Modelling user behaviour at a stochastic road traffic bottleneck

Daphne van Leeuwen
Centrum Wiskunde & Informatica (CWI)
Science Park 123
Amsterdam
dleeuwe@cwi.nl

Peter M. van de Ven
Centrum Wiskunde & Informatica (CWI)
Science Park 123
Amsterdam
ven@cwi.nl

ABSTRACT

Congestion in road traffic has received substantial attention in the research literature. One popular approach to modelling congesting and user response is the seminal bottleneck model introduced by Vickrey [25]. Here traffic is modelled as a fluid, and all travellers are subject to cost for waiting, early departure, and late departure. The travellers' response to the congestion is captured by assuming that they arrive at the bottleneck according to a Wardrop equilibrium, meaning that no traveller can decrease its costs by shifting its arrival time. This model and its extensions have been extensively studied in the research literature, but ignore the fact that road traffic consists of individual travellers with uncertain arrival time and speed. While the fluid approach used in the Vickrey model may be correct when the number of travellers is large, it fails to yield accurate predictions for a small number of travellers.

In the present paper we propose a stochastic version of the bottleneck model, that can also handle smaller number of travellers. We discuss the error made by the fluid approximation, and show that the Wardrop equilibrium results in highly varying costs when applied in the more realistic setting with stochasticity. We then discuss an algorithm for numerically computing the equilibrium arrival rate for the stochastic bottleneck model, and propose a closed-form estimation for this equilibrium. This can be used for future studies into the effect of stochasticity in these bottleneck models.

CCS CONCEPTS

• **Applied computing** → **Transportation**; • **Mathematics of computing** → *Markov processes*; • **Networks** → *Network performance modeling*;

KEYWORDS

Road traffic · Queueing theory · Wardrop equilibrium

1 INTRODUCTION

Bottlenecks are a common phenomenon in road traffic, and arise when the rate of traffic arriving into a stretch of road temporarily exceeds its capacity. The resulting congestion causes economic damages and discomfort to the travellers.

Bottlenecks have been extensively studied in the research literature, starting with the seminal work of Vickrey [25], inspired by a morning commute. Traffic is modelled as a fluid, and travellers experience a penalty for waiting at the bottleneck, as well as for arriving at their office earlier or later than intended. Because the morning commute is a recurring and predictable phenomenon, travellers can learn the behavior of others, and eventually adjust their departure time from home to minimize costs. This strategic behavior is modelled in [2, 25] by assuming that traffic arrives according to a Wardrop equilibrium, meaning that no traveller can shift its arrival into the bottleneck without increasing its costs.

This bottleneck model and its variants have been studied extensively in econometrics and transportation literature, and it remains a popular starting point for many recent studies, see, e.g., [2, 21] for an overview. Extensions include demand elasticity [3], which studies the impact of capacity expansion at the bottleneck. The impacts of heterogeneity among travellers is studied by [1] for which they consider multiple classes of travellers with different cost parameters and target times for departing the bottleneck. In [17] the authors study under which heterogeneity assumptions a bottleneck period still exists. More recent studies consider spatial effects [16], endogenous trip timing effects with respect to group arrival times [10], and the relation between parking facilities and congestion [23]. The Wardrop equilibrium can be computed in closed form for a range of these model variants.

In practice road traffic is not a fluid, but instead consists of individual travellers, each of which may have some uncertainty surrounding its arrival time at the bottleneck and its driving speed. The fluid assumption used in the bottleneck literature is accurate when both the number of travellers at the bottleneck and the bottleneck capacity are large, but is inadequate for smaller bottlenecks. To study the effects of variability and the fact that a bottleneck consists of discrete travellers, we modify the traditional deterministic bottleneck model [2] by considering the traffic waiting at the bottleneck as a stochastic process.

While the resulting stochastic model is less tractable than the deterministic bottleneck model, it allows for more detail and accuracy. We first show that the Wardrop equilibrium computed for the deterministic model does not provide equal costs in the stochastic setting, unless the number of travellers is large. We then compute a similar equilibrium concept for the stochastic model, and discuss how it differs from the Wardrop equilibrium. Using these results we

propose a closed-form approximation for the stochastic equilibrium, and show that it performs well.

This work fits in a larger trend towards modelling uncertainty in the bottleneck literature. Most variants of the bottleneck model that include stochasticity do so exogenously, for instance by including some random additional travel time due to an incident [6, 18]. In our study we investigate the impact of endogenous effects, where the uncertainty of the arrival behaviour is included in the model.

The impact of uncertainty over time due to endogeneity is studied in [7–9]. In particular, in [8] the daily demand and capacity are assumed to be random variables from a known distribution. The authors show that under very general assumptions the variance of the delay is increasing in its expectation. This phenomenon has been observed empirically as well in [5], where the authors demonstrate that travel time variance is strongly correlated with the queue length. A paper by [26] adds these effects by increasing the variance of the error term depending on the queue length. However in our model these effects are implicitly included, which confirms the accuracy of our modelling approach.

Beyond the transportation science literature, this paper is closely related to those on the boundary between queueing and game theory. In [11] the authors consider a queueing system with a finite number of customers that must arrive before some time T. Where each customer tries to minimise its waiting time by strategically determining its arrival time. This model has been extended in [13], where the arrival rate is modelled as a non-homogeneous Poisson process and early arrivals are served at random when the facility opens. This is also related to the so-called concert queueing model [15], where customers aim to arrive at some time T, but incur costs for waiting and for tardiness. Various extensions and generalisations have been studied in [14, 15]. Other related models from game theory are the airport boarding game [24] and the meeting game [12]. Our paper is most closely related to [20], where the authors study a similar model to ours, but consider a different equilibrium concept.

The remainder of this paper is structured as follows. First we discuss the traditional deterministic bottleneck model and the Wardrop equilibrium in Section 2. In Section 3 we introduce the stochastic bottleneck model, and show how to numerically compute the equilibrium arrival rate. In Section 4 we use this to propose a closed-form approximation of the stochastic equilibrium. We conclude in Section 5, and outline future research directions.

2 DETERMINISTIC BOTTLENECK MODEL

In this section we provide some background on the deterministic bottleneck model introduced in [2], and describe the Wardrop equilibrium that ensures that all traffic experiences the same costs.

2.1 Model outline

We consider a single bottleneck, with fluid arriving at time t with rate $\lambda(t)$. The fluid represents identical travellers, and the bottleneck can serve a fixed capacity s of traffic per time unit. Each traveller wants to exit the bottleneck at time t^*, and incurs a penalty for *waiting time* in the queue and departing from the queue *earlier* or *later* then the desired time t^*. This penalty is captured by a linear

cost function, with cost coefficients α (waiting), β (early arrival), and γ (late arrival).

Let t_q denote the time of the first arrival, then the cumulative inflow of traffic at the bottleneck up to time t can be written as $a(t) = \int_{t_q}^{t} \lambda(u)\mathrm{d}u$, and the cumulative outflow as $d(t) = s \max\{0, t - t_q\}$ (assuming that the bottleneck only empties once). The sojourn time of a traveller arriving at time t can be computed as $w(t) = a(t) - d(t)$. The cost incurred by an arrival at time t can then be written as (with $(a)^+ = \max\{a, 0\}$):

$$c(t, \lambda) = \alpha w(t) + \beta(t^* - t - w(t))^+ + \gamma(t + w(t) - t^*)^+. \quad (1)$$

Here the $t + w(t) - t^*$ denotes the difference between the departure time of a traveler $t + w(t)$ and its desired departure time t^*. Observe that c depends on λ through the sojourn time w.

2.2 Wardrop equilibrium

Given a total amount of fluid N, we want to find an inflow curve $\lambda_f(t)$ such that no traveller can decrease its costs by altering its arrival time at the bottleneck. It has been shown (see, e.g., [22]) that such a *Wardrop equilibrium* exists, is unique for $\alpha > \beta$, and is given by

$$\lambda_f(t) = \begin{cases} r_1(t - t_q) & t \in [t_q, t_n), \\ r_1(t_n - t_q) + r_2(t - t_n) & t \in [t_n, t_{q'}], \end{cases} \quad (2)$$

where

$$r_1 = s + \frac{\beta s}{\alpha - \beta}, \quad r_2 = s - \frac{\gamma s}{\alpha + \gamma}, \quad (3)$$

$$t_q = t^* - \frac{\eta N/s}{1 + \eta}, \quad t_{q'} = t^* + \frac{N/s}{1 + \eta}, \quad t_n = t^* - \frac{\delta N/s}{\alpha}, \quad (4)$$

with $\eta = \frac{\gamma}{\beta}$ and $\delta = \frac{\beta\gamma}{\beta+\gamma}$. This arrival curve gives all travellers equal costs

$$c_f = \delta \frac{N}{s}, \quad (5)$$

For $\alpha > \beta$ the inflow rate presented in (2) generates a single busy period, i.e., $w(t) > 0$ for all $t \in (t_q, t_{q'})$ [22]. In this equilibrium, the first and last fluid will only incur costs for early/late arrival, and experience no delay. The fluid leaving exactly at the preferred time t^* encounters costs consisting only of delay.

An example of the Wardrop equilibrium is illustrated in Figure 1. This shows the cumulative inflow $a(t)$ (blue), the cumulative outflow $d(t)$ (black), and the waiting time $w(t)$ (red).

3 STOCHASTIC BOTTLENECK MODEL

In practice road traffic is not a perfect fluid, but consists of individual travellers, which each have some uncertainty surrounding their arrival time at the bottleneck and their speed. We capture this uncertainty by assuming discrete travellers that arrive at the queue according to a time-dependent Poisson process with rate $\lambda(t)$. The arrival rate function is such that the expected total number of travellers in N, i.e., $\int \lambda(t)dt = N$. The bottleneck can serve only a single traveller at a time, which takes an independent and identically distributed exponential time with rate $\mu = s$. Such assumptions are not uncommon in this setting, see [4, 14]. Note that our model is equivalent to an $M_t/M/1$ queue.

Similar to the deterministic model, each traveler prefers to exit the bottleneck at time t^*, and incurs a linear penalty α for waiting,
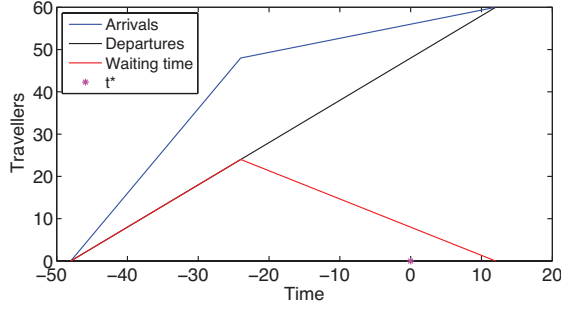
Figure 1: Equilibrium inflow, outflow and waiting time for the deterministic bottleneck model with $T = N/s = 60$, $\alpha = 1$, $\beta = 0.5$, $\gamma = 2$ and $t^* = 0$.

β for arriving early, and γ for tardiness. Let us denote by $W(t)$ the random variable that represents the sojourn time of a traveler arriving at time t, which depends on the past arrival rate through the travellers in the queue upon arrival. The cost function for the stochastic model is identical to that of the deterministic model (1), with the sojourn time replaced by its stochastic counterpart:

$$C(t, \lambda) = \alpha W(t) + \beta(t^* - t - W(t))^+ + \gamma(t + W(t) - t^*)^+. \quad (6)$$

Note that $C(t, \lambda)$ is also a random variable, since it depends on $W(t)$.

A possible equilibrium concept in this situation is the symmetric Nash equilibrium, where we choose the arrival time distribution of each traveler in such a way that none can improve its costs unilaterally, see [20]. Since we are more interested in comparison to the Wardrop equilibrium, we consider the problem of finding an arrival rate such that expected costs $\mathbb{E}C(t, \lambda)$ are the same across the time interval during which arrivals take place:

$$\mathbb{E}C(t, \lambda) = \alpha \mathbb{E}W(t) + \beta \mathbb{E}[(t^* - t - W(t))^+]$$
$$+ \gamma \mathbb{E}[(t + W(t) - t^*)^+]. \quad (7)$$

In contrast to the deterministic model we cannot obtain this equilibrium in closed form, and instead we describe how to compute it numerically in the next sections.

3.1 Computing the expected costs

In order to compute the equilibrium we first compute the expected costs over time for a given arrival rate function $\lambda(t)$. The expected cost of a traveler depends on its sojourn time, which is determined by the queue length upon arrival. Below we describe how to compute the transient queue-length distribution for a given arrival rate $\lambda(t)$, and use this to compute the sojourn time distribution and the expected costs. We assume that there exist some $t_0 < t_1$ such that $\lambda(t) = 0$ outside of $[t_0, t_1]$.

We consider a continuous-time Markov chain representing the number of travellers waiting at the bottleneck. At each state an arrival or departure can take place, except for state 0 in which there is no one waiting. The time-dependent transition matrix $Q(t)$ is

given by

$$Q(t) = \begin{bmatrix} -\lambda(t) & \lambda(t) & 0 & \cdots \\ \mu & -(\lambda(t) + \mu) & \lambda(t) & \cdots \\ 0 & \mu & -(\lambda(t) + \mu) & \lambda(t) \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix}$$

We denote $\bar{\pi}(t) = (\pi_0(t), \pi_1(t), \dots)$ the distribution of the number of travellers waiting at time t. In order to compute this we use uniformisation, where we embed on time instances according to a Poisson process with rate equal to

$$\nu = \sup_t \lambda(t) + \mu, \quad (8)$$

assuming that this supremum exists. Let $\Delta > 0$ and observe that $\frac{(\nu\Delta)^n}{n!} e^{-\nu\Delta}$ denotes the probability that n transitions occur in an interval of length Δ. By conditioning on this we may write

$$\bar{\pi}(t + \Delta) = \bar{\pi}(t) \sum_{n=0}^{\infty} \frac{(\nu\Delta)^n}{n!} e^{-\nu\Delta} P(t)^n, \quad (9)$$

where $P(t)$ denotes the transition probability matrix of the embedded Markov chain given by

$$P(t) = I + \frac{1}{\nu} Q(t). \quad (10)$$

We can then approximate $\bar{\pi}(t)$ by discretizing time into small intervals of length Δ and iterating according to (9), starting from $\bar{\pi}(t_0) = (1, 0, \dots)$.

Having outlined a numerical procedure to obtain the queue-length distribution, we can use this to determine the expected costs over time, by first computing the sojourn time distribution at each time instant t for an arriving traveller.

Let $f(\tau, t)$ denote the density function of sojourn time of a traveller arriving at time t. By conditioning on the number of travellers seen upon arrivals this can be written as

$$f(\tau; t) = \sum_{n=0}^{\infty} \pi_n(t) g_{n+1}(\tau), \quad (11)$$

where $g_n(\tau)$ denotes the sojourn time density given n travellers seen upon arrival, which follows an Erlang-$(n + 1)$ distribution:

$$g_n(\tau) = \frac{\mu(\mu\tau)^n e^{-\tau\mu}}{n!}. \quad (12)$$

In order to obtain the unconditional sojourn time distribution we substitute (9) and (12) into (11).

We are now in position to compute the expected cost incurred by a traveler arriving at time t. In order to do so we evaluate the expected costs (7) by conditioning on the sojourn time of the traveller arriving at time t:

$$\mathbb{E}C(t, \lambda) = \alpha \int_{\tau=0}^{\infty} \tau f(\tau; t) d\tau + \beta \int_{\tau=0}^{(t^*-t)} (t^* - t - \tau) f(\tau; t) d\tau$$
$$+ \gamma \int_{\tau=t^*-t}^{\infty} (t + \tau - t^*) f(\tau; t) d\tau$$

In order to compute the integrals we discretize the sojourn time in small intervals with length Δ to obtain the following approximation:

$$
\mathbb{E}C(t,\lambda) \approx \alpha\Delta \sum_{k=0}^{\infty} k\Delta f(k\Delta;t) + \beta\Delta \sum_{k=0}^{\lfloor(t^*-t)/\Delta\rfloor} (t^*-t-k\Delta)f(k\Delta;t)
$$

$$
+ \gamma\Delta \sum_{k=\lceil(t^*-t)/\Delta\rceil}^{\infty} (t+k\Delta-t^*)f(k\Delta;t). \qquad (13)
$$

	$T = \frac{N}{s}$	t^*	α	β	γ
Set 1	60	0	1	0.5	2
Set 2	60	0	1	0.5	0.5
Set 3	60	0	1	0.75	0.5

Table 1: Parameter sets for numerical experiments.

To illustrate this procedure, we compute the expected costs in the stochastic model for the arrival rate λ_f given by the Wardrop equilibrium (2). In Figure 2 we plot these costs $\mathbb{E}C(t,\lambda_f)$ over time, for each parameter set defined in Table 1. For each parameter set we vary the value of N and s such that N/s remains constant. We see from Figure 2 that $\mathbb{E}C(t,\lambda_f)$ varies significantly between travellers, in particular putting travellers arriving towards the end of the busy period at a disadvantage. This demonstrates that the Wardrop equilibrium is not an accurate equilibrium concept in a realistic setting with stochasticity. As expected, the error is the largest when N is small and disappears as N grows large.

In Figure 3 we plot the decomposition of the costs $\mathbb{E}C(t,\lambda_f)$ into its three components: waiting, early arrival and tardiness. This figure suggests that the large increase in expected costs just before the peak moment is due to the combination of costs for late and early arrival, whereas only early or late costs are encountered in the deterministic model. Moreover, at the end of the bottleneck period in the stochastic model, the queue may not disappear at time $t_{q'}$, giving travellers additional costs $\alpha + \gamma$ for every unit of time spent waiting.
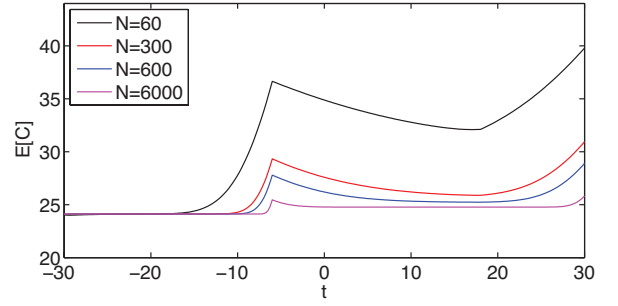
3.2 Stochastic equilibrium

Having demonstrated that the Wardrop equilibrium fails to provide equal costs for all travellers in a stochastic setting, we now present a numerical scheme to determine the equilibrium arrival rate for the stochastic model. That is, we aim to find equilibrium costs c_s and an time-dependent arrival function λ such that $\mathbb{E}C(t,\lambda) = c_s$ for all t with $\lambda(t) > 0$.

Our numerical scheme consists of two phases: first we describe a procedure to obtain an arrival rate λ that satisfies

$$
\mathbb{E}C(t,\lambda) = c \qquad (14)
$$

for any costs $c > 0$. We then scale the arrival rate and the costs to ensure that in expectation N travellers arrive during the bottleneck period. We use the Wardrop equilibrium costs c_f from (5) as a starting point.
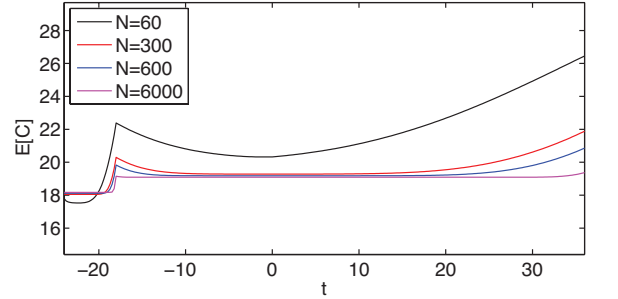
Given target costs c, we can extract the start of the stochastic bottleneck period t_0 using the observation that the first traveller to arrive likely incurs no costs for being late. Instead, the traveller is

(a) Cost function with parameter set 1.
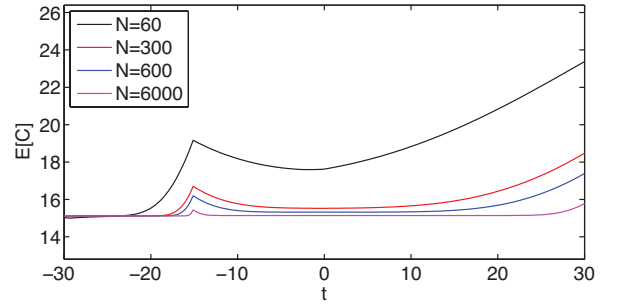
(b) Cost function with parameter set 2.

(c) Cost function with parameter set 3.

Figure 2: Expected costs $\mathbb{E}C(t,\lambda_f)$ for the stochastic model with the arrival rate of the Wardrop equilibrium for different cost parameters function and total number of travellers N.

penalized for being early, and for his or her own service time. We solve

$$
\beta(t^*-t_0-\mathbb{E}W(t_0)) + \alpha\mathbb{E}W(t_0) = c, \qquad (15)
$$

where $\mathbb{E}W(t_0) = 1/s$ as it only depends on the service time duration. Solving this we obtain

$$
t_0 = t^* - \frac{\beta-\alpha+sc}{s\beta}. \qquad (16)
$$

We discretise time into small intervals of length Δ in order to find the time-dependent arrival rate that satisfies (14). We do so iteratively, exploiting the observation that $\mathbb{E}C(t,\lambda)$ only depends on $\lambda(u)$ for $t_0 \leq u \leq t$, due to fact that travellers are served in order
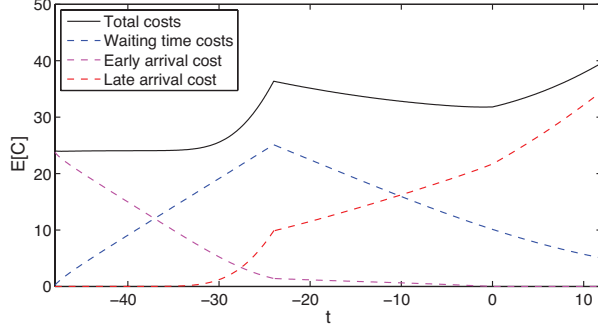
Figure 3: Decomposition of $C(t, \lambda_f)$, with parameter set 1 and $N = 60$.

of arrival. We let $t \geq t_0$ and assume that λ is such that $\mathbb{E}C(u, \lambda) = c$ for all $t_0 \leq u \leq t$. We use this to determine the correct arrival rate for time $t + \Delta$.

In particular, we initialize $\lambda(t + \Delta) = \lambda(t)$ and then adjust the arrival rate by small increments x until we obtain $\mathbb{E}C(u + \Delta, \lambda) = c$ within some small error bound ϵ. The direction of the increments can be obtained from the observation that the cost function is increasing for larger arrival rate $\lambda(t)$. In case of an early arrival the costs change by $\frac{x\Delta}{s}(\alpha - \beta)$, where $\alpha > \beta$. In case of a late arrival the costs change by $\frac{x\Delta}{s}(\alpha + \gamma)$ which is positive as well. We continue this procedure until we first hit a time t_1 such that $\lambda(t_1) = 0$.

The procedure described above yields an arrival rate λ such that $\mathbb{E}C(t) = c$ for all $t \in [t_0, t_1]$, but may not in expectation result in the arrival of N travellers:

$$\Delta \sum_{t=t_0}^{t_1} \lambda(t) = N. \qquad (17)$$

To leverage this procedure to obtain the equilibrium arrival rate for N travellers we modify the target costs c, or equivalently, the bottleneck starting time t_0.

Based on the starting point t_0' and number of travellers N' obtained from an iteration of the algorithm described above, we determine the new starting point by adding the expected service time of the difference in arrivals $(N - N')/s$:

$$t_0 = t_0' - (N - N')/s. \qquad (18)$$

The corresponding equilibrium costs can be computed by (16). We adjust the starting point until $|N - N'| < \epsilon$, for ϵ small. The entire numerical procedure is summarized in pseudo code in Algorithm 1.

Using Algorithm 1 we can compute the stochastic equilibrium arrival rate. We plot this in Figure 4 for parameter set 1 from Table 1 and for various values of N and s, keeping N/s constant. The corresponding Wardrop equilibrium is shown for comparison. From Figure 4a we observe that instead of a sudden transition between the high and low arrival rate seen in the Wardrop equilibrium, the stochastic equilibrium shows a gradual decrease. The smaller the total number of travellers N, the smoother this gradual decrease becomes. In Table 2 the start and end time relative to the corresponding Wardrop equilibrium is shown, as well as the duration of the stochastic equilibrium. Depending on the cost parameters,

---

Algorithm 1 Procedure to obtain stochastic equilibrium
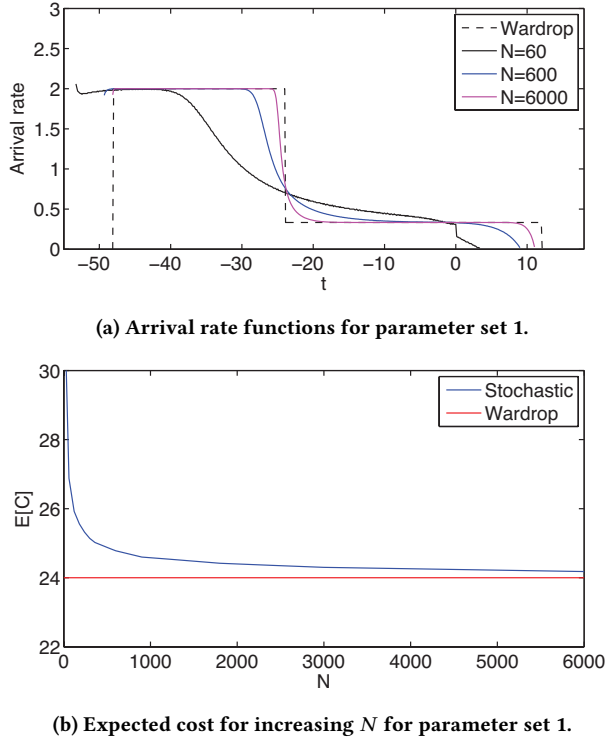
1: **Inputs:**
    $N, s, t^*, \alpha, \beta, \gamma, \epsilon, x$
2: **Initialize:**
    $r_1, r_2, t_q, t_{q'}, t_n, c_f$ from (3) and (4)
    $t_0 \leftarrow t_q$
    $c_s \leftarrow c_f$
    $\lambda(t_0) \leftarrow r_1$
    $N' = 0$
3: **while** $|N' - N| > \epsilon$ **do**
4:     $t \leftarrow t_0$
5:     **while** $\lambda(t) > 0$ **do**
6:         $t \leftarrow t + \Delta$
7:         $\lambda(t) \leftarrow \lambda(t - \Delta)$
8:         **while** $|\mathbb{E}C(t, \lambda(t) - c_s)| > \epsilon)$ **do**
9:             **if** $\mathbb{E}C(t, \lambda(t) - c_s) > 0$ **then**
10:                $\lambda(t) \leftarrow \lambda(t) - x$
11:             **else** $\mathbb{E}C(t, \lambda(t) - c_s) < 0$
12:                $\lambda(t) \leftarrow \lambda(t) + x$
13:             **end if**
14:             obtain $\mathbb{E}C(t, \lambda)$ from (13)
15:         **end while**
16:     **end while**
17:     $N' \leftarrow \Delta \sum_{t=t_0}^{t_1} \lambda(t)$
18:     $t_0 \leftarrow t_0 - (N - N')/s$
19:     $c_s \leftarrow \beta(t^* - t_0 + \frac{1}{s}) + \frac{\alpha}{s}$
20: **end while**
21: $t_1 \leftarrow t$

---

the stochastic bottleneck period can be larger or smaller than the period of the Wardop equilibrium, which is 60 for all three parameter sets. The last column shows the relative increase in costs for the stochastic equilibrium $c_s(N)$ in comparison to the Wardrop equilibrium c_f. We visualised this in Figure 4b for parameter set 1 to show the impact for small values of N, we see that $c_s(N) \rightarrow c_f$ for $N \rightarrow \infty$, as expected. We see empirically that the computation time for tthe stochastic equilibrium grows linearly in N, so there is a trade-off between the computational effort and the benefits from the stochastic equilibrium when N becomes large.

We can use the stochastic equilibrium to investigate the uncertainty over time by plotting the mean waiting time against its standard deviation, see Figure 5. We observe that both the mean and standard deviation of the waiting time increase until peak congestion is reached, after which waiting time decreases but the standard deviation keeps growing. Eventually the standard deviation also decreases as the bottleneck dissappears. This suggests that uncertainty at the end of the bottleneck period has a larger impact than at the beginning. Similar results were shown by Fosgerau in [8] known as the counter-clockwise looping phenomenon also observed in empirical data [5].

(a) Arrival rate functions for parameter set 1.

(b) Expected cost for increasing N for parameter set 1.

Figure 4: Impact of the number of travellers N on the equilibrium arrival rate over time and the impact on the average cost.

	N	$[t_0, t_1] - t_q$	$t_1 - t_0$	$\frac{\mathbb{E}[C]}{c_f}$
Set 1	60	[-5.76 , 51.6]	57.4	12%
	300	[-2.16 , 56.2]	58.3	5%
	600	[-1.44 , 57.4]	58.8	3%
	3000	[-0.48 , 58.8]	59.3	1%
	6000	[-0.24 , 59.3]	59.5	1%
Set 2	60	[-6.72 , 60.7]	67.4	22%
	300	[-2.88 , 60.2]	63.1	10%
	600	[-1.92 , 60.0]	61.9	6%
	3000	[-0.72 , 60.0]	60.7	2%
	6000	[-0.48 , 60.0]	60.5	2%
Set 3	60	[-5.76 , 62.4]	68.2	24%
	300	[-2.40 , 61.0]	63.4	10%
	600	[-1.68 , 60.7]	62.4	7%
	3000	[-0.72 , 60.5]	61.2	3%
	6000	[-0.48 , 60.2]	60.7	2%

Table 2: Comparing the stochastic and Wardrop equilibrium.

4 CLOSED-FORM EXPRESSION OF THE EQUILIBRIUM

The results of Section 4 show the impact of uncertainty over the bottleneck period. Our numerical procedure for computing the
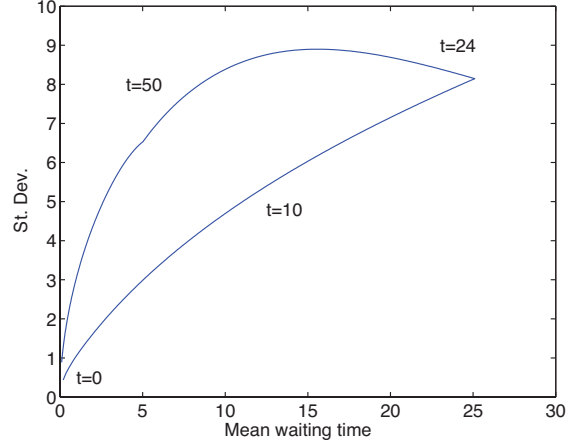
Figure 5: Mean waiting time against its standard deviation over time.

stochastic equilibrium provide useful insights into its behavior, but it lacks the qualitative insights of analytic expressions. In this section we propose a closed-form expression for the stochastic equilibrium.

In Figure 4 we observe that the arrival rate of the stochastic equilibrium shows a gradual decrease between the rate at the beginning and at end of the bottleneck period compared to the instantaneous drop observed in the deterministic model. We propose to approximate this gradual decrease by a sigmoid function. These functions are used in a wide range of fields, for instance in machine learning, biology, and economics.

We use a special case of the Sigmoid function known as the generalised logistic function, which was originally developed as a function to model animal growth [19]. In particular, we choose the following functional form:

$$f(t) = A + \frac{K - A}{(1 + \nu e^{-B(t-M)})^{1/\nu}}, \quad t \in [t_0, t_1]. \quad (19)$$

where A and K are the lower and upper asymptotes respectively, B is the growth rate, $\nu > 0$ represents the symmetry parameter and M defines the point of inflection. We are interested in the period of $[t_0, t_1]$, which indicates the start and end of the bottleneck period.

In order to choose the correct parameter values we draw inspiration from the numerical approximation of the stochastic equilibrium. As we see in Figure 4, the values of the lower and upper asymptotes of the stochastic equilibrium correspond to the lower and upper rate of the Wardrop equilibrium, respectively, and we choose

$$K = r_1, \quad A = r_2. \quad (20)$$

Furthermore, we observe that the inflection point roughly coincides with time t_n of the Wardrop equilibrium, shifted by the difference in starting points of the Wardrop equilibrium and the stochastic equilibrium $t_q - t_0$. This is because the waiting time also starts to decrease at this point. Therefore we set

$$M = t_n - (t_q - t_0). \quad (21)$$

The symmetry parameter v can be related to the ratio between fraction of time the Wardrop equilibrium prescribes the high rate $(t_n - t_q)/(t_{q'} - t_q)$. Multiplying this by the difference in rates of the Wardrop equilibrium we arrive at

$$v = \frac{(r_1 - r_2)}{s} \frac{t_{q'} - t_q}{t_n - t_q}. \tag{22}$$

In contrast to the other parameters, the growth parameter B cannot be readily estimated by relating it to the Wardrop equilibrium, and instead we use nonlinear regression. To this end, first observe from Figure 4 that the steepest descent of the stochastic equilibrium is at the inflection point M, and its derivative at this point k is equal to

$$k := \frac{\mathrm{d}}{\mathrm{d}t} f(t) \Big|_{t=M} = B(K - A)(1 + v e^{-B(t-M)})^{\frac{-1-v}{v}} e^{-B(t-M)} \Big|_{t=M}$$

$$= B(K - A)(1 + v)^{\frac{-1-v}{v}}.$$

The K, A and v can be obtained from (20) and (22), so once we determine k we can compute B as

$$B = \frac{-k(1 + v)^{\frac{-1-v}{v}}}{K - A}. \tag{23}$$

From numerical results we can see that k depends on a combination of the cost parameters α, β and γ, the number of travellers N, and the rate of service s. However this becomes a very complicated function. Therefore, in our regression model we estimate the growth rate B for only a few parameters for the most general form, which is the standard cost value of Set 1 from Table 1. We keep the ratio N/s fixed. Then we adjust the N and s values by taking multiples of 60 for N, ranging from $N \in [60, 3000]$ and $s \in [1, 50]$. For simplicity we divide N by 60 in our regression function. Additionally, we vary $\alpha \in [\beta, \gamma]$ and use the waiting costs expressed as

$$\alpha_{perc} = \frac{\alpha - \beta}{\gamma - \beta}. \tag{24}$$

The resulting values show a linear dependency when a log scaling is applied. To fit our linear regression model, we thus have to solve

$$\log k = a_0 + a_1 log(N/60) + a_2 log(\alpha). \tag{25}$$

We use a least square non-linear regression and obtain the following values for the coefficients $a_0 = -0.0093 \approx 0, a_1 = -1.1896$ and $a_2 = 1.4242 \approx \sqrt{2}$ and with relative residual 0.053. Thus we estimate k by

$$\hat{k} = -\alpha_{perc}^{-1.1896} * (N/60)^{\sqrt{2}}. \tag{26}$$

In Figure 6 we compare $\hat{k}$ with the actual slope k. This slope is computed based on our numerical approximation of the stochastic equilibrium. Figure 6 shows that $\hat{k}$ is a remarkably accurate estimate for k, in particular for small values of N, which is the most relevant regime. Note that the slope estimator $\hat{k}$ is decreasing in N as expected, since as N grows large the stochastic model approaches the deterministic model, where the equilibrium has an instantaneous transition from r_1 to r_2.

By substituting $\hat{k}$ from (26) into (23), along with our estimates for K and A from (20) and v from (22), we obtain an approximation for B.

Having determined all parameters for our approximation of (19), what remains is to find the correct time interval $[t_0, t_1]$ during
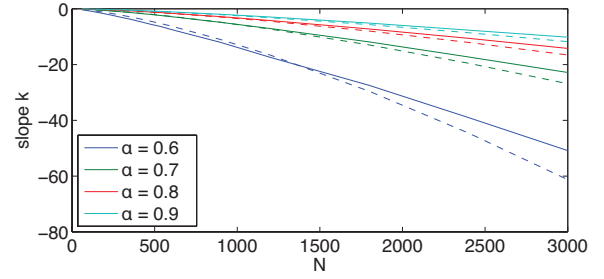
Figure 6: Comparison of the estimator $\hat{k}$ (dashed) against the real values k (solid).

which arrivals occur in the stochastic equilibrium. To this end, we exploit the fact that the expected number of arrivals during the bottleneck duration must add up to N, and that the expected cost throughout must be equal. For simplicity, we do this assuming that $v = 1$, and use this result for general v. Numerically, we find that this approximation works well.

Define

$$F(t) := \int f(t)\mathrm{d}t \Big|_{v=1} = tA + \frac{(K - A)}{B} \log(1 + e^{B(t-M)}), \tag{27}$$

then the fact that the expected number of arriving travellers must equal N can be written as

$$F(t_1) - F(t_0) = N. \tag{28}$$

Since the expected cost in equilibrium $\mathbb{E}[C(t)]$ must be the same throughout the bottleneck duration $t \in [t_0, t_1]$ we have that

$$\mathbb{E}[C(t_0)] = \mathbb{E}[C(t_n - t_q + t_0)]. \tag{29}$$

We can approximate the costs at these two time instances t_0 and $t_n - t_q + t_0$ as follows. Travellers arriving at time t_0 would be the first to enter the system, so its expected sojourn time would be $\frac{1}{s}$ (its own service time), while it would arrive early by an amount of time $\mathbb{E}[t^* - t_0 - X_1]$, where $X_1 \sim \exp(s)$ represents the service time of the traveller. The cost for being late are negligible to the first arrival, so by replacing X_1 by its expectation we can approximate the costs for an arrival at time t_0 as

$$\mathbb{E}[C(t_0)] \approx \frac{\alpha}{s} + \beta(t^* - t_0 + \frac{1}{s}). \tag{30}$$

In the deterministic bottleneck model the travellers arriving at time t_n depart from the bottleneck at exactly time t^*, so they only incur waiting costs. The starting point of the stochastic equilibrium is shifted by $t_0 - t_q$ compared to that of the Wardrop equilibrium, so the costs for travellers arriving at time $t_n + t_0 - t_q$ is dominated by the waiting time, and we approximate $\mathbb{E}[C(t_n - t_q + t_0)] \approx \alpha W(t_n - t_q + t_0)$. In order to approximate the sojourn time at time $t_n - t_q + t_0$ we use that the expected number of arrivals is equal to $F(t_n - t_q + t_0) - F(t_0)$, while the expected service up to that time is $(t_n - t_q)s$. If we also include the service of the traveller itself we obtain

$$\mathbb{E}[C(t_n - t_q + t_0)] \approx \frac{\alpha}{s}(F(t_n - t_q + t_0) - F(t_0) + 1 - (t_n - t_q)s). \tag{31}$$
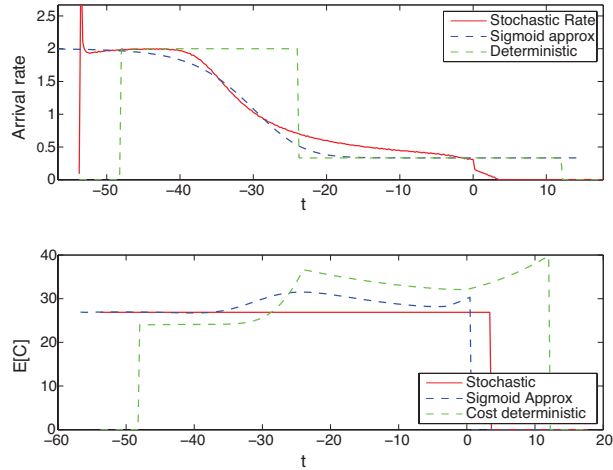
Figure 7: Comparing the Wardrop equilibrium with the analytic and numerical approximation of the stochastic equilibrium by using the parameter values of Set 1 from Table 1 for $N = 60$.

The start and end times of the bottleneck period of the stochastic equilibrium can be obtained by numerically solving t_0 and t_1 from (28) and (31).

We plot our approximation in Figure 7. We can see that the closed form approximation closely follows the stochastic model, and that it does a better job at equalizing the costs among the travellers compared to the Wardrop equilibrium.

5 CONCLUSION

In this paper we presented a general model for predicting the strategic user response to a bottleneck in road traffic. We first reviewed the existing models and results, which rely mostly on deterministic fluid models and Wardrop equilibria. To allow for more realism we proposed to extend these models by considering travellers as discrete entities, which may be subject to randomness. We presented a numerical procedure to compute the equilibrium in this stochastic model, and used these results to find a closed-form approximation for this stochastic equilibrium. Numerically, we showed that this approximation closely follows the equilibrium.

The stochastic bottleneck model gives insight into the effects of strategic arrival behaviour in response to travel times uncertainty. Our approach can be applied to the many extensions that exist of the standard deterministic bottleneck model giving insight into the impact of uncertainty in a broad range of transportation models. Examples include heterogeneity among travellers' departure time, valuation of early and late arrival, and demand elasticity.

REFERENCES

[1] Richard Arnott, André De Palma, and Robin Lindsey. 1987. *Schedule delay and departure time decisions with heterogeneous commuters*. University of Alberta, Department of Economics.
[2] Richard Arnott, André De Palma, and Robin Lindsey. 1990. Economics of a bottleneck. *Journal of urban economics* 27, 1 (1990), 111–130.
[3] Richard Arnott, André De Palma, and Robin Lindsey. 1993. A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *The American Economic Review* (1993), 161–179.
[4] Werner Brilon. 2007. Time dependent delay at unsignalized intersections. In *Proceedings of the 17th International Symposium on Transportation and Traffic Theory (ISTTT17), London, Elsevier*. 555–582.
[5] Chao Chen, Alexander Skabardonis, and Pravin Varaiya. 2003. Travel-time reliability as a measure of service. *Transportation Research Record: Journal of the Transportation Research Board* 1855 (2003), 74–79.
[6] André De Palma and Mogens Fosgerau. 2013. Random queues and risk averse users. *European Journal of Operational Research* 230, 2 (2013), 313–320.
[7] Mogens Fosgerau. 2008. *Congestion costs in bottleneck equilibrium with stochastic capacity and demand*. Technical Report. University Library of Munich, Germany.
[8] Mogens Fosgerau. 2010. On the relation between the mean and variance of delay in dynamic queues with random capacity and demand. *Journal of Economic Dynamics and Control* 34, 4 (2010), 598–603.
[9] Mogens Fosgerau and Anders Karlström. 2010. The value of reliability. *Transportation Research Part B: Methodological* 44, 1 (2010), 38–49.
[10] Mogens Fosgerau and Kenneth A Small. 2017. Endogenous scheduling preferences and congestion. *International Economic Review* 58, 2 (2017), 585–615.
[11] Amihai Glazer and Refael Hassin. 1983. ?/M/1: On the equilibrium distribution of customer arrivals. *European Journal of Operational Research* 13, 2 (1983), 146–150.
[12] Olivier Guéant, Jean-Michel Lasry, and Pierre-Louis Lions. 2011. Mean field games and applications. *Paris-Princeton lectures on mathematical finance* (2011), 205–266.
[13] Refael Hassin and Yana Kleiner. 2010. Equilibrium and optimal arrival patterns to a server with opening and closing times. *IIE Transactions* 43, 3 (2010), 164–175.
[14] Moshe Haviv. 2013. When to arrive at a queue with tardiness costs? *Performance evaluation* 70, 6 (2013), 387–399.
[15] Sandeep Juneja and Rahul Jain. 2009. The concert/cafeteria queueing problem: a game of arrivals. In *Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 1–59.
[16] Raphaël Lamotte and Nikolas Geroliminis. 2017. The morning commute in urban areas with heterogeneous trip lengths. *Transportation Research Procedia* 23 (2017), 591–611.
[17] Gordon F Newell. 1987. The morning commute for nonidentical travelers. *Transportation Science* 21, 2 (1987), 74–88.
[18] Robert B Noland and Kenneth A Small. 1995. Travel-time uncertainty, departure time choice, and the cost of morning commutes. *Transportation research record* 1493 (1995), 150–158.
[19] F J Richards. 1959. A flexible growth function for empirical use. *Journal of experimental Botany* 10, 2 (1959), 290–301.
[20] Eliran Sherzer and Yoav Kerner. 2017. When to arrive at a queue with earliness, tardiness and waiting costs. *arXiv preprint arXiv:1709.03374* (2017).
[21] Kenneth A Small. 2015. The bottleneck model: An assessment and interpretation. *Economics of Transportation* 4, 1 (2015), 110–117.
[22] Michael J Smith. 1984. The existence of a time-dependent equilibrium distribution of arrivals at a single bottleneck. *Transportation science* 18, 4 (1984), 385–394.
[23] Yuki Takayama and Masao Kuwahara. 2016. *Scheduling preferences, parking competition, and bottleneck congestion: A model of trip timing and parking location choices by heterogeneous commuters*. Technical Report. University Library of Munich, Germany.
[24] Rajat Talak, D Manjunath, and Alexandre Proutiere. 2017. Strategic Arrivals to Queues Offering Priority Service. *arXiv preprint arXiv:1704.05986* (2017).
[25] William S Vickrey. 1969. Congestion theory and transport investment. *The American Economic Review* 59, 2 (1969), 251–260.
[26] Shanjiang Zhu, Gege Jiang, and Hong K Lo. 2017. Capturing Value of Reliability through Road Pricing in Congested Traffic under Uncertainty. *Transportation Research Procedia* 23 (2017), 664–678.