Macroscopic modelling and analysis of flows during rush-hour congestion

Dieter Fiems, Balakrishna Prabhu

Macroscopic modelling and analysis of flows during rush-hour congestion

Dieter Fiems[1] Balakrishna J. Prabhu[2]

[1]Ghent University, Dep. TELIN, St-Pietersnieuwstraat 41, 9000 Gent, Belgium
Dieter.Fiems@UGent.be
[2] LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France
balakrishna.prabhu@laas.fr

Abstract

We consider a Markovian queueing model for computing the traffic density and travel times in a city at a macroscopic scale during rush hour. Accounting for the speed/density relation of the macroscopic fundamental diagram of traffic flow, we assume that the service rates of the queueing model at hand are state-dependent. We focus on the fluid limit and obtain a set of differential equations that describe the evolution of the traffic density at the level of neighbourhoods. We also calculate the time-dependent travel times for specific flows in the city and consider the rational time-dependent choice between public and private transport, assuming that there is a congestion-free public alternative to private transportation. Numerical examples reveal that a small reduction in peak traffic can significantly reduce the average travel times.

1 Introduction

Most major urban areas suffer from traffic congestion during the morning and evening commutes. Apart from a waste of time, congestion is a major source of pollution and a health hazard [1]. Several types of remedial measures have been adopted by local authorities to reduce congestion and its impact on the society. These measures take the form of a congestion charge like is the case in cities such as London or Stockholm [2], or, as is done in France and Belgium, a restriction on the entry of vehicles during episodes of peak pollution — either based on the age of the vehicle and its pollution metrics or alternating between even and odd number plates — in certain parts of the city.

Proposing alternative paths is another mechanism for reducing congestion at bottlenecks. This has now become possible in real-time thanks to several online applications that can orient drivers towards the paths which offer the least expected travel times. The users of such applications send travel information to a central server. If there are sufficient users of the application in an area, the aggregation of the collected data yields an accurate assessment of the traffic situation in that area. For each user, optimal paths are then calculated and the corresponding travel advice is returned to the users [3, 4].

A third mechanism is to incentivise public transport. For such schemes to work, the public transport network has to be dense, reliable, and, more importantly, better in terms of travel times and cost. A properly dimensioned public transport network can provide an ecologically friendly alternative to personal vehicle usage. Moreover, public and private transportation can

be successfully combined by park-and-ride systems [5]. A park-and-ride system combines an easily accessible parking area with good connectivity to the public transport network.

For the second and third mechanisms to work, it is important to be able to accurately estimate the travel times between major hubs. If a commuter can see in real-time that it will be faster (or less costly) to travel by public transport, it will give an additional incentive to use the public transport network. Further, if we assume that commuters are rational and choose the alternative with the least cost, we can estimate how many commuters will take the public transport. Such estimates can then be used to dimension the public transport network by appropriately choosing the frequencies and the vehicle sizes on different bus routes.

1.1 Contributions and related work

The main contribution of this paper is a queueing-theory-based method for the computation of travel times for certain flows in road traffic networks. Road traffic modelling has a rich history with models on both macroscopic scales (vehicular flows are modeled as fluids) as well as on microscopic scales (vehicles are discrete entities) [6, 7, 8, 9]. While the link between queueing theory and road traffic networks has been established previously [10, 11, 12], it was only recently that this connection was used to compute travel times at the level of a single neighbourhood [13]. The results were then generalised to networks of neighbourhoods in [14]. This is similar to how Jackson networks [15] or BCMP networks [16] extend single server queueing models.

In this paper, we extend the travel-time analysis of [14] by introducing a richer description of the state of the road traffic network, which allows for separating different flows in the network and for calculating the expected travel times for individual flows. The model in [14] had a single flow with its routing matrix whereas the current paper allows for multiple flows with distinct routing matrices. As in our previous work [13] and [14], based on the travel-time analysis, we also study a public transport game in which commuters can choose between using their own vehicles or taking public transport. The main assumption in this game is that public transport networks are not impacted by congestion and their travel times are thus easier to estimate. This assumption is certainly true for subways and tramways, and is becoming more realistic for buses in cities with dedicated bus corridors. For a given fixed cost of travel by the public transport network, the Wardrop equilibrium is computed in order to determine the fraction of commuters in each neighbourhood that will switch from private vehicles to public transport. One can then compute the influence of pricing on the congestion and, as a consequence, on the travel times and the pollution. In addition, for a certain set of parameters, we also numerically compute the Price of Anarchy (PoA) which is the ratio of the social cost at the Wardrop equilibrium to the optimum social cost. For this, an optimal control problem is first formally defined. Since this problem does not have a closed-form solution, the PoA is obtained by using numerical solvers.

The Wardrop equilibrium concept under study explicitly depends on time. This equilibrium concept differs from the classic Wardrop equilibrium in two important ways. Firstly, the intensity of newly arriving vehicles can vary over time, such that also the decision to opt for public transport varies over time. Secondly, vehicles remain in the neighbourhood for some time, contributing to the vehicle density. This not only means that the time-dependent Wardrop equilibrium relies on a considerably more detailed queueing analysis, but also that public transport decisions at some point in time can be affected by both past and future decisions. Apart from our previous work [13, 14], models with time-dependent demand have been considered in [17, 18, 19]. These authors show the existence of a time-dependent Wardrop equilibrium where the cost of a decision at a point in time depends on the distribution of the load over the complete time-interval under consideration. In particular, the problem is mathematically defined by a cost operator that maps the demand functions on cost functions for the different routes. This operator must be pseudomonotone or hemicontinuous in [17], hemicontinuous in [18], and

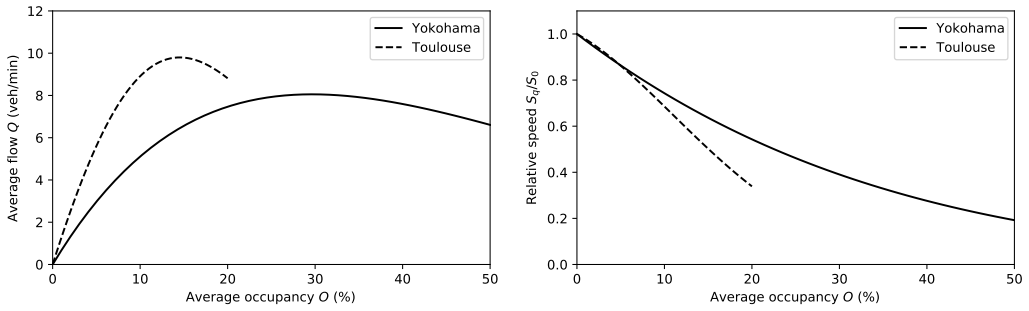


Figure 1: Flow-density (left) and Speed-density (right) macroscopic fundamental diagrams

quasimonotone in [19]. These properties are not easy to verify for our cost operator since, as will be seen later, the mean travel times are obtained as the solution of a set of non-linear differential equations. Therefore, in this paper, we shall assume that the equilibrium exists and satisfies the form in [17].

Finally, we show how to compute a congestion-based cost of the public transport network so that the Wardrop equilibrium matches the social optimum in the fixed cost setting. For time-invariant congestion games, it is known that tolls or taxes can be imposed to achieve a socially optimum behaviour in a game setting (see, e.g., [20]). We derive an analogous result for the time-varying case and for the specific traffic game considered here. This is something that was not done in [13] for the single neighbourhood model.

1.2 The macroscopic fundamental diagram

To compute the travel time of a vehicle in a road traffic network, we shall model the vehicle as a customer in a queueing network, while neighbourhoods will be assigned the role of servers. The specificity of road traffic networks is that the arrival rates of vehicles are time-dependent and the departure rates of vehicles depend on the congestion (or the state) of the network. The higher the congestion, the lower is the rate at which vehicles exit the network. This relationship between the exit rate (or flow in vehicles per minute) and the vehicle density (vehicles per unit length) is captured by what is known as the macroscopic fundamental diagram (MFD) which was first introduced in [21] for single-lane traffic. Here macroscopic refers to the spatial scale of neighbourhoods. This is in contrast to the microscopic fundamental diagram which gives this relationship at the spatial scale of road segments [22, 9, 23]. These macroscopic diagrams have in fact been validated using measurements carried out, for example, in Yokohama [24] and Toulouse [25]. The rescaled and interpolated MFDs for these two cities are shown in Figure 1.

A typical flow-density MFD has an inverted cup shape (like a concave function) as shown in the left pane of Figure 1. When the vehicular density is small (towards the left edge of the curve), the flow initially increases with the density because congestion is yet to form and vehicles are travelling at the maximum permissible speed. As the density increases further, congestion sets in, and the vehicular speed reduces. This leads to an inversion in the average flow which starts to decrease and becomes very small in case of traffic jams. This type of rate curves are not very typical in classic queueing networks.

The flow-density MFD gives the total rate at which vehicles leave the neighbourhood. To compute the travel times, we need the velocity of individual vehicles, which is $\mu(q)/q$. The intuitive explanation of this formula is as follows. Consider a vehicle that enters a small strip of length ϵ. The time it takes for this vehicle to traverse the strip will be the time required

for all the vehicles that were present in this strip at the time of the entry to leave the strip (assuming no overtaking). At density q, there are $q\epsilon$ vehicles, and at an average flow of $\mu(q)$ vehicles/min, it will take $q\epsilon/\mu(q)$ minutes to traverse the strip of length ϵ. This gives a velocity of $\mu(q)/q$. The decrease of the velocities with increasing densities is shown in the speed-density macroscopic diagram in Figure 1. This plot shows the velocity of a vehicle as a function of the density, relative to the free-flow velocity ($q = 0$).

For numerical examples, we shall use the MFD obtained from measurements in the city of Yokohama as the state-dependent service rate of servers.

1.3 Organisation

The rest of the paper is organised as follows. In Section 2, we introduce a Markovian model for vehicles in a road network. A set of partial differential equations for the distribution of the travel times are derived. We then obtain the corresponding equations in the fluid limit when the number of vehicles becomes large. This regime also leads to a simpler set of differential equations for the mean travel times. Section 3 presents the Wardrop equilbrium for the game with public transport as an alternative. The Wardrop equilibrium is then compared to the socially optimal travel choice and the PoA is calculated in the Section 4. We then close with some conclusions in Section 5. Some results are deferred to appendices. In A, we formally prove the fluid limit of Section 2. B then derives the structure of the optimal control policy of Section 4. Finally, C provides in additional illustrative numerical example.

2 Macrosocopic queueing model

We propose a Markovian queueing network with random routing for modelling congestion in a city at a macroscopic level. To be more precise, we model a city as a finite set of interconnected queues, where each queue represents a neighbourhood of the city, the number of vehicles in the neighbourhood being the queue content. In contrast to classic Jackson networks, the service rate of each queue depends on the queue size, to reflect the relation between the speed and density of vehicles as described by the macroscopic fundamental diagram as discussed in section 1.2.

2.1 Markov model

Consider a city and a set of neighbourhoods $\mathcal{N} = \{1, 2, \ldots, N\}$ within that city. A neighbourhood could, in theory, be the entire city itself. Traffic arriving at the neighbourhoods stem from different flows, let $\mathcal{F} = \{1, 2, \ldots, F\}$ denote the set of these flows. For flow $f \in \mathcal{F}$, vehicles arrive at the city in accordance with a Poisson process with time-dependent arrival rate $\lambda^f(t)$. A vehicle from flow f enters the city in neighbourhood $n \in \mathcal{N}$ with probability r_n^f, independent from the other vehicles. The vehicles then move from neighbourhood to neighbourhood till they reach their destination. We introduce the following notation for describing how vehicles are routed through the city. For a vehicle of flow $f \in \mathcal{F}$ in neighbourhood $m \in \mathcal{N}$, let r_{mn}^f denote the probability that it leaves for another neighbourhood n, and let r_{m0}^f denote the probability that the vehicle reaches its destination in neighbourhood m.

The duration of a journey depends on the total distance of the journey as well as on the speed during the journey. The distances the cars travel within each neighbourhood (either to their destination or to another neighbourhood) are assumed to constitute a sequence of independently and identically exponentially distributed random variables with rate 1. With this assumption, vehicles leave with probabilities proportional to their speed. Moreover, taking some other rate is equivalent to rescaling time. The speed of vehicles inside the different neighbourhoods at a

given time instant depends on the average density within the neighbourhood in accordance with the macroscopic fundamental diagram (MFD). Therefore, the number of vehicles that leave the neighbourhood or that reach their destination will also depend on the density, or equivalently, on the number of vehicles in the area. More specifically, as the density in a neighbourhood is proportional to the queue size in that neighbourhood, and the number of departures from the neighbourhood is proportional to the density, the departure rate in neighbourhood m equals,

$$\mu_m(q) = \theta_m F_m(\Theta_m q)\,,$$

when there are q vehicles in neighbourhood m. Here, θ_m and Θ_m are given constants and F_m denotes the macroscopic fundamental diagram of neighbourhood m. As $\mu_m(q)$ is the departure rate of all vehicles in the area, the departure rate of an individual vehicle equals $\mu_m(q)/q$.

We can now study the evolution of the number of vehicles in the different neighbourhoods. Let $Q_n^f(t)$ denote the number of vehicles of flow f (moving around) in neighbourhood n at time t, let $\mathbf{Q}(t) = [Q_n^f(t)]_{n\in\mathcal{N},f\in\mathcal{F}}$, and let $\pi(\mathbf{i};t) = \Pr[Q_n^f(t) = i_n^f, n \in \mathcal{N}, f \in \mathcal{F}] = \Pr[\mathbf{Q}(t) = \mathbf{i}]$ denote the probability that there are i_n^f vehicles of flow f in neighbourhood n at time t, for $n \in \mathcal{N}$ and $f \in \mathcal{F}$, with

$$\mathbf{i} = \left[i_n^f\right]_{n\in\mathcal{N},f\in\mathcal{F}} \in \mathbb{N}^{NF}\,.$$

For further use, for a given state vector $\mathbf{i}$, let i_m denote the number of cars in neighbourhood m,

$$i_m = \sum_{f\in\mathcal{F}} i_m^f\,.$$

In view of the assumptions above, we have the following transitions from state $\mathbf{i} \in \mathbb{N}^{NF}$:

- There is a new arrival of flow $f \in \mathcal{F}$ in neighbourhood $n \in \mathcal{N}$ with rate $\lambda^f(t)r_n^f$. Such an arrival induces a transition to state $\mathbf{i}+\mathbf{e}_n^f$. Here, $\mathbf{e}_m^g = [\mathbb{1}_{\{n=m\}}\mathbb{1}_{\{f=g\}}]_{n\in\mathcal{N},f\in\mathcal{F}}$ is a vector of zeroes, apart from the element corresponding to neighbourhood m and flow g which is one.

- A vehicle of flow f moves from neighbourhood m to neighbourhood n with rate $\mu_m(i_m)(i_m^f/i_m)r_{mn}^f$, the new state being $\mathbf{i}+\mathbf{e}_n^f-\mathbf{e}_m^f$. Here, $\mu_m(i_m)$ is the total outflow from neighbourhood m, the factor (i_m^f/i_m) is the probability that a random car that is leaving from the neighbourhood belongs to flow f, and r_{mn}^f is the probability that this vehicle moves to neighbourhood n. Hence, we implicitly assume that each vehicle in the neighbourhood is equally likely to depart from the neighbourhood (either by arriving at the destination or by leaving for another neighbourhood). In queueing terms, we assume that the "server" uses a processor sharing or random order of service discipline. Since vehicles are moving simultaneously with the same velocity, they can be thought of receiving service simultaneously with the same rate, matching with the processor-sharing discipline.

- Finally, a vehicle of flow f reaches its destination in neighbourhood m with rate $\mu_m(i_m)(i_m^f/i_m)r_{m0}^f$, the new state being $\mathbf{i}-\mathbf{e}_m^f$.

The description of the state transitions then immediately yields the following set of forward Chapmann-Kolmogorov equations,

$$\begin{aligned}\frac{\partial}{\partial t}\pi(\mathbf{i};t) = &\sum_{f\in\mathcal{F}}\sum_{m\in\mathcal{N}}\sum_{n\in\mathcal{N}^*}\pi(\mathbf{i}-\mathbf{e}_n^f+\mathbf{e}_m^f;t)\mu_m(i_m+1)\frac{i_m^f+1}{i_m+1}r_{mn}^f\\ &+\sum_{f\in\mathcal{F}}\sum_{m\in\mathcal{N}}\pi(\mathbf{i}-\mathbf{e}_m^f;t)\lambda^f(t)r_m^f-\sum_{m\in\mathcal{N}}\pi(\mathbf{i};t)\mu_m(i_m)-\sum_{f\in F}\lambda^f(t)\pi(\mathbf{i};t)\,,\end{aligned}$$

with $\mathcal{N}^* = \mathcal{N} \cup \{0\}$, and with $\pi(\mathbf{i};t) = 0$ for $\mathbf{i} \notin \mathbb{N}^{FN}$. In the expression above, we assumed $\mathbf{e}_0^f = \mathbf{0}$ which is in line with our definition above. For Markov chains with non-homogeneous rates, a sufficient condition for the existence of a unique solution to the forward equations is that the transition rates are continuous in time [26]. In the present section, the arrival rates are assumed to be continuous and the MFD is independent of time. However, when the transition rates are not continuous due to, for example, a control mechanism (see Sections 3 and 4), other conditions like measurability of the transition rates and the integrability of the rate of leaving a state need to be verified [26]. In the remainder of this paper, we assume that our control policies satisfy these conditions.

Let $W_n^f(\tau)$ denote the remaining travel time of a vehicle of flow f in neighbourhood n at time τ, and let $\nu_n^f(\tau, t|\mathbf{i}) = \Pr[W_n^f(\tau) > t|\mathbf{Q}(\tau) = \mathbf{i}]$ denote the complementary distribution function of the remaining travel time of this vehicle, conditioned on having $\mathbf{Q}(\tau) = \mathbf{i}$ vehicles of the different flows in the different neighbourhoods at time τ. Conditioning on the possible events in the interval $(\tau, \tau + h]$, we can express $\nu_n^f(\tau, t + h|\mathbf{i})$ as follows,

$$\begin{aligned}
\nu_n^f(\tau, t+h|\mathbf{i}) = &\sum_{g\in\mathcal{F}}\sum_{m\in\mathcal{N}} \nu_n^f(\tau+h, t|\mathbf{i}+\mathbf{e}_m^g)\lambda^g(t) r_m^g h \\
&+ \sum_{g\in\mathcal{F}}\sum_{m\in\mathcal{N}\setminus\{n\}}\sum_{k\in\mathcal{N}^*} \nu_n^f(\tau+h, t|\mathbf{i}-\mathbf{e}_m^g+\mathbf{e}_k^g)\mu_m(i_m)\frac{i_m^g}{i_m} r_{mk}^g h \\
&+ \sum_{g\in\mathcal{F}\setminus\{f\}}\sum_{m\in\mathcal{N}^*} \nu_n^f(\tau+h, t|\mathbf{i}-\mathbf{e}_n^g+\mathbf{e}_m^g)\mu_n(i_n)\frac{i_n^g}{i_n} r_{nm}^g h \\
+ \sum_{m\in\mathcal{N}^*} \nu_n^f(\tau+h, t|\mathbf{i}-\mathbf{e}_n^f+\mathbf{e}_m^f)\mu_n(i_n)\frac{i_n^f-1}{i_n} r_{nm}^f h &+ \sum_{m\in\mathcal{N}} \nu_m^f(\tau+h, t|\mathbf{i}-\mathbf{e}_n^f+\mathbf{e}_m^f)\mu_n(i_n) r_{nm}^f \frac{1}{i_n} h \\
&+ \nu_n(\tau+h, t|\mathbf{i})\left(1 - \sum_{g\in\mathcal{F}}\lambda^g(t)h - \sum_{m\in\mathcal{N}}\mu_m(i_m)h\right) + o(h)\,,
\end{aligned}$$

Note that we again assume that each vehicle in the neighbourhood is equally likely to depart. Sending $h \to 0$, we further have,

$$\begin{aligned}
\frac{\partial}{\partial t}\nu_n^f(\tau, t|\mathbf{i}) - \frac{\partial}{\partial \tau}\nu_n^f(\tau, t|\mathbf{i}) = &\sum_{g\in\mathcal{F}}\sum_{m\in\mathcal{N}} \nu_n^f(\tau, t|\mathbf{i}+\mathbf{e}_m^g)\lambda^g(t) r_m^g \\
&+ \sum_{g\in\mathcal{F}}\sum_{m\in\mathcal{N}\setminus\{n\}}\sum_{k\in\mathcal{N}^*} \nu_n^f(\tau, t|\mathbf{i}-\mathbf{e}_m^g+\mathbf{e}_k^g)\mu_m(i_m)\frac{i_m^g}{i_m} r_{mk}^g \\
&+ \sum_{g\in\mathcal{F}\setminus\{f\}}\sum_{m\in\mathcal{N}^*} \nu_n^f(\tau, t|\mathbf{i}-\mathbf{e}_n^g+\mathbf{e}_m^g)\mu_n(i_n)\frac{i_n^g}{i_n} r_{nm}^g \\
+ \sum_{m\in\mathcal{N}^*} \nu_n^f(\tau, t|\mathbf{i}-\mathbf{e}_n^f+\mathbf{e}_m^f)\mu_n(i_n)\frac{i_n^f-1}{i_n} r_{nm}^f &+ \sum_{m\in\mathcal{N}} \nu_m^f(\tau, t|\mathbf{i}-\mathbf{e}_n^f+\mathbf{e}_m^f)\mu_n(i_n) r_{nm}^f \frac{1}{i_n} \\
&- \nu_n^f(\tau, t|\mathbf{i})\left(\sum_{g\in\mathcal{F}}\lambda^g(t) + \sum_{m\in\mathcal{N}}\mu_m(i_m)\right)\,,
\end{aligned}$$

Finally, the expected travel time of a vehicle that arrives at time τ in neighbourhood n is,

$$\bar{W}_n^f(\tau) = \sum_{\mathbf{i}\in\mathbb{N}^{NF}} \pi(\mathbf{i};t)\int_0^\infty \nu_n^f(\tau, t|\mathbf{i}+\mathbf{e}_n^f)dt\,.$$

Here we used the observation that the mean travel time of a vehicle of flow f arriving at time τ in neighbourhood n while there are $\mathbf{i}$ vehicles of the different flows in the different neighbourhoods equals the mean remaining travel time of a vehicle in neighbourhood n when there are $\mathbf{i} + \mathbf{e}_n^f$ vehicles in the different neighbourhoods.

2.2 Fluid limit

As the fundamental diagram translates to state-dependent service rates, there is no simple analytic solution for the infinite systems of differential equations for $\pi(\mathbf{i}; t)$ and $\nu_n^f(\tau, t|\mathbf{i})$. Moreover, numerically solving these equations is computationally demanding. The process $\mathbf{Q}(t)$ is a time-inhomogenous density dependent population process; see e.g. [27, 28, 29] for recent queueing theoretic applications of such processes. We can study the fluid limit of the system, that is, we consider a sequence of models $\mathbf{Q}^K(t)$, the Kth model having arrival rate $\lambda^f(t)K$ for the fth flow and departure rate $K\bar{\mu}_n(x/K)$ from the nth neighbourhood when there are x vehicles in that neighbourhood. We then study the process $\mathbf{q}(t) = \lim_{K\to\infty} \mathbf{Q}^K(t)K^{-1}$. The parameter K is the scale at which we are making measurements. As the scale becomes large, the number of vehicles also grows proportionally, and hence we have to divide $Q_n^f(t)$ by K to obtain a quantity which goes to a finite limit. This quantity can be thought of as the density of the vehicles. By the functional strong law of large numbers, the randomness of the movement of individual vehicles is washed out by the scaling. This leads to a considerable reduction of the complexity of the analysis since we are no longer dealing with individual cars but instead deal with a large number of cars which behave in a similar way.

Then, by writing the evolution of $Q_n^f(t)$ as the difference of randomly time-changed Poisson processes and applying the functional strong law of large numbers for Poisson processes, we find that the limiting process $q_n^f(t)$ adheres the following set of ordinary differential equations (ode),

$$\dot{q}_n^f(t) = \lambda^f(t) r_n^f - \bar{\mu}_n(q_n(t)) \frac{q_n^f(t)}{q_n(t)} + \sum_{m\in\mathcal{N}, q_m(t)>0} \bar{\mu}_m(q_m(t)) \frac{q_m^f(t)}{q_m(t)} r_{mn}^f \,, \tag{1}$$

for $n \in \mathcal{N}$ and $f \in \mathcal{F}$, and with,

$$q_m(t) = \sum_{f\in\mathcal{F}} q_m^f(t) \,.$$

We formally prove the fluid limit and discuss the conditions on λ_n^f and μ_n in more detail in appendix A.

For the travel time calculations, we scale the arrival process and service process as before. However, we retain the randomness of the random order of service (or processor sharing) discipline. Hence, in the limit, the travel times are still random, let $W_{n,K}^f(\tau)$ denote the travel time from neighbourhood n at time τ for the Kth model, and $w_n^f(\tau, t)$ be the complimentary waiting time distribution in the fluid limit,

$$w_n^f(\tau, t) = \lim_{K\to\infty} \Pr[W_{n,K}^f(\tau) > t] \,.$$

By evaluating the limit above, we obtain the following set of differential equations for these complimentary travel time distributions,

$$\frac{\partial}{\partial t} w_n^f(\tau, t) - \frac{\partial}{\partial \tau} w_n^f(\tau, t) = -w_n^f(\tau, t) \frac{\bar{\mu}_n(q_n(\tau))}{q_n(\tau)} + \sum_{m\in\mathcal{N}} w_m^f(\tau, t) r_{nm}^f \frac{\bar{\mu}_n(q_n(\tau))}{q_n(\tau)} \,. \tag{2}$$

This equation can also be obtained directly by conditioning on whether or not the vehicle under consideration departs, remains in the neighbourhood, or leaves for another neighbourhood in

the interval $(\tau, \tau + h]$. As the remaining travel time at time τ only exceeds $t + h$, if there is no departure in $(\tau, \tau + h]$ and the remaining travel time at time $\tau + h$ exceeds t, we have,

$$w_n^f(\tau, t+h) = \left(1 - \frac{\bar{\mu}_n(q_n(\tau))}{q_n(\tau)} h\right) w_n^f(\tau + h, t) + \sum_{m \in \mathcal{N}} w_m^f(\tau + h, t) r_{nm}^f \frac{\bar{\mu}_n(q_n(\tau))}{q_n(\tau)} h + o(h)\,.$$

Substracting $w_n^f(\tau + h, t)$ from both sides of the equation above, dividing by h and taking the limit $h \to 0$ yields equation (2).

Finally, the mean travel time of a vehicle arriving at time τ in neighbourhood n can then be computed by integrating $w_n(\tau, t)$ over t, that is,

$$\bar{w}_n^f(\tau) = \int_0^\infty w_n^f(\tau, t) dt\,. \tag{3}$$

Integrating both sides of (2) over t, and using (3), we find the following differential equation for the mean travel times,

$$\frac{d}{d\tau} \bar{w}_n^f(\tau) = \bar{w}_n^f(\tau) \frac{\bar{\mu}_n(q_n(\tau))}{q_n(\tau)} - 1 - \sum_{m \in \mathcal{N}} \bar{w}_m^f(\tau) r_{nm}^f \frac{\bar{\mu}_n(q_n(\tau))}{q_n(\tau)}\,. \tag{4}$$

Observe that we do not need to solve (2) to solve (4). The mean travel times can be obtained independently by solving (4) directly.

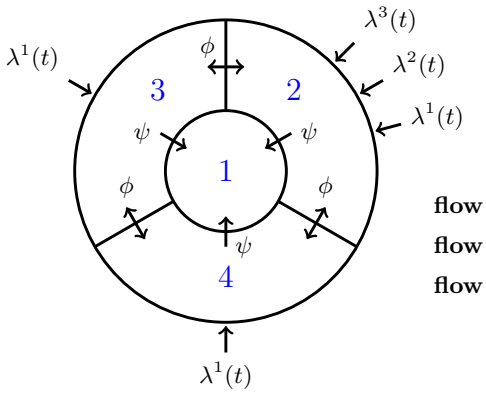
In contrast to the system of differential equations for the traffic densities (1), the differential equation above cannot be used to numerically calculate the time-dependent travel times. Indeed, in order to numerically solve the system of ordinary differential equations, the initial expected travel times $w_n^f(0)$ are needed. Note that these do depend on the future evolution of the traffic densities. To overcome this difficulty, we will reverse time. Assuming that λ_n^f is constant for t sufficiently large, the traffic density in the different neighbourhoods (in the fluid regime) and the corresponding travel times will become approximately constant as well. We therefore choose a sufficiently large T and approximate the mean travel times at time T by the asymptotic mean travel times $w_n^f(T) \approx \lim_{\tau \to \infty} w_n^f(\tau)$. For ease of notation, introduce $\bar{\zeta}_n^f(\tau) = \bar{w}_n^f(T - \tau)$. We then obtain the following system of ordinary differential equations for $\bar{\zeta}_n^f$,

$$\frac{d}{d\tau} \bar{\zeta}_n^f(\tau) = 1 - \bar{\zeta}_n^f(\tau) \frac{\bar{\mu}_n(q_n(T - \tau))}{q_n(T - \tau)} + \sum_{m \in \mathcal{N}} \bar{\zeta}_m^f(\tau) r_{nm}^f \frac{\bar{\mu}_n(q_n(T - \tau))}{q_n(T - \tau)}\,. \tag{5}$$

In contrast to the system of equations (4), the initial value of the ode can now be calculated. Indeed, by assuming that $\lambda_n^f(t)$ is constant and by sending $t \to \infty$, we find that the initial value of the ode solves the linear system of equations,

$$\bar{\zeta}_n^f(0) \frac{\bar{\mu}_n(q_n(T))}{q_n(T)} = 1 + \sum_{m \in \mathcal{N}} \bar{\zeta}_m^f(0) r_{nm}^f \frac{\bar{\mu}_n(q_n(T))}{q_n(T)}\,. \tag{6}$$

Remark 1. The assumptions on the functions λ_n^f do not constrain the problems that can be solved by the method at hand. Indeed, if λ_n^f is not constant for large time values in the time interval under study, it is straightforward to extend the interval such that this is the case. This should not affect the results in the original time interval of interest. The time interval needs to be large enough such that any future traffic patterns do not affect the travel times in the region of interest. For given fixed arrival rates, the differential equations describing the evolution of the densities typically have multiple fixed points. Therefore, one cannot decide on the stationary behaviour by solving a fixed point equation. However, for a given traffic scenario, the solution of the differential equations is unique under mild assumptions (see appendix A), so that there is no problem identifying the fixed point the ode's converge to, if the arrival rates are kept constant for a long time.



flow 1 routing according to the probabilities ϕ, ψ

flow 2 through neighbourhood 2 to neighbourhood 1

flow 3 destination in neighbourhood 2

Figure 2: City with four neighbourhoods. There are arrivals in the outer areas, and only transitions between neighbouring areas, and towards the centre of the city. Three flows are considered, with routing as indicated.

2.3 Numerical example

We now illustrate our approach by a numerical example. We consider a city divided into 4 neighbourhoods as depicted in Figure 2. Traffic only arrives in the outer areas 2 to 4. We consider three distinct flows. The first flow serves as background load for the other flows. Let $\lambda^1(t)$ denote the arrival rate (AR) in area $i \in \{2, 3, 4\}$: $\lambda_2^1(t) = \lambda_3^1(t) = \lambda_4^1(t) = \lambda^1(t)$. Vehicles of this flow leave for the centre neighbourhood 1 with probability $\psi = 0.4$, to any of its neighbours with probability $\phi = 0.1$ or park in the neighbourhood with probability $1 - \psi - 2\phi$. There is no random routing for flows 2 and 3. All vehicles of flow 2 arrive at neighbourhood 2 with rate $\lambda^2(t) = \lambda_2^2(t)$, and cross this neighbourhood to reach their destination in neighbourhood 1. All vehicles of flow 3 arrive at neighbourhood 2 with rate $\lambda^3(t) = \lambda_3^2(t)$ which is their final destination. Finally, departures in each area are governed by the MFD of Yokahama, which is properly rescaled to reflect realistic travel times (around 15 minutes in total for flow 1 when there is no congestion).

Figure 3 shows the evolution of the traffic density and the expected waiting times in the fluid limit regime with arrival rates as depicted. In the left pane, most traffic is background traffic (flow 1). In the right pane, there is still considerable background traffic, but in neighbourhood 2, most of the traffic comes from flows 2 and 3. Notice that the arrival rates in the fluid regime are very small. This is not unexpected: the fluid limit is a deterministic approximation for the city with arrival rates $K\lambda_i^f$, where K is sufficiently large.

On both panes, the arrival rate curve reaches its maximum value at $t = 60$, while the peaks of the traffic density and expected travel times are somewhat later. This is most pronounced for the traffic density in the centre neighbourhood 1. This is expected as all traffic towards the centre needs to cross the outside neighbourhoods, such that peak traffic is delayed. For the travel times, the maximum is reached slightly later than then the peak of the arrival rate, and average travel times typically drop more slowly. Finally, it is readily seen that the maximum traffic density exceeds 30% in some cases. This is typical in real traffic scenarios: at such traffic densities, the average flow decreases for increasing densities, see Figure 1.

To assess the accuracy of the fluid approximation, Figure 4 again shows the time-dependent waiting times, and compares these curves with the expected waiting times in the pre-limit process for different values of K as depicted. We rely on Monte-Carlo simulation to calculate these expected travel times. To be precise, we first simulated a single trajectory for the traffic

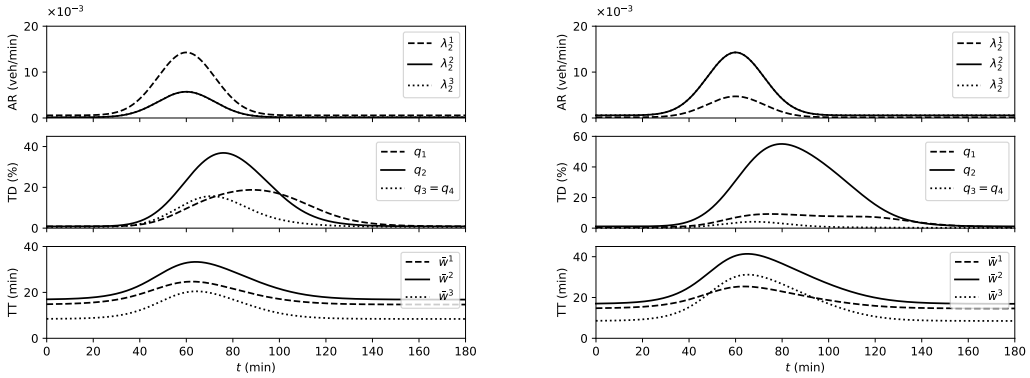


Figure 3: Evolution of the arrival rate (AR) for the flows, the traffic density (TD) in the neighbourhoods and travel times (TT) of the flows during rush hour.

densities of the different flows in the different neighbourhoods. For this trajectory, we then used Monte Carlo simulation to estimate the mean waiting time at various points in time. That is, for each time instant, we sampled the waiting time 5000 times (recall that the waiting time for a given evolution of the traffic densities is a random variable as the vehicle that leaves at a certain instant is randomly selected). We also calculated the 99% confidence intervals for the estimators of the conditional mean waiting times, but omitted these from the figure as their sizes are too small to be properly depicted. We considered 3 distinct scale values K. Practically, the values $K = 1000$, $K = 5000$ and $K = 10000$ correspond to having a peak arrival intensity at the city of 380, 1900 and 3800 vehicles per minute for the rates in the left pane, respectively. The corresponding rates on the right pane are 300, 1500 and 3000 vehicles per minute.

For $K = 1000$, the simulation results considerably deviate from the fluid limit for the curves on the left pane. This is most prominent for the mean travel times for flow 2 and 3. This reflects the remaining randomness in the traffic density process for smaller K, which is most visible for the flows with smaller arrival rates. For $K = 5000$, the correspondence between simulation results and the fluid limit are already good, while the correspondence is very good for $K = 10000$. Even for $K = 10000$, the traffic intensity is not extremely high at the scale of a city. Hence we can conclude that the fluid approximation works in parameter settings of interest.

3 Public transport game

The derivations above not only allow for estimating the travel times of the vehicles in the city. Our results can also be used for dimensioning the capacity of public transport systems, and for assessing the impact of the public transport cost on congestion. In this section, we consider a scenario in which commuters have access to travel time estimates both for commuting by personal vehicles as well as by public transport. If the public transport system (PTS) offers a faster ride, a number of commuters may switch from personal vehicles to the PTS. Assuming commuters are rational, that is, they choose the mode of transport that minimises their expected travel times, we use concepts from game theory to compute the number (or the fraction) of commuters that will switch to the PTS. This can be used to estimate the load on the PTS and dimension its capacity accordingly.

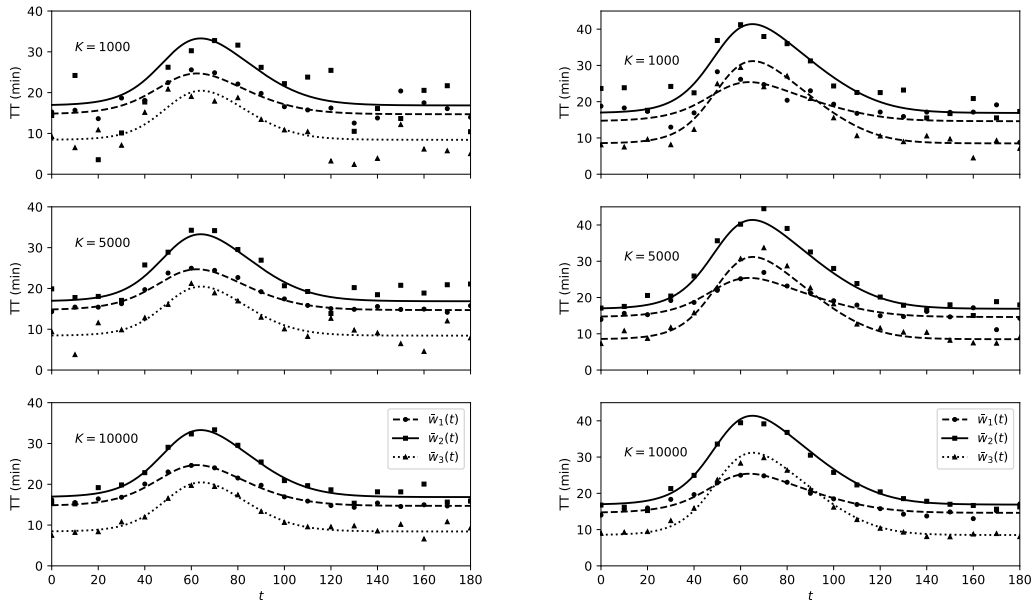


Figure 4: Accuracy of the mean travel times in the fluid approximation for different K as indicated. The mean travel times are calculated for a single sample of the queue size processes.

For simplicity, assume that taking the PTS incurs a cost, which depends on the flow, the neighbourhood and the time of departure, but not on the congestion in the road network. This cost includes the travel time in the public transport system, the travel time to the access points of the public transport system, the inconvenience of using public transport, costs related to possible disruptions of the service, etc. We assume that this cost can be expressed in terms of travel times: the cost of the public transportation system is the maximal travel time one is willing to undergo by private transport. For $n \in \mathcal{N}$ and $f \in \mathcal{F}$, let $C_n^f(t)$ denote the cost of the PTS at time t for a vehicle starting in neighbourhood n.

Remark 2. Note that the assumption that the travel cost does not depend on congestion is natural in case the public transportation system has dedicated lines. This can be either completely separated from the road network like a metro network, or consist of separate lanes embedded in the road network.

3.1 Time-dependent Wardrop equilibrium

We investigate the strategies of rational commuters when the choice of each commuter has a negligible impact on the travel times of the population, a solution concept introduced in game theory by Wardrop [30] in the context of choosing routes in road traffic networks. Known as the Wardrop equilibrium, it says that at the equilibrium choice of routes is such that the cost on each used route is the same and the cost on the unused ones is larger than that of the used ones. In contrast to the classic Wardrop equilibrium, our equilibrium explicitly depends on time. For this *time-dependent Wardrop equilibrium*, the choice at time t not only depends on the choices

of the other commuters at time t, but also on the choices of other commuters (both prior and later than time t).

Let $p_n^f(t) \in [0,1]$ denote the fraction of commuters of flow f entering neighbourhood n that opt for private transport at time t. The function p_n^f is the time-dependent routing strategy for arrivals of flow f in neighbourhood n. For further use, let $p = [p_n^f]_{n\in\mathcal{N},f\in\mathcal{F}}$ be the set of routing strategies for the different flows in the different neighbourhoods. Such a set of routing strategies $\hat{p} = [\hat{p}_n^f]_{n\in\mathcal{N},f\in\mathcal{F}}$ constitutes a Wardrop equilibrium if, for each $n \in \mathcal{N}$ and $f \in \mathcal{F}$ such that $\lambda^f(t)r_n^f > 0$, we have the following constraints on the travel times (see [17], *e.g.*):

$$\begin{cases} \bar{w}_n^f(t;\hat{p}) < C_n^f(t) & \text{then } \hat{p}_n^f(t) = 1\,, \\ \bar{w}_n^f(t;\hat{p}) = C_n^f(t) & \text{then } 0 < \hat{p}_n^f(t) < 1\,, \\ \bar{w}_n^f(t;\hat{p}) > C_n^f(t) & \text{then } \hat{p}_n^f(t) = 0\,. \end{cases} \tag{7}$$

Here $w_n^f(t;\hat{p})$ denotes the expected travel time for flow f in neighbourhood n, assuming that the arrival intensity of flow f in neighbourhood n equals $\lambda^f(t)r_n^f\hat{p}_n^f(t)$. As mentioned in section 1.1, we do not prove the existence and the uniqueness of the equilibrium. It can be seen from (4) that, in general, there is no explicit expression for $\bar{w}_n^f(t;\hat{p})$, and showing properties like quasimonotonicity and/or hemicontinuity is not a trivial task for our cost operator that maps $\hat{p}$ to $\bar{w}$.

The first equation states that all commuters in neighbourhood n prefer private transport if the cost of the PTS exceeds the expected travel time. The second equation states that if a non-zero fraction of commuters opt for PTS and a non-zero fraction of commuters opt for private transport, the cost of both should be equal. Finally, the last equality states that everyone uses the PTS if the cost of private transport exceeds that of the PTS.

To the best of our knowledge, the functions $\hat{p}_n^f$ ($n \in \mathcal{N}$, $f \in \mathcal{F}$) have no nice and easy analytical solution. Therefore, we resort to the following iterative numerical algorithm for their computation. Given the arrival curves $\lambda^f(t)$ and probabilities r_n^f, the fundamental diagrams $\bar{\mu}_n(x)$ and the routing probabilities r_{mn}^f, we set $p_n^f(t) = 1$, and then update these probabilities according to

$$p_n^f(t) \leftarrow \min\left(1, p_n^f(t)e^{-\beta_n(w_n^f(t;p)-C_n^f(t))}\right)\,,$$

for $\lambda^f(t)r_n^f \neq 0$, such that $p_n^f \to \hat{p}_n^f$. Here β_n is a parameter which determines the speed of convergence of the recursion. Small β_n corresponds to conservative updates, such that convergence is slow. In contrast, for large β_n, the update rule may never converge. Finally, note that for $\lambda^f(t)r_n^f = 0$, the value $\hat{p}_n^f(t)$ has no meaning as there is no choice to be made in the absence of traffic.

3.2 Numerical example

We now calculate the Wardop equilibrium for the traffic scenarios of section 2.3 and consider public transport offerings for flows 2 and 3. The scenario on the left pane of Figure 5 corresponds to the scenario on the left pane of Figure 3, but now includes a PTS offering in neighbourhood 2, with a fixed cost of 25 minutes for flow 2 and 15 minutes for flow 3. As before, we depict the arrival rates, the traffic densities and the travel times. For comparison, we depict the arrival rates, traffic densities and travel times without the PTS offering in grey. Similarly, the scenario on the right pane of Figure 5 corresponds to the scenario of the right pane of Figure 3 with a PTS offering in neighbourhood 2, with a cost of 25 and 14 minutes for flow 2 and 3, respectively. For both scenario's, we used the same $\beta_n = 0.15$ for all flows, and observed quick convergence to an approximate solution after about 50 iterations. The more accurate solution which is shown

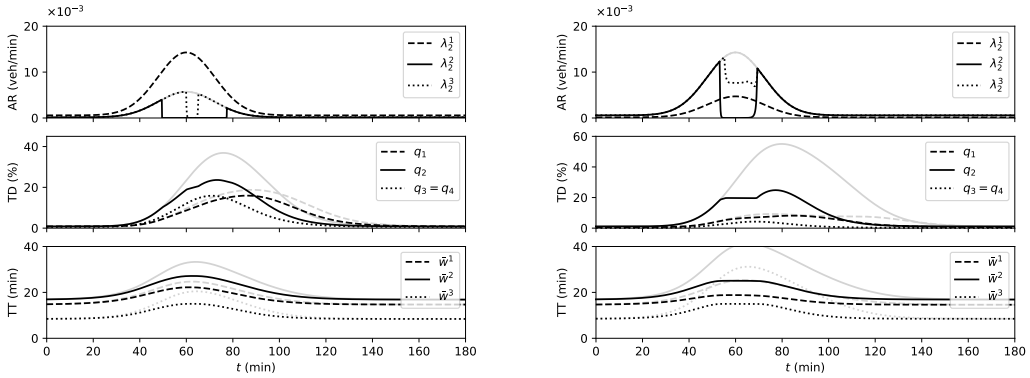


Figure 5: Arrival rate (AR), traffic density (TD) and travel times (TT) during rush hour in Wardrop equilibrium (symmetric public transport costs).

however required considerably more iterations (about 500). When the Wardrop conditions are almost satisfied, the probabilities $\hat{p}_n^f$ are only slightly modified in each iteration, such that many are required.

In both scenarios we can easily verify that the travel times never exceed the PTS cost. Moreover, once the travel times hit the PTS cost, either a fraction of commuters or all commuters of the flow opt for the PTS. This reduction of traffic is not only beneficial for the travel times for these flows. Also the travel time for flow 1 decreases as this flow benefits from the reduction in traffic as well.

4 Social optimum

When commuters choose selfishly between two options, it can lead to inefficiencies in terms of the social cost function defined by taking the average cost over all commuters. Since each commuter is minimising its own cost which is different from that of the other commuters, it fails to sufficiently account for the adverse impact it can have on the other commuters' costs. In our public transport dimensioning problem, there is a natural concept of social cost as well as a central authority (the city hall, for example) that may want to leverage its public transport system to arrive at the optimal social cost. However, implementing the optimal public transport strategy may not be that easy. For example, it may force certain commuters to take public transport which may be unpopular with the electorate or may require upgrading the infrastructure. The central authority may thus have to find a good balance between reducing the social cost while letting the commuters take decisions.

In this section, we investigate two issues. First, we investigate how inefficient the Wardrop equilibrium is. If the inefficiency is small, it may not be worth the effort to look for alternatives to reduce the cost. The concept of Price of Anarchy (PoA) [31] was introduced precisely for quantifying these inefficiencies. It is defined as the ratio of the social cost at the Wardrop equilibrium to that of the socially optimum cost. A higher value of the PoA implies that the inefficiency due to individual decision making is higher, which can be interpreted as a higher price for decentralised decision making.

Second, we investigate how the central authority can adjust the public transport infrastructure (fare, travel times, etc.) so that the Wardrop equilibrium coincides with the social optimum.

In game theory literature, this is known as *mechanism design*. The basic principle is to allow individuals to make decisions but design the game in such a way that the outcome of the game is precisely the social optimum. We shall show how this can be achieved by adjusting the cost of the public transport system during rush hour.

4.1 Optimal control problem

We begin by formulating the optimal control problem that the transport authority has to solve. Let $p_n^f(t)$ be the probability with which a commuter of flow f entering in neighbourhood n uses her vehicle. Then, the mean transportation cost $K_n^f(t)$ at time t of this commuter assuming policy $\mathbf{p}(t) := [p_n^f(t)]_{f,n}$ is

$$K_n^f(t;\mathbf{p}) := \bar{w}_n^f(t;\mathbf{p})p_n^f(t) + (1-p_n^f(t))C_n^f(t)\,. \tag{8}$$

Note that the cost depends upon the entire function $\mathbf{p}$ and not just the function $\mathbf{p}$ evaluated at time t, $\mathbf{p}(t)$. Here, we allow the cost of public transport to depend upon the point of entry and the flow type of the commuter as well as the time of the day. In certain cities (like Paris), the fare depends on the number of subdivisions (called zones) that are traversed by the commuters. A flow can correspond to commuters going between two different zones, for example. Their fare will then depend on their origin and destination zones.

The socially optimal policy is the one that minimises the total weighted mean cost, and is given by

$$\tilde{\mathbf{p}} = \arg\inf_{\mathbf{p}} \int_{T_1}^{T_2} \sum_{n,f} \lambda_n^f(t) K_n^f(t;\mathbf{p})dt \tag{9}$$

where $[T_1, T_2]$ is the rush hour interval. Note that the expression above includes the arrival rates $\lambda_n^f(t)$ as the travel cost impacts more commuters if there are proportionally more arrivals.

With $\hat{\mathbf{p}}(t)$ as the solution of the set of equations (7), the PoA is then defined as,

$$\text{PoA} = \frac{\int_{T_1}^{T_2} \sum_{n,f} \lambda_n^f(t) K_n(t;\hat{\mathbf{p}})dt}{\int_{T_1}^{T_2} \sum_{n,f} \lambda_n^f(t) K_n(t;\tilde{\mathbf{p}})dt}\,.$$

As was the case in [13], (9) is not an optimal control problem for a general $\lambda_n^f(t)$ because $\bar{\mathbf{w}}(0)$ cannot be determined independently of $\mathbf{q}$ (see the discussion below (4)). In order to solve this problem, we shall make the same assumption as was made for obtaining (6). That is, we shall assume that $\lambda_n^f(t)$ goes to a constant as $t \to \infty$. Taking T_2 to be sufficiently large so that the system is stationary, we can then approximate (9) by the solution of following optimal control problem:

$$\min_{\mathbf{p}(t)\in[0,1]} \int_{T_1}^{T_2} \sum_{n,f} \lambda_n^f(t) K_n^f(t;\mathbf{p})dt\,, \tag{10}$$

subject to

$$\frac{d}{dt}q_n^f(t) = \lambda_n^f(t)p_n^f(t) - \bar{\mu}_n(q_n(t))\frac{q_n^f(t)}{q_n(t)} + \sum_{m\in\mathcal{N}} \bar{\mu}_m(q_m(t))\frac{q_m^f(t)}{q_m(t)}r_{mn}^f, \tag{11}$$

$$\frac{d}{dt}\bar{w}_n^f(t) = \bar{w}_n^f(t)\frac{\bar{\mu}_n(q_n(t))}{q_n(t)} - 1 - \sum_{m\in\mathcal{N}} r_{nm}^f \bar{w}_m^f(t)\frac{\bar{\mu}_n(q_n(t))}{q_n(t)}, \tag{12}$$

with the initial values of $q_n(0)$ computed using the steady state solution of (1) and with $\bar{w}(T)$ obtained as the solution of (6).

The optimal control problem (10) can be solved using numerical solvers such as BOCOP [32]. Once the solution is obtained, the PoA can then computed numerically for any set of input parameters (arrival rates, MFD, PTS costs,...). In Section 4.3, we present the solution of (10) and the corresponding PoA for the two scenarios considered in Section 3.2.

4.2 Congestion-based pricing

Now we address the second issue of designing the PTS game in such a way that the Wardrop equilibrium of this new game coincides with the social optimum of the original problem. In the previous subsection, the price (or cost) of public transport was taken to be the same as the one in the game of Section 3. The numerical example in Section 4.3 illustrates that the PoA in this setting is larger than 1. That is, selfish decision making induces inefficiencies from the point of view of the social cost.

One way to remove these inefficiencies is to impose the optimal decisions on individual commuters. However, this forced choice can be seen as too authoritarian. Another method is impose tolls or taxes on certain routes or resources. Here, the individuals still make selfish decisions, but in a game where the costs are modified by tolls. In static congestion games, that is games in which arrival rates are constant, it is known that by introducing appropriate tolls, one can obtain the social optimum in a game setting [20]. In this subsection, we follow this method and compute a congestion-based pricing of public transport such that the equilibrium of the dynamic congestion game with modified costs will be same as the social optimum for the cost setting that was computed in the preceding subsection.

Let $\hat{C}_n^f(t)$ be the congestion-based cost of the public transport network, and let $\hat{p}_n^f(t)$ be the equilibrium probability of flow f at neighbourhood n of the game (7) with the cost C_n^f replaced by $\hat{C}_n^f$. For $\hat{p}_n^f$ to be equal to $\tilde{p}_n^f$ (which is the solution of (9) with cost C_n^f), it needs to satisfy certain conditions that are determined by Pontryagin's maximum principle [33].

When the maximum principle is applied to (9), it follows that (see B)

$$\begin{cases} \bar{w}_n^f(t;\tilde{p}_n^f) < C_n^f(t) - \gamma_n^f(t) & \text{then } \tilde{p}_n^f(t) = 1\,, \\ \bar{w}_n^f(t;\tilde{p}_n^f) = C_n^f(t) - \gamma_n^f(t) & \text{then } 0 < \tilde{p}_n^f(t) < 1\,, \\ \bar{w}_n^f(t;\tilde{p}_n^f) > C_n^f(t) - \gamma_n^f(t) & \text{then } \tilde{p}_n^f(t) = 0\,, \end{cases} \tag{13}$$

where $\gamma_n^f(t)$ is the adjoint (or the co-state) variable for $q_n^f(t)$ and the multiplier of (11).

The mechanism design problem is to construct a PTS cost $\hat{C}_n^f(t)$ for the game so that its equilibrium (7) with PTS cost $\hat{C}_n^f$ is the same as solution of the optimal control problem (13) with the original PTS cost C_n^f.

Comparing (13) and (7), it can be seen that by setting the right-hand side of (13) to $\hat{C}_n^f(t)$, the equilibrium of the modified game can be made the same as the desired optimal control. That is, by setting the PTS cost in the game to be

$$\hat{C}_n^f(t) = C_n^f - \gamma_n^f(t)\,, \quad \forall n \in \mathcal{N}, f \in \mathcal{F}, \tag{14}$$

the social cost at the modified equilibrium will coincide with the social cost of the original optimal control problem. The modified cost for flow f entering neighbourhood n, $\hat{C}_n^f$, is the sum of the original cost C_n^f and a congestion-based cost $\gamma_n^f(t)$. We can interpret $\gamma_n^f(t)$ as the toll (or subsidy if it is positive) that the transport authority should impose to induce a PoA of 1.

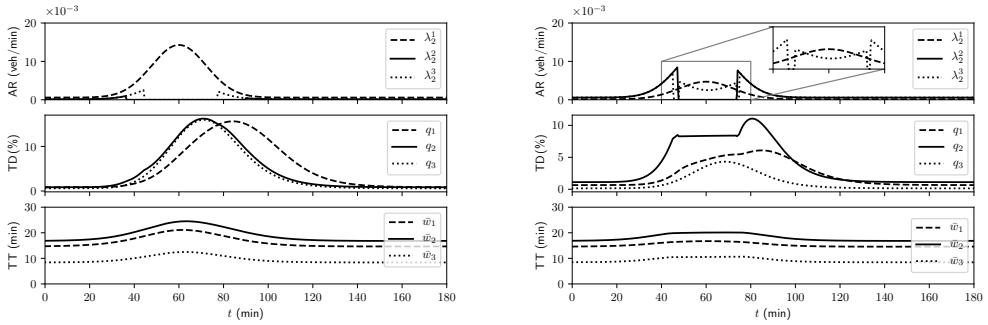


Figure 6: Socially optimal arrival rate (AR), traffic density (TD) and travel times (TT) during rush hour.

4.3 Numerical example

We now give the social optimum policy, PoA, and the congestion-based pricing for the two traffic scenarios of section 3.2. The arrival rate profiles of the flows, the parameters of the MFD, and the routing within the city are the ones described in section 2.3. Recall that in the first scenario, the uncontrolled flow 1 has a much larger volume than the two controlled flows 2 and 3. The PTS cost for flow 2 is $C_2 = 25$, and for flow 3 is $C_3 = 15$. In the second scenario, the controlled flows have a much larger volume compared to the uncontrolled flow. The PTS costs for this scenario are $C_2 = 25$ and $C_3 = 14$. The optimal control problem (10) was solved using the software BOCOP.

The left (resp. right) pane of Figure 6 depicts the effective arrival rates of the three flows in neighbourhood 2 (top figure), the traffic density in the three neighbourhoods (middle figure), and the mean travel times in three neighbourhoods (bottom figure) for scenario 1 (resp. scenario 2). For the second scenario, we add an inset with the effective arrival of flows 1 and 2 since a part of the flow 2 curve is hidden behind that of flow 3. In scenario 1, most of the congestion is due to the uncontrolled flow 1. Since flow 2 traverses two neighbourhoods where flow 1 is present, it starts shifting towards the PTS much earlier than flow 3 which sees flow 1 in only one neighbourhood. In scenario 2, it are flows 2 and 3 that create most of the congestion. Here, the PTS cost of 3 is slightly lower than in scenario 1, so it is not surprising that it moves to PTS earlier. However, surprisingly, once flow 2 shifts to PTS, a fraction of flow 3 goes back to using personal vehicles again since the congestion is now reduced due to the absence of flow 2 vehicles.

The PoA for these particular scenarios was found to be 1.036 and 1.070 for the left and right pane, respectively, that is the average social cost at the Wardrop equilibrium was 3.6% and 7% higher than at the social optimum for the respective scenarios. Here the PoA was calculated over the time interval $[30, 100]$ which covers the congested period.

Finally, Figure 7 shows the congestion-based pricing of (14) for these two scenarios. The adjoint functions $\gamma_2^2(t)$ and $\gamma_2^3(t)$ were computed by BOCOP. For both scenarios, the pricing is in fact a subsidy since it involves reducing the cost of public transport at peak hours. Further, the subsidy is higher during periods of higher travel times (or higher congestion).

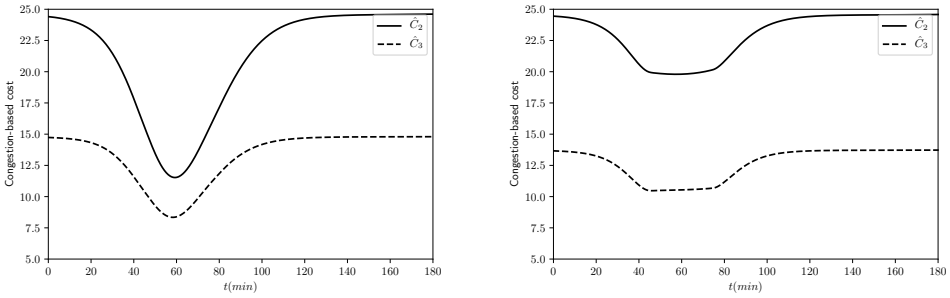


Figure 7: Congestion-based cost function that achieves a PoA of 1.

5 Conclusions

Drawing on queueing network theory and accounting for the macroscopic fundamental diagram of traffic flow, we have proposed a Markovian queueing network model for computing the time-dependent travel times during rush hour in a congested area. The model at hand was characterised by two non-typical properties: queue-dependent departure rates to reflect the macroscopic fundamental diagram and time-dependent arrival intensities which are typical in rush-hour scenarios. As the traffic intensity is high during rush hour, we study the fluid limit of the queueing network model at hand. Numerical experiments show that the travel time can be reduced significantly if a reasonable fraction of commuters switches to public transport. Relying on game theory, we show how our model can be used to assess the impact of public transport systems on congestion and compare the game-theoretical and socially optimal traffic mixes. Finally, a simple congestion-based cost mechanism is proposed that modifies the game and steers its equilibrium to the social optimum of the problem with the original costs.

References

[1] J.I. Levy, J.J. Buonocore, and K. von Stackelberg. Evaluation of the public health impacts of traffic congestion: a health risk assessment. *Environmental Health*, 9(Article 65), 2010.

[2] A. Anas. The cost of congestion and the benefits of congestion pricing: A general equilibrium analysis. *Transportation Research Part B - Methodological*, 136:110–137, 2020.

[3] N. Hoseinzadeh, Y. Liu, L. D. Han, C. Brakewood, and A. Mohammadnazar. Quality of location-based crowdsourced speed data on surface streets: A case study of waze and bluetooth speed data in sevierville, TN. *Computers Environment and Urban Systems*, 83, 2020.

[4] G. de O. Ramos, A. L. C. Bazzan, and B. C. da Silva. Analysing the impact of travel information for minimising the regret of route choice. *Transportation Research Part C - Emerging Technologies*, 88:257–271, 2018.

[5] J. Ortega, J. Toth, T. Peter, and S. Moslem. An integrated model of park-and-ride facilities for sustainable urban mobility. *Sustainability*, 12(11), 2020.

[6] M. Treiber and A. Kesting. *Traffic Flow Dynamics: Data, Models and Simulation*. Springer, Berlin, 2013.

[7] A. Schadschneider, D. Chowdhury, and K. Nishinari. *Stochastic Transport in Complex Systems*. Elsevier, Amsterdam, 2011.

[8] I. Prigogine and R. Herman. *Kinetic theory of vehicular traffic*. Elsevier, 1971.

[9] B. Kerner. *The Physics of Traffic - Empirical Freeway Pattern Features, Engineering Applications, and Theory*. Springer, Berlin, 2004.

[10] N. Vandaele, T. Van Woensel, and A. Verbruggen. A queueing based traffic flow model. *Transportation Research Part D: Transport and Environment*, 5(2):121–135, 2000.

[11] X. Chen, Z. Li, L. Li, and Q. Shi. A traffic breakdown model based on queueing theory. *Networks and Spatial Economics*, 14(3):485–504, Dec 2014.

[12] M. A. A. Boon, R. D. van der Mei, and E. M. M. Winands. Applications of polling systems. *Surveys in Operations Research and Management Science*, 16(2):67–82, 2011.

[13] D. Fiems, B. Prabhu, and K. De Turck. Travel times, rational queueing and the macroscopic fundamental diagram of traffic flow. *Physica A: Statistical Mechanics and its Applications*, 524:412–421, 2019.

[14] D. Fiems and B. Prabhu. Macroscopic modelling and analysis of rush-hour congestion. In *Proceedings of VALUETOOLS '20: 13th EAI International Conference on Performance Evaluation Methodologies and Tools*, Tsukuba Japan, 2020.

[15] J. R. Jackson. Jobshop-like queueing systems. *Management Science*, 10(1):131–142, 1963.

[16] F. Baskett, K. M. Chandy, R. R. Muntz, and F. G. Palacios. Open, closed, and mixed networks of queues with different classes of customers. *Journal of the ACM*, 22(2):248–260, 1975.

[17] P. Daniele, A. Maugeri, and W. Oettli. Time-dependent traffic equilibria. *Journal of Optimization Theory and Applications*, 103(3):543—555, 1999.

[18] A. Nagurney, D. Parkes, and P. Daniele. The internet, evolutionary variational inequalities, and the time-dependent Braess paradox. *Computational Management Science*, 4:355–375, 2007.

[19] D. Aussel and J. Cotrina. Existence of time-dependent traffic equilibria. *Applicable Analysis*, 91(10):1775–1791, 2012.

[20] R. Cole, Y. Dodis, and T. Roughgarden. Pricing network edges for heterogeneous selfish users. In *Proceedings of the Thirty-fifth Annual ACM Symposium on Theory of Computing*, STOC '03, pages 521–530, 2003.

[21] C. F. Daganzo. Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B*, 41:49–62, 2007.

[22] H. Lehmann. Distribution function properties and the fundamental diagram in kinetic traffic flow theory. *Phys. Rev. E*, 54:6058–6064, Dec 1996.

[23] R. Kühne and N. H. Gartner, editors. *Transportation Research Circular E-C149 (75 Years of the Fundamental Diagram for Traffic Flow Theory: Greenshields Symposium)*. Transportation Research Board, 2011.

[24] N. Geroliminis and C. F. Daganzo. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B*, 42:759–770, 2008.

[25] C. Buisson and C. Ladier. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. *Transportation Research Record*, 2124:127–136, 2009.

[26] L. Ye, X. Guo, and O. Hernández-Lerma. Existence and regularity of a nonhomogeneous transition matrix under measurability conditions. *Journal of Theoretical Probability*, 21(3):604–627, Sep 2008.

[27] D. Mukherjee, S. Dhara, S.C. Borst, and J. S. H. van Leeuwaarden. Optimal service elasticity in large-scale distributed systems. In *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, volume 1, pages 1–28, 2017.

[28] I. A. Horváth, Z. Scully, and B. Van Houdt. Mean field analysis of join-below-threshold load balancing for resource sharing servers. In *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, volume 3, 2019.

[29] I. Van Spilbeeck and B. Van Houdt. On the impact of job size variability on heterogeneity-aware load balancing. *Annals of Operations Research*, 293:371–399, 2020.

[30] J. G. Wardrop. Some theoretical aspects of road traffic research. *ICE Proceedings: Engineering Divisions*, 1(3):325–362, 1952.

[31] E. Koutsoupias and C. H. Papadimitriou. Worst-case equilibria. In *Proceeding of STACS*, 1999.

[32] Team Commands, Inria Saclay. BOCOP: an open source toolbox for optimal control. `http://bocop.org`, 2017.

[33] H. P. Geering. *Optimal Control with Engineering Applications*. Springer, 2007.

[34] S. N. Ethier and T. G. Kurtz. *Markov processes, characterization and convergence*. Wiley-Interscience, 1986.

A Fluid limit

For completeness, we here formally state and prove the fluid limit result. To this end, note that the process $\mathbf{Q}(t)$ can be expressed in terms of randomly time-changed Poisson processes

$$Q_n^f(t) = A_n^f\left(\int_0^t \lambda^f(u) r_n^f du\right) - \sum_{m\in\mathcal{N}_*} D_{nm}^f\left(\mu_n(Q_n(u))\frac{Q_n^f(u)}{Q_n(u)} r_{nm}^f\right) + \sum_{m\in\mathcal{N}} D_{mn}^f\left(\mu_m(Q_m(u))\frac{Q_m^f(u)}{Q_m(u)} r_{mn}^f\right).$$

Here the processes A_n^f and D_{nm}^f are unit rate Poisson processes, counting the number of arrivals of flow f in neighbourhood n, and the number of departures of flow f from neighbourhood n to m, respectively.

The stochastic process above is a time-inhomogenous density-dependent population process. A fluid limit result for this type of processes is provided in Theorem 2.1 of [34, Chapter 11],

when the jump rates do not depend on time. We here adapt this theorem to allow for jump rates that depend on both time and density. The fluid result for the traffic densities is then an easy corollary.

Theorem 1. *Consider a sequences of processes in* $\mathbb{R}^d$*,*

$$X_K(t) = X_K(0) + \sum_\ell \ell K^{-1} Y_\ell \left(K \int_0^t \beta_\ell \left(X_K(s), s \right) ds \right) , \tag{15}$$

defined in terms of unit rate Poisson processes Y_ℓ *that count state jumps* $\ell \subset \mathcal{L} \in \mathbb{R}^d$*, with* $|\mathcal{L}| < \infty$*. Further, suppose that the rates* $\beta_\ell(x,t)$ *are bounded by* $\bar{\beta}_\ell$*, and that for* $F(x,t) = \sum_\ell \ell \beta_\ell(x,t)$ *we have the Lipschitz condition,*

$$|F(x,t) - F(y,t)| \leq M|x-y| \,, \quad x, y \in \mathbb{R}^d \,, 0 \leq t \leq T \,.$$

Suppose that X_n *satisfies,* $\lim_{n\to\infty} X_n(0) = x_0$*, and* x *satisfies*

$$x(t) = x_0 + \int_0^t F(x(s), s) ds \,. \tag{16}$$

Then, for every $0 \leq t \leq T$*,*

$$\lim_{K\to\infty} \sup_{0\leq t\leq T} |X_K(t) - x(t)| = 0 \quad a.s. \tag{17}$$

Proof. First note that the Lipschitz condition ensures that the solution of (16) is unique by the Picard-Lindelöf theorem. With $\widehat{Y}_\ell(t) = Y_\ell(t) - t$, subtracting (16) from (15) gives,

$$X_K(t) - x(t) = X_K(0) - x_0 + \sum_\ell \ell K^{-1} \widehat{Y}_\ell \left(K \int_0^t \beta_\ell \left(X_K(s), s \right) ds \right) + \int_0^t \left(F\left(X_K(s), s \right) - F(x(s), s) \right) ds \,,$$

such that for $t \leq T$,

$$|X_K(t) - x(t)| \leq |X_K(0) - x_0| + \eta_K + \int_0^t M |X_K(s) - x(s)| ds \,. \tag{18}$$

with,

$$\eta_K = \sum_\ell |\ell| K^{-1} \sup_{t\leq T} \left| \hat{Y}_\ell \left(K \int_0^t \beta_\ell(X_K(s), s) ds \right) \right|$$

By Gronwall's inequality, (18) implies for $t \leq T$,

$$|X_K(t) - x(t)| \leq \left(|X_K(0) - x_0| + \eta_K \right) \exp(t)$$

or,

$$\sup_{t\leq T} |X_K(t) - x(t)| \leq \left(|X_K(0) - x_0| + \eta_K \right) \exp(T) \,. \tag{19}$$

Moreover, by the bounds on β_ℓ, η_K is bounded by

$$\eta_K \leq \sum_\ell |\ell| K^{-1} \sup_{t\leq T} \left| \hat{Y}_\ell \left(K \bar{\beta}_\ell t \right) \right| \,.$$

The functional strong law of large numbers for Poisson processes then yields $\sup_{t\leq T} \left| \hat{Y}_\ell \left(K \bar{\beta}_\ell t \right) \right| \to 0$ for $K \to \infty$, which shows that $\eta_K \to 0$ for $K \to \infty$ as well. As $|X_K(0) - x_0| \to 0$ a.s. for $K \to \infty$ by assumption, (17) follows from (19).

□

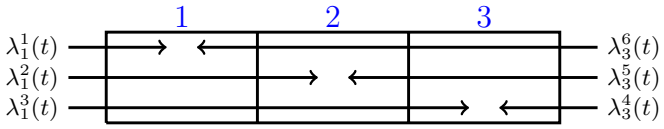


Figure 8: City with three neighbourhoods. There are arrivals in areas 1 and 3, with destinations in all different areas.

Corollary 1. *Assume that (i) that the arrival rates $\lambda^f(t)$ of the different flows are bounded; (ii) that the service rates $\mu_n(q_n)$ are bounded in all neighbourhoods, and (iii) that the service rate per flow $\mu_n(q_n)q_n^f/q_n$ is Lipschitz continuous. Then, with the notation of section 2.2, we have,*

$$\lim_{K\to\infty} \sup_{t\le T} |K^{-1}\mathbf{Q}^K(t) - \mathbf{q}(t)| = 0\,, \quad a.s.$$

B Derivation of (13)

The optimal control problem (10) can be rewritten compactly as

$$\min_{\mathbf{p}(t)\in[0,1]} \int_{T_1}^{T_2} \sum_{n,f} \lambda_n^f(t) K_n^f(t;\mathbf{p}) dt \tag{20}$$

subject to

$$\dot{\mathbf{x}}(t) = g(\mathbf{x},\mathbf{p}) \tag{21}$$

where $\mathbf{x} = [q_n^f, w_n^f]_{n,f}$ is the state of the system and $\mathbf{p}$ is the control variable.

Recall that $\gamma_n^f(t)$ is the co-state vector for $q_n^f(t)$. The Pontryagin Maximum Principle states that if $\mathbf{x}^*$ and γ^* is an optimal trajectory then

$$\mathbf{p}^* = \inf_{\mathbf{p}} \mathcal{H}(\mathbf{x},\mathbf{p},\gamma), \tag{22}$$

where

$$\mathcal{H} = \gamma \cdot g + \sum_{n,f} \lambda_n^f(t) K_n^f(t;\mathbf{p}) \tag{23}$$

is the Hamiltonian.

Since both the running cost (10) as well as the dynamics ((11) and (12)) are linear in p_n^f, so is the Hamiltonian. Substituting for the g and K, we obtain the coefficient of p_n^f in $\mathcal{H}$ as:

$$\lambda_n^t(t) \cdot (\bar{w}_n^f(t) - C_n(t) + \gamma_n^f(t)). \tag{24}$$

That is at each time t, we are solving a linear optimization problem to determine $\mathbf{p}^*(t)$. Since the problem is one of minimization, we set p_n^f to 1 if the (24) is less than 0, and p_n^f to 0 if (24) is positive. This reasoning gives us (13).

C Additional numerical example

To further illustrate our approach, we present a second numerical example. Consider a city divided into 3 neighbourhoods in tandem as depicted in Figure 8, with arrivals in neighbourhoods

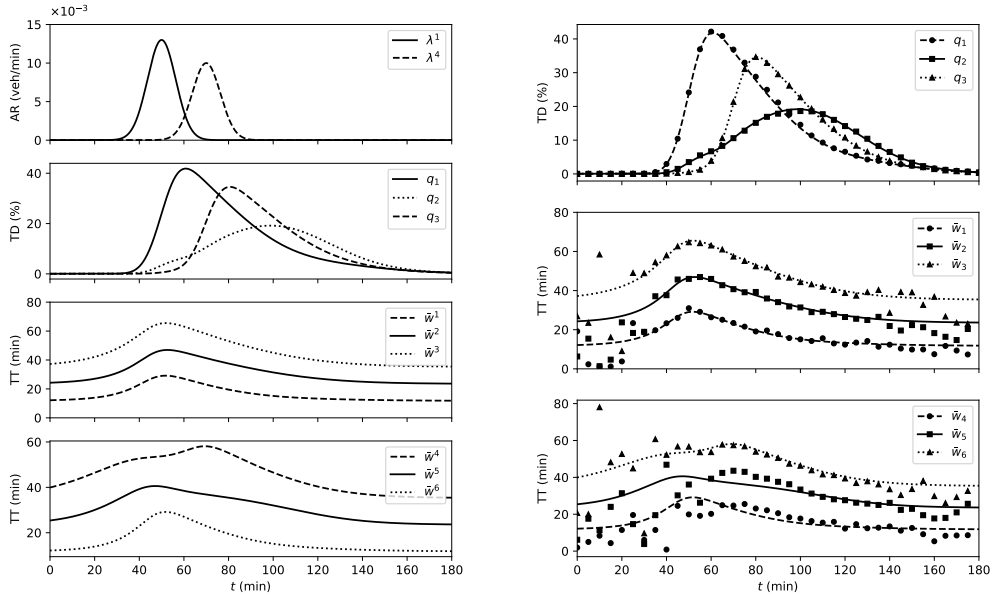


Figure 9: Arrival rate (AR), traffic density (TD) and travel times (TT) during rush hour without public transport alternative. Fluid limit (left). Pre-limit (right).

1 and 3. There are separate flows for each neighbourhood, such that the total number of flows equals 6. We again assume that the departures in the different neighbourhoods are governed by the MFD of Yokohama, as used in the preceding example, which is properly rescaled to reflect realistic travel times (around 15 minutes in total for flow 1 when there is no congestion).

The routes taken by all the flows in this example are deterministic. Vehicles of flow 1 enter in neighbourhood 1 at rate $\lambda_1^1(t)$ and reach their destination in this neighbourhood. Flow 2 enters neighbourhood 1 at rate $\lambda_1^2(t)$ and exits in neighbourhood 2 while flow 3 enters in 1 at rate $\lambda_1^3(t)$ and exits in 3. Flows 4, 5, and 6 have analogous rates and routes as shown Figure 8.

In the game and the optimal control part, only flows 1, 2, and 3, that is those entering neighbourhood 1, participate in the optimisation. The other flows are not controlled but nevertheless influence the queues and the travel times in the neighbourhoods. In the numerical experiments, we assume that the arrival rates of the flows entering the same neighbourhood are the same. That is, $\lambda_1^1(t) = \lambda_1^2(t) = \lambda_1^3(t) = \lambda_1(t)$, and $\lambda_4^3(t) = \lambda_5^3 = \lambda_3^6(t) = \lambda_3(t)$.

Figure 9 shows the arrival rates (top row), the queue lengths in the three neighbourhoods (second row), as well as the mean travel times of the six flows (bottom two rows) in the fluid limit (left pane) when there is no public transport alternative. The notation $\bar{w}^f$ (resp. q_n) indicates the mean travel for flow f (resp. total queue length in neighbourhood n). The accuracy of the fluid limit is shown in the right pane for $K = 10000$. As with the other example, the fluid limit is accurate during congestion periods but not when the traffic densities are low. Recall that the simulation only considers a single trajectory for the densities, and that the depicted travel times are the mean travel times, conditional on that single trajectory.

Next, we introduce a public transport alternative which costs $C^1 = 20$ for flow 1, $C^2 = 35$ for flow 2 and $C^3 = 50$ for flow 3. This reflects higher costs for longer travel distances, as well

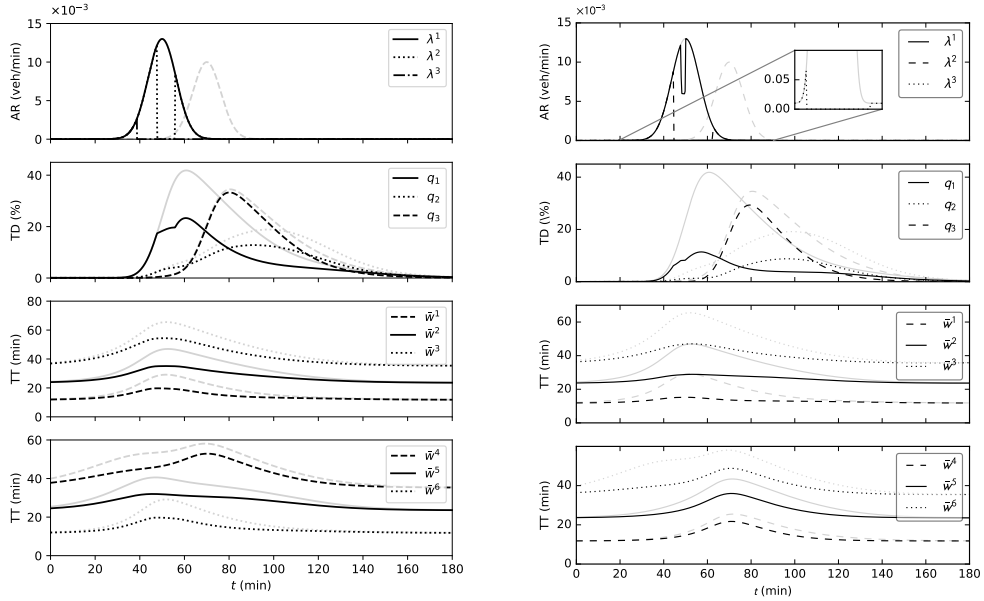


Figure 10: Arrival rate (AR), traffic density (TD) and travel times (TT) during rush hour with a public transport alternative. Wardrop equilibrium (left). Optimal control solution (right).

as a fixed cost for the inconvenience of switching to the PTS. The comparison of the Wardrop equilibrium and optimal control solution is shown in Figure 10. For reference, we also depict the curves without the public transport alternative (as shown in Figure 9) in gray. For better readability, we use the same notation in the legend for both the left and the right panes. Here, it is implicit that the variable in the left pane (resp. right pane) is the one obtained from the Wardrop equilibrium (resp. optimal control solution).

The inset of the arrival rate plot of the optimal control solution shows the rate of flow 3. The travel times for this flow starts to approach and go over its PTS cost of 50 units around $t = 30$. In anticipation of the congestion that will be created by increased arrival rates of the flows 4, 5, 6 when the flow 3 will reach neighbourhood 3, the optimal control solution switches this flow to its PTS at a relatively early point in time and keeps it there until the congestion in neighbourhood 3 starts to wane.

The PoA for this example (for the interval $[30, 120]$) is 1.036, which corresponds to an additional cost of 3.6%. The congestion based pricing curves have a similar shape as the ones of the first example and have been omitted.