Summary

In this dissertation, we develop models and control techniques for road traffic congestion in which the main focus lies on incorporating the impact of uncertainty by means of quantitative stochastic methods.

The overall goal of the research in this dissertation is to gain understanding in the impact of uncertainty on the effectiveness of control mechanisms for road traffic congestion. The effectiveness of traffic management solutions depends on the interaction between travellers and the settings of roadside systems, amongst others. However, the inclusion of uncertainty in modelling large-scale networks often leads to computational intractability. Therefore, it is crucial to partition the road network into a hierarchical structure of manageable subnetworks to keep a scalable solution. We analyse the impact of uncertainty by taking into account the aspects mentioned above. This leads to a division of this dissertation into three parts: actuator control, user behaviour, and lastly, network analysis.

Actuator control
A common goal in traffic management is to keep traffic density near bottleneck junctions low enough to avoid traffic deadlock, but on the other hand, provide sufficient throughput to prevent unnecessary delay in the upstream direction. In Chapters 2-4, we introduce a generic model for such traffic flow control applications. The stochastic nature of traffic flow in both capacity and demand leads to complex system dynamics. This makes it hard to determine effective control mechanisms to reduce or prevent the impact of congestion. Understanding and quantifying the interplay between queues incorporating the stochasticity of the arrival process and capacity is a starting point for stochastic traffic flow control strategies. In these chapters, we study control strategies that avoid accumulation of traffic at strategic points in a network.

In Chapter 2, we introduce two versions of a Markovian tandem model for which the service rate of the first queue can be controlled. In the first model, the control of the service rate at the first queue is limited to being turned on or off. In the second model, the system contains a batch-processing server where the number of jobs to be transferred
Summary

can be specified at all times. For several applications the batch server is a more realistic assumption, for example to model multiple vehicles driving over the same stretch of road simultaneously. For both models, the objective is to keep the mean number of jobs in the second queue as low as possible, without compromising the total system delay (i.e. avoiding starvation of the second queue). The balance between these objectives is governed by a linear cost function of the queue lengths. We formulate this model as a Markov Decision Process (MDP).

It turns out that the optimal strategy for both versions is characterised by a switching curve dividing the state space into two regions. In this case, the state space represents the number of jobs in the first queue on the $x$-axis, and the number of jobs in the second queue on the $y$-axis. The state space in the first version, where jobs are only handled sequentially, is divided by a sub-linear line. Below this line the first queue is processing at full speed, while above this line the service is paused. For the batch-processing server, a similar sub-linear shape is encountered when grouping states with the same optimal decision.

Real-time evaluation of the theoretical optimal strategy under changing conditions can become computationally demanding. Especially when such strategies are to be analysed for a sequence of bottlenecks for which the second queue of the one system serves as the input for the next. Therefore, we introduce two approximation approaches in Chapters 3 and 4.

When the optimal switching curve is rather flat, it can be well approximated by a horizontal one, which corresponds to a fixed threshold strategy. In Chapter 3, we develop an approximation technique to investigate the effectiveness of such fixed threshold strategies. For the 'optimal' threshold level, we verify that it performs very closely to the optimal MDP strategy under medium loaded systems. However, when the load of the system increases, the performance gap between the MDP strategy and the approximation increases. Under heavily loaded systems the structure of the optimal policy becomes more important. To overcome this performance gap, we develop a dynamic approximation strategy in Chapter 4.

In Chapter 4 we exploit the structure of the optimal strategy and develop heuristic policies motivated by the analysis of a related controlled fluid problem. The fluid approach provides excellent approximations, and thus
understanding, of the optimal MDP policy. The computational effort to determine the heuristic policies is much lower and, more importantly, hardly affected by the system load. The heuristic approximations can be extended to models with general service distributions, for which we numerically illustrate the accuracy.

User behaviour
In practice, travellers can strategically choose their departure times and the routes they take. Congestion occurs when more users simultaneously access the infrastructure than can be sustained by that infrastructure. These location are refered as bottlenecks.

The models in Chapters 5 and 6 are based on a popular approach to model congestion and user response. The main goal is to find compatible departure times of travellers, such that all travellers suffer the same discomfort. This discomfort is expressed in a cost function that accounts for three cost components: the cost of being too early at the destination, the cost of arriving too late and the cost of travelling time; the latter component is determined by the delay due to traffic congestion. The compatible departure times are found by the Nash equilibrium, which means that no traveller can improve its costs by shifting its departure time.

In Chapter 5 this model is extended with stochastic (uncertain) arrival times and travelling speeds by using a Poisson arrival process with time-fluctuating rate and exponential travel times. The strategic behaviour of users is captured in the aggregated intensity function of the Poisson arrival process. We discuss the error made by the fluid approximation, and show that the Nash equilibrium of the original model results in highly varying costs when applied in the more realistic setting with stochasticity. We then develop an algorithm to numerically approximate the equilibrium arrival rate for the stochastic bottleneck model, and propose a closed-form estimation for the approximated equilibrium. This approach can be applied to other extensions that have been developed for the standard deterministic bottleneck model. The results give intuition on the impact of uncertainty in a broad range of transportation models. Examples include heterogeneity among travellers' departure time, interpretation of early and late arrival, demand elasticity, etcetera.

In Chapter 6, we use a more detailed model for the rational behaviour of travellers: each can strategically choose a preferred time to join the
Summary

bottleneck, but the actual time at which the bottleneck is reached is subject to a random shift in time. This captures uncertainty with respect to departure and travelling times prior to joining the bottleneck. We show that the arrival density advocated by the Nash equilibrium in Vickrey’s model is not a user equilibrium in the model with random uncertainty. We then investigate the existence of a user equilibrium for the latter and show that, in general, such an equilibrium can neither be a pure Nash equilibrium, nor a mixed equilibrium with a continuous density. With numerical examples, we illustrate the mechanics that prevent existence of such a user equilibrium. Our results demonstrate that when random distortions influence user decisions, the dynamics of standard bottleneck models are inadequate to describe such complex situations.

In Chapter 7 we develop a strategic scheduling model. As in the previous two chapters, the goal is to dynamically spread arrivals, but now travel times are optimised in a joint effort between travellers and a central coordinator. The central coordination allows for effective synchronisation of travellers’ preferences. For this study, we split the travellers into two groups: (1) participating travellers whose departure time interval can be adjusted, and (2) non-participating ‘background’ travellers whose departure times cannot be adjusted. This allows us to assess the impact of the fraction ‘adjustable traffic’ on the total delay. Our results show that a significant decrease in average delay can be established when only a small fraction of the total traffic uses a personal departure advice.

Network analysis

In Chapter 8, we examine the structure of an empirical data set consisting of time-dependent origin-destination pairs in terms of connectedness. A network partitioning algorithm is applied to aggregate travel patterns into high-level partitions of the network. These partitions are composed of historical travel movements in the city of Amsterdam. We show that we can distinguish spatially connected regions when we use a heuristic method that optimises a performance metric called modularity. We proceed to analyse variations in the partitions that arise due to the non-optimal greedy optimisation method. We use a method known as ensemble learning to combine these variations by means of the overlap in community partitions. Ultimately, the combined partition leads to a more consistent result when evaluated again, compared to the individual partitions.