COPING WITH TIME VARIABILITY IN ROUTING PROBLEMS

Tom Van Woensel\textsuperscript{1}, Christophe Lecluyse\textsuperscript{2},
Herbert Peremans\textsuperscript{2}, L. Kerbache\textsuperscript{3} 
and N. Vandaele\textsuperscript{2}

\textsuperscript{1} Eindhoven University of Technology, 
Den Dolech 2, Pav. F05, Eindhoven NL 5600 MB, 
The Netherlands, t.v.woensel@tm.tue.nl
\textsuperscript{2} University of Antwerp, Department of Applied Economics, 
Belgium
\textsuperscript{3} HEC School of Management, 78351 Jouy-en-Josas, France

Abstract

Despite numerous publications on vehicle routing, very few addressed the inherent stochastic nature of this problem. In this paper, a vehicle routing problem with time windows and dynamic travel times due to potential traffic congestion is considered. The approach developed introduces mainly the traffic congestion component based on queueing theory. A test problem is used for illustrative purposes as well as for a discussion around the feasibility of proposed solutions if travel times were not constant.

Keywords: routing problems, time variability, congestion

1 Introduction

Routing problems have been largely studied due to the interest in its different applications in logistic and supply chain management. Not surprisingly, transportation is an important component of supply chain competitiveness since it plays a major role in the inbound, inter-facility, and outbound logistics. The vehicle routing problem (VRP) aims to construct a set of shortest routes for a fleet of vehicles of fixed capacity. Each customer is visited exactly once by one vehicle which delivers the demanded amount of goods to the customer. Each route has to start and end at a depot, and the sum of the demands of the visited customers on a route must not exceed the capacity of the vehicle. Another constraint occurring in the real world is that the customer may specify time intervals in which he will be able to receive the deliveries. This additional restriction leads to the vehicle routing problem with time window constraints.
(VRPTW). The time windows can be either soft or hard indicating whether the time windows can be violated or not (Laporte [4]).

Past results showed that the total travel times can be improved significantly when explicitly taking into account congestion during the optimization. Explicitly making use of the time-dependent congestion results in routes that are (considerably) shorter in terms of travel time. The average improvement achieved was 25.4% when using tabu search (Van Woensel et al. [12]). The assumption that everything in transportation goes according to a schedule is unrealistic which will result in a planning gap (i.e. the difference between the planned route and the actual route). The main consequence of taking into account a congestion function in the route planning is that the planning gap will decrease as the planned route is then much more realistic compared to the classic VRP route only optimized in terms of distances. Some other interesting aspects of the model can be described in detail: how different factors can affect the solutions obtained and the importance of taking into account dynamic travel times. This is important not just because speed profiles can affect the objective(s) of the optimization, but also because best solutions known for a static problem, applied in a dynamic world, are in general infeasible. The analytical approach to congestion based on queueing models however not only allows for the calculation of the expected travel times but also the variance of the travel time can be approximated by queueing theory. Explicitly taking into account the variance of the travel times then allows for evaluating routes on the risk involved. When doing so, depending on the risk profile of the manager/customer (risk averse, neutral or seeking), the planned route could be substantially different. In this paper, the groundwork is laid for going to more advanced models taking into account the variance of the travel times.

In the dynamic routing problem, the key issue is the computation of the travel times on arcs dependent upon the time period. The travel time $T_{ijp}$ during time period $p$ is determined as the distance from $i$ to $j$ divided by the speed on the arc or:

$$T_{ij}^p = \frac{d_{ij}}{v_{ij}^p}$$

Hence, to determine the travel time on arc $(i, j)$, one needs information on the distance between $(i, j)$ and on the travel speed for that arc at time $p$: $v_{ijp}$. The distance is readily available in the static VRP models, but the speed is a new variable that needs to be specified. In this paper, the dynamic speeds are obtained using queueing models for traffic flows (Vandaele et al. [6] and Heidemann [3]). The remainder is organized as follows: first, the queueing approach to traffic flow is demonstrated, then the procedure to obtain the expected travel time is explained in detail, next the determination of the variance of the travel time is elaborated and finally, the probability of meeting a time window is obtained.
Figure 1: The relations between the speed-flow, the speed-density, and the flow-density diagrams

2 Queueing approach

It is often observed that the speed for a certain time period tends to be reproduced whenever the same flow is observed. Based on this observation, it seems reasonable to postulate that, if traffic conditions on a given road are stationary, there should be a relationship between flow, speed, and density. This relationship results in the concept of speed-flow-density diagrams. These diagrams describe the interdependence of traffic flow ($q$), density ($k$) and speed ($v$). The seminal work on speed-flow diagrams was the paper by Greenshields in 1935 [2]. Using well-known formulas of queueing models, these speed-flow-density diagrams can be constructed (Figure 1).

Figure 1 illustrates that, although every speed $v$ corresponds with one traffic flow $q$, the reverse is not true. There are two speeds for every traffic flow: an upper branch ($v_2$) where speed decreases as flow increases and a lower branch ($v_1$) where speed increases. Intuitively it is clear that, as the flow moves from 0 (at maximum speed $v_f$) to $q_{max}$, congestion increases but the flow rises because the decline in speed is over-compensated by the higher traffic density. If traffic tends to grow past $q_{max}$, flow falls again because the decline in speed more than offsets the additional vehicle numbers, further increasing congestion (Daganzo [1]). The flow-density diagram and the speed-density diagrams are an equivalent representation and can be interpreted in the same way.

In a queueing approach to traffic flow analysis, roads are subdivided into
The total time in the system $W$ is then different depending upon the queueing model used. The total time in the system $W$ is then the sum of the waiting time $W_q$ and the service time $W_p$, or $W = W_q + \frac{1}{k_j v_f}$. Table 1 shows the specific form of $W_q$ for the general queueing models.

Results show that the developed queueing models can be adequately used to model traffic flows (Van Woensel and Vandaele [13]). Moreover due to the analytical character of these models, they are very suitable to be incorporated in other models, e.g. the VRP. For a more detailed discussion of the queueing
models and their results, the interested reader is referred to Vandaele et al., [6] and Van Woensel et al. [11]. In general, formula 2 can be rewritten in the following basic form:

\[ v = \frac{v_f}{1 + \Omega} \]  

Formula 3 shows that the speed is only equal to the maximum speed \( v_f \) if the factor \( \Omega \) is zero. For positive values of \( \Omega \), \( v_f \) is divided by a number strictly larger than 1 and speed is reduced. The factor \( \Omega \) is thus the influence of congestion on speed. High congestion (reflected in a high \( \Omega \)) leads to lower speeds than the maximum. The factor \( \Omega \) is a function of a number of parameters depending upon the queueing model chosen: the traffic intensity, the coefficient of variation of service times and coefficient of variation of interarrival times. High coefficients of variation or a high traffic intensity will lead to a value of \( \Omega \) strictly larger than zero. Actions to increase speed (or decrease travel time) should then be focussed on decreasing the variability or on influencing the traffic intensity, for example by manipulating the arrivals (arrival management and ramp metering).

The major strength of using the queueing models, is that given the flow \( q \), the speed can easily be obtained in an analytical way. The flow \( q \) is a parameter that can determined empirically, allowing to determine realistic velocity profiles as a function of time.

3 The expected travel time

To compute the travel times, one should note that in the dynamic case, the travel speeds are no longer constant over the entire length of the arc. More specifically, one has to take into account the change of the travel speed when the vehicle crosses the boundary between two consecutive time periods. For example, the speed changes when going from time period \( p \) to time period \( (p + 1) \) from \( v_{ij}^p \) to \( v_{ij}^{(p+1)} \). The time horizon is discretized into \( P \) time periods of equal length \( \Delta_p \) with a different travel speed associated to each time period \( p \) (\( 1 \leq p \leq P \)).

The travel speeds are obtained using the above discussed queueing models for traffic flows. Formally, the travel time \( T_{ij}^{p_0} \) going from customer \( i \) to customer \( j \), starting at some time \( p_0 \), must satisfy the following condition:

\[ \int_{p_0}^{p_0+T_{ij}^{p_0}} [v^p] \, dp = d_{ij} \]

With \( v^p \) denoting the speed in time period \( p \) and \( d_{ij} \) the distance traveled. Solving this integral for \( T_{ij}^{p_0} \) and making use of the discrete time horizon, results in:

\[ d_{ij} = \Delta_p \left( \sum_{p_0}^{p_0+T_{ij}^{p_0}} v^p + v^{p_0+1} + \ldots + v^{p_0+(k-2)} \right) + \phi_{last} \]

Rewriting as a function of the time slices, gives:
\[ T_{ij}^{p_0} = \varphi \Delta_p + (k - 2) \Delta_p + \phi \Delta_p \]

The travel time is thus the sum of the following components:

1. The fraction of travel time still available in the first time zone, given by
   \((\varphi \Delta_p)\) with \(\varphi\) the fraction parameter \((0 \leq \varphi \leq 1)\).

2. The duration of the \((k - 2)\) intermediate time zones passed: \((k - 2) \Delta_p\).

3. The fraction of the travel time in the last time zone, given by \((\phi \Delta_p)\), with \(\phi\) the fraction parameter \((0 \leq \phi \leq 1)\).

Concluding, in total \(k\) time buckets are crossed. The number \(k\) is totally defined by the equations in the paper and is a function of the distance \(i\) to \(j\) and the speeds in the different time buckets. In practice, the other values are computed incrementally: Starting at time \(p_0\) (part of the first time bucket), one knows the fraction of time left in the first time bucket and consequently \(\varphi = 1 - \frac{p_0}{\Delta_p}\). Then a number of time buckets \(\Delta_p\) is added as necessary to reach the destination city \(j\). Of course, one will spend in the last time bucket only part of the time or the fraction \(\phi\). This fraction is totally depending upon the residual distance that needs to be travelled in the last time bucket. Using this incremental procedure, the travel time \(T_{ij}^{p_0}\) from customer \(i\) to customer \(j\), starting at time \(p_0\) can be determined easily based on the distance \(d_{ij}\) and the speed \(v_p\) for the different time periods \(p\).

### 4 The variance of the Travel Time

In this section, an expression for the variance of the travel time is derived. Again, using the queueing approach to traffic flow [10], the variance can be obtained in a closed form. For each time period \(p\), the variance of the travel time can be determined as follows (using formulae 1 and 2):

\[ Var(T_{ij}^p) = d_{ij}^2 \ast k_{ij}^2 \ast Var(W^p) \]

The distance from \(i\) to \(j\), \(d_{ij}\) and the jam density of the road \(k_{ij}\) are assumed to be known. The variance of the total time in the system \(W\), can be obtained using the two moment approximations from Whitt. As there is no exact form for the variance of the waiting time, one needs to rely on approximations to obtain the variance of the waiting time (Whitt [8]). These approximations have already proven their value and usability in production management (Vandaele [7]; Whitt [8]; and others). Whitt describes an approximation for the variance of the waiting time for the case when expected waiting time is known (either exact or approximated). Only the results necessary for the analysis of the variance of the waiting time are presented here. The approximation has the following general form:
\[ Var(W^p) = |W_\varphi|^2 c_\varphi^2 + c_\varphi^2 k_\varphi^2 v_\varphi^2 \] (4)

with: \( c_\varphi^2 \) the squared coefficient of variation of the waiting times, which is approximated by the set of formulas described in [8]. This approximation depends largely on the approximation of expected waiting time \( W_\varphi \).

5 Managerial Insights

The routing problem discussed in this paper is motivated by the fact that in many circumstances traffic conditions cannot be ignored in order to carry out a realistic optimization. Computational results suggest that in systems where the constant speed approximation is no longer valid, it is crucially important to explicitly consider the variability due to traffic congestion in the model. The impact of dynamic components will be even more important when relating the approach to urban contexts (see also Taniguchi et al. [5] for more insights on the impact of city logistics on routing decisions). Moreover, the capability of taking into account dynamic travel times is extremely valuable, not only because speeds profiles do affect the objective function of the optimization, but also, as demonstrated above, the best solutions for the static problem applied in a dynamic context, are in general suboptimal. Compared to the actual realization in real-life, the planning gap (difference between plan and actual) can be reduced substantially when taking into account a manifestation of the traffic on the roads. Consequently, due to the resulting reduced planning gap, less time and effort needs to be invested in replanning (in real-time) during the day. This paper aims at obtaining a routing solution that performs well in the face of the extra complications due to congestion, which eventually leads to a better solution in practice. These more realistic solutions have the potential to reduce real operating costs for a broad range of industries which daily face routing problems.

6 Conclusions

In this paper, a dynamic vehicle routing problem with time-dependent travel times due to traffic congestion was presented. Recently, the problem considered has received increasing attention due to its relevance to real-life problems. The approach developed here introduces the traffic congestion component in the standard VRP models. The traffic congestion component was modelled using a queueing approach to traffic flows. By making use of this analytical approach to traffic flows, the necessary data to model congestion is easily obtained and very limited which opens the door for real-life applications. Moreover, when including the variance of the travel time, the potential applications are vast: it gives a manager a powerful tool to incorporate and take into account congestion uncertainty in his optimization.
References


