objective can be expressed alternatively as Σ_{ijk}
$_j x_{ijk} + \Sigma_k\ a_2(t_{0,f}^k - t_{0,s}^k)$. Even in situations where
cost per unit time is sunk (ex: driver paid a salary),
time cost must be considered. Waiting time contributes
he elongation of total schedule time which can lead to
time periods. One way to express this contingency in
formulation is to let $c_{ij}=a_1 d_{ij} + a_2[(t_j-T)^+-(t_i-T)^+]$.
can restrict or even eliminate overtime periods by
roducing length of route-time constraints. Such
straints on the length of the scheduling horizon arise
quently in practice from legal or contractual conditions.
ey can be handled very easily in the above formulation by
tting the depot due date d_0 equal the desired route-time
d $d_i=d_0-t_{i0}-s_i$.

The above discussion underscores the importance of the
iting time costs in routing and scheduling problems with
ime windows. In certain cases, the minimization of total
aiting time is even the primary objective (see Pullen and
ebb(1967)). To minimize this criterion, one can define the
aiting time at j as $w_j=[e_j- \Sigma_{ik}(t_{ij}+s_i+t_i)\ x_{ijk}]^+$ and let
$_{ijk}=(e_j-t_i-s_i-t_{ij})^+$.

Our model has several other important features. For
example, it is not always the case that $e_0=0$ and all
vehicles leave the depot at time zero, i.e. $t_{0,s}^k=0$, for
all k. In some situations, vehicles can be used for

additional trips. By defining the model variables for routes rather than for vehicles, a solution to the model will provide the starting and ending time of each route, and, hence, a subsequent assignment of vehicles to routes can be made. Different allowable working periods for the vehicles can be implemented through the use of multiple time windows on the depot $e_0^k \leq t_{0,p}^k \leq d_0^k$, p=s,f. Multiple time windows can also be easily incorporated in the model.

As mentioned above, vehicle routing and scheduling problems involve a predetermined number of vehicles, say K, all of the same capacity. It is therefore assumed that the fixed vehicle costs are sunk, i.e. the fleet of vehicles has already been aquired, and only variable routing and scheduling costs need to be considered. This tactical or operational problem becomes a concern only after the strategic problem of how many vehicles to purchase or lease is solved. In the latter case, both fixed vehicle and variable routing and scheduling costs need to be considered. Our model can accomodate this fleet size problem. To see this, let $f_k=f$, for all k, be the fixed vehicle cost and $c'_{0jk}= f_k+c_{0jk}$. The number of vehicles, K, now becomes a decision variable, and one needs to impose the additional constraint $K^* \leq K \leq n$, where K^* is a lower bound on the number of vehicles used. One possible value for K^* is given by $\sum_{k=1}^{K^*-1} b_k < \sum_{i=1}^{n} a_i \leq \sum_{k=1}^{K^*} b_k$.

If f_k is sufficiently large, it is never worthwhile to increase the number of vehicles in order to reduce the variable routing and scheduling cost. In such cases, the model will first minimize the number of vehicles and for this number, it will minimize the variable cost. Otherwise there exist tradeoffs between fleet size and operating costs. A generalization of the vehicle fleet size problem is the fleet size and mix problem, where vehicles of different costs and capacities have to be considered. The solution to this problem would then precede that of the routing and scheduling problem if this latter problem considers a non-homogeneous fleet.

Considering now the computational complexity of problems with time windows, since the (VRP) is NP-hard (see Lenstra and Rinnooy Kan(1981)), by reduction, the (VRSPTW) is NP-hard. As such, it is very improbable that an optimal solution could be found in polynomial time. Therefore, the development of heuristic algorithms for this problem class is of primary interest. We focus on this topic in the following section.

3.2 Heuristics for (VRSPTW)

We now decribe approximate methods for the solution c (VRSPTW) formulated in the preceding section. Of primar importance to the effectiveness and efficiency of suc methods is the way in which the time window constraints ar incorporated in the solution process. Since we wil concentrate on route building procedures, let us firs examine the necessary and sufficient conditions for tim feasibility when inserting a customer, say 1, as the firs customer on a partially constructed route. Let (0,0) (f_1,t_{f_1}), $(f_2,t_{f_2}),\ldots,$ (f_m,t_{f_m}), $(0,t_0)$ be the schedule fo the partial route currently having $f=f_1$ as the firs customer. Denote by S(i) the set of successors of i on it route. An immediate necessary condition is that the nev arrival time at f_1 be less than the latest time for delivery at f_1, i.e.

$$t^n_{f_1} = t_1 + s_1 + t_{1f_1} \leq d_{f_1} \qquad (3.17)$$

If $t^n_{f_1} > e_{f_1}$, then the arrival times at all customers in $S(f_1)$ could potentially be pushed forward in time, possibly making some of them infeasible. Define the push forward in the schedule at f_1 as:

$$PF_{f_1} = t^n_{f_1} - t_{f_1}$$

$t_{f_2} > e_{f_2}$, then $PF_{f_2} = PF_{f_1}$, and hence the new arrival me at f_2 is:

$$t^n_{f_2} = t_{f_2} + PF_{f_2}$$

ı this case, the push forward is fully propagated at f_2 see the diagram below).

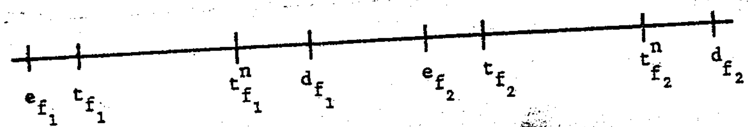

Diagram 3.1

We must require that:

$$t^n_{f_2} \leq d_{f_2}.$$

If $t_{f_2} \leq e_{f_2}$ and $w_{f_2} < PF_{f_1}$ (where w_i is the waiting time at i) ,then $PF_{f_2} = PF_{f_1} - w_{f_2}$, i.e. some of the push forward is absorbed by the waiting time at f_2. We must have that:

$$t^n_{f_2} = t_{f_2} + PF_{f_2} \leq d_{f_2}.$$

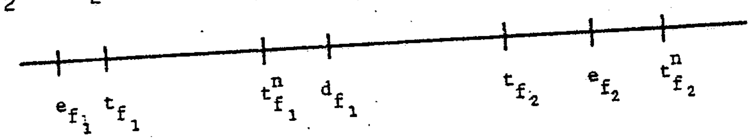

Diagram 3.2

If $t_{f_2} \leq e_{f_2}$ and $w_{f_2} \geq PF_{f_1}$,then obviously $PF_{f_2} = 0$, and therefore, all the customers in $S(f_2)$ will remain feasible. In other words, the push forward in the schedule is

completely absorbed at f_2 (see diagram 3.3, below).

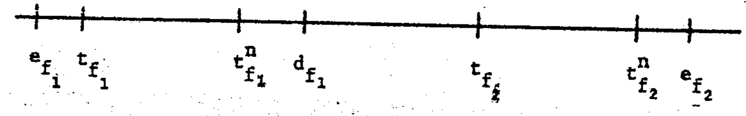

Diagram 3.3

It is clear by a simple induction argument that the process of sequentially examining the customers in $S(f_1)$ for time feasibility continues until we find some customer, say f_p, $p < m$, such that $PF_{f_p} = 0$, or the customer is time infeasible, or, in the worst case, all the customers in $S(f_1)$ are examined. We have just proved:

Lemma 3.2. The necessary and sufficient conditions for a link (l,f) to be time feasible are that:

$$t^n_{f_1} \leq d_{f_1}, \text{ and}$$

$$t_{f_i} + PF_{f_i} \leq d_{f_i}, \qquad (3.18)$$

for all $f_i \varepsilon S(f_1)$ such that $PF_{f_i} > 0$, $1 \leq i \leq m$, where

$$PF_{f_i} = \begin{cases} PF_{f_{i-1}}, & \text{if } t_{f_i} > e_{f_i} \\ PF_{f_{i-1}} - w_{f_i}, & \text{if } t_{f_i} \leq e_{f_i} \text{ and } w_{f_i} < PF_{f_{i-1}}. \end{cases}$$

Let us now describe several heuristic methods for the solution of (VRSPTW).

3.2.1 Savings Heuristics

An approach to solving (VRSPTW) is to extend the savings heuristic of Clarke and Wright(1964), originally proposed for the (VRP). This is a tour building heuristic. Such heuristics are characterized by the sequential addition of a link between two customers, guided by some measure of cost savings, until all customers have been assigned to some route. For the savings method this measure is:

$$s_{ij} = c_{i0} + c_{0j} - \theta c_{ij}$$

The route shape parameter θ was independently introduced by Gaskell(1967) and Yellow(1970). At every step in the heuristic process, the time and capacity constraints are checked for violation. Whereas, in routing, the direction of travel on a tour is unimportant, we now have to account for orientation, due to the existence of time windows. Two routes with end customers i and j, respectively, have compatible orientations if i is first (last), and j is last (first), as the following diagrams illustrate:

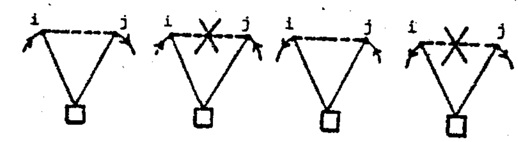

Diagram 3.4 Compatible orientations.

Hence, the admissible links are from the last customer(l) o a route to the first customer(f) on another. We can no describe the extension of the savings approach to th (VRPSTW). Two distinct subtours containing customer i an j, respectively, can be joined if:

a) i and j are not in the same subtour;
b) neither i nor j is an interior point in a subtour;
c) the combined subtour containing both i and j has a total demand not exceeding vehicle capacity;
d) The subtours containing i and j, respectively, hav compatible orientations, and the link (i,j) is tim feasible.

Our computer implementation uses lemma 3.2 to efficiently test time feasibility when linking two customers. In our computational tests, we have found our time feasibility conditions to be clearly superior to the explicit testing of time feasibility at each customer. We have also used list processing and heapsort structures, as proposed by Golden et al.(1977) for the Clarke and Wright heuristic, to reduce the core requirements and computation time. For (VRSPTW), a choice exists as to how to orient the initial partial routes. After preliminary computational experience, we have decided to select the customer with the earlier due date as the first customer on its route.

As presented, the savings heuristic could find it profitable to join two customers very close in distance but far apart in time. Such links lead to the introduction of extended periods of waiting time. As discussed before, waiting time can have a high opportunity cost as far as total schedule cost and the number of vehicles are concerned since the vehicle could be servicing other customers instead of waiting for some customer to open. This suggests the design of a savings approach which accounts for both the spacial and temporal closeness of customers. The enhancement we propose in this direction is to limit the waiting time when joining two customers, i.e., if

$$w_j > W,$$

where W is a prespecified, problem dependent constant, then the link (i,j) should not be used.

3.2.2 A Time Oriented Nearest Neighbor Heuristic

This is a sequential, tour building algorithm. Starting from the depot, this heuristic finds the customer "closest" (in terms of a measure to be described later) to the last customer added to the route. This search is performed among all the customers who can feasibly (with respect to time windows and capacity constraints) be added to the end of the emerging route. A new route is started anytime the search fails, unless there are no more customers to schedule.

The metric we are using tries to account for both geographical and temporal closeness of customers. It is a measure of the distance between two customers, the time difference between their respective delivery times and the urgency of delivery to a customer, given the current point in the schedule.

Formally, if

d_{ij} = distance between customer i and customer j,

$T_{ij} = \max\{e_j, (t_i + s_i + t_{ij})\} - (t_i + s_i)$

= time interval between the departure from i and delivery at j,

$u_{ij} = d_j - (t_i + s_i + t_{ij})$

= time left until j becomes infeasible given that we are currently at i, then

$$c_{ij} = a_1 d_{ij} + a_2 T_{ij} + a_3 u_{ij}\ , \qquad \text{where } a_1 + a_2 + a_3 = 1.$$

The weights attached to the above criteria are dependent on the tightness of the time windows data.

The following example underlines the importance of using our metric, rather than only a distance oriented metric. The example is illustrated in diagram 3.5. The arc weights depicted below are distances. Assume travel times equal distances.

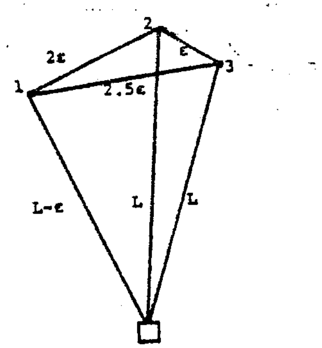

Diagram 3.5.

Let customers 1, 2, 3 have time windows [L-e, L+e], [L+2e, L+4e], [L, L+2e], respectively, where L is a large number, L>>e. We have:

$$c_{12} = a_1(2e) + a_2[(L+2e)-(L-e)] + a_3[(L+4e)-(L-e+2e)]$$
$$= (2a_1 + 3a_2 + 3a_3)e.$$
$$c_{13} = a_1(2.5e) + a_2[(L+1.5e)-(L-e)] + a_3[(L+2e)-(L-e+2.5e$$
$$= (2.5a_1 + 2.5a_2 + .5a_3)e.$$

If $a_1 = 1$, $a_2 = a_3 = 0$, then $c_{12} = 2e$, $c_{13} = 2.5e$, ar hence, the next customer visited will be customer 2. Sir it is time infeasible to add customer 3 at the end of t route (0,1,2) (t_2=L+2e, t_3=L+3e>d_3), the vehicle must retu to the depot and customer 3 will have to be serviced on direct route. On the other hand, by letting a_1=a_2=.4, a_3= , we get: c_{12}=2.6e, c_{13}=2.1e, and hence our heuristic wi find the optimal route (0,1,3,2,0).

3.2.3 Insertion heuristics

This class of sequential, tour building heuristics u: two criteria $c_1(i,u,j)$ and $c_2(i,u,j)$ to insert a ne customer into a route under construction, where i,j ar adjacent customers on the route, and u is yet unrouted. Le $(i_1,i_2,\ldots,i_m)$ be the current route, where $i_1 = i_m = 0$. Fc each unrouted customer, we first compute its best feasibl insertion place in the emerging route as:

$$c_1(i_u,u,j_u) = \underset{p=1,\ldots,m-1}{\text{optimum}} [c_1(i_p,u,i_{p+1})]$$

Inserting u between i_u and i_{u+1} could potentially alter all the arrival times at customers $(i_{u+1},\ldots,i_m)$. One can easily see that this insertion is time feasible if and only if lemma 3.2 holds for $f_1 = u$ and $S(f_1)=S(u)=\{i_{u+1},\ldots,i_m\}$. Next, the best customer u to be inserted in the route is selected as the one for which

$$c_2(i_u{*},u^*,j_u{*}) = \underset{u \text{ unrouted and feasible}}{\text{optimum}} [c_2(i_u,u,j_u)]$$

Customer u^* is then inserted in the route between $i_u{*}$ and $j_u{*}$. When no more customers with feasible insertions can be found, a new route is started, unless all customers have been routed.

Route initialization can have a significant impact on the behavior of a heuristic. Initially, every route is of the form (0,i,0). We have experimented with several criteria for selecting the first customer on the route. These are: the furthest unrouted customer, the unrouted customer with the earliest delivery deadline and the unrouted customer with the minimum weighted combination of direct route time and distance.

Based on the general insertion criteria described above, we have considered three more specific approaches.

i) $c_{11}(i,u,j) = d_{iu} + d_{uj} - a_1 d_{ij}$;

$c_{12}(i,u,j) = t_{j_u} - t_j$,

where t_{j_u} is the new arrival time at j given that u is on the route;

$$c_1(i,u,j) = a_{11}c_{11}(i,u,j) + a_{12}c_{12}(i,u,j),$$

$a_{11}+a_{12}=1$.

$$c_2(i,u,j) = a_2 c_{0u} - c_1(i,u,j)$$

This type of insertion heuristics tries to maximize the benefit derived from servicing a customer on the partial route being constructed rather than on a direct route. The best feasible insertion place for an unrouted customer is the one that minimizes the weighted combination of distance and time insertion. Clearly, different values of a_1 and a_2 lead to possible different criteria for selecting the customer for insertion and its best insertion spot. For example, if $a_{12}=0$ and $c_{0u}=d_{0u}$, one obtains the Mole and Jameson(1976) approach introduced for the (VRP). Furthermore, setting $a_1=1$, $a_2=1$ creates a generalized savings criterion while $1 \leq a_2 \leq 2$, $a_1=a_2-1$, gives a generalized Gaskell's criterion (Gaskell, 1967).

Therefore, the class of heuristics described here, where insertion costs are considered on both distance and time dimensions, are a generalization of the above approaches.

A final comment should be made with respect to the cost of direct routes, c_{0u}. In general, it can be considered as the weighted combination of direct route distance and time cost. Emphasis on the distance component will encourage the early inclusion in the current partial route of customers distant from the depot, while attention to the time component could lead to the insertion of customers close to the depot but with late earliest delivery times.

The second type of insertion heuristics are directed toward the selection of customers whose insertion costs minimize a measure of total route distance and time.

ii) $c_1(i,u,j)$ is defined as before,

$$c_2(i,u,j) = a_d R_d + a_t R_t, \quad a_d + a_t = 1,$$

where R_d and R_t are current estimates of the total route distance and time of the completed route. The distance effect of an insertion on total route distance is independent of what other insertions will be performed

later: the total route distance will increase by $d_{iu}+d_{uj}-d_{ij}$. Hence, we take R_d to be the total route distance resulting from including, say u, in the emerging route. On the other hand, as we have shown in lemma 2.1, the local and global time effects of an insertion on total route time can be different. The local effect, $t_{j_u}-t_j$ is the maximum amount by which the current total route time could be increased by including u. It is conceivable that, if several other customers are added at later stages between u and the depot, the global push forward in the schedule due to u, will equal the local one. This is why we take R_t to equal this maximum total route time value. This choice for R_t encourages the selection of customers who, when inserted, create a small local push forward in schedule. As expected, preliminary computational experience has indicated the superiority of this rule over that of selecting R_t as the actual total route time resulting from inserting u in the current route.

In our third approach, the temporal aspect of the criterion used for insertion also accounts for the urgency of delivery to a customer.

iii) $c_{11}(i,u,j)$ and $c_{12}(i,u,j)$ are defined as before,

$$c_{13}(i,u,j) = d_u - t_u;$$

$$c_1(i,u,j) = a_{11}c_{11}(i,u,j) + a_{12}c_{12}(i,u,j) + a_{13}c_{13}(i,u,j),$$

$_2(i,u,j) = c_1(i,u,j)$,

here $a_{11} + a_{12} + a_{13} = 1$.

It is easy to see that, in fact, this class of heuristics is a generalization of the time oriented nearest neighbor heuristic, in that we allow insertion of an unrouted customer anywhere feasible between a pair of customers on the route, rather than only at the end of the route.

In all the approaches presented, the insertion of unrouted customers is guided by both geographical and temporal criteria. The introduction of time oriented criteria allow these heuristics to account for the time window constraints in the objective function. As a consequence, we expect that the waiting time in the schedules produced by these heuristics will be significantly lower than that produced by distance driven criteria.

Finally, we should remark that the three approaches presented are closely related. One could use the criterion in iii) in any of the other types of heuristics. Moreover, when R_t is estimated using the local push forward , the second approach is in fact a special case of the first, where $a_2=0$.

3.2.4 A Time Oriented Sweep Heuristic.

This heuristic can be viewed as a member of a broad class of approximation methods which decompose the (VRSPTW) into a clustering stage and a scheduling stage. In the first phase, an approximate solution to the nonlinear generalized assignment problem of clustering customers into capacity feasible sets is obtained. We first assign customers to trucks as in the original sweep heuristic (Gillett and Miller(1974)). The idea is to sweep a ray from the central depot serving as the pivot, clockwise or counterclockwise in the plane, to a randomly selected "seed" customer, and add customers to a vehicle as they are encountered until vehicle capacity is exceeded. One can envision other schemes for the partitioning of customers (see Fisher and Jaikumar(1981), Haimovich and Rinnooy Kan(1982)).

In the second phase, we then create a one vehicle schedule for the customers in this sector using one of the insertion heuristics previously described. The insertion heuristic of type 1 was used in our computer implementation. Due to the time window constraints, some customers in this cluster could remain unscheduled.

After eliminating from further consideration the customers that have been scheduled, the above