Transportation Research Part C 36 (2013) 177–195

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Max pressure control of a network of signalized intersections

Pravin Varaiya *

Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720-1770, United States

ARTICLE INFO

Article history:
Received 18 June 2013
Received in revised form 27 August 2013
Accepted 28 August 2013

Keywords:
Max pressure
Local control
Traffic responsive
Adaptive signal control
Bounded queues
Store-and-forward model
Queueing network
Network stability

ABSTRACT

The control of a network of signalized intersections is considered. Vehicles arrive in iid (independent, identically distributed) streams at entry links, independently make turns at intersections with fixed probabilities or turn ratios, and leave the network upon reaching an exit link. There is a separate queue for each turn movement at each intersection. These are point queues with no limit on storage capacity. At each time the control selects a 'stage', which actuates a set of simultaneous vehicle movements at given iid saturation flow rates. Network evolution is modeled as a controlled store-and-forward (SF) queuing network. The control can be a function of the state, which is the vector of all the queue lengths. A set of demands is said to be *feasible* if there is a control that *stabilizes* the queues, that is the time-average of every mean queue length is bounded. The set of feasible demands D is a convex set defined by a collection of linear inequalities involving only the mean values of the demands, turn ratios and saturation rates. If the demands are in the interior D^o of D, there is a fixed-time control that stabilizes the queues. The max pressure (MP) control is introduced. At each intersection, MP selects a stage that depends only on the queues adjacent to the intersection. The MP control does *not* require knowledge of the mean demands. MP stabilizes the network if the demand is in D^o. Thus MP maximizes network throughput. MP does require knowledge of mean turn ratios and saturation rates, but an adaptive version of MP will have the same performance, if turn movements and saturation rates can be measured. The advantage of MP over other SF network control formulations is that it (1) only requires local information at each intersection and (2) provably maximizes throughput. Examples show that other local controllers, including priority service and fully actuated control, may not be stabilizing. Several modifications of MP are offered including one that guarantees minimum green for each approach and another that considers weighted queues; also discussed is the effect of finite storage capacity.

1. Introduction

The evolution of traffic in a signalized road network is modeled in this paper as a network of queues. The state of this network is the vector of all queue lengths at all intersections. The signal control at any time permits certain simultaneous turn movements at each intersection at pre-specified saturation rates. Miller (1963) studies the queue for one approach at a single intersection modeled by the equation

$$x(t+1) = x(t) - c(t) \wedge x(t) + d(t). \quad (1)$$

* Tel.: +1 510 642 5270.
E-mail address: varaiya@berkeley.edu

0968-090X/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.trc.2013.08.014

Here $x(t)$ is the queue length at the beginning of period t, $c(t)$ is the number of vehicles that can potentially depart in period t when the signal is actuated, and $d(t)$ is the demand in period t. $c(t)$ and $d(t)$ are iid (independent, identically distributed) random variables with mean c and d vehicles per period, respectively. $y \wedge z = \min\{y,z\}$. The system (1) is said to be *stable* if the mean queue length is bounded. This system is stable if $d < c$, that is, the mean demand is smaller than the service rate. Under this condition, Miller (1963) estimates the mean and variance of the queue length in equilibrium.

The single-queue model (1) extends to a *network* of signalized intersections. In such a model for every approach k at an intersection the queue length $x_k(t)$ evolves according to

$$x_k(t+1) = x_k(t) - C_k(t+1)S_k(t) \wedge x_k(t) + \sum_l a_{k,l}(t) + d_k(t+1). \tag{2}$$

Here $S_k(t) = 1$ or 0, accordingly as the intersection signal control permits or forbids movement of vehicles from queue $x_k(t)$, $C_k(t+1)$ is the random number of vehicles that could depart in period t if $S_k(t) = 1$, $\sum_l a_{k,l}(t)$ is the sum over all arrivals from other intersections in the network, and $d_k(t+1)$ is the sum of all arrivals from outside the network.

The network model (2) has not been analyzed in the literature, even for a fixed-time control. (A noteworthy exception is the approximate equilibrium analysis of the network model in Osorio and Bierlaire (2009).) In particular, it seems not to be known whether, with stochastic arrivals and service, a particular fixed-time control will stabilize the network, i.e., all queues have bounded mean. That question is settled by Theorem 3 which states that a fixed-time control stabilizes the network if and only if for each queue the mean total arrival rate is smaller that the mean service rate.

Instead of an analysis of the statistical properties of (2) published work has focused on the design of feedback or traffic-responsive controls in which the intersection signals are selected as a function of the current state, the vector of all queue lengths in the network. This large literature is not summarized here in detail, since there are good reviews in Mirchandani and Head (2001), Papageorgiou et al. (2003), Osorio and Bierlaire (2008), and Xie et al. (2012). Here we discuss some major differences between this literature and the contributions of the present paper. A more critical comparison between this literature and max pressure or MP control is available in Varaiya (2013).

The calculation of signal control in systems such as OPAC (Gartner et al., 2001), RHODES (Mirchandani and Head, 2001), and in the widely deployed SCOOT Robertson and Bretherton (1991) is distributed i.e., the control of each intersection is set independently, with the objective of minimizing some measure of upstream queues over some horizon. OPAC uses upstream flow measurements to predict the flow over a rolling horizon (usually one cycle). RHODES uses a 'dynamic network model' to estimate link flows which are used to adjust the control at each intersection based on prediction of vehicle arrivals. The network model includes demands, turn probabilities and saturation flow rates as parameters. SCOOT measures upstream flows at each intersection to update a queue model. None of these models considers the impact of signal control on *downstream* queues. The counter-examples in Section 5 suggest that such controls may therefore be destabilizing. (It is not possible to provide precise counter-examples to these control schemes since they are not mathematically described in the published literature.)

On the other hand, TUC (Diakaki et al., 2003; Aboudolas et al., 2009b) prescribes a centralized control, which may require significant communication infrastructure. By contrast, the calculations in MP are *local*: the evaluation of MP at each intersection at any time requires knowledge only of the queues at adjacent links at that time. According to Lindley (2012), traffic-responsive and adaptive control achieve large benefits but fewer than 10% of intersections in the US use adaptive signals, because of the deployment cost of detection and communication and uncertainty about benefits.

Second, these signal control systems attempt to minimize the cost over an infinite or finite rolling horizon. Calculation of this future cost requires *prediction* of future demands and turn ratios, and if the prediction is biased, the control strategy will not be optimal, see Varaiya (2013). By contrast, MP requires *no* knowledge of the demand, although it does require knowledge of turn ratios. However, the adaptive version of MP, AMP, can estimate these turn ratios.

The third difference is theoretical. None of these systems comes with a guarantee that the resulting closed loop system will be stable. Theorem 2 shows that MP is a stabilizing control if there exists any stabilizing control. Thus MP maximizes throughput.

The paper is organized as follows. Section 2 formulates a static flow problem, with mean (average) demands, turn ratios and saturation rates. In this formulation, the set D of feasible demands is characterized by a set of linear inequalities and each $d \in D^o$ can be supported by a fixed-time controller. Section 3 describes the basic MP control, and shows that it maximizes throughput. Section 4 considers several variations of the basic MP including adaptive MP, and the use of weighted queues. Section 5 is devoted to three examples: a fully actuated control of a two-intersection network, a utilization-maximizing and a priority-based control for a single intersection, all of which are de-stabilizing even though in each case there exists a stabilizing fixed-time control. Section 6 summarizes the conclusions, briefly discusses the limitations of the present formulation and directions for further work. Most of the technical proofs are collected in Appendix A.

A note on max pressure: The max pressure algorithm was first presented in Tassiulas and Ephremides (1992), which considers the routing and scheduling of packet transmission in a wireless network. In that context, packets may not be simultaneously transmitted over two interfering links. (In the traffic context of this paper, vehicles may not make simultaneous movements if these can cause collisions.) In packet networks, the term backpressure policy has been adopted. The name max pressure may have been coined by Dai and Lin (2005), and it seems to be the preferred term in scheduling and routing in flexible manufacturing networks. There is a large literature on max pressure or backpressure algorithms.