Scheduling Flexibility and Insertion Zones in Vehicle Routing

Wout Dullaert*
Ufsia-Ruca Faculty of Applied Economics, University of Antwerp
Prinsstraat 13, 2000 Antwerp, Belgium
wout.dullaert@ua.ac.be

Geert Van de Weyer
Ufsia-Ruca Faculty of Applied Economics, University of Antwerp
Prinsstraat 13, 2000 Antwerp, Belgium
geert.van deweyer@ua.ac.be

Abstract
In this paper, scheduling flexibility and insertion zones are formally defined for less-than-truckload (LTL) and full-truckload (FTL) routing. Scheduling flexibility refers to the flexibility that a time window-constrained customer (LTL) or load (FTL) offers to a dispatcher. Insertion zones indicate the area from which a customer (LTL) or load (FTL) can be inserted into a partially finished route.

1 Introduction
Scheduling flexibility refers to the degrees of freedom that a customer gives to a dispatcher to design routes. The higher the flexibility, the more cost efficient the routes the dispatcher can design. Although some authors have reported on certain aspects of scheduling flexibility in routing before, no systematic approach has been published. In this paper, we try to lay the foundation for a systematic approach to scheduling flexibility.

In the case of less-than-truckload (LTL) routing, scheduling flexibility takes the form of time windows in which customers wish to be serviced. These time windows can differ as far as their moment in time and width is concerned. In the case of a full truckload (FTL) routing problem, the dispatchers objective is to service loads between two nodes at minimal distribution costs, instead of servicing individual customers (nodes). The scheduling flexibility of a load is then determined by both the time windows of the pick-up and the delivery nodes. Here a route no longer consists of nodes (i.e. individual unroute customers), but of loaded route segments. A loaded route segments is defined as an arc on which a load is transported between two nodes. Unloaded arcs are used to

*The authors are indebted to G.R. Janssen and C. Van Meeckens for the constructive comments that improved the presentation in the paper. The usual disclaimer applies.
travel between the loaded arcs and from/to the depot. The scheduling flexibility
of the loaded arc is determined by the time windows of its starting and ending
node.

An insertion zone is the area in which a customer (or a load) can be inserted
between two other ones. For LTL routing, the scheduling flexibility of each of
the two adjacent customers in the route determines whether a new customer can
be inserted between them. If the time windows of customers \( i \) and \( j \) offer more
time than needed to service \( i \) and travel directly to \( j \), there may be time left
to insert an unrouted customer \( a \), located in an insertion zone around \( i \) and \( j \).
For the loaded arcs in FTL routing the same reasoning applies as in the LTL
case. The time available to travel from one loaded arc to another, can be used
to insert a new loaded load.

Flexibility and scheduling flexibility issues have been addressed in a num-
ber of fields such as manufacturing, computing, labor economics etc. In the
literature on the Vehicle Routing Problem a number of authors have modelled
situations with different scheduling flexibility or informally referred to its impact
on routing costs. Therefore a brief literature review on scheduling flexibility will
be presented in Section 2. In the next section, scheduling flexibility is defined for
the LTL case and the insertion zones are shown to be elliptic. In Section 4 the
same is done for full-truckload routing with time windows. Insertion zones be-
tween loaded route segments are proven to be also elliptic. Finally, conclusions
are formulated.

2 Literature review on flexibility and scheduling flexibility

Flexibility is an important issue in the contemporary, globalizing economy. In-
creasing competition forces companies to quickly adapt/react to changes in
the environment. Customers' (variable) demand for high-quality, differentiated
products have — through concepts such as Just-In-Time and Total Quality
Management — spurred academics' and practitioners' interest for operational
or productive flexibility.

Flexible manufacturing systems are designed to combine the efficiency of large
scale production production with the flexibility of a job shop environment. They
can produce different products, while keeping setup times short and work-in-
progress inventories low. As a result, a wide range of products can be offered
while keeping costs at an acceptable level. Because a higher need for flexibility
imposes additional constraints on the production process to be optimized, there
is a trade-off between flexibility and efficiency (i.e. production costs). Tarifa
and Chiotti (1995) study this trade-off in the so-called Flexibility Problem. A
bicriteria optimization approach is used to determine the optimal size of a
plant such that it satisfies all constraints for any of the parameters during the
process operation.

Flexible manufacturing often requires a flexible workforce. However, not
only from the employers' side there is a call for flexibility as flexibility promises

---

2See Martinez Sánchez et al. (2000) for a study of the relation between operational flexibil-
ity, JIT and TQM. Reviews on flexibility in manufacturing can be found in Beckman (1990),
Sethi and Sethi (1990), Hyun and Ahn (1992), and Upton (1995).
improved working conditions and more varied and more interesting jobs (Dyer, 1968). Daniels et al. (1999) study the operational impact of both machine flexibility and labor flexibility. In the field of computing, considerable attention is paid to ways to increase the flexibility of real-time systems (Burns and Fohler, 1991; Burns et al., 2000). Also the flexible delivery of (vocational) training has been addressed in the literature (Evans and Smith, 1999; Smith, 2000).

In activity scheduling models for travel demand a customer’s ability (or flexibility) to revise his schedule under new circumstances and the flexibility of the schedule itself to account for new activities is addressed (Venter and Hansen, 1998). While these activities can be scheduled over shorter and longer time horizons, little attention is paid to the impact of the time window sizes on the flexibility of the schedule.

In transport and logistics, operational flexibility is a driving source for product innovation and cost reduction. A few examples are trucks that can combine with different types of semi-trailers (e.g. for bulk or container transport), multi-compartment, multi-temperature semi-trailers to maximize loading flexibility for supermarkets (Clancy, 2000), flexible bulk palletizing for the optimum bulk pallet for each client or job (DeFayette, 1996), automated transport and inventory systems for aluminium coils (Aluminium, 1994), designing space modules (Bastle et al., 1998) are just a few examples.

The impact of novel equipment or technologies on the operational flexibility of a transport firm has received more attention than the impact of customers’ desiderata. If customers are demanding on the moment of delivery, they offer a dispatcher little flexibility to schedule their order. As a result, rigid customers can lead to cost ineffectives schedules with a lot of waiting time and additional distance to be travelled. Because customers in the Vehicle Routing Problem with Time Windows (VRPTW) have to be serviced in a time window of their choice, the nature and size of the time window reflects their scheduling flexibility.

Four types of time windows have been studied — in decreasing amount of research spent on them — hard, soft, one-sided and flexible time windows. If time windows are hard, service has to start within the specified time window. In the soft time window case, a vehicle is allowed to arrive too late at a customer but a penalty is incurred (see e.g. Balakrishnan (1993) and Taillard et al. (1997)). The rationale behind soft time windows is that by allowing a few (small) time window violations, solution quality can be significantly improved. In both the hard and the soft time window case, a vehicle arriving too early has to wait until the start of the service time window. This is not the case if customers have one-sided time windows without an earliest time (Nygard et al., 1988; Thangiah et al., 1995). One-sided time windows offer more flexibility to the dispatcher in that waiting times before customers can be avoided. However, respecting the latest possible time at which service can start, remains a hard constraint. Chou and Ming (1998) generalized soft time windows by putting a bound on the maximum waiting time and lateness. In the resulting flexible time window, no penalties are incurred in the original (hard) time window. Arriving too early or too late, but within the respective bounds, is penalized. In this paper, we focus on scheduling flexibility for hard time windows.

To our knowledge, Dullaert (1999), Doerner et al. (2000) and Dullaert (2001) are one of the few to draw attention to the effect of the size of time windows on routing costs. Doerner et al. (2000) notice for full-truckload routing that cost savings through larger time windows are larger when the original time
windows are rather tight. They also raise the question on how much a customer should be charged depending on his time window preferences. Independent from Doerner et al. (2000), Dullaert (1999) raised the same question and developed a framework to study the relationship between scheduling flexibility and freight rates for less-than-truckload routing (Dullaert, 2001).

3 LTL Routing

The problem of LTL routing is extensively studied in the Vehicle Routing Problem with Time Windows literature (see e.g. Desrosiers et al. (1995)). In the VRPTW capacitated vehicles located at a depot are required to serve geographically scattered customers over a limited scheduling period (e.g. a day). Each customer \(i\) has a known demand \(d_i\) to be serviced (either for pick-up or delivery but not both) at time \(e_i\) chosen by the carrier. If time windows are hard, \(h_i\) is chosen within a time window, starting at the earliest time \(e_i\) and ending at the latest time \(l_i\) that customer \(i\) permits the start of service. In the soft time window case, a vehicle is allowed to arrive too late at a customer but a penalty is incurred. In both cases, a vehicle arriving too early at customer \(j\), has to wait until \(e_j\). If \(t_{ij}\) represents the direct travel time from customer \(i\) to customer \(j\), and \(s_j\), the service time at customer \(i\), then the moment at which service begins at customer \(j\), \(b_j\), equals \(\max(e_j, b_j + s_j + t_{ij})\) and the waiting time \(w_j\) is equal to \(\max(0, e_j - (b_j + s_j + t_{ij}))\). A time window can also be defined for the depot in order to define a 'scheduling horizon', in which each route must start and end (Potvin and Rousseau, 1993).

**Definition 1 (LTL Scheduling Flexibility)** Given a customer \(i\) with service time \(s_i\) and a hard service time window \([e_i, l_i]\), bounded by the earliest and latest time at which service can start. The scheduling flexibility of customer \(i\) is defined as \((l_i - e_i)\).

Scheduling flexible customers enable a dispatcher to design cost efficient routes in two ways. First, the width of their service time windows allow a dispatcher to schedule them efficiently in a partially finished route. Rigid customers are more difficult to schedule and often lead to schedules with a lot of waiting time and distance to be travelled. Second, a flexible customer in a partially finished route, facilitates inserting unvisited customers in the same route. This last effect is demonstrated by the use of insertion distances and insertion zones.

In the VRPTW literature, a route is traditionally represented as a sequence of nodes \((i_0, i_1, i_2, \ldots, i_m)\) with \(i_0 = i_m = \text{depot}\). Each route can also be considered as a sequence of arcs \((i_0, i_1), (i_1, i_2), \ldots, (i_{m-1}, i_m)\). On each arc \((i, j)\) the time a vehicle has to service \(i\), \(s_i\), and travel directly to \(j\), is larger or equal to \(t_{ij}\). If the customers' time windows would not allow this, the route would be infeasible.

**Definition 2 (LTL Insertion Distance)** Given a customer \(i\) with service time \(s_i\) and a hard service time window \([e_i, l_i]\), bounded by the earliest and latest time at which service can begin. If one unit of time equals one unit of distance\(^2\), an

\(^2\)Without loss of generality we make this common assumption (e.g. see Solomon (1987)) to simplify the analysis.
Figure 1: Inserting $u$ between $i$ and $j$

\[ t_i \quad t_{iu} \quad t_{uj} \quad t_j \]

The insertion distance of $l_j - [b_i + s_i] - s_u$ can be travelled to insert a customer $u$ between customers $i$ and $j$.

Consider Figure 1 in which an unrouted customer $u$ is inserted between nodes $i$ and $j$. The route is feasible up to $j$ if

\[
\begin{align*}
    b_i + s + t_{iu} & \leq l_j \\
    b_i + s + t_{iu} + s + t_{uj} & \leq l_j
\end{align*}
\]

(1)

Assuming $s_i = s_u = s$ rewriting the last inequality, the insertion distance becomes

\[ t_{iu} + t_{uj} \leq l_j - b_i - 2s. \]

(2)

**Proposition 1 (TLT Insertion Zone)** If all customers have the same service time $s = s_i = s_u$, the insertion distance $l_j - b_i - 2s$ defines an elliptic insertion zone having customers $i$ and $j$ as its foci. Any unrouted customer located in the elliptic insertion zone, whose time windows are compatible with those of $i$ and $j$, can be inserted between $i$ and $j$ if the vehicle’s capacity permits and if the route remains feasible after $j$.

**Proof 1** Equal the insertion distance $l_j - b_i - 2s$ to $d$. The $d$ units of time can be used to service an unrouted customer $u$ between $i$ and $j$. At each point at the boundary of the insertion area

\[ t_{iu} + t_{uj} = d \]

(3)

The maximum area that can be covered in $d$ units of time is thus bounded by an ellipse having $i$ and $j$ as its foci. By the definition of the distance insertion, the route remains feasible at least up to $j$. Introducing a new customer between $i$ and $j$ can create a push forward on the begin of service of all subsequent nodes in the route. Time feasibility at the successors of $j$ can be checked by Solomon’s (1987) necessary and sufficient conditions for time window feasibility.

To check the time feasibility of the schedule after inserting an unrouted customer, Solomon (1987) develops necessary and sufficient conditions for time feasibility if time windows are hard.

If we denote by $t_{p}^{new}$ the new time at which service begins at customer $i$ at position $p$ after the insertion of customer $u$ in the partially constructed route $(t_0, t_1, \ldots, t_m)$, and if the triangle inequality holds for both distances and travel times, then the push forward in the schedule at customer $t_p$ is defined as:
and
\[ PF_{v} = b_{v}^{nw} - b_{u} \geq 0 \quad (4) \]

and
\[ PF_{v} = \max \left\{ 0, PF_{u} - w_{u-1} \right\}, p < r < m - 1. \quad (5) \]

Solomon (1987) assumes that all vehicles leave the depot at \( v_{0} \) to use the idea of the maximum push forward generated by inserting an unroute customer \( u \) between two adjacent stops \( t_{p-1} \) and \( t_{p} \). The necessary and sufficient conditions for time feasibility when inserting a customer \( u \) between \( i_{p-1} \) and \( i_{p}, 1 < p < m \), on a partially constructed feasible route \( (i_{0}, i_{1}, \ldots, i_{m}), i_{0} = i_{m} = \) depot, are
\[ b_{u} \leq i_{u} \quad \text{and} \quad b_{u} + PF_{v} \leq i_{v}, \quad p \leq r \leq m \quad (6) \]

Indeed, if \( PF_{v} > 0 \), the schedule at customer \( i \) and some of its successors, i.e., customers \( i_{r}, p \leq r \leq m \) may become infeasible. These customers have to be sequentially examined for time feasibility until we find a customer \( i_{r} \) whose waiting time and the one of its predecessors before the insertion of \( u \), has not lifted the push forward, i.e., \( PF_{v} = 0 \), or which is serviced after \( t_{v} \), making the schedule infeasible.

**Example 1** Consider arc \((2,3)\) in Figure 3. The actual travel time from \( i \) to \( j \), \( t_{ij} = 7.21 \). Suppose that the time available to travel between \( i \) and \( j \), \( t_{ij} - b_{i} - 2s = 10 \). The zone in which customers can be served if their time window permits is elliptical, having \( i \) and \( j \) as its foci. The general equation of an ellipse is given by
\[ \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad (7) \]

Because the axes of the ellipse are not parallel to the two coordinate axes, the axes have to be rotated before the equation of the ellipse can be determined. The original axes have to be rotated by an angle of \( \theta \) with respect to the origin. The new coordinates \( x' \) and \( y' \) are obtained as
\[ x' = x \cos \theta + y \sin \theta \quad (8) \]
\[ y' = -x \sin \theta + y \cos \theta \quad (9) \]

Because \( \cos \theta = \frac{a}{\sqrt{a^2+b^2}} = \frac{4}{5} = 0.8 \) and \( \sin \theta = \frac{b}{\sqrt{a^2+b^2}} = \frac{3}{5} = 0.6 \), the original coordinates \((1,4)\) and \((-3,-4)\) are re-expressed as \((3.83,1.387)\) and \((-3.328,1.387)\). The distance between the two foci, \( 2c = 7.21 \). The length of the major axis of the ellipse equals the insertion distance, \( 2a = t_{ij} - b_{i} - 2s = 10 \). Since \( b = \sqrt{a^2-c^2} \) \( b = \sqrt{7.21^2-(3.608)^2} \approx 3.47 \) and all the necessary information to draw the ellipse is found.

**4 Full truckload routing**

In full truckload routing, a route consists of loads (arcs) instead of individual nodes. The scheduling flexibility of an individual arc is determined by the time windows of its starting and ending node.
Figure 2: The insertion zone of arc \((2, 3)\)
Definition 3 (FTL Scheduling Flexibility) Given a loaded arc \((i, j)\) whose nodes have a service time \(s\) and hard service time windows \([v_i, t_i]\) and \([v_j, t_j]\). The scheduling flexibility of the loaded arc \((i, j)\) is defined as \(t_j - b_j - t_{ij} - s\).

Unloaded arcs provide an opportunity to insert loaded arcs. The area in which loaded arcs can be inserted depends on the length of the arc to be inserted.

Definition 4 (FTL Insertion Distance) Given two adjacent arcs \((i, j)\) and \((p, q)\) in a partially finished route. If each node has a service time \(s\), and time windows \([v_w, t_w]\), \(w = i, j, p, q\), then the insertion distance is equal to \(t_p - b_j - t_{pq} - 3s\).

When inserting a load \((m, n)\) between the two adjacent arcs \((i, j)\) and \((p, q)\), the new route is feasible up to \(p\) if (see Figure 2)

\[
\begin{align*}
 b_j + s + t_{jm} & \leq t_m \\
 b_j + s + t_{jm} + s + t_{mn} & \leq t_n \\
 b_j + s + t_{jm} + s + t_{mn} + s + t_{np} & \leq t_p
\end{align*}
\]

(10)

If all three inequalities are satisfied, the insertion distance can be determined by rewriting the last inequality as

\[
t_{jm} + t_{mn} + t_{np} \leq t_p - b_j - 3s
\]

(11)

and the time feasibility of the schedule after \((p, q)\) can be checked by applying Solomon's (1987) necessary and sufficient conditions for time feasibility on the nodes of the arcs.

Proposition 2 (FTL Insertion Zone) Given 2 route segments \((i, j)\) and \((k, l)\) with each node having a service time \(s\), and time windows \([v_w, t_w]\), \(w = i, j, k, l\). The insertion distance \(t_k - b_j - 3s\) defines an elliptical insertion zone, having nodes \(j\) and \(k\) as its foci. Any unrouted load which is located in the ellipse, can be inserted between \(j\) and \(k\) if the route remains feasible after \(j\).
Figure 4: Inserting \((p,q)\) between \((i,j)\) and \((p,q)\)

\[
\begin{align*}
& t_{jp} + t_{pq} + t_{qk} \leq d \\
& t_{jp} + t_{pq} + t_{qk} \leq d \\
& t_{jp} + t_{pq} + t_{qk} \leq d
\end{align*}
\]

**Proof 2.** Consider in Figure 4 a load \((p,q)\) that falls within the insertion zone between \((i,j)\) and \((k,l)\). If we denote the insertion distance \(t_k - b_j - 3s\) by \(d\), then

The distance from \(p\) to \(k\), \(t_{pk}\), is according to the triangle inequality smaller than \(t_{jk} + t_{jk}\). But then \(t_{jp} + t_{pk} \leq d\) and \(p\) is contained in an ellipse with foci \(j\) and \(k\).

Along the same lines, the distance from \(j\) to \(q\), \(t_{jq}\), is smaller than \(t_{jq} + t_{jk}\) and therefore \(t_{jq} + t_{jq} \leq d\). As a result also \(q\) lies within the ellipse with foci \(j\) and \(k\). \(\blacksquare\)

If nodes are inserted instead of loads, \(t_{pq} = 0\) and only a single service time is inserted between \(j\) and \(k\). As a result, the full truckload insertion distance reduces to the LTL insertion distance, making full truckload routing a special case of LTL routing.

5 Conclusions and directions for further research

Although some authors have already addressed the impact of the size of time windows on solution quality, this paper contains the first formal analysis of scheduling flexibility in routing. Scheduling flexibility has been defined for LTL and FTL routing. As more flexible customers (loads) give a dispatcher more degrees of freedom in designing cost efficient routes, it can be profitable to design a multidimensional tariff\(^2\) on the scheduling flexibility and another characteristic (e.g., weight or distance travelled).

Flexible customers are easier to schedule efficiently, but also facilitate inserting unrouted customers in a route. Insertion distances and insertion zones

\(^2\)See Wilson (1993, pp. 211-235) for an excellent introduction to multidimensional pricing
are used to demonstrate this effect. The insertion distance is defined as the distance that can be travelled between two nodes (two adjacent customers or the beginning or ending node of two adjacent loads). If arcs of length 0 (i.e., nodes) are inserted, the FTL insertion distance simplifies to the LTL insertion distance. This makes the FTL routing problem a special case of LTL routing. The insertion distance can be used to determine the elliptic zone from which customers (or loads) can be inserted in the route.

References


