Massachusetts Institute of Technology
PhD in Transportation Thesis Proposal

Traffic Management in Freeway Networks with Unreliable Link Capacities

Li Jin

Submitted to the thesis committee:
Prof. Saurabh Amin (advisor), Prof. Hamsa Balakrishnan, Prof. Demosthenis Teneketzis, and Prof. Nigel H. M. Wilson (chair)

Abstract

My PhD research focuses on traffic management in freeway networks subject to random disturbances that compromise link capacities. Such disturbances are responsible for a significant portion of congestion in transportation networks, which leads to enormous economic and environmental losses. My research develops a framework that systematically considers the presence of unreliable capacities in the assessment and design of traffic management strategies. Specifically, I model link capacities as variables that stochastically switch between a finite set of values. Two types of traffic management capabilities are considered, viz. routing (network level) and ramp metering (link level). At the network level, I combine the stochastic capacity model and the deterministic queueing model, and develop a novel piecewise-deterministic queueing (PDQ) model, which quantifies the network-wide impact of unreliable capacities. At the link level (ramp metering), I combine the capacity model with the cell transmission model, which results in a stochastic cell transmission model (SCTM). Based on the theory on convergence of Markov processes, I obtained results on stability and steady-state distribution of the PDQ model and the SCTM. I also explored the control design problem by formulating and solving a PDQ network optimization problem. My ongoing work is in three directions. First, I am working on calibrating the stochastic capacity model using real traffic data. Second, I am investigating the consistency between the PDQ model and the SCTM. Third, I am synthesizing the network level and the link level to design joint routing-ramp metering policies.

1 Introduction

Unreliable capacity is one of the major causes of congestion in transportation networks. Over 50% of highway congestion in the United States is non-recurrent, i.e. resulting from random capacity-reducing events such as crashes and work zones [29]. Weather-induced capacity reduction is responsible for over 30% of all flight delays in the US [33] and is a major challenge for air traffic management [28]. Urban public transportation such as metro lines faces similar challenges [5]. Such congestion leads to enormous economic and environmental losses [1, 31].
There have been both theoretical and practical efforts on the analysis of transportation systems with unreliable capacities. One the theoretical side, Ziliaskopoulos [34] proposed a dynamic programming-based method to deal with prescheduled capacity-reducing events. Como et al. [6] studied the resilience of flow networks with unreliable capacities, which focuses on the worst-case performance of the network. One the application side, there is extensive information in the literature on probabilistic models of capacity-reducing events [9, 20, 25, 30]. By incorporating such probabilistic models, Kurzhanskiy and Varaiya [22] developed a macrosimulation-based tool that generates random capacity-reducing events, based on which incident management strategies can be designed. Sumalee et al. [32] proposed a traffic flow model with stochastic capacities, also with an emphasis on its simulation abilities.

Interesting is that neither lines of work mentioned above attempted to study the impact of the stochastic nature of capacity-reducing events on performance from a theoretical perspective. To design traffic management strategies under unreliable capacities, I combine the stochastic capacity model with dynamic flow models to estimate congestion. I propose a two-level hierarchy for control design. At the link level, I developed a stochastic cell transmission model (SCTM), which is a detailed model for ramp metering policy design. At the network level, I propose a piecewise-deterministic queueing (PDQ) model, which is an abstract model for routing policy design. My PhD research focuses on synthesizing the link-level and the network-level analyses to jointly design coordinated routing-ramp metering policies.

This proposal is organized as follows. In Sec. 2, I will motivate my research via an example and overview my approach. In Sec. 3, I will summarize the results that I have obtained so far, with an emphasis on the network-level analysis based on the PDQ model. In Sec. 4, I will introduce my plan for the capacity model calibration, link-level analysis, and joint traffic management policy design, with preliminary results and outlines for next steps. In this proposal, I restrict my analysis to freeway traffic networks. However, I also expect potential application of my approach to other transportation networks.

## 2 Objectives and Hypotheses

### 2.1 Motivating example

First, consider a freeway segment as schematically shown in Fig. 1(a). Traffic arrives at rate \( f \) per unit time. The freeway has a nominal capacity of 1 per unit time. However, accidents frequently happen. When an accident occurs and leads to lane blockage, the freeway’s capacity drops by half to 0.5. On average, accidents occur at rate 1 per unit time. The average duration of the accident-induced lane blockage is 1 unit of time. Here are two questions of interest:

(i) How much traffic can this freeway segment actually handle?
(ii) How much delay do the accidents cause?

Question (i) motivates a capacity model as follows. I use the term *saturation rate* to refer to the freeway’s capability of accommodating incoming flow. I assume that a freeway
may switch between a set of modes, and every mode is associated with a fixed saturation rate. According to this assumption, the saturation rate is a piecewise-constant signal. Furthermore, I assume that intermodal transitions are governed by a continuous-time Markov chain defined over the set of modes.

Question (ii) entails further modeling effort and analysis. I consider two models for this purpose. First, I combine the cell transmission model (CTM, see [7]) and the stochastic capacity model, which results in a stochastic CTM (SCTM1). The CTM is a standard model for freeway traffic, and is used for simulation-based analysis in this research. I also developed a novel piecewise-deterministic queueing (PDQ2) model. The PDQ model can be viewed as an abstraction of the SCTM, which enables more tractable analysis.

The example above can be extended to the network setting. Consider the network of two parallel routes shown in Fig. 1(b). Node $s$ is the origin with a unit inflow. Node $t$ is the destination. Link 1 is subject to disturbances, while link 2 is not. Link 1 has a nominal capacity of 1. Disturbances occur in link 1 at rate 1, and clear also at rate 1. During disturbances, the capacity of link 1 drops to 0.5. Link 2 has a constant capacity of 0.75. In addition, the nominal travel times of links 1 and 2 are 1 and 2, respectively. Suppose that a traffic manager is able to decide the splitting factor $\gamma$, which is the fraction of demand that is sent to link 1. The traffic manager has to respond to the following questions:

(iii) What is the range of $\gamma$ that the network allows?
(iv) What is the $\gamma$ that minimizes the network-wide travel cost?

The above questions can be approached using SCTM/PDQ networks, i.e. dynamical networks where every link is an SCTM/PDQ. Question (iii) actually involves the stability of the network. Simplistically, the network with some routing policy is stable if the delay (or queue length) do not grow unboundedly. Question (iv) is a network optimization problem. Assume that the total cost is the sum of the nominal costs and the delay-induced costs. Thus, the optimization problem is essentially a tradeoff between efficiency (nominal costs) and reliability (delay-induced costs).

Note that real freeway networks are more complex than the abstract setting above. A main complication is that two types of control capabilities are involved, viz. routing and

---

1 The term “SCTM” has been used to refer to CTMs with unreliable parameters [32]. The CTM with unreliable capacities can be viewed a variation of the SCTM in the sense of [32].

2 This proposal also use the term “PDQs” (piecewise-deterministic queues) to refer to stochastic switching systems with dynamics specified the PDQ model.
ramp metering. Although both control capabilities can be considered as “routing” [34], they are very different in practice. Next, I will introduce my approach to establishing the link between control capabilities and performance metrics.

2.2 Research approach

I propose a framework for freeway traffic management as illustrated in Fig. 2. Traffic data from actual freeways can be used for calibration of the stochastic capacity model. Synthesis of the stochastic capacity model and the CTM enables macrosimulation that estimates impacts of unreliable capacities on freeway performance. The SCTM is particularly useful for design of ramp metering control.

However, although the SCTM capacities gives satisfactory simulation capabilities, it provides limited mathematical insights and tractable results, especially in network settings. Therefore, I developed the PDQ model, which abstracts the SCTM. The PDQ model builds on the deterministic queueing (DQ) model proposed by May [24]. The DQ model has been widely utilized for estimate of congestion due to demand-capacity imbalance [2, 27, 26].

The essence of the DQ model is that the rate of change of the traffic volume (queue) in a link is the difference between the inflow and the saturation rate. The DQ model abstracts the SCTM by simplifying the flow-density relationship and the spatial inhomogeneity within freeway sections. The PDQ model combines the stochastic saturation rate model and the DQ model. Between intermodal transitions, the queue evolves deterministically, while the transitions are stochastic.

The main hypothesis of this framework is that the joint routing-ramp metering design problem can be studied in a hierarchical manner. Routing typically depends on the network-wide, aggregate behavior of traffic. In contrast, ramp metering typically depends on local traffic condition and is largely distributed. Therefore, I address the routing problem based on the abstract PDQ model, and study the ramp metering problem based on the detailed SCTM.

Both the PDQ model and the SCTM are piecewise-deterministic Markov processes (PDMPs, see [8]), a class of hybrid-state Markov processes. In fact, Questions (i)–(iv)
can be answered by studying the long-time properties of both models, especially

(v) stability conditions for PDQ/SCTM with routing policies, and
(vi) invariant probability measure of PDQ/SCTM systems.

Analysis of the PDQ/SCTM model is built on recently published results on long-time properties of PDMPs [3, 12]. I derived a necessary condition and a sufficient condition for PDQ stability, and found their equivalence under particular assumptions. I also analytically computed the steady-state distribution of the PDQ model. Long-time properties of the SCTM is approached by comparison with the PDQ model and by numerical simulations.

Based on the framework in Fig. 2, I propose the following parts in my PhD thesis:

Ch.1 Introduction: background, motivation, current practice, literature review, technical approach, and an overview.
Ch.2 Stochastic capacity model: definition, properties, data collection, calibration.
Ch.3 Piecewise-deterministic queueing model and its properties: definition, stability conditions, steady-state behavior, network extension, routing policy design.
Ch.4 Stochastic cell transmission model and its properties: definition, stability conditions, steady-state behavior, network extension, ramp metering design.
Ch.5 Joint routing-ramp metering: comparison between PDQ and SCTM, hierarchical approach.
Ch.6 Evaluation of design: travel cost, resilience/robustness, safety, etc.
Ch.7 Conclusion and future work.

3 Current Work

This section summarizes the main results obtained in the past two years [16], which are mainly associated with the network level (PDQ model and disturbance-aware routing) of the framework in Fig. 2. I will first construct the PDQ model, and then present a series of results regarding the properties of PDQs and their network extensions. I will also demonstrate the intuition of these results via the example settings introduced in Sec. 2.1.

3.1 PDQs and their network extension

Consider the PDQ system illustrated by Fig. 3(a). Traffic arrives at the system at the inflow rate \( f(t) \) at time \( t \in \mathbb{R}_+ \). Before departing, traffic may be temporarily held by the system, which results in a queue. Denote the queue size by \( q(t) \). Let \( u(t) \) denote the saturation rate, i.e. the rate at which the queue is discharged at time \( t \), and let \( r(t) \) denote the discharge rate, i.e. the rate at which traffic departs from the system. If \( q(t) = 0 \) and \( f(t) \leq u(t) \), we have \( r(t) = f(t) \); otherwise \( r(t) = u(t) \). Thus, the queue evolves according to

\[
\dot{q}(t) = f(t) - r(t) = \begin{cases} 
0, & q = 0(t), f(t) \leq u(t), \\
|f(t) - u(t)|, & o.w.
\end{cases}
\]


Figure 3: The PDQ model extends the classical DQ model in that the saturation rate is governed by a finite-state Markov chain.

I refer to Fig. 4 for an illustration of how the queue size $q(t)$ evolves with time. When $f(t) > u(t)$, the queue grows at rate $f(t) - u(t)$. When $f(t) < u(t)$, the queue decreases at rate $u(t) - f(t)$, until it vanishes. Following [24], I assume an infinite queueing space, so $q$ can take arbitrarily large values.

Figure 4: For a given inflow rate, the queue increases or decreases due to stochastic fluctuation of the saturation rate.

I assume that the saturation rate of the PDQ system stochastically switches between a finite set of values. Specifically, let $\mathcal{I}$ be the set of modes of the PDQ system and $z = |\mathcal{I}|$. Every mode $i \in \mathcal{I}$ corresponds to a saturation rate, denoted by $u^i$. To emphasize the dependence of the discharge rate $r$ on the mode $i$, the queue size $q$, and the inflow rate

\[^3\text{Although traffic consists of discrete particles, we have assumed that } q \text{ is real-valued. The continuous approximation is typically made for studying the long-time properties and control of queueing systems [13, 24, 27].}\]
\( f \), we express it as follows:

\[
r(i, q, f) = \begin{cases} 
  f, & q = 0, f \leq u^i, \\
  u^i, & \text{o.w.}
\end{cases}
\]  
\tag{2}

From (1) and (2), the rate of change of \( q \) is governed by a vector field \( F : \mathcal{I} \times \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R} \) defined as

\[
F(i, q, f) = f - r(i, q, f).
\]  
\tag{3}

Let \( i(t) \) denote the mode of the PDQ system at time \( t \). The evolution of \( i(t) \) is governed by a continuous-time, finite-state Markov process with state space \( \mathcal{I} \) and constant transition rates \( \{ \lambda_{ij}, i, j \in \mathcal{I} \} \). Let \( \nu_i = \sum_{j \in \mathcal{I}} \lambda_{ij} \), which is the rate at which the system leaves mode \( i \). Note that \( i(t) \) is a piecewise-constant signal; see Fig. 4. I assume that the embedded chain of the Markov process \( \{ i(t); t \in \mathbb{R}_+ \} \) is irreducible and positive recurrent [10]. Given an initial mode \( i_0 \in \mathcal{I} \) at \( t = t_0 = 0 \), let \( \{ t_k; k = 0, 1, \ldots \} \) be the times at which the intermodal transitions happen. Let \( i_{k-1} \) be the mode during \( [t_{k-1}, t_k) \) and \( s_k = t_k - t_{k-1} \). Then, \( s_k \) follows the exponential distribution with parameter \( \nu_{i_{k-1}} \). We can compactly write the transition rates in a \( z \times z \) matrix

\[
\Lambda = \begin{bmatrix}
-\nu_1 & \lambda_{12} & \cdots & \lambda_{1z} \\
\lambda_{21} & -\nu_2 & \cdots & \lambda_{2z} \\
\vdots & \vdots & \ddots & \vdots \\
\lambda_{z1} & \lambda_{z2} & \cdots & -\nu_z \\
\end{bmatrix}.
\]  
\tag{4}

A point worth clarifying here is the conceptual difference between the PDQ model and the stochastic queueing models (e.g. \( M/M/1 \) queues). The PDQ model focuses on the temporal fluctuation of capacity and ignores the heterogeneity within the traffic flow. In contrast, the stochastic queueing models emphasize the inherent randomness of individual customers’ properties, but the system’s behavior is typically time-homogeneous [4, 19].

The actual capability of a PDQ to accommodate incoming flow is quantified by the effective capacity, which is defined as the maximum (in the sense of limiting time-average value) inflow signal that the system can serve without the queue growing unboundedly. Assume that the limit

\[
\bar{f} = \lim_{t \to \infty} \frac{\int_{t'=0}^{t} f(\tau)d\tau}{t}
\]

exists, and we have the following definition:

**Definition 1 (Effective capacity)** The effective capacity \( \bar{u} \) of a PDQ is defined as

\[
\bar{u} = \sup \left\{ \bar{f} \left| \lim_{t \to \infty} q(t) < \infty, WP1, \forall (i_0, q_0) \in \text{Init} \right. \right\}.
\]

By mass conservation, \( \bar{u} \) is also the maximum time-average rate at which traffic can depart from the system. Intuitively, one may conjecture that \( \bar{u} \) is equal to the time-average value of \( u(t) \) as \( t \to \infty \), which turns out to be true (see Prop. 2).
Let \( p = [p_1, \ldots, p_z] \) be the limiting fraction of time that the system spends in each mode. By Thm. 7.2.7 in [10], \( p \) is the solution to the\footnote{The steady-state equations are given by:
\[ p \Lambda = 0, \quad p e = 1, \quad p \geq 0, \quad (5) \]}
steady-state equations
\[ p \Lambda = 0, \quad p e = 1, \quad p \geq 0, \quad (5) \]
where \( e = [1, \ldots, 1]^T \in \mathbb{R}_+^z \). Then, the limiting time-average saturation rate is \( p U \), where
\[ U = [u^1, \ldots, u^z]^T. \quad (6) \]

Then, we have the following result:

**Proposition 1 (Effective capacity)** The effective capacity \( \bar{u} \) of a PDQ is given by \( \bar{u} = p U \).

In this proposal, I assume that the inflow rate \( f \) is either static or dependent of the state \((i, q)\). In the later case, \( f \) is specified by some state-feedback control policy \( \phi: \mathcal{I} \times \mathbb{R}_+ \rightarrow \mathbb{R}_+ \), i.e. \( f(t) = \phi(i(t), q(t)) \). The next result gives a necessary condition for stability of a PDQ, i.e. the average inflow has to be less than the effective capacity.

**Proposition 2 (Necessary condition for stability)** If a PDQ with control policy \( \phi \) is stable and if the limit
\[ \bar{\phi} = \lim_{t \to \infty} \frac{\int_{\tau=0}^{t} \phi(i(\tau), q(\tau)) \, d\tau}{t} \]
exists, then \( \bar{\phi} < \bar{u} \) WP1, where \( \bar{u} \) is the effective capacity of the PDQ.

I also derived a sufficient condition for stability based on the Harris' theorem [3, 12]:

**Proposition 3 (Sufficient condition for stability)** A PDQ with a state-feedback control policy \( \phi: \mathcal{I} \times \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) is stable if

1. (Drift) There exist \( b > 0 \) and \( k = [k_1, \ldots, k_z]^T > 0 \) such that the Lyapunov function \( V(i, q) = k_1 e^{bq} \) satisfies
\[ \frac{dE[V]}{dt} \leq K - C V, \quad \forall q > 0, \quad (7) \]
for some \( K, C > 0 \).

2. (Minorization) There exist \( i^* \in \mathcal{I} \) and \( \epsilon > 0 \) such that \( \phi(i^*, q) - u^* \) for all \( q \in \mathbb{R}_+ \).

Now let us focus on a particular class of PDQ systems, i.e. bimodal PDQ (BPDQ) systems. A BPDQ is a PDQ switching between exactly two modes. BPDQ systems are particularly interesting, since they can model a transportation facility that randomly switches between a nominal mode and a perturbed mode with reduced capacity. BPDQ systems are the simplest model that allow both initiation and clearance of queues resulting from unreliable capacities. In addition, it might be difficult to calibrate more complex multi-modal PDQ models using readily available data on capacity fluctuations.

For individual BPDQs, I found that the steady-state distribution of the queue is the sum of a probability mass at 0 and a scaled exponential distribution:
Proposition 4 (Invariant measure) The invariant probability measure \( p \) of a BPDQ with static inflow \( f \) such that \( u^1 < f < u \) is given by a pdf \( f : I \times Q \rightarrow \mathbb{R}_+ \) given by

\[
f(1, q) = \begin{cases} 
  z_1 \delta_0, & q = 0, \\
  a_1 e^{-\frac{q}{\theta}}, & q > 0,
\end{cases}
\]

\[
f(2, q) = a_2 e^{-\frac{q}{\theta}}, \quad q \geq 0,
\]

where

\[
z_1 = \frac{1}{\lambda + \mu} \left( \mu - \lambda \frac{f - u^2}{u^1 - f} \right),
\]

\[
a_1 = \frac{\lambda z_1}{u^1 - f}, \quad a_2 = \frac{\lambda z_1}{f - u^2}, \quad \theta = \left( \frac{\mu}{f - u^2} - \frac{\lambda}{u^1 - f} \right)^{-1},
\]

and \( \delta_0 \) is the Dirac delta function centered at 0.

Corollary 5 (Queue length) The expected queue length of a BPDQ is given by

\[
q(f) = \begin{cases} 
  0, & 0 \leq f \leq u^2; \\
  (a_1 + a_2)\theta^2, & u^2 < f < u; \\
  \infty, & f \geq u.
\end{cases}
\]

Furthermore, if \( u^2 < f < u \), we have \( \sigma_q^2 = (a_1 + a_2)\theta^3 \).

---

Figure 5: Relation between average queue length and (static) inflow. Simulation results are consistent with theoretical predictions.

3.2 Network stability and optimization

Both the PDQ model and its properties can be extended to the network setting. Consider a PDQ network whose topology is described by an acyclic, single-origin-single-destination, directed graph \( G = (\mathcal{N}, \mathcal{E}) \), where \( \mathcal{N} \) is the set of nodes and \( \mathcal{E} \) is the set of links. Let \( n = |\mathcal{N}| \) and \( m = |\mathcal{E}| \). Every link \( e \in \mathcal{E} \) is a PDQ. The scalars \( q, u, \) and \( \phi \) in the individual link setting become vectors \( \mathbf{q}, \mathbf{u}, \) and \( \mathbf{\phi} \) in the network setting. The routing policy \( \phi \) is a mapping \( I \times \mathbb{R}_m^+ \rightarrow \mathbb{R}_m^+ \). We say that \( \phi \) is admissible if it satisfies mass conservation. Props. 2 and 3 can be extended to the network setting as follows:
Theorem 6 (Necessary condition for network stability) If a PDQ network with an admissible state-feedback routing policy is stable, then the limiting time-average inflow to every link is less than the link’s effective capacity.

Theorem 7 (Sufficient condition for network stability) A PDQ network with an admissible routing policy is stable if the following conditions hold:

i) (Drift) There exist constant vectors $\mathbf{b} \in \mathbb{R}^m_{>0}$ and $\mathbf{k} \in \mathbb{R}^z_{>0}$ such that the Lyapunov function $V(i, q) = k_i e^{\mathbf{b}^T q}$ satisfies

$$\frac{d\mathbb{E}[V]}{dt} \leq K - CV, \quad \forall q > 0,$$

for some $K, C > 0$.

ii) (Minorization) There exist $i^* \in I$ and $\epsilon > 0$ such that $\|\dot{q}\|_1 < -\epsilon$ for all $q \neq 0$.

Next, I illustrate the results on PDQ systems via the parallel-route example in Fig. 1(b). Fig. 3(b) shows the corresponding PDQ network. First, suppose that $\gamma$ is a constant of the traffic manager’s choice, which however cannot change over time. We examined 101 values of $\gamma$ in the interval $[0, 1]$ by implementing Thm. 7. This involves search for constants $b_k$ and $k_i$ that satisfies the drift condition, which was done by solving optimization problems. We found that the network is stable if $0.25 < \gamma < 0.75$. Note that the effective capacity of link 1 is exactly 0.75. This is not a coincidence:

Proposition 8 For a static inflow $f$, a BPDQ is stable if and only if $f < \overline{u}$.

To determine the $\gamma$ that minimizes total travel time $J$, note that $J$ can be rewritten as

$$J(\gamma) = \overline{q}_1(\gamma) + \gamma + \overline{q}_2(1 - \gamma) + 2(1 - \gamma) = \overline{q}_1(\gamma) + \overline{q}_2(1 - \gamma) - \gamma + 2,$$

where $\overline{q}_1$ and $\overline{q}_2$ can be computed by Corr. 5. Fig. 6(a) illustrates the optimization. The feasible set (i.e. stable range) for $\gamma$ is $(0.25, 0.75)$. The minimum of $J$ is attained at $\gamma^* = 0.57$. One can interpret the optimal solutions as follows: since the faster link 1 is subject to disturbances, part of the traffic should be allocated to the slower link 2 to reduce potential accident-induced congestion, even though the nominal capacity of the faster route is able to handle all the traffic. Thus, the tradeoff between efficiency and reliability is quantified.

Now suppose that the traffic manager is able to change $\gamma$ according to the mode. Let $\gamma_1$ be the splitting factor in mode 1, and $\gamma_2$ in mode 2. Note that, in this case, both links 1 and 2 can viewed as BPDQs. Proceeding as in the static routing case, we obtain the feasible set

$$\{[\gamma_1, \gamma_2]^T \in [0, 1]^2 : (\gamma_1 + \gamma_2)/2 < 0.75, (2 - \gamma_1 - \gamma_2)/2 < 0.75\},$$

4The optimization problems presented in this proposal were solved using YALMIP, a MATLAB-based optimization tool developed by Löfberg [23].
which corresponds to the unshaded region in Fig. 6(b). The optimal solution is \([\gamma^*_1, \gamma^*_2] = [1.00, 0.50]\). This solution implies that the traffic manager should send all traffic to link 1 in the nominal mode, and send only half traffic to link 1 in the accident mode. By doing this, the traffic manager does not allow any queues in either links.

Note that such intuitive routing policy does not always work, especially when the delay cost dominates the nominal cost. Suppose that we modify the setting by reducing the saturation rate of link 2 to 0.45 and letting the nominal travel time of link 2 to equal to that of link 1. In this case, queues are inevitable in either of the links, and the nominal cost becomes irrelevant since it is the the same for both links. The optimal solution becomes \([\gamma^*_1, \gamma^*_2] = [0.91, 0.51]\), as shown in Fig. 6(c). This result suggests that sharing delay in both links is favored over restricting delay in a single link.

4 Ongoing Work

I am extending my previous work in three directions. First, I am working on calibrating the stochastic capacity model using real traffic data. Second, I am investigating the consistency between the PDQ model and the SCTM. Third, I am synthesizing the network level and the link level to design joint routing-ramp metering policies.

4.1 Capacity model calibration

The key of my approach is the stochastic capacity model. Therefore, I am particularly interested in calibrating the capacity model using real traffic data. As a preliminary exploration, I studied a 1.6-mile freeway segment between downtown San Francisco and the San Francisco International Airport [17]. This segment is divided into two cells, with locations indicated on in Fig. 7(a) being the cell boundaries. The data come from the Caltrans Performance and Measurement System (PeMS).

I followed the methodology proposed in [9] to calibrate the fundamental diagrams for the normal case (i.e. \(a = 0\)). To estimate the capacity drop, we retrieved the traffic records during accidents, and determined the reduced capacity using the method in [9]. In Fig. 7(b), the points/curve labeled as “0” correspond to the data reported in the
normal condition. The “1” group is the data reported when there is exactly one accident in this freeway segment. The “2” group is the data reported when there are two or more accidents.

With the preliminary results, I will further study the temporal pattern of capacity. For a particular freeway segment, I will estimate its capacity over a period of time with an appropriate resolution (e.g., every hour). I will then apply machine learning algorithms (e.g., hidden-Markov model) to learn the modes and their respective capacity values. Next, I will examine the intermodal-transition times, and compare the empirical distribution with the hypothesized exponential distribution. Finally, I will validate the learned model on test networks and integrate it with the PDQ/SCTM.

4.2 Further analysis of PDQ and SCTM

Another topic of interest is formalization of the PDQ model as an abstraction of the SCTM. In my previous work [15], we developed the SCTM, a version of CTM with unreliable capacities. The SCTM partitions a freeway into a sequence of cells, as shown in Fig. 8. The vector $\rho = [\rho_1, \ldots, \rho_N]^T$ specifies the state of the freeway, where $\rho_i$ denotes traffic density in the $i$-th cell. Two laws govern the evolution of the traffic, viz. flow-density relation and mass conservation. The SCTM admits a triangular/trapezoidal flow-density relation specified by:

$$f_i = \min \{ \beta_i v_i \rho_i, F_i, w_{i+1}(\rho_{i+1} - \rho_{i+1}) \} ,$$  

(12)
where \( f \) is traffic flow, \( \beta \) is split coefficient, \( v \) is free-flow speed, \( F \) is flow capacity, \( w \) is congestion wave speed, \( \bar{\rho} \) is jam density, and the subscripts are cell indices. Mass conservation yields

\[
\dot{\bar{\rho}}_i(t) = f_{i-1}(t) + r_i(t) - f_i(t)/\beta_i, \quad i = 1, 2, \ldots, N, \quad t \in \mathbb{R}_+.
\]  

(13)

Hence (12) and (13) specifies the evolution of traffic flow. The input variables are the inflow vector \( r = [r_0, \ldots, r_N]^T \).

The capacities \( F_i \) of the SCTM are piecewise-constant signals that switch in a set of values according to Markov chains. The transition rates may or may not depend on the local traffic condition. We studied the sensitivity of delay with respect to parameters of the capacity model, viz. accident intensity (quantified by capacity drop) and accident frequency. Fig. 9(a) illustrates two interesting points. First, there is a threshold intensity (0.2 in this case), below which accidents have insignificant impact. Second, beyond the threshold, the mean VMT decreases approximately linearly with respect to \( \alpha \). The impact of accident rate \( \lambda \) is illustrated in Fig. 9(b). Incident rate is assumed to be proportional to local traffic density. As expected, mean VMT decreases with accident rate. Noteworthy is that the mean VMT is not very sensitive to accident rate. Therefore, throughput is more sensitive to accident intensity than to accident rate.

Ramp metering (i.e. adjusting \( r \)) also impacts throughput. Fig. 10(a) shows the fraction of recovered VMT due to metering at various entrances. The result implies that \( r_7 \) is the most efficient control variable in terms of throughput recovery. However, metering the wrong inflows can result in further loss of VMT. Fig. 10(b) shows the relation between mean VMT and \( r_7 \). At 500 to 800 VPH, the VMT is recovered by approximately 25%. In addition, without considering accidents, we would not expect any congestion until \( r_7 = 2400 \) VPH, but such a high input would actually double the VMT loss. Therefore, operating a freeway near its capacity is very risky under accident-prone conditions.

I am interested in whether the above observations can be extended to PDQ models. Such extension depends on whether the PDQ model abstracts the SCTM in a consistent manner. Specifically, I plan to study the questions below:
(i) How are the parameters of an SCTM related to those of its PDQ counterpart?
(ii) To what extent do the stability conditions for PDQs inform the stability of the SCTM? Can we apply the Harris’ theorem to SCTM as we did for PDQs? Are stability conditions for PDQs over-conservative or over-aggressive with respect to SCTMs?
(iii) Is the delay estimated using the PDQ an upper bound or a lower bound of the delay estimated using the corresponding SCTM?
(iv) In what particular cases would the behavior of a PDQ significantly differ from its SCTM counterpart? When would a PDQ-based decision fail in the SCTM setting?

Proper response to these questions will enable us to better understand and justify the performance of PDQ-based decisions, rather than resorting to the more sophisticated SCTM. To approach the answers, I will utilize known results on the long-time properties of the CTM [11, 15], and perhaps more general results on cooperative/monotone systems [14, 18]. I will proceed with a combination of analytical derivation and numerical simulation.

4.3 Joint routing-ramp metering design and evaluation

First, I plan to use our model to evaluate currently practiced traffic management policies under unreliable capacities. I plan to calibrate a test SCTM/PDQ network using data available from PeMS, including both the capacity model and the traffic flow model. I will try to reverse-engineer the current strategy. One way to do this is by inferring mainline and on-ramp inflows from aggregate flow data [21].

Second, I will try to demonstrate the potential improvements allowed by my approach. This will be a synthesis of the tools that I develop. I will start from the network level, addressing the routing problem under unreliable capacities. On top of the routing policies, I will design ramp metering policies at the link level. I am particularly interested in estimating the benefits of awareness of the unreliable capacities. Based on my preliminary analysis at both the network and the link levels, I hypothesize that ignoring capacity unreliability may lead to inefficient and even very costly decisions.
Finally, I will consider alternative evaluation metrics of traffic management policies. One alternative metric is robustness. For example, a delay-minimizing routing policy may lie on the boundary of the stable region (as in Fig. 6(b)), which means that an infinitesimal deviation in the implementation may affect the network stability. To fix this issue, we need to account for the robustness against errors from various sources. Another possible metric is safety. Sending traffic to a freeway with an accident is sometimes more costly due to the risk of secondary accidents. A third possible metric is the cost of control actions, e.g., redirection of mainline traffic and holding traffic at on-ramps.

References


[16] Li Jin and Saurabh Amin. Control of piecewise-deterministic queueing networks. working paper.


