Queues with server vacations as a model for pre-timed signalised urban traffic

António Pacheco
CEMAT and Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1, 1049-001 LISBOA, Portugal, apacheco@math.tecnico.ulisboa.pt

Maria Lurdes Simões
Universidade do Porto, Faculdade de Engenharia and CONSTRUCT, Rua Dr. Roberto Frias, s/n, P-4200-465 PORTO, Portugal, lurdes.simoes@fe.up.pt

Paula Milheiro-Oliveira
Universidade do Porto, Faculdade de Engenharia and CMUP, Rua Dr. Roberto Frias, s/n, P-4200-465 PORTO, Portugal, poliv@fe.up.pt

A queuing system resulting from a signalized intersection regulated by pre-timed control in a network urban traffic is considered. Modeling the queue length and the delay of vehicles is crucial to evaluate the performance of intersections equipped with traffic signals. Air quality and rational use of energy also depend on an efficient management of the intersection. These traffic systems have the specificity that the server (green signal) is deactivated (red signal) during a fixed period of time. In the present work, an $M/D/1$ queue with a server that occasionally takes vacations is analysed. The mean delays of vehicles and the mean queue length are computed and compared with those obtained by using a detailed simulation model in a case study. We found that, in general, the mean delays of vehicles given by the proposed queuing model provide a good approximation, but it can provide slightly smaller values than those obtained in the simulation model for large traffic flows. This result is of interest for traffic engineers, as the approaches one can find in the literature for large signalised urban traffic flows are subject to criticism.

Queues with server vacations, Markov regenerative processes, signalised urban traffic, traffic models
1 Introduction

Over the last decades, the theory of traffic signals gave a lot of attention to the estimation of delays and queue lengths that result from the adoption of a signal control strategy at individual intersections. Modeling the queue length and the delay (sojourn time in system) of vehicles is fundamental if one wants to study the performance of signalized intersections. The difficulty in estimating vehicle delay at pre-timed signalised intersections is demonstrated by the variety of delay models that have been proposed over many years. Delay models rely, in some cases, on the nature of the way traffic systems are modelled, according to the desired level of detail. Delay estimation may use, in particular, microscopic, mesoscopic, and macroscopic models, which can be either deterministic or stochastic. Microscopic models include mostly car-following models, while queuing systems are of mesoscopic type, and popular models like HCM and Webster’s delay model are macroscopic. As will be seen further in this paper, we will be particularly interested in stochastic mesoscopic models based on some queuing theory model.

The main difficulties involved in the analysis by means of queuing theory come from the need of a good characterization of the circulating vehicles and drivers and from the fact that the deactivation of the server for fixed periods of time (red signal) has to be incorporated in the behavior of the queue. This is essentially the reason why $M/M/1$, $M/G/1$ and $G/M/1$ models do not satisfactorily fit the waiting behavior in pre-timed signalised intersections. As far as we know, no theory is available for the development of a single analytical model to be used for an arbitrary saturation level in such type of intersections.

A detailed study of signalized intersections with fixed size green sig-
nal period can be found in Webster (1958), where the widely accepted Webster’s delay model is proposed:

\[ d = \frac{C(1 - \frac{g}{C})^2}{2(1 - X \frac{g}{C})} + \frac{X^2}{2q(1 - X)} - 0.65 \frac{3}{q^2} X^{(2+5\frac{g}{C})}, \]  

(1)

where \( d \) = average delay per vehicle on a given particular approach (s); \( C \) = cycle length (s), the length of time for which a traffic signal displays a green light added of the the length of time for which the traffic signal displays a red light; \( q \) = flow rate [vehicles per second (veh/s)]; \( g \) = effective green time duration (s); and \( X \) = saturation degree. This formula, which was derived as a semi-empirical expression based upon computer simulation, is only applicable under unsaturated conditions, that is, when demand is smaller than capacity. The first term in the expression can be interpreted as representing the delay of drivers in a scenario of deterministic equally spaced arrivals; the second term intends to take into account randomness in vehicle arrivals, corresponding to the delay in a \( M/D/1 \) queue with service time equal to \( \frac{C}{s.g} \); the third term results from applying a regression model to fit \( d \). The negative sign of the third term is due to the fact that the first two terms in the expression overestimate the delay.

The Highway Capacity Manual 2010 (HCM2010; see Transportation Research Board 2010) presents another model to calculate the average control delay per vehicle for isolated intersections operating with a pre-timed signal in conditions of random arrivals. If no initial queue is assumed the model resumes to:

\[ d = \frac{C \left(1 - \frac{g}{C}\right)^2}{2 \left[1 - \min(1, X) \frac{g}{C}\right]} + 900T \left[(X - 1) + \sqrt{(X - 1)^2 + \frac{4X}{cT}}\right], \]  

(2)

where \( T \) is the duration of the time interval under study (h) and \( c \) is the lane group capacity (veh/h). According to the definition in HCM2010, control delay includes initial deceleration delay, queue move up time, stopped delay and final acceleration delay. Note that the first term is the same in both models (1) and (2).

Early models like those of McNeil and McNeil and Weiss (see Gazis 1974) had major practical limitations. Other formulae like those of
Miller (1968) and Newell (1965) have also been subject to investigation in the past (see also the comparisons by Ohno 1978). Akçelik’s model (Akçelik 1980) and his extension to non-empty initial queues (Akçelik 2002) are, still in our days, very popular among traffic engineers, particularly for dealing with oversaturated intersections. A good summary of the pletoria of early findings on steady state queuing models can be found in Routhail, Tarko and Li (1996, Sec. 9.3).

Models for the dynamic propagation of queues in time and space, extending the Lighthill-Whitham-Richards (LWR) model to signalized intersections, have also been proposed, in early stages as deterministic models (see e.g. Rorbech 1968, Stephanopoulos and Michalopoulos 1979, Daganzo 1995) and more recently for their stochastic instances, either supported by space-discretization of the LWR model (see Sumalee et al. 2011 for the basic stochastic cell transmission model and Zhong et al. 2013 for its extension to signalized intersections) or by vehicle-discretization of the LWR model (see e.g. Osorio et al. 2011, Osorio and Flotterod 2015). These models intend to capture the evolution of link state distributions through the discretization of differential equations. They have wide application opportunities, since they can cope with uncertain supply functions associated with any arrival distribution, which can be non-stationary. They allow for the study of the transient behaviour of the queue.

Although computationally time consuming and demanding a considerable amount of calibration, car-following models have been, for many years, the focus of a vast investigation (see for instance Pandwai and Dia 2005, Brockfeld and Wagner 2006 for considerations on the use of the car-following microscopic simulators and Sun et al. 2016 and references therein for recent developments on car-following models).

Van Woensel and Vandeaele (2007, Sec. 2.3) include an overview of the different efforts made in the use of queuing models for traffic flow, and Viti (2006) and Cheng et al. (2015) provide historical reviews of delay estimation models. The goal of comparing models for vehicle delay also conducted the investigation of Dion et al. (2004), who covered a large variety of analytical delay models of various kinds in their work. They argue that most of the delay models in the literature produce similar results for signalized intersections with low traffic demand (i.e. $q/c < \ldots$)
0.6) but increasing differences occur as the traffic demand approaches saturation. Last decade investigation has also been directed to finding formulae for the delay variability, like the approximate expressions of Fu and Hellinga (2000) or the queue length probability computations and delay variance obtained in Olszewski (1994).

It is worth to remark that models like the $M/D/1$ or the $M/D^X/1$ do not describe well the regular deactivation of the server, even when the traffic flow that reaches the intersection is assumed to follow a Poisson distribution. Indeed, as the signal controlling the traffic in a street alternates between red and green, modeling a signalized intersection is a problem belonging to the class of queuing systems with server vacations (Doshi 1986), with the particularity that the server remains inactive for a pre-timed duration.

Heidemann (1994) proposed an analytic model that includes server vacations, starting from the assumption that the arrival process is a Poisson process, the intersection has a fixed cycle regulation, the interval between departure of vehicles is constant and the traffic capacity is one. With these assumptionss, probability generating functions for performance measures as queue length and the average delay of a vehicle can be derived from specific associated Markov chains. Later, Hu et al. (1997) considered an $M/D^X/1$ vacation model with pre-timed server vacations for a signalized intersection to address the extension of Heidemann’s model to the multi-lane case, using the $M/G/1$ formalism of Neuts (1989) to derive Laplace-Stieltjes transforms of the queue length and vehicle delay, but present neither numerical results nor analytical formulae for the queue length and vehicle delay distributions. More recently, van Leeuwaarden (2006) extended the computation of the probability generating functions of the queue length and delay of a vehicle, as well as the evaluation of performance characteristics as percentiles, to a class of discrete arrival processes which includes Poisson arrivals. Van den Broek et al. (2006) compared the fixed cycle traffic light queue with the bulk service queue and derived new bounds for the mean overflow and mean delay of vehicles.

Alfa and Neuts (1995) suggested the use of discrete time Markov arrival processes to describe the nature of platoons in the traffic flow. Viti and Zuylen (2010) established a formulation for the length of the
queue along time, within a cycle, which is based on Markov chain renewal theory. They provide formulae for the evolution along time of the expected overflow queue and its standard deviation, as well as the expected queue length and the control delay within a signal cycle. The need for transient measures is justified as they are particularly useful if interventions are planned regarding the traffic control.

In the paper, we are concerned with pre-timed intersections in the context of urban traffic, which means that a plan of server vacation times is fixed a priori, setting times of green and red for each traffic light, and the control is executed according exactly to such plan. More precisely, we consider an $M/D/1$ queueing model with pre-timed server vacations. Clients (vehicles) are served in a FIFO regime and the server starts a vacation as soon as a red time $r$ initiates. If the queue is empty when the server returns from a vacation it will remain idle until a new arrival occurs and serves the arriving vehicle(s) until the green time expires. Once the green time expires a new vacation starts. We will explore the specific nature of the $M/D/1$ queue with pre-timed server vacations considered, and in particular its Markov regenerative structure, in order to characterize the distributions of queue length and vehicle delays, for both transient and long-run regimes, and compare the results obtained for the long-run mean delay of a vehicle with those obtained by applying the HCM model and the microscopic simulation model. We used the simulator presented in Simões et al. 2010, in the particular case of pre-timed signalized intersections for the microscopic simulation. An expression for the variability of the length of the queue is also investigated.

The paper is organized as follows. Assumptions and modelling aspects of the system are presented in Section 2, introducing the traffic state at phase transitions as an embedded chain that is characterized in Section 3. Section 4 explores the long-run properties of the traffic process, and Section 5 characterizes the distribution of the delay (i.e., the sojourn time in system) of a vehicle. An application to a particular intersection is studied in Section 6. The paper ends with some brief conclusions drawn in Section 7.
Assumptions and modelling of the traffic system

A signalized intersection regulated by pre-timed control is assumed to be a traffic server system for which each vehicle arriving at the intersection during a green (light) period has to wait if there are vehicles in front of it, or if arriving during a red (light) period. In a detailed way, we consider a traffic system with the following specifications, with time in seconds:

1. Vehicles arrive at the intersection according to an homogeneous Poisson process with rate \( \lambda \);
2. The system has infinite vehicle waiting capacity;
3. The vehicles are served one by one, in order of arrival to the intersection;
4. The service time of a vehicle is constant and equal to \( T_a \);
5. The service of vehicles only starts or ends at time instants \( 0, T_a, 2T_a, \ldots \);
6. Signal cycles have duration \((M + N)T_a\) and are divided in a server working period of length \( MT_a\), a green period, followed by a server vacation period of length \( NT_a\), a red period;
7. Green periods consist of \( M \) equal sub-periods (phases 1, 2, \ldots, \( M \)) of duration \( T_a \), and red periods consist of a single period (phase \( M + 1 \)) of duration \( NT_a \);
8. When returning from a vacation an initial time of duration \( T_a \) is needed for the first stopped vehicle, if there is one, to move across the traffic light.

We should stress that the approach followed in the paper to investigate the traffic system described could be adapted to incorporate the following relaxations of assumptions 1–3: vehicles arriving at the intersection according to a non-homeogeneous compound-Poisson process; the system having finite vehicle waiting capacity; and group service of vehicles, with a maximum group size, being allowed (as considered in Hu et al. 1997). Thus, the approach followed in the paper relies essentially on assumptions 4–8 and the Poissonian (i.e., memoryless) nature of vehicle arrivals.
From the point of view of the reasonableness of the above assumptions to deal with real-life circumstances, we should stress that the Poisson process is currently used as a model for isolated intersections, although this approximation may be ineffective when traffic networks are concerned. The time discretization, with time step $T_a$, that is implicit in the Markov chain we will use, represents a reasonable approximation of the real world traffic; and the use of a constant service time to represent the time spent by a vehicle driving across the intersection is also a fair approximation of the real world behaviour of drivers.

Note that the factor $T_a$ represents the time that a vehicle spends to move through the intersection, i.e., its service time, and that the traffic system has a server (traffic signal) that is available for periods of duration $MT_a$ (seconds), the green periods, and unavailable for periods of duration $NT_a$ (seconds), the red periods. To simplify the exposition, we assume that we start to observe the system at the beginning of a green period (i.e., the signal changes from red to green at time 0) and that there are no vehicles waiting at the signal at time zero.

For $t \geq 0$, we let $(L(t), \xi(t))$ denote the state of the system at time instant $t$, with $L(t)$ denoting the number of vehicles in the system (in brief, the queue size) at time $t$ and $\xi(t)$ the state of the signal (in brief, the phase) at time $t$. In the following, we call $\{(L(t), \xi(t))\}$ the traffic process (at the signalized intersection).

We note that the phase process $\{\xi(t)\}$ is a deterministic process with state space \{1, 2, \ldots, M + 1\}, where phases 1, 2, \ldots, $M$ correspond to the successive intervals of duration $T_a$ in green light periods, and phase $M + 1$ corresponds to the red light periods of duration $NT_a$. In order to characterize the evolution of the phase process, we let $\tau_n$ denote the instant (of time) of occurrence of the $n$-th change of state in the process, with $\tau_0 = 0$. From the description of the traffic model, it readily follows that, for $n > 0$:

$$\tau_n = \begin{cases} \tau_{n-1} + T_a, & n \mod (M + 1) \neq 0 \\ \tau_{n-1} + NT_a, & n \mod (M + 1) = 0 \end{cases}.$$

As a result,

$$\tau_n = [n \text{ div}(M + 1)](M + N)T_a + [n \mod (M + 1)]T_a,$$
and
\[ \xi(t) = (n - 1) \mod (M + 1) + 1, \quad \tau_{n-1} \leq t < \tau_n. \]

The previous equation expresses the fact that the phase at time \( t \), \( \xi(t) \), is equal to: one plus the number of intervals of duration \( T_a \) elapsed since the latest light switch to green taking place before time \( t \), if the traffic light is green at time \( t \); and \( \xi(t) = M + 1 \) if the traffic light is red at time \( t \). Moreover, the traffic light changes from red to green at times \( \tau_{j(M+1)} = j(M + N)T_a, \ j \in \mathbb{N} \), and from green to red at times \( \tau_{j(M+1)+M} = j(M + N)T_a + MT_a, \ j \in \mathbb{N} \).

A careful analysis of the traffic process \( \{(L(t),\xi(t))\} \) leads to the conclusion that it is a Markov regenerative process with state space \( \mathbb{N} \times \{1,2,\ldots,M+1\} \); see, e.g., Kulkarni (1995) for details on Markov regenerative processes. Moreover, by observing the process \( \{(L(t),\xi(t))\} \) at times \( \tau_n, \ n \in \mathbb{N} \), we obtain the embedded Markov chain \( \{X_n\} \), with \( X_n = (L(\tau_n),\xi(\tau_n)), \ n \in \mathbb{N} \), denoting the state of the system immediately after the \( n \)-th phase change.

As a starting point for the study of the traffic process \( \{(L(t),\xi(t))\} \), we will characterize in the next section the embedded chain \( \{X_n\} \); taking profit of the fact that it is an \( M/G/1 \) type Markov chain, a type of chain that was investigated in detail in Neuts (1989).

3 The embedded chain \( \{X_n\} \)

The Markov chain \( \{X_n\} \) has state space \( \mathbb{N} \times \{1,2,\ldots,M+1\} \) and transition probability matrix

\[
Q = \begin{bmatrix}
B_0 & B_1 & B_2 & \cdots \\
A_0 & A_1 & A_2 & \cdots \\
0 & A_0 & A_1 & \cdots \\
\vdots & \vdots & \vdots & \ddots \\
\end{bmatrix},
\] (3)
where the $A_k$ and $B_k$ are $(M+1) \times (M+1)$ nonnegative matrices whose $(i,j)^{th}$ element is given by

$$
(A_k)_{ij} = \begin{cases} 
    e^{-\lambda T_a} \left(\frac{\lambda T_a}{k!}\right)^k, & i = 1, 2, \ldots, M, j = i + 1 \\
    e^{-\lambda N T_a} \left(\frac{\lambda N T_a}{(k-1)!}\right)^{k-1}, & i = M + 1, j = 1, k \geq 1 \\
    0, & \text{otherwise}
\end{cases},
$$

and $(B_k)_{M+11} = (A_{k+1})_{M+11}$, $(B_k)_{ij} = (A_k)_{ij}$, for $(i,j) \neq (M+1,1)$, in view of the structure of the phase transitions, implying in particular that $(A_k)_{ij}$ is zero unless $j = 1 + i(M+1)$, and the fact that vehicles arrive to the system according to a Poisson process of rate $\lambda$. Note that, for $1 \leq i \leq M$, $(A_k)_{ii+1}$ denotes the probability that $k$ vehicles arrive in a time interval, of duration $T_a$, elapsing from a transition to phase $i$ to the next subsequent phase transition, to phase $i+1$. Conversely, for $k \geq 1$, $(A_k)_{M+11}$ denotes the probability that $k-1$ vehicles arrive in a time interval elapsing from a transition to phase $M+1$, starting a red period, to the subsequent phase transition, to phase 1 and starting a green period, and $(A_0)_{M+11} = 0$.

From the structure of the matrix $Q$ in (3), it follows that the Markov chain $\{X_n\}$ is of $M/G/1$ type. A characteristic of such type of Markov chains is the block Hessenberg form of $Q$ and Toeplitz block structure of the sub-matrix that results from removing the first block of lines of $Q$.

We will next characterize the long run properties of the Markov chain $\{X_n\}$, but we first draw our attention to the phase chain $\{\xi(\tau_n)\}$, which is itself a Markov chain with transition probability matrix

$$
A = \sum_{k=0}^{\infty} A_k = \begin{bmatrix}
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1 \\
1 & 0 & 0 & \cdots & 0
\end{bmatrix}.
$$

As a result, the computation of the stationary probability vector $r = [r_1 \ r_2 \ \ldots \ r_{M+1}]$ of $\{\xi(\tau_n)\}$, solution of $rA = r$ with $r1 = 1$, where 1 denotes a vector of ones (of appropriate dimension, in this case $M+1$), is simple
and leads to the (discrete) uniform distribution on the set \{1, 2, \ldots, M + 1\}, i.e.,
\[ r_i = \frac{1}{M + 1}, \quad i = 1, 2, \ldots, M + 1. \]

Defining \( \beta = \sum_{k=1}^{\infty} kA_k1 \) and \( \rho^* = r\beta \), elementary algebra gives us:
\[
\beta = \begin{bmatrix}
\lambda T_a \\
\vdots \\
\lambda T_a \\
\lambda NT_a + 1
\end{bmatrix}
\text{ and } \quad \rho^* = \frac{\lambda T_a (M + N) + 1}{M + 1}.
\]

According to Neuts (1989, Theorem 2.3.1), the process \( \{X_n\} \) is positive recurrent if and only if \( \rho^* < 1 \) or, equivalently,
\[
\rho = \frac{\lambda T_a (M + N)}{M} < 1, \quad (4)
\]
where the parameter \( \rho \) may be regarded as the traffic intensity of the system under study. Note that, as expected, we conclude that \( \{X_n\} \) is positive recurrent if and only if the expected number of vehicles arriving during a traffic cycle (a period of green traffic light followed by the subsequent period of red traffic light), whose duration is \( T_a(M+N) \), is strictly smaller than the maximum number of vehicles that can be served during a traffic cycle, \( M \).

Under condition (4), the procedure described in this section can be used to numerically obtain the invariant probability vector associated with the stochastic matrix \( Q \) (cf. Neuts 1989), further denoted by \( u \). For that purpose let \( G \) be the minimal nonnegative solution of the nonlinear matrix equation \( G = \sum_{i=0}^{\infty} A_i G^i \). Then, \( G \) is a irreducible stochastic matrix (see Latouche 1994 for details) that can be obtained by applying the following recursive procedure presented in Neuts (1989) and briefly described as:
\[
G = \lim_{n \to \infty} G_n,
\]
where
\[
G_n = \begin{cases}
A_0, & n = 0 \\
\sum_{j=0}^{n} A_j G_{n-j}^j, & n > 0
\end{cases}, \quad (5)
\]
truncated at some appropriate \( n \). After computing the matrix \( G \) one can obtain the invariant probability vector of the irreducible Markov chain \( \{X_n\} \) under study.

For this purpose, let \( u = [u^{(0)} \ u^{(1)} \ u^{(2)} \ldots] \) be an infinite row vector such that \( u^{(k)} = [u_{k1} \ u_{k2} \ldots u_{kM+1}] \), \( k \geq 0 \), is an \((M+1)\)-row vector and \( uQ = u \), \( u1 = 1 \). The element \( u_{ki} \) denotes the stationary probability that at the beginning of a period in a phase there are \( k \) vehicles in the system and the system is in phase \( i \). As such, the stationary probability of the number of vehicles in the system at the beginning of a phase being equal to \( k \) is given by \( u_k = \sum_{i=1}^{M+1} u_{ki} \), for \( k \geq 0 \).

In view of (3), one thus gets that

\[
u^{(k)} = u^{(0)} B_k + \sum_{j=1}^{k+1} u^{(j)} A_{k+1-j}, \quad \text{for } k \geq 1. \tag{6}\]

We note that, instead of using (6), we can resort to a more efficient way to compute the vectors \( u^{(k)}, k \geq 1 \), by applying the following recurrence formula due to Ramaswami (1988):

\[
u^{(k)} = \left[u^{(0)} \bar{B}_k + \sum_{j=1}^{k-1} u^{(j)} \bar{A}_{k+1-j}\right] (I - \bar{A}_1)^{-1}, \quad k \geq 1, \tag{7}\]

where

\[
\bar{A}_n = A_n + \bar{A}_{n+1}G \quad \text{and} \quad \bar{B}_n = B_n + \bar{B}_{n+1}G, \quad n \geq 1. \tag{8}\]

From Schellhaas (1990), one gets a fast way to compute \( u^{(0)} \), since \( u^{(0)} = \frac{\kappa}{d} \), with \( \kappa \) denoting the invariant probability vector associated with \( K = B_0 + \bar{B}_1(I - \bar{A}_1)^{-1}A_0 \) (\( \kappa = \kappa K \), with \( \kappa 1 = 1 \)) and \( d = 1 + \frac{k}{1-\rho^*}\left[\sum_{j=1}^{\infty} jB_j 1 + (\sum_{j=0}^{\infty} B_j - I)(I - A + 1r)^{-1} \beta\right] \). \( d = 1 + \kappa[\sum_{j=1}^{\infty} jB_j 1 + (\sum_{j=0}^{\infty} B_j - I)(I - A + 1r)^{-1} \beta]/(1-\rho^*) \) This is a suggestion of improvement in computation that we introduce to the previous method of Neuts (1989).

Note that if the vehicle waiting capacity was finite, then \( \{X_n\} \) would be a Markov chain with finite state space and, as such, its invariant probability vector could be computed directly by solving a finite system of linear equations. Conversely, if the vehicle arrival process was a non-homogeneous Poisson process instead of an homogenous one, then \( \{X_n\} \) would be a non-homogeneous Markov chain instead of an homogeneous one. Moreover, vehicle group service could be dealt with by
partitioning the transition probability matrix \( Q \) in a block-matrix form as done by Hu et al. (1997).

4 Long-run properties of the traffic process

Having described in the previous section a method to compute the long-run fraction of time the embedded chain \( \{X_n\} \) spends in its states, we proceed in this section to characterize the long-run properties of the Markov regenerative traffic process \( \{(L(t), \xi(t))\} \).

We first note that the quantities \( u_{ki} \), derived in the previous section, correspond to the long run fraction of phase transitions that lead to phase \( i \) with \( k \) vehicles staying in the system immediately after the phase transition. This, in turn, implies that the long run fraction of phase \( i \) intervals that are initiated with \( k \) vehicles in the system, denoted by \( \pi_{ki} \), is such that

\[
\pi_{ki} = \frac{u_{ki}}{\sum_{j=0}^{\infty} u_{ji}} = (M + 1) u_{ki},
\]

in view of the fact that \( \sum_{j=0}^{\infty} u_{ji} = r_i = 1/(M + 1) \).

Of particular relevance are the long run (and stationary) distributions of the number of vehicles in the system at the beginning of green light periods, \( \{\pi_{k1}\}_{k \geq 0} \), and at the beginning of red light periods \( \{\pi_{kM+1}\}_{k \geq 0} \). For later use, we let \( E[L_i^m] \) denote the \( m \)-th stationary moment of the number of vehicles in the system immediately after a transition to phase \( i \), corresponding to the \( m \)-th moment of the distribution \( \{\pi_{ki}\}_{k \geq 0} \), i.e.,

\[
E[L_i^m] = \sum_{k=0}^{\infty} k^m \pi_{ki}.
\]

We proceed with the characterization of the long-run properties of the phase process \( \{\xi(t)\} \). This is a semi-Markov process with embedded Markov chain at phase transition epochs \( \{\xi_n\} \), such that the amount of time the process remains in phase \( i \) in each visit to the phase is the constant

\[
T_i = \begin{cases} 
T_a, & i < M + 1 \\
NT_a, & i = M + 1
\end{cases}
\]
We may then use the theory of semi-Markov processes (see, e.g., El-Taha and Stidham, Jr 1999, Theorem 4.6) to conclude that the long-run fraction of time the traffic process spends in phase $i$,

$$p_{\bullet i} = \lim_{t \to \infty} \left[ \frac{1}{t} \int_0^t 1_{\{\xi(s)=i\}} \, ds \right],$$

is such that $p_{\bullet i} = \frac{r_i T_i}{\sum_{j=1}^{M+1} r_j T_j}$. In view of the fact that $r_j = \frac{1}{M+1}$, for $j = 1, 2, \ldots, M + 1$, this leads to

$$p_{\bullet i} = \begin{cases} 
\frac{1}{M+N}, & i = 1, 2, \ldots, M \\
\frac{N}{M+N}, & i = M + 1
\end{cases} \quad (12)$$

We next address the computation of the long-run distribution of the number of vehicles in the system, for which we let $p_{ki}$ denote the long run fraction of time there are $k$ vehicles in the system with the system being in phase $i$, i.e.,

$$p_{ki} = \lim_{t \to \infty} \left[ \frac{1}{t} \int_0^t 1_{\{L(s)=k, \xi(s)=i\}} \, ds \right],$$

implying, in particular, that $p_{\bullet i} = \sum_{k=0}^{\infty} p_{ki}$, for $i = 1, 2, \ldots, M + 1$. The following theorem expresses how the $\{p_{ki}\}$ may be computed from the $\{\pi_{ki}\}$.

For $k \in \mathbb{N}$ and $i \in \{1, 2, \ldots, M + 1\}$,

$$p_{ki} = \sum_{j=0}^{k} \pi_{ji} \frac{\mu_{k-j}(i)}{(M+N)T_a} \quad (13)$$

with $\mu_l(i)$, $l \in \mathbb{N}$, being given by:

$$\mu_l(i) = \begin{cases} 
\frac{1}{\lambda} \left[ 1 - e^{-\lambda T_a} \sum_{m=0}^{l} \frac{(\lambda T_a)^m}{m!} \right], & i < M + 1 \\
\frac{1}{\lambda} \left[ 1 - e^{-\lambda (NT_a)} \sum_{m=0}^{l} \frac{(\lambda (NT_a))^m}{m!} \right], & i = M + 1
\end{cases} \quad (14)$$

Proof. From the theory of Markov regenerative processes (see, e.g., El-Taha and Stidham, Jr 1999, Theorem 4.7), the definition of $p_{ki}$, and
the structure of the traffic process \( \{L(t), \xi(t)\} \), it follows that

\[
p_{ki} = \frac{\sum_{j=0}^{k} u_{ji} \theta_{ji}(k)}{\sum_{l=1}^{M+1} r_l T_l},
\]

with \( \theta_{ji}(k) \) denoting the expected amount of time there are \( k \) vehicles in the system during an interval of time in phase \( i \) initiated with \( j \) vehicles in the system.

As \( \pi_{ji} = (M + 1) u_{ji} \) and \( \sum_{l=1}^{M+1} r_l T_l = \frac{(M+N)T_a}{M+1} \), in order to prove the theorem it remains to show that the quantities \( \theta_{ji}(k) \) are equal to the quantities \( \mu_{k-j}(i) \) defined in (14). This follows, for \( i \in \{1, 2, \ldots, M+1\} \) and \( 0 \leq j \leq k \), from the following set of equalities:

\[
\theta_{ji}(k) = E \left[ \int_0^{T_j} 1\{L(t)=k|L(0)=j,\xi(0)=i\} \, dt \right] = \int_0^{T_j} P(L(t) = k|L(0) = j, \xi(0) = i) \, dt
\]

\[
= \begin{cases} 
\int_0^{T_a} e^{-\lambda t} \frac{(\lambda t)^{k-j}}{(k-j)!} \, dt, & i \neq M + 1 \\
\int_0^{NT_a} e^{-\lambda t} \frac{(\lambda t)^{k-j}}{(k-j)!} \, dt, & i = M + 1 
\end{cases}
\]

\[
= \begin{cases} 
\frac{1}{\lambda} \left[ 1 - e^{-\lambda T_a} \sum_{m=0}^{k-j} \frac{(\lambda T_a)^m}{m!} \right], & i \neq M + 1 \\
\frac{1}{\lambda} \left[ 1 - e^{-\lambda NT_a} \sum_{m=0}^{k-j} \frac{(\lambda NT_a)^m}{m!} \right], & i = M + 1 
\end{cases}
\]

where the last equality may be obtained using induction on \( k - j \) (cf. Kwiatkowska et al. 2006). □

Let \( p_{ki} \) denote the long run fraction of time there are \( k \) vehicles in the system, i.e., \( p_{ki} = \lim_{t \to \infty} \left[ \frac{1}{t} \int_0^t 1\{L(s)=k\} \, ds \right] \). Then, since \( p_{ki} = \sum_{i=1}^{M+1} p_{ki} \), we have the following corollary of the previous theorem.

For \( k \in \mathbb{N} \),

\[
p_{ki} = \frac{\sum_{i=1}^{M+1} \sum_{j=0}^{k} \pi_{ji} \mu_{k-j}(i)}{(M+N)T_a}, \tag{15}
\]

with \( \mu_{k-j}(i) \) given in (14).
The probability distribution \( \{p_k\} \) may be used to compute performance measures of the long run number of vehicles in the system, such as quantiles and moments, including in particular its mean,

\[
E[L] = \lim_{t \to \infty} \left[ \frac{1}{t} \int_0^t \sum_{k=0}^\infty k \mathbf{1}_{\{L(s)=k\}} \, ds \right],
\]

(16)

second moment,

\[
E[L^2] = \lim_{t \to \infty} \left[ \frac{1}{t} \int_0^t \sum_{k=0}^\infty k^2 \mathbf{1}_{\{L(s)=k\}} \, ds \right],
\]

(17)

and variance, \( \text{Var}(L) = E[L^2] - (E[L])^2 \), simply using the fact that

\[
E[L] = \sum_{k=0}^\infty k \ p_k \quad \text{and} \quad E[L^2] = \sum_{k=0}^\infty k^2 \ p_k.
\]

(18)

However, the theory of Markov regenerative processes provides a more direct way to compute \( E[L] \) and \( \text{Var}(L) \), as shown in the next theorem. The long run mean and variance of the number of vehicles in the system are given, respectively, by

\[
E[L] = \sum_{i=1}^{M+1} E[L_i]T_i
\]

and

\[
\text{Var}(L) = \sum_{i=1}^{M+1} \left[ E[L_i^2]T_i + \lambda E[L_i]T_i^2 \right] \frac{1}{(M+N)T_a} + \frac{\lambda(M+N^2)T_a}{2(M+N)} + \frac{\lambda^2(M+N^3)T_a^2}{3(M+N)} - (E[L])^2.
\]

(19)

(20)

Proof. From the theory of Markov regenerative processes (see, e.g., El-Taha and Stidham, Jr 1999, Theorem 4.7), the relation (16) for \( E[L] \), and the structure of the traffic process \( \{L(t), \xi(t)\} \), it follows that

\[
E[L] = \sum_{i=1}^{M+1} \sum_{k=0}^\infty \sum_{l=1}^{M+1} \frac{u_{ki}}{r_lT_l} \delta_{ki}
\]

(21)
\[ \delta_{ki} = E \left[ \int_0^{T_i} \sum_{l=0}^{\infty} (k + l) 1_{\{L(t) = (k+l) \mid L(0) = k, \xi(0) = i\}} \, dt \right]. \]

Thus, as the previous expression implies that
\[ \delta_{ki} = \int_0^{T_i} \sum_{l=0}^{\infty} (k + l) P(\text{L}(t) = (k + l) \mid \text{L}(0) = k, \xi(0) = i) \, dt \]
\[ = \int_0^{T_i} \sum_{l=0}^{\infty} (k + l) e^{-\lambda t} (\lambda t)^l / l! \, dt = k T_i + \frac{\lambda T_i^2}{2}, \]
\[ \text{taking into account (21), we have} \]
\[ E[L] = \sum_{i=1}^{M+1} \sum_{k=0}^{\infty} k \pi_{ki} T_i \sum_{i=1}^{M+1} T_i^2 \sum_{k=0}^{\infty} \frac{\lambda}{2 (M+N) T_i}, \]
\[ \text{in view of the fact that} \quad \pi_{ki} = (M + 1) u_{ki}, \quad \sum_{i=1}^{M+1} r_i T_i = \frac{(M+N) T_i}{M+1}, \quad \text{and} \]
\[ \sum_{k=0}^{\infty} u_{ki} = r_i = \frac{1}{M+1}. \]
\[ \text{The equation (19) for} \ E[L] \ \text{now follows since} \]
\[ \sum_{k=0}^{\infty} k \pi_{ki} = E[L_i] \quad \text{and} \quad \sum_{i=1}^{M+1} T_i^2 = (M + N^2) T_a^2. \]

If we proceed to compute \( E[L^2] \) in a similar way to what we have just used to compute \( E[L] \), we conclude from the theory of Markov regenerative processes (see, e.g., El-Taha and Stidham, Jr 1999, Theorem 4.7), the relation (17) for \( E[L^2] \), and the structure of the traffic process \( \{L(t), \xi(t)\} \), that
\[ E[L^2] = \sum_{i=1}^{M+1} \sum_{k=0}^{\infty} u_{ki} \delta'_{ki} \sum_{i=1}^{M+1} \sum_{k=0}^{\infty} \pi_{ki} \delta'_{ki} = \frac{k^2 T_i + \frac{1}{2} \lambda T_i^2 + \frac{\lambda^2}{3} T_i^3}{(M+N) T_a}, \]
\[ \text{with} \]
\[ \delta'_{ki} = E \left[ \int_0^{T_i} \sum_{l=0}^{\infty} (k + l)^2 1_{\{L(t) = (k+l) \mid L(0) = k, \xi(0) = i\}} \, dt \right], \]
\[ \text{and the previous expression implies that} \]
\[ \delta'_{ki} = \int_0^{T_i} \sum_{l=0}^{\infty} (k + l)^2 P(\text{L}(t) = (k + l) \mid \text{L}(0) = k, \xi(0) = i) \, dt \]
\[ = \int_0^{T_i} \sum_{l=0}^{\infty} (k^2 + 2 kl + l^2) e^{-\lambda t} (\lambda t)^l / l! \, dt \]
\[ = \int_0^{T_i} \left[ k^2 + (2k + 1) \lambda t + (\lambda t)^2 \right] \, dt \]
\[ = k^2 T_i + (k + \frac{1}{2}) \lambda T_i^2 + \frac{\lambda^2}{3} T_i^3, \]
\[ \text{where the third equality above follows taking into account that a Poisson random variable with parameter} \ \lambda t \ \text{has} \]
mean $\lambda t$ and second moment $\lambda t + (\lambda t)^2$. Thus, in view of (22), we have

$$E[L^2] = \frac{\sum_{i=1}^{M+1} \sum_{k=0}^\infty \pi_{ki} \left[ k^2 T_i + k \left( \frac{1}{2} \right) \lambda T_i^2 + \frac{\lambda^2}{3} T_i^3 \right]}{(M+N)T_a}$$

Thus, in view of (22), we have

$$E[L^2] = \frac{\sum_{i=1}^{M+1} \sum_{k=0}^\infty \pi_{ki} \left[ k^2 T_i + (k + 1) \left( \frac{1}{2} \right) \lambda T_i^2 + \frac{\lambda^2}{3} T_i^3 \right]}{(M+N)T_a}$$

5 Vehicle sojourn time in system

Delay at signalized intersections is a primary concern in the assessment of traffic systems. The mean delay in particular is modeled in HCM2010 as a function of many operational parameters, and control delay is used at operational level to define the level of service. This delay is a complex variable that is sensitive to a variety of local and environmental conditions such as driver’s behavior, traffic characteristics, signal setting, etc.

For the model considered in the paper, the mean long run sojourn time of a vehicle (client) in system, $E[W]$, is simply obtained by applying Little’s formula (cf. for instance Kulkarni 1995),

$$E[W] = \frac{E[L]}{\lambda}, \quad (23)$$

to the expression (19) for $E[L]$. However, we are able to characterize the limit distribution function of the long run sojourn time in system of a vehicle as a function of the distribution of arrival times of vehicles to the intersection with respect to the signal cycle.

One relevant component of the sojourn time in system of a vehicle is the waiting time in red of the same vehicle. We start by noting that if we let $\beta_i(k)$ denote the red time elapsed since the finishing of phase
i until the accumulated green time attains \( kT_a \), then

\[
\beta_i(k) = \begin{cases} 
\left\lceil \frac{k - M + i}{M} \right\rceil NT_a , & i < M + 1 \\
\left\lfloor \frac{k - 1}{M} \right\rfloor NT_a , & i = M + 1
\end{cases}
\] (24)

where for a real number \( y \), \( \lfloor y \rfloor \) and \( \lceil y \rceil \) denote the floor and ceiling of \( y \), respectively. We then have the following result.

The probability that the sojourn time of a vehicle in the system does not exceed \( x \) units of time given that at its arrival to the system the vehicle finds \( l \) vehicles waiting for start of service and its arrival takes place in phase \( i \) at exactly \( s \) units of time before the end of the phase, is given by:

\[
\alpha_{l,i}(x; s) = \begin{cases} 
0 , & x < s + T_a \\
1 \{(l + 1)T_a + \beta_i(l + 1) \leq x - s\} , & x \geq s + T_a
\end{cases}
\] (25)

with \( \beta_i(l + 1) \) given by (24).

For \( x \geq 0 \), let \( F_W(x) \) denote the long-run fraction of vehicles whose sojourn time in system is no larger than \( x \), i.e.,

\[
F_W(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} 1\{W_m \leq x\},
\]

with \( W_m \) denoting the sojourn time in system of the \( m \)-th vehicle entering the system. We then have the following result.

For \( x \geq T_a \),

\[
F_W(x) = \sum_{k=0}^{\infty} \sum_{i=1}^{M+1} \pi_{ki} q_{ki}(x) \frac{1}{(M + N)T_a}
\] (26)

with \( q_{ki}(x) \) being given by:

\[
q_{ki}(x) = \begin{cases} 
\sum_{n=1}^{\infty} \int_{0}^{T_a} e^{-\lambda s} \frac{1}{(n-1)!} \alpha(k-1,+,n-1,i)(x; T_i - s)ds , & 1 \leq i \leq M \\
\sum_{n=1}^{\infty} \int_{0}^{NT_a} e^{-\lambda s} \frac{1}{(n-1)!} \alpha_{k+n-1,M+1}(x; NT_a - s)ds , & i = M + 1
\end{cases}
\] (27)

with \( m^+ = \max(m, 0) \).

Proof. From the theory of Markov regenerative processes (see, e.g., El-Taha and Stidham, Jr 1999, Theorem 4.7), the
definition of $F_W(x)$, and the structure of the traffic process $\{L(t), \xi(t)\}$, it follows that
\[
F_W(x) = \sum_{k=0}^{\infty} \sum_{i=1}^{M+1} u_{ki} q^*_k(x) \sum_{l=1}^{M+1} r_l \lambda T_l,
\]
with $q^*_k(x)$ denoting the expected number of vehicles arriving during a phase $i$ time interval initialized with $k$ vehicles that remain in the system for no more than $x$ units of time.

As $\pi_{ji} = (M + 1)u_{ji}$ and $\sum_{l=1}^{M+1} r_l T_l = \frac{(M+N)T_a}{M+1}$, the theorem follows from the fact that the quantities $q^*_k(x)$ are equal to $\lambda q_k(x)$ with the quantities $q_k(x)$ defined in (27). □

6 Application

We consider an intersection with 4 traffic streams having 2 phases, as illustrated in Figure 1. The following 3 timing plans with red time $r = C - g$ are considered:

Example 1: $g = 20\ s$, $r = 40\ s$.

Example 2: $g = 30\ s$, $r = 30\ s$.

Example 3: $g = 40\ s$, $r = 20\ s$.

The value $T_a = 2\ s$ has been used as it is the most common in a one lane approach. The delays per vehicle that we present correspond only to stream $S_1$.

The mean waiting times (mean delays) estimated by the model presented in the previous sections are shown in Figure 2 together with the results obtained by applying the Webster model (1), the HCM model (2) and the simulation model ran by using the package described in Simões et al. (2010), considering different saturation degrees. The proposed model is named “Markov” in the figures.

The following set up was used in the simulator by Simões et al. (2010):
vehicle’s characteristics: desired speed - Gaussian \((13.9 m/s, 0.2 m/s)\); maximum acceleration - Gaussian \((1.7 m/s^2, 0.3 m/s^2)\); length of a vehicle - Gaussian \((4.0 m, 0.3 m)\); 

number of replications: between 100 and 1000, depending on saturation degree \(x\), controlling for the Monte Carlo error to be smaller than 1;

warm up time: 600s;

run time: 2 hours/replica.

This simulator is based on a microscopic stochastic simulation model that emulates the traffic movements at signalized isolated intersections with a specified type of signal operations. The basic component of the simulator is a car-following model. All parameters were set in the simulator for the particular case of pre–timed control, according to the above examples.

The results suggest that, for small to moderate values of the degree of saturation (approximately \(X < 0.8\)), the estimates of the mean delay of drivers given by the Markov model through expressions (19)–(23) are quite close to the simulation results. For considerably large \(X\) (approximately \(X > 0.8\)), however, the estimates given by the model that we
Figure 2: Comparison between the mean delay estimated by the Markov chain model and other models: (a) Example 1; (b) Example 2; (c) Example 3.
present in this paper remain smaller than those obtained by numerical simulation but much closer to it than the HCM or Webster’s model, except in Example 1. This fact may be explained by the diversity of reactions that is typical of drivers’ behavior and of interactions between vehicles which is mimicked in the simulation model quite closely (cf. Simões et al. 2010) but is hardly taken into account in a Markov or renewal type process modelling.

Note that the number of vehicles in the system at any signal phase can be computed using expression (10). Figure 3 presents the results obtained, in the three examples, at the start of the green signal, at the start of the red signal (that is the overflow queue) and at any time instant, for different saturation degrees. We can see the exponential increase of the queues when the saturation degree approaches 1, as expected, and a reduction of all queues with the increasing of green time. Variances can also be computed resorting to the expression (10).

7 Conclusions

A detailed probabilistic description of the delay of vehicles in pre-timed control of traffic, as well as of the waiting time and the virtual waiting time were achieved by applying the theory of queues with server vacations. This description includes formulae for the computation of the stationary distributions, the means and variances of queue length and waiting time.

When compared to other models existing in the literature the expressions that we give in this paper provide realistic estimates of the mean delay of vehicles particularly when the saturation degree is below 70%. For large traffic flows (congestion scenarios) the estimates provided by the expressions that we propose tend to be smaller than those returned by the numerical simulator and in one case overestimated it, but still give better approximations than those provided by the HCM or Webster’s formulae. We could also derive a formula that enables the analysis of the variability of the queue length.

We are very grateful to the referees for their constructive criticism when reviewing this paper. Research partially supported by CMUP (UID/MAT/00144/2013) and UID/Multi/04621/2013, funded
Figure 3: Mean length estimated by the Markov chain model at the start of the green signal (LSgreen), at the start of the red signal (LSred) and at any time instant (L): (a)Example 1; (b)Example 2; (c)Example 3.
by FCT (Portugal) with National (MEC) and European structural funds through the programs FEDER, under partnership agreement PT2020.

References


