European Journal of Operational Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Stochastics and Statistics

Congestion analysis of unsignalized intersections: The impact of impatience and Markov platooning

Abhishek[a], Marko A. A. Boon[b,*], Michel Mandjes[a], Rudesindo Núñez-Queija[a]

[a] Korteweg-de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands
[b] Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

ARTICLE INFO

Article history:
Received 1 July 2018
Accepted 26 September 2018
Available online xxx

Keywords:
Queuing
Unsignalized intersection
Gap acceptance with impatience
Stochastic capacity analysis
Markov platooning

ABSTRACT

This paper considers an unsignalized intersection used by two traffic streams. The first stream of cars is using a primary road, and has priority over the other stream. Cars belonging to the latter stream cross the primary road if the gaps between two subsequent cars on the primary road are larger than their critical headways. A question that naturally arises relates to the capacity of the secondary road: given the arrival pattern of cars on the primary road, what is the maximum arrival rate of low-priority cars that can be sustained? This paper addresses this issue by considering a compact model that sheds light on the dynamics of the considered unsignalized intersection. The model, which is of a queueing-theoretic nature, reveals interesting insights into the impact of the user behavior on the capacity.

The contributions of this paper are threefold. First, we introduce a new way to analyze the capacity of the minor road. By representing the unsignalized intersection by an appropriately chosen Markovian model, the capacity can be expressed in terms of the solution of an elementary system of linear equations. The setup chosen is so flexible that it allows us to include a new form of bunching on the main road that allows for dependence between successive gaps, which we refer to as *Markov platooning*; this is the second contribution. The tractability of this model facilitates studying the impact that driver impatience and various platoon formations on the main road have on the capacity of the minor road. Finally, in numerical experiments we observe various surprising features of the aforementioned model.

1. Introduction

Unsignalized priority-controlled intersections are very common in urban networks. In the first place there is a high-priority class that consists of cars that use a major (or primary) road. These cars pass the intersection according to some inherently random process. In the second place there is a low-priority stream, consisting of vehicles on the minor (or secondary) road, trying to cross the major road, not affecting the high-priority stream. More specifically, vehicles on the minor road are only allowed to cross the major road when there is a sufficiently large gap between two successive cars on the major road. This *critical headway* T can be either deterministic or a (possibly car-specific) random variable. In practice, a low-priority driver's critical gap typically decreases over time, as he becomes increasingly impatient while scanning for a sufficiently large gap.

Since the high-priority cars on the main road are not hindered by the low-priority cars on the minor road, the capacity of the system is fully determined by the traffic flow on the minor road. We are interested in the capacity of this secondary road, which is defined as the maximum possible number of departures (per time unit) of vehicles from this road. Heidemann and Wegmann (1997) show that this definition implies that the capacity can be expressed in terms of the stability of the corresponding queue: what is the maximum arrival rate of low-priority cars for which it can be guaranteed that the queue (on the minor road) does not become systematically congested? The answer to this question evidently depends on the distribution of the gaps between subsequent cars on the primary road. In particular, the capacity of the minor road is greatly influenced by the clustering of vehicles in platoons on the main road.

In addition to the above, specific features of the low-priority car drivers play a crucial role, in terms of the way that individual car drivers choose their critical headways. We distinguish three mechanisms, which we call (for consistency with Abhishek, Boon, Mandjes, & Núñez Queija, 2016) B_1, B_2, and B_3. The first model assumes that T is constant and the same for all drivers (Guo & Lin, 2011). In

* Corresponding author.
E-mail addresses: abhishek@uva.nl (Abhishek), m.a.a.boon@tue.nl (M.A.A. Boon), m.r.h.mandjes@uva.nl (M. Mandjes), nunezqueija@uva.nl (R. Núñez-Queija).

https://doi.org/10.1016/j.ejor.2018.09.049
0377-2217/

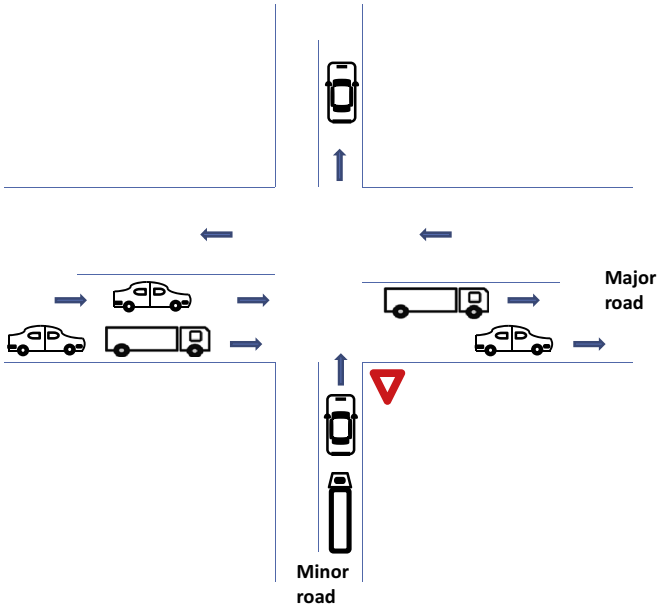


Fig. 1. An example of a situation that can be analyzed using the model in this paper.

the second model T is considered to be a random variable, where the value of T is resampled for any new attempt to cross the main road (Guo, Wang, & Wang, 2014; Wu, 2012). The randomness captures the heterogeneity in the preferences (and driving styles) of the low-priority car drivers. Due to the resampling, this model is often referred to as *inconsistent behavior*: the same driver can have different critical headways in different attempts. The third model assumes that different drivers have different thresholds T, but each driver persistently uses a single driver-specific value of T for all attempts. This is called *consistent behavior*.

The main objective of this paper is to set up and analyze a model that includes all features mentioned above: (a) driver impatience and (b) platooning on the main road, (c) for the behavior types B_1, B_2, and B_3 described above. Notably, previous studies did not succeed in simultaneously incorporating this array of features. The emphasis is on computing the capacity of the minor road.

Various aspects of gap acceptance models have been studied before. The main applications concern unsignalized intersections (e.g. Catchpole & Plank, 1986; Cheng & Allam, 1992; Heidemann & Wegmann, 1997; Tanner, 1962), pedestrian crossings (e.g. Mayne, 1954; Tanner, 1951; Wei, Kumfer, Wu, & Liu, 2016), and freeways (e.g. Drew, Buhr, & Whitson, 1967; Drew, LaMotte, Buhr, & Wattleworth, 1967). Although the queueing aspects in these three application areas might differ slightly, the gap acceptance process exhibits similar features and one common procedure can be used to determine the capacity of the system. In this paper, we focus on the setting of an unsignalized intersection, or T-junction, as depicted in Fig. 1, but all results regarding the capacity of the minor road can be applied to pedestrian crossings or freeway merging, possibly after making small application-specific adjustments as described in the aforementioned papers. Heidemann and Wegmann (1997) give an excellent overview of the existing results in gap acceptance theory, including the three types of user behavior that were discussed above.

As mentioned above, this paper aims at computing the capacity of the minor road in our setting with impatience, platooning, and three different behavior types. In more detail, our contributions are the following.

- We introduce a new method to analyze the capacity of an unsignalized priority-controlled intersection by representing it by a suitably chosen Markovian model. Evaluating the capacity of the minor road thus reduces to solving an elementary system of linear equations, for which we provide a computationally efficient implementation. For any instance considered in this paper, we obtain the capacity nearly instantaneously, thus providing a substantial advantage over performing microsimulations.
- Importantly, the tractability of this model allows us to incorporate simultaneously driver impatience *and* bunched arrivals on the major road. To achieve the latter, we introduce a new model for vehicle clustering, which we will refer to as *Markov platooning* throughout this paper. We study the impact of various platoon formations (Gaur & Mirchandani, 2001; Jia, Lu, Wang, Zhang, & Shen, 2016; Li, 2017) on the main road on the capacity of the minor road. We do so for the three different behavior types introduced above.
- Through numerical examples we present a sequence of surprising insights regarding the capacity of the minor road. In Abhishek et al. (2016) it was observed that for the model without impatience and platooning the capacities that correspond with the three different types of driver behavior that we introduced above, are strictly ordered: B_2 has the largest capacity, then B_1, and the capacity of B_3 is the smallest (with the mean critical headway of models B_2 and B_3 chosen equal to the deterministic critical headway of model B_1). In the present paper we empirically observe that in the setting platooning the ordering still holds, whereas the ordering is lost when impatience is added to the model.
  Another important insight is that one needs to be extremely careful when determining the capacity of an unsignalized intersection when the traffic flow on the major road switches between multiple regimes. More specifically, for this situation with platoons we show that one should build *one* model with various background states (corresponding to the different regimes) to determine the overall capacity, rather than computing the capacities for the different regimes separately and then taking an average.

Platoon forming has also been studied in the existing literature on gap acceptance models before. The most common models that include clustering on the major road are so-called gap-block models. In these models, vehicles tend to form platoons, most commonly arriving according to Poisson processes. The lengths of these platoons are i.i.d. random variables with general distributions, which can be chosen carefully to mimic real-life clustering behavior. Tanner (1962) considers a model where platoon lengths are distributed as the busy period of a single-server queue. Wegmann (1991) and Wu (2001) analyze the capacity under even less restrictive assumptions. However, all of these models assume no (or a very weak form of) dependence between successive block sizes and gap sizes. By introducing Markov platooning, an arrival process based on Markov modulation, we allow for a more refined way of bunching on the major road that includes dependence between successive gap sizes.

As mentioned above, our model also incorporates impatience of the drivers that are waiting to cross the major road. Such behavior was widely encountered in practice; see e.g. Abou-Henaidy, Teply, and Hund (1994). The impact of impatience has been studied before in e.g. Drew, Buhr et al. (1967), Drew, LaMotte et al. (1967), and Weiss and Maradudin (1962), but (to the best of our knowledge) not in a context with platooning and randomness in the critical headway T.

This paper is structured as follows. In the next section, we describe in more detail the variations of the gap acceptance model, including the aforementioned types of gap acceptance behavior, impatience, and platooning on the major road. In Sections 3 and 4, we study the impact of Markov platooning and impatience on

Please cite this article as: Abhishek et al., Congestion analysis of unsignalized intersections: The impact of impatience and Markov platooning, European Journal of Operational Research (2018), https://doi.org/10.1016/j.ejor.2018.09.049

the capacity of the minor road, respectively. In these sections we also present numerical results, exhibiting interesting features of the model variations. Section 5 concludes the paper.

2. Preliminaries

2.1. Arrival process

The situation analyzed in this paper is depicted in Fig. 1. We consider an intersection used by two traffic streams, both of which wishing to cross the intersection. There are two priorities: the cars on the major road have priority over cars on the minor road (and hence do not notice the presence of the minor road). The low-priority cars on the minor road cross the intersection as soon as the gap between two subsequent high-priority cars has a duration larger than T, commonly referred to as the *critical headway*.

Cars on the minor road arrive according to a Poisson process with rate λ. In this paper we distinguish between two types of arrival processes on the major road. The arrival process for the high-priority car drivers is a generalization of the Poisson process, viz. the Markov modulated Poisson process (MMPP). The MMPP, which will be discussed in greater detail in Section 3, is a well-studied object in applied probability which is generally used to model dependencies between inter-arrival times. In an MMPP, at time t the time till the next arrival is exponentially distributed with mean $1/q_i$ if an independently evolving Markov process (usually referred to as the *background process*) is in state i at time t. The flexibility of the MMPP allows us to vary the inter-arrival times in such a way, that we can create platoons, single arrivals, or combinations thereof.

2.2. Gap acceptance behavior

We have not yet exactly defined the criterion by which the low-priority cars decide to cross. In this paper, we distinguish three types of 'behavior' when making this decision.

- B_1: The first model is the most simplistic: the critical headway T is deterministic, and uniform across all low-priority car drivers.
- B_2: Clearly B_1 lacks realism, in that there will be a substantial level of heterogeneity in terms of driving behavior: one could expect a broad range of 'preferences', ranging from very defensive to very reckless drivers. In B_2 this is modeled by the car driver at the front end of the queue *resampling* T (from a given distribution) at any new attempt (where an 'attempt' amounts to comparing this sampled T to the gap between the two subsequent cars that he is currently observing).
- B_3: In the third model an alternative type of driver behavior is assumed. More specifically, it reflects *persistent* differences between drivers, in that each driver selects a random value of T, but then sticks to that same value for all attempts, rather than resampling these.

Note that inconsistent and consistent behavior as defined in B_2 and B_3, respectively, is a well-studied feature in queueing systems where low-priority jobs can be interrupted by high-priority jobs (cf. Takagi, 1991). When service is resumed, depending on the model, the remaining processing time might be a new (identical) full service time, a new random service time, or just the remaining part of the original service time. In gap acceptance models the last type of model does not make sense, which is the reason why we restrict ourselves to the first two variants.

2.3. Impatience

For each of the aforementioned behavior types, we also consider a variant that includes impatience. With impatience, the critical headway decreases after each failed attempt, reflecting the impatience of drivers, resulting in the willingness to accept smaller and smaller gaps. In more detail, we define a critical headway T_m for the mth attempt to enter the main road ($m = 1, 2, \ldots$). Note that, depending on the distributions of $T_1, T_2, \ldots$, in model B_2 situations might occur where $T_{m+1} > T_m$, despite T_{m+1} being stochastically smaller than T_m. This is a typical feature of the model with resampling. Exact details regarding the manner in which impatience is incorporated will be given in Section 4.

A few remarks are in place here. In the first place, above we positioned this setup in the context of an unsignalized intersection, but various other applications could be envisioned. One of these could correspond to the situation in which the low-priority cars have to merge with the stream of high-priority cars (e.g. from a ramp or a roundabout). Also in the context of pedestrians crossing a road, the model can be used. We also stress that in the case the primary road actually consists of two lanes that have to be crossed (without a central reservation), with cars arriving (potentially in opposite directions) at Poisson rates (say) $q^{\leftarrow}$ and $q^{\rightarrow}$, our model applies as well, as an immediate consequence of the fact that the superposition of two Poisson processes is once again a Poisson process with the parameter $q := q^{\leftarrow} + q^{\rightarrow}$; see also the discussion in Wu (2001, Section 5).

3. Markov platooning on the major road

In this section we introduce a new method to compute the capacity of the minor road, based on the principle of departure cycles (cf. Heidemann & Wegmann, 1997, Section 5); we here consider the impact of platooning, whereas in the next section impatience is included as well. The main idea is to cast the traffic situation into a Markovian model. After introducing a novel way of platoon forming, we analyze the three types of driver behavior B_1–B_3 introduced in the previous section. Since many results for the variants without impatience have been known in the existing literature (see, for example, Heidemann & Wegmann, 1997 for an overview), we will mainly focus on the additional insights that can be obtained for the capacity of the minor road under different circumstances, which turns out to lead to a few interesting new insights depicted in the numerical examples at the end of this section.

The assumption of Poisson arrivals on the major road is realistic in periods of free traffic flow, where any individual vehicle does not affect vehicles behind it. In more congested situations, however, vehicles form platoons. As described in the introduction, several papers have looked into the effect of platooning. Heidemann and Wegmann (1997) propose a general framework based on gap-block models, relying on results by Tanner (1962). In such models, vehicles form platoons which arrive according to a Poisson process, where the lengths of these platoons are i.i.d. random variables with a general distribution, which can be suitably chosen such that it matches real-life clustering behavior. Wu (2001) observed that, in practice, the traffic flow in the major stream is in one of four different regimes: free space (no vehicles), free flow (single vehicles), bunched traffic (platoons of vehicles), and queueing. By conditioning on the current regime, he applies the framework of Heidemann and Wegmann (1997) to set up a heuristic argument that provides a general capacity formula that is valid under all four regimes; we return to this approach below.

In this paper, we assume that the arrival process on the major road can be modeled by a *Markov modulated Poisson process* (MMPP). In an MMPP arrivals are generated at a Poisson rate q_i when an exogenous, autonomously evolving continuous-time

Please cite this article as: Abhishek et al., Congestion analysis of unsignalized intersections: The impact of impatience and Markov platooning, European Journal of Operational Research (2018), https://doi.org/10.1016/j.ejor.2018.09.049

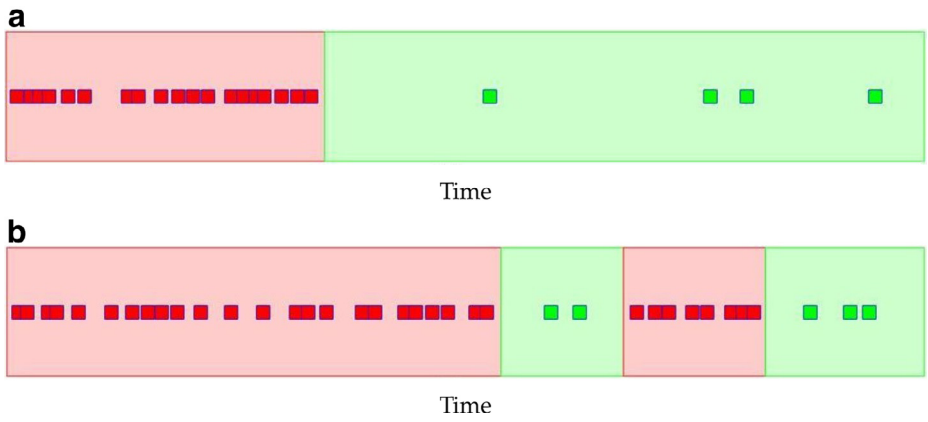


Fig. 2. Simulated examples of two MMPP's with two background states. On the horizontal axis we depict the time, while the squares (red or green) mark the arrivals. In (a), we have chosen $\mu_1 = 1/20, \mu_2 = 1/40$ and arrival rates $q_1 = 1, q_2 = 1/15$ vehicles per time unit. In (b) we use $\mu_1 = 1/20, \mu_2 = 1/20$ and arrival rates $q_1 = 1, q_2 = 1/5$ vehicles per time unit. The red areas indicate that the background process is in state 1 (more platooning) and the green areas correspond to state 2 (less platooning). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Markov process (commonly referred to as the *background process*) is in state i. We denote by $d \in \{1, 2, \ldots\}$ the number of states of the background process (where $d = 1$ corresponds to a non-modulated, ordinary Poisson process). We assume the background process to be irreducible; the corresponding stationary distribution is given by the vector π. In the sequel we denote by $\mathcal{M} = (\mu_{ij})_{i,j=1}^{d}$ the transition rate matrix of background process, and define $\mu_i := -\mu_{ii}$. Therefore, an MMPP allows different traffic-flow regimes on the major road. For example, in Fig. 2, we show the arrival patterns of two MMPP's, each with two background states. The red squares mark arrivals during the high traffic intensity (q_1), while the green squares mark arrivals during the low intensity (q_2). It can be seen that platoons are generally longer when the background process is in state 1, corresponding to a high arrival rate. Additionally, we observe in Fig. 2(a) that the background process stays longer in state 2 ($\mu_2 = 1/40$) than in state 1 ($\mu_1 = 1/20$). Another difference between the two sub-figures is that we choose $q_2 = 1/15$ in Fig. 2(a) and $q_2 = 1/5$ in Fig. 2(b). This explains why, in state 2, we see no platooning at all in Fig. 2(a), but Fig. 2(b) still shows some mild platoon forming.

The main objective of this section is to develop methods that determine the capacity of the minor road under MMPP arrivals on the major road, for the models B_1 up to B_3; here 'capacity' is defined as the maximum arrival rate λ such that the corresponding queue does not grow beyond any bound. Because of this focus on the capacity, we can simplify the model by taking away the queueing aspect on the minor road, assuming that this road is *saturated*: there are *always* low-priority cars waiting for gaps. The reason underlying this reduction is that capacity is a quantity that corresponds to stability of the associated queue, and stability essentially amounts to the queue being able to process all input in the long run.

The capacity, to be denoted by $\bar{\lambda}$, is the ratio of the mean number of arrived cars in a cycle (which we define below) to the mean duration of a cycle, which equals (due to renewal theory) the number of cars that can be served per unit time. The system is stable when λ, the arrival intensity on the minor road, is less than $\bar{\lambda}$. Again, we distinguish between the three behavior types B_1–B_3 introduced in Section 2, each with its own capacity $\bar{\lambda}_i$, for $i = 1, 2, 3$.

Our objective is to assess the impact of the three types of the driver's behavior on stability of the underlying queueing model. The capacity can be interpreted as the reciprocal of the time it takes for an arbitrary car to cross the major road (the 'service time'). At first sight, the following procedure seems to provide us with $\bar{\lambda}$. Define S_i as the time it takes for an arbitrary car to cross the major road, given the background process is in state i when the car (which has reached the head of the queue) starts his attempt. Recalling that π_i represents the long-run fraction of time that the background process resides in state i, it is tempting to conclude that the capacity would equal

$$\sum_{i=1}^{d} \frac{\pi_i}{\mathbb{E}[S_i]}. \tag{3.1}$$

Alternatively, one might try to first take a weighted average of the mean service times, and then take the reciprocal capacity:

$$\frac{1}{\sum_{i=1}^{d} \pi_i \mathbb{E}[S_i]}. \tag{3.2}$$

There is, however, a conceptual mistake in these (naïve) approaches. The crucial point is that there is a difference between the invariant distribution π of the background process (that underlies the platooning mechanism) and the distribution of the background process seen by a car that has reached the head of the queue (which we denote by $\pi^{(q)}$). Put differently, the distribution of the background process when a car *arrives at the intersection* (which is π due to the PASTA property) does not coincide with the distribution of the background process when a car *reaches the head of the queue*. To see this, think of a situation in which the arrival rates (i.e., the rates q_1 and q_2 that pertain to the two states of the background process) are chosen very differently. More specifically, q_1 has some moderate value q, whereas q_2 has some huge value Q (meaning that when the background process is in state 2, the cars at the highway pass by at a high frequency). It implies that minor road cars that find the queue non-empty are, with high probability, faced with the background process being in state 1 when reaching the head of the queue (as the minor road cars in front of it can only leave the queue when the background state is 1); only cars that find the queue empty have a chance that the background process is in state 2 when reaching the head of the queue. We thus observe that there is an evident difference between π and $\pi^{(q)}$.

This reasoning illustrates how careful one should be when weighing capacities that belong to different regimes by the fractions of time in which those regimes apply. A very similar decomposition approach was followed by Wu (2001); he distinguishes four different regimes, as described above, each with an own capacity, which are then combined into a single capacity. The formulas obtained by Wu (2001) likely provide a reasonable indication of the capacity across a wide range of parameters, but there are also many cases in which the approach fails to do so. Later on, we

Please cite this article as: Abhishek et al., Congestion analysis of unsignalized intersections: The impact of impatience and Markov platooning, European Journal of Operational Research (2018), https://doi.org/10.1016/j.ejor.2018.09.049

provide an example which illustrates that following such naïve approaches may lead to substantial errors.

B_1 (constant gap): In this model, every driver on the minor road needs the same constant critical headway T to enter the major road. In our analysis we use the renewal reward theorem, which entails that the capacity can be written as the mean number of cars arriving in a regenerative cycle divided by the mean duration of that cycle, see also Heidemann & Wegmann, 1997, Section 5.1. For our purposes, an appropriate definition of a cycle is: the time elapsed between two consecutive epochs such that (i) the background process is in a reference state (say state 1, but the choice of the reference state is arbitrary), and (ii) a service is completed (i.e., a low-priority car is served).

To make our model Markovian, we approximate this deterministic T by an Erlang random variable with k phases of average length T/k. It is well known that a deterministic T can be approximated by the sum of k independent exponential random variables, each with parameter $\kappa := k/T$, with k large; to see this, observe that this Erlang random variable has mean T (as desired), and variance $k/\kappa^2 = T^2/k$, which goes to 0 as k grows large. Define h_{ij} as the mean number of cars that is served till the cycle ends, given that the current state of the background process is $i \in \{1, \ldots, d\}$ and the car in service has finished $j \in \{0, \ldots, k-1\}$ phases of the Erlang distribution. To find the mean number of arrived cars in a cycle, we need to find h_{10}. This can be done as follows. Exploiting the memorylessness of the exponential distribution, we distinguish between all the possible jumps that can take place. Jumps may occur due to the background process changing state (with rate $\mu_{i\ell}$ from state i to ℓ), an arrival on the major road (with rate q_i), or the completion of a phase of the Erlang critical gap distribution (with rate κ). In the sequel we write $\varrho_i := \mu_i + q_i + \kappa$. Relying on 'standard Markovian reasoning', by conditioning on the first jump,

$$h_{ij} = \sum_{\ell \neq i} \frac{\mu_{i\ell}}{\varrho_i} h_{\ell j} + \frac{q_i}{\varrho_i} h_{i0} + \frac{\kappa}{\varrho_i} \cdot \begin{cases} h_{i,j+1} & \text{if } j < k-1, \\ 1 & \text{if } j = k-1, i = 1, \\ 1 + h_{i0} & \text{if } j = k-1, i > 1. \end{cases} \tag{3.3}$$

This can be written as a linear system of dk equations with dk unknowns of the form $A\vec{h} = \vec{b}$, where entries of the matrix $A = [a_{mn}]$, $\vec{h}$ and $\vec{b} = [b_m]$ are given as follows, with $i = \lceil m/k \rceil$,

$$a_{mn} = \begin{cases} -\frac{\kappa}{\varrho_i}, & \text{if } n = m+1 \text{ and } m \neq k, 2k, \ldots, dk; \\ -\frac{q_i}{\varrho_i}, & \text{if } n=(i-1)k+1 \text{ and } m \neq 1, k+1, 2k, 2k+1, 3k, \ldots, (d-1)k+1, dk; \\ -\frac{\kappa + q_i}{\varrho_i}, & \text{if } n = (i-1)k+1 \text{ and } m = 2k, \ldots, dk; \\ -\frac{\mu_{i,\ell+1}}{\varrho_i}, & \text{if } n = (\ell - i + 1)k + m \text{ and } \ell \in \{0, 1, \ldots, d-1\} \setminus \{i-1\}; \\ 1 - \frac{q_i}{\varrho_i}, & \text{if } n = m \text{ and } m = 1, k+1, \ldots, (d-1)k+1; \\ 1, & \text{if } n = m \text{ and } m \neq 1, k+1, \ldots, (d-1)k+1; \\ 0, & \text{else,} \end{cases}$$

$\vec{h} = [h_{10}, h_{11}, \ldots, h_{1,k-1}, h_{20}, h_{21}, \ldots, h_{2,k-1}, \ldots, h_{d0}, h_{d1}, \ldots, h_{d,k-1}]^{\mathrm{T}}$, and

$$b_m = \begin{cases} \frac{\kappa}{\varrho_\ell}, & \text{if } m = \ell k, \ell = 1, 2, \ldots, d \\ 0, & \text{else.} \end{cases}$$

It is noted that $|a_{mm}| = \sum_{n \neq m} |a_{mn}|$ for all $m \neq k$ and for $m = k$, $|a_{kk}| > \Sigma_{n \neq k} |a_{kn}|$. Therefore, the matrix A is weakly diagonally dominant with one row being strictly dominant. Moreover, A is also irreducible and, hence, invertible (Horn & Johnson, 1986). Therefore, the solution of the system of equations $A\vec{h} = \vec{b}$ is $\vec{h} = A^{-1}\vec{b}$. We thus find the desired quantity h_{10}.

To determine the capacity we need, in addition to the mean number of arrived cars in a cycle, also the mean duration of a cycle. To this end we define τ_{ij} as the mean time till the end of the current cycle, given that the current state of the background process is $i \in \{1, \ldots, d\}$ and the car in service has finished $j \in \{0, \ldots, k-1\}$ phases of the Erlang distribution. The objective is now to find the mean duration of a cycle, which is given by τ_{10}.

Similarly to the procedure we set up above,

$$\tau_{ij} = \frac{1}{\varrho_i} + \sum_{\ell \neq i} \frac{\mu_{i\ell}}{\varrho_i} \tau_{\ell j} + \frac{q_i}{\varrho_i} \tau_{i0} + \frac{\kappa}{\varrho_i} \cdot \begin{cases} \tau_{i,j+1} & \text{if } j < k-1, \\ 0 & \text{if } j = k-1, i = 1, . \\ \tau_{i0} & \text{if } j = k-1, i > 1. \end{cases} \tag{3.4}$$

Also this system can be written as dk linear equations with dk unknowns. More precisely, with $\vec{\tau} = [\tau_{10}, \tau_{11}, \ldots, \tau_{1,k-1}, \tau_{20}, \tau_{21}, \ldots, \tau_{2,k-1}, \ldots, \tau_{d0}, \tau_{d1}, \ldots, \tau_{d,k-1}]^{\mathrm{T}}$, we have $A\vec{\tau} = \vec{c}$ with A as defined before and

$$c_m = \frac{1}{\varrho_\ell} \quad \text{for } (\ell-1)k + 1 \leq m \leq \ell k, \text{ and } \ell = 1, 2, \ldots, d.$$

We already proved that A is invertible, and therefore the unique solution of the system of equations $A\vec{\tau} = \vec{c}$ is $\vec{\tau} = A^{-1}\vec{c}$. We thus find τ_{10}.

The capacity of this system can now be evaluated as $\bar{\lambda}_1 := h_{10}/\tau_{10}$, meaning that the stability condition of the low-priority queue is $\lambda < \bar{\lambda}_1$. In the numerical procedure, the value of k should be chosen large, to ensure that the Erlang distribution is sufficiently 'close-to-deterministic'.

B_2 (sampling per attempt): As pointed out before, in this behavior type every driver samples a 'fresh' random T for every attempt to enter the major road. Let us assume that the gap size T equals some deterministic T_n with probability p_n , for $n \in \{1, 2, \ldots, N\}$. Analogously to what we did in the procedure to evaluate the capacity for B_1, we approximate T_n by an Erlang random variable with k_n phases; each of the phases is exponentially distributed with parameter $\kappa_n = k_n/T_n$. Let $K := \sum_{n=1}^{N} k_n$.

We write $\varrho_i^{(n)} := \mu_i + q_i + \kappa_n$. Let $h_{ij}^{(n)}$ be the mean number of cars that is served till the cycle ends, given that the current state of the background process is $i \in \{1, \ldots, d\}$, the car in service has gap size T_n and the car in service has finished $j \in \{0, \ldots, k_n - 1\}$ phases. We wish to find h_{10} where

$$h_{i0} = \sum_{n=1}^{N} p_n h_{i0}^{(n)}, \qquad \text{for } i = 1, 2, \ldots, d. \tag{3.5}$$

Then

$$h_{ij}^{(n)} = \sum_{\ell \neq i} \frac{\mu_{i\ell}}{\varrho_i^{(n)}} h_{\ell j}^{(n)} + \frac{q_i}{\varrho_i^{(n)}} h_{i0} + \frac{\kappa_n}{\varrho_i^{(n)}} \cdot \begin{cases} h_{i,j+1}^{(n)} & \text{if } j < k_n - 1, \\ 1 & \text{if } j = k_n - 1, i = 1, \\ 1 + h_{i0} & \text{if } j = k_n - 1, i > 1. \end{cases} \tag{3.6}$$

Observe how the resampling is incorporated in this system: when an attempt has failed a 'fresh' new gap size is sampled, explaining the h_{i0} (rather than $h_{i0}^{(n)}$) in the right hand side. The last occurrence of h_{i0} in (3.6), when $i > 1$, $j = k - 1$, corresponds with the event that an attempt has succeeded, after which a new gap size is sampled.

The above equations can be written as a linear system of the type $A\vec{h} = \vec{b}$ for a matrix A and vector $\vec{b}$ (which evidently differ from the matrix A and vector $\vec{b}$ that were used in the model B_1) consisting of dK equations with dK unknowns. With the same argument as we have used for B_1, it follows that the coefficient matrix A is invertible. Using (3.5), this facilitates the computation of $\vec{h}$ and in particular the desired quantity h_{10} (from $h_{10} = p_1 h_{10}^{(1)} + p_2 h_{10}^{(2)} + \cdots + p_N h_{10}^{(N)}$).

We then define $\tau_{ij}^{(n)}$ as the mean time till the current cycle ends, given that the current state of the background process is $i \in \{1, \ldots, d\}$, the car in service has gap size T_n, and has finished

$j \in \{0, \ldots, k_n - 1\}$ phases. The objective is to set up a numerical procedure to evaluate τ_{10} where $\tau_{i0} := \sum_{n=1}^{N} p_n \tau_{i0}^{(n)}$ for $i = 1, \ldots, d$. Using the same argumentation as above,

$$\tau_{ij}^{(n)} = \frac{1}{\varrho_i^{(n)}} + \sum_{\ell \neq i} \frac{\mu_{i\ell}}{\varrho_i^{(n)}} \tau_{\ell j}^{(n)} + \frac{q_i}{\varrho_i^{(n)}} \tau_{i0} + \frac{\kappa_n}{\varrho_i^{(n)}} \cdot \begin{cases} \tau_{i,j+1}^{(n)} & \text{if } j < k_n - 1, \\ 0 & \text{if } j = k_n - 1, i = 1, \\ \tau_{i0} & \text{if } j = k_n - 1, i > 1. \end{cases} \tag{3.7}$$

Again, this system can be written as a linear system of dK equations with dK unknowns, say $A\vec{\tau} = \vec{c}$, with A as above (and hence invertible). Therefore, the solution of the system of equations $A\vec{\tau} = \vec{c}$ is $\vec{\tau} = A^{-1}\vec{c}$, and we can compute τ_{10}. The capacity of the low-priority queue under B_2 is therefore $\bar{\lambda}_2 = h_{10}/\tau_{10}$.

B_3 (sampling per driver): We finally consider the model with consistent behavior, i.e., each driver sticks to the gap size he or she initially sampled. The procedure is similar to the ones we developed for B_1 and B_2, and therefore we restrict ourselves to the main steps.

Define, as before, $h_{i0} = \sum_{n=1}^{N} p_n h_{i0}^{(n)}$ for $i = 1, 2, \ldots, d$. The mean number of cars served during the cycle follows from

$$h_{ij}^{(n)} = \sum_{\ell \neq i} \frac{\mu_{i\ell}}{\varrho_i^{(n)}} h_{\ell j}^{(n)} + \frac{q_i}{\varrho_i^{(n)}} h_{i0}^{(n)} + \frac{\kappa_n}{\varrho_i^{(n)}} \cdot \begin{cases} h_{i,j+1}^{(n)} & \text{if } j < k_n - 1, \\ 1 & \text{if } j = k_n - 1, i = 1, \\ 1 + h_{i0} & \text{if } j = k_n - 1, i > 1; \end{cases}$$

it is instructive to compare this equation with the corresponding one for B_2: when the attempt has failed the gap size is *not* resampled. Resampling is only done when an attempt has been successfully completed.

Similarly, the system of equations for the mean cycle length is

$$\tau_{ij}^{(n)} = \frac{1}{\varrho_i^{(n)}} + \sum_{\ell \neq i} \frac{\mu_{i\ell}}{\varrho_i^{(n)}} \tau_{\ell j}^{(n)} + \frac{q_i}{\varrho_i^{(n)}} \tau_{i0}^{(n)} + \frac{\kappa_n}{\varrho_i^{(n)}} \cdot \begin{cases} \tau_{i,j+1}^{(n)} & \text{if } j < k_n - 1, \\ 0 & \text{if } j = k_n - 1, i = 1, \\ \tau_{i0} & \text{if } j = k_n - 1, i > 1; \end{cases}$$

with $\tau_{i0} := \sum_{n=1}^{N} p_n \tau_{i0}^{(n)}$ for $i = 1, \ldots, d$. The linear system can be solved as before, yielding h_{10} and τ_{10}. Therefore, the capacity of the system can be evaluated as $\bar{\lambda}_3 = h_{10}/\tau_{10}$.

Example 1 (Convergence to deterministic critical caps)**.** In our model, we approximate deterministic critical gaps by an Erlang distribution with a number of phases that we increase until the variance becomes practically zero. In this example we suggest how many phases to take. To this end, we try to reproduce the results from Example 1 in Abhishek et al. (2016), which studies the impact of driver behavior on the capacity of the system and on the queue lengths. We introduce the following three scenarios, with the parameters chosen such that the system exhibits interesting features:

(1) All drivers search for a gap between consecutive cars on the major road, that is at least 7 seconds long.
(2) A driver on the minor street, waiting for a suitable gap on the major street, will sample a new (random) critical headway every time a car passes on the major street. With probability 9/10 this critical headway is 6.22 seconds, and with probability 1/10 it is exactly 14 seconds. Note that the expected critical headway is $0.9 \times 6.22 + 0.1 \times 14 = 7$ seconds, ensuring a fair comparison between this scenario and the previous scenario.
(3) In this scenario we distinguish between slow and fast traffic. We assume that 90% of all drivers on the minor road need a gap of (at least) 6.22 seconds. The other 10% need at least 14 seconds. Again, the mean critical gap is 7 seconds.

Note that these three scenarios correspond to, respectively, B_1, B_2, and B_3. In this simple example, where $d = 1$, closed-form expressions are available (cf. Abhishek et al., 2016):

$$\bar{\lambda}_1 := \frac{q}{e^{qT} - 1}, \quad \bar{\lambda}_2 = \frac{q}{(\mathbb{E}[e^{-qT}])^{-1} - 1}, \quad \bar{\lambda}_3 = \frac{q}{\mathbb{E}[e^{qT}] - 1}. \tag{3.8}$$

Fig. 3 depicts the capacity (vehicles per hour) of the minor street as a function of q, the flow rate on the main road (vehicles per hour). As a sanity check, we vary q between 0 and 3600 so we can validate our model. When $q \downarrow 0$ the limiting capacity is $1/T$ (because every arriving vehicle can immediately cross the major road), while the capacity drops to zero when q becomes large. We stress that we included large values of q for reasons of validation; values in the top end of the range [0,3600] are evidently not realistic.

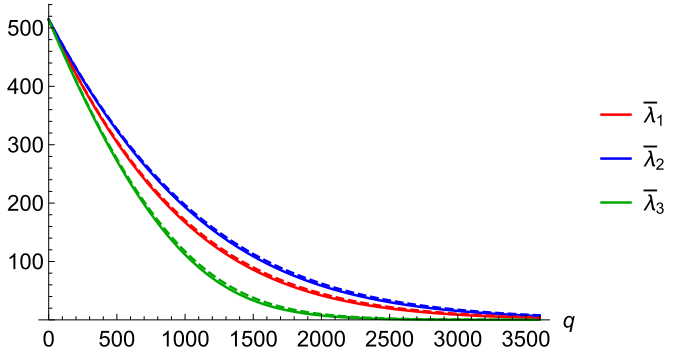


Fig. 3. The capacities (vehicles per hour) in Example 1. The solid lines correspond to the exact expressions, the dashed lines to the approximations with $k = 100$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The solid lines are obtained using the exact results given by (3.8); the dashed lines are the approximations based on an Erlang distribution with $k = 100$ phases. It can be seen that for $k = 100$ the capacities determined by our approximations are very close to the exact values. There is still a small difference visible, which is probably acceptable for all practical purposes. If a higher accuracy is desired, it is straightforward to increase k which results in more accurate approximations. Every data point can be computed nearly instantaneously; this remains the case when performing similar calculations for more complex variants of the model considered in this example. This is obviously a great advantage of our analysis compared to microsimulations.

Finally, note that the relation $\bar{\lambda}_2 \geqslant \bar{\lambda}_1 \geqslant \bar{\lambda}_3$ is clearly visible in Fig. 3. In Abhishek et al. (2016) a rigorous proof is given for this ordering, for models with Poisson arrivals and no impatience. The interpretation is that *consistent* random gap acceptance behavior always decreases the capacity compared to non-random behavior, while the *inconsistent* random model leads to an increase in capacity. Unfortunately, it is quite unlikely that this type of behavior is frequently encountered in practice. In Abhishek et al. (2016, Example 2) it was already observed that this ordering of capacities is not preserved in situations with driver impatience. In the next example we will assess the impact of platooning on the validity of this ordering.

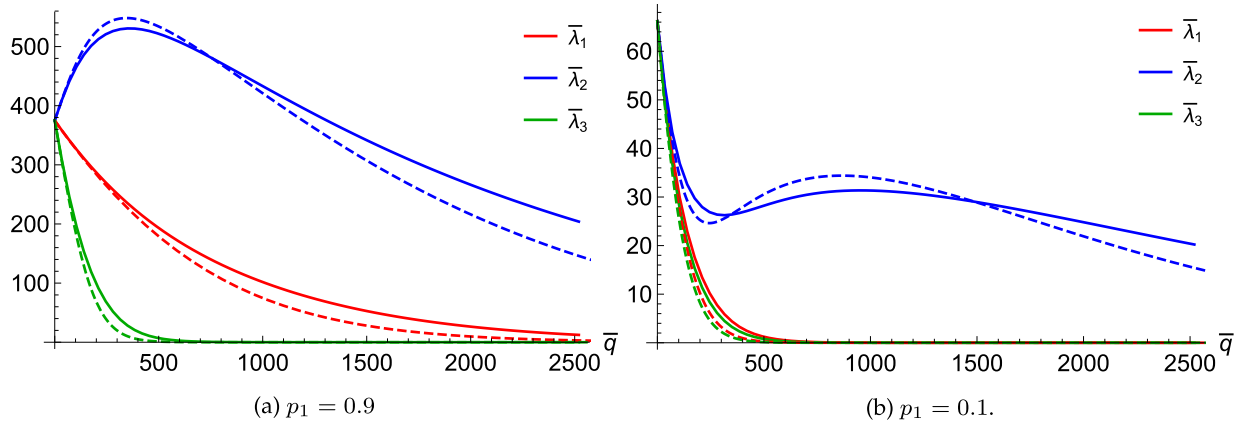
Example 2 (The impact of Markov platooning)**.** The purpose of this collection of numerical examples is to exhibit specific, interesting features of gap acceptance models that relate to the impact of Markov platooning. In the literature it has already been observed that platoon forming on the major road may have a positive impact on the capacity of the minor road. For the first example, which is similar to Example 1 but now with Markov platooning, we compare the capacity of the minor road for the three behavior types B_1–B_3. For the last two behavior types, we assume that a driver requires either a short gap of $T_1 = 4$ seconds, or an extremely long gap of $T_2 = 60$ seconds. Obviously these values are not chosen with the intention to mimic realistic behavior, but to point out extreme situations that might occur. For behavior type B_1, we take $T = p_1 T_1 + p_2 T_2$ seconds long, where $p_2 := 1 - p_1$.

For these settings, we compare the model with and without Markov platooning. With platooning, we take $\mu_1 = 1/60$ and $\mu_2 = 1/240$, resulting in exponential periods of, on average, one minute where the arrival rate on the major road is q_1, followed by exponential periods of, on average, four minutes, with arrival rate q_2. We assume a fixed ratio of q_1 and q_2, namely $q_1 = 3q_2$. The long-term average arrival rate equals

$$\bar{q} := \frac{q_1/\mu_1 + q_2/\mu_2}{1/\mu_1 + 1/\mu_2} = \frac{q_1\mu_2 + q_2\mu_1}{\mu_1 + \mu_2}.$$

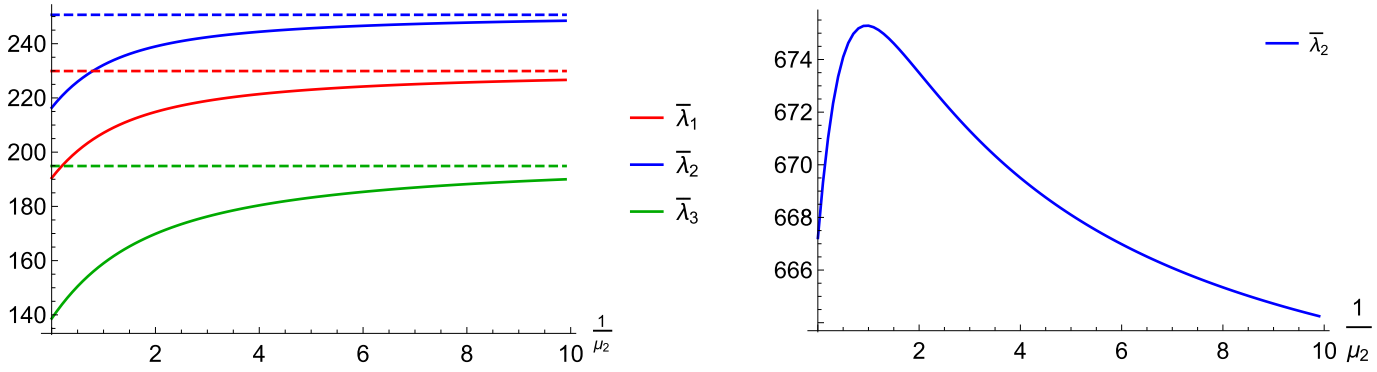
We compare the capacities with those obtained from the model without platooning, where we assume Poisson arrivals with rate $\bar{q}$.

Please cite this article as: Abhishek et al., Congestion analysis of unsignalized intersections: The impact of impatience and Markov platooning, European Journal of Operational Research (2018), https://doi.org/10.1016/j.ejor.2018.09.049



(a) $p_1 = 0.9$ (b) $p_1 = 0.1$.

Fig. 4. Capacity of the minor street (vehicles per hour) as a function of the average flow rate on the main road (vehicles per hour) in Example 2. The solid lines correspond to the model with Markov platooning; the dashed lines correspond to the model without platooning. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) $T_1 = 6.22, T_2 = 14, p_1 = 0.9$ (b) $T_1 = 3, T_2 = 60, p_1 = 0.9$

Fig. 5. Capacity of the minor street (vehicles per hour) as a function of the mean platoon length (seconds) in Example 3. The dashed lines in (a) indicate the limiting capacities for $\mu_2 \downarrow 0$ while keeping the ratio μ_1/μ_2 fixed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Figure 4 depicts the capacity (vehicles per hour) of the minor street as a function of $\bar{q}$, the average flow rate on the main road (vehicles per hour), for $p_1 = 0.9$ and $p_1 = 0.1$, respectively. As in the non-modulated case, we observe the relation $\bar{\lambda}_2 \geqslant \bar{\lambda}_1 \geqslant \bar{\lambda}_3$. Due to the lack of explicit expressions for $\bar{\lambda}_1, \bar{\lambda}_2$, and $\bar{\lambda}_3$, we cannot prove the strict ordering now. We did, however, observe it in all numerical examples that we conducted, and conjecture the ordering to hold true in general.

Based on the results of this example (and many other examples that are not discussed in the present paper) we are inclined to believe that platooning has a positive effect on the capacity of the minor road, but *only* for models B_1 and B_3. In a model with inconsistent behavior, it really depends on the model parameters whether platooning increases or decreases the capacity. This is nicely illustrated in Fig. 4(a) and even better in Fig. 4(b).

Example 3 (Platoon lengths)**.** In this example we fix the overall arrival rate on the major road, but we vary the platoon sizes. In more detail, we assume that $q_1 = 600$ vehicles per hour and $q_2 = 2400$ vehicles per hour. This means that phase 1 can be considered as a situation of moderate traffic (every 6 seconds a car passes), whereas phase 2 can be considered as one big platoon (on average every 1.5 seconds a car passes). The overall arrival rate $\bar{q}$ is fixed at 900 vehicles per hour, which implies that $\mu_1/\mu_2 = 1/5$. By varying the mean platoon length $1/\mu_2$ (in seconds) between 0 and 10, we will get better insight in the relation between platoon lengths and the capacity. Wegmann (1991, Section 5) conducted a very similar experiment, varying the mean number of vehicles per bunch. He observed that the capacity increases with increasing variance of gaps.

We consider two different distributions for the critical headways. First, we consider the situation with $T_1 = 6.22$, $T_2 = 14$, and $p_1 = 0.9$, which can be considered as a quite realistic situation that we have used before. In Fig. 5(a) we show the results for behavior types B_1–B_3. The relation between the capacity and the mean platoon length is in line with Wegmann (1991, Fig. 3). Our numerical experiments confirm that this is indeed typical behavior for B_1–B_3. Nevertheless, we want to show that it is possible to create a situation where model B_2 exhibits completely different behavior. When changing the distribution of the critical headway such that $T_1 = 3$ and $T_2 = 60$, we no longer see a monotonous relation between the capacity and the mean platoon length; see Fig. 5(b). Considering the fact that this inconsistent behavior type in combination with the extreme values for T_1 and T_2 might not be all too realistic, we do not find it likely that this type of behavior occurs in practical situations, but the model shows that it is not entirely impossible. For completeness, we want to mention that under extreme circumstances such as mean platoon lengths of 1000 seconds, the capacity with consistent behavior B_3 will also exhibit a drop, but not as drastically as in Fig. 5(b).

The final conclusion that can be drawn from this example, is that one should be cautious when developing capacity estimates based on Eqs. (3.1) and (3.2). This type of reasoning may create a substantial bias, due to the fact that the vehicle at the head of the queue typically does not see the background process in equilibrium. It is noted that such argumentation underlies the capacity formulae in e.g. Wu (2001), where the capacity is calculated by conditioning on the state of the background process, i.e., the state of the traffic on the major road (free space, free flow, bunching, or queueing). This example, and also Wegmann's example, clearly show that there is a clear dependency between the mean platoon size and the capacity. The parameters in these examples are carefully chosen, such that the steady-state distribution of the background process (the vector π) remains unchanged. In our case, the major road is in state 'free flow' for a fraction $\pi_1 = 5/6$ of the time, and in

Please cite this article as: Abhishek et al., Congestion analysis of unsignalized intersections: The impact of impatience and Markov platooning, European Journal of Operational Research (2018), https://doi.org/10.1016/j.ejor.2018.09.049

state 'bunched' for a fraction $\pi_2 = 1/6$ of the time. If one would use the naïve approach and determine $\mathbb{E}[S_1]$ and $\mathbb{E}[S_2]$ by considering two separate models with regular Poisson arrivals, with intensities respectively q_1 and q_2, and use Eq. (3.1), the capacities for models B_1, B_2, and B_3, respectively, would be

$$\bar{\lambda}_1 = 229.91, \bar{\lambda}_2 = 250.65, \bar{\lambda}_3 = 194.89,$$

independent of μ_1 and μ_2. From Figs. 5(a) and 3 in Wegmann (1991), it is clearly visible that these values (indicated by the dashed lines in Fig. 5(a)) may differ substantially from the actual capacities. In fact, the capacities calculated from (3.1) can be interpreted as the limiting capacities from our MMPP model when $\mu_2 \downarrow 0$ while keeping the ratio μ_1/μ_2 fixed. When using (3.2) to compute the capacities, one would obtain

$$\bar{\lambda}_1 = 96.28, \bar{\lambda}_2 = 130.74, \bar{\lambda}_3 = 11.63,$$

leading to even more substantial errors.

4. Impatience

The goal of this section is to add more realism to the model by also incorporating driver's impatience. As evidenced by Abou-Henaidy et al. (1994), drivers tend to grow more impatient as the number of rejected gaps increases. This impatience may result in an increased willingness to accept smaller gaps. To the best of our knowledge, Abhishek et al. (2016) was the first to present new results for gap acceptance models that include impatience *and* randomness in the critical headways.

As discussed in Section 2, we incorporate impatience by letting the critical headway depend on the number of failed attempts. Denote by T_m the critical headway for the mth attempt to enter the main road ($m = 1, 2, \ldots$). For models B_1 and B_3, we assume that $T_1 \geq T_2 \geq \cdots \geq T_{\min}$ (more details below). Due to the resampling in model B_2, we cannot make this assumption for this model, but we can assume that $T_i \geqslant_{st} T_{i+1}$, where $\geq_{st}$ is used to denote that T_i is stochastically greater than T_{i+1}. We impose the (reasonable) assumption that after a certain attempt number, say the Mth attempt, the critical gap does not decrease any further, i.e., $T_M = T_{M+1} = T_{M+2} = \cdots$. For reasons of compactness, we mainly focus on the differences with Section 3.

B_1 (constant gap): Every driver on the minor road needs the same constant critical headway T_m for the mth attempt to enter the main road ($m = 1, 2, \ldots, M$). We follow the same approach as before, defining a cycle as the time between two consecutive epochs such that the background process is in state 1 and the service of a low-priority car is completed. Compared to Section 3, we now need to make several variables dependent on the attempt number m. First, to make our model Markovian, we approximate each (deterministic) T_m by an Erlang random variable with k_m phases and rate $\kappa_m := k_m/T_m$. Define $\varrho_i^{(m)} := \mu_i + q_i + \kappa_m$ and let $h_{ij}^{(m)}$ be the mean number of cars that is served till the cycle ends, given that the current state of the background process is $i \in \{1, \ldots, d\}$ and the car in service is in its mth attempt, having finished $j \in \{0, \ldots, k_m - 1\}$ phases of the Erlang distribution. The equivalent of Eq. (3.3), for the model with impatience, becomes

$$h_{ij}^{(m)} = \sum_{\ell \neq i} \frac{\mu_{i\ell}}{\varrho_i^{(m)}} h_{\ell j}^{(m)} + \frac{q_i}{\varrho_i^{(m)}} h_{i0}^{(m+1)} + \frac{\kappa_m}{\varrho_i^{(m)}} \times \begin{cases} h_{i,j+1}^{(m)} & \text{if } j < k_m - 1, \\ 1 & \text{if } j = k_n - 1, i = 1, \\ 1 + h_{i0}^{(1)} & \text{if } j = k_m - 1, i > 1. \end{cases} \tag{4.1}$$

It is clearly seen that the arrival of a new car on the major road (with rate q_i in background state i) increases the attempt number, while a service completion (with rate κ_m when $i > 1$ and $j = k_m - 1$) resets the attempt number to one.

To determine the mean cycle duration, we make similar adaptations, defining $\tau_{ij}^{(m)}$ as the mean time till the end of the current cycle, given that the current state of the background process is $i \in \{1, \ldots, d\}$ and the car in service is in its mth attempt, having finished $j \in \{0, \ldots, k_m - 1\}$ phases of the Erlang distribution. The equivalent of (3.4) is

$$\tau_{ij}^{(m)} = \frac{1}{\varrho_i^{(m)}} + \sum_{\ell \neq i} \frac{\mu_{i\ell}}{\varrho_i^{(m)}} \tau_{\ell j}^{(m)} + \frac{q_i}{\varrho_i^{(m)}} \tau_{i0}^{(m+1)} + \frac{\kappa_m}{\varrho_i^{(m)}} \times \begin{cases} \tau_{i,j+1}^{(m)} & \text{if } j < k_m - 1, \\ 0 & \text{if } j = k_n - 1, i = 1, \\ \tau_{i0}^{(1)} & \text{if } j = k_m - 1, i > 1. \end{cases} \tag{4.2}$$

To solve the sets of Eqs. (4.1) and (4.2), we now use that $T_M = T_{M+1} = T_{M+2} = \cdots$, which also implies that

$$h_{ij}^{(M)} = h_{ij}^{(M+1)} = h_{ij}^{(M+2)} = \cdots$$

and

$$\tau_{ij}^{(M)} = \tau_{ij}^{(M+1)} = \tau_{ij}^{(M+2)} = \cdots.$$

If we replace $h_{i0}^{(m+1)}$ by $h_{i0}^{(M)}$ and $\tau_{i0}^{(m+1)}$ by $\tau_{i0}^{(M)}$, for $m = M$, we have two sets of dKM equations each. After solving these, we can evaluate the capacity of the system as

$$\bar{\lambda}_1 := h_{10}^{(1)}/\tau_{10}^{(1)}.$$

B_2 (sampling per attempt): Now the driver samples a 'fresh' random T_m for the mth attempt. We model this randomness by assuming that the critical headway equals some deterministic $T_{n,m}$ with probability $p_{n,m}$, for $n \in \{1, \ldots, N\}$ and $m = 1, 2, \ldots$. As before, we approximate $T_{n,m}$ by an Erlang random variable with $k_{n,m}$ phases; each of the phases is exponentially distributed with parameter $\kappa_{n,m} = k_{n,m}/T_{n,m}$. Define $\varrho_i^{(n,m)} := \mu_i + q_i + \kappa_{n,m}$ and let $h_{ij}^{(n,m)}$ be the mean number of cars that is served till the cycle ends, given that the current state of the background process is $i \in \{1, \ldots, d\}$ and the car in service is in its mth attempt with gap size $T_{n,m}$, having finished $j \in \{0, \ldots, k_{n,m} - 1\}$ phases of the Erlang distribution. We wish to find $h_{i0} := h_{i0}^{(1)}$, defined as

$$h_{i0}^{(m)} = \sum_{n=1}^{N} p_{n,m} h_{i0}^{(n,m)}, \qquad \text{for } i = 1, 2, \ldots, d; m = 1, 2, \ldots. \tag{4.3}$$

Then

$$h_{ij}^{(n,m)} = \sum_{\ell \neq i} \frac{\mu_{i\ell}}{\varrho_i^{(n,m)}} h_{\ell j}^{(n,m)} + \frac{q_i}{\varrho_i^{(n,m)}} h_{i0}^{(m+1)} + \frac{\kappa_{n,m}}{\varrho_i^{(n,m)}} \times \begin{cases} h_{i,j+1}^{(n,m)} & \text{if } j < k_{n,m} - 1, \\ 1 & \text{if } j = k_{n,m} - 1, i = 1, \\ 1 + h_{i0}^{(1)} & \text{if } j = k_{n,m} - 1, i > 1. \end{cases} \tag{4.4}$$

The required changes to determine the expected cycle length are similar. The modification of Eq. (3.7) that incorporates impatience is given below, where we have used $\tau_{ij}^{(n,m)}$ to denote the mean time till the current cycle ends, given that the current state of the background process is $i \in \{1, \ldots, d\}$, the car in the service has critical gap size $T_{n,m}$, and the number of completed service time phases is $j \in \{0, \ldots, k_{n,m} - 1\}$.

$$\tau_{ij}^{(n,m)} = \frac{1}{\varrho_i^{(n,m)}} + \sum_{\ell \neq i} \frac{\mu_{i\ell}}{\varrho_i^{(n,m)}} \tau_{\ell j}^{(n,m)} + \frac{q_i}{\varrho_i^{(n,m)}} \tau_{i0}^{(m+1)} + \frac{\kappa_n}{\varrho_i^{(n,m)}} \times \begin{cases} \tau_{i,j+1}^{(n,m)} & \text{if } j < k_{n,m} - 1, \\ 0 & \text{if } j = k_{n,m} - 1, i = 1, \\ \tau_{i0}^{(1)} & \text{if } j = k_{n,m} - 1, i > 1, \end{cases} \tag{4.5}$$

Please cite this article as: Abhishek et al., Congestion analysis of unsignalized intersections: The impact of impatience and Markov platooning, European Journal of Operational Research (2018), https://doi.org/10.1016/j.ejor.2018.09.049

with $\tau_{i0}^{(m)} = \sum_{n=1}^{N} p_{n,m}\tau_{i0}^{(n,m)}$. The capacity of the low-priority queue is

$$\bar{\lambda}_2 = h_{10}^{(1)}/\tau_{10}^{(1)}.$$

B_3 (sampling per driver): In this variant, every car driver samples a random T_1 at his first attempt. We model this randomness analogously to model B_2, assuming that the first critical headway T_1 equals some deterministic $T_{n,1}$ with probability $p_{n,1}$, for $n \in \{1, \ldots, N\}$. The difference with B_2 is that this (random) value determines the complete sequence of critical headways at subsequent attempts. We denote these critical headways by $T_{n,m}$, for $n \in \{1, \ldots, N\}$ and $m = 1, 2, \ldots$. The definitions of $\kappa_{n,m}$, $k_{n,m}$, $\varrho_i^{(n,m)}$ and even $h_{i0}^{(m)}$, $h_{ij}^{(n,m)}$ and $\tau_{ij}^{(n,m)}$ remain unchanged compared to the previous model. The latter should be computed differently, though. The capacity of the low-priority queue is

$$\bar{\lambda}_3 = h_{10}^{(1)}/\tau_{10}^{(1)}.$$

The unknowns in the right-hand side follow from the following two systems of equations.

$$h_{ij}^{(n,m)} = \sum_{\ell\neq i} \frac{\mu_{i\ell}}{\varrho_i^{(n,m)}} h_{\ell j}^{(n,m)} + \frac{q_i}{\varrho_i^{(n,m)}} h_{i0}^{(n,m+1)} + \frac{\kappa_{n,m}}{\varrho_i^{(n,m)}} \times \begin{cases} h_{i,j+1}^{(n,m)} & \text{if } j < k_{n,m}-1, \\ 1 & \text{if } j = k_{n,m}-1, i = 1, \\ 1 + h_{i0}^{(1)} & \text{if } j = k_{n,m}-1, i > 1, \end{cases} \tag{4.6}$$

$$\tau_{ij}^{(n,m)} = \frac{1}{\varrho_i^{(n,m)}} + \sum_{\ell\neq i} \frac{\mu_{i\ell}}{\varrho_i^{(n,m)}} \tau_{\ell j}^{(n,m)} + \frac{q_i}{\varrho_i^{(n,m)}} \tau_{i0}^{(n,m+1)} + \frac{\kappa_n}{\varrho_i^{(n,m)}} \times \begin{cases} \tau_{i,j+1}^{(n,m)} & \text{if } j < k_{n,m}-1, \\ 0 & \text{if } j = k_{n,m}-1, i = 1, \\ \tau_{i0}^{(1)} & \text{if } j = k_{n,m}-1, i > 1. \end{cases} \tag{4.7}$$

Again, the only differences with Eqs. (4.4) and (4.5), for model B_2, are the terms corresponding to the arrival of a new vehicle on the major road (with rate q_i in background state i) which does *not* lead to resampling.

Example 4 (Impatience and platooning)**.** We now revisit the model from Example 1, but we add driver impatience and Markov platooning. Impatience is added in the following specific form:

$$T_{m+1} = \alpha(T_m - \Delta) + \Delta, \qquad m = 1, 2, \ldots; 0 < \alpha < 1, \tag{4.8}$$

which means that the critical headway decreases after each failed attempt. Eventually, it will approach the limiting value of Δ. The parameter α determines the speed at which the patience decreases. In Scenario 1 all T_k are fixed, with $T_1 = 7$ seconds. In Scenarios 2 and 3, each of the T_k is a random variable, but with a slightly different distribution than in Example 1: T_1 equal to 4 or 14 seconds, with probability 7/10 and 3/10, respectively. The distribution of T_m for $m > 1$ can be determined from Eq. (4.8). In Scenario 2, the impatience is a new random sample at each attempt, independent of the value of T_{m-1}. In Scenario 3, as before, each driver samples a random impatience T_1 exactly once. The value of T_1 (which is again either 4 or 14 seconds) determines the whole sequence of critical gap times at the subsequent attempts according to (4.8). In the numerical algorithm we take $k = 200$ and $M = 10$, meaning that the effect of impatience stops after the tenth attempt (when T_1 has dropped from 7 to 4.006 seconds).

Besides impatience we also include platooning, which we do in a similar way as in Example 2, taking $\mu_1 = 1/60$ and $\mu_2 = 1/240$, while varying arrival rates q_1 and q_2. Fig. 6 shows the capacity as a function of the average flow rate $\bar{q}$, when $\alpha = 9/10$ and $\Delta = 4$ seconds. The solid lines correspond to the model with impatience and platooning; the dashed lines correspond to a model with only platooning. In this example it can be seen that the capacity increases significantly due to the impatience. Although we have chosen not to include it in this figure, it turns out that platooning had a much smaller impact on the capacities. Note that, compared to Example 1, we have drastically increased the variability of the random variable representing a critical gap, while keeping the expected value equal. We have done this to show that an interesting phenomenon can occur. Strikingly, Model B_1 now performs *worst* in terms of capacity. It was already observed in other experiments (cf. Abhishek et al., 2016) that impatience might destroy the relation $\bar{\lambda}_2 \geqslant \bar{\lambda}_1 \geqslant \bar{\lambda}_3$ for some values of q, but in this example we observe that $\bar{\lambda}_1$ is smaller than the other two for *all* arrival rates on the major road. If we would have taken the same critical gap distributions as in Example 1 (with T_1 equal to 6.22 or 14 seconds, with probability 9/10 and 1/10, respectively), $\bar{\lambda}_1$ would still be the smallest of the three capacities, but the difference would then be almost negligible. In fact, it is the value of α that affects the ordering most significantly. If $\alpha = 0.9$, model B_1 is only worse than B_3 for very high flow rates on the major road. We refer to Abhishek et al. (2016, Fig. 3) for more details.

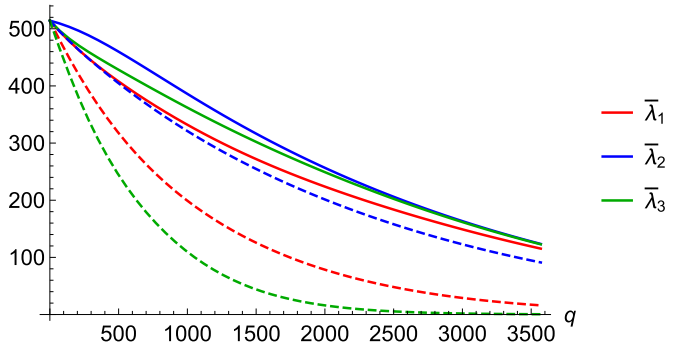


Fig. 6. Capacity of the minor street (vehicles per hour) as a function of the average flow rate on the main road (vehicles per hour) in Example 4. The solid lines correspond to the model with platooning and impatience. The dashed lines correspond to only platooning. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Critical gaps T_m (in seconds) for different values of α and m.

m	1	2	3	4	5	10
$\alpha = 0.2$	7.000	4.600	4.120	4.024	4.005	4.000
$\alpha = 0.5$	7.000	5.500	4.750	4.375	4.188	4.006
$\alpha = 0.8$	7.000	6.400	5.920	5.536	5.229	4.403

To conclude this example, we discuss the impact of the parameters α and M on the capacity. The parameter α represents the rate at which the critical gap decreases towards its limiting value Δ, while M is the attempt number after which the critical gap does not decrease any further. Table 1 presents T_m for $m \in \{1, 2, 3, 4, 5, 10\}$ for $\alpha \in \{0.2, 0.5, 0.8\}$, computed using (4.8).

To get more insight in the impact of these parameters on the capacities, we take the aforementioned values for α and compute the capacities for model B_1 without Markov platooning, for various values of M. The results for models B_2 and B_3, and for models with Markov platooning are omitted because their behavior is very similar. Table 2 shows the results, for $q = 300$ and $q = 1200$. Clearly the capacity is decreasing in α (because larger values of α correspond to larger critical gaps) and increasing in M (because the critical gaps decrease further when M is larger). It makes sense that the impact of α and M on the capacities is heavily dependent on the arrival rate on the major road. If q is small, vehicles on the minor road will quickly find an acceptable gap and M will

Table 2
Capacities (vehicles per hour) for different values of α and M.

$q = 300$ vehicles per hour					
M	2	3	4	5	10
$\alpha = 0.2$	463.3	469.2	469.5	469.5	469.5
$\alpha = 0.5$	429.9	439.7	441.2	441.5	441.5
$\alpha = 0.8$	398.9	405.5	407.5	408.1	408.3
$q = 1200$ vehicles per hour					
M	2	3	4	5	10
$\alpha = 0.2$	288.9	326.4	332.3	333.1	333.3
$\alpha = 0.5$	214.6	263.0	284.1	292.3	297.1
$\alpha = 0.8$	159.4	183.0	200.7	213.1	233.1

have hardly any influence on the capacity. The influence of α is also much smaller than when $q = 1200$. For large values of q, vehicles on the minor road will need more attempts before succeeding in finding a suitable gap. In this case, the capacity can be greatly improved by having smaller values of α, or larger values of M. We want to emphasize, though, that α is a parameter that should be obtained from the empirical data, whereas M can be considered as a model parameter that one can choose, depending on the observed value of α.

5. Concluding remarks

In this paper we have developed tools to evaluate the capacity of an unsignalized, priority-controlled intersection. Importantly, the model developed incorporates impatience and platooning, whereas various types of driver behavior (related to the low-priority drivers) are covered. The underlying Markov model facilitates the evaluation of the capacity by solving an elementary system of linear equations (which is a standard numerical operation that can be done highly efficiently).

Through a sequence of numerical examples we have been able to validate various theoretical results. It is confirmed that in the setting without platooning and impatience inconsistent behavior always leads to a higher capacity than strictly deterministic behavior, while consistent behavior performs worst in terms of capacities. Empirically we observe the same ordering when platooning is added. When impatience is added, however, this ordering is lost; cf. also the preliminary findings in Abhishek et al. (2016).

In a platooning model in which the arrival rate at the major road alternates between two values, it is tempting to compute the capacity as an average of the two capacities pertaining to the two individual arrival rates. An important insight from this paper, is that one needs to be extremely careful with this naïve procedure, as it may lead to significant errors. It stresses that our approach, featuring *one* model with various arrival rates, produces more reliable estimates of the capacity of the minor road.

Acknowledgments

The research of Abhishek and M. Mandjes is partly funded by NWO Gravitation project NETWORKS, grant number 024.002.003. The authors thank Onno Boxma (Eindhoven University of Technology) and Bart van Arem (Delft University of Technology) for helpful discussions.

References

Abhishek, Boon, M. A. A., Mandjes, M., & Núñez Queija, R. (2016). Congestion analysis of unsignalized intersections. In *Proceedings of the intelligent transportation systems workshop, COMSNETS 2016* (pp. 1–6).
Abou-Henaidy, M., Teply, S., & Hund, J. H. (1994). Gap acceptance investigations in Canada. In R. Akçelik (Ed.), *Proceedings of the second international symposium on highway capacity: 1* (pp. 1–19).
Catchpole, E. A., & Plank, A. W. (1986). The capacity of a priority intersection. *Transportation Research-B, 20B*(6), 441–456.
Cheng, T. E. C., & Allam, S. (1992). A review of stochastic modelling of delay and capacity at unsignalized priority intersections. *EJOR, 60*(3), 247–259.
Drew, D. R., Buhr, J. H., & Whitson, R. H. (1967). The determination of merging capacity and its applications to freeway design and control. *Report 430-4*. Texas Transportation Institute.
Drew, D. R., LaMotte, L. R., Buhr, J. H., & Wattleworth, J. (1967). Gap acceptance in the freeway merging process. *Report 430-2*. Texas Transportation Institute.
Gaur, A., & Mirchandani, P. (2001). Method for real-time recognition of vehicle platoons. *Transportation Research Record, 1748*, 8–17.
Guo, R. J., & Lin, B. L. (2011). Gap acceptance at priority-controlled intersections. *Journal of Transportation Engineering, 137*(4), 269–276.
Guo, R.-J., Wang, X.-J., & Wang, W.-X. (2014). Estimation of critical gap based on Raff's definition. *Computational Intelligence and Neuroscience, 2014* Article ID 236072, 7 pages. 10.1155/2014/236072.
Heidemann, D., & Wegmann, H. (1997). Queueing at unsignalized intersections. *Transportation Research-B, 31*(3), 239–263.
Horn, R. A., & Johnson, C. R. (1986). *Matrix analysis*. New York, NY, USA: Cambridge University Press.
Jia, D., Lu, K., Wang, J., Zhang, X., & Shen, X. (2016). A survey on platoon-based vehicular cyber-physical systems. *IEEE Communications Surveys & Tutorials, 18*(1), 263–284.
Li, B. (2017). Stochastic modeling for vehicle platoons (I): Dynamic grouping behavior and online platoon recognition. *Transportation Research Part B, 95*, 364–377.
Mayne, A. J. (1954). Some further results in the theory of pedestrians and road traffic. *Biometrika, 41*(3/4), 375–389.
Takagi, H. (1991). Queueing analysis: A foundation of performance evaluation. *Vacation and priority systems, Part 1*: 1. Amsterdam: North-Holland.
Tanner, J. C. (1951). The delay to pedestrians crossing a road. *Biometrika, 38*(3/4), 383–392.
Tanner, J. C. (1962). A theoretical analysis of delays at an uncontrolled intersection. *Biometrika, 49*(1/2), 163–170.
Wegmann, H. (1991). Intersections without traffic signals II. In W. Brilon (Ed.), *A general capacity formula for unsignalized intersections* (pp. 177–191). Springer-Verlag.
Wei, D., Kumfer, W., Wu, D., & Liu, H. (2018). Traffic queuing at unsignalized cross-walks with probabilistic priority. *Transportation Letters, 10*(3), 129–143.
Weiss, G. H., & Maradudin, A. A. (1962). Some problems in traffic delay. *Operations Research, 10*(1), 74–104.
Wu, N. (2001). A universal procedure for capacity determination at unsignalized (priority-controlled) intersections. *Transportation Research Part B, 35*, 593–623.
Wu, N. (2012). Equilibrium of probabilities for estimating distribution function of critical gaps at unsignalized intersections. *Transportation Research Record, 2286*, 49–55.