**Generalized Gap Acceptance Models
for Unsignalized Intersections**

Abishek, M. Boon, M. Mandjes

GENERALIZED GAP ACCEPTANCE MODELS
FOR UNSIGNALIZED INTERSECTIONS

ABHISHEK[1], MARKO BOON[2], AND MICHEL MANDJES[1]

ABSTRACT. This paper contributes to the modeling and analysis of unsignalized intersections. In classical gap acceptance models vehicles on the minor road accept any gap greater than the *critical* gap, and reject gaps below this threshold, where the gap is the time between two subsequent vehicles on the major road. The main contribution of this paper is to develop a series of generalizations of existing models, thus increasing the model's practical applicability significantly. First, we incorporate driver impatience behavior while allowing for a realistic merging behavior; we do so by distinguishing between the critical gap and the merging time, thus allowing *multiple* vehicles to use a sufficiently large gap. Incorporating this feature is particularly challenging in models with driver impatience. Secondly, we allow for multiple classes of gap acceptance behavior, enabling us to distinguish between different driver types and/or different vehicle types. Thirdly, we use the novel $M^X/SM2/1$ queueing model, which has batch arrivals, dependent service times, and a different service-time distribution for vehicles arriving in an empty queue on the minor road (where 'service time' refers to the time required to find a sufficiently large gap). This setup facilitates the analysis of the service-time distribution of an arbitrary vehicle on the minor road and of the queue length on the minor road. In particular, we can compute the *mean* service time, thus enabling the evaluation of the capacity for the minor road vehicles.

Keywords: unsignalized intersection, priority-controlled intersection, merging time, gap acceptance with impatience, stochastic capacity analysis, queueing theory.

1. INTRODUCTION

In this era of rapidly emerging new technologies for urban traffic control, the vast majority of urban traffic intersections are still priority-based unsignalized intersections, in which the major roads have priority over the minor roads. Due to the increasing level of traffic congestion, the need for state-of-the-art quantitative analysis methods is greater than ever. Performance analysis of unsignalized traffic intersections is traditionally based on so-called *gap acceptance models*, with roots that can be traced back to classical queueing models. The basis of these models is the assumption that the crossing decision of a driver on the minor road is based on the gap between two successive vehicles on the major road [16]. This model can be applied in other contexts as well, e.g. when analyzing freeways [17, 18, 24] and pedestrian crossings [31, 34].

In this paper we contribute to the modeling and analysis of unsignalized intersections. In classical gap acceptance models vehicles on the minor road accept any gap greater than the *critical* gap (or critical headway), and reject gaps below this threshold, where the gap is defined as the time between two subsequent vehicles on the major road. However, in reality, drivers typically do not need the full critical gap; the remaining part can be used by the next vehicle on the minor road.

In the literature, the part of the critical gap that is really used by the vehicle is referred to as the *merging* time.

Although gap acceptance models have improved greatly since their introduction more than fifty years ago, there are still some fundamental limitations to their practical applicability. As pointed out recently by Liu *et al.* [23], two of its main limitations are the inability to incorporate (1) different driver behaviors and (2) heterogeneous traffic; see also [28, 29]. The main reason why these features could not be included is that the currently used analysis techniques are based on a single-server queueing model with exceptional first service (see also [36, 38, 39]). In the gap acceptance literature, this model is commonly referred to as the M/G2/1 queue, a term seemingly introduced by Daganzo [14]. Importantly, in this queue, the 'service times' (corresponding to the time required to search for a sufficiently large gap and crossing the intersection or, depending on the application, merging with the high-priority traffic flow) are assumed independent. Incorporating more realistic features, such as driver impatience or different types of driver behavior, creates dependencies that make the model significantly more difficult to analyze.

An important development, that helps us overcome this obstacle, is the recently developed framework for the analysis of queueing models that allows a highly general correlation structure between successive service times (the $M^X/SM2/1$ queue [4]). Although set up with general applications in mind, this framework turns out to be particularly useful in road traffic models, where the dependencies can be used to model clustering of vehicles on the major road, to differentiate between multiple types of driver behavior, or to account for heterogeneous traffic. The present paper focuses on the latter two features.

In the existing literature, various quantitative methods have been used to study unsignalized intersections. The two most common procedures are empirical regression techniques and gap-acceptance models [8], but other methods have been employed as well (for example the additive conflict flows technique by Brilon and Wu [9]). A topic of particular interest concerns the capacity of the minor road [7], which is defined as the maximum possible number of vehicles per time unit that can pass through an intersection from the minor road. Other relevant performance measures are the queue length and the delay on the minor road.

Due to the driver characteristics such as age, gender, years of driving experience, as well as the vehicle's features, in particular the size and acceleration speed of the vehicle, there are different types of driver behavior. First, we distinguish between fast vehicles and slow vehicles. The differences in speed/acceleration can be caused by human factors (risk avoidance) or by different vehicle types (e.g. cars, trucks). Second, we assume that drivers grow more and more impatient while waiting for a sufficiently large gap on the major road, which was actually confirmed in an empirical data study by Abou-Henaidy *et al.* [5]. Only very few papers on gap acceptance models incorporate driver impatience in the model ([2, 17, 18, 35]), because it is known to complicate the analysis drastically [22], in particular when trying to maintain a realistic merging behavior.

In earlier papers, such as [2], three variations of gap acceptance models can be distinguished. The first is the basic model in which all low-priority drivers are assumed to use the same fixed critical gap. In the second model, critical gaps are random, with drivers sampling a new critical gap at each new attempt. This is typically referred to as *inconsistent* driver behavior. The third model is also known as the *consistent* model, in which a random critical gap is sampled by each driver for his first attempt only, and the driver then uses that same value at his subsequent attempts. The

present paper generalizes all three gap acceptance models into a more realistic model that covers various realistic driving behavior features.

One of the first studies using queueing models for unsignalized intersections was by Tanner [32], who assumed constant critical gap and move-up time. Tanner first determines the mean delay of the low priority vehicles, and characterizes the capacity of the minor road as the arrival rate of the low-priority vehicles, at which the mean delay grows beyond any bound. Tanner's model has been generalized in various ways [10, 12, 20, 21, 33, 35] by allowing random critical gaps and move-up times, also analyzing performance measures such as the queue length and the waiting time on the minor road. Heidemann and Wegmann [22] give an excellent overview of the earlier existing literature. Moreover, they add a stochastic dependence between the critical gap and the merging (move-up) time, and study the minor road as an M/G2/1 queue, to determine the queue length, the delay and the capacity on the minor road. A further generalization, dividing the time scale of the major stream into four regimes (viz. free space, single vehicle, bunching, and queueing) was investigated by Wu [37].

Although the literature on gap acceptance models is relatively mature and the existing gap acceptance models have proven their value, they are sometimes criticized for *looking like* rigorous mathematics, but in reality being based on pragmatic simplifications. As a consequence, the produced results might be of a correct magnitude, but are, however, only of approximative nature [9]. The main goal of this paper is to increase the significance of gap acceptance models by taking away some of these concerns by including essential new features that are typically encountered in practice. In more detail, the main novelties of this paper are the following.

- First, we incorporate driver impatience behavior while allowing for a realistic merging behavior. Put more precisely, we distinguish between the critical gap and the merging time, thus allowing *multiple* vehicles to use a sufficiently large gap. Incorporating this feature is particularly challenging in models with driver impatience.
- Secondly, we allow for multiple classes of gap acceptance behavior, enabling us to distinguish between different driver types and/or different vehicle types.
- Thirdly, we use a queueing model in which vehicles arriving in an empty queue on the minor road have different service-time distributions than the queueing vehicles (where 'service time' is meant in the sense introduced above). This setup facilitates the analysis of the queue length on the minor road as well as the service-time distribution of an arbitrary low-priority vehicle. The capacity for the minor road vehicles is then derived from the expectation of the service time for queuers.

The remainder of this paper is organized as follows. In Section 2 we introduce our model. Then the full queue-length analysis is presented in Section 3, whereas the capacity of the minor road is determined in Section 4. In Section 5 numerical examples demonstrate the impact of driver behavior on the capacity of the minor road.

2. Model description

In this section we provide a detailed mathematical description of our new model. We study an unsignalized, priority controlled intersection where drivers on the major road have priority over the drivers on the minor road. For notational convenience, we will focus on the situation with one traffic stream on each road, but several extensions also fall in our framework (see, for example, Figure 1). Vehicles on the major road arrive according to a Poisson process with intensity q. On

the minor road, we have a *batch* Poisson arrival process with λ denoting the arrival intensity of the batches (platoons) and B denoting the (random) platoon size. We assume that traffic on the major road is not hindered by traffic on the minor road, which is a reasonable assumption in most countries (but not all, see [23]).
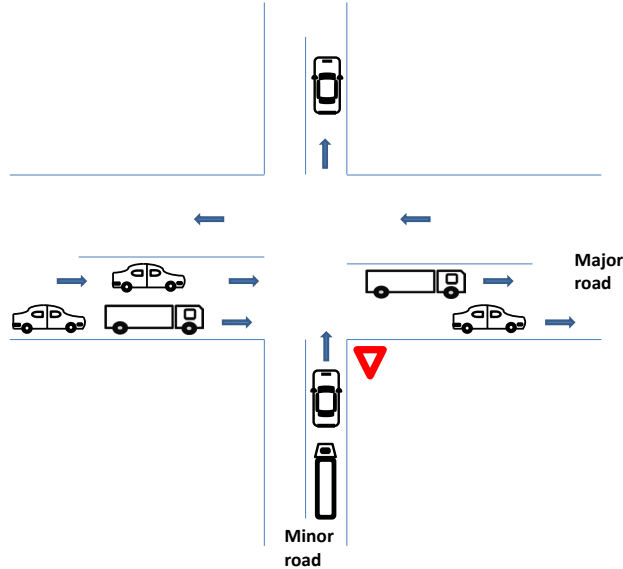

FIGURE 1. The situation analyzed in this paper.

The drivers on the minor road cross the intersection as soon as they come across a sufficiently large gap between two subsequent vehicles on the major road. Any gap that is too small for a driver on the minor road is considered a *failed attempt* to cross the intersection and adds to the impatience of the driver.

In the existing literature, *three* model variants have been introduced: (1) the standard model with constant critical gaps, (2) inconsistent gap acceptance behavior where a new random critical gap is sampled at each attempt, and (3) consistent behavior where drivers sample a random critical gap which they use for all attempts (cf. [2, 22]). In this paper we apply a framework that allows us to create a *single* model that generalizes all three aforementioned model variations, with a combination of consistent *and* inconsistent behavior, allowing for heterogeneous traffic and driver impatience. This is a major enhancement of the existing models, as such a general model has not been successfully analyzed so far. Unfortunately, this requires a significant adaptation of the underlying M/G2/1 queueing model that has been the basis of all gap acceptance models so far.

We will now describe the full model dynamics in greater detail. We distinguish between $R \in \{1, 2, \ldots\}$ driver/vehicle profiles. The profile of any arriving vehicle on the minor road is modeled as a random variable that equals r with probability p_r (with $\sum_{r=1}^{R} p_r = 1$). The profile represents the *consistent part* of the gap selecting behavior. Each driver profile has its own critical gap acceptance behavior, allowing us to distinguish between drivers that are willing to accept small gaps and (more cautious) drivers that require longer gaps, but also to allow for heterogeneous traffic with, for example, trucks requiring larger gaps than cars or motor cycles.

We introduce $T_{(i,r)}$ to denote the critical gap of a driver with profile r during his i-th attempt. The *impatient behavior* is incorporated in our framework by letting the critical gap $T_{(i,r)}$ depend on the attempt number $i = 1, 2, 3, \ldots$. This generalizes classical gap acceptance models where the critical gap is the same for all driver types and throughout all attempts.

Finally, we introduce the *inconsistent* component of our model by allowing the critical gap of a driver of type r in attempt i ($T_{(i,r)}$, that is) to be a discrete random variable which can take any of the following values:

$$T_{(i,r)} = \begin{cases} u_{(i,1,r)} & \text{with probability } p_{(i,1,r)}, \\ u_{(i,2,r)} & \text{with probability } p_{(i,2,r)}, \\ \vdots \\ u_{(i,M,r)} & \text{with probability } p_{(i,M,r)}, \end{cases}$$

with $\sum_{k=1}^{M} p_{(i,k,r)} = 1$ for all (i,r). Inconsistent behavior is used in the existing literature when the driver samples a new random critical gap after each failed attempt. It is generally being criticized for not being very realistic, because it is unlikely that a driver's gap acceptance behavior fluctuates significantly throughout multiple attempts. In our model, however, it is an excellent way to include some randomness in the gap selection process (which is due to the fact that not every driver from the same profile will have *exactly* the same critical gap at his i-th attempt), because we can limit the variability by cleverly choosing the values $u_{(i,k,r)}$.

Realistic values for the critical gaps $u_{(i,k,r)}$ can be empirically obtained by measuring the lengths of all rejected and accepted gaps for each vehicle type. One needs to distinguish between systematic fluctuations due to different driver/vehicle types and (smaller) random fluctuations within the same driver profile. Since the index i represents impatience, it makes sense that $u_{(i,k,r)}$ is decreasing in i. Moreover, since r denotes the driver profile, $u_{(i,k,r)}$ is relatively large for driver profiles r that represent slow vehicles/drivers and vice versa. The variability in $u_{(i,1,r)}, \ldots, u_{(i,M,r)}$ can be limited to avoid unrealistic fluctuations. For example: if driver profile $r = 1$ corresponds to trucks and $r = 2$ corresponds to motor cycles, then it may be quite realistic that

$$\min_{1 \leq k \leq M} u_{(i,k,1)} > \max_{1 \leq k \leq M} u_{(i,k,2)}.$$

Paraphrasing, the slowest motor cycle will always be faster than the fastest truck.

To make the model even more realistic, we assume that the actual vehicle *merging time*, denoted by Δ_r for profile r, is less than the critical gap $T_{(i,r)}$. As a consequence, the remaining part of the critical headway can be used by following drivers.

At first sight, our model has a few seeming limitations. In the first place, we assume the critical gap has a discrete distribution with M possible values. In the second place, as will be discussed in great detail in the next section, the analysis technique requires that there is a maximum number of possible attempts: $i \in \{1, 2, \ldots, N\}$. It is important to note, however, that both issues do not have big practical consequences, since one can choose M and N quite large. This does come at the expense of increased computation times, obviously.

3. Queue length analysis

The main objective of this section is to determine the stationary queue length on the minor road at departure epochs of low-priority vehicles. In the next section we will use these results to determine the service-time distribution of the low-priority vehicles, which enables us to evaluate the capacity of the minor road.

3.1. **Preliminaries.** We start by describing the queueing process on the minor road. All arriving vehicles experience three stages before leaving the system: queueing, scanning and merging. The queueing phase is defined as the time between joining the queue and reaching the front of the queue (for queuers this is the moment that the preceding driver accepts a gap and departs). The scanning phase commences when the driver starts scanning for gaps and ends when a sufficiently large gap is found. The scanning phase is followed by the merging phase, which ends at the moment that the vehicle merges into the major stream. Note that phases 1 and 2 may have zero length. The length of phase 1 is called the *delay*. The length of phase 3 for a vehicle of profile r is denoted by the profile-specific constant Δ_r. We define the *service time* as the total time spent in phases 2 and 3 (scanning and merging).

Remark 1. In some papers, an additional move-up phase is distinguished but, depending on the application (freeway or unsignalized intersection), this phase can be incorporated in one of the other phases. See Heidemann and Wegmann [22, Section 3] for more details.

The queue-length analysis is based on observing the system at *departure epochs*. In contrast to the classical M/G2/1 queueing models, in our current model we now have *dependent* service times. This dependence is caused by the fact that the length of the lag (i.e., the remaining gap) left by the previous driver, due to the impatience, now depends on the attempt number *and* the profile of the previous driver. Using the notation introduced in the previous section: we need to know the critical gap $T_{(i,r)}$ of the previous driver of profile r that led to a successful merge/crossing of the intersection. A queueing model with this type of dependencies is called a queueing model with *semi-Markovian* service times, sometimes referred to as the M/SM/1 queue. Although several papers have studied such a queueing model (cf. [1, 6, 11, 15, 19, 25, 26, 27, 30]), we still need to make several adjustments in order to make it applicable to our situation. The analysis below is based on the framework that was recently introduced in [1].

Denote by X_n the queue length at the departure epoch of the n-th vehicle, i.e. the number of vehicles *behind* the n-th driver at the moment that he merges into the major stream. We use $G^{(n)}$ to denote the service time of the n-th vehicle and A_n to denote the number of arrivals (on the minor road) during this service time. Since vehicles arrive in batches (platoons), we denote by B_n the size of the batch in which the n-th vehicle arrived. Let $X(\cdot)$, $A(\cdot)$, and $B(\cdot)$ denote, respectively, the probability generating functions (PGFs) of the limiting distributions of these random variables,

$$X(z) = \lim_{n \to \infty} \mathbb{E}[z^{X_n}], \quad A(z) = \lim_{n \to \infty} \mathbb{E}[z^{A_n}], \quad B(z) = \lim_{n \to \infty} \mathbb{E}[z^{B_n}], \qquad |z| \le 1,$$

whereas $\tilde{G}(\cdot)$ is the limiting Laplace-Stieltjes transform (LST) of $G^{(n)}$:

$$\tilde{G}(s) = \lim_{n \to \infty} \mathbb{E}[e^{-sG^{(n)}}], \qquad s \ge 0.$$

Note that the arrivals constitute a batch Poisson process. Therefore, $A(z)$ can be expressed in terms of $B(z)$ and $\tilde{G}(s)$,

$$(3.1) \qquad\qquad\qquad A(z) = \tilde{G}(\lambda(1 - B(z))),$$

but finding an expression for $\tilde{G}(s)$ is quite tedious. For this reason, we split the analysis in two parts. This section discusses how to find $X(z)$ in terms of $A(z)$, while Section 4 shows how to find $G(z)$ for the gap acceptance model considered in this paper.

The starting-point of our analysis is the following recurrence relation:

$$(3.2) \qquad\qquad X_n = \begin{cases} X_{n-1} - 1 + A_n & \text{if } X_{n-1} \ge 1 \\ A_n + B_n - 1 & \text{if } X_{n-1} = 0 \end{cases}, \quad n = 1, 2, 3, \ldots.$$

As argued before, $\{X_n\}_{n \geq 1}$ is not a Markov chain. In order to obtain a Markovian model, we need to keep track of all the characteristics of the $(n-1)$-th driver. Let $J_n = (i, k, r)$ contain these characteristics of the $(n-1)$-th driver, where i is the 'succeeded attempt number', k indicates which random gap the driver sampled, and r denotes the profile of the driver.

In [4] it is shown that the ergodicity condition of this system is given by

$$(3.3) \qquad \lim_{n \to \infty} \mathbb{E}[A_n | X_{n-1} \geq 1] < 1.$$

Throughout this paper we assume that this condition holds.

3.2. A queueing model with semi-Markovian service times and exceptional first service.

To find the probability generating function (PGF) of the queue-length distribution at departure epochs, we use the framework introduced in [4]. There an analysis is presented of the $\mathrm{M}^X/\mathrm{SM2}/1$ queue, which is a very general type of single-server queueing model with batch arrivals and semi-Markovian service times with exceptional service when the queue is empty, extending earlier results (cf. also [1, 11, 19, 25]) to make it applicable to situations as the one discussed in the present paper. To improve the readability of this paper, we briefly summarize the most important results from [4] that are also valid for our model.

In [4] the type of the n-th customer is denoted by $J_n \in \{1, 2, \ldots, N\}$. In order to apply these results correctly to our model, we need to make one small adjustment. In order to determine the service time of the n-th customer, we need to know exactly how much of the critical gap of the $(n-1)$-th customer remains for this n-th customer. As it turns out in the next section, this requires the following knowledge about the $(n-1)$-th customer:

(A) we need to know the sampled critical gap $u_{(i,k,r)}$ of the $(n-1)$-th customer;

(B) we need to know whether the $(n-1)$-th customer emptied the queue at the minor road and the n-th customer was the first in a new batch arriving some time after the departure of his predecessor.

To start with the latter: a consequence of (B) is that we need a queueing model with so-called *exceptional first service*. As a consequence of (A), we use the triplet (i, k, r), extensively discussed in the previous subsection, to denote the 'customer type', which should be interpreted as a vehicle of profile r that accepted the i-th gap, where it sampled $u_{(i,k,r)}$ as its critical headway. It means that if the n-th customer is of type (i, k, r), then his *predecessor* (which is the $(n-1)$-th customer) succeeded in his i-th attempt, while being of profile r and having critical gap $u_{(i,k,r)}$.

A translation from our model to the model in [4] can easily be made by mapping our $\bar{N} := NMR$ customer types onto their N customer types (for instance, if $j = (i-1)MR + (k-1)R + r$, then $j \in \{1, 2, \ldots, \bar{N}\}$ can serve as the customer type in [4]).

Mimicking the steps in [4], it is immediate that

$$(3.4) \quad \mathbb{E}[z^{X_n} 1_{\{J_{n+1}=(j,l,r_1)\}}]$$

$$= \frac{1}{z} \sum_{r_0=1}^{R} \sum_{k=1}^{M} \sum_{i=1}^{N} \mathbb{E}[z^{X_{n-1}} 1_{\{J_n=(i,k,r_0)\}}] \mathbb{E}[z^{A_n} 1_{\{J_{n+1}=(j,l,r_1)\}} | J_n = (i,k,r_0), X_{n-1} \geq 1] +$$

$$\frac{1}{z} \sum_{r_0=1}^{R} \sum_{k=1}^{M} \sum_{i=1}^{N} \Big(B(z) \mathbb{E}[z^{A_n} 1_{\{J_{n+1}=(j,l,r_1)\}} | J_n = (i,k,r_0), X_{n-1} = 0] -$$

$$\mathbb{E}[z^{A_n} 1_{\{J_{n+1}=(j,l,r_1)\}} | J_n = (i,k,r_0), X_{n-1} \geq 1] \Big) \mathbb{P}(X_{n-1} = 0, J_n = (i,k,r_0)),$$

for $n \in \mathbb{N}$, $j \in \{1, 2, \ldots, N\}$, $l \in \{1, 2, \ldots, M\}$, and $r_1 \in \{1, 2, \ldots, R\}$.

To further evaluate these expressions, we need to introduce some additional notation. Define, for $i, j \in \{1, 2, \ldots, N\}$ $k, l \in \{1, 2, \ldots, M\}$, $r_0, r_1 \in \{1, 2, \ldots, R\}$, and $|z| \leq 1$,

$$A_{(i,k,r_0)}^{(j,l,r_1)}(z) = \lim_{n\to\infty} \mathbb{E}[z^{A_n} 1_{\{J_{n+1}=(j,l,r_1)\}} | J_n = (i, k, r_0), X_{n-1} \geq 1],$$

$$A_{(i,k,r_0)}^{*(j,l,r_1)}(z) = \lim_{n\to\infty} \mathbb{E}[z^{A_n} 1_{\{J_{n+1}=(j,l,r_1)\}} | J_n = (i, k, r_0), X_{n-1} = 0],$$

where A^* corresponds to the number of arrivals during the (exceptional) service of a vehicle arriving when there is no queue in front of him. In addition,

$$(3.5) \qquad\qquad f_{(j,l,r_1)}(z) = \lim_{n\to\infty} \mathbb{E}[z^{X_n} 1_{\{J_{n+1}=(j,l,r_1)\}}],$$

such that

$$(3.6) \qquad\qquad f_{(j,l,r_1)}(0) = \lim_{n\to\infty} \mathbb{P}(X_n = 0, J_{n+1} = (j, l, r_1)).$$

The above definitions entail that

$$(3.7) \qquad\qquad X(z) = \sum_{r_1=1}^{R} \sum_{l=1}^{M} \sum_{j=1}^{N} f_{(j,l,r_1)}(z).$$

In steady state, Equation (3.4) thus leads to the following system of $\bar{N}$ linear equations: for any j, l, and r_1,

$$(3.8) \qquad \left(z - A_{(j,l,r_1)}^{(j,l,r_1)}(z)\right) f_{(j,l,r_1)}(z) - \sum_{\substack{r_0=1,\\ r_0\neq r_1}}^{R} A_{(j,l,r_0)}^{(j,l,r_1)}(z) f_{(j,l,r_0)}(z)$$

$$- \sum_{r_0=1}^{R} \sum_{\substack{k=1,\\ k\neq l}}^{M} A_{(j,k,r_0)}^{(j,l,r_1)}(z) f_{(j,k,r_0)}(z) - \sum_{r_0=1}^{R} \sum_{k=1}^{M} \sum_{\substack{i=1,\\ i\neq j}}^{N} A_{(i,k,r_0)}^{(j,l,r_1)}(z) f_{(i,k,r_0)}(z)$$

$$= \sum_{r_0=1}^{R} \sum_{k=1}^{M} \sum_{i=1}^{N} (B(z) A_{(i,k,r_0)}^{*(j,l,r_1)}(z) - A_{(i,k,r_0)}^{(j,l,r_1)}(z)) f_{(i,k,r_0)}(0).$$

From the above we conclude that when we know $A_{(i,k,r_0)}^{(j,l,r_1)}(z)$ and $A_{(i,k,r_0)}^{*(j,l,r_1)}(z)$, the PGFs of our interest can be evaluated. In the next section we show how to find the Laplace-Stieltjes transforms of the conditional service-time distributions, which lead to $A_{(i,k,r_0)}^{(j,l,r_1)}(z)$ and $A_{(i,k,r_0)}^{*(j,l,r_1)}(z)$.

We refer to [4, Section 3] for more details on how to write the linear system of equations (3.8) in a convenient matrix form and solve it numerically. In [4], and in more detail in [1], it is also discussed how the PGF of the queue-length at *arbitrary epochs* (denoted by X^{arb}) can be found; in particular the clean relation between X^{arb} and X:

$$(3.9) \qquad\qquad \mathbb{E}[z^{X^{\mathrm{arb}}}] = \mathbb{E}[z^X] \frac{\mathbb{E}[B](1-z)}{1-B(z)}$$

is useful in this context. The interpretation is that the number of customers at an *arbitrary epoch* is equal to the number of customers left behind by an arbitrary *departing* customer, excluding those that arrived in the same batch as this departing customer. We refer to [1] for more details and a proof of (3.9).

4. Capacity

In this section we derive the Laplace-Stieltjes transform of the service-time distribution of our gap acceptance model. This service-time analysis serves two purposes. Firstly, it is used to complete the queue-length analysis of the previous section. Secondly, it is an important result by itself, because it is directly linked to the capacity of the intersection, which is defined as the maximum possible number of vehicles per time unit that can pass from the minor road. An alternative but equivalent definition of capacity, also employed by Heidemann and Wegmann [22], is the maximum arrival rate for which the corresponding queue remains stable. Combining (3.1) and (3.3), and denoting

$$g := \lim_{n \to \infty} \mathbb{E}[G^{(n)} \mid X_{n-1} \geq 1],$$

we can rewrite the stability condition as

$$(4.1) \qquad \qquad \lambda \mathbb{E}[B] \, g < 1,$$

where λ is the arrival rate of batches, implying that the arrival rate of individual vehicles is $\lambda \mathbb{E}[B]$. As a consequence, the capacity of the minor road is given by

$$(4.2) \qquad \qquad C = \frac{1}{g}.$$

Every driver (of profile r) samples a random $T_{(i,r)}$ at his i-th attempt, irrespective of his previous attempts. However, he uses only a constant Δ_r of the gap to cross the main road. As a consequence, when accepting the gap, the remaining part $(T_{(i,r)} - \Delta_r)$ can be used by subsequent drivers. In order to keep the analysis tractable, we need the following assumption.

Assumption 1 (Limited gap reusability assumption). We assume that *at most one* driver can reuse the remaining part of a critical gap accepted by his predecessor. The means that $(T_{(i,r)} - \Delta_r)$ should *not* be large enough for more than one succeeding vehicle to use it for his own critical headway.

To avoid confusion, we stress that this assumption does *not* mean that a large gap between successive vehicles on the major road cannot be used by more than one vehicle from the minor road. It *does* mean, however, that the difference between *critical* gaps of two successive vehicles cannot be so large that the second vehicle's critical gap is completely contained in its predecessor's critical gap (excluding the merging time).

Remark 2. This assumption seems to be realistic in most, but not all, cases, because one can envision for instance situations in which (parts of) a critical gap accepted by, say, a slow truck might be reused by two fast accelerating vehicles following this truck. It is noted, however, that if Assumption 1 is violated, despite our method no longer being exact, the computed capacity is still very close to the real, simulated capacity; numerical evidence of this property is provided in Section 5. It implies that our method can still be used as a very accurate approximation in those (rare) cases where Assumption 1 is violated.

We recall that there is a complicated dependence structure between service times of two successive vehicles, due to our assumption that a vehicle can use part of the gap left behind by its predecessor *combined* with the assumption that drivers have different profiles and tend to become impatient. This obstacle can be overcome by analyzing the service time of a vehicle of type (j, l, r_1) in the situation that its predecessor was of type (i, k, r_0).

We proceed by introducing some notation. Let E_y be the event that there is a predecessor gap available of length y. Denote by $\pi_{(i,k,r_0)}$ the probability that an arbitrary driver with profile r_0 succeeds in his i-th attempt with critical headway $u_{(i,k,r_0)}$. In addition, let

$$\text{(4.3)} \qquad \tilde{G}_{(j,l,r_1)}(s,y) = \mathbb{E}\left[e^{-sG^{(n)}}1_{\{J_{n+1}=(j,l,r_1)\}} \mid E_y\right],$$

for $0 \le y \le u_{(i,k,r_0)} - \Delta_{r_0}$, $j = 1,2,\dots,N$, and $l = 1,2,\dots,M$. The LST of an arbitrary service time can be found by conditioning on the type of the current vehicle and its predecessor:

$$\text{(4.4)} \qquad \tilde{G}(s) = \mathbb{E}[e^{-sG^{(n)}}] = \sum_{r_0=1}^{R}\sum_{r_1=1}^{R}\sum_{j=1}^{N}\sum_{l=1}^{M}\sum_{i=1}^{N}\sum_{k=1}^{M} \tilde{G}_{(i,k,r_0)}^{(j,l,r_1)}(s)\mathbb{P}(J_n=(i,k,r_0)),$$

where

$$\text{(4.5)} \qquad \tilde{G}_{(i,k,r_0)}^{(j,l,r_1)}(s) = \mathbb{E}\left[e^{-sG^{(n)}}1_{\{J_{n+1}=(j,l,r_1)\}}|J_n=(i,k,r_0)\right],$$

and

$$\text{(4.6)} \qquad \mathbb{P}(J_n=(i,k,r_0)) = \pi_{(i,k,r_0)}p_{(i,k,r_0)}p_{r_0}.$$

We are now ready to state the main result of this paper. Define

$$\bar{f}_{(i,k,r)} := \frac{f_{(i,k,r)}(0)}{\pi_{(i,k,r)}p_{(i,k,r)}p_r}, \quad \bar{u}_{(i,k,r)} := u_{(i,k,r)} - \Delta_r.$$

Theorem 1. *The partial service-time LST of a customer, jointly with the event that he is of type (j,l,r_1), given that his predecessor was of type (i,k,r_0), is given by*

$$\text{(4.7)} \qquad \tilde{G}_{(i,k,r_0)}^{(j,l,r_1)}(s) = \tilde{G}_{(j,l,r_1)}(s,\bar{u}_{(i,k,r_0)})\left(1 - \bar{f}_{(i,k,r_0)}\right) +$$

$$\left(\int_{x=0}^{\bar{u}_{(i,k,r_0)}} \lambda e^{-\lambda x}\tilde{G}_{(j,l,r_1)}(s,\bar{u}_{(i,k,r_0)}-x)\mathrm{d}x + \tilde{G}_{(j,l,r_1)}(s,0)e^{-\lambda\bar{u}_{(i,k,r_0)}}\right)\bar{f}_{(i,k,r_0)},$$

where

- $f_{(i,k,r_0)}(0)$ *is obtained by solving the system of equations* (3.8), *and*
- *expressions for $\tilde{G}_{(j,l,r_1)}(s,y)$ and $\pi_{(i,k,r_0)}$ are provided in Lemmas 1 and 2 below.*

Proof This partial LST $\tilde{G}_{(i,k,r_0)}^{(j,l,r_1)}(s)$ can be expressed in terms of $\tilde{G}_{(j,l,r_1)}(s,y)$ by first distinguishing between the case where there was no queue upon arrival of the n-th customer ($X_{n-1}=0$), and the case where there was a queue ($X_{n-1} \ge 1$). If the customer arrives in an empty system at time t and the previous arrival took place (merged on the major road) at time $t-x$, then we need to check whether the remaining gap $u_{(i,k,r_0)} - \Delta_{r_0} - x = \bar{u}_{(i,k,r_0)} - x$ is greater than zero, or not. If the system was not empty at arrival time, the lag is simply $\bar{u}_{(i,k,r_0)}$ due to Assumption 1 (we provide more details on this in Remark 3). This, combined with (3.6) and (4.6), yields

$$\tilde{G}_{(i,k,r_0)}^{(j,l,r_1)}(s) = \lim_{n\to\infty}\mathbb{E}\left[e^{-sG^{(n)}}1_{\{J_{n+1}=(j,l,r_1)\}}|X_{n-1}\ge 1, J_n=(i,k,r_0)\right]\mathbb{P}(X_{n-1}\ge 1|J_n=(i,k,r_0)) +$$

$$\lim_{n\to\infty}\mathbb{E}\left[e^{-sG^{(n)}}1_{\{J_{n+1}=(j,l,r_1)\}}|X_{n-1}=0, J_n=(i,k,r_0)\right]\mathbb{P}(X_{n-1}=0|J_n=(i,k,r_0))$$

$$= \mathbb{E}\left[e^{-sG^{(n)}}1_{\{J_{n+1}=(j,l,r_1)\}}|E_{(u_{(i,k,r_0)}-\Delta_{r_0})}\right]\left(1 - \bar{f}_{(i,k,r_0)}\right) +$$

$$\left(\int_{x=0}^{\infty}\lambda e^{-\lambda x}\mathbb{E}\left[e^{-sG^{(n)}}1_{\{J_{n+1}=(j,l,r_1)\}}|E_{\max(u_{(i,k,r_0)}-\Delta_{r_0}-x,0)}\right]\mathrm{d}x\right)\bar{f}_{(i,k,r_0)},$$

which can easily be rewritten to (4.7).   □

It thus remains to find expressions for $\tilde{G}_{(j,l,r_1)}(s,y)$ and $\pi_{(i,k,r_0)}$, which we do in the following two lemmas. We denote by $E_r^{(n)}$ the event that the n-th driver does not succeed in his first attempt, given that he has profile r.

Lemma 1. *The conditional service-time LST of a vehicle, given the event E_y, is given by*

$$\tilde{G}_{(1,l,r)}(s,y) = p_r p_{(1,l,r)} e^{-q\max\{u_{(1,l,r)}-y,0\}} e^{-s\Delta_r},$$

$$\tilde{G}_{(j,l,r)}(s,y) = p_r p_{(j,l,r)} \left(1 - \mathbb{E}\left[e^{-(s+q)\max\{T_{(1,r)}-y,0\}} \right] \right)$$

$$\times e^{-(s(y+\Delta_r)+qu_{(j,l,r)})} \left(\frac{q}{s+q} \right)^{j-1} \prod_{m=2}^{j-1} \left(1 - \mathbb{E}\left[e^{-(s+q)T_{(m,r)}} \right] \right),$$

for $j = 2, 3, \ldots, N$.

Proof For $j = 1$, the expression $\tilde{G}_{(j,l,r)}(s,y)$ simply follows from computing the probability that a customer is of type r *and* merges successfully during its first attempt with critical gap $u_{(1,l,r)}$, for $l = 1, 2, \ldots, M$, and faces a gap that is greater than $u_{(1,l,r)} - y$. In this case, the customer's service time is Δ_r. For the case $j \geq 2$ we find

$$\tilde{G}_{(j,l,r)}(s,y) = p_r \sum_{l_1=1}^{M} p_{(1,l_1,r)} 1_{\{u_{(1,l_1,r)}>y\}} \int_0^{u_{(1,l_1,r)}-y} qe^{-qt} e^{-s(y+t)} \mathbb{E}\left[e^{-sG^{(n)}} 1_{\{J_{n+1}=(j,l,r)\}} | E_r^{(n)} \right] \mathrm{d}t$$

$$= p_r \sum_{l_1=1}^{M} p_{(1,l_1,r)} 1_{\{u_{(1,l_1,r)}>y\}} \frac{qe^{-sy}}{s+q} (1 - e^{-(s+q)(u_{(1,l_1,r)}-y)}) \mathbb{E}\left[e^{-sG^{(n)}} 1_{\{J_{n+1}=(j,l,r)\}} | E_r^{(n)} \right]$$

$$= p_r \sum_{l_1=1}^{M} p_{(1,l_1,r)} \frac{qe^{-sy}}{s+q} (1 - e^{-(s+q)\max\{u_{(1,l_1,r)}-y,0\}}) \mathbb{E}\left[e^{-sG^{(n)}} 1_{\{J_{n+1}=(j,l,r)\}} | E_r^{(n)} \right]$$

$$= p_r \frac{qe^{-sy}}{s+q} \left(1 - \mathbb{E}\left[e^{-(s+q)\max\{T_{(1,r)}-y,0\}} \right] \right) \mathbb{E}\left[e^{-sG^{(n)}} 1_{\{J_{n+1}=(j,l,r)\}} | E_r^{(n)} \right];$$

here it should be kept in mind that

$$\mathbb{E}\left[e^{-sG^{(n)}} 1_{\{J_{n+1}=(j,l,r)\}} | E_r^{(n)} \right]$$

does not depend on n and can be rewritten as

$$\left(\prod_{m=2}^{j-1} \sum_{k=1}^{M} p_{(m,k,r)} \int_0^{u_{(m,k,r)}} qe^{-qt} e^{-st} \mathrm{d}t \right) p_{(j,l,r)} \int_{u_{(j,l,r)}}^{\infty} qe^{-qt} e^{-s\Delta_r} \mathrm{d}t$$

$$= p_{(j,l,r)} e^{-(s\Delta_r+qu_{(j,l,r)})} \left(\frac{q}{s+q} \right)^{j-2} \prod_{m=2}^{j-1} \sum_{k=1}^{M} p_{(m,k,r)} (1 - e^{-(s+q)u_{(m,k,r)}})$$

$$= p_{(j,l,r)} e^{-(s\Delta_r+qu_{(j,l,r)})} \left(\frac{q}{s+q} \right)^{j-2} \prod_{m=2}^{j-1} \left(1 - \sum_{k=1}^{M} p_{(m,k,r)} e^{-(s+q)u_{(m,k,r)}} \right)$$

$$= p_{(j,l,r)} e^{-(s\Delta_r+qu_{(j,l,r)})} \left(\frac{q}{s+q} \right)^{j-2} \prod_{m=2}^{j-1} \left(1 - \mathbb{E}\left[e^{-(s+q)T_{(m,r)}} \right] \right).$$

Slightly rewriting this last expression completes the proof of Lemma 1. $\qquad\square$

The next lemma shows how to compute the probabilities $\pi_{(i,k,r)}$. We denote by τ_q the generic interarrival time between two vehicles on the *major* road, which is exponentially distributed with parameter q. Let $\hat{v}$ be defined as $\max\{v,0\}$, and $\check{v}$ as $\max\{-v,0\}$. In addition, $v_{(i,l,r_0)}^{(1,k,r)} = u_{(1,k,r)} - u_{(i,l,r_0)} + \Delta_{r_0}$.

Lemma 2. *The probability that a driver of profile r is served in his i-th attempt with critical headway $u_{(i,k,r)}$ is given by*

$$(4.8) \qquad \pi_{(i,k,r)} = P_{(i,k,r)} \prod_{m=1}^{i-1} \left(1 - \sum_{k_m=1}^{M} p_{(m,k_m,r)} P_{(m,k_m,r)} \right),$$

where, for $i \in \{2, 3, \ldots, N\}$ and $k \in \{1, 2, \ldots, M\}$,

$$(4.9) \qquad P_{(i,k,r)} = \mathbb{P}(\tau_q \geq u_{(i,k,r)}) = e^{-q u_{(i,k,r)}}.$$

The remaining $P_{(1,k,r)}$ can be found by solving the system of linear equations

$$(4.10) \qquad \left(1 + p_{(1,k,r)} \left(-p_r e^{-q\hat{v}_{(1,k,r)}^{(1,k,r)}} + c_{(k,r,r)} \right) \right) P_{(1,k,r)}$$

$$+ \sum_{\substack{r_0=1, \\ r_0 \neq r}}^{R} \left(-p_{r_0} e^{-q\hat{v}_{(1,k,r_0)}^{(1,k,r)}} + c_{(k,r_0,r)} \right) p_{(1,k,r_0)} P_{(1,k,r_0)}$$

$$+ \sum_{r_0=1}^{R} \sum_{\substack{l_1=1, \\ l_1 \neq k}}^{M} \left(-p_{r_0} e^{-q\hat{v}_{(1,l_1,r_0)}^{(1,k,r)}} + c_{(k,r_0,r)} \right) p_{(1,l_1,r_0)} P_{(1,l_1,r_0)}$$

$$= \sum_{r_0=1}^{R} \sum_{l=1}^{M} \sum_{i=1}^{N} \left(1 - e^{-\lambda \check{v}_{(i,l,r_0)}^{(1,k,r)}} - e^{-q\hat{v}_{(i,l,r_0)}^{(1,k,r)}} + \frac{\lambda}{\lambda+q} e^{-(\lambda+q)\check{v}_{(i,l,r_0)}^{(1,k,r)} - q v_{(i,l,r_0)}^{(1,k,r)}} \right.$$

$$\left. + \frac{q}{\lambda+q} e^{-\lambda \bar{u}_{(i,l,r_0)} - q u_{(1,k,r)}} \right) f_{(i,l,r_0)}(0) + \sum_{r_0=1}^{R} c_{(k,r_0,r)},$$

for $k \in \{1, 2, \ldots, M\}$ and $r \in \{1, 2, \ldots, R\}$, where

$$c_{(k,r_0,r)} = \sum_{l=1}^{M} \sum_{i=2}^{N} p_{r_0} p_{(i,l,r_0)} e^{-q(u_{(i,l,r_0)} + \hat{v}_{(i,l,r_0)}^{(1,k,r)})} \prod_{m=2}^{i-1} \left(1 - \mathbb{E}[e^{-qT_{(m,r_0)}}] \right).$$

Proof Expression (4.8) simply follows from the definition of $\pi_{(i,k,r)}$, by multiplying the probabilities that the driver does *not* succeed in attempts $1, 2, \ldots, i-1$ (each time distinguishing between all M possible random critical gap values) and finally multiplying with the probability that he succeeds in the i-th attempt with critical headway $u_{(i,k,r)}$.

For $i \in \{2, 3, \ldots, N\}$ it is easy to determine $P_{(i,k,r)}$ because we do not have to take into account any gap left by the predecessor; we simply compute the probability that the critical gap $u_{(i,k,r)}$ is smaller than the remaining interarrival time τ_q (which, due to the memoryless property, is again exponentially distributed with rate q).

The case $i = 1$ is considerably more complicated. Using (4.6) and noting that $\pi_{(1,k,r)} = P_{(1,k,r)}$, we can find a system of equations for $P_{(1,k,r)}$ by conditioning on the type of the *predecessor* of the vehicle under consideration. Define

$$\chi_{(i,l,r_0)}^{(j,k,r)} := \mathbb{P}(J_n = (j,k,r) | J_{n-1} = (i,l,r_0)),$$

and

$$(4.11) \qquad \phi_{(i,l,r_0)}^{(j,k,r)} := \mathbb{P}(J_n = (j,k,r) | J_{n-1} = (i,l,r_0), X_{n-2} = 0),$$

$$(4.12) \qquad \psi_{(i,l,r_0)}^{(j,k,r)} := \mathbb{P}(J_n = (j,k,r) | J_{n-1} = (i,l,r_0), X_{n-2} \geq 1).$$

Evidently,

$$(4.13) \qquad P_{(1,k,r)} = \frac{1}{p_{(1,k,r)}p_r} \mathbb{P}(J_n = (1,k,r))$$

$$= \frac{1}{p_{(1,k,r)}p_r} \sum_{r_0=1}^{R} \sum_{l=1}^{M} \sum_{i=1}^{N} \chi_{(i,l,r_0)}^{(1,k,r)} \mathbb{P}(J_{n-1} = (i,l,r_0))$$

$$= \frac{1}{p_{(1,k,r)}p_r} \sum_{r_0=1}^{R} \sum_{l=1}^{M} \sum_{i=1}^{N} \chi_{(i,l,r_0)}^{(1,k,r)} \pi_{(i,l,r_0)} p_{(i,l,r_0)} p_{r_0}$$

$$= \frac{1}{p_{(1,k,r)}p_r} \sum_{r_0=1}^{R} \sum_{l=1}^{M} \sum_{i=1}^{N} \left(\phi_{(i,l,r_0)}^{(1,k,r)} \mathbb{P}(X_{n-2} = 0 | J_{n-1} = (i,l,r_0)) + \right.$$

$$\left. \psi_{(i,l,r_0)}^{(1,k,r)} \mathbb{P}(X_{n-2} \geq 1 | J_{n-1} = (i,l,r_0)) \right) \pi_{(i,l,r_0)} p_{(i,l,r_0)} p_{r_0},$$

which can be further evaluated, using notation introduced before, as

$$\frac{1}{p_{(1,k,r)}p_r} \sum_{r_0=1}^{R} \sum_{l=1}^{M} \sum_{i=1}^{N} \left(\phi_{(i,l,r_0)}^{(1,k,r)} \bar{f}_{(i,l,r_0)} + \psi_{(i,l,r_0)}^{(1,k,r)} \left(1 - \bar{f}_{(i,l,r_0)} \right) \right) \bar{\pi}_{(i,l,r_0)},$$

with $\bar{\pi}_{(i,l,r_0)} := \pi_{(i,l,r_0)} p_{(i,l,r_0)} p_{r_0}$. We now consider the individual terms separately. Observe that, by conditioning on the value of τ_q,

$$\frac{1}{p_{(1,k,r)}p_r} \phi_{(i,l,r_0)}^{(1,k,r)} \bar{f}_{(i,l,r_0)} \bar{\pi}_{(i,l,r_0)} = \left(\int_0^{\max\{\bar{u}_{(i,l,r_0)} - u_{(1,k,r)}, 0\}} \lambda e^{-\lambda x} \mathrm{d}x \right.$$

$$+ \int_{\max\{\bar{u}_{(i,l,r_0)} - u_{(1,k,r)}, 0\}}^{\bar{u}_{(i,l,r_0)}} \mathbb{P}(\tau_q \geq u_{(1,k,r)} - \bar{u}_{(i,l,r_0)} + x) \lambda e^{-\lambda x} \mathrm{d}x$$

$$\left. + \int_{\bar{u}_{(i,l,r_0)}}^{\infty} \mathbb{P}(\tau_q \geq u_{(1,k,r)}) \lambda e^{-\lambda x} \mathrm{d}x \right) f_{(i,l,r_0)}(0).$$

Also,

$$\frac{1}{p_{(1,k,r)}p_r} \psi_{(i,l,r_0)}^{(1,k,r)} (1 - \bar{f}_{(i,l,r_0)}) \bar{\pi}_{(i,l,r_0)}$$

$$= \mathbb{P} \left(\tau_q \geq \max\{u_{(1,k,r)} - \bar{u}_{(i,l,r_0)}, 0\} \right) \left(\pi_{(i,l,r_0)} p_{(i,l,r_0)} p_{r_0} - f_{(i,l,r_0)}(0) \right).$$

Combining the above, it follows that $P_{(1,k,r)}$ equals, with $v_{(i,l,r_0)}^{(1,k,r)}$ as defined above,

$$\sum_{r_0=1}^{R} \sum_{l=1}^{M} \sum_{i=1}^{N} \left(1 - e^{-\lambda \check{v}_{(i,l,r_0)}^{(1,k,r)}} - e^{-q\hat{v}_{(i,l,r_0)}^{(1,k,r)}} + \frac{\lambda}{\lambda+q} e^{-(\lambda+q)\check{v}_{(i,l,r_0)}^{(1,k,r)} - qv_{(i,l,r_0)}^{(1,k,r)}} \right.$$

$$\left. + \frac{q}{\lambda+q} e^{-\lambda(u_{(i,l,r_0)} - \Delta_{r_0}) - qu_{(1,k,r)}} \right) f_{(i,l,r_0)}(0) + \sum_{r_0=1}^{R} \sum_{l=1}^{M} \sum_{i=1}^{N} e^{-q\hat{v}_{(i,l,r_0)}^{(1,k,r)}} \pi_{(i,l,r_0)} p_{(i,l,r_0)} p_{r_0}.$$

These expressions can be written in a more convenient form. To this end, it is noted that

$$\sum_{r_0=1}^{R} \sum_{l=1}^{M} \sum_{i=1}^{N} e^{-q\hat{v}_{(i,l,r_0)}^{(1,k,r)}} \pi_{(i,l,r_0)} p_{(i,l,r_0)} p_{r_0}$$

$$= \sum_{r_0=1}^{R} \sum_{l=1}^{M} e^{-q\hat{v}_{(1,l,r_0)}^{(1,k,r)}} P_{(1,l,r_0)} p_{(1,l,r_0)} p_{r_0} + \sum_{r_0=1}^{R} \sum_{l=1}^{M} \sum_{i=2}^{N} e^{-q\hat{v}_{(i,l,r_0)}^{(1,k,r)}} \pi_{(i,l,r_0)} p_{(i,l,r_0)} p_{r_0}$$

$$= \sum_{r_0=1}^{R} \sum_{l=1}^{M} e^{-q\hat{v}_{(1,l,r_0)}^{(1,k,r)}} P_{(1,l,r_0)} p_{(1,l,r_0)} p_{r_0}$$

$$+ \sum_{r_0=1}^{R} \sum_{l=1}^{M} \sum_{i=2}^{N} e^{-q\hat{v}_{(i,l,r_0)}^{(1,k,r)}} p_{(i,l,r_0)} p_{r_0} P_{(i,l,r_0)} \prod_{m=1}^{i-1} \left(1 - \sum_{l_m=1}^{M} p_{(m,l_m,r_0)} P_{(m,l_m,r_0)} \right).$$

After some basic algebraic manipulations, and using (4.9) we obtain

(4.14)

$$\sum_{r_0=1}^{R} \sum_{l=1}^{M} \sum_{i=1}^{N} e^{-q\hat{v}_{(i,l,r_0)}^{(1,k,r)}} \pi_{(i,l,r_0)} p_{(i,l,r_0)} p_{r_0} = \sum_{r_0=1}^{R} \sum_{l_1=1}^{M} \left(e^{-q\hat{v}_{(1,l_1,r_0)}^{(1,k,r)}} \right.$$

$$- \sum_{l=1}^{M} \sum_{i=2}^{N} p_{(i,l,r_0)} e^{-q(u_{(i,l,r_0)}+\hat{v}_{(i,l,r_0)}^{(1,k,r)})} \prod_{m=2}^{i-1} \left(1 - \sum_{l_m=1}^{M} p_{(m,l_m,r_0)} e^{-qu_{(m,l_m,r_0)}} \right) \right) P_{(1,l_1,r_0)} p_{(1,l_1,r_0)} p_{r_0}$$

$$+ \sum_{r_0=1}^{R} \sum_{l=1}^{M} \sum_{i=2}^{N} p_{r_0} p_{(i,l,r_0)} e^{-q(u_{(i,l,r_0)}+\hat{v}_{(i,l,r_0)}^{(1,k,r)})} \prod_{m=2}^{i-1} \left(1 - \sum_{l_m=1}^{M} p_{(m,l_m,r_0)} e^{-qu_{(m,l_m,r_0)}} \right).$$

Using Equation (4.14) in (4.13), and after simplification, we obtain the linear system of equations (4.10), which proves Lemma 2.   □

Remark 3. We now discuss in more detail Assumption 1, and how it plays a role in our result. The assumption entails that the gap $T_{(i,r_0)} - \Delta_{r_0}$ left by a driver of profile r_0, while being successful in the i-th attempt, can be used by the subsequent driver only. If we would allow more than one following vehicle use a gap left behind by a merging vehicle, the analysis becomes much more complicated due to the fact that we need to distinguish between all possible cases where multiple drivers fit into this lag.

For example, in the proof of Theorem 1 we use explicitly that, if the system was not empty at arrival time, the gap left behind by a predecessor of type (j,l,r_1) is simply $\bar{u}_{(j,l,r_1)} = u_{(j,l,r_1)} - \Delta_{r_1}$. However, a situation where this is not necessarily true, occurs when $j = 1$ (meaning that the predecessor succeeded at his first attempt) and required a critical gap of $u_{(1,l,r_1)}$ that is *smaller* than the remaining gap of the *predecessor's predecessor*, say $\bar{u}_{(i,k,r_0)}$. In this case, the gap left behind by the predecessor is equal to

$$u_{(i,k,r_0)} - \Delta_{r_0} - \Delta_{r_1} > u_{(1,l,r_1)} - \Delta_{r_1},$$

meaning that the current vehicle and its predecessor both used the same gap $u_{(i,k,r_0)}$. In this sense, Assumption 1 ensures the tractability of the model.

Assumption 1 has not been stated in terms of the model input parameters, and is therefore not straightforward to check. To remedy this, we give an equivalent definition in terms of the critical headway and the merging time. Assumption 1 is satisfied if and only if

(4.15) $$u_{(1,l,r_1)} \geq \bar{u}_{(i,k,r_0)} = u_{(i,k,r_0)} - \Delta_{r_0},$$

for all $i \in \{1, 2, \ldots, N\}$, $k, l \in \{1, 2, \ldots, M\}$, and $r_0, r_1 \in \{1, 2, \ldots, R\}$. In the next section, we will numerically show that in cases that Assumption 1 is not fulfilled, the estimated capacity of the minor road is slightly smaller, but still highly accurate. The reason why the estimated capacity is a *lower* bound for the true capacity, is the fact that we waste some capacity by not allowing more than one vehicle to use the remaining part of a critical gap. The astute reader might have noticed that the maximum operators in Lemma 1 are not really needed due to condition (4.15). It turns out, however, that by preventing the corresponding terms from becoming negative, the capacity is approximated much better even when the condition is violated.

5. Numerical results

The analysis from the previous sections facilitates the evaluation of the performance of the system, including the assessment of the sensitivity of the capacity when varying model parameters. In this section, we present numerical examples to demonstrate the impact of these model parameters and of different driver behavior.

5.1. **Example 1:** In this illustrative example, we distinguish between two driver profiles. Profile 1 represents 'standard' traffic, whereas profile 2 can be considered as 'slower' traffic (for example large, heavily loaded vehicles). We assume that the ratio between profile 1 and profile 2 vehicles is 90%/10%. The fact that profile 1 vehicles are faster than profile 2 vehicles is captured in their merging times (respectively Δ_1 and Δ_2) and in the length of their critical gaps. From the profile 1 drivers, we assume that 40% need a gap of at least 5 seconds, upon arrival at the intersection, while the remaining 60% need a gap of at least 6 seconds. If, however, the drivers do not find an acceptable gap right away, they will grow impatient and be more and more prepared to accept slightly smaller gaps. We introduce an impatience rate $\alpha \in (0, 1)$ that determines how fast critical gaps decrease. Profile 2 vehicles are assumed to be slower, meaning that they have a need critical gaps of respectively 8 seconds (50%) or 9 seconds (50%) at their first attempt. Summarizing, we have the following model parameters in this example:

Profile 1	Profile 2
$p_1 = 0.9$	$p_2 = 0.1$
$p_{i,1,1} = 0.4$	$p_{i,1,2} = 0.5$
$p_{i,2,1} = 0.6$	$p_{i,2,2} = 0.5$
$u_{1,1,1} = 5.0$	$u_{1,1,2} = 8.0$
$u_{1,2,1} = 6.0$	$u_{1,2,2} = 9.0$

and, due to the impatience, we have for both profiles:

$$(5.1) \qquad u_{(i+1,k,r)} = \alpha(u_{(i,k,r)} - \Delta_r) + \Delta_r, \quad i = 1, 2, \ldots, N-1, \ k = 1, 2, \ r = 1, 2.$$

We vary α, Δ_1 and Δ_2 to gain insight into the impact of these parameters on the capacity of the minor road, while also varying q, the arrival rate on the major road. We take $\alpha \in \{0.6, 0.8, 0.9, 1.0\}$, $\Delta_1 \in \{4, 5\}$ and $\Delta_2 \in \{5, 6, 7\}$ and vary q between 0 and 1000 vehicles per hour. Note that $\alpha = 1$ corresponds to a model *without* impatience. The results are depicted in Figures 2(a) and 2(b). In Figure 2(a), we can observe how the capacity of the minor road increases when merging time corresponding to at least one of the driver profiles decreases. From Figure 2(b) we conclude by how much the capacity of the minor road increases when drivers become more impatient. The biggest capacity gain is caused by a decrease in Δ_1 from 5 to 4, obviously because profile 1 constitutes the vast majority of all vehicles.

5.2. **Example 2:** The model parameters in Example 1 are such that condition (4.15) is met. In practice, however, situations might arise where this is not the case. The purpose of this example is to quantify errors made when computing the capacity using the methods proposed in Section 4, since these methods rely on condition (4.15). We take the same settings as in Example 1, but we adjust the critical gaps for profile 2 drivers from 8 and 9 to, respectively, 10 and 12 seconds. We keep the original merging times for both profile drivers of $\Delta_1 = 4$ seconds and $\Delta_2 = 5$ seconds. In order to quantify the errors, we use both the analysis from Section 4 and a computer program that simulates the model and gives accurate results. Table 1 compares the capacities obtained with

(a) Impact of merging times of drivers for $\alpha = 0.9$.

(b) Impact of impatient behaviors of drivers for $\Delta_1 = 4$ seconds and $\Delta_2 = 5$ seconds
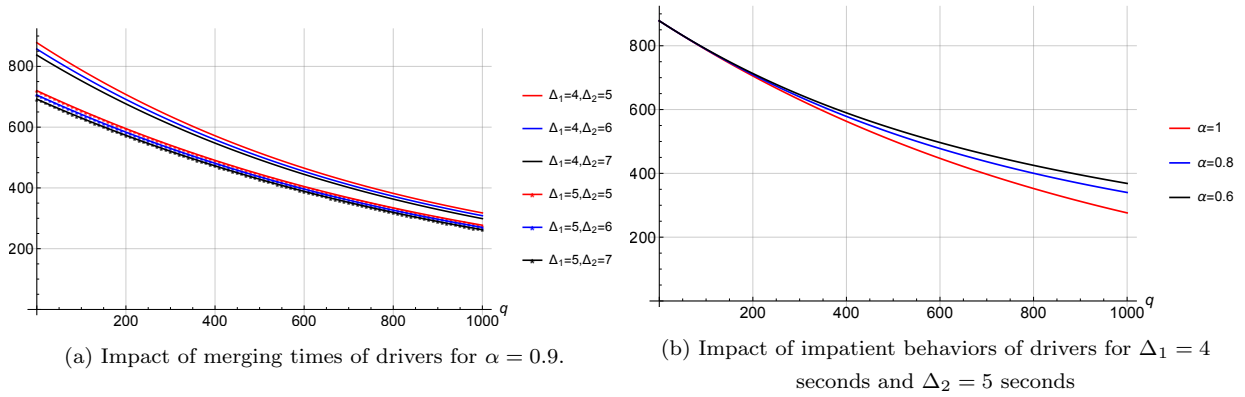
FIGURE 2. Capacity of the minor road (veh/h) as a function of the flow rate on the major road (veh/h) in Example 1.

both methods, for various values of q and α. It immediately becomes apparent that the relative error is below 0.5 percent in all cases. We conducted similar experiments for different settings and in all cases we obtained relative errors below one percent, which clearly justifies using the analysis of this paper in all practical situations, including those violating (4.15). As expected (see Remark 3), the approximated capacities are (slightly) underestimating the actual capacities.

		$q = 250$	$q = 500$	$q = 750$	$q = 1000$
$\alpha = 1.0$	Approximation	646.2	466.4	328.9	225.8
	Simulation	647.2	467.7	330.0	226.5
$\alpha = 0.9$	Approximation	652.8	491.0	377.8	298.9
	Simulation	653.7	491.5	378.0	299.0

TABLE 1. Capacity of the minor road (veh/h) for various values of the flow rate, q, on the major road (veh/h) in Example 2.

6. DISCUSSION AND CONCLUDING REMARKS

In this paper we have presented a gap acceptance model for unsignalized intersections that considerably generalizes the existing models. The model proposed incorporates various realistic aspects that were not taken care of in previous studies: driver impatience, heterogeneous driving behavior, and the service time being dependent on the vehicle arriving at an empty queue or not. Despite the rather intricate system dynamics, we succeed in providing explicit expressions for the stationary queue-length distribution of vehicles on the minor road, which facilitate the evaluation of the corresponding capacity (of the minor road, that is).

We have concluded our paper by presenting a series of numerical results, which are representative of the extensive experiments that we performed. Our techniques facilitate the quantitative evaluation of the system at hand, including the assessment of the sensitivity of the capacity when varying the model parameters (which is typically considerably harder when relying on simulation).

An interesting direction in which to pursue this line of research, would be to allow a more general arrival process on the major road. In the present paper, we assume that the arrival process on the major road is a Poisson process. However, in practice there might be platoon forming on this road and several papers have shown that this clustering of vehicles will influence the capacity of

the *minor* road (cf. [3, 32, 33, 37]). It would certainly be interesting to combine the framework in this paper with, for example, the gap-block arrival process in Cowan [13] or Markov platooning introduced in [3].

REFERENCES

[1] Abhishek, M. A. A. Boon, O. J. Boxma, and R. Núñez Queija. A single server queue with batch arrivals and semi-Markov services. *Queueing Systems*, 86(3–4):217–240, 2017.

[2] Abhishek, M. A. A. Boon, M. Mandjes, and R. Núñez Queija. Congestion analysis of unsignalized intersections. In *COMSNETS 2016: Intelligent Transportation Systems Workshop*, pages 1–6, 2016.

[3] Abhishek, M. A. A. Boon, M. R. H. Mandjes, and R. Núñez Queija. Congestion analysis of unsignalized intersections: The impact of impatience and Markov platooning. `ArXiv` report, University of Amsterdam, 2017.

[4] Abhishek, M. A. A. Boon, and R. Núñez Queija. Heavy-traffic analysis of the M^X/semi-Markov/1 queue. `ArXiv` report, University of Amsterdam, 2018.

[5] M. Abou-Henaidy, S. Teply, and J. H. Hund. Gap acceptance investigations in Canada. In R. Akçelik, editor, *Proceedings of the Second Int. Symp. on Highway Capacity*, volume 1, pages 1–19, 1994.

[6] I. J. B. F. Adan and V. G. Kulkarni. Single-server queue with Markov-dependent inter-arrival and service times. *Queueing Systems*, 45:113–134, 2003.

[7] W. Brilon and T. Miltner. Capacity at intersections without traffic signals. *Transportation Research Record*, 1920:32–40, 2005.

[8] W. Brilon, R. Troutbeck, and M. Tracz. Review of international practices used to evaluate unsignalized intersections. *Transportation Research Circular*, 468, 1997. Transportation Research Board, Washington, DC.

[9] W. Brilon and N. Wu. *Unsignalized Intersections - A Third Method for Analysis*, chapter 9, pages 157–178. 2002.

[10] E. A. Catchpole and A. W. Plank. The capacity of a priority intersection. *Transportation Research-B*, 20B(6):441–456, 1986.

[11] E. Çinlar. Time dependence of queues with semi-Markovian services. *J. Appl. Probab.*, 4:356–364, 1967.

[12] T. E. C. Cheng and S. Allam. A review of stochastic modelling of delay and capacity at unsignalized priority intersections. *European Journal of Operational Research*, 60(3):247–259, 1992.

[13] R. J. Cowan. Useful headway models. *Transportation Research*, 9(6):371–375, 1975.

[14] C. F. Daganzo. Traffic delay at unsignalized intersections: clarification of some issues. *Transportation Science*, 11(2):180–189, 1977.

[15] J. H. A. de Smit. The single server semi-Markov queue. *Stochastic Processes and their Applications*, 22:37–50, 1986.

[16] D. R. Drew. *Traffic flow theory and control*. McGraw-Hill, New York, 1968.

[17] D. R. Drew, J. H. Buhr, and R. H. Whitson. The determination of merging capacity and its applications to freeway design and control. Report 430-4, Texas Transportation Institute, 1967.

[18] D. R. Drew, L. R. LaMotte, J. H. Buhr, and J. Wattleworth. Gap acceptance in the freeway merging process. Report 430-2, Texas Transportation Institute, 1967.

[19] D. P. Gaver. A comparison of queue disciplines when service orientation times occur. *Naval Res. Logist. Quart.*, 10:219–235, 1963.

[20] A. G. Hawkes. Queueing for gaps in traffic. *Biometrika*, 52(1/2):79–85, 1965.

[21] D. Heidemann. Queue length and delays distributions at traffic signals. *Transportation Research-B*, 28(5):377–389, 1994.

[22] D. Heidemann and H. Wegmann. Queueing at unsignalized intersections. *Transportation Research-B*, 31(3):239–263, 1997.

[23] M. Liu, G. Lu, Y. Wang, and Z. Zhang. Analyzing drivers' crossing decisions at unsignalized intersections in China. *Transportation Research-F: Traffic Psychology and Behaviour*, 24:244–255, 2014.

[24] P. K. Munjal and L. A. Pipes. Propagation of on-ramp density perturbations on unidirectional two- and three-lane freeways. *Transportation Research*, 5(4):241–255, 1971.

[25] M. F. Neuts. The single server queue with Poisson input and semi-Markov service times. *J. Appl. Probab.*, 3:202–230, 1966.

[26] M. F. Neuts. The M/G/1 queue with several types of customers and change-over times. *Adv. in Appl. Probab.*, 9:604–644, 1977.

[27] M. F. Neuts. Some explicit formulas for the steady-state behavior of the queue with semi-Markovian service times. *Adv. in Appl. Probab.*, 9:141–157, 1977.

[28] J. Prasetijo. *Capacity and Traffic Performance of Unsignalized Intersections under Mixed Traffic Conditions*. Phd thesis, Ruhr-University Bochum, 2007.

[29] J. Prasetijo and H. Ahmad. Capacity analysis of unsignalized intersection under mixed traffic conditions. *Procedia - Social and Behavioral Sciences*, 43:135–147, 2012. 8th International Conference on Traffic and Transportation Studies (ICTTS 2012).

[30] P. Purdue. A queue with Poisson input and semi-Markov service times: busy period analysis. *J. Appl. Probab.*, 12:353–357, 1975.

[31] J. C. Tanner. The delay to pedestrians crossing a road. *Biometrika*, 38(3/4):383–392, 1951.

[32] J. C. Tanner. A theoretical analysis of delays at an uncontrolled intersection. *Biometrika*, 49(1/2):163–170, 1962.

[33] H. Wegmann. Intersections without traffic signals II. In W. Brilon, editor, *A General Capacity Formula for Unsignalized Intersections*, pages 177–191. Springer-Verlag, 1991.

[34] D. Wei, W. Kumfer, D. Wu, and H. Liu. Traffic queuing at unsignalized crosswalks with probabilistic priority. *To appear in Transportation Letters*, 2016.

[35] G. H. Weiss and A. A. Maradudin. Some problems in traffic delay. *Operations Research*, 10(1):74–104, 1962.

[36] P. D. Welch. On a generalized M/G/1 queuing process in which the first customer of each busy period receives exceptional service. *Operations Research*, 12(5):736–752, 1964.

[37] N. Wu. A universal procedure for capacity determination at unsignalized (priority-controlled) intersections. *Transportation Research-B*, 35:593–623, 2001.

[38] G. F. Yeo. Single server queues with modified service mechanisms. *Journal of the Australian Mathematical Society*, 2(4):499–507, 1962.

[39] G. F. Yeo and B. Weesakul. Delays to road traffic at an intersection. *Journal of Applied Probability*, 1:297–310, 1964.