Vehicle routing under time-dependent travel times: the impact of congestion avoidance; details of the computational tests

A.L. Kok[a,*], E.W. Hans[b], J.M.J. Schutten[b]

[a] *Algorithmic R&D, ORTEC, P.O. Box 490, 2800AL Gouda, Netherlands*

[b] *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 costs for hiring the truck drivers. Such costs caused by traffic congestion can be reduced by taking into account and avoiding predictable traffic congestion within vehicle route plans. In the 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

*Corresponding author. Tel.: +31 182 540 500; fax: +31 182 540 540
E-mail adresses: leendert.kok@ortec.com, e.w.hans@utwente.nl, m.schutten@utwente.nl

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;

1 Introduction

Due to a growing amount of traffic and a limited capacity of the road network, traffic congestion has become a daily phenomenon. In the USA, the annual travel delay has grown from 2.5 billion delay hours in 1995 to 4.2 billion delay hours in 2005 (Schrank and Lomax [1]). Since traffic congestion causes heavy delays, it is very costly for intensive road users such as logistic service providers and distribution firms. The Dutch Organization for Transport and Logistics (TLN) estimated that over 10% of the truck drivers working hours are lost due to delays as a result of traffic congestion. This causes 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 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 predictable and they constitute a large part (70 to 87%) of all traffic congestion delays (Skabardonis et al. [2]), 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, we can change the visit sequence of a vehicle or even move a customer from one vehicle to another. These strategies can be optimized by solving a vehicle routing problem (VRP) with time-dependent travel times (TDVRP). Branch and cut and price methods have been very successful in solving vehicle routing problems with timing restrictions such as time windows (see, amongst others, Kohl et al. [3] and Chabrier [4]). Although the literature on the VRP with time-independent travel times is exhaustive (for an extensive overview, see

Toth and Vigo [5]), the literature on the TDVRP is rather scarce.

To account for traffic congestion effects within vehicle routing, Malandraki and Daskin [6] introduce the TDVRP and propose an ILP formulation for it. They model time-dependent travel times as a travel time step function for customer links. A drawback of this model is that it has the unrealistic property that a later departure time may result in an earlier arrival time, i.e., it does not satisfy the so-called *non-passing property*. Malandraki and Dial [7] propose a restricted dynamic programming heuristic for the Time-Dependent Traveling Salesman Problem (TDTSP), which is the special case of the TDVRP with only one vehicle. Bentner et al. [8] and Schneider [9] also consider the TDTSP. Both works consider one peak period and propose local search methods such as simulated annealing to solve the TDTSP. Ichoua et al. [10] resolve the non-passing property by considering a speed step function for customer links instead of a travel time step function. Since with a speed step function vehicles drive the same speed when traversing the same link at the same time, they can never overtake each other. Ichoua et al. [10] also show that a time-dependent model may lead to substantial improvements over a time-independent one, by computational tests on a set of modified VRP benchmarks. Fleischmann et al. [11] propose a modified savings heuristic for the TDVRP. They test their heuristic using travel time data obtained from a traffic information system in the city of Berlin. Haghani and Jung [12] propose a genetic algorithm for the TDVRP, which they validate with an exact algorithm for problem instances up to 30 customers. Donati et al. [13] propose an ant colony optimization algorithm for the TDVRP, which they test on some theoretical benchmarks, and a test case in the district of Padua in Italy. Van Woensel et al. [14] use a queueing model to derive time-dependent travel speeds. In addition, they show potential improvements by optimizing departure times from the depot. Hashimoto et al. [15] also consider both determining vehicle tours under time-dependent travel times and departure time optimization. Moreover, they consider time windows in their model.

In addition to solving a TDVRP, business traffic may also avoid traffic congestion by selecting alternative routes between customers at problematic hours. These routes are optimized by solving a shortest path problem with time-dependent travel times (TDSPP). Orda and Rom [16] show that solving a time-dependent shortest path problem for a given departure time can be done using a modified Dijkstra [17] search. Note that for solving a TDSPP a full representation of the road network is needed, whereas for solving a

TDVRP inter-customer time-dependent travel times suffice.

We compare four strategies to avoid traffic congestion. The strategies are applied within the development of off-line vehicle route plans. We test the impact of these plans in a realistic setting. We restrict ourselves to computing travel routes for some centrally controlled fraction of the traffic demand. With vehicle routing, we refer to routing of the vehicles of a certain company, such as a logistic service provider or a distribution firm. Within the development of the route plans, we combine time-dependent shortest path problems and time-dependent vehicle routing problems in one model. This is a difference with existing literature, in which these problems are generally considered separately.

We test the impact of the strategies on large cases, which is another difference with existing literature, since the majority of the papers dealing with the TDVRP consider theoretical benchmarks (except for a few small cases). For this purpose, we develop a number of VRP instances on real road networks and a speed model representing peak hour traffic congestion. We evaluate the quality of the vehicle route plans by executing them with the actual speeds in the road network obtained from the speed model.

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, we evaluate the impact of four congestion avoidance strategies with respect to different cost measures such as number of vehicles used, total travel distance, total duty time (sum of total traveling, serving, and waiting time), and total number of late arrivals. Since these measures all play a role in practice, we obtain a better indication of the performance of the strategies in practice than when resorting to only one or two objectives. Third, this paper proposes a speed model on large road networks that reflects the key elements of peak hour traffic congestion. This speed model differs from existing ones in that it considers the entire road network, it considers inter-urban traffic congestion effects, it considers two peak periods, and it considers different road types.

This paper is organized as follows. In Section 2 we propose the speed model, and in Section 3 we present the four congestion avoidance strategies. In Section 4 we discuss our solution approach to solve the problem instances with these strategies. In Section 5 we compare the impact of the strategies, 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 represents peak hour traffic congestion. The speed model defines for each arc in the road network a speed step function. This function defines speeds for different time intervals, such that in each time interval the speed is constant. We use five different time intervals: the morning and evening peak periods, and the periods before, in between, and after the peak periods. By defining speeds for each single road, the model is more refined than the majority of the papers dealing with the TDVRP, in which speeds are usually defined for customer links. Note that the speed model still contains the unrealistic effect that the speed suddenly changes when entering a new time interval. This effect can be reduced by considering more time intervals, but to keep the analysis simple, we choose to use five time intervals.

The speed model reflects the key elements of peak hour traffic congestion, as observed by the Dutch motorists' organization ANWB [18] and the English Highways Agency of the Department for Transport [19]. These key elements are: 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). The common observation by ANWB and the Highways Agency indicate that the key elements hold in general.

We develop the speed model for the road network data used in this paper. However, the methodology can be applied to other road network data. Note that we base the speed model on the key elements of peak hour traffic congestion; we do not base it on real (historical) travel time data. Therefore, for practical use, the speed model should be tailored to the road networks under consideration. This tailoring is beyond the scope of this paper: the objective of this paper 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 [20], which consist of road network data of each of the 50 US states. These data sets are suitable for our analysis, because all properties needed to develop the speed model are present: the data sets contain urban and rural areas, and they contain different road categories. Moreover, the sizes are appropriate for our analysis, which is restricted to one-day planning.

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 assume the morning peak to last from 6:30AM until 9:30AM and the evening peak from 3:30PM until 7PM, as indicated by the Dutch motorists' organization ANWB. The speed outside the peak hours is set to the speed provided by the TIGER/Line files. During the peak hours, the speed drops by a fraction of the speed outside the peak hours. We base this speed drop 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).

Therefore, 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.

Table 1 presents the maximum (relative) speed drops during the morning peak for each road category (for the evening peak the two rows are swapped). We express the speed drops as a fraction of the free-flow speed. 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. This analysis will illustrate the robustness of the results in Section 5, showing that these results do not depend on our choices for the actual numbers in Table 1.

	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 fraction of the free-flow speed

To account for the degree of urbanization, we multiply the maximum speed drops with a fraction. Table 2 presents these fractions, in which degree 1 represents the highest degree of urbanization.

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

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 (×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

Vehicle route plans are generally constructed in two phases. First, optimal travel times and distances between customer locations are calculated by solving a (time-dependent) shortest path problem for each pair of locations. Second, a vehicle routing problem (VRP) is solved based on the calculated travel times in phase one. The classical VRP can be formulated as follows. Given a homogeneous vehicle fleet located at a central depot and a set of customer locations, find an optimal set of vehicle tours, each starting and ending at the depot, such that all customers are served by exactly one vehicle. Several variations on the classical VRP have been proposed to account for various practical restrictions, such as multiple depots, heterogeneous vehicle fleets, capacity restrictions, and customer service time windows. In this paper, we consider the VRP with capacity restrictions and customer service time windows, which implies that the total demand along each vehicle tour may not exceed the vehicle's capacity, and that each service must start within the given time window at the customer. Moreover, we consider time-dependent travel times to account for traffic congestion effects.

We consider four different strategies in which congestion avoidance is applied to an increasing extent. In the first two strategies, we model travel times as time-independent. Strategy 1 completely ignores traffic congestion, Strategy 2 accounts for traffic congestion by including some slack travel time. In the other two strategies we consider time-dependent travel times. We first describe the four strategies. Next, we validate our choices in Section 3.1.

Strategy 1 ignores traffic congestion. In this strategy, arc speeds are set to their maximum value. Time-independent travel times between customers are obtained using a shortest path algorithm. This corresponds to the classic VRP solution.

Strategy 2 accounts for traffic congestion by solving a VRP based on average travel times. The shortest paths are obtained in the same way as in Strategy 1. Next, average travel times are calculated for each path, which serve as input for the VRP. The average travel times are obtained by calculating the exact travel times for a large number of different departure times (the inter-departure times are set to 15 minutes) and taking the average.

Strategy 3 avoids traffic congestion by solving a TDVRP. The time-dependent travel times are obtained in a similar way as with Strategy 2, with the difference that the travel times (with inter-departure times of 15 minutes) are not averaged. Interpolation is used each time the TDVRP solver requires the travel time between two customers for a given departure time.

Strategy 4 avoids traffic congestion by solving a TDSPP and a TDVRP. Time-dependent travel times are obtained by using a time-dependent shortest path algorithm. The travel times are calculated for the same inter-departure times as with Strategies 2 and 3. Again, interpolation is used to calculate the travel time for a given departure time when the TDVRP is being solved. Strategy 4 requires an extra phase to determine the complete vehicle route plan. Since the planned departure times in the TDVRP solution 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 these planned departure times in the last phase. Table 5 gives an overview of the four strategies.

Strategy	Shortest paths	Travel times input for VRP	Accounting for congestion	Avoiding congestion
1	time-indep.	time-indep.	no	no
2	time-indep.	average	yes	no
3	time-indep.	time-dep.	yes	yes
4	time-dep.	time-dep.	yes	yes

Table 5: Strategy overview

3.1 Validation

After the VRP instances are solved with each strategy, we evaluate the performances of the resulting vehicle route plans according to the speeds resulting from the speed model. Note that with Strategies 2, 3, and 4 travel time estimations are used to solve a (TD)VRP. The reason for using travel time estimations with Strategy 3 is that calculating the exact travel times would be too time-consuming. With this strategy, travel time calculations depend on the number of arcs along the route under consideration and, therefore, generally take much longer than using interpolation. In practice, however, the TDVRP needs to be solved within limited computation times (e.g., customer demands are often known only one day ahead), such that exact travel time calculations allows the TDVRP solver to compute fewer solutions. Calculating exact travel times with Strategy 4 would require a different shortest path algorithm with a higher complexity than with the other strategies (Orda and Rom [16] show that calculating time-dependent shortest paths for a given departure time can be done with a standard Dijkstra [17] search, whereas calculating time-dependent shortest paths for all departure times requires a shortest path algorithm of type Ford and Fulkerson [21]). Finally, we choose to use a similar mechanism to derive travel time estimations with Strategy 2 as with Strategies 3 and 4 for clarity reasons.

The required computation times in phase 1 differ with each strategy. Strategy 4 requires the longest computation times since with this strategy more shortest paths have to be determined (i.e., for multiple departure times). Since companies generally have a (partially) fixed set of customers, phase 1 is only fully deployed once for the development of multiple vehicle route plans. Only when customers are added to the company's customer set, or when road network changes, phase 1 has to be (partially) deployed again. Moreover, shortest paths can often be determined in a pre-processing phase (e.g., customer demands do not have to be exactly known yet, only there location), such that computation times in phase 1 play only 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 a shortest path algorithm that can solve both SPPs and TDSPPs, and a VRP algorithm that can solve both VRPs and TDVRPs, without tailoring the solution methods. 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.

We solve the (TD)SPPs with a modified Dijkstra [17] 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 Strategies 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 non-passing property 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. The speed model satisfies the non-passing property, since vehicles traversing the same arc at the same time drive the same speed.

We provide (TD)VRP solutions using a restricted dynamic programming (DP) heuristic, introduced by Gromicho et al. [22]. We select this method for two reasons. First, Kok et al. [23] show that this DP heuristic is successful in solving vehicle routing problems with complex timing restrictions. In addition, the DP heuristic can evaluate the same number of VRP solutions as TDVRP solutions within similar computation times, without tailoring the algorithm. This is not only very valuable for practical use, but it also provides a fair comparison of the quality of the different strategies in practice. Second, ORTEC - a key-player in the vehicle routing systems market - has implemented this DP heuristic in their software, because of its high performance on practical vehicle routing problems. Since we aim at gaining insight in which strategies are best in practice, it is highly valuable to use methods that have proved their value in practice. We provide a short explanation of the restricted DP heuristic of Gromicho et al. [22].

The restricted DP heuristic for the VRP is based on the exact DP algorithm for the TSP of Held and Karp [24] and Bellman [25]. 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. [26]. 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.

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 [7]. 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. Gromicho et al. [22] show that this restriction on the number of state expansions results in a running time complexity of $O(nHlog(H))$ (where n equals the number of customer nodes and vehicle origin and destination nodes). We store in each state the departure time from the last visited node, such that we can apply the restricted DP heuristic 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 parameter settings are motivated from practice through discussions with ORTEC.

In practice, customer locations are often clustered in urban areas, i.e., the density of customer locations is higher in urban areas than in rural areas. Since the density of nodes in the road network show a similar effect, we randomly select nodes as customer locations to obtain a realistic spread of the customer locations. To obtain a diverse set of problem instances, we randomly select a node in the road network as the depot location.

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. In our setting, computation times are practical if they are limited to a few minutes, since we consider a planning horizon of one day. In practice, customer demands are in this setting often known one day in advance at the end of that day, and planners are often only willing to wait a few minutes for a response from a VRP solver.

We add time windows to 50 percent of the customers indicating the period in which service must start. In practice, some customers require strict service time intervals, whereas other customer sites are open all day. Selecting 50 percent of the customers to have a strict time window gives a fair mix. 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 durations of the time windows at the selected customers are randomly drawn from $\{2, 3, 4, 5, 6\}$ quarters of an hour. ORTEC indicated that these bounds on the time window durations (2 and 6 quarters of an hour) are common in practice. We randomize the durations to obtain a diverse set of problem instances. In practice, time window violations are sometimes allowed. However, it is very hard to quantify the costs of such violations, since these costs are often subject to subjective measures (e.g., relationship with the customer). Therefore, ORTEC's vehicle route optimization software only considers strict service time windows (on which planners may set tolerances, but then again these tolerances are considered strict), and we follow this approach.

The customer service times are randomly drawn from $\{1, 2\}$ quarters of an hour. Again, ORTEC indicated that with day planning these service time durations are common. 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. To obtain a diverse set of problem instances in which both restrictions play a role, we set the capacity such that sometimes the time windows are restrictive and sometimes the capacities. Initial experiments indicated that a capacity of 55 suits 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.

In practice, often multiple objectives are considered when optimizing vehicle routes. ORTEC indicated that a minimal number of vehicles and a minimal total duty time of the truck drivers are generally considered the most important ones. To handle multiple objectives, they use a lexicographical objective function. We follow a similar approach by setting the primary objective to minimize the total number of vehicles used, 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 (note that in general drivers are also paid during waiting times, so they cannot be ignored). To use this lexicographic objective function, we set the cost factor of the states in the DP heuristic to a tuple (number of vehicles used, total duty time), where two states are first compared with respect to the first element (number of vehicles used) and in only case of equality they are compared with respect to the second element (total duty time).

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. [27]), 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, except for the problem instances generated on Kentucky, since we will use these instances for comparison (recall that Kentucky is much less urbanized than the other states). Between brackets, we present the relative changes 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. For practice, all 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.

Problem size	Performance measure	Strat. 1[a]	Strat. 2[b]	Strat. 3[b]	Strat. 4[b]
15 cust.	# vehicle routes	3.06	16.34%	0.00%	0.00%
	Total duty time (hrs)	28.9	5.45%	-6.97%	-8.55%
	Total travel distance	12.36	0.12%	-0.17%	-2.59%
	# late arrivals	2.32	-85.34%	-100.00%	-100.00%
	# late return times	0.050	-60.00%	-100.00%	-100.00%
	Max. late time (hrs)	2.012	-86.26%	-100.00%	-100.00%
	Total late time (hrs)	2.37	-93.12%	-100.00%	-100.00%
50 cust.	# vehicle routes	7.05	10.35%	1.70%	0.99%
	Total duty time (hrs)	69.6	2.67%	-6.44%	-7.41%
	Total travel distance	25.41	-0.41%	-0.89%	0.22%
	# late arrivals	9.51	-77.18%	-99.79%	-99.47%
	# late return times	0.190	-84.21%	-100.00%	-100.00%
	Max. late time (hrs)	2.210	-64.63%	-99.94%	-99.94%
	Total late time (hrs)	8.25	-89.08%	-100.00%	-82.96%
100 cust.	# vehicle routes	13.13	5.94%	0.15%	0.23%
	Total duty time (hrs)	128.1	0.90%	-6.17%	-6.71%
	Total travel distance	42.45	-2.20%	-1.06%	-0.38%
	# late arrivals	14.91	-69.35%	-99.80%	-99.73%
	# late return times	0.270	-48.15%	-100.00%	-100.00%
	Max. late time (hrs)	3.051	-67.56%	-99.97%	-99.97%
	Total late time (hrs)	13.56	-82.68%	-100.00%	-100.00%

[a]absolute figures
[b]relative change compared to Strategy 1

Table 6: Main results, aggregated over all problem instances, except for Kentucky

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

Note that the increase in number of vehicle routes with Strategy 2 reduces with the size of the problem instances (and as a result, also the increase in total duty time reduces with the problem size). The only difference between Strategies 1 and 2 is that in Strategy 2 a VRP is solved based on average travel times instead of free-flow travel times. As a result, it often happens that a feasible solution with Strategy 1 is not feasible with Strategy 2, such that extra vehicle routes are required with Strategy 2. Since the number of solutions exponentially increases with the problem size, the relative increase in extra vehicles required reduces with the problem size.

The duty times show a similar pattern, which can be partially explained by the number of vehicle routes. Furthermore, the additional congestion information with Strategies 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 Strategies 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 extra duty times caused by traffic congestion. Table 7 presents the average amount of extra 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% of the total duty time is due to traffic congestion delays. Strategy 2 results in larger extra duty times than Strategy 1, due to the extra extra vehicles required with Strategy 2. Strategies 3 and 4, however, reduce the extra duty time of Strategy 1 substantially by 75% and 87% on average, respectively.

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. Note that if all roads are of the same type, or if they would all behave similarly during peak hours, this tendency would not be present.

The reliability of the route plans strongly increases if the level of con-

Problem size	Strat. 1[a]	Strat. 2[b]	Strat. 3[b]	Strat. 4[b]
15 customers	2.70	58.28%	-74.78%	-91.30%
50 customers	6.22	29.86%	-72.04%	-82.96%
100 customers	10.05	11.47%	-78.71%	-85.58%

[a]absolute figures
[b]relative change compared to Strategy 1

Table 7: Average extra duty time (hrs) caused by traffic congestion

gestion 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 8 presents the results for the sixth state: Kentucky. We scale the results of the 15-, and 50-customer problem instances to 100-customer problem instances by multiplying the performance measures (except for the maximum late time) with 100/15 and 100/50, respectively. We report averages over all problem instance, except for the maximum late time which is the (unscaled) maximum late time over all problem instances. 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 8 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 extra 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 total duty time, and may still substantially increase the reliability of the vehicle route plans.

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%
extra duty time (hrs)	0.992	91.65%	-10.75%	-83.94%

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

Table 8: Change of Strategies 2, 3 and 4, relative to Strategy 1 for Kentucky

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 at one city or dispersed across many cities), difference in speed limit between trucks and cars, and even culture. To verify the robustness of the results obtained the speed model, we conduct a sensitivity analysis by repeating the computational experiments for a number of extreme alternatives: a) only speed drops on highways, b) all roads have the same maximum relative speed drops, c) the same speed drops during the morning and evening peak, d) small speed drops. The first three alternatives verify the robustness with respect to the three key elements of peak hour traffic congestion, the last alternative verifies the robustness with respect to the amount of traffic congestion. Table 9 presents the maximum relative speed drops for the four alternatives.

We run all computational experiments again for all alternatives. We compare the results for the four alternatives with the results for the original speed drops. We do not report the results with respect to the reliability of the route plans, since they are similar to the results with the original speed drops.

(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 9: 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 have the same maximum relative speed drops, c) the same speed drops during the morning and evening peak, d) small speed drops.

Table 10 presents the results for Alternative (a) 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 needed with Strategy 4 with respect to Strategy 3 is almost twice as big as this reduction 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 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.

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 10: Results Alternative (a): only speed drops on highways.

Table 11 presents the results for Alternative (b) 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 because of the bigger speed drops, on average. 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, 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 smaller *absolute* speed drops on roads with lower maximum speeds, which typically appear more often on alternative paths.

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 11: Results Alternative (b): all roads have the same maximum relative speed drop.

Table 12 presents the results for Alternative (c) 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 (c) 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 13 presents the results for Alternative (d) 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 caused by traffic congestion is also smaller. As a consequence, Strategy 3 and 4 find solutions with approximately the same number of vehicle routes

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 12: Results Alternative (c): speed drops during the morning and evening peak are similar.

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 noticed 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.

6 Conclusions

We compared four strategies for avoiding traffic congestion by developing better vehicle route plans. We proposed a speed model on real road networks that reflects the key elements of peak hour traffic congestion. We used this speed model to generate a set of realistic VRP instances for testing the impact of the different strategies.

The test results indicated that the reliability of route plans strongly in-

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 13: Results Alternative (d): the speed drops are half the original speed drops.

crease 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. We thank the anonymous referees for their helpful comments to improve this paper.

References

[1] D. Schrank and T. Lomax. The 2007 urban mobility report. *The Texas A&M University System*, pages 1–53, 2007.

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

[3] Niklas Kohl, Jacques Desrosiers, Oli B. G. Madsen, Marius M. Solomon, and Francois Soumis. 2-Path Cuts for the Vehicle Routing Problem with Time Windows. *TRANSPORTATION SCIENCE*, 33(1):101–116, 1999.

[4] Alain Chabrier. Vehicle routing problem with elementary shortest path based column generation. *Computers & Operations Research*, 33(10):2972 – 2990, 2006. Part Special Issue: Constraint Programming.

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

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

[7] 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.

[8] Johannes Bentner, Günter Bauer, Gustav M. Obermair, Ingo Morgenstern, and Johannes Schneider. Optimization of the time-dependent traveling salesman problem with monte carlo methods. *Phys. Rev. E*, 64(3):036701, Aug 2001.

[9] Johannes Schneider. The time-dependent traveling salesman problem. *Physica A: Statistical Mechanics and its Applications*, 314(1-4):151 – 155, 2002.

[10] 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.

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

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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

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

[19] English Highways Agency of the Department for Transport. Traffic Forecastes. World Wide Web, last checked on October 6$^{\text{th}}$ 2010, 2010. http://www.highways.gov.uk/traffic/forecast.aspx.

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

[21] L. R. Ford and D. R. Fulkerson. *Flows in Networks*. Princeton University Press, Princeton, N.J., 1962.

[22] J. Gromicho, J. van Hoorn, A. L. Kok, and J. M. J. Schutten. Restricted dynamic programming: a flexible framework for solving realistic VRPs. *Beta working paper series*, 266, 2008. http://beta.ieis.tue.nl/node/1154.

[23] A. L. Kok, C. M. Meyer, H. Kopfer, and J. M. J. Schutten. A dynamic programming heuristic for the vehicle routing problem with time windows and european community social legislation. *Transportation Science*, Articles in Advance (September 24):1–13, 2010.

[24] 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.

[25] R. E. Bellman. Dynamic programming treatment of the travelling salesman problem. *Journal of the Association for Computing Machinery*, 9(1):61–63, 1962.

[26] 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.

[27] A. L. Kok, E. W. Hans, and J. M. J. Schutten. Optimizing departure times in vehicle routes. *European Journal of Operational Research*, to appear.