Analysis of a Stochastic Switched Model of Freeway Traffic Incidents

Li Jin and Saurabh Amin

Abstract

This article models the interaction between freeway traffic dynamics and capacity-reducing incidents as a stochastic switched system, and analyzes its long-time properties. Incident events on a multi-cell freeway are modeled by a Poisson-like stochastic process. Randomness in the occurrence and clearance of incidents results in traffic dynamics that switch between a set of incident modes (discrete states). The rates of occurrence and clearance depend on the traffic densities (continuous state). The continuous state evolves in each incident mode according to mode-dependent macroscopic flow dynamics. At steady state, the system state resides in its accessible set, which supports an invariant probability measure. Behavior of the accessible set is studied in terms of inputs (on-ramp inflows) and incident parameters (incident intensity). An over-approximation of accessible set and some useful bounds on the performance metrics (throughput and travel time) are also derived using the limiting states of individual incident modes. These results provide following insights about the steady-state system behavior: (i) Expected loss of throughput increases with the incident intensity for fixed incident rate, but this loss is less sensitive to changes in the occurrence rate for fixed intensity; (ii) Operating an incident-prone freeway close to its capacity significantly increases the expected travel time; (iii) The impact of incidents reduces when certain inputs upstream of incident-prone sections are metered.

Index Terms

Stochastic switched systems, Freeway Incident management, Traffic control.

I. Introduction

Freeway traffic networks are prone to capacity disturbances, such as crashes, road damage, and other incident events. These disturbances introduce considerable *non-recurrent* congestion;

L. Jin and S. Amin are with the Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mails: jnl@mit.edu; amins@mit.edu)

i.e., congestion due to factors other than the regular fluctuations of demand [31]. To explore control policies that enable a freeway network to survive such disturbances, one has to study congestion dynamics due to the initiation of incidents and propagation of the induced congestion. This article introduces a stochastic switched model of freeway incident dynamics, and studies long-time properties of this model.

The current practice of freeway incident management largely utilizes *scenario-based* strategies. In this approach, traffic managers specify a set of *scenarios* based on the severity or likelihood of occurrence, and assess the likely impact on the current traffic conditions using simulations of traffic models [23], [35]. For high-impact scenarios, specific control strategies are implemented; however their design is typically ad hoc. The main limitation of this approach is its reactive nature; i.e. the response to an incident is triggered only after it has been detected. Moreover, a control strategy designed for a specific scenario might not be robust to a broader range of incident scenarios. Consequently, scenario-based approaches are prone to delayed or suboptimal responses.

In contrast, this article argues for a model-based control approach for incident management. Towards this end, it proposes a stochastic switched system to model random transitions in freeway traffic dynamics due to the occurrence and clearance of incidents. In this model, the occurrence (resp. clearance) of an incident triggers an instantaneous switch, which results in decrease (resp. increase) in the capacity at the location of incident. The capacity reduction during an incident is directly proportional to its intensity. At a given freeway location, incidents occur and clear (or resolve) in a Poisson-like manner. Between transitions, the traffic density (continuous state) evolves according to the cell transmission model (CTM) [9], where the capacity at each location is governed by the currently active incident event at that location.

We focus on analysis of the long-time behavior of the stochastic switched system. We first show that the *accessible set* assumes qualitatively different geometries depending on the input: a singleton or a connected set. When the accessible set is a connected set, a full characterization is hard, but we derive a useful over-approximation based on the properties of individual modes. Finally, we show that every invariant probability measure of the stochastic switched system is supported by the accessible set, and study the range of performance metrics (throughput and total travel time) at steady state.

Our model is a *stochastic switched system*. Recently, Benaïm et al. [3] examined the long-time

properties of this class of systems, specially piecewise-deterministic Markov process (PDMP) without continuous state reset. Our analysis utlizes their results (as well as some ideas from [2] and [15]) in the context of incident dynamics. Alternatively, one can also view the stochastic switched model as an SHS with a set of affine dynamics and with forced transitions. This affine representation has been reported in the context of real-time estimation and control [22], [34]. We mainly consider the former representation, while the latter is introduced in our earlier work [17].

A property of the CTM dynamics that is frequently utilized in our analysis is *monotonicity*.[1] Gomes et al. [14] utilized the monotonicity of the CTM and arrived at two main conclusions: first, given a stationary input, there always exists a set of equilibrium states; second, all trajectories converge to some equilibrium. We examine the relation between the accessible set and these equilibria, which we refer to as limiting states in this article. Coogan and Arcak [7] reported that network CTMs may not admit monotone dynamics. However, the lack of monotonicity does not affect the convergence (for acyclic networks), and in fact can be utilized to improve network throughput via ramp metering. Como et al. [5], [6] also utilized monotonicity to study the behavior of flow networks under perturbations. Their focus was analysis of throughput that can be achieved by distributed routing policies. In contrast to the robust routing framework of [5], [6], we focus on the analysis of stochastic occurrence and clearance of perturbations.

Our model is consistent with previously reported studies on freeway incident/accident statistics. Previous work (e.g. [13], [17], [18], [32]) consistently report that freeway incidents can be modeled as stochastic processes, typically as Poisson arrivals. Knoop [21] found that capacity reduction adequately captures the impact of incidents on freeway traffic. Khattak et al. [20] developed a stochastic queuing model that estimates the consequences of capacity reduction. Recent work by Miller and Gupta [28] used a classification model to assess the severity and induced delay due to reported incidents. Similar statistical learning procedures can be used to estimate the parameters of our SHS model.

The main contributions of our work are as follows. First, we propose a stochastic switched model for traffic dynamics in incident-prone freeways. The main idea of the switched model is an underlying stochastic process governing the initiation and termination of capacity-reducing

[1] in the sense of Hirsch [16]

incidents. The system randomly switches among a set of pre-identified dynamics, which models the impact due to capacity reduction/recovery. We use the CTM to estimate incident-induced congestion. (Sec. II)

Second, we analyze the steady-state behavior of the stochastic switched model. Behavior of the accessible set is studied in terms of inputs (on-ramp inflows) and incident parameters (incident intensity). An over-approximation of accessible set and some useful bounds on the performance metrics are also derived using the limiting states of individual incident modes. (Sec. III–V)

Third, we develop computational results that provide insights for incident management. We perform a simulation-based study to validate the theoretical results. We also investigate the sensitivity of performance metrics with respect to incident intensity and the incident rate, and explore (static) ramp metering strategies under incident-prone conditions. (Sec. VI)

II. Incident Model and Switched System

In this section, we construct a stochastic switched system to model the random transitions (jumps) in traffic dynamics due to the occurrence and clearance of incidents on a freeway.

A. Model background

1) Deterministic CTM: To model the evolution of traffic density (continuous state), we adopt the classical cell transmission model (CTM) [9]. Let us recall the deterministic (un-switched) CTM for an $N-$cell freeway; see Fig. 1. Let $x_i(t)$ denote the traffic density (in vehicles-per-mile, or vpm) in the i-th cell at time t. The $N-$dimensional state $x(t) = [x_1(t), x_2(t), \ldots, x_N(t)]^T$ is defined over the closed set $\mathcal{X} = \prod_{i=1}^{N}[0, \bar{x}_i] \subset \mathbb{R}^N$, where $\bar{x}_i$ is the i-th cell's jam density. With unit cell length, the state evolves as follows:

$$\begin{aligned} \dot{x}_i(t) &= f_{i-1}(t) + r_i(t) - f_i(t) - s_i(t) \\ &= f_{i-1}(t) + r_i(t) - f_i(t)/\beta_i,\ i = 1, 2, \ldots, N, \end{aligned} \tag{1}$$

where $f_i(t)$ is the flow from the $i-$th to the $(i+1)-$th cell, $r_i(t)$ is the on-ramp flow (input) into the i-th cell, and $s_i(t)$ is the off-ramp flow out of the i-th cell. We assume that for each cell i, the ratio $\beta_i \equiv f_i/(f_i + s_i)$ is a constant [14]. All flows are in vehicles-per-hour (vph). The $(N+1)-$dimensional input vector $r(t) = [r_0(t), r_1(t), \ldots, r_N(t)]^T$ is defined over the set $\mathcal{R} = \prod_{i=0}^{N}[0, \bar{r}_i]$, where $\bar{r}_i$ is the maximum *admissible* input for the $i-$th cell; see [14].

Let F_i^{nom} denote the *nominal capacity* of the $i-$th cell in the nominal (i.e. no-incident) conditions. For each cell i, we define the sending flow S_i and the receiving flow R_i as follows:

$$\begin{aligned} &S_0(x) = r_0,\ S_i(x) = \beta_i v_i x_i,\ i = 1, \dots, N \\ &R_i(x) = \omega_i(\bar{x}_i - x_i),\ i = 1, \dots, N, \end{aligned} \tag{2}$$

where v_i denotes the free-flow speed and ω_i the congestion wave speed, both in miles per hour (*mph*). The flow $f_i(t)$ is given by the flow function (i.e., the so-called *fundamental diagram*):

$$f_i(x) := \begin{cases} \min\{S_i(x), F_i^{\text{nom}}, R_{i+1}(x)\}, & 0 \le i < N \\ \min\{S_i(x), F_i^{\text{nom}}\}, & i = N. \end{cases} \tag{3}$$

Eqn. (1)–(3) define the deterministic CTM. We say that the i-th cell is *uncongested* if $x_i \le x_i^{\text{c}}$ and *congested* if $x_i > x_i^{\text{c}}$, where

$$x_i^{\text{c}} = \frac{F_i^{\text{nom}}}{\beta_i v_i} = \overline{x}_i - \frac{F_i^{\text{nom}}}{\beta_i w_i},\ i = 1, 2, \dots, N \tag{4}$$

is the *critical density* of cell i [14].

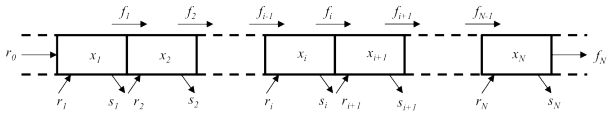
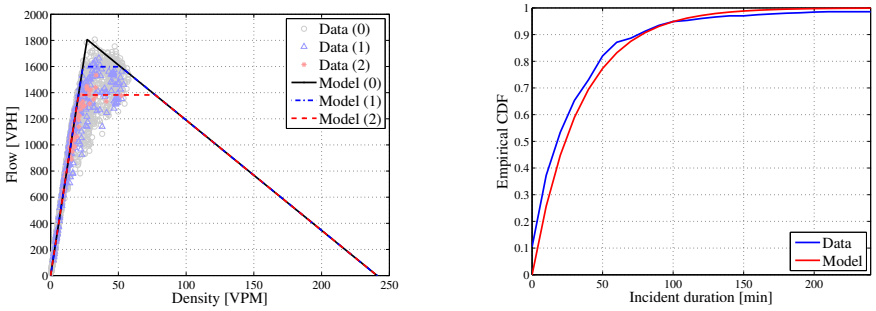


Fig. 1. A freeway segment with no incidents.

2) Data: To motivate our incident model, we studied real-world incidents that occurred on a segment of the US-101S freeway in the California Bay Area during a 3-month period. Our analysis is based on the incident and flow data obtained from the Caltrans Performance Measurement System (PeMS) [33].

First, we observe that the effective capacity during an incident, say, in the $i-$th cell, is less than the nominal value F_i^{nom}. However, incidents do not cause a significant change in other parameters of (2)-(3); i.e., v_i, ω_i, $\bar{x}_i$ are *relatively* unaffected by incidents. Fig. 2(a) shows calibrated fundamental diagrams for three types of events: no incident (model 0), low intensity incidents (model 1), and high intensity incidents (model 2).[2] The calibration was performed

[2]By low intensity incident events, we mean single accidents at a cell. By high intensity incident events, we mean events with multiple accidents at a given location, e.g. chain-reaction crashes.



(a) Capacity reduces during incidents (Asm. 1). (b) Incidents clear in a Poisson-like manner.

Fig. 2. Observations from incident data.

using the statistical procedure described in Dervisoglu et al. [11]. Fig. 2(a) indicates that the effective capacity is maximum during nominal condition ($1,800$ vph), but reduces to $1,600$ vph (resp. $1,400$ vph) during low (resp. high) intensity incident condition.

Second, we observe that the interarrival time and duration of incidents are roughly exponentially distributed, which suggests a Poisson-like model of incident occurrence/clearance events; see Fig. 2(b). This observation is consistent with the previously reported statistical models of incident rates; e.g. [30], [32]. In addition, previous literature suggests that the incident occurrence (resp. clearance) rates are positively (resp. negatively) correlated with the local traffic densities [1], [26].

B. Model

Our stochastic switched system is a hybrid process where the state $(X(t), Y(t))$ of the system has a continuous component (i.e., traffic densities) $X(t)$ taking values in the set $\mathcal{X} \subset \mathbb{R}^N$, and a discrete component taking values in a finite set $\mathcal{Y} := \{y_{\text{nom}}, \ldots, y_{\max}\}$. The set $\mathcal{Y}$ represents the set of incident configurations or equivalently, the set of *incident modes* for the $N-$cell freeway. For ease of presentation, we assume that every cell has only two states: with or without incident. Typically, incidents are concentrated only at certain locations of the freeway, which we call *incident hotspots*. We denote the set of incident hotspots by H (note that $H \subset \{1, 2, \ldots, N\}$). Thus, $|\mathcal{Y}| = 2^{|H|} \ll 2^N$. We label the nominal mode as y_{nom} and the mode with incidents in all incident hotspots as $y_{\max}$. The stochastic process $(X(t), Y(t))$ is given by:

$$X(t) = X(0) + \int_0^t G(X(\tau), Y(\tau), r(\tau))d\tau, \tag{5}$$

$$\begin{aligned}&\mathsf{Pr}\{Y(t+\delta)=y|Y(t)=w,X(s),Y(s),s\leq t\}\\&\qquad=\lambda(y,X(t),w)\delta+o(\delta),\ y\neq w,\\&X(0)=\bar{x}\in\mathcal{X},Y(0)=\bar{y}\in\mathcal{Y},\end{aligned}\tag{6}$$

where $G:\mathcal{X}\times\mathcal{Y}\times\mathcal{U}\to\mathcal{X}$ is a vector-valued function and $\lambda:(\mathcal{Y}\times\mathcal{X})\times\mathcal{Y}\to\mathbb{R}_+$ is the transition rule. We will precisely define G and λ in Sec. II-B1 and II-B2, respectively. We impose standard regularity assumptions [10]: for each $y\in\mathcal{Y}$, $G(\cdot,y,\cdot)$ is bounded, continuous and Lipschitz in the first argument uniformly with respect to the third.

We now specilize (5)-(6) to model switched dynamics. Starting from $(X(0),Y(0))=(\bar{x},\bar{y})$, the cumulative distribution function (CDF) of the time U at which the next mode transition occurs is given by:

$$\mathsf{F}_U^y(u)=1-\exp\left(-\int_{\tau=0}^{u}\lambda^y\left(\phi_\tau^y(\bar{x})\right)\tau d\tau\right),\ t\in\mathbb{R}_+,$$

where

$$\lambda^y(\cdot):=\sum_{w\in\mathcal{Y}}\lambda(y,\cdot,w)\tag{7}$$

is the rate at which the system leaves the incident mode y and $\phi_\tau^y(\bar{x})$ is the orbits (or integral curves) induced by $G(\cdot,y,r)$, i.e.

$$\frac{d}{d\tau}\phi_\tau^y(\bar{x},r)=G(\phi_\tau^y(\bar{x},r),y,r),\ \tau>0,\ \phi_0^y(\bar{x},r)=\bar{x}.$$

With a slight abuse of notation, we use $G(x,y)$ and $\phi_t^y(\bar{x})$ for short to denote the mode-dependent dynamics and orbits when the inputs are stationary.

Let U_n be the time between the $(n-1)$-th and the n-th transitions, and let $T_0=0$. Let $y_n\in\mathcal{Y}$ denote the realized incident mode between the $(n-1)$-th and the n-th transitions. Let T_n be the epoch, i.e. $\mathrm{T}_n=\sum_{l=0}^{n}U_l$; the discrete mode switches only at epochs. Then, $\{\mathrm{T}_n,n>0\}$ is a stochastic arrival process, and for the random process $\{Y(t),t\geq 0\}$, we have $Y(t)=y_{n-1}$, for $t\in[\mathrm{T}_{n-1},\mathrm{T}_n),\ n\in\mathbb{Z}_+$.

1) Effective capacity: For a given $y\in\mathcal{Y}$, let m_i^y denote the $i-$th cell's incident state: $m_i^y=1$ if there is an incident in the i-th cell, and $m_i^y=0$ otherwise. If the $i-$th cell is operating under an active incident, i.e., $m_i^y=1$ in mode $y\in\mathcal{Y}$, its *effective capacity*, denoted F_i^y, is less than the nominal capacity F_i^{nom} (recall Fig. 2(a)). The following assumption formally specifies such capacity reduction.

Assumption 1. *In an incident mode $y \in \mathcal{Y}$, the effective capacity upstream of the i-th cell reduces according to:*

$$F_{i-1}^y = (1-\alpha_{i-1}^y)F_{i-1}^{\text{nom}} = (1-\alpha_{i-1}m_i^y)F_{i-1}^{\text{nom}},\ i=1,\cdots,N, \tag{8}$$

where $\alpha_{i-1} \in (0,1)$ is the incident intensity parameter when the $i-$th cell is under the influence of incidents. Also define $\alpha_N^y = 0$ for all $y \in \mathcal{Y}$.

For each $y \in \mathcal{Y}$ and given $S_i(\cdot)$ and $R_i(\cdot)$ for $i = 0,\ldots,N$, the mode-dependent flow function is given by:

$$f_i(\cdot,y) = \begin{cases} \min\{S_i(\cdot), F_i^y, R_{i+1}(\cdot)\}, & 0 \le i < N; \\ \min\{S_i(\cdot), F_i^y\}, & i = N. \end{cases} \tag{9}$$

The following technical assumption guarantees that every configuration of incidents is allowed, i.e $|\mathcal{Y}| = 2^{|H|}$:

Assumption 2. *For every mode $y \in \mathcal{Y}$ and every incident hotspot $i \in H$, there exists $w \in \mathcal{Y}$ such that $m_i^w + m_i^y = 1$ and $m_j^w = m_j^y$ for all $j \neq i$.*

We call $\alpha^y := [\alpha_0^y,\ldots,\alpha_N^y]^T$ the incident intensity vector for mode y, and denote the collection of intensity vectors by $(\alpha^y)_{y\in\mathcal{Y}}$. For a random transition from mode y to mode w, if $m_i^y = 0$ and $m_i^w = 1$ (resp. $m_i^y = 1$ and $m_i^w = 0$), the occurrence (resp. clearance) of incident in the $i-$th cell results in an *instantaneous* reduction (resp. recovery) of capacity from F_i^y to F_i^w. Moreover, we define the *reduced capacity* as

$$\underline{F}_i := (1-\alpha_i)F_i^{\text{nom}},\ i=0,1,\cdots,N. \tag{10}$$

Under Asm. 1 and given admissible inputs $r = [r_0, r_1, \ldots, r_N]^T$, the dynamics of continuous state in the $i-$th cell, $G_i(x,y,r)$, is described as follows:

$$G_i(x,y,r) = f_{i-1}(x,y) + r_i - f_i(x,y)/\beta_i. \tag{11}$$

Then $G(x,y,r) = [G_1(x,y,r),\ldots,G_N(x,y,r)]^T$ denotes the continuous dynamics in a given mode y. Thus, (8)–(11) along with appropriate initial conditions specify the dynamics of continuous state in each mode. Note that, under our assumptions, the continuous state $X(t)$ is not reset after discrete mode transitions. This is consistent with the observation that, in most real-world scenarios, only a small number of vehicles are affected immediately after an incident event.

2) Transition rate: We adopt the standard regularity and irreducibility assumption about the transition rules.

Assumption 3. *The transition rate $\lambda : (\mathcal{Y}\times\mathcal{X})\times\mathcal{Y} \to \mathbb{R}_+$ is bounded, continuous, and Lipschitz (with respect to the continuous argument). Furthermore, we assume that, given x, the transition matrix $\Lambda = (\lambda(i,x,j))_{ij}$ is irreducible.*

The transition rate can be interpreted as the rate at which incidents occur and clear. Note that if the transition rates are independent of the densities, the inter-transition times will be exponentially distributed, and the occurrence-clearance cycles form a renewal process [12]; in addition, incidents at different locations will be statistically independent. In this article, we maintain the dependence of transition rates on densities. Other simpler models (e.g., [25], [32]) are special cases of our switched system model.

Thus, we view the *incident model* for an N-cell freeway segment as a triple $\langle \mathcal{Y}, (\alpha^y)_{y\in\mathcal{Y}}, \lambda\rangle$, where $\mathcal{Y}$ denotes the set of incident modes, $(\alpha^y)_{y\in\mathcal{Y}}$ the collection of incident intensity vectors, and λ the transition rule. Under these assumptions, it is easy to check that, $\{(X(t), Y(t)), t \geq 0\}$ is a Piecewise-Deterministic Markov process (PDMP) [10], and $X(t)$ is continuous with respect to t.

III. Main Results

A. Preliminaries

1) Invariant probability measure and accessible set: Following Benaïm et al. [3] (also see [27]), we construct a "thinning" representation of the PDMP $\{(X(t), Y(t)); t \geq 0\}$. Select a $\lambda > 0$ such that

$$\lambda > \max_{x\in\mathcal{X}, y\in\mathcal{Y}} \lambda^y(x),$$

where λ^y is given by (7). Given $x \in \mathcal{X}$, for every $y, w \in \mathcal{Y}$ (and $y \neq w$), let

$$Q(y,x,w) = \lambda(y,x,w)/\lambda,$$
$$Q(y,x,y) = 1 - \sum_{w\in\mathcal{Y}, w\neq y} \lambda(y,x,w)/\lambda.$$

For a function $g : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ that is bounded, measurable, and smooth with respect to x, define

$$Ag(x,y) = \langle G(x,y), \nabla_x g(x,y)^T\rangle, \tag{12}$$

$$Qg(x,y) = \sum_{w \in \mathcal{Y}} Q(y,x,w) g(x,w), \tag{13}$$

where $\nabla_x g(x,y)$ is a 1-by-N row vector. The infinitesimal generator [10] is given by

$$Lg(x,y) = Ag(x,y) + \lambda \left(Qg(x,y) - g(x,y) \right), \tag{14}$$

where the first term accounts for the deterministic dynamics and the second term accounts for the stochastic transitions.

Let $\{N(t); t \geq 0\}$ be a homogeneous Poisson process with rate λ, and let $\{U_n; n \in \mathbb{Z}_+\}$ and $\{T_n; n \in \mathbb{Z}_+\}$ be the inter-transition times and epochs, respectively. With the initial condition $(\widetilde{X}_0, \widetilde{Y}_0) \in \mathcal{X} \times \mathcal{Y}$, define $\{(\widetilde{X}_n, \widetilde{Y}_n); n \in \mathbb{Z}_+\}$ such that

$$\begin{aligned}
&\widetilde{X}_n = \phi_{U_n}^{\widetilde{Y}_{n-1}} \left(\widetilde{X}_{n-1} \right), \\
&\mathsf{Pr}\{\widetilde{Y}_n = w | \widetilde{X}_n, \widetilde{Y}_{n-1} = y\} = Q\left(y, \widetilde{X}_n, w \right), \ n = 1, 2, \ldots
\end{aligned}$$

The discrete stochastic process $\left(\widetilde{X}_n, \widetilde{Y}_n \right)$ is called the *embedded chain* of the PDMP [3]. The PDMP can be defined as the interpolation of the embedded chain:

$$\begin{aligned}
&X(0) = \widetilde{X}_0, \ Y(0) = \widetilde{Y}_0, \\
&X(t) = \phi_{t-T_{n-1}}^{\widetilde{Y}_{n-1}} \left(\widetilde{X}_{n-1} \right), \\
&Y(t) = \widetilde{Y}_{n-1}, \ \forall t \in [T_{n-1}, T_n), \ n = 1, 2, \ldots
\end{aligned}$$

Both the PDMP and its embedded chain can be characterized using transition operators K_t and $\widetilde{K}$ defiend as follows[3]

$$K_t g(x,y) = g\left(\phi_t^y(x), y \right), \tag{15}$$

$$\widetilde{K} g(x,y) = \int_{t=0}^{\infty} \lambda e^{-\lambda t} K_t g(x,y) dt. \tag{16}$$

Let $P = (P_t)_{t \geq 0}$ be the transition operator induced by $\{(X(t), Y(t)); \ t \geq 0\}$ such that

$$P_t g(x,y) = \mathsf{E}\left[g(X(t), Y(t)) | (X(0), Y(0)) = x, y \right].$$

[3]One can interpret K_t as an operator propagating the continuous dynamics and $\widetilde{K}$ as an operator accounting for the stochastic transitions in addition to the continuous dynamics.

Let $\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ be the set of probability measures on $\mathcal{X}\times\mathcal{Y}$. For $\mu\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ and $h\in L^1(\mu)$, we write

$$\mu h=\sum_{y\in\mathcal{Y}}\int_{x\in\mathcal{X}}h(x,y)\mu(dx,y).$$

We say that μ is an *invariant probability measure* for P_t if, for every $t\geq 0$, $\mu P_t=\mu$; the set of invariant probability measures is denoted by $\mathcal{P}_{\text{inv}}$. Similarly, define the transition operator $\widetilde{P}$ such that

$$\widetilde{P}g(x,y)=\mathsf{E}\left[g\left(\widetilde{X}_1,\widetilde{Y}_1\right)|(X(0),Y(0))=x,y\right]=\widetilde{K}Qg(x,y), \tag{17}$$

and the set of invariant probability measures $\widetilde{\mathcal{P}}_{\text{inv}}$ such that $\mu\widetilde{P}=\mu$ for $\mu\in\widetilde{\mathcal{P}}_{\text{inv}}$.

A natural candidate for the support of the invariant probability measure is the *accessible set* of the stochastic switched system, which is the closure of the set of states that can be reached from arbitrary initial conditions. Again, following [3], the accessible set is defined as follows:

Definition 1 (Accessible set)**.** *For all $n\in\mathbb{N}^+$, let $\mathscr{T}_n=\mathcal{Y}^{n+1}\times\mathbb{R}_+^n$. Given $(\boldsymbol{y},\mathbf{u})=((y_0,y_1,\cdots,y_n);(u_1,u_2,\cdots,u_n))\in\mathscr{T}_n$ and $x\in\mathcal{X}$, let $\boldsymbol{\phi}_{\mathbf{u}}^{\boldsymbol{y}}(x)=\phi_{u_n}^{y_{n-1}}\circ\phi_{u_{n-1}}^{y_{n-2}}\circ\cdots\circ\phi_{u_1}^{y_0}$, the positive trajectory of x is the set*

$$\gamma^+(x)=\left\{\boldsymbol{\phi}_{\mathbf{u}}^{\boldsymbol{y}}(x)|(\boldsymbol{y},\mathbf{u})\in\bigcup_{n\in\mathbb{N}^+}\mathscr{T}_n\right\}.$$

The accessible set $\Gamma\subset\mathcal{X}$ is a compact *set defined as*

$$\Gamma=\bigcap_{x\in\mathcal{X}}\overline{\gamma^+}(x). \tag{18}$$

2) Basic properties and metrics: Given a fixed $y\in\mathcal{Y}$, consider the model specified by (8)–(11) and a stationary input vector $r(t)\equiv[r_0,r_1,\ldots,r_N]^T,\forall t\in\mathbb{R}_+$, which induces an *uncapacitated* (i.e. without capacity constraints) equilibrium flow $f^{\text{unc}}(r)=[f_0^{\text{unc}}(r),f_1^{\text{unc}}(r),\ldots,f_N^{\text{unc}}(r)]^T$ given by

$$f_0^{\text{unc}}(r)=r_0,\ f_i^{\text{unc}}(r)=\beta_i(f_{i-1}^{\text{unc}}(r)+r_i),\ i=1,\ldots,N. \tag{19}$$

Also, define $z^{\text{unc}}=[z_1^{\text{unc}},z_2^{\text{unc}},\ldots,z_N^{\text{unc}}]^T$ such that $z_i^{\text{unc}}=f_i^{\text{unc}}/(\beta_i v_i)$ for every i.

Definition 2 (Input feasibility)**.** *If $f_i^{\text{unc}}(r)\leq F_i^y$ for every i, then r is* feasible for mode y*; if $f_i^{\text{unc}}(r)<F_i^y$ for every i, then r is* strictly feasible for mode y*; otherwise, r is* infeasible for mode y.

We call $z \in \mathcal{X}$ a *limiting state of mode* y if $G_i(z, y) \equiv 0$ for all i. Similarly, define *limiting flow of mode* y as $f^y = f(z, y)$. Note that f^y is generally different from f^{unc}, since f^y is subject to the capacity constraints. Limiting flow is unique, while limiting state is not necessarily unique [14]. We denote the set of limiting states of mode y as $\mathcal{Z}^y$. We now argue that the convergence results of the discrete-time CTM shown by [14] also hold for the individual modes of the switched system, using the monotonicity of the continuous-time dynamics. The key is to show that, for every i, $G_i(x, y)$ is nondecreasing in x_j for all $j \neq i$. Note that, for $i = 2, 3, \ldots, N-1$, we have $f_{i-1}(x, y) = \min\{\beta_{i-1} v_{i-1} x_{i-1}, F^y_{i-1}, w_i(\overline{x}_i - x_i)\}$, which is non-decreasing in x_j for all $j \neq i$. Similarly, f_i is non-increasing in x_j for all $j \neq i$. Thus, for $i = 2, 3, \ldots, N-1$, $G_i(x, y)$ is nondecreasing in x_j for all $j \neq i$. The same result can be shown for $i = 1$ and $i = N$. Thus, the dynamics are *cooperative* and thus *monotone* [16] (also see [19]). [4] The monotone CTM dynamics lead to convergence towards the limiting states.

Definition 3 (Bottleneck)**.** *Consider $y \in \mathcal{Y}$ and $z \in \mathcal{Z}^y$. The i-th cell is a* bottleneck in mode y *if $z_i \geq x^{\text{c}}_i$ and $z_{i+1} < x^{\text{c}}_i$, where x^{c}_i is as defined in* (4)*. The set of bottlenecks is denoted as $\mathcal{I}^y$.*

Remark 1. *Both input feasibility and bottleneck are defined with respect to stationary input, limiting states, and particular modes.*

Remark 2. *Gomes et al. [14] defined bottlenecks as locations where the capacity constraint is binding, which is essentially equivalent to our definition. They also showed that strictly feasible inputs never result in bottlenecks, and that infeasible inputs must induce bottlenecks. At a bottleneck, due to the insufficient capacity, demand upstream of this location cannot be fully served, which leads to congestion in the upstream cells and queues at the upstream entrances. In addition, a bottleneck restricts the flow into the downstream cells, which typically makes those cells uncongested.*

[4] Monotone dynamics are defined as follows. For given $y \in \mathcal{Y}$ and $x, \xi \in \mathcal{X}$, let $\{\phi^y_t(x); t \geq 0\}$ and $\{\phi^y_t(\xi); t \geq 0\}$ be the respective integral curves; if $x < \xi$ (respectively $x \leq \xi$), then $\phi^y_t(x) < \phi^y_t(\xi)$ (respectively $\phi^y_t(x) \leq \phi^y_t(\xi)$) for every $t > 0$, then the dynamics are monotone.

B. Statement of main results

Theorem 1 (Geometry of accessible set)**.** *For an N-cell freeway segment, consider the mode-dependent dynamics $G : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^N$, the incident model $\langle \mathcal{Y}, (\alpha^y)_{y\in\mathcal{Y}}, \lambda \rangle$, and a stationary input $r \in \mathcal{U}$. Let $\mathcal{Z}^{\mathrm{nom}}$ and f^{nom} be the set of limiting states and the limiting flow for the nominal mode $y_{\mathrm{nom}} \in \mathcal{Y}$, respectively. Let $\underline{F}_i$ be as defined in* (10)*.*

(i) If

$$(\forall i \in \{0, 1, \ldots, N\})\ f_i^{\mathrm{nom}} \leq \underline{F}_i \tag{20}$$

and if $\mathcal{Z}^{\mathrm{nom}}$ is a singleton, then $\Gamma = \mathcal{Z}^{\mathrm{nom}}$.

(ii) If

$$(\exists i \in \{0, 1, \ldots, N\})\ f_i^{\mathrm{nom}} > \underline{F}_i \tag{21}$$

and if $\exists y \in \mathcal{Y}$ such that $\mathcal{Z}^y$ is a singleton, then Γ is a connected set.

The above result can be interpreted as follows. In case (i), the capacity constraint is not active at any of the limiting states. Therefore, all modes have identical limiting states. This case arises when r is very low or very high. In case (ii), capacity reduction changes the limiting state. Consequently, the limiting states will be distinct across modes, and the accessible set is a connected set. This case arises when r is intermediate.

Remark 3. *There are two additional cases which Thm. 1 does not cover:*

(iii) if (20) *holds and if $\mathcal{Z}^{\mathrm{nom}}$ is not a singleton;*

(iv) if (21) *holds and if there exists no $y \in \mathcal{Y}$ such that $\mathcal{Z}^y$ is a singleton.*

We will elaborate on these two cases in Sec. IV-A. Since these two cases are of limited practical interest, unless otherwise stated, we will assume that, for every $y \in \mathcal{Y}$, $\mathcal{Z}^y$ is a singleton.

When the accessible set is a connected set, the analytical expression of the accessible set is difficult, especially for high-dimensional systems. In Thm. 2, we give an over-approximation of accessible set. Consider the case where the demand is feasible in the nominal mode but infeasible in some other modes.[5] To arrive at this over-approximation, we will show (in Sec. IV, Prop. 8)

[5]This is the case of interest, since incidents are the major concern; if the freeway were congested even in the absence of incidents, then excessive demand would be the primary source of congestion.

that, if r is infeasible for mode y, then z^y has only one bottleneck (see Def. 3), which we denote by $i(y_{\rm nom}, y)$.

Then, for mode y, define $a_1^y, \ldots, a_N^y$ and $b_1^y, \ldots, b_N^y$ as

$$a_i^y = \begin{cases} z_i^{\rm nom}, & i = 1, \ldots, i(y_{\rm nom}, y); \\ z_i^y, & i = i(y_{\rm nom}, y) + 1, \ldots, N; \end{cases} \tag{22}$$

$$b_i^y = \begin{cases} \bar{x}_N + r_N/\omega_N - F_N^y/(\beta_N \omega_N), & i = N; \\ \bar{x}_i + r_i/\omega_i - \overline{b}_i^y, & i = 1, \ldots, N-1, \end{cases} \tag{23}$$

where

$$\overline{b}_i^y = \frac{\min\{F_i^y, \omega_{i+1}(\bar{x}_{i+1} - b_{i+1})\}}{\beta_i \omega_i}, \ i = 1, \ldots, N-1. \tag{24}$$

The following result gives an N-dimensional orthotope that "attracts" all integral curves in all modes, and this orthotope is thus an over-approximation of the accessible set.

Theorem 2 (Over-approximation)**.** *Consider a stationary input $r \in \mathcal{U}$. If r is strictly feasible for $y_{\rm nom}$ but infeasible for some $y \in \mathcal{Y}$, and if $\mathcal{Z}^y$ is a singleton for all $y \in \mathcal{Y}$, then Γ is a connected set bounded by the orthotope specified by*

$$\mathcal{H} := \{x \in \mathcal{X} | a_i \leq x_i \leq b_i, i = 1, 2, \cdots, N\}, \tag{25}$$

where $a_i = \min_{y \in \mathcal{Y}}\{a_i^y\}$ and $b_i = \max_{y \in \mathcal{Y}}\{b_i^y\}$; a_i^y and b_i^y are given by (22)–(24)*.*

Remark 4. *The lower bounds are tight, while the upper bounds may or may not be tight.*

We consider two performance metrics that are typically defined for an instantaneous state [24]. The *throughput* is measured in vehicle-miles-traveled (VMT) and the *total travel time* is measured in vehicle-hours-traveled (VHT). For a given state $(x, y) \in \mathcal{X} \times \mathcal{Y}$, the instantaneous throughput J_F and instantaneous total travel time J_T are defined as follows:

$$J_F(x, y) = \sum_{i=1}^{N} f_i(x, y) l_i, \ J_T(x, y) = \sum_{i=1}^{N} x_i l_i, \tag{26}$$

where $f_i(x, y)$ is the flow function and l_i is the length of the i-th cell. For an invariant probability measure $\mu \in \mathcal{P}_{\rm inv}$, the expected performance metrics are denoted as J_F^μ and J_T^μ, respectively, i.e.

$$J_F^\mu = \mu J_F = \textstyle\sum_{y \in \mathcal{Y}} \int_{x \in \mathcal{X}} J_F(x, y) d\mu,$$

$$J_T^\mu = \mu J_T = \sum_{y\in\mathcal{Y}} \int_{x\in\mathcal{X}} J_T(x,y) d\mu.$$

The following result states that the accessible set supports invariant probability measures of the switched system, and that the bounding box given by Thm. 2 provides bounds on the performance metrics at steady state. Let $a = [a_1, \ldots, a_N]^T$ and $b = [b_1, \ldots, b_N]^T$. Let $V(\mathcal{H})$ be the set of vertices of $\mathcal{H}$. Define

$$\underline{J}_F = \min_{x\in V(\mathcal{H})} J_F(x, y_{\max}), \tag{27}$$

and define $\overline{J}_F$ as the optimal value of the linear program

$$\begin{aligned} \max \ & \sum_{i=1}^{N} f_i l_i \\ s.t. \ & f_i \le S_i(x), \ f_i \le F_i^{\text{nom}}, \ \forall i \in \{0,1,\ldots,N\}, \\ & f_i \le R_{i+1}(x), \ \forall i \in \{0,1,\ldots,N-1\}, \\ & x \in \mathcal{H}. \end{aligned} \tag{28}$$

Theorem 3 (Bounds for performance metrics)**.** *Consider an $r \in \mathcal{U}$ which is strictly feasible for y_{nom} but infeasible for some $y \in \mathcal{Y}$. Assume that $\mathcal{Z}^y$ is a singleton for all $y \in \mathcal{Y}$. Then, Γ supports every $\mu \in \mathcal{P}_{\text{inv}}$. Furthermore, for $\mu \in \mathcal{P}_{\text{inv}}$, we have*

$$\begin{aligned} J_T(a, y_{\text{nom}}) &\le \min_{(x,y)\in\Gamma\times\mathcal{Y}} J_T(x,y) < J_T^\mu \\ &< \max_{(x,y)\in\Gamma\times\mathcal{Y}} J_T(x,y) \le J_T(b, y_{\max}), \\ \underline{J}_F \le \min_{(x,y)\in\Gamma\times\mathcal{Y}} J_F(x,y) &< J_F^\mu < \min_{(x,y)\in\Gamma\times\mathcal{Y}} J_F(x,y) \le \overline{J}_F, \end{aligned}$$

where $\underline{J}_F$ and $\overline{J}_F$ are given by (27)–(28)*.*

C. A two-cell example

We illustrate the main results using a two-cell system (i.e. $N = 2$) in which incidents are only allowed in cell 2; see Fig. 3. Tab. I lists the parameters of the traffic flow model. The incident model is:

(i) $\mathcal{Y} = \{y_{\text{nom}}, y_{\max}\}$;

(ii) $\alpha^{\text{nom}} = [0,0]^T$, $\alpha^{y_{\max}} = [0, 0.4]^T$; and

(iii) $\lambda(y_{\rm nom}, x, y_{\rm max}) = 0.5 + 0.015x_2$, $\lambda(y_{\rm max}, x, y_{\rm nom}) = 2$, both in hr^{-1}.

The capacity subject to change is F_1. The inputs are r_0 and r_2.

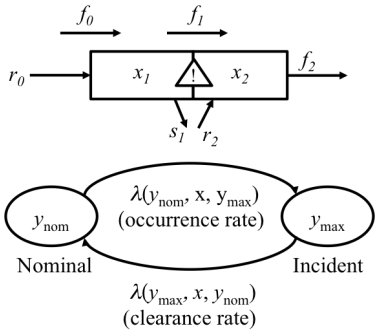


Fig. 3. A two-cell, two-mode system and its transition.

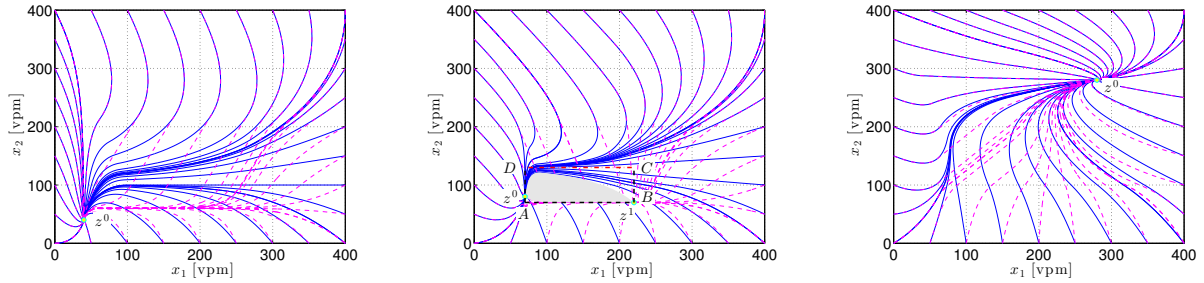
TABLE I

PARAMETERS OF THE TRAFFIC FLOW MODEL.

Name	Symbol	Value	unit
Cell length	l_i	1	mi
Free-flow speed	v_i	60	mph
Congestion wave speed	w_i	20	mph
Splitting coefficient	β_i	1	N/A
Jam density	$\overline{x}_i$	400	vpm
Critical density	x_i^c	100	vpm
Nominal capacity	$F_i^{\rm nom}$	6000	vph

Fig. 4 illustrates that the accessible set in this example assumes qualitatively different geometry depending on the stationary input $r = [r_0, r_2]^T$. When r is strictly feasible in both modes, the single limiting state constitutes the accessible set (i.e. Thm. 1, case (i); also see Fig. 4(a)). As r increases, the input remains feasible for $y_{\rm nom}$, but becomes infeasible for $y_{\rm max}$. The accessible set becomes a connected region containing both limiting states (Fig. 4(b)). Even in this two-cell example, analytic characterization of the accessible set is not straightforward. However, following Thm. 2, we can find an over-approximation of the accessible set given by $\mathcal{H} = [a_1, a_2] \times [b_1, b_2]$, where

$$a_1 = z_1^{\rm nom} = \frac{r_0}{v} = 70[vpm],$$



(a) Low input: $r = [2400, 0]^T$. (b) Intermediate input: $r = [4200, 600]^T$. (c) High input: $r = [6000, 1200]^T$.

Fig. 4. Integral curves of a two-cell, two-mode system. Solid (blue) curves correspond to $y = 0$, and dashed (magenta) curves correspond to $y = 1$. Fig. 4(b) shows the bounding box (rectangle $ABCD$).

$$b_1 = \overline{x} - \frac{(1-\alpha_1^{y_{\max}})F}{w} = 220[vpm],$$
$$a_2 = z_2^{y_{\max}} = \frac{(1-\alpha_1^{y_{\max}})F + r_2}{v} = 70[vpm],$$
$$b_2 = \overline{x} + \frac{r_2}{w} - \frac{(1-\alpha_1^{y_{\max}})F}{w} = 130[vpm].$$

If we further increase r to make it infeasible for y_{nom} (and indeed infeasible for $y_{\max}$), both cells are congested. The accessible set is again a singleton (i.e. Thm. 1, case (i); also see Fig. 4(c)).

Fig. 5 shows the empirical steady-state distributions of the continuous states (x_1, x_2). These distributions approximate the invariant probability measures.

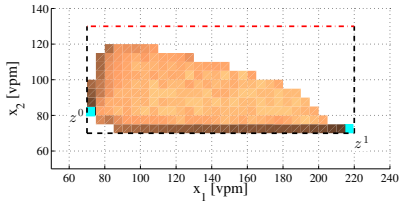


Fig. 5. Invariant probability measure depends on transition rate λ. Dark shade stands for high probability density.

This probability measure tends to concentrate near the limiting states. The support of the invariant probability does not depend on transition rate λ. The invariant probability is supported by the accessible set, which is specified by the deterministic dynamics provided that $(\lambda(i, x, j))_{ij}$ is irreducible for all $x \in \mathcal{X}$. However, the probability measure itself does depend on λ. Intuitively, high occurrence rate (or equivalently low clearance rate) tends to result in high traffic density

upstream to an incident hotspot and low traffic density downstream.

IV. Accessible Set

In this section, we study the geometry of the accessible set (Thm. 1) and derive an over-approximation for the accessible set when it is not a singleton (Thm. 2).

A. Geometry of accessible set

We prove Thm. 1, which gives the conditions for which the accessible set is a singleton or a connected set. First, we state a trivial observation:

Lemma 4. *If* (20) *holds, then, for all* $z \in \mathcal{Z}^{\mathrm{nom}}$ *and for all* $y \in \mathcal{Y}$*, we have* $z \in \mathcal{Z}^{y}$.

Proof. Note that $G(z, y_{\mathrm{nom}}) = 0$, $\forall z \in \mathcal{Z}^{\mathrm{nom}}$. If (20) holds, then $G(z, y) = G(z, y_{\mathrm{nom}}) = 0$, $\forall z \in \mathcal{Z}^{\mathrm{nom}}$, $\forall y \in \mathcal{Y}$. This implies that $z \in \mathcal{Z}^{y}$. □

We will use the following known results in the proof of Thm. 1. Thm. 5 is from [14], and Prop. 6 is from [3].

Theorem 5 (Convergence of CTM)**.** *For all* $x \in \mathcal{X}$ *and all* $y \in \mathcal{Y}$*, we have* $\lim_{t\to\infty} \phi_t^y(x) \in \mathcal{Z}^y$.

Proposition 6 (Positive invariance)**.** *For all* $x \in \Gamma$ *and all* $t \geq 0$*, we have* $\overline{\gamma}^+(x) \subset \Gamma$.

The proof of Thm. 1 is as follows:

Proof of Thm. 1. (i) Let $\mathcal{Z}^{\mathrm{nom}} = \{z^{\mathrm{nom}}\}$ (singleton). On one hand, by Thm. 5, we have $(\forall x \in \mathcal{X})$ $z^{\mathrm{nom}} \in \overline{\gamma}^+(x)$, and thus $z^{\mathrm{nom}} \in \Gamma$ (see Def. 1). On the other hand, by Lmm. 4, since (20) holds, we have $z^{\mathrm{nom}} \in \mathcal{Z}^y$, and thus $G(z^{\mathrm{nom}}, y) = 0$, $\forall y \in \mathcal{Y}$. Therefore, $\overline{\gamma}^+(z^{\mathrm{nom}}) = \{z^{\mathrm{nom}}\}$, and thus $\Gamma \subset \{z^{\mathrm{nom}}\}$. In conclusion, $\Gamma = \mathcal{Z}^{\mathrm{nom}} = \{z^{\mathrm{nom}}\}$.

(ii) Let y be a mode such that $\mathcal{Z}^y = \{z\}$ is a singleton. by Thm. 5, $z \in \overline{\gamma}^+(x)$, $\forall x \in \mathcal{X}$. Hence $z \in \Gamma$, which implies that Γ is not an empty set. If $\exists i$ such that $f_i^{\mathrm{unc}} > \underline{F}_i$, then there exists $w \in \mathcal{Y}$ such that $f^w \neq f^{\mathrm{nom}}$, and thus that $z^w \notin \mathcal{Z}^{\mathrm{nom}}$.

If y is such that $z \in \mathcal{Z}^{\mathrm{nom}}$, then $\overline{\gamma}^+(z) \supset \{\phi_t^w(z); t \geq 0\} \neq \{z\}$. By positive invariance of Γ (see Prop. 6), we have $\overline{\gamma}^+(z) \subset \Gamma$. Therefore, Γ is not a singleton. Finally, because of continuity of the trajectories, Γ is a connected set.

If y is such that $z \notin \mathcal{Z}^{\rm nom}$, then $\overline{\gamma}^+(z) \supset \{\phi_t^{\rm nom}(z); t \geq 0\} \neq \{z\}$, which in turn implies that Γ is a connected set. □

Now we elaborate on Rem. 3. In case (iii), if (20) holds and if $\mathcal{Z}^{\rm nom}$ is not a singleton, consider $z, \zeta \in \mathcal{Z}^{\rm nom}$, $z \neq \zeta$. By Lmm. 4, since (20) holds, we have $z \in \mathcal{Z}^y$, and thus $G(z, y) = 0, \forall y \in \mathcal{Y}$. Therefore, given the initial condition $(X(0), Y(0)) = (z, \cdot)$, we have $G(X(t), Y(t)) = 0$ and $X(t) = z$ for all $t \geq 0$. Therefore, $\overline{\gamma}^+(z) = \{z\}$, and thus $\Gamma \subset \{z\}$. Similarly, we can show that $\Gamma \subset \{\zeta\}$. Hence $\Gamma \subset \{z\} \cap \{\zeta\} = \emptyset$. In case (iv), if there exists i such that $f_i^{\rm nom} > \underline{F}_i$ and if there is no $y \in \mathcal{Y}$ such that $\mathcal{Z}^y$ is a singleton, then proof of the existence of accessible set is an open question which is not pursued here due to its limited significance for incident management.

In fact, for a given y, $\mathcal{Z}^y$ is a singleton unless there exist some $z \in \mathcal{Z}^y$ and some i such that $F_i^y = R_{i+1}(z)$, i.e. both the capacity and the receiving flow constraints are simultaneously binding. Only a small set of peculiar values of r will lead to this situation. An arbitrarily small perturbation in r will prevent this from happening. This technicality causes substantial mathematical complexity but brings limited practical insights. Therefore, in the rest of this article, we will only consider r such that $\mathcal{Z}^y$ is a singleton for all $y \in \mathcal{Y}$. Also, we denote this unique limiting state by z^y.

B. Over-approximation

We now consider the case where r is strictly feasible for $y_{\rm nom}$ but infeasible for some $y \in \mathcal{Y}$. Prop. 8 and Cor. 9 study how capacity reduction impacts the location of bottlenecks. Location of bottlenecks implies congestion pattern at limiting states, and enables derivation of the over-approximation.

We first recall a known result from [14]:

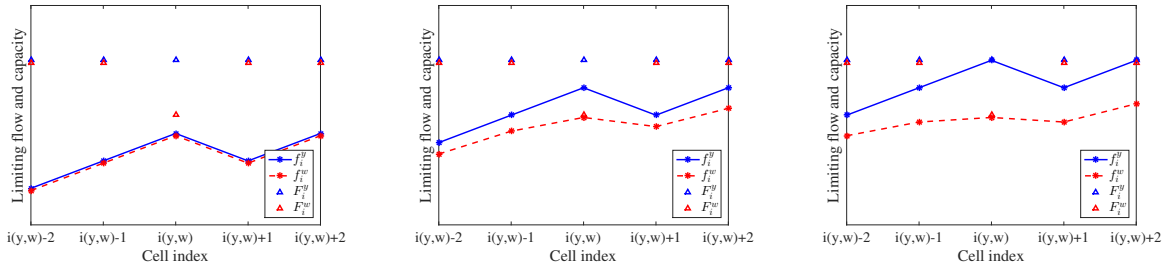
Theorem 7 (Structure of limiting states)**.** *At limiting states, a freeway can be partitioned into sections of uncongested and congested segments. Furthermore, a congested section must be located immediately upstream of a bottleneck, while an uncongested section must be located immediately downstream of a bottleneck.*

Next, we study how capacity reduction changes structure of limiting states.

Proposition 8 (Impact on bottlenecks)**.** *Consider two modes $y, w \in \mathcal{Y}$ such that $\alpha^y_{i(y,w)} < \alpha^w_{i(y,w)}$ for a unique $i(y,w)$ and $\alpha^y_i = \alpha^w_i$ for every $i \neq i(y,w)$.*

(i) If $\alpha^w_{i(y,w)} < 1 - f^y_{i(y,w)}/F^y_{i(y,w)}$, then $z^w = z^y$, and f^w has the same bottlenecks, if any, as f^y does.

(ii) If $\alpha^w_{i(y,w)} > 1 - f^y_{i(y,w)}/F^y_{i(y,w)}$, the $i(y,w)$-th cell is the only bottleneck for f^w. Moreover, every cell upstream to the $i(y,w)$-th cell is congested, while every cell downstream is uncongested. [6]



(a) Case (i): identical limiting flows. (b) Case (ii)a: distinct limiting flows. (c) Case (ii)b: distinct limiting flows.

Fig. 6. Illustration of various cases in Prop. 8.

Proof. Parts (i) is obvious, since the capacity reduction does not affect the limiting flow, and thus $f^y = f^w$; see Fig. 6(a). To show Part (ii), we need to account for two sub-cases:

Case (i)a: If $f^y = [f^y_0, f^y_1, \cdots, f^y_N]^T$ has no bottlenecks, then $f^y_i < F^y_i$ for all i; see Fig. 6(b). At limiting state, by conservation of flow, we have, for all i,

$$f^y_i = \beta_i(f^y_{i-1} + r_i), \quad f^w_i = \beta_i(f^w_{i-1} + r_i).$$

Since $\alpha^w_{i(y,w)} > 1 - f^y_{i(y,w)}/F^y_{i(y,w)}$, i.e. $f^y_{i(y,w)} > F^w_{i(y,w)}$, we have

$$\begin{aligned} f^w_{i(y,w)+1} &\leq \beta_{i(y,w)+1}(F^w_{i(y,w)} + r_{i(y,w)+1}) \\ &< \beta_{i(y,w)+1}(f^y_{i(y,w)} + r_{i(y,w)}) < F^y_{i(y,w)+1} = F^w_{i(y,w)+1}. \end{aligned}$$

Since $f^w_{i(y,w)+1} < f^y_{i(y,w)+1}$, we can show in a similar manner that $f^w_{i(y,w)+2} < F^y_{i(y,w)+2}$, $f^w_{i(y,w)+3} < F^y_{i(y,w)+3}$, and so forth. Hence, we have shown that $f^w_i < F^w_i$ for all $i > i(y,w)$. Since $f^w_N < F^w_N$,

[6] If $\alpha^w_{i(y,w)} = 1 - f^y_{i(y,w)}/F^y_{i(y,w)}$, then $\mathcal{Z}^y$ may not be a singleton. We do not consider this case.

the N-th cell has to be uncongested. Since $f_i^w < F_i^w$ for all $i > i(y,w)$, by Thm. 7, all cells downstream to the incident location have to be uncongested.

Next, we show that the upstream cells, including the incident location, have to be congested. By conservation of flow, we have $f_{i(y,w)}^w \leq F_{i(y,w)}^w$ and $f_i^w < F_i^w$ for $i = i(y,w)+1, \ldots, N$. If $i(y,w)$ were uncongested, then, since none of the first $i(y,w)$ cells is a bottleneck, they would have to be uncongested (see Thm. 7). Thus, the first cell would be uncongested, and it would have to be true that $f_0^w = r_0 = f_0^y$, which contradicts with $f_0^w < f_0^y$. Therefore, $i(y,w)$ has to be congested, and thus we have $f_{i(y,w)}^w = F_{i(y,w)}^w$, which means that $i(y,w)$ is a bottleneck. Since $f_0^w < f_0^y$, the first cell has to be congested. Since there are no bottlenecks between the first cell and the incident location, by Thm 7, all of the first $i(y,w)$ cells have to be congested.

Case (ii)b: If f^y has bottlenecks, the arguments above still hold; see Fig. 6(c). Since $f_{i(y,w)}^w < f_{i(y,w)}^y$, we can deduce that $f_i^w < f_i^y$ for all i, using the conservation of flow argument. Therefore, we have $f_i^w < F_i^y = F_i^w$ for every $i \neq i(y,w)$. Therefore, by definition, every $i \neq i(y,w)$ cannot be a bottleneck. Then i is uncongested for every $i \geq i(y,w)+1$. Since $f_0^w < f_0^y = r_0$, the first cell has to be congested, the implication being that one of its downstream cells has to be a bottleneck. Since $i(y,w)$ is the only possible bottleneck, it has to be so. □

For a multi-incident mode, one only needs to consider it as a superposition of a sequence of individual incidents, and apply Prop. 8 repeatedly. The result is as follows:

Corollary 9. *For an incident mode $y \in \mathcal{Y}$ with multiple incident locations, if $z^y \neq z^{\mathrm{nom}}$, then z^y has at most one bottleneck.*

Proof. This proof is based on the fact that the limiting flow of a multi-incident mode can be derived from the limiting flow of modes with fewer incidents. Specifically, we can prove this result by sequentially adding incidents and repeatedly applying Prop. 8.

Suppose that mode y has K incidents, which are located at cells $i_1, i_2, \cdots, i_K$. Let y_1 denote the mode with an incident at i_1, y_2 denote the mode with incidents at i_1 and i_2, y_3 denote the mode with incidents at i_1, i_2, and i_3, and so forth; then $y_K = y$.[7]

Next, consider z^{y_1} and f^{y_1}. If $f_i^{\mathrm{nom}} < F_i^{y_1}$ for all i, then $f^{y_1} = f^{\mathrm{nom}}$ is the limiting flow of mode y_1 and $z^{y_1} = z^{\mathrm{nom}}$. If $f_i^{\mathrm{nom}} > F_i^{y_1}$ for some i, by Prop. 8, $z^{y_1} \neq z^{\mathrm{nom}}$, and i_1 is the only

[7]Asm. 2 makes this construction possible.

bottleneck.

If $f_i^{y_{k-1}} < F_i^{y_k}$ for all i, then $f^{y_k} = f^{y_{k-1}}$ is the limiting flow of mode y_k and $z^{y_k} = z^{y_{k-1}}$. If $f_i^{y_{k-1}} > F_i^{y_k}$ for some i, by Prop. 8, $z^{y_k} \neq z^{y_{k-1}}$, and i_k is the only bottleneck. Then, by applying the inductive arguments above from y_1 to y_K, we complete the proof.[8] □

When Γ is a connected set, it is hard to analytically express the exact shape of Γ. Fortunately, we are able to determine a bounding box for the accessible set. In Fig. 4(b), we have a rectangle that approximates the accessible set. Analogously, in high dimensions, we can derive a high-dimensional "rectangle", or an orthotope [8], to approximate an accessible set. Before presenting the proof of Thm. 2, we construct the bounding box for an N-cell, two-mode system.

Proposition 10 (Bounding box for two modes)**.** *Consider an N-cell freeway system with $\mathcal{Y} = \{y_{\text{nom}}, y\}$ such that $\alpha^{y_{\text{nom}}}_{i(y_{\text{nom}},y)} < \alpha^{y}_{i(y_{\text{nom}},y)}$ for a unique $i(y_{\text{nom}}, y)$ and $\alpha_i^{y_{\text{nom}}} = \alpha_i^y$ for every $i \neq i(y_{\text{nom}}, y)$. Let r be an input that is strictly feasible for y_{nom} but infeasible for y. Let $z^{y_{\text{nom}}}$ and z^y be the unique limiting states of y_{nom} and y, respectively. Then $\Gamma \subset \mathcal{H}^y := \{x \in \mathcal{X} | a_i^y \leq x_i \leq b_i^y, i = 1, 2, \cdots, N\}$, where a_i^y and b_i^y are given by* (22)-(24)*.*

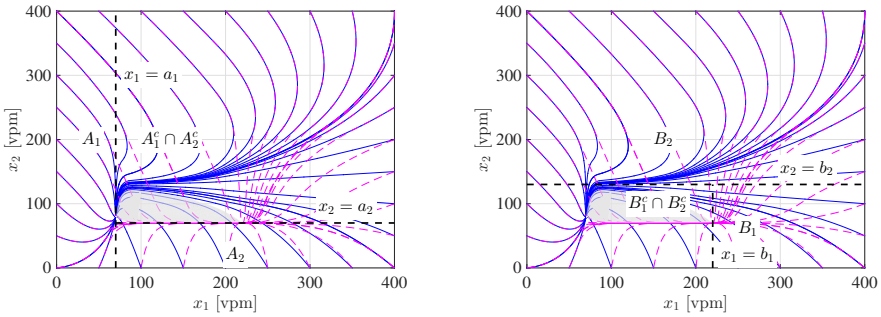
Proof. We need to show that the orthotope "absorbs" every admissible trajectory $\boldsymbol{\phi}_{\mathbf{u}}^{\boldsymbol{y}}(x)$ for all $x \in \mathcal{X}$. Fig. 7 illustrates the intuition of this proof via the two-cell example (see Sec. III-C). Define

$$
\begin{aligned}
A_1 &= \{\xi \in \mathcal{X} : \ \xi_1 < a_1^y\}, \\
A_i &= \{\xi \in \mathcal{X} : \ \xi_1 \geq a_1^y, \xi_2 \geq a_2^y, \ldots, \xi_i < a_i^y\}, \\
B_N &= \{\xi \in \mathcal{X} : \ \xi_N > b_N^y\}, \\
B_i &= \{\xi \in \mathcal{X} : \ \xi_N \leq b_N^y, \xi_{N-1} \leq b_{N-1}^y, \ldots, \xi_i > b_i^y\},
\end{aligned}
$$

We need to show that the accessible set $\Gamma \subset (A_1^c \cap \ldots A_N^c \cap B_1^c \cap \ldots \cap B_N^c)$. Fig. 7 shows that the orbits in all modes are absorbed by a subset of the continuous state space $\mathcal{X}$. This behavior results from the cooperative property of the dynamics G. The details are as follows.

(i) Lower bounds: We first show that $\Gamma \subset A_1^c$. Since r is strictly feasible for y_{nom}, we know that z^{nom} is uncongested everywhere. Therefore, for all $x \in A_1^c$, we have $f_0(x, \cdot) = r_0 = f_0^{\text{nom}}$

[8]Once again, we omit the case where $f_i^{y_{k-1}} = F_i^{y_k}$ for some k and some i; otherwise the limiting states may not be unique for y_k.



(a) All orbits are absorbed by $A_1^c \cap A_2^c$. (b) All orbits are absorbed by $B_1^c \cap B_2^c$.

Fig. 7. Construction of the bounding box for the two-cell example.

and $f_1(x,\cdot) \le S_1(x) \le f_1^{\text{nom}}$. Also note that $f_i(x, y_{\text{nom}}) \ge f_i(x,y)$ for all $x \in \mathcal{X}$. Hence, for all $x \in A_1^c$, we have $G_1(x,\cdot) \ge f_0^{\text{nom}} + r_1 - f_1^{\text{nom}}/\beta_1 = 0$, where the second argument of G_1 can be either y_{nom} or y. Now we show that every $x \in A_1$ is not accessible from A_1^c. Consider the initial condition $(x,\cdot)$ such that $x \in A_1^c$. Assume by contradiction that there exists $\zeta \in A_1$ such that $\zeta \in \overline{\gamma}^+(x)$; then there exists $n > 0$ and $(\boldsymbol{y}, \mathbf{u}) \in \mathscr{T}_n$ such that $\zeta = \boldsymbol{\phi}_{\mathbf{u}}^{\boldsymbol{y}}(x)$ (see Def. 1 for notations). Since $x_1 \ge z_1^{y_{\text{nom}}}$ and $\zeta_1 < z_1^{y_{\text{nom}}}$, continuity implies that there exists $k \le n$ and $0 \le s_1 < s_2 < u_k$ such that $(\phi_{s_1}^{y_k})_1 = z_1^{y_{\text{nom}}}$ and $(\phi_s^{y_k})_1 < z_1^{y_{\text{nom}}}$ for all $s \in [s_1, s_2]$. Thus $G_1(\phi_\tau^{y_k}, y_k) \ge 0$ for all $s \in [s_1, s_2]$, no matter whether $y_k = y_{\text{nom}}$ or $y_k = y$. Then we have

$$(\phi_{s_2}^{y_k})_1 = \int_{\tau=s_1}^{s_2} G_1(\phi_\tau^{y_k}, y_k) d\tau + (\phi_{s_1}^{y_k})_1 \ge z_1^{y_{\text{nom}}},$$

which contradicts with $(\phi_s^{y_k})_1 < z_1^{y_{\text{nom}}}$. Therefore, $A_1 \cap \Gamma = \emptyset$, and thus $\Gamma \subset A_1^c$; see Fig. 7(a).

Next, we show that $\Gamma \subset A_1^c \cap A_2^c$. Note that, for any $x \in A_2$, we have $f_1(x,\cdot) \ge \min\{S_1(x), F_1^y\} \ge f_1^{\text{nom}}$ and $f_2(x,\cdot) \le S_2(x) \le f_2^{\text{nom}}$. Hence, for all $x \in A_2$, we have

$$G_2(x,\cdot) \ge f_1^{\text{nom}} + r_1 - f_2^{\text{nom}}/\beta_2 = 0.$$

Following the contradiction argument that we made for A_1, we can conclude that $\Gamma \subset A_1^c \cap A_2^c$. We can similarly show that $\Gamma \subset A_1^c \cap A_2^c \cap \ldots \cap A_i^c$ for all $i \le i(y,w)$.

For $i = i(y,w) + 1$, note that every cell downstream of $i(y,w)$ is uncongested at z^y. Thus, for all $x \in A_{i(y,w)+1}$, we have $G_{i(y,w)+1}(x,\cdot) \ge f_{i(y,w)}^y + r_{i(y,w)+1} - f_{i(y,w)+1}^y/\beta_{i(y,w)+1} = 0$, where we have utilized the fact that $f_{i(y,w)+1}^y = f_{i(y,w)+1}(z^y, y_{\text{nom}})$. Once again, by following the contradiction argument, one can show that $\Gamma \subset A_{i(y,w)+1}^c$. Similarly, one can show that

$\Gamma \subset A_1^c \cap A_2^c \cap \ldots \cap A_i^c$ for all $i(y,w)+1 \leq i \leq N$; see Fig. 7(a). Here we have also covered the case where $i(y_{\rm nom}, y) = 0$.

In conclusion, $\Gamma \subset A_1^c \cap A_2^c \cap \ldots \cap A_N^c$, which implies the lower bounds.

(ii) Upper bounds: We first show that $\Gamma \subset B_N^c$. We claim that $b_N = \bar{x}_N + r_N/\omega_N - F_N^y/(\beta_N \omega_N)$ is a valid upper bound. Now we verify this bound. By (4), we know that $b_N > x_N^c$. Also note that $f_N(x, y_{\rm nom}) \geq f_N(x,y)$ for all $x \in \mathcal{X}$. Thus, for $x \in B_N$, we have $G_N(x,\cdot) \leq w_N(\overline{x}_N - b_N) + r_N - F_N^y \leq 0$, Then we can conclude $\Gamma \subset B_N^c$ by following the contradiction arguments that we made in part (i).

Next, we search for b_{N-1} such that $G_{N-1}(x,\cdot) \leq 0$ over B_{N-1}. By the cooperative property, the worst case is when $x_N = b_N$. We can apply the arguments for $i = N$ to $i = N-1$, by simply shifting the cell indices, and replacing F_{N-1} with $\min\{F_{N-1}, \omega_N(\bar{x}_N - b_N)\}$. Thus one can show $\Gamma \subset B_{N-1}^c \cap B_N^c$. The above logic applies to every $i = N-2, \ldots, 2$. For $i = 1$, one can replace F_{i-1} with r_0 and proceed similarly.

In conclusion, $\Gamma \subset B_1^c \cap B_2^c \cap \ldots \cap B_N^c$, which implies the upper bounds.

□

Proof for Theorem 2. We need to show that $\Gamma \subset \mathcal{H} = \{x \in \mathcal{X} | a_i \leq x_i \leq b_i, i = 1, 2, \cdots, N\}$, where $a_i = \min_{y \in \mathcal{Y}}\{a_i^y\}$ and $b_i = \max_{y \in \mathcal{Y}}\{a_i^y\}$.

If $a_1 = z_1^{\rm nom}$, then we have $z_1^y \geq z_1^{\rm nom}$ for all y. By the way that we constructed a_1^y, we have $G_1(x,y) \geq 0$ for all $x \in \{\xi : \xi_1 < a_1\}$ and for all $y \in \mathcal{Y}$. Thus, by contradiction method (see the proof of Prop. 10), we can show that $x \in \{\xi : \xi_1 < a_1\}$ is disjoint with Γ. Similarly, we can show $\{\xi \in \mathcal{X} : \xi_1 \geq a_1, \ldots, \xi_{i-1} \geq a_{i-1}, \xi_i < a_i\}$ is disjoint with Γ for all i. The same logic applies to the upper bounds.

By Prop. 8, if $a_1 = z_1^y \neq z_1^{\rm nom}$ for some y, then we must have that $F_0^y < F_0^{\rm nom}$ and $z_i^y < x_i^c$ for all i. Now consider an arbitrary mode $w \neq y$. If w has no incident in cell 1, then $f_0(x,w) > f_0(x,y)$. Apparently we have $f_1(x,w) \leq f_1(x,y)$. Thus, as we showed in the proof of Prop. 10, $G_1(x,w) > G_1(x,y) \geq 0$ for all $x \leq a_1 = z_1^y \leq z_1^{\rm nom}$. If w has incident in cell 1, then $f_0(z^y, y) = f_0(z^y, w)$ and $f_1(z^y, y) \leq f_1(z^y, w)$. Thus, $G_1(x,w) \geq G_1(x,y) \geq 0$. Then one can proceed as the previous case for $i = 2, \ldots, N$ and then for the upper bounds.

□

V. Steady-State Performance Metrics

In this section, we first show that the accessible set supports the invariant probability measure(s), by following [3]. Then we utilize properties of the flow dynamics to derive the bounds for performance metrics.

A. Support of invariant probability measure

The authors of [3] developed their results for smooth vector fields. Here we show that their results regarding the support of the invariant probability measure also hold for the dynamics G defined by (8)–(11), a continuous but non-smooth vector field. Specifically, we will show in sequence that

(i) The discrete process $\left(\widetilde{X}_n, \widetilde{Y}_n\right)$ and the continuous process $(X(t), Y(t))$ has the same support for their invariant probability measures (Prop. 11);

(ii) For the discrete process $\left(\widetilde{X}_n, \widetilde{Y}_n\right)$, every state in the accessible set can be reached with positive probability (Lmm. 12);

(iii) The accessible set of the discrete process $(\widetilde{X}_n, \widetilde{Y}_n)$ supports the invariant probability measure (Lmm. 13).

First, we verify the correspondence between $\mathcal{P}_{\text{inv}}$ and $\widetilde{\mathcal{P}}_{\text{inv}}$

Proposition 11. *If $\mu \in \widetilde{\mathcal{P}}_{\text{inv}}$, then $\mu\widetilde{K} \in \mathcal{P}_{\text{inv}}$. If $\mu \in \mathcal{P}_{\text{inv}}$, then $\mu Q \in \widetilde{\mathcal{P}}_{\text{inv}}$. Furthermore,* $\mathsf{supp}(\mu) = \mathsf{supp}\left(\mu\widetilde{K}\right)$

Proof. (i) Consider a smooth (with respect to the continuous argument) function $g : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. By (15), we have

$$\frac{d}{dt} K_t g(x,y) = (\nabla g(\phi_t^y(x), y)) G(\phi_t^y(x), y) = K_t A g(x,y),$$

where A is defined by (12). Therefore,

$$\begin{aligned}\widetilde{K}(\lambda I - A) g(x,y) &= \int_{t=0}^{\infty} \lambda e^{-\lambda t} K_t (\lambda I - A) g(x,y) dt \\ &= \int_{t=0}^{\infty} \lambda e^{-\lambda t} \left(\lambda K_t g(x,y) - \frac{d}{dt} K_t g(x,y) \right) dt = \lambda g(x,y),\end{aligned}$$

where I is the identity matrix. Similarly, one can show that $(\lambda I - A)\widetilde{K} g = \lambda g$. Then, recalling (13)-(17), we have

$$\mu\widetilde{K} L g(x,y) = -\lambda \left(\mu - \mu\widetilde{P} \right) g(x,y).$$

If $\mu \in \widetilde{\mathcal{P}}_{\text{inv}}$, then $\mu - \mu\widetilde{P} = 0$, and thus $\mu\widetilde{K}Lg(x,y) = 0$. Hence, $\mu\widetilde{K} \in \mathcal{P}_{\text{inv}}$. Also note that

$$\mu L\widetilde{K}g(x,y) = -\lambda\left(\mu - \mu Q\widetilde{K}\right)g(x,y).$$

If $\mu \in \mathcal{P}_{\text{inv}}$, then $\mu L\widetilde{K}g(x,y) = 0$, and thus $\mu = \mu Q\widetilde{K}$. Hence $\mu Q = \mu Q\widetilde{K}Q = \mu Q\widetilde{P}$, i.e. $\mu Q \in \widetilde{\mathcal{P}}_{\text{inv}}$.

(ii) Then we show that the invariant probability measures of the discrete process $\left(\widetilde{X}_n, \widetilde{Y}_n\right)$ and that of the continuous process $(X(t), Y(t))$ are supported by the same set. We first show that $\mu\widetilde{K}$ is strictly positive everywhere over $\mathsf{supp}(\mu)$. Let $(x,y) \in \mathsf{supp}(\mu)$ and $\mathcal{B}(x)$ be a neighborhood of x. Then, for a sufficiently small $t_0 > 0$ and $t \in [0, t_0]$, continuity of trajectories implies that $\phi^y_{-t}(\mathcal{B}(x)) \subset \mathcal{B}(x)$. Thus

$$\begin{aligned}\left(\mu\widetilde{K}\right)(\mathcal{B}(x)\times\{y\}) &= \int_{t=0}^{\infty} \lambda e^{-\lambda t}\mu\left(\mathcal{B}(x)\times\{y\}\right)dt \\ &\geq \lambda\int_{t=0}^{t_0} e^{-\lambda t}\mu\left(\mathcal{B}(x)\times\{y\}\right)dt > 0,\end{aligned}$$

which proves that $\mathsf{supp}(\mu) \subset \mathsf{supp}\left(\mu\widetilde{K}\right)$. Conversely, let $\nu = \mu\widetilde{K}$, $(x,y) \in \mathsf{supp}(\nu)$, and $\mathcal{B}(x)$ be a neighbothood of x. Then, by part (i),

$$\begin{aligned}\mu\left(\mathcal{B}(x)\times\{y\}\right) &= \sum_{w\in\mathcal{Y}}\int_{\mathcal{B}(x)} Q(y,x,w)\nu\left(\mathcal{B}(x)\times\{y\}\right) \\ &\geq \int_{\mathcal{B}(x)} Q(y,x,y)\nu\left(\mathcal{B}(x)\times\{y\}\right) > 0,\end{aligned}$$

which proves that $\mathsf{supp}\left(\mu\widetilde{K}\right) \subset \mathsf{supp}(\mu)$.

□

Next, we construct the set of all "possible" trajectories starting from some $x \in \mathcal{X}$.

Definition 4 (Adapted sequences)**.** *For all $n \in \mathbb{Z}_+$, let*

$$\mathscr{T}_n^{y,w} = \left\{(\mathbf{y},\mathbf{u}) \in \mathscr{T}_n : y_0 = y, y_n = w\right\},$$

where $\mathscr{T}_n$ is defined in Def. 1. Given $(\mathbf{y},\mathbf{u}) \in \mathscr{T}_n$, define $(x_k)_{0\leq k\leq n}$ by induction by setting $x_0 = x$ and $x_{k+1} = \phi^{y_{k-1}}_{u_k}(x_k)$. Define $\mathbf{t} = (t_0, t_1, \ldots, t_n)$ such that $t_0 = 0$ and that $t_k = t_{k-1} + t_k$ for $k = 1, 2, \ldots, n$. Define

$$\eta_{x,\mathbf{y},\mathbf{u}}(t) = \begin{cases} x & \text{if } t = 0, \\ \phi^{y_{k-1}}_{t-t_{k-1}}(x_{k-1}) & \text{if } t_{k-1} < t \leq t_k,\ k = 1,\ldots,n, \\ \phi^{y_n}_{t-t_n}(x_n) & \text{if } t > t_n. \end{cases}$$

Finally, let $p(x, \mathbf{y}, \mathbf{u}) = \prod_{k=1}^{n} Q(y_{k-1}, x_k, y_k)$*, and define*

$$\mathscr{T}_{n,\mathrm{ad}(x)} = \{(\mathbf{y}, \mathbf{u}) \in \mathscr{T}_n : p(x, \mathbf{y}, \mathbf{u}) > 0\}.$$

An element of $\mathscr{T}_{n,\mathrm{ad}(x)}$ *is said to be* adapted to $x \in \mathcal{X}$.

Let $\mathsf{Pr}_{x,y}$ be the probability conditional on initial condition (x, y). Next, we show that every adapted trajectory can be realized with a positive probability.

Lemma 12. *Let* $\xi \in \Gamma$ *and* $\mathcal{B}(\xi)$ *be a neighborhood of* ξ*. There exist* $m \in \mathbb{Z}_+$ *and* $\delta > 0$ *such that, for all* $y, w \in \mathcal{Y}$ *and* $x \in \mathcal{X}$*,*

$$\mathsf{Pr}_{x,y}\left\{\left(\tilde{X}_m, \tilde{Y}_m\right) \in \mathcal{B}(\xi) \times \{w\}\right\} \geq \delta. \tag{29}$$

In particular, if Γ *is a connected set, then*

$$\mathsf{Pr}\left\{(\exists s \geq 0)(\forall t \geq s) X(t) \in \gamma\right\} = 1. \tag{30}$$

Proof. We first show that, for any $(\mathbf{y}, \mathbf{u}) \in \mathbb{T}_n$, $(x, y) \in \mathcal{X} \times \mathcal{Y}$, any $T \geq 0$, and any $\delta > 0$, we have

$$\mathsf{Pr}_{x,y}\left\{\sup_{0 \leq t \leq T} \|X(t) - \eta_{x,\mathbf{y},\mathbf{u}}(t)\| \leq \delta\right\} > 0. \tag{31}$$

Consider an arbitrary $(\mathbf{y}, \mathbf{u}) \in \mathbb{T}_{n,\mathrm{ad}(x)}$. Let $\mathbf{v} \in \mathbb{R}^n_+$. If $\max_{i=1,\ldots,n} |v_i - u_i| \leq \delta_1$ for some $\delta_1 > 0$, then there exists some $\delta_2 > 0$ such that

$$\sup_{0 \leq t \leq T} \|\eta_{x,\mathbf{y},\mathbf{v}}(t) - \eta_{x,\mathbf{y},\mathbf{u}}(t)\| \leq \delta, \quad p(x, \mathbf{y}, \mathbf{v}) \geq \delta_2.$$

Let $(U_1, \ldots, U_{n+1})$ be $n+1$ independent rv s with $U_i \sim \mathsf{Exp}(\lambda)$ for all i. Then

$$\begin{aligned}
&\mathsf{Pr}_{x,y}\left\{\sup_{0 \leq t \leq T} \|X(t) - \eta_{x,\mathbf{y},\mathbf{u}}(t)\| \leq \delta\right\} \\
&\geq \delta_2 \mathsf{Pr}\left\{\max_{i=1,\ldots,n} |U_i - u_i| \leq \delta_1, U_{n+1} \geq T - t_n + \delta_1\right\} \\
&\geq \delta_2 \left(\prod_{i=1}^{n} \left(e^{-\lambda(u_i - \delta_1)} - e^{-\lambda(u_i + \delta_1)}\right)\right) e^{-\lambda(T - t_n + \delta_1)} > 0.
\end{aligned}$$

Now consider the case where $(\mathbf{y}, \mathbf{u})$ is not adapted, say, starting from an arbitrary initial mode w. Then, since Q is by construction irreducible and aperiodic, a transition from w to y_0 can happen in any arbitrarily short time interval with positive probability. Finally, (29) and (30) follow from (31) and Lmm. 3.16 in [3]. □

By summarizing the above results, we have the following lemma:

Lemma 13 (Support of invariant probability measure)**.** *Every invariant probability measure $\mu \in \mathcal{P}_{\rm inv}$ is supported by the accessible set Γ.*

Proof. If Γ is a singleton, the proof is trivial. If Γ is a connected set, by Lmm. 12, every state in Γ (of the continuous-time process) is positively recurrent. Therefore, every $\mu \in \widetilde{\mathcal{P}}_{\rm inv}$ is supported by Γ. Thanks to the equivalence between $\widetilde{\mathcal{P}}_{\rm inv}$ and $\mathcal{P}_{\rm inv}$ (see Prop. 11), the invariant probability measure(s) of the continuous-time process is (are) supported by Γ. □

B. Bounds for performance metrics

Now we consider the long-time behavior of performance metrics. Once again, we only consider the case where r is strictly feasible for $y_{\rm nom}$ but infeasible for some $y \in \mathcal{Y}$, and $\mathcal{Z}^y$ is a singleton for all $y \in \mathcal{Y}$. Without of loss of generality, we assume unit length of cells.

Lemma 14. *If $\mathcal{Z}^y = \{z\}$, then $z \in \Gamma$.*

Proof. Recall that, when z is the unique limiting state in mode y, every orbit $\{\phi_t^y(x); t \geq 0\}$ converges to z. Therefore, $z \in \overline{\gamma}^+(x)$ for all $x \in \mathcal{X}$. Hence $z \in \Gamma$. □

Lemma 15. *for $\mu \in \mathcal{P}_{\rm inv}$, we have*

$$\begin{aligned} J_T(a, y_{\rm nom}) &< \min_{(x,y)\in\Gamma\times\mathcal{Y}} J_T(x,y) < J_T^\mu \\ &< \max_{(x,y)\in\Gamma\times\mathcal{Y}} J_T(x,y) < J_T(b, y_{\max}), \qquad (32) \\ \min_{x\in V(\mathcal{H})} J_F(x, y_{\max}) &\leq \min_{(x,y)\in\Gamma\times\mathcal{Y}} J_F(x,y) < J_F^\mu \\ &< \max_{(x,y)\in\Gamma\times\mathcal{Y}} J_F(x,y) \leq \overline{J}_F, \qquad (33) \end{aligned}$$

where $\overline{J}_F$ is the optimal value of the LP (28)*.*

Proof. The inequalities (32) naturally follow from Thm. 2 and Lmm. 13. The proof of (33) is as follows.

Note that, for every $i \in \{0, 1, \ldots, N\}$ and every $y \in \mathcal{Y}$, $f_i(x, y)$ is concave in x (see (9)). Therefore, $J_F(x, y)$ is concave in x. Since $\mathcal{H}$ is a polyhedron, by standard results from

optimization [4], we have

$$\min_{x\in V(\mathcal{H})} J_F(x, y_{\max}) = \min_{x\in\mathcal{H}} J_F(x, y_{\max})$$
$$\leq \min_{x\in\Gamma} J_F(x, y_{\max}) = \min_{(x,y)\in\Gamma\times\mathcal{Y}} J_F(x, y),$$

where $V(\mathcal{H})$ is the set of vertices of $\mathcal{H}$. Hence the lower bound follows.

To show the upper bound, note that

$$\max_{(x,y)\in\Gamma\times\mathcal{Y}} J_F(x, y) \leq \max_{x\in\Gamma} J_F(x, y_{\text{nom}}) \leq \max_{x\in\mathcal{H}} J_F(x, y_{\text{nom}}).$$

Since $J_F(x, \cdot)$ is concave and piecewise-affine in x, standard results in linear programming (LP, see e.g. [4]) imply that $\max_{x\in\mathcal{H}} J_F(x, y_{\text{nom}}) = \overline{J}_F$, where $\overline{J}_F$ is the optimal value of the LP (28). □

Thm. 3 follows from Lmm. 13 and 15.

VI. Computational Study

We present a computational study to numerically analyze both qualitative and quantitative properties of the stochastic switched model of random incidents.

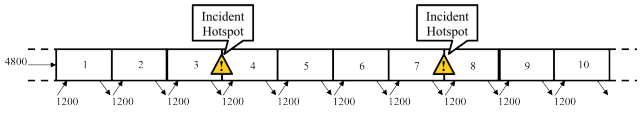


Fig. 8. A ten-cell freeway segment with two incident hotspots, upstream of cell 4 and cell 8.

A. Freeway segment with two incident hotspots

Consider a freeway segment with ten cells ($N = 10$) and two incident hotspots, i.e. $H = \{4, 8\}$; see Fig. 8. This system has four modes: no incident ($y = 0$), incident in cell 4 ($y = 1$), incident in cell 8 ($y = 2$), and incidents in both cells ($y = 3$); let $\mathcal{Y} = \{0, 1, 2, 3\}$. The model transitions between these discrete modes, as depicted in Fig. 9. The incident rate at each hotspot is given by $kx_i + b$, $i \in \{4, 8\}$, where k and b are constant parameters. For simplicity, the incident clearing rate is assumed to be a constant μ (independent of x_i) for each hotspot i. Numerical values of the parameters are given in Fig. 9. Unless otherwise specified, we nominally assume that an

incident in either location results in a 1/3 capacity drop, i.e. $\alpha_3 = 0.33$ and $\alpha_7 = 0.33$. All other parameters are specified in Tab. I. The aforementioned parameters completely specify the stochastic switched model (8)–(11).

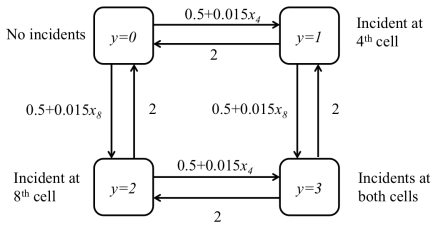


Fig. 9. Incident modes and mode transitions.

TABLE II

LIMITING STATES AND PERFORMANCE METRICS IN EACH INCIDENT MODE.

Mode (y)	Limiting states										J_F^y	J_T^y
	z_1^y	z_2^y	z_3^y	z_4^y	z_5^y	z_6^y	z_7^y	z_8^y	z_9^y	z_{10}^y		
0	100	100	100	100	100	100	100	100	100	100	52800	1000
1	**238**	**222**	**210**	87	89	91	93	95	96	97	45658	1317
2	**351**	**313**	**282**	**258**	**238**	**223**	**210**	87	89	91	35364	2141
3	**351**	**313**	**282**	**258**	**238**	**223**	**210**	87	89	91	35364	2141

This freeway segment is subject to a stationary demand specified by the $(N+1)$-dimensional vector $r = [4800, 1200, \ldots, 1200]^T$. Here r_i denotes the fixed arrival rate (in vph) at the on-ramp of cell i. Table II shows the values of the limiting states and the performance metrics (VHT and VMT) obtained by running a deterministic model simulation for each mode.

We conducted $100,000$ simulation runs of the stochastic switched model with random initial conditions. To study the invariant measure, we consider the empirically estimated steady-state marginal distribution of individual continuous states (i.e. traffic densities in specific cells). The empirical distributions for the i-th cell approximates the invariant measure, μ, marginalized over the continuous states x_j in other cells ($j \neq i$) and over all incident modes $y \in \mathcal{Y}$.

Fig. 10 shows the empirical steady-state marginal distribution of traffic densities in cell 3, which is located upstream of an incident hotspot. The figure also shows the limiting states given in Table II (columns for z_3^y). Since each of the individual mode dynamics converge to the

corresponding limiting state, the continuous state of the stochastic switched system, in steady state, resides in the vicinity of these limiting states with higher probability (compared to the regions in the accessible set that are further away from the limiting states). Fig. 10 validates this observation.

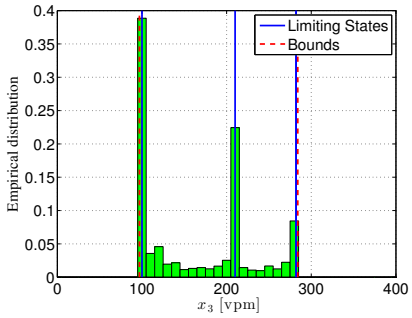
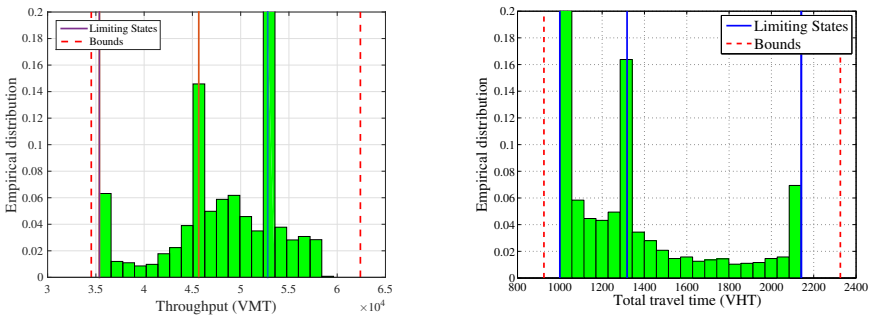


Fig. 10. Marginal distribution of density in the 3rd cell.

In addition, Fig. 10 shows the bounding intervals $[a_3, b_3]$ and $[a_7, b_7]$ defined by (22)-(23) in Thm. 2. These intervals bound the projections of the accessible set Γ (see Def. 18) on the respective dimensions. Consequently, these intervals bound the support of corresponding steady-state marginal distributions. These bounds are tight for this particular example.

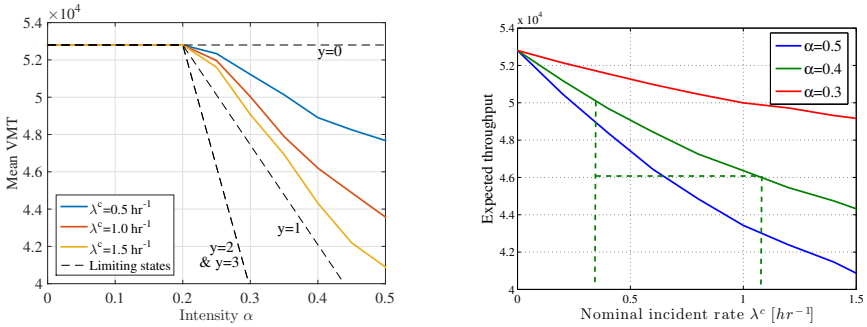
In Fig. 11, we plot the empirical distributions of the performance metrics (VMT and VHT), which also exhibit similar properties: these metrics are concentrated in the neighborhood of the mode-specific values (columns for J_F^y and J_T^y in Tab. II). Fig. 11 also shows the accuracy of the bounds given by Thm. 3.



(a) Distribution of throughput (VMT). (b) Distribution of total travel time (VHT).

Fig. 11. Empirical distribution of limiting-state throughput and VHT, both concentrated at the limiting states.

B. Impact of Incident Characteristics (α and λ)



(a) Impact of incident intensity.

(b) Impact of incident rate.

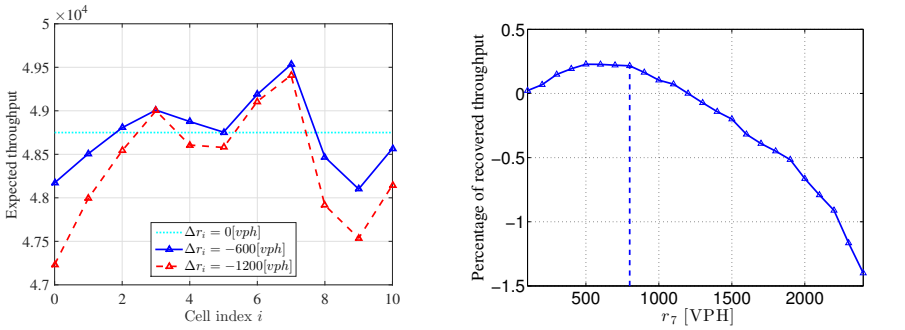
Fig. 12. Sensitivity of throughput with respect to incident characteristics: changing α from 0.3 to 0.4 is equivalent to changing λ^c from 0.35 to 1.1 hr^{-1}.

Impact of incident intensity: We now assume $\alpha \equiv \alpha_3 = \alpha_7$, and vary α from 0 (no capacity drop) to 0.5 (50% capacity reduction). Fig. 12(a) shows how the expected throughput (VMT) in steady state varies with α. The figure mainly illustrates two points. First, there is a critical intensity ($\alpha^c = 0.2$), below which incidents do not reduce expected throughput; see Prop. 8, case (i). Second, for $\alpha \geq \alpha^c$, the expected throughput decreases with α. Again, this is a consequence of the behavior of the limiting states.

Impact of incident rate: Fig. 12(b) shows that, in steady state, the expected throughput decreases with incident rate. We varied the coefficient k for the density-dependent incident rate. For the purpose of presentation, w define the *nominal incident rate* λ^c, which is the λ evaluated at the critical densities. Note that the sensitivity of expected throughput to incident rate is *lesser* in comparison to its sensitivity to incident intensity. Even at a high incident rate (e.g. more than 1 incident per hour), the expected throughput only reduces to 46,610 vehicle-mile, merely 10% less than the nominal value (52,800 vehicle-mile). In contrast, an increase from 0.2 to 0.3 in the intensity α would cause roughly 25% loss of expected throughput. *This observation indicates that less frequent but severe incidents could lead to more loss in expected throughput relative to a more frequent but less intense incidents.*

C. Impact of static ramp control in the wake of incidents

From a traffic operations viewpoint, a traffic controller has the capability of controlling inputs r by metering on-ramp flows. Previous control policies [14], [24], [29] do not explicitly account for the incident-induced congestion. Using our stochastic switched model, we would like to investigate the impact of metering inputs upstream of incident hotspots on long-term expected performance metrics. Unfortunately, the dependence of these metrics on r is not straightforward. So we use model simulation to determine which entrances/exits should be metered, and how much traffic should be metered at these entrances/exits.[9]



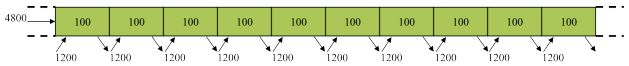
(a) Ramp metering at different locations. (b) Ramp metering of different magnitudes.

Fig. 13. Ramp metering impacts throughput.

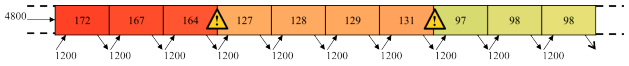
Fig. 13(a) shows the fraction of recovered throughput as a result of a 600/1200 vph reduction at each entrance metered individually. We note that the closer an input to a downstream incident hotspot, the more significantly it improves the expected throughput. In addition, r_7 plays a more significant role than r_3, because metering r_7 reduces congestion in cells 1 through 7, while metering r_3 only affects cells 1 through 3. Thus, r_7 is most effective for throughput recovery. If multiple entrances can be metered, the inputs r_3, r_5, and r_6 can all be metered. However, metering wrong locations could result in loss of throughput! For example, if r_0 were reduced, the loss would increase: reducing r_0 wastes available capacity in cells 1 through 3. Similar behavior is observed when the inputs r_8, r_9, and r_{10} are inappropriately metered. This also indicates that reallocating some demand from cell 7 to the downstream cells would improve throughput.

[9]We only consider static inputs, but our analysis also provides hints for dynamic traffic control.

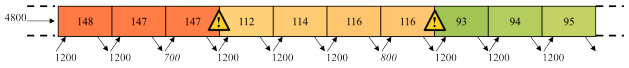
Next, we studied the impact of varying r_7; see Fig. 13(b). For low inflow rates, the capacity is underutilized, so throughput increases with r_7. Between 500 and 800 vph, roughly 25% throughput is recovered. Further increasing r_7 quickly results in throughput drop. As shown in Fig. 14(c), if we reduce r_7 from 1200 vph to 800 vph, the congestion in the upstream cells is reduced in comparison to the unmetered case in Fig. 14(b). Although this metering policy will cause a queue (growing at a rate of 400 vph) at entrance of ramp 7, it increases the total throughput by 500 vph. If traffic can be diverted from entrance 7 to entrance 8, then even more throughput can be recovered. Further, the nominal capacity allows r_7 to increase to 2,400 vph without causing congestion; however, the throughput loss would be increased by almost 150% if r_7 were increased to 2,400 vph, due to the impact of incidents. Therefore, *operating a freeway close to its nominal capacity can be highly suboptimal under incident-prone conditions.*



(a) In the nominal mode, the unmetered demand induces no congestion.



(b) Due to incidents, the unmetered demand tends to cause congestion. Incidents cause congestion in the cells upstream to hotspots; and cells 1 through 3 have the most severe congestion, since they are affected by both hotspots.



(c) Ramp metering strategy alleviates congestion induced by incidents.

Fig. 14. The idea of incident-aware ramp metering.

Our analysis is consistent with the arguments regarding the effect of ramp metering location on throughput made by Gomes et al. [14], although they did not focus on control in incident-prone conditions. They showed that throughput can be improved by reallocating demand at a bottleneck to an upstream location. We showed that, in incident-prone conditions, an incident hotspot is a potential bottleneck. Therefore, controlling the demand at incident hotspots will

improve the expected throughput.

VII. Concluding Remarks

The main contributions of this article are: (i) a switched stochastic model for traffic dynamics in incident-prone freeways, (ii) an analysis of steady-state properties of the switched system model, and (iii) simulation-based validation of the technical results and insights for incident management. The main idea of the switched model is an underlying stochastic process governing the initiation and termination of capacity-reduction incidents. The system randomly switches among a set of incident modes, and captures the impact due to mode-dependent dynamics. The accessible set (for the continuous state) of the stochastic switched system may be a singleton or a connected set. When the input is sufficiently low or sufficiently high (see Thm. 1), the accessible set is a singleton. When the accessible set is a connected set, we provide a bounding box to approximate it. Analytical characterization of the steady-state probability distribution of the continuous state is generally hard, but this distribution tends to concentrate near the limiting states. Interestingly, the expected performance is more sensitive to incident intensity than to incident rate. For stationary ramp metering, the incident-aware strategy is able to improve expected throughput.

Acknowledgment

This work was supported by NSF Grant No. 1239054 CPS Frontiers: Foundations Of Resilient CybEr-Physical Systems (FORCES) and MIT CEE Graduate Fellowship. We also appreciate the valuable inputs from Dr. Alex A. Kurzhanskiy.

References

[1] Mohamed A Abdel-Aty and A Essam Radwan. Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention*, 32(5):633–642, 2000.

[2] Yuri Bakhtin and Tobias Hurth. Invariant densities for dynamical systems with random switching. *Nonlinearity*, 25(10):2937–2952, 2012.

[3] Michel Benaïm, Stéphane Le Borgne, Florent Malrieu, and Pierre-André Zitt. Qualitative properties of certain piecewise deterministic Markov processes. *Annales de l'IHP*. to appear.

[4] Dimitris Bertsimas and John N Tsitsiklis. *Introduction to linear optimization*, volume 6. Athena Scientific Belmont, MA, 1997.

[5] Giacomo Como, Ketan Savla, Daron Acemoglu, Munther A Dahleh, and Emilio Frazzoli. Robust distributed routing in dynamical networks Part I: Locally responsive policies and weak resilience. *Automatic Control, IEEE Transactions on*, 58(2):317–332, 2013.

[6] Giacomo Como, Ketan Savla, Daron Acemoglu, Munther A Dahleh, and Emilio Frazzoli. Robust distributed routing in dynamical networks Part II: Strong resilience, equilibrium selection and cascaded failures. *Automatic Control, IEEE Transactions on*, 58(2):333–348, 2013.

[7] Samuel Coogan and Murat Arcak. A compartmental model for traffic networks and its dynamical behavior. *arXiv preprint arXiv:1409.6354*, 2014.

[8] Harold Scott Macdonald Coxeter. *Regular polytopes*. Courier Dover Publications, 1973.

[9] Carlos F Daganzo. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological*, 28(4):269–287, 1994.

[10] Mark H A Davis. Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society. Series B. Methodological*, 46(3):353–388, 1984.

[11] Gunes Dervisoglu, Gabriel Gomes, Jaimyoung Kwon, Roberto Horowitz, and Pravin Varaiya. Automatic calibration of the fundamental diagram and empirical observations on capacity. In *Transportation Research Board 88th Annual Meeting*, number 09-3159, 2009.

[12] Robert G Gallager. *Stochastic Processes: Theory for Applications*. Cambridge University Press, 2013.

[13] Genevieve Giuliano. Incident characteristics, frequency, and duration on a high volume urban freeway. *Transportation Research Part A: General*, 23(5):387–396, 1989.

[14] Gabriel Gomes, Roberto Horowitz, Alexandr A Kurzhanskiy, Pravin Varaiya, and Jaimyoung Kwon. Behavior of the cell transmission model and effectiveness of ramp metering. *Transportation Research Part C: Emerging Technologies*, 16(4):485–513, 2008.

[15] Martin Hairer. Convergence of markov processes. *lecture notes*, 2010.

[16] Morris W Hirsch. Systems of differential equations that are competitive or cooperative ii: Convergence almost everywhere. *SIAM Journal on Mathematical Analysis*, 16(3):423–439, 1985.

[17] Li Jin and Saurabh Amin. A piecewise-deterministic Markov model of freeway accidents. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*. IEEE, 2014.

[18] Bryan Jones, Lester Janssen, and Fred Mannering. Analysis of the frequency and duration of freeway accidents in Seattle. *Accident Analysis & Prevention*, 23(4):239–255, 1991.

[19] Erich Kamke. Zur Theorie der Systeme gewöhnlicher Differentialgleichungen. II. *Acta Mathematica*, 58(1):57–85, 1932.

[20] A Khattak, X Wang, and H Zhang. Incident management integration tool: dynamically predicting incident durations, secondary incident occurrence and incident delays. *IET Intelligent Transport Systems*, 6(2):204–214, 2012.

[21] Victor L Knoop. *Road Incidents and Network Dynamics: Effects on driving behaviour and traffic congestion*. PhD thesis, Technische Universiteit Delft, 2009.

[22] Alexandr A Kurzhanskiy. Set-valued estimation of freeway traffic density. In *Control in Transportation Systems*, pages 271–277, 2009.

[23] Alexandr A Kurzhanskiy. Online traffic simulation service for highway incident management. Technical Report SHRP 2 L-15(C), Relteq Systems, Inc., Albany, CA, February 2013.

[24] Alexandr A Kurzhanskiy and Pravin Varaiya. Active traffic management on road networks: a macroscopic approach. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1928):4607–4626, 2010.

[25] Richard C Larson and Amedeo R Odoni. *Urban Operations Research*. Prentice-Hall, NJ, 1981.

[26] Chris Lee, Bruce Hellinga, and Frank Saccomanno. Real-time crash prediction model for application to crash prevention in freeway traffic. *Transportation Research Record: Journal of the Transportation Research Board*, 1840(1):67–77, 2003.

[27] Peter A Lewis and Gerald S Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, 1979.

[28] Mahalia Miller and Chetan Gupta. Mining traffic incidents to forecast impact. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, pages 33–40, Beijing, China, August 2012. ACM.

[29] Markos Papageorgiou, Habib Hadj-Salem, and Jean-Marc Blosseville. ALINEA: A local feedback control law for on-ramp metering. *Transportation Research Record*, (1320), 1991.

[30] Francisco C Pereira, Filipe Rodrigues, and Moshe Ben-Akiva. Text analysis in incident duration prediction. *Transportation Research Part C: Emerging Technologies*, 37:177–192, 2013.

[31] David Schrank, Bill Eisele, and Tim Lomax. TTI's 2012 urban mobility report. *Proceedings of the 2012 annual urban mobility report. Texas A&M Transportation Institute, Texas, USA*, 2012.

[32] Alexander Skabardonis, Karl F Petty, Robert L Bertini, Pravin P Varaiya, Hisham Noeimi, and Daniel Rydzewski. I-880 field experiment: analysis of incident data. *Transportation Research Record: Journal of the Transportation Research Board*, 1603(1):72–79, 1997.

[33] State of California. Caltrans performance measurement system. Retrieved in September, 2013.

[34] Jérôme Thai and Alexandre M Bayen. State estimation for polyhedral hybrid systems and applications to the godunov scheme for highway traffic estimation. *Automatic Control, IEEE Transactions on*, 60(2):311–326, 2015.

[35] Qi Yang, Haris N Koutsopoulos, and Moshe E Ben-Akiva. Simulation laboratory for evaluating dynamic traffic management systems. *Transportation Research Record: Journal of the Transportation Research Board*, 1710(1):122–130, 2000.