INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

# Models of Bus Queueing at Curbside Stops

**Weihua Gu, Michael J. Cassidy, Yuwei Li**

WORKING PAPER
UCB-ITS-VWP-2012-3

UC Berkeley Center for Future Urban Transport
A VOLVO Center of Excellence

**July 2012**

# Models of Bus Queueing at Curbside Stops

Weihua Gu [*a], Michael J. Cassidy [a], and Yuwei Li [b]

[a] 416 Mclaughlin Hall, Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720, United States

[b] Department of Transportation Planning, Shanghai Maritime University, 1550 Haigang Boulevard, Shanghai 201306, China

* Corresponding author. Tel.: +1 (510) 931-6646; fax: +1 (510) 643-8919.
E-mail address: weihuagu@berkeley.edu.

We consider curbside bus stops of the kind that serve multiple bus routes, and that are isolated from the effects of traffic signals and other stops. A Markov chain embedded in the bus queueing process is used to develop steady-state queueing models for two special cases of this stop type. The models estimate the maximum rate that buses can arrive to, and be served by the stop, and still satisfy a specified target of average bus delay. These models can be used to determine, for example, a stop's suitable number of bus berths, given the bus demand and the specified delay target. The solutions for the two special cases are used to derive a closed-form, parsimonious approximation model for general cases. Our approximations closely match simulations for a range of conditions that arise in real settings. And the approximations unveil how suitable choices for the number of bus berths are influenced by both, the variation in the time that buses spend serving passengers at the stop, and the specified delay target. The models further show why the proxy measure that is used for the delay target in other bus-stop studies is a poor one.

**Keywords**: bus-stop capacity, bus-stop congestion, bus-stop queueing models

## 1 Introduction

Buses that travel on busy routes can be temporarily blocked from entering or exiting a stop by other buses that are serving passengers there. The resulting reduction in the bus discharge flow from the stop (i.e., the bus-stop capacity) can degrade performance system-wide.

Though professional handbooks (e.g. TRB, 2000) have long offered formulas and tables for estimating bus-stop discharge flow, these are known to be unreliable (Gibson, et al, 1989; Fernandez, 2010; Gu, et al, 2011; Estrada, et al, 2011). And though the latter-cited works are useful in their own rights, most have relied to greater or lesser degrees on computer simulation. The only analytical models produced from these cited works are applicable to simple cases involving either stops with a single bus berth (Gu, et al, 2011), or double-berth stops where bus queues are always present at the entrances (Estrada, et al, 2011). The dearth of analytical methods for evaluating general cases has this far hampered our understanding of cause-and-effect relations at busy stops.

In light of this, the present paper furnishes analytical models to predict the maximum bus flow that can be served by a curbside stop while still maintaining a target service level, where the latter is to be specified by some agency charged with oversight. The models are for steady-state conditions. The bus flow that can be served by the stop is therefore equivalent to the flow that can arrive to, and depart from it. Hence, the literature commonly refers to this maximum bus flow as the stop's "capacity," though it is invariably one that is subject to the specified service target. We shall therefore use the term "allowable bus flow" to describe this capacity. Our intent is to underscore the constraining role played by the service target.

Though our methodology can be applied using almost any metric for this target, average bus delay is the one adopted here. This seems to us the most meaningful choice: it is a readily observable measure that accounts for the delay that buses suffer both, in a queue at the stop's entrance and while waiting to depart the stop after having served their passengers there. It may therefore be the measure that both bus operators and users tend to care about most in regard to a stop's operational quality.

The models to be described are suitable for curbside stops that are isolated from the effects of traffic signals and other bus stops; and where buses arrive at the stop as a Poisson process, as may occur when a moderately busy stop serves multiple bus routes.[1] It is assumed that bus overtaking maneuvers are prohibited, both within an entry queue and within the stop itself, should multiple bus berths exist there. Overtaking restrictions of this kind are common in cities, because an overtaking bus can disrupt car traffic in adjacent travel lanes.

Our solutions feature a Markov chain that is embedded in the bus queueing process. The queueing models formulated as a result furnish exact solutions for two special cases, as described in the following section. These solutions are used to derive a closed-form approximation for general cases, and the approximations agree nicely with simulations, all as shown in Section 3. Our models unveil that the variations in the time that buses spend serving passengers at the stop, and the specified target for average bus delay, can both influence suitable choices for the stop's number of bus berths; and that these influences act in opposing directions. The models further show why a certain proxy that is commonly used for average bus delay is a poor one. These insights are furnished in Section 4. Implications are discussed in Section 5.

## 2 Methodology for Two Special Cases

The literature on queueing theory pertains to systems with servers that are either: in parallel, such that a customer (e.g. a bus) can use any empty server (a bus berth) without being blocked by customers that are occupying the other servers; or in tandem, such that customers will be served by *all* servers in sequence. Systems of these kinds are distinct from the one of present interest. Upon reaching the head of a stop's entry queue, a bus: enters the stop when its upstream-most berth is vacated; advances until encountering either an occupied berth or the stop's downstream-most berth to serve its passengers; and eventually discharges from the stop only after any further downstream berths are all emptied of other buses. Hence, berths in our models are laid-out in

[1] The Poisson assumption may not be valid when the bus flow is very high, as described in Newell (1982).

serial fashion along the stop, but a bus is served at only one berth, and can be blocked by buses that occupy other berths.

Bus service times for loading and unloading passengers at a stop are assumed to be independent and identically distributed across berths. These service-time distributions may, in reality, vary across buses, especially among buses of distinct size or that operate on distinct routes. However, for simplicity, our models are formulated by assuming that each bus' service time conforms to a combined service-time distribution for all buses that use the stop. Adjustments to address more complicated scenarios will be described in Section 3.2.

Two classical methods in queueing theory are utilized for our purposes: embedded Markov chains and the z-transform method. Exact solutions are obtained in these ways for two special cases: (i) a stop with deterministic bus service time and $c$ tandem berths, where $c$ can be any positive integer; and (ii) a stop with a general service time distribution but only two berths. These are denoted as $M/D/c/SERIAL$ and $M/G/2/SERIAL$ systems, respectively, as per the notation of Kendall (1953); and $SERIAL$ is used to denote the queue discipline at a curbside stop.

We next describe the embedded Markov chain in the bus queueing process (Section 2.1), and compute its transition probabilities (Section 2.2). The balance equations are then formulated and solved for the Markov chain's limiting probabilities (Section 2.3), and the models are used to calculate the average bus delay for the two special cases (Section 2.4). The discussion will emphasize the logic used for these matters. Details of the derivations are largely relegated to appendices.

## 2.1 The Embedded Markov Chain

We define a *regenerative point* as the instant in time when the buses in all berths have discharged from the stop. (Though the stop is empty at each regenerative point, it may be filled immediately thereafter if a bus queue is present at the entrance.) A *cycle* is defined as the time interval between two successive regenerative points.

Let $\tilde{L}_n$ be the number of buses queued at the stop's entrance at the beginning of the $n$-th cycle (i.e. the $n$-th regenerative point); $\lambda$ the rate of Poisson bus arrivals; and recall that $c$ is the stop's number of serial berths. We have the following result:

Claim: given $\lambda$, $c$, and the distribution of bus service times at the stop, the stochastic process $\{\tilde{L}_n\}$ is a Markov chain.

To see that the claim is true, note first that from the $n$-th regenerative point and thereafter, the bus arrival process, by virtue of being Poisson, is independent of any historical information. The service process is also independent of historical information, except for $\tilde{L}_n$, because at the regenerative point all berths are empty, and only $\tilde{L}_n$ (and possibly $c$) determines the number of buses that enter the stop and the length of any residual bus queue immediately thereafter.

The bus stop's allowable bus flow (and other of its steady-state properties) can be obtained using the Markov chain once its transition probabilities are determined.

## 2.2 Transition Probabilities

We scale a berth's service rate, $\mu$, to be 1; and define the load rate, $r = \lambda/\mu = \lambda$. This $r$ has a supremum (i.e. a minimum upper bound) above which the system operates in an unstable state with infinite bus queues. As is customary, we consider only steady-state cases in which $r$ is less than its supremum. We further define the Markov chain's transition probabilities $P_{i,j} = Pr\{\tilde{L}_{n+1} = j | \tilde{L}_n = i\}$.

For the special case of the $M/D/c/SERIAL$ system, the $P_{i,j}$ are formulated as functions of $r$ and $c$, and details of this formulation are given in Appendix A.1 of the manuscript. For the $M/G/2/SERIAL$ case, the $P_{i,j}$ are formulated as functions of $r$ and the cumulative distribution function of bus service time, $F_S(t)$, as described in Appendix A.2.

Based on these transition probabilities, we next find the limiting probabilities of the Markov chain by formulating and solving the balance equation.

## 2.3 Balance Equation of Limiting Probabilities and its Solution

Let $\boldsymbol{P} = [P_{i,j}, i \geq 0, j \geq 0]$ be the matrix of transition probabilities; $\pi_i$ $(i \geq 0)$ the limiting probability that the Markov chain is in state $i$, i.e. $\pi_i = \lim_{n\to\infty} Pr\{\tilde{L}_n = i\}$; and $\boldsymbol{\pi} = [\pi_1, \pi_2, ...]$ the limiting distribution of the Markov chain. Thus, $\boldsymbol{\pi}$ is the solution to the balance equation $\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}$.

Our solution method uses the z-transform of $\boldsymbol{\pi}$, $\tilde{\pi}(z) = \sum_{i=0}^{\infty} \pi_i z^i$, to consolidate the infinite-size balance equation into a single functional equation (Crommelin, 1932). Its solution can be converted (e.g. using the inverse z-transform) back to $\boldsymbol{\pi}$, the original distribution. Details on the solutions for the two special cases are given in Appendix B.1 and B.2.

The results obtained thus far are now used to estimate the bus-stop's service level.

## 2.4 Average Bus Delay

Recall that our service-level metric is the average delay imparted to each bus by the stop. This delay is denoted $\overline{W}$ and is normalized to be the (unitless) multiple of the mean service time, $1/\mu$, by again taking $\mu$ to be 1. The $\overline{W}$ is the sum of two unitless average delays: in the entry queue, $\overline{W}_q$; and in the berth after the bus has finished serving its passengers, $\overline{W}_b$.

Determining $\overline{W}_q$ requires the calculation of the average number of buses in queue over time, $\overline{L}_q$. The latter is equal to the average of the queue length seen by each Poisson bus arrival, thanks to the Poisson Arrivals See Time Averages (PASTA) property (Wolff, 1982). The $\overline{L}_q$ is obtained by taking the ratio of two averages: (i) the average of the sum of the queue lengths seen by each bus arrival in cycle $n$, $TL_n$, which we denote as $\overline{TL}$; and (ii) the average of the number of bus arrivals in that cycle, $A_n$, denoted as $\bar{A}$. This is because:

$$\bar{L}_q = \lim_{T\to\infty} \frac{\text{sum of queue lengths seen by all arrivals during } T}{\text{number of arrivals during } T}$$
$$= \lim_{T\to\infty} \frac{\text{sum of queue lengths seen by all arrivals during } T/\text{number of cycles during } T}{\text{number of arrivals during } T/\text{number of cycles during } T}$$
$$= \frac{\lim_{T\to\infty} \frac{\text{sum of queue lengths seen by all arrivals during } T}{\text{number of cycles during } T}}{\lim_{T\to\infty} \frac{\text{number of arrivals during } T}{\text{number of cycles during } T}}$$
$$= \frac{\overline{TL}}{\bar{A}}.$$

To obtain $\overline{TL}$ and $\bar{A}$, we consider the following four scenarios which describe the possible states of the system at the start and end of each $n$-th cycle.

Scenario 1: No bus queues are present at the stop's entry, both at the start and the end of the $n$-th cycle; i.e. $\tilde{L}_n = \tilde{L}_{n+1} = 0$. In this scenario, no bus arriving during cycle $n$ encounters a queue, meaning that $TL_n = 0$; and the number of buses that arrive during cycle $n$, $A_n$, is the number served during that cycle, $M_n$; i.e. $A_n = M_n$.

Scenario 2: A bus queue is present at the start of cycle $n$ ($\tilde{L}_n = i > 0$), but not at the end of that cycle ($\tilde{L}_{n+1} = 0$). To satisfy $\tilde{L}_{n+1} = 0$, the $i$ can be no greater than $c$, and $TL_n = 0$. Further note that $A_n = M_n - i$.

Scenario 3: A queue of size $i \leq c$ is present at the start of cycle $n$ ($\tilde{L}_n = i \leq c$), and a queue persists at the end of that cycle ($\tilde{L}_{n+1} = j > 0$). In this scenario, the stop is filled during the cycle; i.e. $M_n = c$ and $A_n = c + j - i$. The first $c - i$ arrivals fill unused berths, such that the first $(c - i + 1)$ arrivals see no entry queue. The $j - 1$ arrivals to follow will see successively longer queues that range from 1 to $j - 1$. Thus, $TL_n = \frac{1}{2} j(j-1)$.

Scenario 4: A queue size greater than $c$ is present at the start of cycle $n$ ($\tilde{L}_n = i \geq c + 1$), and a queue thus persists at the end of that cycle ($\tilde{L}_{n+1} = j \geq i - c > 0$). In this scenario, as in the previous one, $M_n = c$ and $A_n = c + j - i$. And since the earliest moments of cycle $n$ are characterized by $i - c$ buses that remain in the entry queue, arrivals thereafter will see queue lengths in the sequence $i - c, i - c + 1, \dots, j - 1$. Thus $TL_n = \frac{1}{2}(i + j - c - 1)(c + j - i)$.

Note from the above that, for known $c$, the $TL_n$ and $A_n$ depend only on $\tilde{L}_n$, $\tilde{L}_{n+1}$, and $M_n$. Thus, $\overline{TL}$ and $\bar{A}$ can be obtained by taking weighted averages:

$$\overline{TL} = \sum_{i,j,k} Pr\{\tilde{L}_n = i, \tilde{L}_{n+1} = j, M_n = k\} \cdot TL_n$$

$$\bar{A} = \sum_{i,j,k} Pr\{\tilde{L}_n = i, \tilde{L}_{n+1} = j, M_n = k\} \cdot A_n$$

where $Pr\{\tilde{L}_n = i, \tilde{L}_{n+1} = j, M_n = k\}$ is the long-run probability of a cycle where $\tilde{L}_n = i$, $\tilde{L}_{n+1} = j$, and $M_n = k$. This probability is equal to $\pi_i \cdot Pr\{\tilde{L}_{n+1} = j, M_n = k | \tilde{L}_n = i\}$; and the derivations of $Pr\{\tilde{L}_{n+1} = j, M_n = k | \tilde{L}_n = i\}$ for the two special cases are furnished in Appendix A.

Therefore, we have:

$$\bar{L}_q = \frac{\sum_{i,j,k} Pr\{\tilde{L}_n = i, \tilde{L}_{n+1} = j, M_n = k\} \cdot TL_n}{\sum_{i,j,k} Pr\{\tilde{L}_n = i, \tilde{L}_{n+1} = j, M_n = k\} \cdot A_n} . \quad (1)$$

The average bus delay in the queue is then obtained from Little's formula: $\bar{W}_q = \bar{L}_q/\lambda$.

The above recipe for $\bar{W}_q$ is applied to $M/D/c/SERIAL$ systems in Appendix C.1. Given that the service time is deterministic, the average bus delay in berths, $\bar{W}_b$, is 0 in this case.

For $M/G/2/SERIAL$ systems, $\bar{W}_b$ can be obtained similarly to the above, as shown in Appendix C.2. This latter appendix also describes the determination of $\bar{W}_q$ for these systems.

In theory, the methodology presented above can be applied to any $M/G/c/SERIAL$ system. However, when $c > 2$ and the bus service time follows certain distributions (e.g., gamma), computational complexity increases dramatically. An approximation for $M/G/c/SERIAL$ systems is therefore sought using the results obtained above.

# 3 An Approximation for General Service Time Distribution and Arbitrary *c*

An approximation formula for $M/G/c/SERIAL$ systems is developed in Section 3.1, and is tested against simulation in Section 3.2.

## 3.1 Approximation Formula

Recall that the exact solution for the $M/D/c/SERIAL$ system accounts for the effect of $c$ on $\bar{W}$, and that the solution for the $M/G/2/SERIAL$ system uses the cumulative distribution function of bus service time, $F_S$, as input. The exact solution for this latter system consequently accounts for the effect on $\bar{W}$ imposed by the coefficient of variation in bus service time, denoted $C_S$.

With the above in mind, an approximation for general $M/G/c/SERIAL$ systems can be formulated by applying an idea similar to those previously used in the queueing literature (e.g. Maalφe, 1973; Nozaki and Ross, 1975):

$$\bar{W}_{M/G/c/SERIAL}(\rho, c, C_S) \approx \frac{\bar{W}_{M/D/c/SERIAL}(\rho,c)}{\bar{W}_{M/D/c/SERIAL}(\rho,2)} \cdot \bar{W}_{M/G/2/SERIAL}(\rho, C_S), \quad (2)$$

where the $\bar{W}$ denote the average bus delays for the queueing systems described by the subscripts; and the service ratio, $\rho$, is the ratio of bus inflow to the supremum of the bus discharge flow from the stop; i.e. the bus discharge flow when a bus queue is always present. Thus, $\rho$ takes a value between 0 and 1.

We found a closed-form approximation by (2) by first obtaining approximations for $\bar{W}_{M/D/c/SERIAL}$ and $\bar{W}_{M/G/2/SERIAL}$ by fitting least squares models to the exact solutions. Both of these $\bar{W}$'s increase as $\rho$ increases; they are 0 when $\rho = 0$, and they approach infinity as $\rho \rightarrow 1$. With this in mind, combinations of algebraic fractions and exponential functions, monomials and trigonometric functions were taken as candidate forms for least squares models. From the many candidates tried, the following models were found to have the minimum square errors:

$$\bar{W}_{M/D/c/SERIAL}(\rho, c) = \frac{0.19c+0.061}{c-0.54}\left(\tan\frac{\pi}{2}\rho\right)^{(-0.065c+1.21)} \quad (3)$$

and

$$\bar{W}_{M/G/2/SERIAL}(\rho, C_S) = (0.29C_S + 0.29)\left(\tan\frac{\pi}{2}\rho\right)^{(0.046C_S+1.10)}, \quad (4)$$

where for $M/G/2/SERIAL$ systems, we assume that bus service time is uniformly distributed.

Inserting (3) and (4) into (2), we obtain an approximation for average bus delay:

$$\bar{W}_{M/G/c/SERIAL}(\rho, c, C_S) = \frac{0.63c+0.20}{c-0.54}(0.29C_S + 0.29)\left(\tan\frac{\pi}{2}\rho\right)^{(-0.065c+0.046C_S+1.23)} \quad (5)$$

Equation (5) can be used to estimate the stop's allowable bus flow for a specified target of $\bar{W}$ by exploiting a relationship between the load rate, $r$, and $\rho$. For uniformly distributed bus service time, that relationship is

$$r = \frac{c}{1+\sqrt{3}C_S\frac{c-1}{c+1}}\rho, \quad (6)$$

as derived in Appendix D.

Combining (5) and (6):

$$r = \frac{2c}{\pi(1+\sqrt{3}C_S\frac{c-1}{c+1})}\tan^{-1}\left[\left(\frac{c-0.54}{(0.63c+0.20)(0.29C_S+0.29)}\bar{W}\right)^{\frac{1}{-0.065c+0.046C_S+1.23}}\right]. \quad (7)$$

When $\mu$ is normalized to 1, $r$ is the bus-stop's normalized allowable flow for a specified $\bar{W}$. If $\mu \neq 1$, the allowable flow can be estimated by $\lambda = r\mu$, where $r$ is determined from (7). If the specified target for $\bar{W}$ is not normalized but instead has units (e.g. minutes), then $\bar{W} = \mu\bar{w}$, where $\bar{w}$ is the un-normalized target average bus delay.

The accuracy of this approximation is examined next.

## 3.2 Tests

Results from (7) are compared against outcomes from simulation. The latter were obtained from a simulation model already described in the literature (Gu, et al, 2011), and each simulated scenario entailed the arrival and departure of at least 500,000 buses to ensure that predictions converged to expected values in the steady state. Comparisons begin with scenarios involving multi-berth stops with conditions that are consistent with, or nearly consistent with, the assumptions in (7). More complicated scenarios will be examined thereafter.

### *3.2.1 Homogeneous Bus Size and Service Times*

Consider cases in which all buses are of a single size and have service times that conform to a single distribution with a normalized mean of 1. Four distributional forms for service time are explored: two of these, deterministic (D) and uniform (U), were used in the derivation of (7); and two others, gamma (G) and Weibull (W) were not.

Outcomes are furnished in Table 1. Each cell presents the percent difference between approximated and simulated values of normalized allowable bus flow, $r$. These are shown for ranges of $c$ and $C_S$ that occur in real settings (e.g. St. Jacques and Levinson, 1997), and for reasonable range of $\overline{W}$. Negative-valued outcomes denote cases in which the approximation is lower than the simulated prediction.

Reassuringly, the predictions for most of the scenarios agree to within 10%. Numerous exceptions to this arise, however. These are highlighted with shading in the figure and they often reflect the influence of the assumptions behind (7). For example, we see that larger differences can occur when service times are gamma- or Weibull distributed, which are not the distributions that led to (7). Larger differences also tend to occur as $c$ and $C_S$ jointly grow large. This is not surprisingly, since (7) is the compilation of two exact solutions that assume either $c$ or $C_S$ is small.

Finally, larger differences occur when $\overline{W}$ is small. This is an artifact of having minimized squared errors when the predictions (of $r$) were themselves rather small.

Table 1 Relative errors of the approximation – homogeneous buses

| **Service time distributions** | | $c = 2$ | | | $c = 4$ | | |
|---|---|---|---|---|---|---|---|
| | | $\overline{W} = 0.2$ | $\overline{W} = 0.5$ | $\overline{W} = 2$ | $\overline{W} = 0.2$ | $\overline{W} = 0.5$ | $\overline{W} = 2$ |
| **D** | $C_S = 0$ | -6% | 2% | 0.8% | -0.8% | 2% | -0.2% |
| **U** | $C_S = 0.25$ | -17% | -4% | 0.3% | -12% | 0.5% | -0.1% |
| | $C_S = 0.5$ | -16% | -3% | 0.8% | -9% | 4% | 1% |
| **G** | $C_S = 0.25$ | -16% | -3% | 0.3% | -11% | 1% | 0.4% |
| | $C_S = 0.5$ | -14% | -3% | 0.4% | -6% | 8% | 5% |
| | $C_S = 0.75$ | -6% | 3% | 2% | 10% | 19% | 10% |
| **W** | $C_S = 0.25$ | -17% | -4% | -0.2% | -13% | -0.5% | -0.7% |
| | $C_S = 0.5$ | -15% | -3% | 0.8% | -8% | 7% | 4% |
| | $C_S = 0.75$ | -7% | 2% | 2% | 10% | 18% | 9% |

*3.2.2 Heterogeneous Service Times*

Consider now cases in which homogeneously-sized buses exhibit distinct distributions of service time, as may occur when buses using the same stop serve distinct routes. For illustration, we explore scenarios involving two fleets of standard-size buses. Service times for both will be gamma distributed, but the mean for fleet "A" is greater than that of "B" by 50%, and the $C_S$ is 0.6 and 0.4 for "A" and "B," respectively.

The Poisson arrival processes of both fleets are assumed to be independent of the service times. To generate the approximations, the mean of the combined service-time distributions is normalized to 1, and the $C_S$ of the combined distributions were used in (7).

Outcomes are presented in Table 2. The approximations and simulated results for most of the scenarios agree rather closely. Exceptions again can occur for small $\overline{W}$ and when $c$ and $C_S$ jointly grow large.

Table 2 Relative errors of the approximation – heterogeneous service time distributions

| **Flow ratio between fleets A and B** | $c = 2$ | | | $c = 4$ | | |
|---|---|---|---|---|---|---|
| | $\overline{W} = 0.2$ | $\overline{W} = 0.5$ | $\overline{W} = 2$ | $\overline{W} = 0.2$ | $\overline{W} = 0.5$ | $\overline{W} = 2$ |
| **20% / 80% (combined $C_S = 0.522$)** | -18% | -5% | -0.5% | -13% | 2% | 0.6% |
| **50% / 50% (combined $C_S = 0.592$)** | -14% | -3% | 1% | -8% | 6% | 4% |
| **80% / 20% (combined $C_S = 0.606$)** | -11% | 0.4% | 3% | -2% | 11% | 7% |

*3.2.3 Heterogeneous Buses*

For the final battery of tests, consider multi-berth stops that serve buses of distinct sizes (standard-sized and articulated) and where service times tend to be longer for the larger, articulated buses. Assume that a stop's berths are sized to fit the standard buses and that, in reality, articulated buses occupy two berths while dwelling at a stop.

We use (7) by treating each articulated bus as two standard-sized buses.[2] The simulation model in Gu, et al (2011) was modified for this round of tests, as furnished in Appendix E. The arrivals of both bus classes are assumed to be independent Poisson processes with gamma distributed service times. The mean service time of the articulated fleet is 1.5 times that of the standard-sized one; and the $C_S$ are 0.6 and 0.4 for the articulated and standard buses, respectively.

Outcomes are given in Table 3, where the case of $c = 3$ is added to show the fitness of the approximation for stops with an odd number of berths. For the most part, the agreements between approximated and simulated $r$ are quite good. Note that large differences arise when a stop with an odd number of berths ($c$ = 3) serves a vehicle stream that is dominated by articulated buses (80%). Whether scenarios of this kind often arise in real settings seems an open question; e.g. a bus agency might do well to expand the berth sizes at stops that primarily serve articulated vehicles.

[2] Our use of (7) implicitly assumes that the two adjoining portions of an articulated bus operate independently and with distinct service times. Hence, for example, the front portion of an articulated bus can: enter a stop's upstream-most berth while the rear portion remains in the entry queue; and thereafter temporarily block the rear portion from serving passengers.

Table 3 Relative errors of the approximation – heterogeneous buses

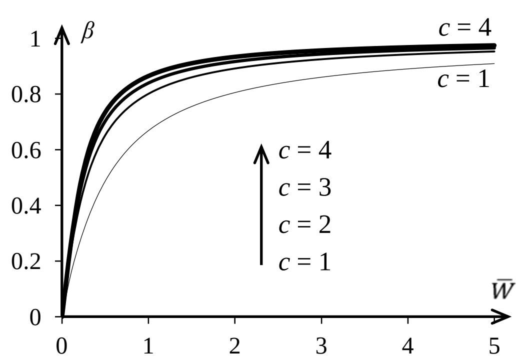
| **Flow ratio between articulated and standard-sized buses** | $\boldsymbol{c = 2}$ | | | $\boldsymbol{c = 3}$ | | | $\boldsymbol{c = 4}$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\boldsymbol{\overline{W} = 0.2}$ | $\boldsymbol{\overline{W} = 0.5}$ | $\boldsymbol{\overline{W} = 2}$ | $\boldsymbol{\overline{W} = 0.2}$ | $\boldsymbol{\overline{W} = 0.5}$ | $\boldsymbol{\overline{W} = 2}$ | $\boldsymbol{\overline{W} = 0.2}$ | $\boldsymbol{\overline{W} = 0.5}$ | $\boldsymbol{\overline{W} = 2}$ |
| **20% / 80% (combined $\boldsymbol{C_S = 0.564}$)** | -8% | 0.6% | -0.2% | -11% | 2% | 2% | -11% | 5% | 3% |
| **50% / 50% (combined $\boldsymbol{C_S = 0.604}$)** | 7% | 8% | 0.1% | -0.1% | 12% | 7% | -6% | 10% | 5% |
| **80% / 20% (combined $\boldsymbol{C_S = 0.605}$)** | 5% | 4% | -7% | 17% | 21% | 10% | -8% | 5% | -0.8% |

# 4 Insights

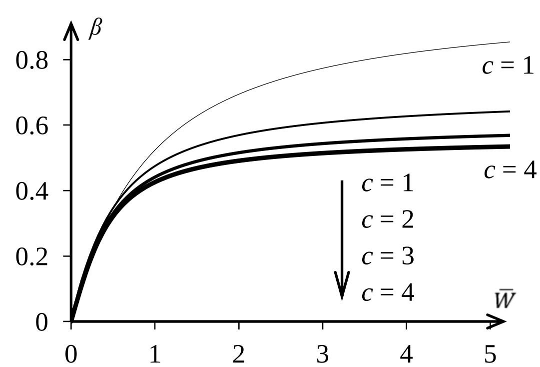
We use (7) to examine how $\overline{W}$ and $C_S$ can affect suitable choices of $c$ for a stop (Sections 4.1 and 4.2), and to explore the nature of these influences (Section 4.3). We also demonstrate the pitfalls of specifying service-level targets by means of proxy measure, rather than $\overline{W}$ itself (Section 4.4).

## 4.1 Returns in Allowable Flow

Consider first the allowable bus flows that can be returned to a stop by adding berths to it. Figures 1a and b present predictions of normalized allowable flow per berth, $\beta = \lambda/(c\mu)$, as functions of normalized (unit-less) $\overline{W}$. These are shown in both figures for $c = 1{\sim}4$, and for $C_S = 0$ in Figure 1a and $C_S = 0.75$ in Figure 1b.



(a) $C_S = 0$



(b) $C_S = 0.75$

Figure 1 – Average allowable flow per berth versus $\overline{W}$

The figures collectively unveil a key role played by $C_S$. When $C_S$ = 0 (Figure 1a), $\beta$ always increases with $c$, meaning that the returns in allowable bus flow increase with added berths, for any choice of $\bar{W}$ examined here. However, when $C_S$ is high (Figure 1b), the opposite occurs, meaning that adding berths will never bring an increasing return in allowable flow.[3]

## 4.2 Maximizing Berth Productivity

We next identify the values of $c$ that bring the highest $\beta$, given $\bar{W}$ and $C_S$. The curves in Figure 2 collectively delineate regions where certain $c$ for a stop (as shown in ellipses) result in the greatest $\beta$.

Visual inspection of this figure reveals that, for any choice of the $\bar{W}$ examined, the value of this $c$ diminishes as $C_S$ grows large. This observation is consistent with the findings of the previous section.

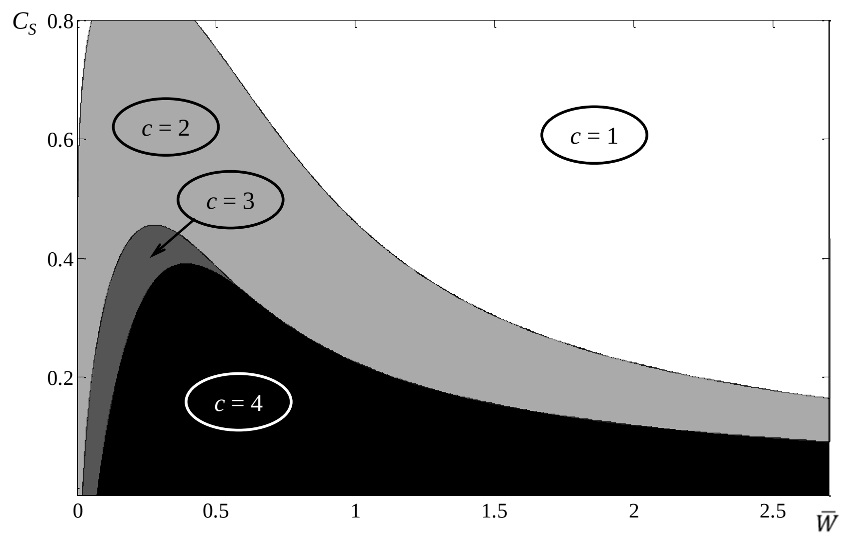



Figure 2 – The most productive values of $c$ in the range: $C_S = 0\sim0.8$, $\bar{W} = 0\sim2.7$, and $c = 1\sim4$

## 4.3 Blocking among Buses

The observations of the previous two sections arise because dwelling buses can temporarily block other buses from entering or exiting the stop. We explore this matter by means of simple experiment: allowable bus flows, r, for ordinary curbside stops are compared against the r for idealized stops where each bus freely overtakes other dwelling buses and thereby enter vacant berths, and later depart those berths, without delay. The r for an idealized stop will, of course, exceed that of its ordinary counterpart, and this difference is the loss in allowable bus flow due to bus blocking.

We note that an idealized stop is equivalent to the $M/G/c$ queueing system with parallel servers. Solutions for this system were obtained via simulation; a well-known algorithm (e.g. see Cooper, 1981) was used for that purpose. Predictions for the ordinary stops were obtained from our exact

[3] These present findings run contrary to those of an earlier study; see Gu, et al (2011). This is probably because the earlier study relied upon a proxy measure for $\bar{W}$, and the pitfalls of this will be discussed momentarily.

solutions except for cases with high computational costs; e.g., as occurs for gamma-distributed service times. These latter cases were generated from simulation and are denoted with shading in Table 4, which presents the outcomes of our comparisons. The values in that table reveal the percent reductions in allowable flow due to blocking. Three observations emerge.

First, downward comparisons of the values in any single column of the table show that the fractional losses in allowable bus flow grow larger with growing $C_S$. This unveils how bus blocking grows more disruptive as bus service times grow more varied.

Second, for a given $C_S$ and $\overline{W}$, one can see that the percent reduction in allowable flow is more pronounced for $c = 4$ than for $c = 2$. Thus we see how a greater number of berths can exacerbate the blocking problem.

And third, comparing the values across any single row and for a given $c$, one observes that fractional flow reductions attenuate with the selection of larger $\overline{W}$. This reveals how a reservoir of excess demand (i.e., a more persistent entry queue owing to a choice of large $\overline{W}$) can lessen the dominating effect of blocking.

Table 4 Relative losses in allowable flow due to bus blocking

| **Service time distributions** | | $c = 2$ | | | $c = 4$ | | |
|---|---|---|---|---|---|---|---|
| | | $\overline{W} = 0.2$ | $\overline{W} = 0.5$ | $\overline{W} = 2$ | $\overline{W} = 0.2$ | $\overline{W} = 0.5$ | $\overline{W} = 2$ |
| **D** | $C_S = 0$ | 24.9% | 7.3% | 0.2% | 35.1% | 11.9% | 1.3% |
| **U** | $C_S = 0.25$ | 31.4% | 25.1% | 14.9% | 50.4% | 35.2% | 23.1% |
| | $C_S = 0.5$ | 43.3% | 36.8% | 25.1% | 64.2% | 49.4% | 37.1% |
| **G** | $C_S = 0.75$ | 56.7% | 46.4% | 34.5% | 76.9% | 65.2% | 51.4% |

An implication of the above findings will be noted in Section 5. Before pursuing that discussion, we offer a final insight.

## 4.4 Pitfalls by Proxy

In the literature on bus-stop capacity, one commonly finds that the metric for target service level is not $\overline{W}$, but rather an apparent proxy known as failure rate (TRB, 2000, 2003; Levinson and St. Jacques, 1998; Gu, et al, 2011). The latter metric is defined as the probability that a bus arriving at a stop is temporarily blocked from using it by another bus. This proxy, which we denote as $FR$, turns out to be a poor one.

To see this, consider the relations between $FR$ and normalized $\overline{W}$ shown in Figures 3a and b. These were obtained using our exact solutions of Section 2. Figure 3a presents $M/D/c/SERIAL$ systems for $c = 1$–$4$. Figure 3b displays $M/G/2/SERIAL$ systems for the range of $C_S$ that occurs in real settings (e.g. St. Jacques and Levinson, 1997).

The relations shown in both figures are non-linear. And both reveal that for modest values of $\overline{W}$, small changes in that metric coincide with disproportionally large changes in $FR$. Conversely, $FR$ becomes rather insensitive to changes in $\overline{W}$ when $\overline{W}$ is large.

Further note from Figure 3a that for distinct values of $c$, the same $FR$ corresponds to different values of $\overline{W}$. A dashed vertical line is shown in the figure to highlight this point. Notice that if an analyst chooses an invariant value of $FR$ as the service-level target, then reductions in $\overline{W}$ brought by adding berths to a stop may not be captured in the model.

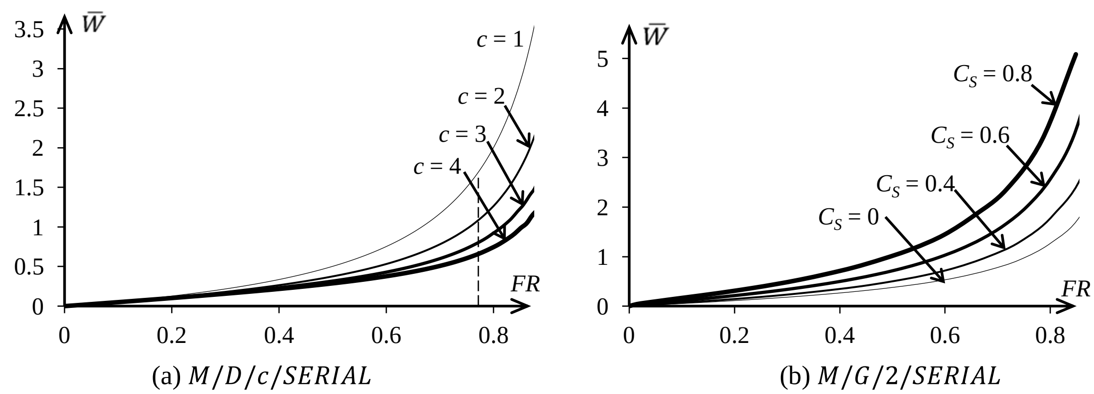



Figure 3 – Failure rate versus average bus delay

# 5 Conclusions

Formulas were developed to predict the maximum bus flows that can be served by a select class of bus stops while still maintaining target service levels. Alternatively, the models can be used to determine a stop's suitable number of serial berths to achieve a specified service level. The formulas use a Markov chain that is embedded in the bus queueing process at these stops. Exact solutions were derived for two special cases, and used to develop a closed-form and parsimonious approximation for the general case. Approximations compared nicely with simulated results.

The present models are limited in that they apply to curbside stops that are isolated from traffic signals and other bus stops, and where bus overtaking maneuvers are prohibited. Furthermore, the models are idealized in that they describe operation in the steady-state. And though prediction errors were found to be modest in many cases, the approximation for the general case introduces error nonetheless.

Still, the models shed light on cause-and-effect relations. Most notably, they unveil the role played by the variation in bus service-time at a stop, and point to the value of management schemes (e.g. off-line fare collection) that limit this variation. The findings show that schemes of this kind can be especially valuable at stops with a large number of berths or when the bus agency opts for a low target for service level.

As a practical matter, our models offer improvements over existing methods of bus-stop design and analysis. Further, the modeling framework can be used to predict other outputs of interest, such as the distribution of queue lengths at a stop. Predictions of this latter kind can be especially useful at stops with very limited queue-storage space. And we have seen that with modest adjustments, our models can be used to approximate rather complex scenarios, such as when the buses served by a stop come in a variety of sizes and display varying service-time distributions.

With other small modifications, the present models could account for the lost times due to bus deceleration and acceleration into and out of a stop, and the extra times required of boarding passengers as they walk to appropriate berths at a multi-berth stop. Modest modifications would also make our models suitable for bus stops with different queue disciplines, e.g., at stops where a bus that has finished serving its passengers can exit by overtaking other buses as they dwell in downstream berths.

Finally, the present models can be applied to other serial queueing systems, e.g., taxi queues, personal rapid transit stations, and highway toll stations with tandem service berths. Models for bus-stops that are affected by neighboring traffic signals or other bus stops nearby are currently being developed.

## Acknowledgements

Funding for this work was provided by the University of California Transportation Center, by the Volvo Research and Educational Foundations, and by the Dwight David Eisenhower Transportation Graduate Fellowship.

### Appendix A

### Calculating the Transition Probabilities in Section 2.2

We calculate the transition probabilities of the embedded Markov chain for $M/D/c/SERIAL$ systems in Section A.1, and for $M/G/2/SERIAL$ systems in Section A.2.

#### A.1 Transition Probabilities for $M/D/c/SERIAL$

We obtain the transition probabilities, $P_{i,j}$, by conditioning on the number of buses that get served in the $n$-th cycle, $M_n$, and then by finding all the $Pr\{\tilde{L}_{n+1} = j, M_n = k|\tilde{L}_n = i\}$ for $1 \le k \le c$. These $Pr\{\tilde{L}_{n+1} = j, M_n = k|\tilde{L}_n = i\}$ are also useful when calculating the average bus delay (see Section 2.4). Note that service times are normalized to 1 for this system.

First, we obtain the following equations by conditioning:
$P_{0,0} = \sum_{k=1}^{c} Pr\{\tilde{L}_{n+1} = 0, M_n = k|\tilde{L}_n = 0\}$;
$P_{0,j} = Pr\{\tilde{L}_{n+1} = j, M_n = c|\tilde{L}_n = 0\}$, for $j \ge 1$;
$P_{i,0} = \sum_{k=i}^{c} Pr\{\tilde{L}_{n+1} = 0, M_n = k|\tilde{L}_n = i\}$, for $i \ge 1$;
$P_{i,j} = Pr\{\tilde{L}_{n+1} = j, M_n = c|\tilde{L}_n = i\}$, for $i \ge 1, j \ge 1$.

Then we calculate $Pr\{\tilde{L}_{n+1} = j, M_n = k | \tilde{L}_n = i\}$ as follows:
$Pr\{\tilde{L}_{n+1} = 0, M_n = k | \tilde{L}_n = 0\}$
$= Pr\{H_l < 1, l = 1,2, \dots, k-1, \text{and } H_k > 1\}$
$= \prod_{l=1}^{k-1} Pr\{H_l < 1\} \cdot Pr\{H_k > 1\}$
$= (1 - e^{-r})^{k-1} e^{-r}, 1 \leq k \leq c,$
where $H_l$ $(l = 1,2, \dots, k)$ is the exponential headway following the $l$-th bus arrival in the cycle; Similarly,

$Pr\{\tilde{L}_{n+1} = j, M_n = c | \tilde{L}_n = 0\} = (1 - e^{-r})^{c-1} \cdot \frac{e^{-r} r^j}{j!}$, for $j \geq 1$;
$Pr\{\tilde{L}_{n+1} = 0, M_n = k | \tilde{L}_n = i\} = (1 - e^{-r})^{k-i} e^{-r}$, for $1 \leq i \leq c-1, i \leq k \leq c$;
$Pr\{\tilde{L}_{n+1} = j, M_n = c | \tilde{L}_n = i\} = (1 - e^{-r})^{c-i} \cdot \frac{e^{-r} r^j}{j!}$, for $1 \leq i \leq c-1, j \geq 1$;
$Pr\{\tilde{L}_{n+1} = j, M_n = c | \tilde{L}_n = i\} = \frac{e^{-r} r^{j-i+c}}{(j-i+c)!}$, for $i \geq c, j \geq i-c$;
for all other combinations of $i$, $j$, and $k$, $Pr\{\tilde{L}_{n+1} = j, M_n = k | \tilde{L}_n = i\} = 0$.

If we let $q_m = \frac{e^{-r} r^m}{m!}$, for $m = 0, 1, 2, \dots$, then $P_{i,j}$ can be written as follows:
$P_{i,j} = q_{j-i+c}$, for $i \geq c$, and $j \geq i-c$;
$P_{i,j} = 0$, for $i \geq c$, and $j < i-c$;
$P_{i,0} = 1 - (1 - q_0)^{c+1-i}$, for $1 \leq i \leq c-1$;
$P_{i,j} = (1 - q_0)^{c-i} q_j$, for $1 \leq i \leq c-1$, and $j \geq 1$;
$P_{0,j} = P_{1,j}$, for $j \geq 0$.

### A.2 Transition Probabilities for $M/G/2/SERIAL$

As in Section A.1, we have:
$P_{0,0} = P_{1,0} = Pr\{\tilde{L}_{n+1} = 0, M_n = 1 | \tilde{L}_n = 0\} + Pr\{\tilde{L}_{n+1} = 0, M_n = 2 | \tilde{L}_n = 0\}$
$= Pr\{\tilde{L}_{n+1} = 0, M_n = 1 | \tilde{L}_n = 1\} + Pr\{\tilde{L}_{n+1} = 0, M_n = 2 | \tilde{L}_n = 1\}$;
$P_{0,j} = P_{1,j} = Pr\{\tilde{L}_{n+1} = j, M_n = 2 | \tilde{L}_n = 0\} = Pr\{\tilde{L}_{n+1} = j, M_n = 2 | \tilde{L}_n = 1\}$, for $j > 0$;
$P_{i,j} = Pr\{\tilde{L}_{n+1} = j, M_n = 2 | \tilde{L}_n = i\}$ , for $i \geq 2$, and $j \geq i-2$;
$P_{i,j} = 0$, otherwise.

Let $F_S(t)$ be the cumulative distribution function (CDF) of bus service time. We have:
$Pr\{\tilde{L}_{n+1} = 0, M_n = 1 | \tilde{L}_n = 1\} = Pr\{H_1 > S_1\} = \int_{t=0}^{\infty} e^{-rt} \mathrm{d}F_S(t)$,
where $H_1$ is the headway following the first (and the only) bus arrival in the cycle; and $S_1$ is that bus' service time.

Given that $H_1 < S_1$, there would be at least 2 arrivals in the cycle. Let $\tau$ be the time between the 2nd arrival and its departure in this cycle:
$\tau = \max\{S_1 - H_1, S_2 | H_1 < S_1\}$,
where $S_2$ is the service time of the 2nd bus to arrive in the cycle.

We can derive the CDF of $\tau$ as:

$F_\tau(t) = Pr\{\tau < t\} = Pr\{\max\{S_1 - H_1, S_2 | H_1 < S_1\} < t\}$

$= \frac{Pr\{H_1 < S_1 < H_1 + t\}}{Pr\{H_1 < S_1\}} \cdot Pr\{S_2 < t\}$

$= \frac{\int_{h=0}^{\infty}(F_S(h+t) - F_S(h)) r e^{-rh} \mathrm{d}h}{Pr\{H_1 < S_1\}} \cdot F_S(t),$

where the third equality holds because $H_1$, $S_1$, and $S_2$ are mutually independent.

Thus for $j \geq 0$,

$Pr\{\tilde{L}_{n+1} = j, M_n = 2 | \tilde{L}_n = 1\} = Pr\{H_1 < S_1\} \cdot \int_{t=0}^{\infty} \frac{e^{-rt}(rt)^j}{j!} \mathrm{d}F_\tau(t)$

$= \int_{t=0}^{\infty} \frac{e^{-rt}(rt)^j}{j!} \mathrm{d}\left[F_S(t) \cdot \int_{h=0}^{\infty}(F_S(h+t) - F_S(h)) r e^{-rh} \mathrm{d}h\right].$

Finally, the CDF of the platoon service time of 2 buses entering the stop simultaneously would be ${F_S}^2(t)$. Then for $i \geq 2$, and $j \geq i - 2$:

$Pr\{\tilde{L}_{n+1} = j, M_n = 2 | \tilde{L}_n = i\} = \int_{t=0}^{\infty} \frac{e^{-rt}(rt)^{j-i+2}}{(j-i+2)!} \mathrm{d}{F_S}^2(t).$

In summary, we have:

$P_{0,0} = P_{1,0} = \int_{t=0}^{\infty} e^{-rt} \mathrm{d}\left[F_S(t) \cdot \left(1 + \int_{h=0}^{\infty}(F_S(h+t) - F_S(h)) r e^{-rh} \mathrm{d}h\right)\right];$

$P_{0,j} = P_{1,j} = \int_{t=0}^{\infty} \frac{e^{-rt}(rt)^j}{j!} \mathrm{d}\left[F_S(t) \cdot \int_{h=0}^{\infty}(F_S(h+t) - F_S(h)) r e^{-rh} \mathrm{d}h\right]$, for $j > 0$;

$P_{i,j} = \int_{t=0}^{\infty} \frac{e^{-rt}(rt)^{j-i+2}}{(j-i+2)!} \mathrm{d}{F_S}^2(t)$, for $i \geq 2$, and $j \geq i - 2$;

$P_{i,j} = 0$, otherwise.

## Appendix B

## Solving the Balance Equations in Section 2.3

We continue from Appendix A, and solve the balance equations for the limiting probabilities for $M/D/c/SERIAL$ and $M/G/2/SERIAL$ systems in B.1 and B.2, respectively. The z-transform method is used for this.

### B.1 Limiting Probabilities for $M/D/c/SERIAL$

From the transition probabilities given in the end of Appendix A.1, the balance equations can be written as:

$$\pi_k = \begin{cases} \pi_0(1-(1-q_0)^c) + \sum_{i=1}^{c-1} \pi_i (1-(1-q_0)^{c+1-i}) + \pi_c q_0, & \text{for } k = 0 \\ \pi_0(1-q_0)^{c-1} q_k + \sum_{i=1}^{c-1} \pi_i (1-q_0)^{c-i} q_k + \sum_{i=c}^{k+c} \pi_i q_{k+c-i}, & \text{for } k = 1, 2, \ldots \end{cases}.$$

We first define:

$\tilde{q}(z) = \sum_{k=0}^{\infty} z^k q_k = \sum_{k=0}^{\infty} z^k \frac{e^{-r} r^k}{k!}$

$= e^{r(z-1)} \cdot \sum_{k=0}^{\infty} \frac{e^{-rz}(rz)^k}{k!} = e^{r(z-1)},$

then multiply $z^k$ to both sides of the balance equations, and add them together. Thus we have:

$\tilde{\pi}(z) = \pi_0[1 + (1 - q_0)^{c-1}(\tilde{q}(z) - 1)] + \sum_{i=1}^{c-1} \pi_i\big[1 + (1 - q_0)^{c-i}(\tilde{q}(z) - 1)\big]$
$+ \sum_{k=0}^{\infty} z^k \sum_{i=c}^{k+c} \pi_i q_{k+c-i}$.

For the last term of the above equation, swap the summation subscripts, so that:

$\sum_{k=0}^{\infty} z^k \sum_{i=c}^{k+c} \pi_i q_{k+c-i}$
$= \sum_{i=c}^{\infty} \sum_{k=i-c}^{\infty} \pi_i q_{k+c-i} z^k$
$= \sum_{i=c}^{\infty} \sum_{j=0}^{\infty} \pi_i q_j z^{i+j-c}$ (Let $j = k + c - i$)
$= z^{-c}\big(\sum_{i=c}^{\infty} \pi_i z^i\big)\big(\sum_{j=0}^{\infty} q_j z^j\big)$
$= z^{-c}\big(\tilde{\pi}(z) - \sum_{i=0}^{c-1} \pi_i z^i\big)\tilde{q}(z)$.

Hence,

$\tilde{\pi}(z) = \pi_0[1 + (1 - q_0)^{c-1}(\tilde{q}(z) - 1)] + \sum_{i=1}^{c-1} \pi_i\big[1 + (1 - q_0)^{c-i}(\tilde{q}(z) - 1)\big]$
$+z^{-c}\big(\tilde{\pi}(z) - \sum_{i=0}^{c-1} \pi_i z^i\big)\tilde{q}(z)$.

So,

$$\tilde{\pi}(z) = \frac{\pi_0\left[1+(1-q_0)^{c-1}\left(e^{r(z-1)}-1\right)-z^{-c}e^{r(z-1)}\right]+\sum_{i=1}^{c-1}\pi_i\left[1+(1-q_0)^{c-i}(e^{r(z-1)}-1)-z^{i-c}e^{r(z-1)}\right]}{1-z^{-c}e^{r(z-1)}}. \qquad \text{(B.1)}$$

We need $c$ equations to solve for $c$ unknowns: $\pi_i (i = 0,1, \dots, c - 1)$ in (B.1). By applying Rouché's theorem (Crommelin, 1932), we show that the denominator $1 - z^{-c}e^{r(z-1)}$ has exactly $c$ zeros in $|z| \leq 1$ (see Section B.1.1 for a proof). Note that by definition, $\tilde{\pi}(z)$ converges for any $z$ such that $|z| \leq 1$, so these $c$ zeros must be the zeros of the numerator as well. One of them is $z_0 = 1$, which is obviously a zero in the numerator; the other $c - 1$ zeros are:

$z_k = -\frac{c}{r} \cdot \text{LambertW}\left(-\frac{r}{c} e^{-\frac{r}{c}+i\frac{2k\pi}{c}}\right)$, for $k = 1, 2, \dots, c - 1$,

where the function LambertW$(\cdot)$ is the inverse function of $\varphi(w) = we^w$. We solve for the $c - 1$ zeros numerically and insert them into the numerator, and obtain $c - 1$ equations:

$\pi_0\big[1 + (1 - q_0)^{c-1}\big(e^{r(z_k-1)} - 1\big) - z_k{}^{-c}e^{r(z_k-1)}\big]$
$+\sum_{i=1}^{c-1} \pi_i\big[1 + (1 - q_0)^{c-i}(e^{r(z_k-1)} - 1) - z_k{}^{i-c}e^{r(z_k-1)}\big] = 0,$ for $k = 1, 2, \dots, c - 1$.

The last equation comes from taking the derivative of (B.1) with respect to $z$, and letting $z = 1$ (note that $\tilde{\pi}(1) = 1$):

$\pi_0[r(1 - q_0)^{c-1} + c - r] + \sum_{i=1}^{c-1} \pi_i\big[r(1 - q_0)^{c-i} + c - i - r\big] = c - r$.

With these $c$ linear equations, it is easy to solve $\pi_i (i = 0,1, \dots, c - 1)$ using numerical tools. With the first $c$ limiting probability values known, there are two ways to obtain the entire limiting probability distribution: (i) since $\tilde{\pi}(z)$ is determined by (B.1), the remaining limiting probabilities can be computed as $\pi_k = \frac{1}{k!}\frac{\mathrm{d}^k\tilde{\pi}(z)}{\mathrm{d}z^k}\Big|_{z=0}$, for $k = c, c + 1, \dots$; or (ii) we can solve for $\pi_c, \pi_{c+1}, \dots$ in turn by iteratively applying the balance equations given at the beginning of this section.

### B.1.1 Proof that $1 - z^{-c}e^{r(z-1)}$ has Exactly $c$ Zeros inside of, or on the Unit Circle of the Complex Plane

Proof: It suffices to show that $z^c - e^{r(z-1)} = 0$ has exactly $c$ zeros inside of, or on the unit circle.

Let $f(z) = z^c$, $g(z) = e^{r(z-1)}$, we want to show that for any $p > 1$ and close enough to 1, $|g(z)| < |f(z)|$ for all $z$ such that $|z| = p$.

We first show that for any $p > 1$ and close enough to 1, $g(p) = e^{r(p-1)} < p^c = f(p)$. This is because $g(1) = f(1) = 1$, and $g'(1) = r < c = f'(1)$, where $r < c$ is a necessary condition for a queueing system to operate in steady-state.

Now let $z = p(\cos\alpha + i\sin\alpha)$,
$|g(z)| = \left|e^{r(p(\cos\alpha + i\sin\alpha)-1)}\right|$
$= \left|e^{r(p\cos\alpha - 1)}\right| \le e^{r(p-1)} < p^c = |f(z)|$.

Since $f(z)$ and $g(z)$ are both holomorphic, Rouché's theorem (Crommelin, 1932) tells us that $f$ and $f - g$ have the same number of zeros inside the circle $|z| = p$, counting multiplicities. Given that $p$ can be arbitrarily close to 1, we conclude that $f$ and $f - g$ have the same number of zeros on and inside of the unit circle $|z| = 1$. Obviously $f(z)$ has exactly $c$ zeros on and inside of the unit circle ($z = 0$ with multiplicity $c$), so $f(z) - g(z) = z^c - e^{r(z-1)}$ has $c$ zeros as well.

### B.2 Limiting Probabilities for $M/G/2/SERIAL$

Similarly, from the transition probabilities given at the end of Appendix A.2, we can write the balance equation in the $z$-domain as:
$\tilde{\pi}(z) = (\pi_0 + \pi_1)\sum_{j=0}^{\infty} P_{0,j} z^j + \sum_{i=2}^{\infty} \pi_i \cdot \sum_{j=i-2}^{\infty} P_{i,j} z^j$.

Let $G(t) = F_S(t) \cdot \int_{h=0}^{\infty} \big(F_S(h+t) - F_S(h)\big) r e^{-rh} \mathrm{d}h$, we have:

$\sum_{j=0}^{\infty} P_{0,j} z^j = P_{0,0} + \sum_{j=1}^{\infty} P_{0,j} z^j$

$= Pr\{M_n = 1 | \tilde{L}_n = 1\} + Pr\{\tilde{L}_{n+1} = 0, M_n = 2 | \tilde{L}_n = 1\}$

$+ \sum_{j=1}^{\infty} Pr\{\tilde{L}_{n+1} = j, M_n = 2 | \tilde{L}_n = 1\} z^j$

$= \int_{t=0}^{\infty} e^{-rt} \mathrm{d}F_S(t) + \sum_{j=0}^{\infty} z^j \int_{t=0}^{\infty} \frac{e^{-rt}(rt)^j}{j!} \mathrm{d}G(t)$

$= \int_{t=0}^{\infty} e^{-rt} \mathrm{d}F_S(t) + \int_{t=0}^{\infty} e^{-rt} \left(\sum_{j=0}^{\infty} \frac{(rt)^j}{j!} z^j\right) \mathrm{d}G(t)$

$= \int_{t=0}^{\infty} e^{-rt} \mathrm{d}F_S(t) + \int_{t=0}^{\infty} e^{rt(z-1)} \mathrm{d}G(t)$.

And,

$\sum_{i=2}^{\infty} \pi_i \cdot \sum_{j=i-2}^{\infty} P_{i,j} z^j$

$= \sum_{i=2}^{\infty} \pi_i \cdot \sum_{j=i-2}^{\infty} z^j \int_{t=0}^{\infty} \frac{e^{-rt}(rt)^{j-i+2}}{(j-i+2)!} \mathrm{d}{F_S}^2(t)$

$= \sum_{i=2}^{\infty} \pi_i \cdot \sum_{k=0}^{\infty} z^{k+i-2} \int_{t=0}^{\infty} \frac{e^{-rt}(rt)^k}{k!} \mathrm{d}{F_S}^2(t)$ (let $k = j - i + 2$)

$= z^{-2} \sum_{i=2}^{\infty} (\pi_i z^i) \cdot \int_{t=0}^{\infty} e^{-rt} \left( \sum_{k=0}^{\infty} z^k \frac{(rt)^k}{k!} \right) \mathrm{d}F_S{}^2(t)$
$= z^{-2} (\tilde{\pi}(z) - \pi_0 - \pi_1 z) \cdot \int_{t=0}^{\infty} e^{rt(z-1)} \mathrm{d}F_S{}^2(t).$

So,

$$\tilde{\pi}(z) = \frac{(\pi_0+\pi_1)\left(\int_{t=0}^{\infty} e^{-rt}\mathrm{d}F_S(t) + \int_{t=0}^{\infty} e^{rt(z-1)}\mathrm{d}G(t)\right) - z^{-2}(\pi_0+\pi_1 z)\int_{t=0}^{\infty} e^{rt(z-1)}\mathrm{d}F_S{}^2(t)}{1 - z^{-2}\int_{t=0}^{\infty} e^{rt(z-1)}\mathrm{d}F_S{}^2(t)}. \quad \text{(B.2)}$$

For any given $F_S(t)$, we can follow the steps introduced in Appendix B.1, i.e., we (i) numerically compute a root of the denominator in the region $|z| \leq 1$; (ii) insert the root into the numerator and let it be zero to obtain an equation for the unknowns $\pi_0$ and $\pi_1$; (iii) take derivative of (B.2) and let $z = 1$ to get a second equation; and (iv) solve these equations for $\pi_0$ and $\pi_1$ to obtain the limiting distribution $\tilde{\pi}(z)$.

## Appendix C

## Calculating the Average Bus Delay in Section 2.4

In Sections C.1 and C.2, we calculate the average bus delay from the previous results for $M/D/c/SERIAL$ and $M/G/2/SERIAL$ systems, respectively.

### C.1 Average Bus Delay for $M/D/c/SERIAL$

When bus service time is deterministic, $\bar{W}_b = 0$, thus we only need to calculate $\bar{W}_q$.

From Appendix A.1, we have:
$Pr\{\tilde{L}_n = 0, \tilde{L}_{n+1} = 0, M_n = k\} = \pi_0(1 - e^{-r})^{k-1}e^{-r}$, for $k = 1,2, \dots c$;
$Pr\{\tilde{L}_n = i, \tilde{L}_{n+1} = 0, M_n = k\} = \pi_i(1 - e^{-r})^{k-i}e^{-r}$, for $k = i, i+1, \dots c$, $1 \leq i \leq c$;
$Pr\{\tilde{L}_n = i, \tilde{L}_{n+1} = j, M_n = c\} = \pi_i P_{i,j}$, for $j \geq 1$.

By inserting the above into Equation (1) in Section 2.4, we obtain:

$$\bar{L}_q = \frac{\frac{1}{2}\pi_0 r^2(1-e^{-r})^{c-1} + \frac{1}{2}\sum_{i=1}^{c-1}\pi_i r^2(1-e^{-r})^{c-i} + \frac{1}{2}\sum_{i=c}^{\infty}\pi_i\left[r^2+2r(i-c)\right]}{r+\pi_0[e^r - r + (r+1-e^r)(1-e^{-r})^{c-1}] - \sum_{i=1}^{c-1}\pi_i(r+1-e^r)(1-(1-e^{-r})^{c-i})}.$$

Hence,

$$\bar{W} = \bar{W}_q = \frac{\bar{L}_q}{\lambda}$$
$$= \frac{1}{\mu} \cdot \frac{\frac{1}{2}\pi_0 r(1-e^{-r})^{c-1} + \frac{1}{2}\sum_{i=1}^{c-1}\pi_i r(1-e^{-r})^{c-i} + \frac{1}{2}\sum_{i=c}^{\infty}\pi_i(r-2c) + \sum_{i=c}^{\infty} i\pi_i}{r+\pi_0[e^r - r + (r+1-e^r)(1-e^{-r})^{c-1}] - \sum_{i=1}^{c-1}\pi_i(r+1-e^r)(1-(1-e^{-r})^{c-i})}$$
$$= \frac{\frac{1}{2}\pi_0 r(1-e^{-r})^{c-1} + \frac{1}{2}\sum_{i=1}^{c-1}\pi_i r(1-e^{-r})^{c-i} + \frac{1}{2}(r-2c)\left(1-\sum_{i=0}^{c-1}\pi_i\right) + \left(\left.\frac{d\tilde{\pi}(z)}{dz}\right|_{z=1} - \sum_{i=1}^{c-1} i\pi_i\right)}{r+\pi_0[e^r - r + (r+1-e^r)(1-e^{-r})^{c-1}] - \sum_{i=1}^{c-1}\pi_i(r+1-e^r)(1-(1-e^{-r})^{c-i})}.$$

In the last equation, we eliminate $\pi_i$ $(i \geq c)$ to avoid computing the numerous limiting probabilities.

### C.2 Average Bus Delay for $M/G/2/SERIAL$

We first calculate $\bar{W}_q$.

From Appendix A.2, we know that:

$Pr\{\tilde{L}_{n+1} = 0, M_n = 1|\tilde{L}_n = 0 \text{ or } 1\} = \int_{t=0}^{\infty} e^{-rt}\mathrm{d}F_S(t);$

$Pr\{\tilde{L}_{n+1} = j \geq 0, M_n = 2|\tilde{L}_n = 0 \text{ or } 1\} = \int_{t=0}^{\infty} \frac{e^{-rt}(rt)^j}{j!}\mathrm{d}G(t);$

$Pr\{\tilde{L}_{n+1} = j \geq i-2, M_n = 2|\tilde{L}_n = i \geq 2\} = \int_{t=0}^{\infty} \frac{e^{-rt}(rt)^{j-i+2}}{(j-i+2)!}\mathrm{d}{F_S}^2(t).$

So,

$$\overline{TL} = \sum_{i,j,k} Pr\{\tilde{L}_n = i, \tilde{L}_{n+1} = j, M_n = k\} \cdot TL_n$$
$$= \sum_{j=0}^{\infty} (\pi_0 + \pi_1) \int_{t=0}^{\infty} \frac{e^{-rt}(rt)^j}{j!}\mathrm{d}G(t)\frac{j(j-1)}{2}$$
$$+ \sum_{i=2}^{\infty} \pi_i \sum_{j=i-2}^{\infty} \int_{t=0}^{\infty} \frac{e^{-rt}(rt)^{j-i+2}}{(j-i+2)!}\mathrm{d}{F_S}^2(t)\frac{(i+j-3)(j-i+2)}{2}$$
$$= (\pi_0 + \pi_1) \int_{t=0}^{\infty} e^{-rt}\left[\sum_{j=0}^{\infty} \frac{(rt)^j}{j!} \cdot \frac{j(j-1)}{2}\right]\mathrm{d}G(t)$$
$$+ \sum_{i=2}^{\infty} \pi_i \int_{t=0}^{\infty} e^{-rt}\left[\sum_{k=0}^{\infty} \frac{(rt)^k}{k!} \cdot \left(\frac{k(k-1)}{2} + (i-2)k\right)\right]\mathrm{d}{F_S}^2(t) \quad \text{(let } k = j-i+2\text{)}$$
$$= (\pi_0 + \pi_1) \int_{t=0}^{\infty} e^{-rt}\left[\sum_{j=0}^{\infty} \frac{(rt)^j}{j!} \cdot \frac{j(j-1)}{2}\right]\mathrm{d}G(t)$$
$$+ \sum_{i=2}^{\infty} \pi_i \int_{t=0}^{\infty} e^{-rt}\left[\sum_{k=0}^{\infty} \frac{(rt)^k}{k!} \cdot \left(\frac{k(k-1)}{2} + (i-2)k\right)\right]\mathrm{d}{F_S}^2(t)$$
$$= (\pi_0 + \pi_1) \int_{t=0}^{\infty} \frac{(rt)^2}{2}\mathrm{d}G(t) + \sum_{i=2}^{\infty} \pi_i \int_{t=0}^{\infty} \left[\frac{(rt)^2}{2} + (i-2)rt\right]\mathrm{d}{F_S}^2(t);$$

$$\bar{A} = \sum_{i,j,k} Pr\{\tilde{L}_n = i, \tilde{L}_{n+1} = j, M_n = k\} \cdot A_n$$
$$= \pi_0 \int_{t=0}^{\infty} e^{-rt}\mathrm{d}F_S(t) \cdot 1 + \pi_0 \sum_{j=0}^{\infty} \int_{t=0}^{\infty} \frac{e^{-rt}(rt)^j}{j!}\mathrm{d}G(t)\,(j+2)$$
$$+\pi_1 \sum_{j=0}^{\infty} \int_{t=0}^{\infty} \frac{e^{-rt}(rt)^j}{j!}\mathrm{d}G(t)\,(j+1) + \sum_{i=2}^{\infty} \pi_i \sum_{j=i-2}^{\infty} \int_{t=0}^{\infty} \frac{e^{-rt}(rt)^{j-i+2}}{(j-i+2)!}\mathrm{d}{F_S}^2(t)\,(j-i+2)$$
$$= \pi_0 \int_{t=0}^{\infty} e^{-rt}\mathrm{d}F_S(t) + \pi_0 \int_{t=0}^{\infty}(rt+2)\mathrm{d}G(t) + \pi_1 \int_{t=0}^{\infty}(rt+1)\mathrm{d}G(t) + \sum_{i=2}^{\infty} \pi_i \int_{t=0}^{\infty} rt\mathrm{d}{F_S}^2(t).$$

According to Equation (1) in Section 2.4, $\overline{W}_q = \frac{\bar{L}_q}{\lambda} = \frac{\overline{TL}}{r\bar{A}}$, where $\overline{TL}$ and $\bar{A}$ are given above.

We then calculate $\overline{W}_b$ in a similar way. Let $TW_b$ be the total wait time in berths for all buses served in a cycle.

$$E[TW_b|\tilde{L}_n = 0 \text{ or } 1]$$
$$= E[TW_b|H_1 > S_1, \tilde{L}_n = 0 \text{ or } 1] \cdot Pr\{H_1 > S_1\}$$
$$+E[TW_b|H_1 < S_1, \tilde{L}_n = 0 \text{ or } 1] \cdot Pr\{H_1 < S_1\}$$
$$= E[0|H_1 > S_1] \cdot Pr\{H_1 > S_1\} + E[\max\{0, S_1 - H_1 - S_2\}|H_1 < S_1] \cdot Pr\{H_1 < S_1\}$$
$$= E[\max\{0, S_1 - H_1 - S_2\}|H_1 < S_1] \cdot Pr\{H_1 < S_1\}$$
$$= E[0|H_1 < S_1 < H_1 + S_2] \cdot Pr\{H_1 < S_1 < H_1 + S_2\}$$
$$+E[S_1 - H_1 - S_2|S_1 > H_1 + S_2] \cdot Pr\{S_1 > H_1 + S_2\}$$
$$= \frac{\int_{h=0}^{\infty}\left[\int_{s=0}^{\infty}\left(\int_{t=0}^{\infty} t\mathrm{d}F_S(h+s+t)\right)\mathrm{d}F_S(s)\right]re^{-rh}\mathrm{d}h}{Pr\{S_1 > H_1 + S_2\}} \cdot Pr\{S_1 > H_1 + S_2\}$$

$= \int_{h=0}^{\infty}\left[\int_{s=0}^{\infty}\left(\int_{t=0}^{\infty} t\mathrm{d}F_S(h+s+t)\right)\mathrm{d}F_S(s)\right] re^{-rh}\mathrm{d}h,$
where $H_1$, $S_1$, and $S_2$ are defined in Appendix A.2.

$E[TW_b|\tilde{L}_n \geq 2] = \frac{1}{2}E[TW_b|\tilde{L}_n \geq 2, S_1 > S_2]$ (by symmetry)
$= \frac{1}{2}E[S_1 - S_2|S_1 > S_2]$
$= \frac{1}{2}\frac{\int_{s=0}^{\infty}\left(\int_{t=0}^{\infty} t\mathrm{d}F_S(s+t)\right)\mathrm{d}F_S(s)}{Pr\{S_1 > S_2\}}$
$= \int_{s=0}^{\infty}\left(\int_{t=0}^{\infty} t\mathrm{d}F_S(s+t)\right)\mathrm{d}F_S(s).$

So,
$\bar{W}_b = \frac{\overline{TW}_b}{E[M_n]} = \frac{\overline{TW}_b}{\bar{A}}$
$= \frac{(\pi_0+\pi_1)E[TW_b|\tilde{L}_n = 0 \text{ or } 1]+(1-\pi_0-\pi_1)E[TW_b|\tilde{L}_n \geq 2]}{\bar{A}},$
where $E[TW_b|\tilde{L}_n = 0 \text{ or } 1]$, $E[TW_b|\tilde{L}_n \geq 2]$, and $\bar{A}$ are given above. The second equality holds because the total number of buses served over a long period equals the total number of arrivals over that same period.

Therefore, $\bar{W} = \bar{W}_q + \bar{W}_b = \frac{\overline{TL}/r+(\pi_0+\pi_1)E[TW_b|\tilde{L}_n = 0 \text{ or } 1]+(1-\pi_0-\pi_1)E[TW_b|\tilde{L}_n \geq 2]}{\bar{A}}$.

## Appendix D

## Derivation of (6) for Uniformly Distributed Service Time, as in Section 3.1

Suppose that the bus service time follows a uniform distribution in $[1-\alpha, 1+\alpha]$ and recall that the mean service time is normalized to 1. Then $C_S = \frac{\alpha}{\sqrt{3}}$. The CDF of the service time is

$$F_S(t) = \begin{cases} 0, & \text{if } t < 1-\alpha \\ \frac{t-(1-\alpha)}{2\alpha}, & \text{if } 1-\alpha < t < 1+\alpha \\ 1, & \text{if } t > 1+\alpha \end{cases}.$$

From Equation (4) in Gu, et al (2011), we calculate the supremum of bus discharge rate from the stop:
$Q(c) = \frac{c}{\int_{t=0}^{\infty}(1-(F_S(t))^c)\mathrm{d}t}$
$= \frac{c}{\int_{t=0}^{1-\alpha} dt + \int_{t=1-\alpha}^{1+\alpha}\left(1-\left(\frac{t-(1-\alpha)}{2\alpha}\right)^c\right)\mathrm{d}t}$
$= \frac{c}{1+\alpha\frac{c-1}{c+1}} = \frac{c}{1+\sqrt{3}C_S\frac{c-1}{c+1}}.$

Therefore, $r = \rho \cdot Q(c) = c\rho/(1+\sqrt{3}C_S\frac{c-1}{c+1}).$

## Appendix E

## Simulation Algorithm for Multi-Berth Stops with Heterogeneous Bus Sizes

The variables used in this simulation are listed as follows:

$S_i$ – Service time of the $i$-th bus ($i = 1, 2, 3, ...$), not including the time that the bus waits to depart the stop after it has finished serving passengers;

$P_i$ – Position (number) of the berth where the $i$-th bus dwells to serve passengers; the number starts from the downstream-most berth; if the current bus is an articulated bus, $P_i$ indicates the berth that is occupied by the second section of the bus;

$M$ – Maximum number of buses;

$H_i$ – Headway between the arrivals of the $(i-1)$-th bus and the $i$-th bus; $H_1$ is the system idle time before the first bus arrives;

$W_{i,q}$ – Waiting time in the queue of the $i$-th bus before it enters the stop;

$W_{i,b}$ – Waiting time in the berth of the $i$-th bus after its service is finished;

$W_i$ – Total waiting time of the $i$-th bus;

$T_i$ – Type of the $i$-th bus, $T_i = 1$ if it is an articulated bus, and 0 if not;

The algorithm of the simulation model is as follows:

Step 1: For a given flow ratio of articulated / regular buses, and distributions of bus service times and headways of the two bus types, generate random variable sequences $\{T_i\}$, $\{S_i\}$ and $\{H_i\}$.

Step 2: Set $i = 1$, $P_i = 1 + T_i$, $W_{i,q} = W_{i,b} = W_i = 0$. Set $i = i + 1$.

Step 3: $W_{i,q} = \begin{cases} \max\ \{0, W_{i-1,q} + S_{i-1} + W_{i-1,b} - H_i\}, & \text{if } P_{i-1} + T_i \geq c \\ \max\ \{0, W_{i-1,q} - H_i\}, & \text{otherwise} \end{cases}$.

Step 4: $P_i = \begin{cases} 1 + T_i, & \text{if } W_{i-1,q} + S_{i-1} + W_{i-1,b} - H_i - W_{i,q} \leq 0 \\ P_{i-1} + 1 + T_i, & \text{otherwise} \end{cases}$.

Step 5: $W_{i,b} = \max\ \{0, W_{i-1,q} + S_{i-1} + W_{i-1,b} - H_i - W_{i,q} - S_i\}$.

Step 6: $W_i = W_{i,q} + W_{i,b}$.

Step 7: If $i = M$, go to Step 9; otherwise, set $i = i + 1$, and go to Step 3.

Step 8: The average bus delay is calculated by averaging $\{W_i\}$. The allowable bus flow is given by the inverse of the mean of bus headway. ■

## References


Cooper, R.B. 1981. *Introduction to Queueing Theory*, second edition. Elsevier North Holland, Inc., New York.

Crommelin, C.D. 1932. Delay probability formulae when the holding times are constant. *Post Office Electr. Engineers J*. **25** 41–50.

Estrada, M., J. Ortigosa, F. Robusté. 2011. Tandem bus stop capacity. TRB 90th Annual Meeting Compendium of Papers, 11-1747. Transportation Research Board of the National Academies, Washington D.C.

Fernandez, R. 2010. Modelling public transport stops by microscopic simulation. *Transportation Res.* **18C** 856-868.

Gibson, J., I. Baeza, L.G. Willumsen. 1989. Bus stops, congestion, and congested bus stops. *Traffic Engineering and Control* **30** 291-302.

Gu, W., Y. Li, M.J. Cassidy, J.B. Griswold. 2011. On the capacity of isolated, curbside bus stops. *Transportation Res.* **45B**, 714-723.

Kendall, D.G. 1953. Stochastic processes occurring in the theory of queues and their analysis by the method of imbedded Markov chains. *The Annals of Mathematical Statistics* **24** 338-354.

Levinson, H.S., K.R. St. Jacques. 1998. Bus lane capacity revisited. *Transportation Res. Record, Journal of Transportation Res. Board* **1618** 189-199.

Maaløe, E. 1973. Approximation formulae for estimation of waiting-time in multiple-channel queueing system. *Management Sci.* **19** 703-710.

Nozaki, S.A., S.M. Ross. 1975. Approximations in multi-server Poisson queues. Report, University of California, Berkeley.

St. Jacques, K., H.S. Levinson. 1997. Operational analysis of bus lanes on arterials. *TCRP Report* 26. Transportation Research Board of the National Academies, Washington D.C.

TRB. 2000. *Highway Capacity Manual*. Transportation Research Board of the National Academies, Washington D.C.

TRB. 2003. *Transit Capacity and Quality of Service Manual*, 2nd Ed. Transportation Research Board of the National Academies, Washington D.C.

Wolff, R.W. 1982. Poisson arrivals see time averages. *Oper. Res.* **30** 223-231.