Real-Time Traffic Network State Estimation and Prediction with Decision Support Capabilities: Application to Integrated Corridor Management

Hossein Hashemi
Graduate Research Assistant
shashemi@smu.edu

Khaled Abdelghany
(Corresponding Author)
Associate Professor and Chair
khaled@lyle.smu.edu

Ahmed Hassan
Visiting Assistant Professor
aehassan@lyle.smu.edu

M. Maverick Lezar
Undergraduate Research Assistant
mlezar@smu.edu

Department of Civil and Environmental Engineering
Southern Methodist University
P.O. Box 750340
Dallas, TX 75275-0340
Phone: (214) 768-3024
Fax: (214) 768-2164

Tables and Figures: 7
Text: 5729
Total: 7479

Prepared for presentation at the 92nd Annual Meeting of the Transportation Research Board, January 2013, Washington D.C. and publication in Transportation Research Record: Journal of Transportation Research Board

August 1, 2012
Real-Time Traffic Network State Estimation and Prediction with Decision Support Capabilities: Application to Integrated Corridor Management

ABSTRACT

This paper presents a real-time traffic network state estimation and prediction system with built-in decision support capabilities for traffic network management. The system seeks to provide traffic network managers with the capabilities to estimate the current network conditions, predict congestion dynamics, and generate efficient traffic management schemes for recurrent and non-recurrent congestion situations. The system adopts a closed-loop rolling horizon framework in which network state estimation and prediction modules are integrated. The system is applied in the context of Integrated Corridor Management (ICM), which is envisioned to provide a system-based approach for managing congested urban corridors. A genetic algorithm methodology is developed to generate efficient traffic management schemes that integrate preapproved control actions by all managing agencies. The system is applied to a section of a commuter corridor in Dallas, Texas. The results show the ability of the system to improve the overall network performance during hypothetical incident scenarios.

Keywords: Real-Time Traffic Network Management, Estimation and Prediction, Dynamic Traffic Assignment, and Genetic Algorithms
Real-Time Traffic Network State Estimation and Prediction with Decision Support Capabilities: Application to Integrated Corridor Management

BACKGROUND

Traffic congestion has reached alarming levels in most urban areas. With the limited ability to expand the physical capacity of the roadway network to meet the ever-growing travel demand, there are increasing calls to develop Advanced Traffic Management Systems (ATMS) to achieve optimal utilization of the roadway network. Such systems are envisioned to provide traffic network managers with decision support capabilities through developing efficient traffic management schemes that integrate a wide range of traffic control and traveler information strategies. Nonetheless, developing such capabilities requires modeling the traffic network at high fidelity by capturing the tempo-spatial demand-supply interactions and associated congestion phenomena. In addition, it requires short-term prediction of the network traffic dynamics in order to provide proactive traffic management strategies. Considering their numerous offline applications, simulation-based Dynamic Traffic Assignment (DTA) models have proved to be a valuable tool for modeling congestion dynamics in large-scale urban transportation networks. Thus, extensive research effort has been devoted to extending the capabilities of these models for real-time traffic management applications (1-4). In addition, effort has been devoted to developing other essential complementary modules including traffic data fusion and consistency checking (5-7), Origin-Destination (OD) demand estimation and prediction (8-13), and optimal traffic management (14-27).

The two pioneer systems, DYNASMART-X and DYNAMIT, are examples of DTA real-time simulation-based traffic management systems. They integrate capabilities for traffic network state estimation and prediction, and real-time traffic network management. However, developed traffic management strategies using these systems have primarily focused on providing either descriptive or normative route guidance strategies, and less on integrating routing strategies with the optimal setting of the traffic control devices. Abdelfatah and Mahmassani (1998) presented a modeling framework for solving the system optimal time-dependent path assignment and signal timing (14). While the solution provides a benchmark to evaluate other traffic management strategies, the assumption that all drivers comply with the provided normative route guidance is unrealistic. Additionally, the developed solution algorithm of the problem is computationally cumbersome which precludes its application for real-time applications. Abdelghany et al. (1999) introduced the path-based signal coordination strategy to provide additional capacity along dominant routes used by travelers diverted from the freeway due to accidents (15). A traffic network prediction module is used to determine the dominant diversion routes along which coordination is provided. Mirchandani and Head (1998) developed a real-time traffic signal control system based on adaptive signal timing using a hierarchical architecture (16). The system predicts traffic flow at appropriate resolution levels to enable proactive control.

Park et al. (1999) presented a Genetic Algorithm (GA) traffic signal optimization program for oversaturated conditions (17). The approach optimizes signal timing by using essential components of traffic signal such as cycle length, green split, offset, and phase sequence. Abu-Lebdeh and Benekohal (2003) presented a similar approach which adopts a genetic algorithm to determine the optimal signal setting in a traffic network (18). Ceylan and Bell (2004) used a bi-level optimization approach to solve the combination of signal timing and traffic assignment problem (19). The lower level obtains equilibrium link flows based on drivers’
routing, while the upper level solves the signal timing using GA approach. Varia and Dhingra (2004) employed a simulation-based approach to solve the dynamic system optimal traffic assignment in a congested network with signalized intersections (20). The paper adopts GA to minimize the travel time and optimize signal timing simultaneously. Lee and et al. (2005) proposed an optimization approach for adaptive traffic signal control using GA (21). The adaptive system generates an efficient signal timing strategy to respond to changing temporal-spatial traffic demand. The system is not evaluated in oversaturated conditions as it lacks the capability to adequately present queue spillback from adjacent intersections. Teklu et al. (2007) extends the GA approach for optimizing traffic control signals while considering drivers rerouting behavior (22). Choy et al. (2003) presented a cooperative, hybrid agent architecture for real-time traffic control (23). The problem is divided into various subproblems managed by an agent with a fuzzy neural decision making module. A cooperative distributed problem solving approach is considered to achieve coordination between agents. Etemadnia et al. (2012) presented an autonomic architecture for traffic network management (24). Similar to the work of Choy et al. (2003), the network is divided into subareas where a controller manages each area (23). Different team formation strategies are evaluated to determine the best cooperation schemes among controllers. Finally, traffic management schemes have also extended to optimize the operation of ramp metering systems. For example, Papageorgiou et al. (1990) presented a real-time coordinated, feedback ramp metering strategy (25). The proposed approach was shown to be more efficient than local ramp metering strategies when there are multiple bottlenecks on the freeway or restricted ramp storage spaces. Papamichail et al. (2010) used a new heuristic traffic-responsive feedback control strategy that coordinates local ramp metering actions along a freeway network (26). The developed strategy was shown to increase the traffic throughput and reduce the travel time.

This paper presents a real-time traffic network state estimation and prediction system with built-in decision support capabilities for traffic network management. The system seeks to provide traffic network managers with the capabilities to estimate the current network conditions, predict congestion dynamics, and generate efficient traffic management schemes for recurrent and non-recurrent congestion situations. The system adopts a closed-loop rolling horizon framework, which integrates network state estimation and prediction modules. The network state estimation module is in the form of a real-time simulation-based DTA model which is synchronized with the real clock and receives real-time traffic data updates to ensure consistency between the estimated and observed states. The prediction module periodically activates another instance of the network simulator running faster than real-time to predict the network state over a predefined horizon. Given the predicted network conditions, a traffic management scheme generator is activated to provide an efficient traffic management scheme that is consistent with the predicted network conditions (27).

The system is applied in the context of Integrated Corridor Management (ICM), which is envisioned to provide a system-based approach for managing congested urban corridors. It integrates the operations of multiple adjacent multimodal networks and facilities typically managed by multiple agencies with limited or no coordination (28). It seeks to improve the overall corridor performance through developing integrated traffic management schemes jointly implemented by the different agencies. A key factor to the success of ICM schemes is that they must be approved by all agencies. An agency approves a scheme only if all control actions in that scheme are feasible to implement. In addition, the plan must not result in deteriorating the level of service on the network/facility managed by this agency. In most current practices, a set of
traffic management schemes are generated and a priori approved by all agencies. The generated schemes are usually designed to fit predetermined combinations of traffic flow patterns and incident scenarios that are frequently observed along the corridor. Nonetheless, several limitations are associated with such approach. First, traffic networks are highly dynamic with numerous sources of stochasticity on the demand and the supply sides. Determining common traffic flow patterns and incident scenarios could be a challenge especially without the availability of adequate historical data describing the network conditions over a reasonable period. Second, the approach requires considerable effort to develop efficient traffic management schemes for these scenarios, and to achieve consensus among all agencies to approve these schemes. These two tasks must be frequently updated to account for changes in the network traffic flow pattern associated with growth/decline in the economical activities in the region. Finally, the effectiveness of the developed schemes to manage a new traffic flow pattern not a priori considered is questionable.

This research introduces an alternative approach in which agencies are assumed to provide pre-approved control actions that could be implemented in their jurisdictions. The system develops a traffic management scheme using these control actions. The system’s traffic management module adopts a genetic algorithm methodology in which a traffic management scheme is modeled as a chromosome. Genes in this chromosome represent the set of control actions that constitute the scheme. The prediction module is used to evaluate the performance (fitness) of the generated schemes to determine the scheme with the best performance.

The paper contributes to the literature in several aspects. First, the developed methodology suits current practice in regional traffic management where multiple agencies are involved. Second, the iterative use of traffic network simulation to evaluate the performance of the generated traffic management schemes ensures that the optimal scheme is consistent with the drivers’ route choice behavior. Third, the framework demonstrates a closed-loop implementation of the traffic management process. Thus, it demonstrates the importance of providing responsive traffic management schemes to cope with changes in the current network conditions associated with previously implemented control actions. Additionally, the framework presents a moving horizon approach, more suitable for real-time traffic management applications, for reporting the network performance. Finally, the framework can integrate a wide range of control actions implemented through different control devices.

The next section describes the overall framework of the developed real-time traffic management system. Next, the decision support capabilities of the system are described. Then, the results of a set of simulation-based experiments conducted to evaluate the proposed system are presented. Finally, concluding comments and considered research extensions are discussed.

OVERALL FRAMEWORK OF THE REAL-TIME TRAFFIC MANAGEMENT SYSTEM

Figure 1 illustrates the overall framework of the implemented real-time traffic management system. The system adopts a rolling horizon approach, which integrates network state estimation and prediction modules. The network state estimation module is synchronized with the real clock and provides an estimate of the current network conditions at any point in time. It consists of a real-time simulation-based DTA model capable of capturing the network congestion dynamics resulting from the network’s demand-supply interaction. The network simulation model consists of several modules including (a) demand generation; (b) travel behavior; (c) shortest path algorithm; (d) vehicle simulation; and (e) statistics collection.
model can accept as demand input a file listing the population of travelers, their attributes (including origin, destination, time of departure), or a time-dependent origin-destination trip table. Each generated traveler is assigned a set of attributes including his/her trip starting time, generation link, final destination, and a distinct identification number. Prevailing travel times on each link are estimated using the vehicle simulation component which adopts a mesoscopic simulation approach. The model utilizes measures that travelers may use as criteria to evaluate the different route options, including travel time, private car operation cost and highway tolls. These measures are combined in a generalized cost formula utilized in a route decision module activated at fixed intervals to provide travelers with a set of superior route options. The activation interval (usually in the range of three to ten minutes) is set such that the variation in network conditions is captured, while retaining desirable computational performance. Vehicles move in the network subject to the prevailing traffic conditions until they reach their final destinations along the pre-specified routes. If a driver receives en-route information and the traveler presumably complies with the provided information, the route of this traveler is updated accordingly. To ensure consistency between the simulation and the real network, the simulation model receives continuous data feeds in the form of speed and density observations for roadway links equipped with surveillance devices. A consistency checking module compares the simulation results to the received data and updates the model’s parameters to minimize any inconsistencies. As illustrated in Figure 1, the estimation module implements a moving horizon approach to report the estimated network performance. Statistics are continuously collected and are reported at each roll for a pre-defined past horizon (e.g., 30 minutes). This approach is considered more suitable for real-time applications in which the system is continuously running to monitor and manage the traffic network.

![Diagram](attachment:image.png)

**FIGURE 1: Real-time network state estimation and prediction framework**
The prediction module is periodically activated to predict the network conditions over a predefined horizon (30 minutes to one hour). The prediction module consists of another instance of the network simulation model running faster than real-time. The initial conditions for each prediction horizon are obtained from the estimation module which provides a snapshot of the network conditions at the start time of each prediction horizon. This snapshot defines the current location, speed, and assigned route for all vehicles in the network. The new vehicles to be loaded during the prediction horizon are obtained through either a dynamic demand estimation and prediction module, or historical demand data. Vehicles already in the network at the start of the prediction horizon and newly generated vehicles are simulated for the pre-specified horizon. In case a traffic management scheme is evaluated, the parameters of the control devices are updated to replicate this scheme. Statistics are collected for each vehicle to compute an array of performance measures that describes the overall network performance over the simulated horizon.

DECISION SUPPORT CAPABILITIES

As mentioned above, the system provides decision support capabilities by developing efficient traffic management schemes that are consistent with the predicted network conditions. The traffic management scheme determines the optimal settings for available traffic control devices in the network. In the current implementation, we adopt a GA approach to generate efficient traffic management schemes. GA is a machine-learning model, which adopts its behavior from the processes of evolution in nature. The process starts with the creation of a population of individuals represented by chromosomes. Chromosomes in this population continuously pass through a process of evolution to increase their fitness and adaptiveness to their environments. The evolution occurs by exchanging characteristics with other chromosomes of the population (crossover) or through self-changes in the chromosome (mutation). New generations appear from clones of the current population, in proportion to their fitness. The fitness is a single objective function of the chromosome that returns a numerical value to differentiate between good and bad chromosomes.

A traffic management scheme is modeled in the form of a chromosome. As illustrated in Figure 2, a gene in a chromosome defines a control action implemented as part of the scheme. A timing plan at a signalized intersection, a route diversion message on a dynamic message sign, a speed limit advisory, and a ramp meter flow rate are examples of possible control actions. Figure 2 gives examples of multiple schemes with different combinations of actions. The figure illustrates the structure of two parent schemes (1 and 2) in a generation. These two schemes are used to produce three new schemes as part of a subsequent generation. Children 1 and 2 are two new schemes formed by the crossover of Parents 1 and 2. A crossover point is randomly selected to execute the action exchange. Child 3 is obtained by the mutation of Parent 1 by randomly changing one or more of its actions. In the presented example, the DMS action is mutated.

Each scheme is evaluated by its fitness, measured as the average travel time over the prediction horizon when the scheme represented by this chromosome is implemented. The prediction module is activated to estimate the average traveler travel time for each considered scheme. The traffic network is simulated after modifying the settings of the control devices to represent their corresponding values in the generated scheme. The use of the DTA simulation model to evaluate the fitness of each scheme not only ensures accurate evaluate of the performance of the generated schemes but also ensures that the scheme is consistent with the drivers route choice behavior.
The steps of the used GA are as follows. First, an initial population and the fitness values of all its schemes are obtained. Schemes in the population are sorted based on their fitness value and top elements are used to produce the next generation using crossover and mutation strategies. Schemes in the new population are again evaluated and ranked. The process continues until the improvement in the fitness of the best scheme in two successive generations is smaller than a pre-defined threshold.

**Steps of the GA**

*Step 1:* Set iteration number $itr = 0$.

*Step 2:* Generate initial feasible population of traffic management schemes $P(itr)$.

*Step 3:* Using the prediction module, identify the fitness of each scheme in the population.

*Step 4:* While convergence is not obtained:

  *Step 4a:* Update the counter.
  
  *Step 4b:* Select a sub-population with the highest fitness from the population $P(itr-1)$. 

**FIGURE 2: GA representation of the traffic management schemes**
**Step 4c**: Elements of the sub-population are then used to generate a new population $P^{(itr)}$ using crossover and mutation strategies.

**Step 4d**: Each traffic management scheme in the population is evaluated using the simulation model.

**Step 5**: Output the traffic management scheme with the best fitness.

**EXPERIMENTS, RESULTS AND ANALYSIS**

This section describes the settings and results of the different experiments that are conducted to examine the performance of the developed system. As illustrated in Figure 3, the study area includes a section of Dallas North Tollway (DNT) surrounded by parallel and perpendicular arterials. The network consists of about 400 links and 150 junctions. A demand pattern that approximately represents a typical evening rush period is considered. Following this pattern, the majority of the simulated traffic presumably moves from the south boundaries to the north boundaries of the area (from Dallas’ downtown to its northern suburbs). Four congestion levels are considered, low, medium, medium-high and high, respectively. The demand is loaded over 2.5 hours and follows a symmetric triangular loading pattern. The average travel speed is recorded as 30, 25, 23, and 21 miles/hour at these demand levels. DNT represents the principal arterial in the network and carries about two-thirds of the traffic in the northbound direction. The other one-third of the northbound traffic uses all parallel arterials.

![FIGURE 3: The test-bed network](image)

Travelers are assumed to follow their historical routes and receive no pre-trip or in-vehicle information. To obtain these historical routes, travelers’ route choice is assumed to follows a user-equilibrium (UE) behavior. A simulation-assignment DTA model is used to...
determine the UE traffic flow pattern in the network for the entire horizon. The solution gives the departure time and the UE route for each traveler in the network. In case of a freeway incident, travelers are assumed to follow their historical routes unless a) they receive en-route information in the form of route guidance through active DMSs along the freeway, or b) they evaluate the congestion in their localities and divert to parallel arterials based on their historical knowledge of the network. In the first case, each DMS is assumed to divert a certain percentage of the traffic as a function of the travelers’ compliance rate with the provided information, and the period of time where the route advisory message displays on the DMS. In the second case, it is assumed that a certain percentage of the travelers will exit the freeway if the observed traffic density along the two subsequent freeway links reaches the jam density level. A hypothetical incident scenario is assumed to close two out of three lanes of the northbound freeway and to start 30 minutes after the start of demand loading and continue for one hour. In all experiments, three scenarios are compared. The first scenario represents the normal operation conditions (no incident). In the second scenario, travelers follow their historical routes and experience the delay due to the incident on the freeway while the traffic management system remains inactive. In the third scenario, the traffic management system is activated to manage the incident. In all cases, the moving horizon approach is used to report the network performance assuming a roll period of five minutes and a past horizon of 30 minutes. A traffic management scheme is assumed to consist of timing a subset of the signalized intersections along parallel arterials (30 intersections), and activating the DMS upstream of the incident. Five different timing plans are assumed to be pre-approved for each intersection. Additionally, seven possible activation periods, corresponding to different traffic diversion percentages, are assumed for this DMS. The considered diversion percentages range from 10% to 70% of the traffic on the freeway. The traffic management module is activated with the start of the incident through the end of the simulation horizon.

FIGURE 4: Effects of increasing the population size and number of generations
The first set of experiments examines the impact of the GA parameters (e.g., population size, \( \alpha \), and number of generations, \( \beta \)) on the quality of the generated traffic management schemes. The solution quality is compared where \( \alpha=5 \) and \( \beta=5 \) versus the case where \( \alpha=10 \) and \( \beta=10 \). In both cases, 10 different runs with different random seeds are conducted. In this set of experiments, the high demand level is assumed, and the traffic management module is activated every five minutes. Figure 4 presents the results of this set of experiments. It compares the percentage increase in the average vehicle travel time due to the incident to that of the normal operation conditions, both before and after activating the traffic management. The figure shows the average network performance of the ten simulation runs. As illustrated in the figure, activating the traffic management system results in a saving in the average travel time. In addition, increasing \( \alpha \) and \( \beta \) generally improves the quality of the generated traffic management schemes as depicted by more saving in the average travel time. At a point representing the peak congestion, a saving of about 5% is recorded when \( \alpha=10 \) and \( \beta=10 \). This saving is recorded as approximately 3% when \( \alpha=5 \) and \( \beta=5 \). In this experiment, an average execution time of 4 minutes is recorded for 30 minutes prediction horizon when \( \alpha=5 \) and \( \beta=5 \), which is increased to 8 minutes when \( \alpha=10 \) and \( \beta=10 \). The increase in the execution time is due to the increase in the number of schemes that need to be simulated.

The second set of experiments examines the ability of the system to produce efficient traffic management schemes at different network congestion levels. The four demand levels mentioned above are considered in this set of experiments. The traffic management module is activated every five minutes and the GA parameters are set at \( \alpha=5 \) and \( \beta=5 \). Figure 5 illustrates the results of this set of experiments. The normal operation scenario in the low demand case is considered as a base of comparison. As shown in the figure, the system is able to provide efficient schemes at the different demand levels. For all cases, a saving in the average travel time is recorded compared to the corresponding do-nothing scenario. In general, more travel time savings are recorded as the network congestion level increases.

![FIGURE 5: Effect of different demand levels](image-url)
The next set of experiments examines the effect of the spatial coverage of the control actions on the quality of the generated traffic management schemes. Traffic signals along the parallel arterials are considered in three different scenarios. The first scenario considers traffic signals along the frontage road (6 intersections). The second scenario considers traffic signals along the frontage road and the next two parallel arterials (21 intersections). The last scenario considers all 30 intersections. In this experiment, the traffic management module is activated every five minutes, and a high demand scenario is assumed. The results of this set of experiments are depicted in Figure 6, which presents the percentage saving in the average travel time for the three scenarios. The figure shows more saving in the average travel time with more spatial coverage of the control actions. A percentage saving in the average travel time over the entire horizon of about 9% is recorded for the case when all 30 intersections are considered. In the case where only signalized intersections along the frontage road are considered, the saving in the average travel time was limited to about 3%. More spatial coverage allows improvement of the signal timing plans and increases the capacity of more diversion routes during the incident. Improving the capacity of the diversion routes enhances the overall network throughput and results in more travel time saving as shown in the figure.

![Figure 6: Effect of the spatial coverage of the control actions](image)

The last set of experiments evaluates the effect of changing the frequency at which the traffic management module is activated. In this set of experiments, we compare different activation intervals; 3, 10, 15, and 20 minutes, respectively. The GA parameters are set at $\alpha=5$ and $\beta=5$, and a high demand scenario is assumed. Figure 7 gives the saving in the average travel time over the simulated horizon considering different activation intervals. As shown in the figure, the saving in the average travel time is highest at 5 and 10 minutes activation intervals. Savings in the average travel time of 9.13% and 9.26% are recorded at 5, and 10 minutes activation intervals, respectively. When using a small activation interval (e.g., 3 minutes) or a large activation interval (e.g., 15 and 20 minutes), the system results in less travel time saving.
For example, at the 20 minutes activation interval, the saving is recorded at 4.94%, and a saving of 6.91% is recorded for the 3 minutes activation interval case. The result of this experiment shows that frequent response to disruption in dynamic systems does not necessarily result in efficient recovery of these systems. The results indicate that better performance are achievable if implemented control actions are allowed more time to affect the system as planned. On the other hand, infrequent response could result in losing recovery opportunities. Thus, careful examination of the system dynamics is recommended through offline testing to determine the most suitable wait-and-act strategy (30).

FIGURE 7: Effect of the activation frequency of the traffic management module

CONCLUSIONS

This paper presents a real-time traffic management system for congested urban networks. The system estimates the current network conditions, provides short term prediction, and generates efficient traffic management schemes to improve the overall network performance. A genetic algorithm approach generates efficient traffic management schemes. Each scheme is modeled in the form of a chromosome, where gene in this chromosome defines a control action that can be implemented as part of the scheme. A timing plan at a signalized intersection, a route diversion message on a dynamic message sign, and a ramp metering strategy are examples of possible actions. The system is applied in the context of Integrated Corridor Management (ICM).
The results show the ability of the system to improve the overall network performance during a hypothetical incident scenario. Several extensions are considered for this research work. First, we consider different optimization-based formulations and solution methodologies to generate the optimal traffic management schemes and to compare their performances against the GA approach presented in this paper. Second, the developed framework provides a platform for testing the effectiveness of time-dependent decision making strategies that are commonly used in the control of complex dynamic systems. Finally, in the ICM context, it is assumed that all agencies collaborate to improve the overall corridor performance irrespective of the performance of their local jurisdictions. In most cases, agencies could require the generated schemes to maintain a certain level of service for a certain subarea or a facility. Considering such constraints while developing the traffic management schemes is another extension.

ACKNOWLEDGEMENTS

The research presented in this paper is supported in part by the Dallas Area Rapid Transit (DART) and the Bobby B. Lyle School of Engineering at Southern Methodist University. The authors would like to thank Koorosh Olyai at DART, Ed Seymour, Chris Poe and Minh Lee at Texas Transportation Institute, and Robert Saylor at City of Richardson for their valuable comments. The authors are of course responsible for all results and opinions expressed in this paper.

REFERENCES


**TRB 2013 Annual Meeting**

**Paper revised from original submittal.**


