Title
Delay-Based Traffic Signal Control for Throughput Optimality and Fairness at an Isolated Intersection

Journal
IEEE Transactions on Vehicular Technology, 67(2)

ISSN
0018-9545

Authors
Wu, Jian
Ghosal, Dipak
Zhang, Michael
et al.

Publication Date
2018

DOI
10.1109/tvt.2017.2760820

Delay-based Traffic Signal Control for Throughput Optimality and Fairness at an Isolated Intersection

Jian Wu, Dipak Ghosal, *Member, IEEE,* H. Michael Zhang, *Senior Member, IEEE,* and Chen-Nee Chuah, *Fellow, IEEE*

***Abstract*—With the attractive feature of guaranteeing maximum network throughput, backpressure routing has been widely used in wireless communication networks. Motivated by the backpressure routing, in this paper, we propose a delay-based traffic signal control algorithm in transportation networks. We prove that this delay-based control achieves optimal throughput performance, same to the queue-based traffic signal control in literature. However, a vehicle at a lane whose queue length remains very small may be excessively delayed under queue-based signal control. Our delay-based backpressure control can deal with the excessive delays, and achieve better fairness with respect to delay, while still guaranteeing throughput optimality. Moreover, a general weighted control scheme combining the queue-based and delay-based schemes is also investigated, to provide a more flexible control according to the quality of service requirements. Numerical results explore their performance under both homogeneous and heterogeneous traffic scenarios.**

***Index Terms*—Traffic signal control, backpressure control, throughput optimality, fairness**

I. Introduction

Traffic signal control at signalized intersections is indispensable for urban traffic networks. Current adaptive signal control system relies mostly on data from infrastructure-based sensors, including in-pavement or video based loop detectors, which need to be installed and maintained properly. Moreover, the rapid development of technologies on connected vehicles (CVs) and vehicular networks introduces great opportunities of reforming the conventional traffic signal operation [1]. Millions of roadside units and vehicles equipped with communication and positioning devices will be connected, and vehicles are able to communicate with each other (V2V) and with the infrastructure (V2I) through IEEE 802.11p and dedicated short-range communication technologies, which allows finer data transmission [2], [3]. The data will provide real-time vehicle location, speed, acceleration and other vehicle information, based on which traffic controllers should be able to make "smarter" decisions [4]–[6]. With all the support, the efficient traffic signal control systems will focus on improving network throughput, reducing vehicle delays, avoiding network congestion (guaranteeing stability), and so on.

Authors are affiliated with University of California, Davis, CA 95616, USA.

Jian Wu and Dipak Ghosal are with the Department of Computer Science (email: eejwu@ucdavis.edu; dghosal@ucdavis.edu).

H. Michael Zhang is with the Department of Civil and Environmental Engineering (email: hmzhang@ucdavis.edu).

Chen-Nee Chuah is with the Department of Electrical and Computer Engineering. (email: chuah@ucdavis.edu).

A. Related work

With the help of the current loop detectors and the upcoming connected vehicle technologies mentioned above, real-time traffic information, such as vehicle queue lengths and vehicle waiting times, can be estimated [7]–[9]. Based on this, various adaptive signal control methods have been studied.

A longest queue first maximal weight matching algorithm is presented in [10] for scheduling the signal to minimize the queue sizes at each approach. It derives the stability conditions using Lyapunov function-based analysis, and considers scenarios in which differentiated services are offered to vehicle classes with different priorities. An oldest arrival first algorithm is proposed in [11] to minimize the delay across the intersection, which needs to group vehicles into approximately equal-sized platoons for its implementation. However, the longest queue scheduling may be unfair for vehicles in a short queue which cannot accumulate long enough to be scheduled. On the other hand, the oldest arrival scheduling cannot guarantee optimal throughput and has limitations in its implementation.

With the aim to guarantee optimal throughput, the "backpressure control" has been introduced into the traffic signal scheduling. The backpressure routing was first studied in [12], which considers a multi-hop radio network with random arrivals, and controls the system through link selection and job assignment. It has been mainly applied to wireless communication networks [13]–[16]. The idea of backpressure routing was first adapted to traffic signal control systems in [17]. For a network of intersections, it determines the phase for each junction to be activated during each time slot in a distributed manner using the queue length information. Based on this work, the capacity-aware backpresssure control is proposed in [18], where the finite capacity constraint is considered and a normalized pressure is utilized to mitigate congestion propagation. Feedback based traffic signal control has been studied considering deterministic arrivals in [19] and stochastic arrivals in [20]. For the latter case, the queue-length based backpressure control is studied using "network calculus", which is called "max-pressure controller" therein. Fixed-cycle backpressure control is studied in [21], where its throughput-optimal property can still be guaranteed. All these works study the backpressure traffic control using only queue length information.

B. Our contributions

The queue-based backpressure traffic signal control does not consider explicit delays of waiting vehicles and is completely based on the queue length, and thus it also suffers from the "last packet problem", which is well-known for the queue-based backpressure scheduling in wireless networks [15]. A vehicle at a queue whose queue length remains very small may be starved for a long time and delayed excessively, because the queue-based scheme gives a higher priority to lanes with larger queue lengths. To deal with the critical fairness issue is one of the main concerns in our work.

In this work, we build on the backpressure traffic signal control, and aim to not only guarantee optimal throughput, but also take into account the fairness with respect to delay. Similar to existing studies, we assume that certain vehicle information is available due to the V2V, V2I and CV technologies. To the best of our knowledge, the delay-based backpressure control, which potentially could deal with the excessive delay in queue-based backpressure control due to lack of subsequent vehicle arrivals, has not yet been explored in the traffic signal scheduling literature. In this paper, we start with the isolated intersection scenario, and study the delay-based backpressure traffic signal control which uses Head-Of-Line (HOL) delay information instead of queue lengths. Then we generalize it to a weighted backpressure control considering both HOL delays and queue lengths. We explore whether they are throughput optimal, check their stability regions, and evaluate the fairness with respect to mean delay as well as tail delay performance.

Our scheduling discipline is a rule that chooses the active phase at each time slot based on the network environment, e.g., road capacity, and the state of the vehicle traffic. The key challenge is to determine if the throughput optimality can still be guaranteed under this delay-based control. This requires that the vehicle queues are stable as long as the traffic arrival falls into the system maximum stability region. We use the fluid limit technique similar to that in [15], [22] to investigate this. Roughly speaking, the key idea to show throughput optimality of the delay-based backpressure traffic signal control is to exploit that, in the "fluid limit" and after some initial period of time, there exists a linear relation between queue lengths and HOL delays. Based on this, we can prove that the delay-based control achieves optimal throughput performance, same as the queue-based traffic signal control in literature. For the generalized weighted backpressure control, it allows for a more flexible control of queue lengths and delay distributions, to satisfy a variety of quality of service requirements.

The main contributions of this paper include the following.

- We propose the delay-based backpressure traffic signal control scheme at isolated intersections, with the aim to provide better fairness experience regarding delay while still guaranteeing maximum network throughput.
- We explore the properties of the delay-based control scheme, by adopting the fluid limit model from [15], [22]. We obtain several key results: (i) The delay-based backpressure traffic signal control scheme is throughput optimal, i.e., it makes all queues stable as long as the traffic arrival rates are within the system stability region. (ii) The delay-based control scheme shares the same stability region with that of the queue-based control scheme in literature. However, it achieves better fairness due to the ability in dealing with excessive delays, experienced by vehicles whose queue length remains very small under queue-based signal control. (iii) In most traffic cases, the delay-based and queue-based controls share almost the same average queue lengths. However, the average queue length of the delay-based scheme may be higher when there is high burstiness and/or heterogeneity in the traffic.
- We investigate a weighted backpressure scheme combining the delay-based and queue-based controls, which is proved also to be throughout optimal. The weighted control scheme allows a tradeoff between delay-based control and queue-based control. We can achieve moderate fairness and average queue length performance by choosing appropriate weighting parameters, especially when dealing with bursty or heterogeneous traffic.
- We compare the performance of different schemes under both homogeneous and heterogeneous random traffic arrivals, regarding the stability region, Jain's fairness index of delay, average queue length, as well as the delay tail distribution. Particularly, for the heterogeneous vehicle arrivals, the tail of the delay distribution under delay-based control vanishes much faster than that of queue-based control.

The remainder of this paper is organized as follows. Sec. II gives the system model. The fluid limit model is discussed in Sec. III. The delay-based traffic signal control, as well as the weighted control scheme, are proposed in Sec. IV. The throughput optimality is proved in Sec. V. Sec. VI provides the numerical results. The conclusion and future work are given in Sec. VII.

II. System model

We model an isolated intersection using a queueing network with nodes and links. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{M})$. $\mathcal{V}$ denotes the set of nodes which correspond to different roads/lanes, and $\mathcal{M}$ denotes the set of links which correspond to traffic movements/transfers between nodes.

$A_i(t)$ denotes the amount of traffic that exogenously arrives to the network at node i during time slot t. With probability one, it satisfies

$$\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\{A_i(\tau)\} = \lambda_i, \tag{1}$$

where λ_i is the average arrival rate at node i. The arrival rate vector of the network is $\vec{\lambda} = \{\lambda_1, \lambda_2, \cdots, \lambda_{|\mathcal{V}|}\}$. (For road i w/o arrivals, $\lambda_i = 0$.) Let $Q_{i,j}(t)$ denote the number of vehicles on lane i (or at node i) at the beginning of slot t waiting to leave lane i for lane j (or node j). For notational ease, we also use $Q_{i,j}$ to denote the queue itself. We let $\vec{Q}(t) \triangleq [Q_{i,j}(t), (i,j) \in \mathcal{M}]$ denote the queue length vector at slot t, and use $\|\cdot\|$ to denote the L_1-norm of a vector, e.g., $\|\vec{Q}(t)\| =$

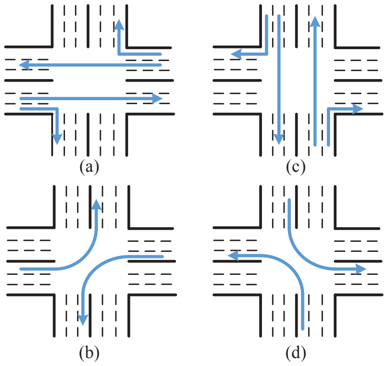


Fig. 1. The four phases, $|\mathcal{V}| = 24$ lanes and $|\mathcal{M}| = 12$ movements at an isolated intersection.

$\sum_{(i,j)\in\mathcal{M}} Q_{i,j}(t)$. Let $\Pi_{i,j}(t)$ represent the number of vehicles transferred from lane i to lane j during slot t.

Let $F_{i,j}(t)$ denote the total number of vehicles that arrive at lane i for lane j until time slot $t \geq 0$, including those present at slot 0, and let $\hat{F}_{i,j}(t)$ denote the total number of vehicles that are served at $Q_{i,j}$ until time slot $t \geq 0$. Obviously, $\hat{F}_{i,j}(0) = 0$ for all $(i,j) \in \mathcal{M}$. We have

$$Q_{i,j}(t) = F_{i,j}(t) - \hat{F}_{i,j}(t) \tag{2}$$

hold based on their definitions.

Let $T_{i,j,k}(t)$ denote the sojourn time of the k-th vehicle of $Q_{i,j}$ in the network at slot t, where the time is measured from the time when the vehicle arrives in the network. $W_{i,j}(t)$ denotes the sojourn time of the Head-Of-Line (HOL) vehicle of $Q_{i,j}$ in the network at slot t, and thus $W_{i,j}(t) \triangleq T_{i,j,1}(t)$. Furthermore, $W_{i,j}(t) = 0$ if $Q_{i,j}(t) = 0$. Similarly, $\vec{W}(t) \triangleq [W_{i,j}(t), (i,j) \in \mathcal{M}]$ denote the HOL sojourn time vector at slot t. Let us denote

$$U_{i,j}(t) \triangleq t - W_{i,j}(t), \tag{3}$$

which is the time when the HOL vehicle of $Q_{i,j}$ arrives in the network. With the definition of $U_{i,j}(t)$, for $t \geq 0$, we have

$$U_{i,j}(t) = \inf\{\tau \leq t : F_{i,j}(\tau) > \hat{F}_{i,j}(t)\}. \tag{4}$$

A schedule is a set of movements that can be active at the same time, which is denoted as $\vec{p} \in \{0,1\}^{|\mathcal{M}|}$. We will use the term "schedule" and "phase" interchangeable. Let $\mathcal{S}_P$ denote the set of all feasible schedules, and $Co(\mathcal{S}_P)$ denote its convex hull. For example, if $p_{i,j}(t) = 1$, the vehicle transfer from lane i to lane j is active at slot t; otherwise, if $p_{i,j}(t) = 0$, the transfer from lane i to lane j is inactive at slot t. We use $\mu_{i,j}(\vec{p})$ to denote the rate at which vehicles can go from lane i to lane j under schedule $\vec{p}$. Here we omit t for simplicity. Fig. 1 gives an example of four phases at an isolated intersection. In this paper, we assume that vehicles have fixed routing and all the routing information is known beforehand.

Similar to [15], [17], [22], the closed region Λ of arrival rate vectors $\vec{\lambda}$ is defined with the following properties: (1) $\vec{\lambda} \in \Lambda$ is a necessary condition for network stability; (2) $\vec{\lambda} \in \mathrm{int}(\Lambda)$, which means the inequalities in (5) are all strict, is a sufficient condition for the network stability.

$$\Lambda = \left\{ \vec{\lambda} \mid \exists\ \vec{\phi} \in Co(\mathcal{S}_P) \quad s.t.\ \lambda_i \leq \phi_{i,j}, \forall (i,j) \in \mathcal{M} \right\}. \tag{5}$$

The set of all arrival rates strictly inside Λ, $\vec{\lambda} \in \mathrm{int}(\Lambda)$, is usually called the system **maximum stability region**, or just **stability region**. A scheduling algorithm is said to **maximize the network throughput**, or to be **throughput optimal**, if it stabilizes the network for all arrival rates strictly inside Λ, $\vec{\lambda} \in \mathrm{int}(\Lambda)$, i.e., if it ensures all queues are stable as long as the arrival rates are within the system stability region.

The summary of notations is listed in Table I.

TABLE I
SUMMARY OF NOTATIONS

Symbol	Definition
$\mathcal{V}$	set of nodes (different lanes)
$\mathcal{M}$	set of links (traffic movements/transfers between nodes)
$\vec{p}$	a schedule (movements that can be active simultaneously)
$\mathcal{S}_P$	set of feasible schedules
$Co(\mathcal{S}_P)$	convex hull of $\mathcal{S}_P$
$\mu_{i,j}(\vec{p})$	rate at which vehicles go from lane i to lane j under schedule $\vec{p}$
Λ	optimal throughput region (stability region)
$A_i(t)$	# of vehicles arrive to the network for node i at time slot t
λ_i	average arrival rate for lane i
$F_{i,j}(t)$	total # of vehicles arrive at lane i for lane j until time slot t
$\hat{F}_{i,j}(t)$	total # of vehicles served at $Q_{i,j}$ until time slot t
$\Pi_{i,j}(t)$	# of vehicles served at $Q_{i,j}$ during time slot t
$Q_{i,j}(t)$	queue length of $Q_{i,j}$ at time slot t
$T_{i,j,k}(t)$	sojourn time of the k-th vehicle of $Q_{i,j}$ in the network at time slot t
$W_{i,j}(t)$	sojourn time of the HOL vehicle of $Q_{i,j}$ in the network at time slot t, i.e., $T_{i,j,1}(t)$
$U_{i,j}(t)$	time when the HOL vehicle of $Q_{i,j}$ arrives in the network, i.e., $t - W_{i,j}(t)$

III. Fluid Limit Model

Define the process describing the behavior of the system as $\mathcal{X} = (\mathcal{X}(t), t = 0, 1, 2, \cdots)$, where

$$\mathcal{X}(t) \triangleq \big((T_{i,j,1}(t), \cdots, T_{i,j,Q_{i,j}(t)}(t)), (i,j) \in \mathcal{M}\big), \tag{6}$$

whose norm is defined as

$$\|\mathcal{X}(t)\| \triangleq \|\vec{Q}(t)\| + \|\vec{W}(t)\|. \tag{7}$$

Let $\mathcal{X}^{(x)}$ denote a process $\mathcal{X}$ with an initial condition such that $\|\mathcal{X}^{(x)}(0) = x\|$. In the following, all variables associated with a process $\mathcal{X}^{(x)}$ will be marked with the upper index (x). As in [22] and [15], we extend the definition of $F_{i,j}^{(x)}(t)$ to the negative interval $t \in [-x, 0)$ by assuming that the vehicles present in the system in its initial state $\mathcal{X}^{(x)}(0)$ arrived in the past at some of the time instants $-(x-1), -(x-2), \cdots, 0$, according to their delays in the state $\mathcal{X}^{(x)}(0)$. By this convention $F_{i,j}^{(x)}(-x) = 0$ for all $(i,j) \in \mathcal{M}$ and x, and $\sum_{(i,j)\in\mathcal{M}} F_{i,j}^{(x)}(0) = x$.

We follow the techniques in [15], [22] to establish the fluid limit model. Define the process $X^{(x)} = (F^{(x)}, \hat{F}^{(x)}, \Pi^{(x)}, Q^{(x)}, W^{(x)}, U^{(x)})$, a sample of which uniquely defines the sample path of $\mathcal{X}^{(x)}$. Next, we adopt the convention $Y(t) \triangleq Y(\lfloor t \rfloor)$ for $Y = F^{(x)}$ with $t \geq -x$, and for $Y = \hat{F}^{(x)}, \Pi^{(x)}, Q^{(x)}, W^{(x)}, U^{(x)}$ with $t \geq 0$, making them as continuous time processes. Then using the techniques in the proof for [17, Lemma 1], we can show that with probability one, for any sequence of processes $\{X^{(x)}, x \in \mathcal{N}\}$, there exists a subsequence $\{X^{(k)}, k \in \mathcal{K} \subseteq \mathcal{N}\}$, such that as $k \to \infty$, the scaled subsequence $\{x^{(k)}, k \in \mathcal{K}\}$, which is defined as $x^{(k)}(t) = \frac{1}{k}X^{(k)}(kt)$, has the following convergence properties hold uniformly over compact (u.o.c.) intervals:

$$\frac{1}{k}F_{i,j}^{(k)}(kt) \to f_{i,j}(t), \tag{8}$$

$$\frac{1}{k}\hat{F}_{i,j}^{(k)}(kt) \to \hat{f}_{i,j}(t), \tag{9}$$

$$\frac{1}{k}Q_{i,j}^{(k)}(kt) \to q_{i,j}(t), \tag{10}$$

$$\frac{1}{k}\int_0^{kt}\Pi_{i,j}^{(k)}(\tau)d\tau \to \int_0^t \pi_{i,j}(\tau)d\tau, \tag{11}$$

and has the following hold at every continuous point of the corresponding limit functions:

$$\frac{1}{k}W_{i,j}^{(k)}(kt) \Rightarrow w_{i,j}(t), \tag{12}$$

$$\frac{1}{k}U_{i,j}^{(k)}(kt) \Rightarrow u_{i,j}(t). \tag{13}$$

The functions $(f, \hat{f}, \pi, q, w, u)$ are "fluid" limits of the corresponding scaled subsequences. These functions can be explained as follows. $f_{i,j}(t)$ is the amount of fluid from lane i to lane j that arrived into the system by the (scaled) time t. $\hat{f}_{i,j}(t)$ is the amount of fluid from lane i to lane j served by the system by the (scaled) time t. $q_{i,j}(t)$ is the amount of unserved fluid from lane i to lane j at the (scaled) time t. $\pi_{i,j}(t)$ is the amount of fluid transferred from lane i to lane j, $w_{i,j}(t)$ is the "head-of-the-line" fluid delay, and $u_{i,j}(t)$ is the "head-of-the-line" fluid arrival time at the (scaled) time t.

Definition 1. *The fluid model equations of the system are as follows:*

$$\sum_{(i,j)\in\mathcal{M}} f_{i,j}(0) \quad \leq \quad 1; \tag{14}$$

$$q_{i,j}(t) = f_{i,j}(t) - \hat{f}_{i,j}(t), \quad t \geq 0; \tag{15}$$

$$f_{i,j}(t) = f_{i,j}(0) + \lambda_i t, \quad t \geq 0; \tag{16}$$

$$u_{i,j}(t) = t - w_{i,j}(t); \tag{17}$$

$$\frac{d}{dt}q_{i,j}(t) = \lambda_i - \pi_{i,j}(t), \quad q_{i,j}(t) > 0. \tag{18}$$

Remark: In Eq. (15), the amount of unserved fluid $q_{i,j}(t)$ from lane i to lane j at the (scaled) time t follows from Eq. (2). Eq. (16) means that after time 0 the fluid from lane i to lane j arrives at constant rate λ_i. Eq. (17) follows from Eq. (3), and Eq. (18) is based on Eqs. (15)-(16).

Lemma 1. *For any fixed $t_1 > 0$, the two conditions $u_{i,j}(t_1) > 0$ and $\hat{f}_{i,j}(t_1) > f_{i,j}(0)$ are equivalent for every movement $(i,j) \in \mathcal{M}$. Furthermore, if they hold, then we have*

$$\lambda_i w_{i,j}(t) = q_{i,j}(t), \tag{19}$$

for all $t \geq t_1$ with probability 1.

Proof: Based on Eq. (4), we have $U_{i,j}(t) = \inf\{\tau \leq t : F_{i,j}(\tau) > \hat{F}_{i,j}(t)\}$ for all $t \geq 0$. Combining this with the definition of fluid limits, the equivalence of the two conditions can be obtained. When $\hat{f}_{i,j}(t_1) > f_{i,j}(0)$, by the definition of $u_{i,j}(t)$, we have $\hat{f}_{i,j}(t) = f_{i,j}(u_{i,j}(t))$ for $t \geq t_1$, since all vehicles arrive before the HOL vehicles have already been served. Then we have

$$\begin{aligned} q_{i,j}(t) &\overset{(a)}{=} f_{i,j}(t) - \hat{f}_{i,j}(t) \\ &= f_{i,j}(t) - f_{i,j}(u_{i,j}(t)) \\ &\overset{(b)}{=} [f_{i,j}(0) + \lambda_i t] - [f_{i,j}(0) + \lambda_i u_{i,j}(t)] \\ &= \lambda_i(t - u_{i,j}(t)) \\ &\overset{(c)}{=} \lambda_i w_{i,j}(t) \end{aligned}$$

where (a), (b) and (c) are based on Eq. (15), Eq. (16) and Eq. (17), respectively. ■

Remark: Similar to that in [22], property (19) states that if by some fixed (scaled) time t_1, the amount of served fluid from lane i to lane j is greater than its initial amount, i.e., $\hat{f}_{i,j}(t_1) > f_{i,j}(0)$, (or, equivalently, if by time t_1, the "head-of-the-line" fluid arrival time is larger than 0, i.e., $u_{i,j}(t_1) > 0$), which means all the "initial fluid" is "gone" by time t_1, then for all $t > t_1$, the linear relationship $\lambda_i w_{i,j}(t) = q_{i,j}(t)$ exists between the amount of fluid $q_{i,j}(t)$ and the "head-of-the-line" fluid delay $w_{i,j}(t)$. This is important for the proof of throughput optimality later.

IV. Delay-based backpressure traffic signal control at an isolated intersection

A. Delay-based Traffic Signal Control

For the signal control at isolated intersections, since the traffic backlog/delay at the downstream is considered zero, the backpressure control can be simplified and modeled using a general maximum weight problem.

The delay-based traffic signal control is given in Algorithm 1, where the optimal phase at slot t is chosen as follows:

$$\vec{p}^*(t) \in \underset{\vec{p}\in\mathcal{S}_P}{\operatorname{argmax}} \sum_{p_{i,j}=1} \gamma_{i,j} \cdot W_{i,j}(t) \cdot \mu_{i,j}(\vec{p}). \tag{20}$$

Algorithm 1 Delay-based backpressure traffic signal control (**DBPC**).

Input:
Set of feasible phases $\mathcal{S}_P$, HOL vehicle sojourn time $W_{i,j}(t)$ for all $(i,j) \in \mathcal{M}$.
Output:
Phase $\vec{p}^* \in \mathcal{S}_P$ to be activated during slot t.

1: Set $\mathcal{O}^*_{\vec{p}} = -\infty$, $\vec{p}^* = \emptyset$;
2: **for** each phase $\vec{p} \in \mathcal{S}_P$ **do**
3: $\quad\mathcal{O}_{\vec{p}} = \sum_{p_{i,j}=1} \gamma_{i,j} \cdot W_{i,j}(t) \cdot \mu_{i,j}(\vec{p})$;
4: $\quad$**if** $\mathcal{O}_{\vec{p}} > \mathcal{O}^*_{\vec{p}}$ **then**
5: $\quad\quad\mathcal{O}^*_{\vec{p}} = \mathcal{O}_{\vec{p}}$;
6: $\quad\quad\vec{p}^* = \vec{p}$;
7: $\quad$**end if**
8: **end for**

Here for the total pressure release $\mathcal{O}_{\vec{p}} = \sum_{p_{i,j}=1}\gamma_{i,j} \cdot W_{i,j}(t) \cdot \mu_{i,j}(\vec{p})$ allowed by $\vec{p}$ in line 3 of Algorithm 1, it is the sum of HOL vehicle sojourn time weighted by the flow of vehicles that can be transferred through the corresponding link when phase $\vec{p}$ is activated. We could further use positive constants $\gamma_{i,j}$, $(i,j) \in \mathcal{M}$, to give more emphasis to certain movements, and $\gamma_{i,j}$ is for the vehicles from lane i to lane j[1]. Finally, the returned phase is the one that maximizes the total pressure release, as shown in lines 4-7.

For the performance comparison later, the queue-based backpressure control in literature [17], [18] is also given in Algorithm 2, and the optimal phase is chosen as follows:

$$\vec{p}^*(t) \in \underset{\vec{p} \in \mathcal{S}_P}{\operatorname{argmax}} \sum_{p_{i,j}=1} \gamma_{i,j} \cdot Q_{i,j}(t) \cdot \mu_{i,j}(\vec{p}). \tag{21}$$

For the total pressure release allowed by $\vec{p}$ in line 3 of Algorithm 2, it is the sum of queue length weighted by the flow of vehicles that can be transferred through the corresponding link when phase $\vec{p}$ is activated, which is further weighted using positive constants $\gamma_{i,j}$ to give different emphases of movements.

Algorithm 2 Queue-based backpressure traffic signal control [17], [18] (**QBPC**).

Input:

Set of feasible phases $\mathcal{S}_P$, queue length $Q_{i,j}(t)$ for all $(i,j) \in \mathcal{M}$.

Output:

Phase $\vec{p}^* \in \mathcal{S}_P$ to be activated during slot t.

1: Set $\mathcal{O}^*_{\vec{p}} = -\infty$, $\vec{p}^* = \emptyset$;
2: **for** each phase $\vec{p} \in \mathcal{S}_P$ **do**
3: $\quad \mathcal{O}_{\vec{p}} = \sum_{p_{i,j}=1} \gamma_{i,j} \cdot Q_{i,j}(t) \cdot \mu_{i,j}(\vec{p})$;
4: $\quad$ **if** $\mathcal{O}_{\vec{p}} > \mathcal{O}^*_{\vec{p}}$ **then**
5: $\quad\quad \mathcal{O}^*_{\vec{p}} = \mathcal{O}_{\vec{p}}$;
6: $\quad\quad \vec{p}^* = \vec{p}$;
7: $\quad$ **end if**
8: **end for**

B. Weighted Backpressure Traffic Signal Control

Besides the delay-based and queue-based traffic control schemes, a weighted backpressure traffic signal control scheme is proposed in Algorithm 3, which uses both the queue length and delay information. Its optimal phase at slot t is chosen as follows:

$$\vec{p}^*(t) \in \underset{\vec{p} \in \mathcal{S}_P}{\operatorname{argmax}} \sum_{p_{i,j}=1} \gamma_{i,j} \cdot [\eta^{(W)}_{i,j} W_{i,j}(t) + \eta^{(Q)}_{i,j} Q_{i,j}(t)] \cdot \mu_{i,j}(\vec{p}). \tag{22}$$

Such joint design with parameters $\eta^{(W)}_{i,j}$, $\eta^{(Q)}_{i,j} \in [0,1]$ allows a more flexible control of queue lengths and delay distributions, which can be modified according to the relative importance between queue length and fairness of delay. For example, when lane i has a low vehicle arrival rate heading for lane j and we need to guarantee fair delay performance, we can set $\eta^{(W)}_{i,j}$ to be large relative to $\eta^{(Q)}_{i,j}$; conversely, when the vehicle arrival rate is high and we want to bound its queue length, $\eta^{(Q)}_{i,j}$ should be large relative to $\eta^{(W)}_{i,j}$. The throughput optimality of this weighted control will be discussed in the end of Sec. V.

[1]In the simulation of Sec. VI we make $\gamma_{i,j} = 1$. However, we will prove the optimal throughput for the general case of $\gamma_{i,j}$. The performance exploration for traffic with different emphases, e.g., multiple types of traffic with different priorities, will be left for the future work.

Algorithm 3 Weighted backpressure traffic signal control (**WBPC**).

Input:

Set of feasible phases $\mathcal{S}_P$, queue length $Q_{i,j}(t)$ and HOL vehicle sojourn time $W_{i,j}(t)$ for all $(i,j) \in \mathcal{M}$.

Output:

Phase $\vec{p}^* \in \mathcal{S}_P$ to be activated during slot t.

1: Set $\mathcal{O}^*_{\vec{p}} = -\infty$, $\vec{p}^* = \emptyset$;
2: **for** each phase $\vec{p} \in \mathcal{S}_P$ **do**
3: $\quad \mathcal{O}_{\vec{p}} = \sum_{p_{i,j}=1} \gamma_{i,j} \cdot [\eta^{(W)}_{i,j} W_{i,j}(t) + \eta^{(Q)}_{i,j} Q_{i,j}(t)] \cdot \mu_{i,j}(\vec{p})$;
4: $\quad$ **if** $\mathcal{O}_{\vec{p}} > \mathcal{O}^*_{\vec{p}}$ **then**
5: $\quad\quad \mathcal{O}^*_{\vec{p}} = \mathcal{O}_{\vec{p}}$;
6: $\quad\quad \vec{p}^* = \vec{p}$;
7: $\quad$ **end if**
8: **end for**

V. Throughput Optimality of Delay-based Traffic Signal Control

In this section we prove that the delay-based traffic signal control achieves optimal throughput performance. First, through Lemma 2 we provide the linear relation between queue lengths and delays in the fluid limits, which follows from Eq. (19) of Lemma 1. Then from this linear relation, we prove the throughput optimality in Proposition 3.

Lemma 2. *Consider the delay-based backpressure traffic signal control. For $\vec{\lambda}$ strictly inside Λ, there exists $T > 0$ such that the fluid limits satisfy the following property with probability* 1

$$\hat{f}_{i,j}(T) > f_{i,j}(0). \tag{23}$$

for all $(i,j) \in \mathcal{M}$.

Proof: The proof is in Appendix A. ■

Remark: This lemma states that under the delay-based backpressure traffic signal control, by some fixed time T, the amount of fluid served from lane i to lane j, i.e., $\hat{f}_{i,j}(T)$, will be larger than its initial amount $f_{i,j}(0)$, which means all the "initial fluid" will be gone by time T. Therefore, with Lemma 1, the linear relationship $\lambda_i w_{i,j}(t) = q_{i,j}(t)$ exists between the amount of fluid $q_{i,j}(t)$ and the "head-of-the-line" fluid delay $w_{i,j}(t)$ for all $t > T$. We prove this by induction following the techniques described in [17, Lemma 7], and the detail is in Appendix A.

The throughput optimality of the delay-based backpressure control is then presented in the following proposition, where we prove the stability using fluid limits and standard Lyapunov techniques.

Proposition 3. *The traffic signal control with delay-based backpressure scheduling can support any traffic with arrival rate vector that is strictly inside Λ.*

Proof: Based on Lemma 2 and Lemma 1, we know that Eq. (19) holds for the delay-based traffic signal control. Let

$L(\vec{q}(t))$ denote the Lyapunov function as

$$L(\vec{q}(t)) = \frac{1}{2} \sum_{(i,j)\in\mathcal{M}} \gamma_{i,j} \frac{(q_{i,j}(t))^2}{\lambda_i}. \tag{24}$$

Assume $\vec{\lambda}$ is strictly inside Λ, and then there exists a vector $\vec{\phi} \in Co(\mathcal{S}_P)$ such that $\vec{\lambda} < \vec{\phi}$, i.e. $\lambda_i < \phi_{i,j}$ for all $(i,j) \in \mathcal{M}$.

$$\begin{aligned}
\frac{d}{dt}L(\vec{q}(t)) &= \sum_{(i,j)\in\mathcal{M}} \frac{\gamma_{i,j}}{\lambda_i} q_{i,j}(t) \frac{d}{dt} q_{i,j}(t) \\
&= \sum_{(i,j)\in\mathcal{M}} \frac{\gamma_{i,j}}{\lambda_i} q_{i,j}(t)(\lambda_i - \pi_{i,j}(t)) \\
&= \sum_{(i,j)\in\mathcal{M}} \frac{\gamma_{i,j}}{\lambda_i} q_{i,j}(t)(\lambda_i - \phi_{i,j}(t)) \\
&\quad + \sum_{(i,j)\in\mathcal{M}} \gamma_{i,j} \frac{q_{i,j}(t)}{\lambda_i}(\phi_{i,j}(t) - \pi_{i,j}(t)).
\end{aligned} \tag{25}$$

Let us choose $\delta_3 > 0$ such that $L(\vec{q}(t)) \geq \delta_1 > 0$ implies $\max_{(i,j)\in\mathcal{M}} q_{i,j}(t) \geq \delta_3$. Based on the fact that the sum of non-negative elements must be no less than each element, we have

$$\begin{aligned}
&\sum_{(i,j)\in\mathcal{M}} \frac{\gamma_{i,j}}{\lambda_i} q_{i,j}(t)(\phi_{i,j}(t) - \lambda_i) \\
\geq\; & \frac{\gamma_{i',j'}}{\lambda_{i'}} [\max_{(i,j)\in\mathcal{M}} q_{i,j}(t)](\phi_{i',j'}(t) - \lambda_{i'})\Big|_{(i',j')=\underset{(i,j)\in\mathcal{M}}{\operatorname{argmax}}\, q_{i,j}(t)} \\
\geq\; & (\min_{(i,j)\in\mathcal{M}} \frac{\gamma_{i,j}}{\lambda_i}) \max_{(i,j)\in\mathcal{M}} q_{i,j}(t) \min_{(i,j)\in\mathcal{M}} (\phi_{i,j}(t) - \lambda_i) \\
\geq\; & (\min_{(i,j)\in\mathcal{M}} \frac{\gamma_{i,j}}{\lambda_i}) \delta_3 \min_{(i,j)\in\mathcal{M}} (\phi_{i,j}(t) - \lambda_i).
\end{aligned} \tag{26}$$

Then the first sum in Eq. (25) is bounded as follows:

$$\begin{aligned}
&\sum_{(i,j)\in\mathcal{M}} \frac{\gamma_{i,j}}{\lambda_i} q_{i,j}(t)(\lambda_i - \phi_{i,j}(t)) \\
\leq\; & -(\min_{(i,j)\in\mathcal{M}} \frac{\gamma_{i,j}}{\lambda_i}) \delta_3 \min_{(i,j)\in\mathcal{M}} (\phi_{i,j}(t) - \lambda_i) \\
\triangleq\; & -\delta_2 < 0.
\end{aligned} \tag{27}$$

The second term in Eq. (25) is non-positive if

$$\vec{\pi}(t) \in \underset{\vec{\phi}\in Co(\mathcal{S}_P)}{\operatorname{argmax}} \sum_{(i,j)\in\mathcal{M}} \gamma_{i,j} \cdot \frac{q_{i,j}(t)}{\lambda_i} \cdot \phi_{i,j}(t), \tag{28}$$

which means

$$\sum_{(i,j)\in\mathcal{M}} \gamma_{i,j} \frac{q_{i,j}(t)}{\lambda_i} \phi_{i,j}(t) \leq \sum_{(i,j)\in\mathcal{M}} \gamma_{i,j} \frac{q_{i,j}(t)}{\lambda_i} \pi_{i,j}(t). \tag{29}$$

We know that (28) holds because that the underlying delay-based scheduler indeed maximizes $\sum_{(i,j)\in\mathcal{M}} \gamma_{i,j} \frac{q_{i,j}(t)}{\lambda_i} \phi_{i,j}(t)$ due to the linear relationship $w_{i,j}(t) = \frac{q_{i,j}(t)}{\lambda_i}$ for $t \geq T$.

As a result, for any $\delta_1 > 0$, there exists $\delta_2 > 0$ and a finite time $T > 0$ such that $L(\vec{q}(t)) \geq \delta_1$ implies $\frac{d}{dt}L(\vec{q}(t)) \leq -\delta_2$ for any regular time $t \geq T$. Then it follows that for any $\zeta > 0$, there exists a large enough integer $T_1 > 0$ such that for any fluid limit with $\|\vec{q}(0)\| \leq 1$, there is

$$\|\vec{q}(t)\| \leq \zeta \tag{30}$$

for any time $t \geq T_1$. Then

$$\hat{f}_{i,j}(T_1) = f_{i,j}(T_1) - q_{i,j}(T_1) > f_{i,j}(0) \tag{31}$$

for all $(i,j) \in \mathcal{M}$, and we have $w_{i,j}(t) = \frac{q_{i,j}(t)}{\lambda_i}$ from Lemma 1. As a result,

$$\begin{aligned}
\|\vec{q}(t)\| + \|\vec{w}(t)\| &\leq \|\vec{q}(t)\| + \frac{1}{\min_i \lambda_i} \|\vec{q}(t)\| \\
&\leq (1 + \frac{1}{\min_i \lambda_i}) \zeta \\
&\triangleq 1 - \epsilon < 1,
\end{aligned} \tag{32}$$

and therefore, with probability 1,

$$\limsup_{x\to\infty} \frac{1}{x} \|\mathcal{X}^{(x)}(xT)\| \leq 1 - \epsilon. \tag{33}$$

We can make $1-\epsilon$ arbitrarily small by choosing small enough ζ. Since it is easy to show that the sequence $\{\frac{1}{x}\|\mathcal{X}^{(x)}(xT)\|\}$ is uniformly integrable, along with Eq. (33), it verifies the following condition

$$\limsup_{x\to\infty} \mathbf{E}\left[\frac{1}{x}\|\mathcal{X}^{(x)}(xT)\|\right] \leq 1 - \epsilon. \tag{34}$$

Then we introduce Theorem 4 in [22], which states that suppose there exist $\epsilon > 0$ and an integer $T > 0$ such that for any sequence of processes $\{\mathcal{X}^{(x)}, x = 1, 2, \cdots\}$, we have $\limsup_{x\to\infty} \mathbf{E}\left[\frac{1}{x}\|\mathcal{X}^{(x)}(xT)\|\right] \leq 1 - \epsilon$, then $\mathcal{X}$ is stable. As a result, based on the condition in (34), the proof is complete. ∎

For the queue-based backpressure traffic signal control, its throughput optimality has already been proved in [17]. We can revisit its throughput optimality using the fluid limit techniques, similar to the proof of Proposition 3, by defining the Lyapunov function to be $\frac{1}{2}\sum_{(i,j)\in\mathcal{M}} \gamma_{i,j}(q_{i,j}(t))^2$.

Proposition 4. *The traffic signal control with weighted backpressure scheduling can support any traffic with arrival rate vector that is strictly inside* Λ.

Proof: The proof is in Appendix B. ∎

Remark: Since we have already proved that under the delay-based backpressure traffic signal control, after some fixed time T, the linear relationship $\lambda_i w_{i,j}(t) = q_{i,j}(t)$ holds for all $t > T$. For the weighted backpressure control, which is a linear weighted combination between the queue-based control and the delay-based control, it's intuitive that, after a finite time, this linear relationship between the queue length and the delay in the fluid limit model still holds. Actually, a slightly adjusted proof of Lemma 2 allows us to prove (23) for the weighted backpressure traffic signal control, which means $\lambda_i w_{i,j}(t) = q_{i,j}(t)$, and the detailed proof is omitted here.

VI. Numerical Results

We simulated a single intersection with 4 phases as shown in Fig. 2. The slot length is denoted with T_s. Assume that all lanes have infinite queue capacity.

Based on [17], [23], [24], the number of passing vehicles during slot t on each lane is $R_m(1 - e^{-\frac{Q(t)+I^a(t)}{R_m}})$, where

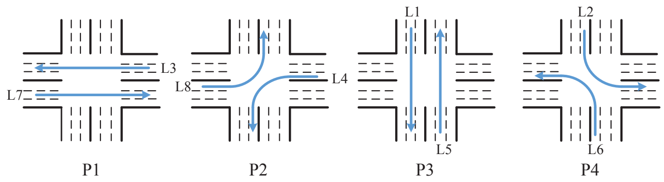


Fig. 2. A 4-phase intersection with 8 links in the simulation.

$R_m = \mu_s T_s$ is the maximum number of passing vehicles per slot. Here μ_s is the saturation flow rate. $Q(t)$ is the queue length of this lane at the beginning of this slot, and $I^a(t)$ is the number of vehicles arriving at this lane during time slot t. In the following simulation, we set $\mu_s = 0.5$ vehicles per second per lane (v/s/l) and $T_s = 5$s unless there is specific explanation. Therefor for the 8 lanes in Fig. 2, we have $\vec{\mu_s} = [1,1,1,1,1,1,1,1] * 0.5$ v/s/l.

We will first compare the performance of Queue-based BackPressure Control (QBPC) and Delay-based BackPressure Control (DBPC) in terms of maximum stability region (i.e., stability region, or optimal throughput region)[2], fairness, as well as tail delay, under different traffic patterns and network settings. Then taking into account the Weighted BackPressure Control (WBPC), we compare these three schemes together, and explore how this weighted scheme plays its tradeoff role.

A. *Maximum stability region, fairness and tail delay*

It is hard to find a closed form expression for the boundary of the stability region (i.e., the optimal throughput region) for each control scheme. As a result, similar to the method in [15], we probe the boundary by scaling the traffic load. After we choose the average arrival rate vector $\vec{\lambda}$ of the 8 lanes, we run the simulations with traffic load $\alpha\vec{\lambda}$ changing α, which scales the load. So the actual arrival rate vector $\alpha\vec{\lambda}$ is adjusted through the parameter α. When α increases, the traffic load of the system increases. Since the optimal throughput region is defined as the set of arrival rates under which the queue lengths remain finite, we consider the traffic load, under which the queue length increases rapidly, as the boundary of the optimal throughput region. Note that when we study the fairness and tail delay performance later, we only use traffic load within this stability region.

One potential advantage of the delay-based control is to resolve the large delay that the queue-based control may suffer from. To explore the fairness of each scheme regarding delay performance, we first use the Jain's fairness index [25]. Denote d_i to be the delay of vehicle i, and M is the total number of vehicles. Then the Jain's fairness index is given by

$$f(\vec{d} = [d_1, d_2, ..., d_M]) = \frac{\left(\sum_{i=1}^{M} d_i\right)^2}{M\sum_{i=1}^{M}(d_i{}^2)}. \quad (35)$$

If all vehicles experience the same delay, the fairness index is 1, and the system is 100% fair in terms of delay. The fairness decreases as the disparity in the delay faced by different vehicles increases.

[2]These three terms: maximum stability region, stability region, and optimal throughput region, will be used interchangeably.

Besides the fairness index, we also study the tail performance of the delay distribution. Specifically, we use the probability that the delay suffered by the vehicles is greater than a threshold (i.e., the tail distribution of the delay) as a metric.

B. *Homogeneous arrivals*

First, we study the case that vehicles arrive at each lane following Poisson process, with homogeneous average arrival rates. We set $\vec{\lambda} = [1,1,1,1,1,1,1,1] * 0.125$ v/s/l.

The average queue length under different traffic loads is illustrated in Fig. 3(a) to examine the performance limits of the backpressure control schemes. As stated above, we consider the traffic load, under which the queue length increases rapidly, as the boundary of the optimal throughput region. Fig. 3(a) shows that DBPC achieves the same stability region as QBPC. Moreover, QBPC and DBPC perform similarly in terms of the average queue length per lane, which also means they perform almost the same regarding the average delay due to the Little's Law.

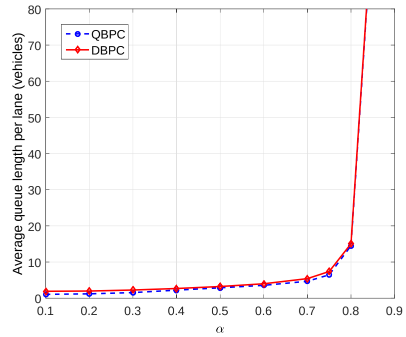
However, as shown in Fig. 3(d), DBPC has a better fairness performance than QBPC. This is because, even under homogeneous Poisson arrivals, for QBPC, the random arrivals may lead to large delay for some vehicles due to lack of subsequent vehicle arrivals.

Next, we want to study the case with bursty vehicle arrivals. Here we consider that vehicles arrive at each lane following Interrupted Poisson Process (IPP), whose burstiness can be measured using the coefficient of variation C^2. We still assume homogeneous average arrival rates $\vec{\lambda} = [1,1,1,1,1,1,1,1] * 0.125$ v/s/l. The parameters of IPP can be found in [26]. We can achieve different values of C^2 while guaranteeing the average arrival rate unchanged, by adjusting the parameters of IPP. For the traffic of each lane, given the average arrival rate, we observe the performance under different coefficients of variance. The larger C^2 is, the more bursty the traffic will be.

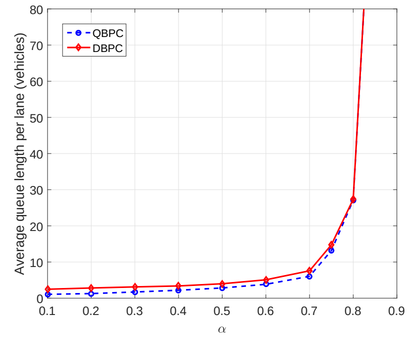
From Fig. 3(b) and Fig. 3(c), DBPC still achieves the same stability region as QBPC whatever the value of C^2 is. In Fig. 3(b) and Fig. 3(e) for IPP arrivals with $C^2 = 2$, they share similar performance with the Poisson arrivals. Compared with Fig. 3(d), the fairness is a little worse in Fig. 3(e) for both DBPC and QBPC, and the difference between DBPC and QBPC is slightly larger. The impact of burstiness is more obvious when $C^2 = 5$. In Fig. 3(f), the gap between DBPC and QBPC regarding fairness is greatly increased under these more bursty traffic, with a maximum improvement of almost 0.5, compared with the results in Fig. 3(d) and Fig. 3(e). However, as shown in Fig. 3(c), for the more bursty arrivals, DBPC experiences a slightly larger average queue length than QBPC.

C. *Heterogeneous Arrivals*

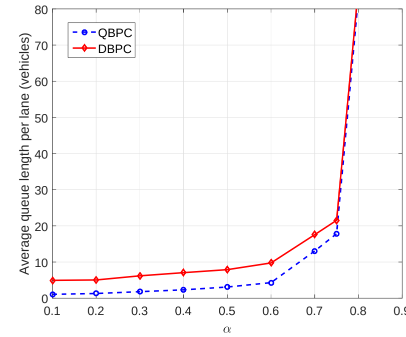
We study the case that the vehicles at each lane arrive following Poisson process, with heterogeneous average arrival rates. Specifically, we set $\vec{\lambda} = [0.2,1,1,0.5,0.2,1,1,0.5] *$



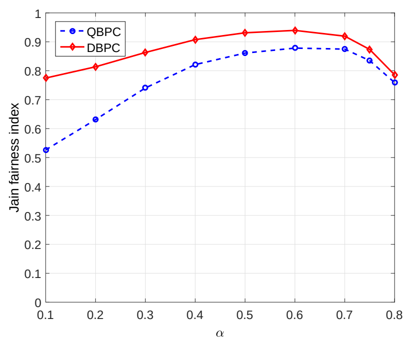
(a) Stability region, Poisson.



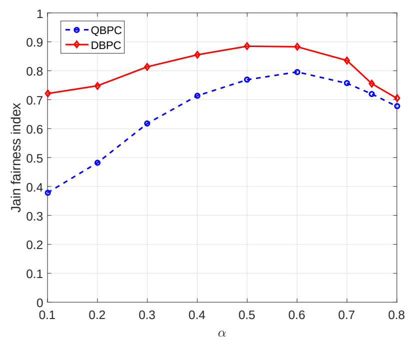
(b) Stability region, IPP, $C^2 = 2$.



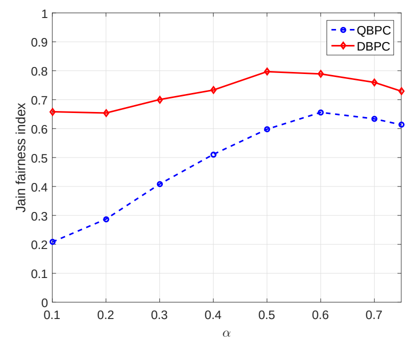
(c) Stability region, IPP, $C^2 = 5$.



(d) Fairness, Poisson.

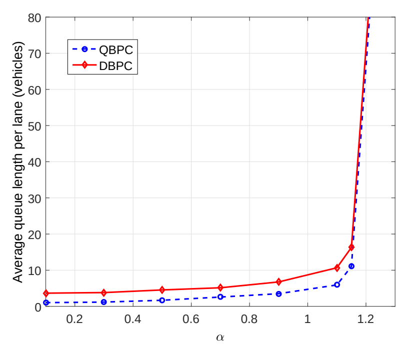


(e) Fairness, IPP, $C^2 = 2$.

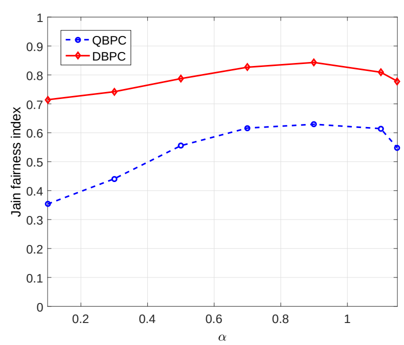


(f) Fairness, IPP, $C^2 = 5$.

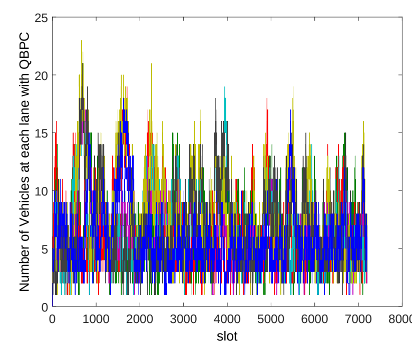
Fig. 3. Performance comparison between QBPC and DBPC with homogeneous arrivals at all lanes. ($\vec{\lambda} = [1, 1, 1, 1, 1, 1, 1, 1] * 0.125$ v/s/l.)



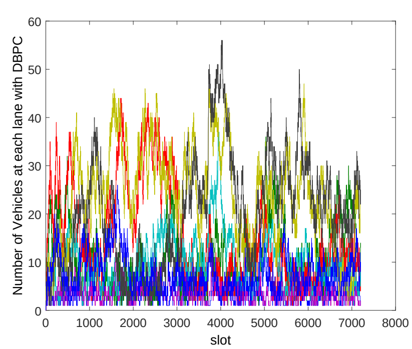
(a) Stability region.



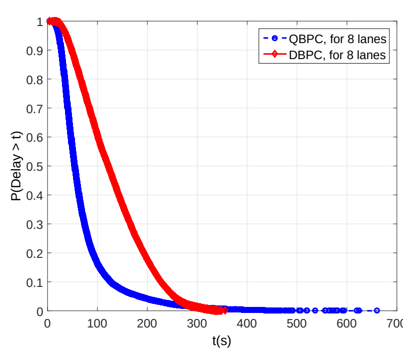
(b) Fairness.



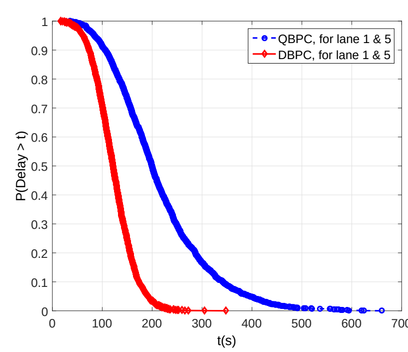
(c) Queue length of QBPC, $\alpha = 1.1$.



(d) Queue length of DBPC, $\alpha = 1.1$.

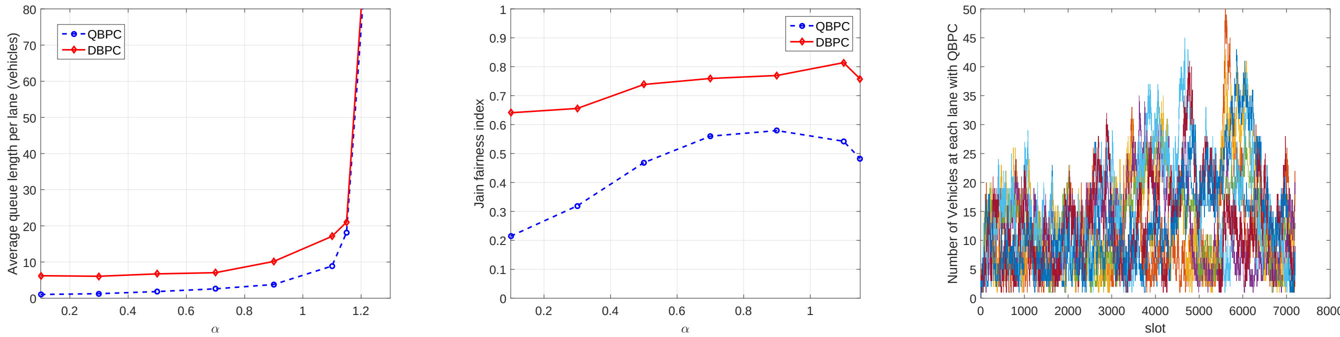


(e) Tail delay for 8 lanes, $\alpha = 1.1$.

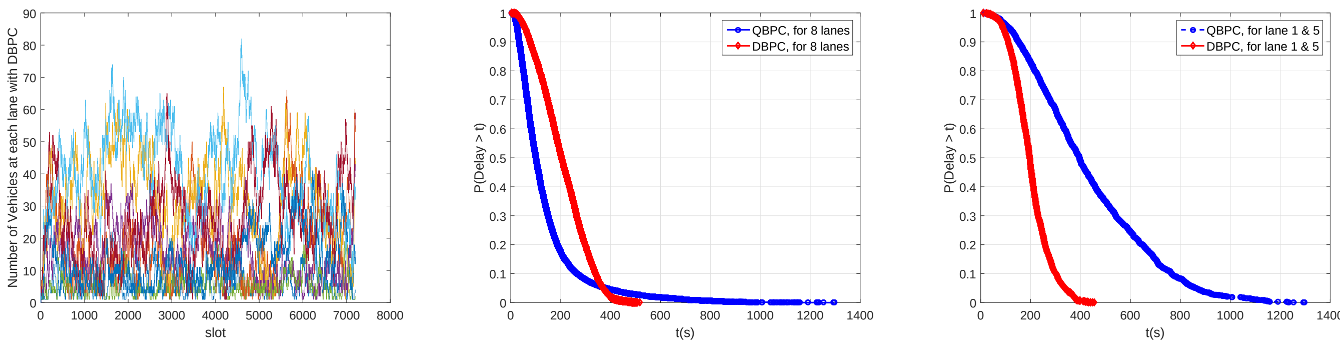


(f) Tail delay for lane 1 & 5, $\alpha = 1.1$.

Fig. 4. Performance comparison between QBPC and DBPC with heterogeneous Poisson arrivals. ($\vec{\lambda} = [0.2, 1, 1, 0.5, 0.2, 1, 1, 0.5] * 0.125$ v/s/l.)



(a) Stability region. (b) Fairness. (c) Queue length of QBPC, $\alpha = 1.15$.



(d) Queue length of DBPC, $\alpha = 1.15$. (e) Tail delay for 8 lanes, $\alpha = 1.15$. (f) Tail delay for lane 1 & 5, $\alpha = 1.15$.

Fig. 5. Performance comparison between QBPC and DBPC with heterogeneous IPP arrivals. ($\vec{\lambda} = [0.2, 1, 1, 0.5, 0.2, 1, 1, 0.5] * 0.125$ v/s/l, $C^2 = 2$.)

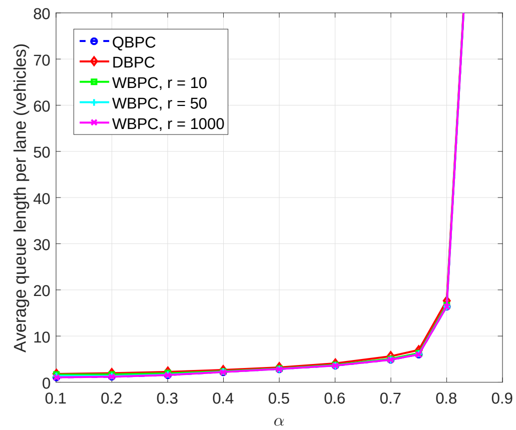
0.125 v/s/l. The results are given in Fig. 4. Fig. 4(a) shows that DBPC and QBPC have the same stability region, and the average queue length of DBPC is a little larger than that of QBPC. In Fig. 4(b), DBPC has much better fairness index than QBPC, with a maximum improvement of over 0.3. Moreover, compared with that in Fig. 3(d) under homogeneous arrivals, the advantage of DBPC in the delay fairness is more obvious under heterogeneous arrivals. We expect that for the vehicles of lanes with low average arrival rates, they may experience a very large delay under QBPC. Because the lanes with low arrival rates may lack subsequent vehicle arrivals in a long period of time compared with lanes with high arrival rates. This does not make their queues to grow, and thus leads to a large tail in the delay distribution. This is verified in Fig. 4(f) for the vehicles of lane 1 and lane 5 with $\alpha = 1.1$, where the tail of the delay distribution under DBPC vanishes much faster than that of QBPC. Taking $\alpha = 1.1$ as an example, the queue length variations of each lane are shown in Fig. 4(c) and Fig. 4(d). Even for the total traffic, as shown in Fig. 4(e), QBPC still has a much longer tail than that of DBPC.

Fig. 5 shows the performance of heterogeneous IPP arrivals with $\vec{\lambda} = [0.2, 1, 1, 0.5, 0.2, 1, 1, 0.5] * 0.125$ v/s/l. The results are similar to those in Fig. 4. The same stability region is achieved for both DBPC and QBPC in Fig. 5(a), and the average queue length of DBPC is a little larger. In Fig. 5(b), DBPC has larger fairness index than QBPC, with a maximum improvement of over 0.4. Still the advantage of DBPC in the delay fairness is more obvious under heterogeneous arrivals than homogeneous arrivals, comparing Fig. 5(b) and Fig. 3(e). In Fig. 5(f), for the vehicles of lane 1 and lane 5 with relatively less arrivals, the tail of the delay distribution under DBPC vanishes much faster than that of QBPC. Moreover, even for the total traffic in Fig. 5(e), DBPC still has a much shorter tail than that of QBPC.

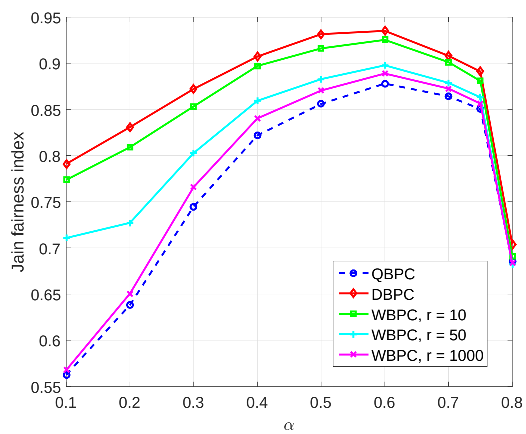
D. Comparison of three schemes

We compare the three control schemes in this section. Fig. 6 considers homogeneous Poisson arrivals, while Fig. 7 shows the comparison under heterogeneous Poisson arrivals. For the WBPC, we assume that $\eta_{i,j}^{(W)} = 1 - \eta_{i,j}^{(Q)}$ and $r = \eta_{i,j}^{(Q)} / \eta_{i,j}^{(W)}$ for all $(i, j) \in \mathcal{M}$. Note that r works as a performance tradeoff parameter, where an increasing r means we value the queue length more, while more emphasis is put on the fairness of delay when r decreases. When $r = 0$, it is DBPC, and when $r \to \infty$, it becomes QBPC.

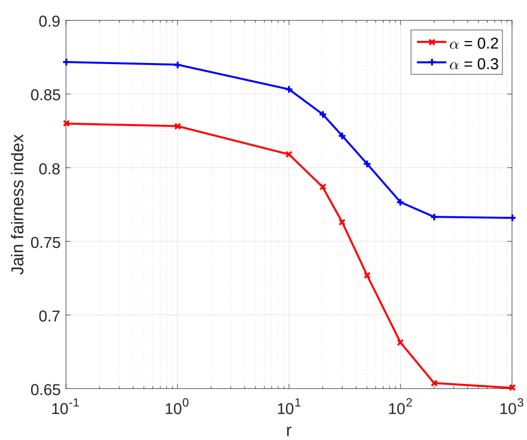
Fig. 6(a) shows that they share the same stability region, and specifically, almost the same average queue length performance. However, in Fig. 6(b), the delay-based control has the best fairness performance, while the queue-based control has the smallest values. The performance of the weighted control varies among them with different values of r. For example, when r is large, e.g., $r = 1000$, which means more emphasis is put on the queue length, the index curve of WBPC is close to that of QBPC; when r is small, e.g., $r = 10$, we value the fairness of delay more, and the index curve of WBPC is close to that of DBPC. Taking $\alpha = 0.2$ and $\alpha = 0.3$



(a) Stability region.



(b) Fairness.



(c) Impact of tradeoff parameter r.

Fig. 6. Comparison under homogeneous Poisson arrivals at all lanes. ($\vec{\lambda} = [1, 1, 1, 1, 1, 1, 1, 1] * 0.125$ v/s/l.)

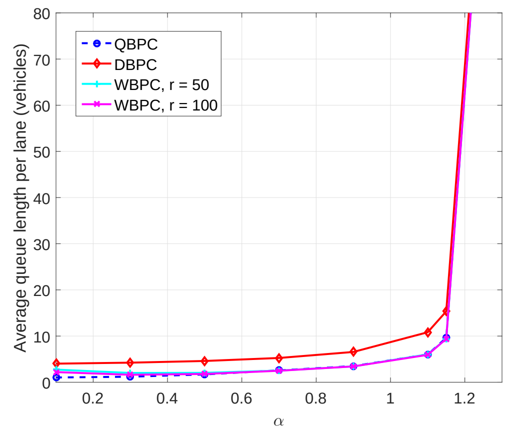
as examples, Fig. 6(c) further shows how the fairness index varies with r. We can see that as r decreases, the index approaches the value of DBPC, and as r increases, the index approaches the value of QBPC. Based on these, we can see that, under homogeneous Poisson arrivals, it's beneficial to adopt the delay-based backpressure control, which not only guarantees the same stability region, but also brings the best delay fairness while still keeps almost the same average queue length.

Under heterogeneous Poisson arrivals, in Fig. 7(a), although they share the same stability region, we can see that there is a small gap between DBPC and QBPC for the average queue length. However, for the weighted control scheme, especially when $\alpha > 0.3$, its average queue length is almost the same with the best QBPC. Moreover, the fairness of WBPC in Fig. 7(b) is much better than that of QBPC. The impact of r is also given in Fig. 7(c). From Fig. 7, we can see that as the heterogeneity of the traffic increases, while still guaranteeing the best fairness performance, the DBPC gradually shows its weakness in the average queue length performance compared with QBPC. In this case, however, the weighted control scheme plays a tradeoff role between DBPC and QBPC, and we can achieve moderate fairness and average queue length performance by choosing appropriate weighting parameters.

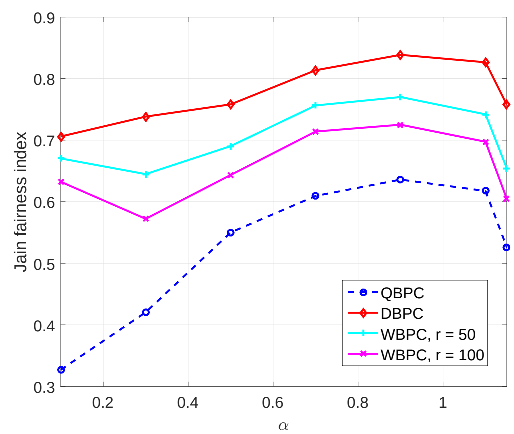
VII. Conclusion and Future Work

In this paper, we studied the delay-based backpressure traffic signal control at isolated intersections. First, the throughput optimality is proved for this delay-based control using the fluid limits technique. Then the advantages of the delay-based control in dealing with the potential excessive delays over the queue-based control are explored under different traffic patterns. Moreover, a weighed backpressure control is proposed and proved to be throughput optimal. Simulation results show that, while guaranteeing the same stability region as the queue-based control, the delay-based control achieves better fairness regarding the delay performance, and this improvement is more obvious under bursty and heterogeneous traffic. Even when there is a difference between the average queue length of the delay-based and queue-based schemes, we can still use the weighted scheme to achieve moderate fairness and average queue length performance by choosing appropriate weighting parameters.

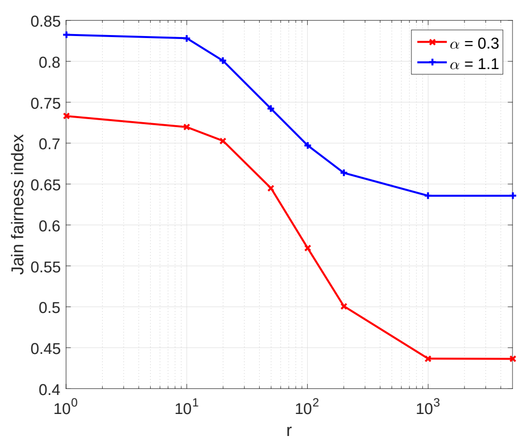
However, since the delay-based backpressure traffic signal control is quite a new approach, there are several directions in which this work can be extended. First, it is important to extend this control for multiple types of traffic, e.g., emergency vehicles like fire truck and ambulance, bicycles, pedestrians and so on, where a more elaborate priority-based structure should be considered on the service of different traffic types. Second, this work can be applied to different shapes of intersection when certain conditions are satisfied, such as, signalized intersection with predefined phases composed of non-conflicting movements. Last but not the least, it is interesting to study this considering network of intersections. However, since the delays of vehicles at neighboring intersections are coupled together, this further complicates the problem. First, the definition of pressure regarding delay needs to be enhanced by incorporating the coupling. Moreover, due to the coupling, a totally distributed design as proposed for the queue-based control in [17] is no longer feasible for the delay-based backpressure control, and it will be centralized over the whole



(a) Stability region.



(b) Fairness.



(c) Impact of tradeoff parameter r.

Fig. 7. Comparison under heterogeneous Poisson arrivals. ($\vec{\lambda} = [0.2, 1, 1, 0.5, 0.2, 1, 1, 0.5] * 0.125$ v/s/l.)

network. Therefore we need to find a way to "decentralize" this for practice usage. For example, a potential approach is to start with a centralized control, where the network can be partitioned into small autonomous regions, and then design heuristic distributed control algorithms.

Appendix A
Proof of Lemma 2

We need to show that there exists a finite time $T > 0$ such that the fluid limits $\hat{f}_{i,j}(T) > f_{i,j}(0)$ hold for all movements $(i,j) \in \mathcal{M}$. This is proved by induction here. First, we show the existence of a finite T that satisfies this condition for at least one movement; then, we prove that for a given set of movements satisfying this condition, at least one additional movement will satisfy this condition by increasing T.

Let us fix an arbitrary $\epsilon_1 > 0$. In the fluid limit model, we have

$$f_{i,j}(\epsilon_1) = f_{i,j}(0) + \lambda_i \epsilon_1 > f_{i,j}(0), \quad \forall (i,j) \in \mathcal{M}, \tag{A.1}$$

and based on Eqs. (15) and (14), there is

$$\begin{aligned} \sum_{(i,j)\in\mathcal{M}} q_{i,j}(\epsilon_1) &\le \sum_{(i,j)\in\mathcal{M}} f_{i,j}(\epsilon_1) \\ &= \sum_{(i,j)\in\mathcal{M}} f_{i,j}(0) + \epsilon_1 \sum_{(i,j)\in\mathcal{M}} \lambda_i \\ &\le 1 + \epsilon_1 \sum_{(i,j)\in\mathcal{M}} \lambda_i \\ &\le K_1. \end{aligned} \tag{A.2}$$

Here we define $K_1 \triangleq 1 + \epsilon_1 \max\{\sum_{(i,j)\in\mathcal{M}} \lambda_i\}$ as a constant, which is the value of $1 + \epsilon_1 \sum_{(i,j)\in\mathcal{M}} \lambda_i$ when the system is at its maximum total arrival rates.

Based on (A.1), we will show by induction the existence of T such that

$$\hat{f}_{i,j}(T) \ge f_{i,j}(\epsilon_1), \quad \forall (i,j) \in \mathcal{M}. \tag{A.3}$$

Base Case: There exists $T_1 > 0$ such that for at least one movement $(i,j) \in \mathcal{M}$,

$$\hat{f}_{i,j}(T_1) \ge f_{i,j}(\epsilon_1). \tag{A.4}$$

Let $T_1 \triangleq \epsilon_1 + K_1$. Suppose (A.4) does not hold, which means there exists at least one vehicle that arrives before slot $\lfloor k\epsilon_1 \rfloor + 1$ and does not leave the system by the end of slot $\lfloor kT_1 \rfloor$. Therefore, at each time slot in $[\lfloor k\epsilon_1 \rfloor + 1, \lfloor kT_1 \rfloor]$, there exists at least one schedule $\vec{p}$ that makes $\mathcal{O}_{\vec{p}}$ in Algorithm 1 positive. As a result, for the delay-based traffic signal control, the schedule determined by Algorithm 1 must serve at least one vehicle each time slot. Then for sufficiently large k, we must have

$$\sum_{(i,j)\in\mathcal{M}} \left(\hat{F}_{i,j}^{(k)}(kT_1) - \hat{F}_{i,j}^{(k)}(k\epsilon_1) \right) \ge \lfloor kT_1 \rfloor - \lfloor k\epsilon_1 \rfloor. \tag{A.5}$$

Dividing both sides of (A.5) by k and letting $k \to \infty$, we obtain

$$\sum_{(i,j)\in\mathcal{M}} \left(\hat{f}_{i,j}(T_1) - \hat{f}_{i,j}(\epsilon_1) \right) \ge T_1 - \epsilon_1 = K_1. \tag{A.6}$$

Based on (A.2) and Eq. (15), there is

$$\begin{aligned} \sum_{(i,j)\in\mathcal{M}} \hat{f}_{i,j}(T_1) &\ge \sum_{(i,j)\in\mathcal{M}} \hat{f}_{i,j}(\epsilon_1) + K_1 \\ &\ge \sum_{(i,j)\in\mathcal{M}} \hat{f}_{i,j}(\epsilon_1) + \sum_{(i,j)\in\mathcal{M}} q_{i,j}(\epsilon_1) \\ &= \sum_{(i,j)\in\mathcal{M}} f_{i,j}(\epsilon_1). \end{aligned} \tag{A.7}$$

Therefore, $\hat{f}_{i,j}(T_1) \geq f_{i,j}(\epsilon_1)$ needs to hold for at least one movement $(i,j) \in \mathcal{M}$, to make $\sum_{(i,j)\in\mathcal{M}} \hat{f}_{i,j}(T_1) \geq \sum_{(i,j)\in\mathcal{M}} f_{i,j}(\epsilon_1)$ hold. This contradiction proves the base case.

Induction Step: Suppose that there exists T_l and a subset $S_l \subseteq \mathcal{M}$ of cardinality l, such that for all $(i,j) \in S_l$ we have

$$\hat{f}_{i,j}(T_l) \geq f_{i,j}(\epsilon_1). \tag{A.8}$$

Then there exists $T_{l+1} \geq T_l$ and a movement $(i,j) \in \mathcal{M}\backslash S_l$ such that

$$\hat{f}_{i,j}(T_{l+1}) \geq f_{i,j}(\epsilon_1). \tag{A.9}$$

The subset S_{l+1} of cardinality $l+1$ is defined as $S_{l+1} = S_l \cup \{(i,j)\}$. We prove the induction step for $l=1$. The generalization for arbitrary $l>1$ is straightforward. Therefore, we need to prove that given S_1 and T_1, there exists $T_2 \geq T_1$ such that for at least two different movements (A.9) holds with T_2.

Let (i',j') denote the movement that satisfies (A.8) with T_1, and we have

$$\hat{f}_{i',j'}(t) \geq f_{i',j'}(\epsilon_1), \quad t \geq T_1, \tag{A.10}$$

according to the base case. Based on (A.2), we observe that

$$\sum_{(i,j)\neq(i',j')} \Big(f_{i,j}(\epsilon_1) - \hat{f}_{i,j}(T_1)\Big) \leq K_1. \tag{A.11}$$

Suppose that for all $t \geq T_1$, we have

$$\hat{f}_{i,j}(t) < f_{i,j}(\epsilon_1), \quad \forall (i,j) \neq (i',j'). \tag{A.12}$$

In the following, we provide a choice of $T_2 > T_1$ such that the assumption (A.12) leads to a contradiction.

We view each path $X^{(k)}(t)$ after time slot $\lceil kT_1 \rceil$ as a generalized system with movements in $S_1 = \{(i',j')\}$, and consider the time slots unavailable to S_1 when vehicles of movements $(i,j) \in \mathcal{M}\backslash S_1$ are served. Based on (A.11) and (A.12), we obtain

$$\begin{aligned} h_{S_1}(t) &\leq \sum_{(i,j)\neq(i',j')} \Big(\hat{f}_{i,j}(t) - \hat{f}_{i,j}(T_1)\Big) \\ &< \sum_{(i,j)\neq(i',j')} \Big(f_{i,j}(\epsilon_1) - \hat{f}_{i,j}(T_1)\Big) \\ &\leq K_1. \end{aligned} \tag{A.13}$$

for $t \geq T_1$, where $h_{S_1}(t)$ denotes the amount of time unavailable to S_1 in $(T_1, t]$ in the scaled system. Since the time unavailable to S_1 is bounded, there exists a sufficiently large $t \geq T_1$ such that the time given to $(i,j) \in \mathcal{M}\backslash S_1$ is negligible. Therefore we can focus on S_1. Based on Lemma 1, we know that $\lambda_{i'} w_{i',j'}(t) = q_{i',j'}(t)$ for $t \geq T_1$. Following from Proposition 3, it is easy to prove that the movements in S_1 are stable under the delay-based backpressure control, and thus $q_{i',j'}(t) \leq C_1$ for $t \geq T_1$. As a result,

$$w_{i',j'}(t) \leq \frac{C_1}{\lambda_{i'}}, \quad \forall t \geq T_1, \tag{A.14}$$

where the constant C_1 depends only on T_1 and K_1, but not time t. Based on (A.12), we know that for each movement $(i,j) \in \mathcal{M}\backslash S_1$ there are vehicles that arrive at the system by time ϵ_1 and have not been served by time t, and thus

$$t - \epsilon_1 \leq w_{i,j}(t) \leq t, \quad \forall (i,j) \in \mathcal{M}\backslash S_1. \tag{A.15}$$

Note that $w_{i',j'}(t)$ is bounded in (A.14), and $w_{i,j}(t)$ increases linearly on the order of t for $(i,j) \in \mathcal{M}\backslash S_1$ from (A.15). As a result, there exists a large T_2' such that for $t > T_2'$, the movements in $\mathcal{M}\backslash S_1$ will be scheduled at all the time slots between $[\lfloor kT_2' \rfloor + 1, \lfloor kt \rfloor]$ under the delay-based backpressure control. Then we can choose $T_2 > T_2'$ and make

$$h_{S_1}(T_2) \geq T_2 - T_2' > K_1. \tag{A.16}$$

This contradicts with (A.13), and thus the assumption (A.12) does not hold. As a result, there exists a large T_2 such that

$$\hat{f}_{i,j}(T_2) \geq f_{i,j}(\epsilon_1), \tag{A.17}$$

for at least one movement $(i,j) \in \mathcal{M}\backslash S_1$.

Appendix B
Proof of Proposition 4

Let $L(\vec{q}(t))$ denote the Lyapunov function as

$$L(\vec{q}(t)) = \frac{1}{2} \sum_{(i,j)\in\mathcal{M}} \gamma_{i,j} [\eta_{i,j}^{(Q)} (q_{i,j}(t))^2 + \eta_{i,j}^{(W)} \frac{(q_{i,j}(t))^2}{\lambda_i}]. \tag{B.1}$$

Assume $\vec{\lambda}$ is strictly inside Λ, and then there exists a vector $\vec{\phi} \in Co(\mathcal{S}_P)$ such that $\vec{\lambda} < \vec{\phi}$, i.e. $\lambda_i < \phi_{i,j}$ for all $(i,j) \in \mathcal{M}$.

$$\begin{aligned} &\frac{d}{dt} L(\vec{q}(t)) \\ &= \sum_{(i,j)\in\mathcal{M}} \gamma_{i,j} [\eta_{i,j}^{(Q)} + \frac{\eta_{i,j}^{(W)}}{\lambda_i}] q_{i,j}(t) \frac{d}{dt} q_{i,j}(t) \\ &= \sum_{(i,j)\in\mathcal{M}} \gamma_{i,j} [\eta_{i,j}^{(Q)} + \frac{\eta_{i,j}^{(W)}}{\lambda_i}] q_{i,j}(t) (\lambda_i - \pi_{i,j}(t)) \\ &= \sum_{(i,j)\in\mathcal{M}} \gamma_{i,j} [\eta_{i,j}^{(Q)} q_{i,j}(t) + \eta_{i,j}^{(W)} \frac{q_{i,j}(t)}{\lambda_i}] (\lambda_i - \phi_{i,j}(t)) \\ &\quad + \sum_{(i,j)\in\mathcal{M}} \gamma_{i,j} [\eta_{i,j}^{(Q)} q_{i,j}(t) + \eta_{i,j}^{(W)} \frac{q_{i,j}(t)}{\lambda_i}] (\phi_{i,j}(t) - \pi_{i,j}(t)). \end{aligned} \tag{B.2}$$

Let us choose $\delta_3 > 0$ such that $L(\vec{q}(t)) \geq \delta_1 > 0$ implies $\max_{(i,j)\in\mathcal{M}} [\eta_{i,j}^{(Q)} + \frac{\eta_{i,j}^{(W)}}{\lambda_i}] q_{i,j}(t) \geq \delta_3$. Based on the fact that the sum of non-negative elements must be no less than each element, we have

$$\begin{aligned} &\sum_{(i,j)\in\mathcal{M}} \gamma_{i,j} [\eta_{i,j}^{(Q)} q_{i,j}(t) + \eta_{i,j}^{(W)} \frac{q_{i,j}(t)}{\lambda_i}] (\phi_{i,j}(t) - \lambda_i) \\ &\geq \gamma_{i',j'} \Big[\max_{(i,j)\in\mathcal{M}} [\eta_{i,j}^{(Q)} + \frac{\eta_{i,j}^{(W)}}{\lambda_i}] q_{i,j}(t)\Big] (\phi_{i',j'}(t) - \lambda_{i'}) \\ &\geq \min_{(i,j)\in\mathcal{M}} \gamma_{i,j} \Big[\max_{(i,j)\in\mathcal{M}} [\eta_{i,j}^{(Q)} + \frac{\eta_{i,j}^{(W)}}{\lambda_i}] q_{i,j}(t)\Big] \min_{(i,j)\in\mathcal{M}} (\phi_{i,j}(t) - \lambda_i) \\ &\geq \min_{(i,j)\in\mathcal{M}} \gamma_{i,j} \delta_3 \min_{(i,j)\in\mathcal{M}} (\phi_{i,j}(t) - \lambda_i). \end{aligned} \tag{B.3}$$

We have $(i',j') = \underset{(i,j)\in\mathcal{M}}{\text{argmax}}[\eta_{i,j}^{(Q)} q_{i,j}(t) + \eta_{i,j}^{(W)} \frac{q_{i,j}(t)}{\lambda_i}]$ in the first inequality.

Then the first sum in Eq. (B.2) is bounded as follows:

$$\begin{aligned} & \sum_{(i,j)\in\mathcal{M}} \gamma_{i,j}[\eta_{i,j}^{(Q)} q_{i,j}(t) + \eta_{i,j}^{(W)} \frac{q_{i,j}(t)}{\lambda_i}](\lambda_i - \phi_{i,j}(t)) \\ \leq & \ -\min_{(i,j)\in\mathcal{M}} \gamma_{i,j}\delta_3 \min_{(i,j)\in\mathcal{M}} (\phi_{i,j}(t) - \lambda_i) \\ \triangleq & \ -\delta_2 < 0. \end{aligned} \tag{B.4}$$

The second term in Eq. (B.2) is non-positive if

$$\vec{\pi}(t) \in \underset{\vec{\phi}\in Co(\mathcal{S}_P)}{\text{argmax}} \sum_{(i,j)\in\mathcal{M}} \gamma_{i,j}\cdot[\eta_{i,j}^{(Q)} q_{i,j}(t) + \eta_{i,j}^{(W)} \frac{q_{i,j}(t)}{\lambda_i}]\cdot\phi_{i,j}(t), \tag{B.5}$$

which means

$$\begin{aligned} & \sum_{(i,j)\in\mathcal{M}} \gamma_{i,j}[\eta_{i,j}^{(Q)} q_{i,j}(t) + \eta_{i,j}^{(W)} \frac{q_{i,j}(t)}{\lambda_i}]\phi_{i,j}(t) \leq \\ & \sum_{(i,j)\in\mathcal{M}} \gamma_{i,j}[\eta_{i,j}^{(Q)} q_{i,j}(t) + \eta_{i,j}^{(W)} \frac{q_{i,j}(t)}{\lambda_i}]\pi_{i,j}(t). \end{aligned} \tag{B.6}$$

We know that (B.6) holds because that the underlying weighted backpressure scheduler indeed maximizes $\sum_{(i,j)\in\mathcal{M}} \gamma_{i,j}[\eta_{i,j}^{(Q)} q_{i,j}(t) + \eta_{i,j}^{(W)} \frac{q_{i,j}(t)}{\lambda_i}]\phi_{i,j}(t)$ due to the linear relationship $w_{i,j}(t) = \frac{q_{i,j}(t)}{\lambda_i}$ for $t \geq T$.

As a result, for any $\delta_1 > 0$, there exists $\delta_2 > 0$ and a finite time $T > 0$ such that $L(\vec{q}(t)) \geq \delta_1$ implies $\frac{d}{dt}L(\vec{q}(t)) \leq -\delta_2$ for any regular time $t \geq T$. Then the following proof will be the same with that of Proposition 3, and we can finish the proof of Proposition 4.

References

[1] E. Uhlemann, "Introducing connected vehicles [connected vehicles]," *IEEE Vehicular Technology Magazine*, vol. 10, no. 1, pp. 23–31, 2015.

[2] X. Wu, S. Subramanian, R. Guha, R. G. White, J. Li, K. W. Lu, A. Bucceri, and T. Zhang, "Vehicular communications using dsrc: challenges, enhancements, and evolution," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 399–408, 2013.

[3] J. A. Fernandez, K. Borries, L. Cheng, B. V. Kumar, D. D. Stancil, and F. Bai, "Performance of the 802.11 p physical layer in vehicle-to-vehicle environments," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 1, pp. 3–14, 2012.

[4] Y. Feng, K. L. Head, S. Khoshmagham, and M. Zamanipour, "A real-time adaptive signal control in a connected vehicle environment," *Transportation Research Part C: Emerging Technologies*, vol. 55, pp. 460–473, 2015.

[5] K. Yang, S. I. Guler, and M. Menendez, "Isolated intersection control for various levels of vehicle technology: Conventional, connected, and automated vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 72, pp. 109–129, 2016.

[6] J. Zheng and H. X. Liu, "Estimating traffic volumes for signalized intersections using connected vehicle data," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 347–362, 2017.

[7] G. Comert, "Queue length estimation from probe vehicles at isolated intersections: Estimators for primary parameters," *European Journal of Operational Research*, vol. 252, no. 2, pp. 502–521, 2016.

[8] X. Ban, R. Herring, P. Hao, and A. Bayen, "Delay pattern estimation for signalized intersections using sampled travel times," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2130, pp. 109–119, 2009.

[9] P. Hao, X. Ban, and J. Whon Yu, "Kinematic equation-based vehicle queue location estimation method for signalized intersections using mobile sensor data," *Journal of Intelligent Transportation Systems*, vol. 19, no. 3, pp. 256–272, 2015.

[10] R. Wunderlich, C. Liu, I. Elhanany, and T. Urbanik, "A novel signal-scheduling algorithm with quality-of-service provisioning for an isolated intersection," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 9, no. 3, pp. 536–547, 2008.

[11] K. Pandit, D. Ghosal, H. M. Zhang, and C.-N. Chuah, "Adaptive traffic signal control with vehicular ad hoc networks," *Vehicular Technology, IEEE Transactions on*, vol. 62, no. 4, pp. 1459–1471, 2013.

[12] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *Automatic Control, IEEE Transactions on*, vol. 37, no. 12, pp. 1936–1948, 1992.

[13] M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic power allocation and routing for time-varying wireless networks," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 1, pp. 89–103, 2005.

[14] M. J. Neely and R. Urgaonkar, "Optimal backpressure routing for wireless networks with multi-receiver diversity," *Ad Hoc Networks*, vol. 7, no. 5, pp. 862–881, 2009.

[15] B. Ji, C. Joo, and N. B. Shroff, "Delay-based back-pressure scheduling in multihop wireless networks," *Networking, IEEE/ACM Transactions on*, vol. 21, no. 5, pp. 1539–1552, 2013.

[16] A. Sinha and E. Modiano, "Optimal control for generalized network-flow problems," *To appear in IEEE INFOCOM*, 2017.

[17] T. Wongpiromsarn, T. Uthaicharoenpong, Y. Wang, E. Frazzoli, and D. Wang, "Distributed traffic signal control for maximum network throughput," in *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*. IEEE, 2012, pp. 588–595.

[18] J. Gregoire and et al, "Capacity-aware backpressure traffic signal control," *Control of Network Systems, IEEE Transactions on*, vol. 2, no. 2, pp. 164–173, 2015.

[19] P. Varaiya, "A universal feedback control policy for arbitrary networks of signalized intersections," *Published online. URL: http://paleale. eecs. berkeley. edu/~ varaiya/papers\ _ps. dir/090801-Intersectionsv5. pdf*, 2009.

[20] ——, "The max-pressure controller for arbitrary networks of signalized intersections," in *Advances in Dynamic Network Modeling in Complex Transportation Systems*. Springer, 2013, pp. 27–66.

[21] T. Le, P. Kovács, N. Walton, H. L. Vu, L. L. Andrew, and S. S. Hoogendoorn, "Decentralized signal control for urban road networks," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 431–450, 2015.

[22] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, "Scheduling in a queuing system with asynchronously varying service rates," *Probability in the Engineering and Informational Sciences*, vol. 18, no. 02, pp. 191–217, 2004.

[23] P. Pecherková, M. Flidr, and J. Dunik, "Application of estimation techniques on queue lengths estimation in traffic network," in *Cybernetic Intelligent Systems, 2008. CIS 2008. 7th IEEE International Conference on*. IEEE, 2008, pp. 1–6.

[24] P. Pecherková, J. Dunık, and M. Flıdr, "Modelling and simultaneous estimation of state and parameters of traffic system," *Robotics, Automation and Control*, vol. 1, no. 1, pp. 319–336, 2008.

[25] R. Jain, D.-M. Chiu, and W. R. Hawe, *A quantitative measure of fairness and discrimination for resource allocation in shared computer system*. Eastern Research Laboratory, Digital Equipment Corporation Hudson, MA, 1984, vol. 38.

[26] J. Wu, Y. Bao, G. Miao, S. Zhou, and Z. Niu, "Base-station sleeping control and power matching for energy–delay tradeoffs with bursty traffic," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3657–3675, 2016.

Jian Wu received her B.S. and Ph.D. degrees in Electronic Engineering from Beijing Jiaotong University in 2009 and Tsinghua University in 2015, respectively. She is currently a Postdoc in Computer Science Department at University of California, Davis, CA, USA. She received the Best Paper Award from the 23th IEEE International Conference on Communication Technology (ICCT) in 2011. Her research interests include vehicular networks, cloud computing, radio access networks, and green wireless communications.

Dipak Ghosal received the B.Tech degree in electrical engineering from the Indian Institute of Technology, Kanpur, India, in 1983, the M.S. degree in computer science from the Indian Institute of Science, Bangalore, India, in 1985, and the Ph.D. degree in computer science from the University of Louisiana, Lafayette, in 1988. From 1988 to 1990, he was a Research Associate at the Institute for Advanced Computer Studies, University of Maryland at College Park. From 1990 to 1996, he was a Member of Technical Staff at Bell Communications Research (Bellcore and later Telcordia), Red Bank, NJ. Since 1996, he has been with the faculty of the Department of Computer Science, University of California at Davis. His research interests include control and management of high-speed networks, wireless and vehicular adhoc networks, parallel and distributed computing, and performance evaluation of computer and communication systems.

Michael Zhang is currently a professor in the Civil and Environmental Engineering Department at University of California Davis. His research is in traffic operations and control, transportation network analysis and intelligent transportation systems. Professor Zhang received his BS degree in Civil Engineering from Tongji University, and MS and PhD degrees in Engineering from University of California Irvine. He is an Area Editor of the journal Network and Spatial Economics, and an Associate Editor of Transportation Research, Part B: Methodological, and Transportation Science.

Chen-Nee Chuah is a Professor in Electrical and Computer Engineering at the University of California, Davis. She received her B.S. in Electrical Engineering from Rutgers University, and her M. S. and Ph.D. in Electrical Engineering and Computer Sciences from the University of California, Berkeley. Her research interests include Internet measurements, network management, and applying data and network science techniques to online social networks, security detection, digital healthcare, and intelligent transportation systems. Chuah is a Fellow of the IEEE and an ACM Distinguished Scientist. She was a recipient of the NSF CAREER Award (2003) and the UC Davis College of Engineering Outstanding Junior Faculty Award (2004). Chuah was named a Chancellors Fellow of UC Davis in 2008. She has served as an Associate Editor for IEEE/ACM Transactions on Networking and IEEE Transactions on Mobile Computing.