamada T, Duin JHR. City logistics: Network modelling and Intelligent Transportation Systems. Pergamon: Elsevier Science; 2001.

[3] Schrank D, Lomax T. The 2007 annual urban mobility report. Urban Mobility Information: Texas Transportation Institute; 2007.

[4] Texas Transportation Institute, Cambridge Systematics Inc. Monitoring Urban Freeways in 2003: Current conditions and trends from archived operations data. Washington, DC: US FHWA Office of Operations; 2004.

[5] Cambridge Systematics Inc., Texas Transportation Institute. Traffic congestion and reliability: Linking solutions to problems. Washington, D.C: US FHWA Office of Operations; 2004.

[6] Sigal CE, Pritsker AAB, Solberg JJ. The stochastic shortest route problem. Operations Research 1980;28:1122-9.

[7] Frank H. Shortest paths in probabilistic graphs. Journal of Operational Research 1969;17:583-99.

[8] Loui RP. Optimal paths in graphs with stochastic or multidimensional weights. Communications of the ACM 1983;26(9):670-6.

[9] Eiger AP, Mirchandani P, Soroush H. Path preferences and optimal paths in probabilistic networks. Transportation Science 1985;19:75–84.

[10] Dijkstra E. A note on two problems in connection with graphs. Numerical Mathematics 1959;1:269-71.

[11] Hall R. The fastest path through a network with random time-dependent travel time. Transportation Science 1986;20(3):182-8.

[12] Bertsekas DP, Tsitsiklis J. An analysis of stochastic shortest path problems. Mathematics of Operations Research 1991;16:580-95.

[13] Fu L, Rilett LR. Expected shortest paths in dynamic and stochastic traffic networks. Transportation Research Part B 1998;32B(7):499-516.

[14] Bander JL, White III CC. A heuristic search approach for a nonstationary shortest path problem with terminal costs. Transportation Science 2002;36:218-30.

[15] Fu L. An adaptive routing algorithm for in vehicle route guidance systems with real-time information. Transportation Research Part B 2001;35B(8):749-65.

[16] Waller ST, Ziliaskopoulos AK. On the online shortest path problem with limited arc cost dependencies. Networks 2002;40 (4):216-27.

[17] Gao S, Chabini I. Optimal routing policy problems in stochastic time-dependent networks. Transportation Research Part B 2006;40B:93-122.

[18] Miller-Hooks ED, Mahmassani HS. Least possible time paths in stochastic, time-varying networks. Computers & Operations Research 1998;12:1107-25.

[19] Miller-Hooks ED, Mahmassani HS. Least expected time paths in stochastic, time-varying transportation networks. Transportation Science 2000;34 198-215.

[20] Psaraftis HN, Tsitsiklis JN. Dynamic shortest paths in acyclic networks with Markovian arc costs. Operations Research 1993;41:91-101.

[21] Polychronopoulos GH, Tsitsiklis JN. Stochastic shortest path problems with recourse. Networks 1996;27(2):133-43.

[22] Azaron A, Kianfar F. Dynamic shortest path in stochastic dynamic networks: Ship routing problem. European Journal of Operational Research 2003;144:138-56.

[23] Kim S, Lewis ME, White III CC. Optimal vehicle routing with real-time traffic information. IEEE Transactions on Intelligent Transportation Systems 2005;6(2):178-88.

[24] Kim S, Lewis ME, White III CC. State space reduction for non-stationary stochastic shortest path problems with real-time traffic congestion information. IEEE Transactions on Intelligent Transportation Systems 2005;6(3):273-84.

[25] Wirasinghe SC. Determination of traffic delays from shock-wave analysis. Transportation Research 1978;12:343-8.

[26] Al-Deek H, Garib A, Radwan AE. New method for estimating freeway incident congestion. Transportation Research Record 1995;1494:30-9.

[27] Mongeot H, Lesort JB. Analytical expressions of incident-induced flow dynamics perturbations. Transportation Research Record 2000;1710:58-68.

[28] Morales J. Analytical procedures for estimating freeway traffic congestion. Public Roads 1986;50(2):55-61.

[29] Fu L, Rilett LR. Real-time estimation of incident delay in dynamic and stochastic networks. Transportation Research Record 1997;1603:99-105.

[30] Cohen H, Southworth F. On the measurement and valuation of travel time variability due to incidents on freeways. Journal of Transportation And Statistics 1999;2(2):123-31.

[31] HCM. Highway Capacity Manual. Washington, D.C.: Transportation Research Board; 2000.

[32] Li J, Lan C-J, Gu X. Estimation of incident delay and its uncertainty on freeway networks. Transportation Research Record 2006;1959:37-45.

[33] Baykal-Gürsoy M, Xiao W, Ozbay K. Modeling traffic flow interrupted by incidents. European Journal of Operational Research 2008, In Press.

[34] Lindley JA. Urban freeway congestion: quantification of the problem and effectiveness of potential solutions. Institute of Transportation Engineers Journal 1987;57(1):27-32.

[35] Garib A, Radwan AE, Al-Deek H. Estimating magnitude and duration of incident delays. ASCE Journal of Transportation Engineering 1997;123(6):459-66.

[36] Sullivan EC. New model for predicting incidents and incident delay. ASCE Journal of Transportation Engineering 1997;123(5):267-75.

[37] Gaver DP. Highway delays resulting from flow stopping conditions. Journal of Applied Probability 1969;6:137-53.

[38] Golob TF, Recker WW, Leonard JD. An analysis of the severity and incident duration of truck-involved freeway accidents. Accident Analysis & Prevention 1987;19(4):375-95.

[39] Giuliano G. Incident characteristics, frequency, and duration on a high volume urban freeway. Transportation Research Part A 1989;23A(5):387-96.

[40] Khattak AJ, Schofer JL, Wang M-H. A simple time sequential procedure for predicting freeway incident duration. IVHS Journal 1995;2(2):113-38.

[41] Noland RB, Polak JW. Travel time variability: A review of theoretical and empirical issues. Transport Reviews 2002;22(1):39-54.

[42] Hensher D, Mannering F. Hazard-based duration models and their application to transport analysis. Transport Reviews 1994;14:63-82.

[43] Nam D, Mannering F. An exploratory hazard-based analysis of highway incident duration. Transportation Research Part A 2000;34A(2):85-102.

[44] Ferris MC, Ruszczynski A. Robust path choice in networks with failures. Networks 2000;35:181-94.
[45] Thomas BW, White III CC. The dynamic shortest path problem with anticipation. European Journal of Operational Research 2007;176:836-54.
[46] Bertsekas DP. Dynamic programming and optimal control. Athena Scientific; 2001.
[47] Ross SM. Introduction to Probability Models. Academic Press, Inc.; 2006.
[48] Verbeek JJ, Vlassis N, Kröse B. Efficient greedy learning of Gaussian Mixture Models. Neural Computation 2003;5(2):469-85.

14. LIST OF ACRONYMS

ADP	Adaptive Decision Policy
ATIS	Advanced Traveler Information Systems
DP	Dynamic Programming
GMM	Gaussian mixture model
GPS	Global Positioning System
ITS	Intelligent Transportation Systems
JIT	Just-in-Time
LET	Least Expected Travel
MDOT	Michigan Department of Transportation
MDP	Markov decision process
MITS	Michigan Intelligent Transportation System Center
MITS	Michigan Intelligent Transportations Systems
OD	Origin and Destination
OEM	Original equipment manufacturers
SDP	Stochastic Dynamic Programming
SP	Shortest Path
STD-SP	Stochastic Time-Dependent Shortest Path
vph	Vehicle per hour