Vehicle routing under time-dependent travel times: the impact of congestion avoidance

A.L. Kok*, E.W. Hans, J.M.J. Schutten

Operational Methods for Production and Logistics, University of Twente,

P.O. Box 217, 7500AE, Enschede, Netherlands

Abstract

Daily traffic congestion forms a major problem for businesses such as logistic service providers and distribution firms. It causes late arrivals at customers and additional hiring costs for the truck drivers. Such costs caused by traffic congestion can be reduced by taking into account and avoid well-predictable traffic congestion within vehicle route plans. In literature, various strategies are proposed to avoid traffic congestion, such as selecting alternative routes, changing the customer visit sequences, and changing the vehicle-customer assignments. We investigate the impact of these and other strategies in off-line vehicle routing on the performance of vehicle route plans in reality. For this purpose, we develop a set of vehicle routing problem instances on real road networks, and a speed model that reflects the key elements of peak hour traffic congestion. The instances are solved for different levels of congestion avoidance using a modified Dijkstra algorithm and a restricted dynamic programming heuristic. Computational experiments show that 99% of late arrivals at customers can be eliminated if traffic congestion is accounted for off-line. On top of that, about 87% of the extra duty time caused by traffic congestion can be eliminated by clever congestion avoidance strategies.

Keywords: Congestion avoidance; Time-dependent VRP; Time-dependent SPP; Speed model;

*Corresponding author. Tel.: +31 53 489 4164; fax: +31 53 489 2159

E-mail adresses: a.l.kok@utwente.nl, e.w.hans@utwente.nl, j.m.j.schutten@utwente.nl

1 Introduction

Due to a growing amount of traffic and a limited capacity of the road network, traffic congestion has become a daily phenomenon. Since traffic congestion causes heavy delays, it is very costly for intensive road users such as logistic service providers and distribution firms. In particular, such delays cause large costs for hiring the truck drivers and the use of extra vehicles, and if they are not accounted for in the vehicle route plans they may cause late arrivals at customers or even violations of driving hours regulations. Therefore, accounting for and avoiding traffic congestion has a large potential for cost savings.

Traffic congestion may have several causes. Some are well-predictable, such as the large amount of commuter traffic during the daily peak hours, and others are less predictable, such as the weather or road accidents. Since delays caused by peak hour traffic congestion are well-predictable and they constitute a large part (70 to 87%) of all traffic congestion delays (Skabardonis et al. [1]), we focus on avoiding peak hour traffic congestion.

Given a certain realization of the factors causing traffic congestion, peak hour traffic congestion depends on location and time of the day. Therefore, congestion avoidance is all about not being at the wrong place at the wrong time. There are several strategies to achieve this. For example, changing the visit sequence of a vehicle may avoid a large traffic jam. Another example is to remove a customer from one vehicle route and insert it into another. These congestion avoidance strategies can be optimized by solving a vehicle routing problem (VRP) with time-dependent travel times (TDVRP). Although literature on the VRP with time-independent travel times is exhaustive (for an extensive overview, see Toth and Vigo [2]), literature on the TDVRP is scarce. The only papers we are aware of are: Malandraki and Daskin [3], Ichoua et al. [4], Fleischmann et al. [5], Haghani and Jung [6], Donati et al. [7], Van Woensel et al. [8], and Hashimoto et al. [9].

Another strategy to avoid traffic congestion is to select an alternative route between two customers at problematic hours. This strategy implies that the route between two customers depends on the chosen departure time, which can be optimized by solving a shortest path problem with time-dependent travel times (TDSPP). Orda and Rom [10] show that solving a time-dependent shortest path problem for a given departure time can be done using a modified Dijkstra [11] search.

We investigate the impact of different congestion avoidance strategies within off-line vehicle routing on the performance of vehicle route plans in reality. We propose four strategies in which congestion avoidance is applied to an increasing extent. Within these strategies, time-dependent shortest path problems and time-dependent vehicle routing problems are combined in one model for off-line vehicle routing. To the best of our knowledge, this is the first paper that considers both these problems in one model, which is needed to solve practical vehicle routing problems including traffic congestion.

To test the impact of the four strategies in a realistic setting, we develop a number of VRP instances on real road networks and a speed model representing peak hour traffic congestion. The VRP instances include customer service time windows, which have been generally ignored in TDVRP literature. The speed model reflects the key elements of peak hour traffic congestion, based on observations of the Dutch motorists' organization ANWB [12] of peak hour traffic congestion in the Netherlands: large delays in urban areas, large delays on road lanes towards urban areas during the morning peak and in the opposite direction during the evening peak, and large delays on roads with a high speed limit (highways). We assume that these key elements are similar for other road networks with similar densities. We determine for each VRP instance an off-line vehicle route plan for each different congestion avoidance strategy. Then, we evaluate the quality of these plans by executing them with the actual speeds in the road network obtained from the speed model. Quality is measured in terms of, among other things, number of late arrivals at customers, number of vehicles needed, total duty time (the duty time of a truck driver is defined as the difference between his start and completion time), and total travel distance. This methodology evaluates the impact of the different levels of congestion avoidance in reality.

The contributions of this paper are the following. First, to the best of our knowledge, this is the first paper that considers shortest path problems and vehicle routing problems including traffic congestion in one model. Second, this paper quantifies the performance improvement with respect to both duty times and travel distances by accounting for and avoiding traffic congestion. Due to the simplicity of the classical vehicle routing models (in which client networks are considered in which travel distances are set to euclidean distances and travel times are proportional to these travel distances), quantifying both duty times and travel distances is not possible within these models. Third, this paper quantifies the number of late arrivals and the late times at customers resulting from a lack of accounting for traffic congestion. Since this is done for different congestion avoidance strategies, we get a clear indication of which congestion avoidance strategies are necessary or beneficial in gaining high quality vehicle route plans in practice. Fourth, this paper proposes a speed model on large road networks that reflects the key elements of peak hour traffic congestion. Therefore, it can be used to evaluate the quality of new vehicle routing models and solution methods for VRPs including traffic congestion.

This paper is organized as follows. In Section 2 we propose a speed model representing the key elements of peak hour traffic congestion. In Section 3 we propose four strategies with different levels of congestion avoidance for solving off-line vehicle routing problems. In Section 4 we propose a general solution approach to solve the problem instances with these four congestion avoidance strategies. In Section 5 we present the impact of the different congestion avoidance strategies on the real-life performance, and in Section 6 we give some concluding remarks.

2 Speed model

To investigate the impact of the different congestion avoidance strategies in a realistic setting, we propose a speed model for real road networks that reflects the key elements of peak hour traffic congestion, as observed by the Dutch motorists' organization ANWB. The speed model is designed for the road network data used in this paper, but the methodology can be applied to any other road network data. Note that the speed model is based on the key elements of peak hour traffic congestion; it is not based on real (historical) travel time data. There is no empirical evaluation of the performance of this speed model on real road networks in particular. To apply vehicle routing methods in practice, the speed model should be tailored to the road networks under consideration. This tailoring is beyond the scope of this study: the objective of this study is to get a good estimation of the performance of different congestion avoidance strategies in a broad and realistic setting.

Our road network data is a selection of the TIGER/Line files [13], which consist of road network data of each of the 50 US states. We select the states Rhode Island, Connecticut, Maryland, Massachusetts, and New Jersey, because they have a high degree of urbanization overall, resulting in many traffic congestion problems during the peak hours. On top of that, the sizes of these states are comparable to some smaller countries in Europe such as the Netherlands and Belgium, which have to face large congestion problems since they are densely populated. Furthermore, we select Kentucky for comparison reasons: this state has next to a relatively small urban area also large rural areas.

The TIGER/Line data contain geoinformation on nodes in the road network (a node may represent an intersection of different roads or a change in average speed on the same road), and distance and road category information on directed arcs connecting these nodes. There are four road categories with their corresponding (normalized) average speeds: 1, 0.8, 0.6, and 0.4. These speeds are time-independent.

We propose a time-dependent speed model that incorporates peak hour traffic congestion. This speed model defines for each arc a travel speed during the peak hours and a travel speed outside the peak hours. We assume the morning peak to last from 6:30AM until 9:30AM and the evening peak from 3:30PM until 7PM, since observations of the Dutch motorists' organization ANWB indicate the peak periods as such. The speed outside the peak hours is set to the speed provided by the TIGER/Line files. The speed drops during the peak hours are based on the key elements of peak hour traffic congestion.

Peak hour congestion is mainly caused by a large amount of commuter traffic. Since commuter traffic needs to be at the same time (at the start of the working day) at the same place (large cities), the most common roads get congested during the peak hours. With respect to peak hour traffic congestion, the following elements are relevant:

1. Degree of urbanization. Within urban areas, there is much more traffic congestion than in rural areas. Therefore, there is a positive correlation between the degree of urbanization and the amount of speed drop during the peak hours.

2. Direction of commuter traffic. During the morning peak, commuter traffic is traveling toward working areas. Therefore, during the morning peak much more traffic congestion appears on road lanes directed to urban areas than on road lanes in the opposite direction (and during the evening peak vice versa).

3. Speed limit. In general, roads with a high speed limit (highways) are more heavily used than roads with a lower speed limit (rural roads). Because these roads lie on many (free-flow) shortest paths, there is a positive correlation between a road's speed limit and the amount of traffic congestion during the peak hours.

We propose the following approach to quantify the speed drops during the peak hours on each arc in the road network, based on the three observations described above. First we determine the degree of urbanization of the source- and destination-node of the arc under consideration. We determine this, by counting the number of network nodes in the proximity area of each node. We refer to such nodes as proximity nodes. In Section 2.1, we explain in detail how this proximity area is defined and how we use it to determine the degree of urbanization of each node. We set the degree of urbanization of each arc to the maximum of the degrees of its source- and destination-node. Next, we determine the direction of the arc, i.e., toward or from an urban area. If the destination-node has a higher number of proximity nodes than the source-node, then the arc is directed toward an urban area. Finally, the speed limit on the arc is given by the road category of the arc under consideration.

Using the calculated degree of urbanization, the arc direction, and the road category, we calculate the speed drops as follows. Table 1 presents the maximum (relative) speed drops, expressed as a percentage of the free-flow speed, during the morning peak for each road category (for the evening peak the two rows are swapped). These maximum speed drops depend both on the arc direction and on the road category. In Section 5.1, we conduct a sensitivity analysis of these speed drops by repeating all computational experiments for a selected number of alternative maximum speed drop patterns. We multiply the corresponding speed drop with a fraction, depending on the degree of urbanization. Table 2 presents these fractions. Degree 1 represents the highest degree of urbanization.

	road cat. 1	road cat. 2	road cat. 3	road cat. 4
Arcs toward urban areas	0.9	0.65	0.4	0.15
Arcs from urban areas	0.3	0.25	0.2	0.15

Table 1: Maximum speed drop during the morning peak as a percentage of the free-flow speed

Using this approach, the speed model reflects the three key elements of peak hour traffic congestion. Table 1 presents the dependency of the speed drops on the road direction and the speed limit, and Table 2 shows that arcs

Degree of urbanization	Fraction of speed drop
1	1
2	2/3
3	1/3
4	0

Table 2: Degree of urbanization and corresponding speed drop fraction

located in high degrees of urbanization encounter larger speed drops than arcs in low degrees of urbanization.

2.1 Determining the degree of urbanization of a node

We propose the following methodology for determining the degree of urbanization of each node in the road network. We define the proximity area of a node to a circle centered at this node with a radius of 10 km, such that urban areas are identified if a node is in a 10 km range of this urban area. To get an indication of the number of proximity nodes for a node that lies in the center of a large city and, therefore, has the highest degree of urbanization, we determine for each state the maximum number of proximity nodes over all nodes that lie in the largest city of that state.

State	Max # proximity nodes ($\times$1,000)
Connecticut	15
Kentucky	19
Rhode Island	24
Maryland	28
New Jersey	32
Massachusetts	38

Table 3: Maximum # proximity nodes in the largest city

Table 3 shows that nodes which have the largest degree of urbanization contain 15 thousand or more proximity nodes. Therefore, we set the degree of urbanization of a node to 1 if it contains at least 15 thousand proximity nodes. The numbers of proximity nodes corresponding to the other degrees of urbanization are evenly spread between 0 and 15 thousand. Table 4 presents the resulting correspondence between the number of proximity nodes and the degree of urbanization.

# proximity nodes ($\times$1,000)	degree of urbanization
15+	1
10 - 15	2
5 - 10	3
0 - 5	4

Table 4: Number of proximity nodes and degree of urbanization

3 Strategies

We propose four different strategies in which congestion avoidance is applied to an increasing extent. The first two strategies are closest to the VRP literature, in which the travel times are modeled as time-independent. The only difference between these two strategies is that Strategy 2 accounts to some extent for traffic congestion, while Strategy 1 completely ignores traffic congestion. In the other two strategies, time-dependent travel times are accounted for and, therefore, traffic congestion can be avoided. This is done to a greater extent with Strategy 4 than with Strategy 3. Each strategy consist of two phases. In phase one the (time-dependent) shortest paths are determined between the customers. In phase two a (time-dependent) VRP is solved, based on the travel times resulting from phase one. We now describe the four strategies in more detail.

Strategy 1 is the base strategy in which traffic congestion is completely ignored. In phase one, all free-flow shortest paths between the customers are determined. For this purpose, the speeds at the arcs in the road network are set to their maximum speed of the day (the speed outside the peak hours). The travel times along the paths are calculated using these maximum speeds. Next, a VRP is solved based on the resulting free-flow travel times between the customers. This strategy does not account for traffic congestion to any extent, and, therefore, there is no traffic congestion avoidance.

With Strategy 2, traffic congestion is accounted for by solving a VRP based on average travel times over the day. As with Strategy 1, the free-flow shortest paths are determined in the first phase by setting all arcs to their maximum speed. These paths are fixed and they will be used if they are selected in phase two. Next, the average travel times along these paths are determined by evaluating the exact travel times for a large number of different departure times (we set the inter-departure time to 15 minutes). For each departure time, the travel times are calculated using the time-dependent arc speeds resulting from the speed model. The averages over these time-dependent travel times are used as input for a VRP in phase two. With this strategy, traffic congestion is accounted for by averaging the delays along a route over the entire day. Traffic congestion could be avoided to a small extent if routes with smaller average travel times are selected due to smaller delays during the peak hours.

With Strategy 3, traffic congestion is avoided by solving a TDVRP. The time-dependent travel times which form the input for the TDVRP are determined in a similar way as with Strategy 2. The only difference is that the large number of time-dependent travel times (with inter-departure times of 15 minutes) is not averaged over the day. Interpolation is used each time the TDVRP solver requires the travel time between two customers for a given departure time. Note that with this strategy the routes between the customers are still fixed at the free-flow shortest paths. This strategy accounts for traffic congestion to a high extent and allows congestion avoidance by selecting alternative customer visit sequences and alternative vehicle-customer assignments.

Strategy 4 adopts the highest level of congestion avoidance. With this strategy, time-dependent shortest paths are determined for a large number of different departure times in phase 1 (the inter-departure times are again set at 15 minutes). Next, the time-dependent travel times are determined for all these different departure times along the corresponding time-dependent shortest paths and used as an input for solving a TDVRP. Again, interpolation is used to estimate the travel time for a given departure time when the TDVRP is being solved. Finally, since the resulting planned departure times generally do not coincide with a departure time for which the time-dependent shortest path has already been determined in phase 1, we determine the shortest paths for the planned departure times as a last step with this strategy. This strategy accounts for traffic congestion to the same extent as Strategy 3, but adopts a higher level of congestion avoidance: at problematic hours alternative routes between customers are chosen. Table 5 gives an overview of the four strategies.

After the VRP instances are solved with each strategy, the performances

Strategy	Shortest paths	Travel times input for VRP	Accounting for congestion	Avoiding congestion
1	free-flow	free-flow	no	no
2	free-flow	average	a small extent	no
3	free-flow	time-dep.	yes	yes
4	time-dep.	time-dep.	yes	yes

Table 5: Strategy overview

of the resulting vehicle route plans are evaluated according to the speeds resulting from the speed model in the road network. Note that with each strategy travel time estimations are used to solve a (TD)VRP. The reason for using travel time estimations with Strategy 3 and 4 is that calculating the exact travel time based on the chosen route and the speed model each time a travel time is needed would be too time-consuming. On top of that, these travel time calculations depend on the number of arcs along the route under consideration and, therefore, on the size of the road network. In practice, however, the TDVRP often needs to be solved within limited computation times. Therefore, the proposed travel time estimations, which do not depend on the size of the road network, are suitable for practice.

The required computation times to solve the shortest path problems also depend on the size of the road networks. However, in practice shortest paths are re-used many times: only the addition of new customers or road network changes require shortest path (re-)calculations. Therefore, these shortest paths can often be determined in a pre-processing phase in which computation times play a minor role.

4 Solution methods

With each strategy, we need to solve a (TD)SPP and a (TD)VRP. To make a fair comparison of the different strategies, we solve the problems with solution methods that do not need to be tailored for each specific problem. This means that the shortest path algorithm can solve both SPPs and TDSPPs, and the VRP algorithm can solve both VRPs and TDVRPs. We also require that the computation times of the VRP solution methods are (approximately) the same with each strategy, such that this does not affect the applicability of the different strategies in practice. For the shortest path algorithms we are less restrictive, since shortest path calculations are generally done as a pre-processing in practice.

We solve the (TD)SPPs with a modified Dijkstra [11] algorithm. The only adaptation we make is that we initiate the searches with a given departure time from the source node, and we keep track of the departure times at each reached node. This is necessary to determine the time-dependent travel times when the labels of the nodes need to be updated. This approach allows us to solve the shortest path problems with Strategy 1, 2, and 3 in (approximately) the same computation times. Only with Strategy 4 the computation time increases, since we have to rerun the algorithm for each possible departure time. However, this is all done in phase 1, a pre-processing phase in which computation times play a minor role in practice. Note that the speed model satisfies the non-passing property, which is necessary to guarantee optimality of Dijkstra's algorithm, since violations of this property may allow an optimal path to contain non optimal sub-paths.

We solve the (TD)VRPs with a restricted dynamic programming heuristic, introduced by Gromicho et al. [14]. We select this solution method, since it can solve VRPs and TDVRPs with similar computation times. Moreover, no tailoring is necessary for the problems at hand. This is not only very valuable for practical use, but it also provides a fair comparison of the quality of the different strategies. Finally, this method has the advantage that it can solve small problem instances to optimality, and only when computation times become restrictive it acts as a heuristic. We provide a short explanation of the restricted DP heuristic of Gromicho et al. [14].

The restricted DP heuristic for the VRP is based on the exact DP algorithm for the TSP of Held and Karp [15] and Bellman [16]. This DP algorithm defines states $(S, j), j \in S, S \subseteq V \backslash 0$, which represent a minimum-length tour with cost $C(S, j)$, and in which V represents the entire set of nodes to be visited. This tour starts at node 0 and visits all nodes in S, which is a proper subset of V, and it ends in node $j \in S$. The costs of the states in the first stage are calculated by $C(\{j\}, j) = c_{0j}, \forall j \in V \backslash 0$, in which c_{ij} is the cost of traveling directly from node i to node j. Next, the costs of the states in all subsequent stages are calculated by the recurrence relation $C(S, j) = \min_{i \in S \backslash j} \{C(S \backslash j, i) + c_{ij}\}$.

The DP algorithm for the TSP is applied to the VRP through the giant-tour representation of vehicle routing solutions introduced by Funke et al. [17]. In this representation, the vehicles are ordered and for each vehicle k a unique origin node o_k and destination node d_k are introduced. Next, the destination node of each vehicle is connected to the origin node of its successive vehicle, as well as the destination node of the last vehicle with the origin node of the first vehicle, creating a giant-tour. The DP algorithm is applied to the extended node set with the vehicle origin and destination nodes, in which each node addition now requires a feasibility check.

The feasibility checks ensure that an origin node of a vehicle o_k can be added to a partial route represented by a state if and only if the last visited node is d_{k-1}. Furthermore, these checks only allow d_k to be added if o_k is already in the visited node set S. To account for other restrictions, such as capacity restrictions or time windows, state dimensions are added. For example, in case of capacity restrictions a state dimension c is added that keeps track of the accumulated demand of the active vehicle k. With active vehicle we refer to the last vehicle for which o_k has been added to the set of visited nodes. Each time a vehicle origin node o_k is added to a state, c is reset to zero. Furthermore, a customer addition is only allowed if the accumulated demand c together with the customer demand does not exceed the capacity of the active vehicle. Many other restrictions such as time windows, sequencing restrictions (pickup and delivery), multiple depots, and heterogeneous vehicle fleets can be incorporated by adding state dimensions or control via the input, allowing for a general framework for solving VRPs.

Since the (unrestricted) DP algorithm does not run in practically acceptable computation times for problem instances of realistic sizes, the state space is restricted by a parameter H. The value of H specifies the maximum number of states to be taken to the next stage, such that the smallest cost states are maintained, as proposed by Malandraki and Dial [18]. Since states in the same stage represent partial tours of the same length, states with smaller costs are more likely to lead to good overall solutions.

These restrictions on the number of state expansions results in the following running time complexity of the restricted DP heuristic. In each stage, at most H states are expanded to at most nH states (where n equals the number of customer nodes and vehicle origin and destination nodes). Since we have to select the H best states for the next stage, each stage requires $O(nHlog(H))$ time. The total number of stages equals the number of nodes in the network, which is $O(n)$. Therefore, the running time complexity of the restricted DP heuristic is $O(n^2Hlog(H))$. Since customers are always added to the end of the partial vehicle route and we store in each state the departure time from the last visited node, we can apply the restricted DP heuristic directly to the TDVRP without affecting the running time complexity.

5 Computational experiments

We test the impact of the four congestion avoidance strategies on a large number of VRP instances. These VRP instances are developed on the road networks of the six selected states and the speeds resulting from the speed model. The customer locations are uniformly randomly selected from the nodes in the road network. Clustering of the customer locations is a natural result, since the road network is denser in urban areas. Furthermore, the first selected node is considered to be the depot.

We develop 15, 50, and 100 customer problem instances. The 15 customer problem instances are small enough to be solved to optimality in practical computation times. The 100 customer problem instances are approximately the largest instances for which we can still solve a sufficiently large number of problem instances within practical computation times.

We add time windows to 50 percent of the customers indicating the period in which service must start. We set the time window of the depot to $[0, 14]$, indicating a working day of 14 hours from 6AM until 8PM. The morning and evening peak last from 6:30AM until 9:30AM and from 3:30PM until 7PM, respectively. The width of the time windows at the selected customers are randomly drawn from $\{2, 3, 4, 5, 6\}$ quarters of an hour. The customer service times are randomly drawn from $\{1, 2\}$ quarters of an hour and the demands are randomly drawn from $\{1, 2, ..., 10\}$. If the vehicle capacities are set too low, then the length of vehicle routes are only restricted by these capacities. If they are set too high, then the length of the routes are only restricted by the time windows. Initial experiments showed that a capacity of 55 results in a good trade off and is therefore used in our experiments. We generate 20 problem instances for each combination of state and number of customers resulting in 360 VRP instances in total.

The primary objective is set to minimize the total number of vehicle routes, and the secondary objective to minimize the total duty time of the truck drivers. Duty times are the sum of travel, service, and waiting times. We choose to minimize duty times as a secondary objective instead of travel times or travel distances, since small duty times are generally more important in practice, as duty times define the hiring costs of the truck drivers and the periods in which the trucks are not available for other services. We choose to dispatch the trucks at time zero. Although postponing the departure times at the depot may substantially reduce duty times (see Kok et al. [19]), it is beyond the scope of this paper to also optimize the departure times of the vehicles. Therefore, extending this study with the impact of departure optimization on the performance of vehicle route plans is one of our recommendations for future research.

We implemented the data-structures and solution algorithms in Delphi 7 on a PC with a Core 2 Quad, 2.83 GHz CPU and 4 GB of RAM. Table 6 presents the average results for the four strategies over all problem instances (we scale the results of the 15- and 50-customer instances to 100-customer instances), except for the problem instances generated on Kentucky. We scale and average the results, since the scaled results of the 15-, 50-, and 100-customer problem instances are similar. Table 7 presents the relative differences of the performance measures of the last three strategies with respect to Strategy 1. All performance measures are derived by evaluating the developed route plans with each strategy against the speeds resulting from the speed model. Next to the two objectives 'minimizing number of vehicle routes' and 'minimizing duty times', we also report on the following performance measures: total travel distance, total number of late arrivals at customers, total number of late return times at the depot, maximum late time over all customers, and total late time over all customers. Although we do not optimize on travel distances, we also report the impacts of the strategies on this performance measure. For practice, both measures are relevant, since they cause different transport costs. The last four performance measures indicate the reliability of the route plans. Note that all performance measures present averages over all problem instances, except for 'Maximum late time' which presents the maximum over all problem instances and all vehicle routes.

Performance measure	Scen. 1	Scen. 2	Scen. 3	Scen. 4
# vehicle routes	15.88	17.73	15.96	15.93
Total duty time (hrs)	153.2	158.3	143.1	141.4
Total travel distance	58.56	58.21	58.21	57.83
# late arrivals	16.47	3.726	0.023	0.047
# late return times	0.328	0.111	0.000	0.000
Maximum late time (hrs)	3.051	0.990	0.001	0.001
Total late time (hrs)	15.28	1.745	0.000	0.000

Table 6: Main results, aggregated over all problem instances

The number of vehicle routes is larger with Strategy 2, 3, and 4 than with Strategy 1. This can be explained by the travel time estimations with

Performance measure	Strat. 2	Strat. 3	Strat. 4
# vehicle routes	11.70%	0.55%	0.36%
Total duty time	3.34%	-6.59%	-7.69%
Total travel distance	-0.60%	-0.59%	-1.24%
# late arrivals	-77.37%	-99.86%	-99.72%
# late return times	-66.10%	-100.00%	-100.00%
Maximum late time	-67.56%	-99.96%	-99.95%
Total late time	-88.58%	-100.00%	-100.00%

Table 7: Strategy 2, 3 and 4 relative to Strategy 1

Strategy 1, which are based on free-flow travel conditions and result in lower bounds to the travel time estimations with the other strategies. However, the results indicate that avoiding traffic congestion to an increasing extent reduces the number of vehicles again and approaches the number of vehicles required with Strategy 1.

The duty times show a similar pattern, which can be partially explained by the number of vehicle routes. Furthermore, the additional congestion information with Strategy 2, 3, and 4 with respect to Strategy 1 reduces the total duty time. This even results in an overall decrease of total duty time for Strategy 3 and 4 with respect to Strategy 1, despite the larger number of vehicle routes. Note that the *estimated* duty time with Strategy 1 is the best (optimal for the 15-customer instances) for flee-flow travel conditions. Therefore, if we subtract the estimated duty times found with Strategy 1 from the real duty times with each strategy, then we obtain estimations of the additional duty times caused by traffic congestion. Table 8 presents the average amount of additional duty time for each strategy with respect to the estimated duty time found with Strategy 1. The results show that with Strategy 1 about 8.8% of the total duty time is due to traffic congestion delays. Strategy 2 results in an even larger additional duty time than Strategy 1. Strategy 3 and 4, however, reduce the additional duty time of Strategy 1 substantially by 75% and 87%, respectively.

	Strat. 1	Strat. 2	Strat. 3	Strat. 4
Additional duty time (hrs)	13.49	18.61	3.403	1.711

Table 8: Average additional duty time caused by traffic congestion

The travel distances are similar with each strategy. The smallest travel distances are obtained with Strategy 4. This can be explained by choosing alternative paths between customer locations at bad hours with this strategy. Such alternative paths typically contain arcs with smaller speed drops than arcs on the free-flow shortest paths due to, e.g., arcs with lower maximum speeds. However, such lower speeds have to be compensated by smaller travel distances.

The reliability of the route plans strongly increases if the level of congestion avoidance increases. All reliability measures show a strong reduction with respect to Strategy 1. This is not surprising, since the strategies account for traffic congestion to an increasing extent. However, the huge improvement of Strategy 2 with respect to Strategy 1 in comparison with the additional improvements of the other two strategies is less obvious. The explanation for this is that an underestimation of a travel time at the start of a vehicle route with Strategy 1 propagates along all arrival times at successive customers of that vehicle route. With Strategy 2, such an underestimation is generally compensated by an overestimation of later travel times.

Table 9 presents the results for the sixth state: Kentucky. As mentioned before, Kentucky contains a large rural area compared to the other 5 states. Therefore, the main part of the routes of the problem instances generated on this state do not contain heavy delays caused by traffic congestion. Table 9 shows that this has a strong impact on the results. The increase in number of vehicle routes with Strategy 2 with respect to Strategy 1 is much smaller than for the other states. Moreover, with Strategy 4 almost the same number of vehicle routes is attained as with Strategy 1.

The differences in duty times are also much smaller. With Strategy 1, only about 1 hour of duty time is caused by congestion delays, as opposed to the 13.5 hours for the other states. Congestion avoidance strategies lead to reductions of this additional duty time up to 84% with Strategy 4. The reliability measures show similar results as for the other states. In conclusion, Kentucky leads to less congestion problems than highly urbanized states, but congestion avoidance may still substantially reduce costs in terms of number of vehicle routes and total duty time, and may still substantially increase the reliability of the vehicle route plans.

5.1 Sensitivity analysis of the speed model

The actual speed drops on specific road networks depends on several factors, such as landscape (hilly or flat), urban organization (companies centered

Performance measure	Strat. 1[a]	Strat. 2[b]	Strat. 3[b]	Strat. 4[b]
# vehicle routes	15.48	1.69%	0.93%	0.11%
Total duty time (hrs)	135.4	0.67%	-0.08%	-0.61%
Total travel distance	52.50	0.93%	2.27%	-0.98%
# late arrivals	2.011	-68.51%	-100.00%	-92.82%
# late return times	0.000[c]	-[c]	-[c]	-[c]
Maximum late time (hrs)	0.349	-80.24%	-100.00%	-99.93%
Total late time (hrs)	0.558	-88.29%	-100.00%	-99.94%
Additional duty time (hrs)	0.992	91.65%	-10.75%	-83.94%

[a]absolute figures
[b]relative change with Strategy 1
[c]there were no late return times with each strategy

Table 9: Change of Strategy 2, 3 and 4, relative to Strategy 1 for Kentucky

at one city or dispersed across many cities), difference in speed limit between trucks and cars, and even culture. Therefore, we conduct a sensitivity analysis of the speed model we proposed in this study by repeating the computational experiments for a selected number of alternatives. We selected the following four alternatives for the maximum speed drops, which can be seen as extreme cases of the key elements of peak hour traffic congestion:

1. Only speed drops on highways. In this alternative, there are no speed drops during the peak hours on secondary roads, but only on the highways. This alternative can be seen as a case in which the majority of the road users does not consider alternative paths through secondary roads as an option to avoid peak hour traffic congestion. It may also be the case that secondary roads are less sensitive to traffic congestion due to better traffic flows as a result of a lower maximum driving speed. Table 10[a] presents the maximum relative speed drops for this alternative.

2. All roads the same maximum relative speed drop. This alternative is basically the opposite of Alternative 1. It represents the case in which many road users choose alternative paths during the peak hours, such that also secondary roads get congested. Therefore, the maximum relative speed drops depend no longer on the road category. Table 10[b] presents the maximum relative speed drops for this alternative.

3. Similar speed drops during the morning and evening peak. Some cities lay between two areas that both attract much commuter traffic. Such an urban organization may imply that roads passing these 'in-between cities' get congested in both directions, and both during the morning and evening peak. To account for this alternative, we set the maximum relative speed drops independent of the road direction (which also implies that they are the same for the morning and evening peak), as Table 10[c] presents.

4. Small speed drops. To investigate the sensitivity of the amount of the relative speed drops, we propose a fourth alternative in which the maximum relative speed drops are half the original maximum relative speed drops. Table 10[d] presents the resulting maximum relative speed drops.

We run all computational experiments again for all alternatives. We compare the results of the different alternatives with the original speed drop pattern.

Table 11 presents the results for Alternative 1 in which speed drops during the peak hours only appear on highways. Strategy 2 results in a larger number of vehicle routes than Strategy 1, but this increase is smaller than with the original speed drops. This can be explained by the smaller speed drops, on average, in this alternative. The same holds for the increase in duty time with Strategy 2 with respect to Strategy 1. The reduction in the number of vehicles with Strategy 4 with respect to Strategy 3 is larger than with the original speed drops. This can be explained by the free-flow travel times on secondary roads, which are only exploited with Strategy 4. The higher reliability of the route plans when higher levels of congestion avoidance are adopted is similar to the results with the original speed drops. The reduction in additional duty time caused by traffic congestion is in this alternative even more impressive than with the original speed drops: 95% instead of 87% with Strategy 4. This larger reduction is due to the bigger opportunities for selecting alternative paths during the peak hours.

Table 12 presents the results for Alternative 2 in which all roads have the same maximum relative speed drop. For this alternative, the reliability measures with Strategy 1 are worse than with the original speed drops, which is caused by the bigger speed drops in Alternative 2, on average. As a consequence, the increase in number of vehicle routes with Strategy 2, 3,

(a)

	road cat. 1	road cat. 2	road cat. 3	road cat. 4
Arcs towards urban areas	0.9	0.0	0.0	0.0
Arcs from urban areas	0.3	0.0	0.0	0.0

(b)

	road cat. 1	road cat. 2	road cat. 3	road cat. 4
Arcs towards urban areas	0.9	0.9	0.9	0.9
Arcs from urban areas	0.3	0.3	0.3	0.3

(c)

	road cat. 1	road cat. 2	road cat. 3	road cat. 4
Arcs towards urban areas	0.9	0.65	0.4	0.15
Arcs from urban areas	0.9	0.65	0.4	0.15

(d)

	road cat. 1	road cat. 2	road cat. 3	road cat. 4
Arcs towards urban areas	0.45	0.325	0.2	0.075
Arcs from urban areas	0.15	0.125	0.1	0.075

Table 10: Maximum speed drop during the morning peak as a percentage of the free-flow speed for the four alternatives: a) only speed drops on highways, b) all roads the same maximum relative speed drop, c) similar speed drops during the morning and evening peak, d) small speed drops.

Performance measure	Strat. 1[a]	Strat. 2[b]	Strat. 3[b]	Strat. 4[b]
# vehicle routes	15.88	9.43%	0.59%	0.10%
Total duty time (hrs)	154.8	1.75%	-7.64%	-9.31%
Total travel distance	58.56	-0.24%	-0.26%	-1.44%
# late arrivals	17.15	-62.83%	-100.00%	-97.82%
# late return times	0.490	-67.35%	-100.00%	-100.00%
Maximum late time (hrs)	3.244	-57.08%	-100.00%	-99.22%
Total late time (hrs)	18.05	-79.74%	-100.00%	-99.92%
Additional duty time (hrs)	15.10	17.94%	-78.29%	-95.46%

[a]absolute figures
[b]relative change with Strategy 1

Table 11: Results Alternative 1: only speed drops on highways.

and 4 with respect to Strategy 1 is larger than for the original speed drops, especially with Strategy 2 (18% vs. 12%). Also the changes in duty times are more extreme than with the original speed drops: a larger increase with Strategy 2, and a larger decrease with Strategy 3 and 4. Although the maximum relative speed drops are the same for each road category, Strategy 4 still results in better vehicle route plans than Strategy 3. This can be explained by simply having more alternatives to choose from, but also by the fact that the same *relative* speed drop results in a smaller *absolute* speed drop on roads with lower maximum speeds, which is only exploited with Strategy 4.

Table 13 presents the results for Alternative 3 in which speed drops during the morning and evening peak are similar. Due to the larger speed drops, on average, the reliability measures for Strategy 1 are worse than with the original speed drops. As a consequence, the increase in number of vehicle routes with Strategy 2, 3, and 4 with respect to Strategy 1 is larger than for the original speed drops. The larger speed drops in Alternative 3 offer, on the other hand, more possibilities for avoiding them, which results in larger duty time reductions with Strategy 3 and 4 with respect to Strategy 1 than with the original speed drops. The other results are similar to the results with the original speed drops.

Table 14 presents the results for Alternative 4 in which speed drops are half the original speed drops. Due to the smaller speed drops, there are fewer late arrivals than with the original speed drops, and the additional duty time

Performance measure	Strat. 1[a]	Strat. 2[b]	Strat. 3[b]	Strat. 4[b]
# vehicle routes	15.88	18.23%	1.18%	0.65%
Total duty time	158.8	5.73%	-8.57%	-9.90%
Total travel distance	58.56	0.08%	0.60%	0.05%
# late arrivals	20.42	-83.64%	-99.54%	-99.85%
# late return times	0.823	-84.35%	-97.30%	-100.00%
Maximum late time	4.878	-82.36%	-98.60%	-99.96%
Total late time	23.48	-94.48%	-99.87%	-100.00%
Additional duty time	19.08	47.72%	-71.32%	-82.36%

[a]absolute figures
[b]relative change with Strategy 1

Table 12: Results Alternative 2: all roads have the same maximum relative speed drop.

Performance measure	Strat. 1[a]	Strat. 2[b]	Strat. 3[b]	Strat. 4[b]
# vehicle routes	15.88	15.31%	0.78%	0.57%
Total duty time	159.4	4.20%	-9.20%	-10.82%
Total travel distance	58.56	-0.29%	-0.94%	-1.70%
# late arrivals	20.97	-75.79%	-99.92%	-99.70%
# late return times	0.752	-63.96%	-97.05%	-100.00%
Maximum late time	3.589	-68.74%	-100.00%	-99.97%
Total late time	23.71	-89.06%	-100.00%	-100.00%
Additional duty time	19.67	33.99%	-74.52%	-87.66%

[a]absolute figures
[b]relative change with Strategy 1

Table 13: Results Alternative 3: speed drops during the morning and evening peak are similar.

caused by traffic congestion is also smaller. As a consequence, Strategy 3 and 4 find solutions with approximately the same number of vehicle routes as Strategy 1. Note that the smaller number of vehicle routes with Strategy 3 than with Strategy 1 and 4 is due to the heuristic solution method: for the 15 customer problem instances (which are solved to optimality) the number of vehicle routes is the same for all strategies.

Even with the small speed drops in Alternative 4, the reductions of the additional duty time caused by traffic congestion with Strategy 3 and 4 with respect to Strategy 1 are still substantial (more than 60%). We saw this also in the results on Kentucky, for which also the average congestion delays are much smaller than for the other states and alternatives. This strongly indicates that high levels of congestion avoidance leads to substantial cost savings for a broad range of different road networks. The reliability improvements with respect to Strategy 1 are similar to the improvements with the original speed drops.

Performance measure	Strat. 1[a]	Strat. 2[b]	Strat. 3[b]	Strat. 4[b]
# vehicle routes	15.88	2.97%	-0.06%	0.02%
Total duty time (hrs)	141.2	1.72%	-0.67%	-0.68%
Total travel distance	58.56	-0.25%	0.40%	-1.26%
# late arrivals	3.394	-87.33%	-99.41%	-97.94%
# late return times	0.000[c]	-[c]	-[c]	-[c]
Maximum late time (hrs)	0.328	-87.34%	-99.79%	-99.22%
Total late time (hrs)	0.494	-95.26%	-99.98%	-99.93%
Additional duty time (hrs)	1.528	158.94%	-62.11%	-63.07%

[a]absolute figures
[b]relative change with Strategy 1
[c]there were no late return times with each strategy

Table 14: Results Alternative 4: the speed drops are half the original speed drops.

6 Conclusions

We proposed four different strategies in which traffic congestion is accounted for and avoided to an increasing extent. Next, we proposed a speed model on real road networks, which reflects the key elements of peak hours traffic congestion. This speed model provided us a platform for testing the impact of the different strategies in a realistic setting.

The test results indicated that the reliability of route plans strongly increase if traffic congestion is accounted for. However, if VRPs are modeled with time-independent travel times, then this reliability increase is achieved against more vehicle routes and larger duty times. By adopting higher levels of congestion avoidance - such as solving VRPs with time-dependent travel times and solving time-dependent shortest path problems - these cost measures can be reduced substantially. Solving a combination of these two problems is particularly effective, resulting in huge reliability improvements, substantial duty time reductions (about 87% of the additional duty times caused by traffic congestion can be eliminated), and substantially reducing the number of vehicles needed (almost all extra vehicles needed to account for congestion delays can be eliminated).

We conducted a sensitivity analysis of the speed model, which indicated that under various scenarios the improvements with the higher levels of congestion avoidance remain. Even in case of small speed drops during the peak hours, congestion avoidance results in substantially more reliable route plans and substantial reductions of duty times and number of vehicle routes. In certain extreme cases, such as only speed drops on highways, congestion avoidance is even more powerful resulting in reductions of the additional duty time caused by traffic congestion of almost 95%.

Acknowledgment

This work was financially supported by Stichting Transumo through the project ketensynchronisatie.

References

[1] A. Skabardonis, P. Varaiya, and K. F. Petty. Measuring recurrent and nonrecurrent traffic congestion. *Transportation Research Record*, 1856:118–124, 2003.

[2] P. Toth and D. Vigo. *The vehicle routing problem.* SIAM Monographs on Discrete Mathematics and Applications. Philadelphia, 2002.

[3] C. Malandraki and M. S. Daskin. Time dependent vehicle routing problems: Formulations, properties and heuristic algorithms. *Transportation Science*, 26(3):185–200, 1992.

[4] S. Ichoua, M. Gendreau, and J. Y. Potvin. Vehicle dispatching with time-dependent travel times. *European Journal of Operational Research*, 144(2):379–396, 2003.

[5] B. Fleischmann, M. Gietz, and S. Gnutzmann. Time-varying travel times in vehicle routing. *Transportation Science*, 38(2):160–173, 2004.

[6] A. Haghani and S. Jung. A dynamic vehicle routing problem with time-dependent travel times. *Computers & Operations Research*, 32(11):2959–2986, 2005.

[7] A. V. Donati, R. Montemanni, N. Casagrande, A. E. Rizzoli, and L. M. Gambardella. Time dependent vehicle routing problem with a multi ant colony system. *European Journal of Operational Research*, 185(3):1174–1191, 2008.

[8] T. Van Woensel, L. Kerbache, H. Peremans, and N. Vandaele. Vehicle routing with dynamic travel times: A queueing approach. *European Journal of Operational Research*, 186(3):990–1007, 2008.

[9] H. Hashimoto, M. Yagiura, and T. Ibaraki. An iterated local search algorithm for the time-dependent vehicle routing problem with time windows. *Discrete Optimization*, 5:434–456, 2008.

[10] A. Orda and R. Rom. Shortest-path and minimum-delay algorithms in networks with time-dependent edge-length. *Journal of the Association for Computing Machinery*, 37(3):607–625, 1990.

[11] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, pages 269–271, 1959.

[12] ANWB Reisinformatie. (In Dutch) Dagelijks drukke trajecten. World Wide Web, last checked on January 14th 2010, 2010. `http://www.anwb.nl/verkeer/nederland/verkeersinformatie, /verkeersverwachting/Dagelijkse-drukke-trajecten.html`.

[13] Tiger/line. United states road networks, 2002. `http://www.dis.uniroma1.it/~challenge9/data/tiger/`.

[14] J. Gromicho, J. van Hoorn, A. L. Kok, and J. M. J. Schutten. The flexibility of restricted dynamic programming for the VRP. *Beta working paper series*, 266, 2008.

[15] M. Held and R. M. Karp. A dynamic programming approach to sequencing problems. *Journal of the Society for Industrial and Applied Mathematics*, 10(1):196–210, 1962.

[16] R. Bellman. Dynamic programming treatment of the travelling salesman problem. *Journal of the Association for Computing Machinery*, 9(1):61, 1962.

[17] B. Funke, T. Grünert, and S. Irnich. Local search for vehicle routing and scheduling problems: Review and conceptual integration. *Journal of Heuristics*, 11(4):267–306, 2005.

[18] C. Malandraki and R. B. Dial. A restricted dynamic programming heuristic algorithm for the time dependent traveling salesman problem. *European Journal of Operational Research*, 90(1):45–55, 1996.

[19] A. L. Kok, E. W. Hans, and J. M. J. Schutten. Optimizing departure times in vehicle routes. *Beta working paper series*, 236, 2008.