SOME EXPERIMENTS WITH A SAVINGS HEURISTIC AND A TABU SEARCH APPROACH FOR THE VEHICLE ROUTING PROBLEM WITH MULTIPLE DELIVERYMEN

Vanessa de Oliveira Ferreira* and Vitória Pureza

ABSTRACT. In this work we consider a variant of the vehicle routing problem that allows the assignment of multiple deliverymen to one or more routes. A practical motivation for this variant arises, for example, in the distribution of beverages in highly dense urban areas, characterized by the difficulty in serving daily requests within regular working day hours with a single deliveryman per vehicle. We present a mathematical model and a savings algorithm in order to generate low cost routes that maximize the number of requests served in compliance with the maximum route time. The impact of the extra deliverymen on the solutions provided by the proposed heuristic is assessed by means of sets of generated examples based on classical instances of literature. It is also presented the results obtained by an adaptation of a tabu search approach from the literature.

Keywords: Vehicle routing, multiple deliverymen, heuristics, commercial routing systems.

1 INTRODUCTION

The management of goods and services distribution presents a variety of decisions at strategic, tactical and operational levels. Decisions regarding facilities location, such as plants, depots and warehouses, can be defined as strategic, while the definition of the fleet size and mix is seen as tactic. At the operational level, decisions on routing and scheduling of vehicles and distribution personnel are taken day by day.

In vehicle routing problems the objective is to generate routes for serving demand sites, that is, to provide for each vehicle, the sequence of visits (routing) and the times the activities should take place (scheduling). In some situations, the size of the fleet is also defined. For those problems, a key challenge for practitioners and researchers is to apply or develop methodologies to elaborate the routes so that one or more objectives are optimized. Most vehicle routing problems have combinatorial nature, and it is considered difficult to solve (Lenstra & Rinnooy-Kan, 1981) which explains the continuous research.

*Corresponding author
Universidade Federal de São Carlos – UFSCar, Rodovia Washington Luis, km 235, SP-310 13565-905 São Carlos, SP, Brasil. E-mails: vanoferr@gmail.com / vpureza@dep.ufscar.br

With the advance of computer technology and increasing awareness of the impact of distribution costs on product prices, several commercial routing systems have been developed. In addition to data and location of customers, these systems store road network information, average speed and traffic restrictions. From these data, exact and heuristic methods elaborate efficient routes under different operational conditions (such as soft or hard time windows, multiple depots, heterogeneous fleet, node or arc routing, among others), allowing companies logistic departments to carry out the distribution with satisfactory results. A recent survey on the performance and characteristics of commercial routing software can be found in Hall & Partyka (2008).

Despite these advances, new practical situations arise and define variations of the problem for which existing methods may not be the most suitable. For instance, large service times at each demand site may prevent service to all customers within the established working hours. This situation is faced, for example, by Brazilian companies when delivering beverages to small retailers in urban areas. Due to parking difficulties, demand sites close to each other (cluster) have a parking site associated with the vehicle that will serve them. The delivery of the cargo to each demand site of the cluster is then performed by a deliveryman that visits customers by foot, very often by making more than one trip between the vehicle and each customer. As this market segment is highly competitive, the usual policy is to serve all daily requests on the same day, despite the resulting overtime costs and fines from regulatory agencies, and the inconvenience caused to the customers.

Note, however, that if the number of deliverymen assigned to each route is greater than one, service times at each customer will most likely be reduced, therefore affecting the total service time in the cluster, and increasing the number of customers that can be served within regular working hours. This new reality encourages the study of a problem little or not yet explored in the literature, and here called the vehicle routing problem with multiple deliverymen (VRPMD).

This paper aims to analyze the impact of the assignment of multiple deliverymen in route generation so as to minimize the number of unserved customers on a given work day. To this end, we propose an extension of the classic savings algorithm (Clarke & Wright, 1964). We chose to extend the savings algorithm since its effectiveness and the simple manipulations of the data set involved in its application allow large problems to be satisfactorily solved, making it the basis
of most commercial VRP systems (Laporte & Semet, 2002). The extension is applied to sets of examples based on data provided by classic instances of Solomon (1987), and the results are compared to those provided by the original savings algorithm according to a set of performance criteria. In order to assess possible gains from more sophisticated methods we also provide the results obtained with an adaptation of a tabu search approach from the literature.

The remainder of this paper is organized as follows. Section 2 describes the problem of vehicle routing with multiple deliverymen tackled in this research, and Section 3 presents the associated formulation. Sections 4 and 5 respectively discuss the savings heuristic and the tabu search approach. Section 6 describes the computational experiments and results for the sets of generated instances, followed by conclusions and perspectives for future research in Section 7.

2 THE VEHICLE ROUTING PROBLEM WITH MULTIPLE DELIVERYMEN

The capacitated time-constrained VRPMD can be defined as follows: given a set of nodes, each of which requiring a given demand of goods to be delivered, a homogeneous vehicle fleet, and a set of deliverymen, define a set of routes to be carried out by the vehicles in order to serve the nodes demand. The routes should minimize operation costs and satisfy the following restrictions: (a) each route must start and finish at a central depot; (b) all visited nodes must be served by a single vehicle; (c) the demand of each served node must be completely met; (d) the vehicle load, at any time, must not exceed the vehicle capacity (without loss of generality, it is admitted that the demand of each node does not surpass the vehicle capacity); (e) the crew in each vehicle must not exceed the capacity of the vehicle’s cabin; and (f) the total time spent on each route (consisting of the vehicle travel times and service times at each node) should not exceed a predetermined limit associated to end of the work day. The service time at each demand node includes the deliverymen walking time and the cargo unloading time in the node.

Note that these restrictions have a reasonable degree of generality so as to allow the inclusion of additional constraints describing different variations of the problem. It may be required, for instance, that the fleet and/or set of deliverymen has (have) unlimited size. In this case, and assuming that the instance is feasible (the problem data allow any node \( i \) to be served by the direct route \( i \rightarrow \Delta \rightarrow i \)), we can impose that all demand nodes must be served. Alternatively, one can consider limitations on the size of one or both resources so that service cannot be guaranteed to all nodes. For the application dealt in this paper, we assume that the fleet size is limited to \( F \) vehicles and the number of deliverymen is large enough to allow each vehicle to have the maximum crew size. We also assume that the direct route \( \Delta \rightarrow i \rightarrow \Delta \) does not require more than one deliveryman to ensure that the maximum route time constraint is satisfied.

The costs to be minimized in the VRPMD decisions also depend on the application. If no limitations are imposed on the fleet size and the available number of deliverymen, the main objective is to minimize the number of vehicles (fixed cost), followed by the number of deliverymen, and the distance/total time of the operation (variable cost). For the application considered in this paper,
these objectives are preceded by the minimization of unserved nodes. Priority is given, thus, to optimizing the level of customer service.

Note that for the beverage delivery example described in Section 1, the nodes to be routed correspond to the vehicles parking sites, and the demand of each node is the sum of the cluster’s total demand. Service times at each node depend on the size of the vehicle crew, the delivery strategy (more than one deliveryman serves one customer at a time/customers are served simultaneously, each by a single deliveryman), and cluster characteristics, such as demand and customers geographic dispersion.

Despite its potential practical applications in urban areas (other examples include the tobacco and snacks distribution), as far as we know, the use of extra deliverymen was only recently formalized for the vehicle routing problem with time windows (Pureza et al., 2012), hence resulting in the variant VRPTWMD. In that paper, the authors propose two heuristic procedures: a tabu search approach and an ant colony optimization algorithm. Their procedures are tailored to address situations for which there are limitations on the number of available deliverymen (rather than on the size of the fleet), which means that all demand nodes are required to be visited.

3 FORMULATION

The VRPMD described in Section 2 can be modeled based on the two-index variable model presented in Pureza et al. (2012) for the VRPTWMD. Since fleet and crew size in each vehicle may prevent a subset of nodes to be served, the model also include some features presented in the three-index variable model of Tang & Wang (2006) for the prize-collecting vehicle routing problem (PCVRP). The network comprises \( n \) nodes indexed by \( i = 1, \ldots, n \); \( i = 1 \) represents the depot and \( i = 2, \ldots, n \) refer to the parking sites (clusters). For the sake of simplicity, it is assumed that service times in each node are provided by the user.

The following notation is used for the model’s description:

**Input Data**

- \( n \) Number of nodes or clusters \((i = 1, \ldots, n)\); the depot is represented by node \( i = 1 \), and the parking sites by nodes \( i = 2, \ldots, n \).
- \( L \) Maximum crew size (the driver plus the extra deliverymen) that can be assigned to a single vehicle \((l = 1, \ldots, L)\). If the crew size assigned to a vehicle is \( l \), we say that this vehicle travels in mode \( l \).
- \( F \) Fleet size
- \( p \) Prize of one served node
- \( c_1 \) Cost of a vehicle
- \( c_2 \) Cost of one deliveryman
- \( c_3 \) Cost of an unitary distance traveled by a vehicle
- \( Q \) Capacity of each vehicle
- \( T \) Maximum route time
Average speed of the vehicles

d_{ij}

Distance between nodes i and j (i, j = 1, ..., n, i ≠ j); we consider that d_{ij} may be different from d_{ji} due to, e.g., road restrictions and one-way directions (asymmetric)

t_{v,i,j}

Average direct travel time between nodes i and j (i, j = 1, ..., n, i ≠ j) (e.g., \( t_{v,i,j} = \frac{d_{ij}}{v} \))

st_{il}

Service time in node i = 1, ..., n with l = 1, ..., L deliverymen; it is assumed that st_{il} = 0

q_{i}

Non-negative demand (in the same units of capacity Q) of node i = 1, ..., n, with q_{i} = Q; if j = 2, ..., n, q_{i} is equal to the quantities to be delivered to the cluster associated to parking site i; it is assumed that q_{1} = 0

a

Vehicles departure time at node i = 1; it is assumed that a = 0

b

Vehicles latest arrival time at node i = 1; it is assumed that b = T

Variables

\( x_{ijl} = \begin{cases} 1, & \text{if the vehicle travels directly from node } i \text{ to node } j \text{ in mode } l \\ 0, & \text{otherwise} \end{cases} \) \( (i, j = 1, \ldots, n; i ≠ j; l = 1, \ldots, L) \)

\( w_{il} = \begin{cases} 1, & \text{if node } i \text{ is served by a vehicle in mode } l \\ 0, & \text{otherwise} \end{cases} \) \( (i = 1, \ldots, n; l = 1, \ldots, L) \)

\( t_{il} \)

Service start time of the vehicle in mode l = 1, ..., L at node i = 1, ..., n; \( t_{il} \) corresponds to the time the vehicle in mode l = 1, ..., L returns to the depot

\( y_{il} \)

Vehicle load in mode l = 1, ..., L right after serving node i = 1, ..., n.

The VRPMD is formulated according to the following mixed 0-1 linear model:

\[
\begin{align*}
\min \quad & z = \sum_{i=2}^{n} \sum_{k=1}^{K} \sum_{l=1}^{L} w_{ikl} + \sum_{i=2}^{n} \sum_{j=2}^{n} x_{ijl} + \sum_{i=1}^{n} \sum_{j=2}^{n} \sum_{l=1}^{L} l x_{1jl} + \sum_{i=1}^{n} \sum_{j=2}^{n} \sum_{l=1}^{L} d_{ij} x_{ijl} \\
\text{subject to} \quad & \sum_{j=1}^{n} \sum_{l=1}^{L} x_{ijl} \leq 1, \quad j = 2, \ldots, n \\
& \sum_{j=1}^{n} \sum_{l=1}^{L} x_{ijl} \leq 1, \quad j = 2, \ldots, n \\
& \sum_{j=1}^{n} x_{ijl} = w_{il}, \quad i = 2, \ldots, n; \quad l = 1, \ldots, L
\end{align*}
\]
\[
\sum_{j=1 \atop j \neq i}^{n} x_{jl} = \sum_{j=2}^{n} x_{jl}, \quad i = 1, \ldots, n; \quad l = 1, \ldots, L
\]  

(5)

\[
t_{jl} \geq t_{ij} + (s_{il} + t_{ij})x_{jl} - M_{ij}(1 - x_{ij}), \quad i = 2, \ldots, n; \quad j = 1, \ldots, n; \quad i \neq j; \quad l = 1, \ldots, L
\]  

(6)

\[
y_{jl} \geq y_{ij} + q_{j}x_{jl} - Q(1 - x_{ij}), \quad i = 1, \ldots, n; \quad j = 2, \ldots, n; \quad i \neq j; \quad l = 1, \ldots, L
\]  

(7)

\[
\sum_{j=2}^{L} \sum_{l=1}^{L} x_{jl} \leq F
\]  

(8)

\[
u_{ij}, x_{ij} \in \{0, 1\}, \quad i, j = 1, \ldots, n; \quad i \neq j; \quad l = 1, \ldots, L
\]  

(9)

\[a \leq t_{il} \leq b, \quad l = 1, \ldots, L
\]  

(10)

\[q_{i} \leq y_{jl} \leq Q; \quad i = 1, \ldots, n; \quad i \neq j; \quad l = 1, \ldots, L
\]  

(11)

where \(M_{ij}\) is a sufficiently large number. The objective function (1) minimizes the number of unserved nodes, the fixed and variable costs of the required fleet as well as the cost of assigned deliverymen. Constraints (2) ensure that at most one vehicle in one mode enters in each node \(j(j \neq 1)\), whereas constraints (3) ensure that one vehicle in one mode at most leaves each node \(l(l \neq 1)\). Constraints (4) state that the traversal of an arc \(i-j\) by a vehicle in mode \(l(l = 2, \ldots, n; j = 1, \ldots, n)\), that is, \(x_{ij} = 1\), implies that node \(i\) is served in mode \(l(u_{ij} = 1)\). If \(x_{ij} = 0(j = 1, \ldots, n; l = 1, \ldots, L)\) then node \(i\) is not served. Constraints (5) are the flow conservation equations that ensure that the same vehicle that enters in a node \(i\) in mode \(l\) should leave this node \(i\) in the same mode \(l\). Note that these constraints are also valid if node \(i\) is not visited (along with constraints (4), they imply that \(u_{ij} = x_{ij} = 0, j = 1, \ldots, n; l = 1, \ldots, L\)).

Constraints (6) define the relationship between the flow variables \(x_{ij}\) and the variables of service start time \(t_{ij}\); since increasing service start times are required along the tour, these constraints also prevent against the formation of subtours that do not contain the depot. Note that these constraints cannot be defined for \(i = 1\) (i.e., departure of the depot, given that \(s_{il} = 0\) and \(t_{il} \geq t_{ij}, j = 2, \ldots, n\), but they are defined for \(j = 1\) (return to the depot), in order to guarantee that the time window of the depot is met (given that \(a \leq t_{ij} \leq b\). The formation of subtours that do not contain the depot is also similarly ensured by constraints (7) which define the relationship between the flow variables \(x_{ij}\) and the load variables \(y_{jl}\) (note that the vehicle load increases along a tour). These constraints are defined for \(i = 1\) (i.e., departure from the depot), but need not be defined for \(j = 1\) (return to the depot), given that \(q_{1} = 0\). Constraints (8) ensure that the size of the available fleet \(F\) is not exceeded. Constraints (9) define the domain of the variables, while constraints (10) and (11) respectively ensure that the time window of the depot is satisfied, and the maximum capacity \(Q\) of each vehicle is not exceeded. Without loss
of generality, the model size can be reduced based on the input data; for instance, by fixing the variables $x_{ijl} = 0$ if $q_i + q_j > Q$ (capacity constraint violated).

The problem described by formulation (1)-(11) is NP-hard since the VRP can be seen as its particular case for which there is only one mode (i.e., $L = 1$), $p \gg c_1 \gg c_3$, $c_2 = 0$, and $F$ is sufficiently large. As pointed out in Bráysy and Gendreau (2005a, 2005b), the VRP is NP-hard.

4 A CONSTRUCTIVE HEURISTIC APPROACH

In the following we show how the savings algorithm proposed by Clarke & Wright (1964) can be extended for solving the VRPMD described in the previous sections. The original savings algorithm is a simple and popular constructive heuristic that applies to problems for which the number of vehicles is a decision variable (unlimited fleet). It computes the savings that arise from merging two existent routes, and iteratively selects the option with the largest savings that do not violate capacity and temporal restrictions. Note that in addition to the reduction of the total distance, each merge decrements the current number of vehicles in one unity.

Figure 1 briefly describes the sequential version of the savings algorithm (henceforth called SAV). Some of the model’s notation applies in this description.

1. Let $S$ be a solution comprising $r = n$ routes, each of which serving a demand node (with one deliveryman).
2. (Savings computation) Compute the savings in distance resulting from serving the pair of nodes $i$ and $j$ (i.e., $i \neq j$) in the same route (that is, $s_{ij} = d_{i1} + d_{1j} - d_{ij}$), and store the resulting values (ordered in a non-increasing fashion) in a list $L$.
3. While $L \neq \emptyset$:
   
   3.1. Starting from the first element of $L$, select the pair of nodes $i$ and $j$ that satisfies one of the following conditions:
      
      a) Neither $i$’s nor $j$’s routes have already been merged. In this case, merging would include both $i$ and $j$ in the same route.
      b) Exactly one of the two nodes’ routes (i’s or j’s) has already been merged and the corresponding node is not interior to that route. In this case, the link $i-j$ would be added to that same route.
      c) Both i’s and j’s routes are distinct from each other, have already been merged, and neither node is interior to its route. In this case, the two routes would be merged.
      
      3.2. Obtain a new solution $S$ by implementing the selected merge, rescheduling the resulting route, and making $r = r - 1$ if the vehicle capacity and maximum route time constraints are not violated.
      
      3.3. Eliminate the element associated to the selected merge from $L$.

4. Return solution $S$.

Figure 1 – Sequential savings algorithm (SAV).
Note that if the fleet size is limited to $F$ vehicles, the $r - F$ routes resulting from the application of algorithm SA V actually comprise unrouted nodes. On the other hand, if multiple deliverymen could be assigned to one or more routes, service times would be reduced, favoring the increase of the number of feasible merges. As a result, the number of required vehicles is reduced, and the number of served nodes increases.

The main feature of algorithm SA V with multiple deliverymen (henceforth called algorithm SA V md) consists in the generation of candidate solutions in which the number of deliverymen in the current solution $S$ is iteratively incremented until the maximum crew size in each vehicle is tested or reached, or all nodes are routed in the solution. For a given iteration, the candidate solution that produces the largest number of served nodes is selected to replace $S$. The description of SA V md is presented in Figure 2 and discussed in more detail in the following paragraphs.

1. Let $F$ be the fleet size ($F < n$), and $L$ be the maximum crew size that can be assigned to a vehicle. Apply algorithm SA V (described in Fig. 1), obtaining a solution with one deliveryman per route. Furthermore, for each route $k (k = 1, \ldots, r)$, let $server_k$ and $tserver_k$ be respectively the number of deliverymen assigned to route $k$ (in the current candidate solution $S$), and the last tested number of deliverymen in route $k$ (that is, $server_k = tserver_k = 1$).

2. While ($r > F$) and ($server_k < L$ for some $k = 1, \ldots, r$) and ($tserver_k < L$ for some $k = 1, \ldots, r$):
   2.1. Let $server_{rk} = server_k$ for $k = 1, \ldots, r$ (store the number of deliverymen of each route in the current solution $S$).
   2.2. For each route $k (k = 1, \ldots, r)$ such that ($server_k < L$) and ($tserver_k < L$):
      2.2.1. Make $server_k = server_k + 1$ (assign an extra deliveryman to route $k$).
      2.2.2. Reschedule route $k$.
      2.2.3. Apply algorithm SA V (described in Fig. 1) from step 2 on, obtaining a candidate solution $S_k$ with $r_k$ routes.
      2.2.4. Make $tserver_k = server_k$ and $server_k = server_{rk}$ (restore $k$’s original crew size).
   2.3. If at least one feasible merge was produced in step 2.2, select $S_k$ (with incremented number of deliverymen) that serves the largest number of nodes and make $S = S_k$.
   3. If $r > F$, select the $F$ routes with the largest number of nodes to comprise solution $S$.
   4. Return $S$.

Figure 2 – Sequential savings algorithm with multiple deliverymen (SA V md).

In step 1 of SA V md a starting solution $S$ with $r$ routes is produced, each of which served by a single deliveryman. At this point, if all nodes are served ($r \leq F$) the route construction ends. Otherwise, an iterative procedure that computes the impact of an additional deliveryman in the solution (steps 2.1-2.2) is employed.

In the latter case, an extra deliveryman is temporarily assigned to each route $k$, one route at a time, as long as the maximum crew size ($L$) is not exceeded in $k$ and the resulting provisional number of deliverymen ($tserver_k$) has not yet been tested in $k$. Each assignment requires the
associated route to be rescheduled (step 2.2.2), hopefully increasing its temporal slack. The application of algorithm SAV is then resumed in step 2.2.3, resulting in the candidate solution $S_t$. The candidate solution with the largest number of served nodes is selected to replace solution $S$, and the additional deliveryman is permanently assigned to the associated route (step 2.3). No improvement is obtained if all candidate solutions have the same number of routes as solution $S$ (no feasible merge is produced), hence no assignment is performed.

Steps 2.1-2.3 are repeated as long as $r > F$ and there is at least one route for which both the actual number of deliverymen and the last tested crew size is less than the maximum crew in a vehicle. If the resulting solution has $r > F$, the $F$ routes with the largest number of nodes are chosen to comprise $S$ while the nodes of the $r - F$ remaining routes are not served (step 3). It should be noted, however, that many real time applications allow some percentage of violation of the maximum route time. In these cases, algorithm SAV$_{md}$ can be modified by increasing the maximum route time and penalizing merges according to the amount of the violation.

5 A TABU SEARCH APPROACH

This research mainly aims at proposing a basic routing method for environments that require a rapid response upon processing a substantial amount of data. This is the reality of many distribution companies with the difficult task of planning routes for hundreds of trucks in order to serve thousands of clients in a daily basis. Nevertheless, when the application allows longer computational time, further improvement in the solutions provided by a constructive heuristic can be attained by more sophisticated methods such as local search or metaheuristics. In order to observe the impact of such methods in relation to SAV$_{md}$, the tabu search approach proposed in Pureza et al. (2012) for the VRPTWMD was modified to tackle the variation of the VRPMD addressed in this work and simplified to some extent. A high-level description of the adapted approach (henceforth called TS$_{vrpmd}$) is described in Figure 3.

1. Apply algorithm SAV$_{md}$ (described in Section 4), followed by the tabu search algorithm for number of routes reduction (described in Section 5.2).
2. If all nodes are routed in the best found solution $S^*$, go to step 4.
3. Repeat
   1. Apply the insertion heuristic (described in Section 5.1) to include unroute nodes in $S^*$.
   2. Apply the tabu search algorithm for number of routes reduction.
   Until there is no available vehicle or all nodes are routed in $S^*$.
4. If at least one route in $S^*$ is served by more than one deliveryman, apply the tabu search algorithm for crew size reduction.
5. Apply the tabu search algorithm for distance reduction.
6. Return $S^*$.

Figure 3 – The tabu search approach (TS$_{vrpmd}$).
In step 1 of the tabu search approach, reducing the number of routes of the starting solution addresses not only this objective itself, but also the minimization of the number of unrouted nodes and the total crew size. The deliverymen assigned to the eliminated routes become unnecessary, and in case of unrouted nodes (step 3) an insertion-type heuristic (step 3.1) can use the released vehicles to serve them. In order to strengthen the approach, the insertion heuristic and the number of routes reduction algorithm (step 3.2) are alternated until all vehicles are utilized or all nodes are routed. The application of the tabu search algorithm for crew size reduction (step 4) and distance reduction (step 5) provide additional improvement.

As one can see, the structure of the tabu search algorithm is general enough to be used either for reducing the number of routes, the number of deliverymen or the total distance. It should be noted that in Pureza et al. (2012), the fleet size can be incremented so as to guarantee that all nodes are routed.

The following two subsections describe the insertion heuristic and the tabu search algorithm employed in TS\_vrpmd.

5.1 The insertion heuristic

We use a variant of the cheapest insertion heuristic (Lawler et al., 1985) to include unrouted nodes in the solution. The heuristic is a simple but effective method that sequentially chooses the unrouted node and its insertion position in a current route that minimize the resulting extra distance. In case of no feasible insertion positions for all nodes but one or more available vehicles, a new route is created with the farthest node from the depot. A more formal description of the insertion heuristic is presented in Figure 4. Note that in Pureza et al. (2012), unrouted nodes are included in the solution by a procedure based on Solomon (1987)’s I1 insertion heuristic.

1. Let U be the set of unserved nodes in the best found solution S* comprising r routes.
2. Repeat
   2.1. Repeat
      2.1.1. For each node i ∈ U obtain the feasible insertion position p in a current route v(p = 1, ..., r) that minimizes the resulting extra distance.
      2.1.2. If there is at least one feasible insertion position for some node in the current routes, select the node i* and its insertion position p* in route v* with the smaller extra distance. Insert i* accordingly and reschedule v*. Make U = U – {i*}.

   Until there is no feasible insertion position in the current routes.

2.2. If U ≠ ∅ and r < P, let i* be the farthest node from the depot (i* ∈ U). Start a new route with i*, and make U = U – {i*} and r = r + 1.

   Until there are no available vehicles (r = P) or all nodes are routed (U = ∅).

3. Return S*.

Figure 4 – The insertion heuristic.
5.2 The tabu search algorithm

As in the work of Pureza et al. (2012) we use an integrated intensification/diversification mechanism that modifies selected tabu search parameters (Glover & Laguna, 1997) based on the analysis of search trajectory patterns (the curve of solution cost vs. iteration). The core of the mechanism is the assumption that trajectory patterns reflect the restrictiveness level imposed by the tabu framework design. Under this assumption, setting parameter values that strategically respond to the observed patterns can be a means for improving search effectiveness.

In order to identify the current trajectory pattern, the search process is divided into stages, and for each two consecutive stages \( \theta - 1 \) and \( \theta \), the current average solution cost in stage \( g \) is compared to the previous average solution cost in stage \( \theta - 1 \). The search describes a stagnated trajectory, an ascent trajectory or a descent trajectory depending on whether the two averages are approximately the same, the current average is larger than the previous average, or the current average is smaller than the previous average, respectively. Once the trajectory pattern is identified, the approach reacts to it by imposing a new operational setting for some specified number of iterations. In Pureza et al. (2012), the number of nodes of the smaller route is used as the solution cost, and parameter changes are applied for \( \sigma \ast \text{hoz} \) iterations, where \( \text{hoz} \) is an integer randomly generated within a pre-specified range and \( \sigma \) is a tuning factor whose value depends on the prescribed change. Prior to the trajectory evaluation, it is verified if the last stage corresponds to an improvement phase; in this case, the tabu parameters and the tuning factor are also set accordingly. Note that the period of application of the new operational setting corresponds to a new stage \( (\theta + 1) \).

A high-level description of the algorithm is presented in Figure 5. For further details, the reader should refer to Pureza et al. (2012).

1. From the best found solution \( S^* \), proceed with the search for two stages using standard parameter values.
2. Repeat until no feasible moves are available or some stopping criterion is met:
   2.1. If any solution improvement is verified in the last stage, update \( S^* \). Otherwise, identify the current pattern described by the last two stages.
   2.2. Set parameters according to the trajectory pattern or to an improvement phase.
   Apply the setting for the prescribed period of application, obtaining a new stage.
3. Return \( S^* \).

**Figure 5** – The tabu search algorithm.

In our adapted approach the tabu search algorithm is first used to reduce the required fleet size (steps 1 and step 3.2 in Fig. 3), then to reduce the crew sizes on the vehicles (step 4 in Fig. 3), and finally to decrease the total traveled distance (step 5 in Fig. 3). As in the previous work, four inter-route move operators are used in these steps to generate the neighborhood structure: relocation of a single node to a different route (single insertion), relocation of two nodes in the same route to...
a different route (double insertion), exchange of two nodes in different routes (exchange), and an exchange type move followed by an insertion type move (exchange/insertion). These operators are instantiations of the $\lambda$-interchange operators proposed by Osman (1993) with $\lambda = 2$. Only moves that maintain the solution feasibility are allowed.

Added (removed) arcs in recent moves are classified as tabu-active and restrictiveness levels are controlled by the tabu activation rule. The latter prescribes the tolerable number of tabu-active arcs for each move type in a given search stage. Standard tolerance values are applied along the first two search stages (step 1 of Fig. 5); the first stage comprises all solutions up to the first local optimum, while the second stage ends with the solution found $h_0z$ iterations ahead. Minimum tolerance values are applied in the occurrence of a stagnated trajectory as means to induce diversification. Stagnated trajectories are identified when the absolute value of the percent deviation of the solution’s current average from the previous average ($apd$) is less than or equal to a given threshold $ST$. In the occurrence of ascent trajectories, the prescribed parameter values depend on the range of $apd$ values. The general idea is that the larger $apd$ is, the more ascending is the trajectory, and more vigorous should be the relaxation imposed. Finally, if an improvement phase or a descent trajectory is identified, intensification is induced by setting large tolerance values.

Since we also use the number of nodes of the smaller route as solution cost, and the same datasets to generate test instances (see Section 6) as in the previous work, identical tabu parameter settings are employed in our experiments. These settings are summarized in Tables 1 and 2.

Table 1 – Stagnation threshold, move type tolerances and tuning factors for some trajectory types or search phase.

<table>
<thead>
<tr>
<th>Trajectory type or search phase</th>
<th>Tolerances for move type</th>
<th>Tuning factor ($\sigma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single insertion</td>
<td>Double insertion</td>
</tr>
<tr>
<td>First two stages</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Stagnation ($ST = 2%$)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Descent</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Improvement phase</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 2 – Move type tolerances and tuning factors for ascent trajectories.

<table>
<thead>
<tr>
<th>Apd range ($%$)</th>
<th>Tolerances for move type</th>
<th>Tuning factor ($\sigma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single insertion</td>
<td>Double insertion</td>
</tr>
<tr>
<td>$[12, \infty)$</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>$[10, 12)$</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>$[8, 10)$</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>$[6, 8)$</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>$[2, 6)$</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
For the reduction of the required fleet, the lexicographic ordering used for move selection consists of the following two criteria: (maximum reduction in the number of vehicles, maximum difference of route sizes in the resulting solution). The basis of the second criterion is that the relocation of nodes from small routes to larger routes may eventually empty and eliminate some of the smaller routes. Note that in Pureza et al. (2012), two additional criteria are used to include node urgency considerations due to the existence of time windows.

For the reduction of the crew size (step 4 in Fig. 3), the four types of moves have an additional step in which the number of deliverymen of each affected route is iteratively decremented until a further reduction results in an infeasible route. The lexicographic ordering described above thus includes the resulting crew sizes as a third criterion. For distance reduction (step 5 in Fig. 3), moves are selected according to the lexicographic order: (number of vehicles, number of deliverymen, total distance).

Each individual application of the algorithm of reduction of the number of vehicles, crew or distance is performed for 15,000 iterations in case there are feasible moves. After each application, the best solution found so far is further subjected to a simple local search provided by intra-route 2-opt moves (Croes, 1958).

Differently from the work of Pureza et al. (2012), the period in which changes in parameter setting are applied (hocz) and the tabu tenure are not randomly drawn from some specified ranges, but set to \( \left\lceil \frac{3}{2} \sqrt{n} \right\rceil \) and \( \left\lceil \sqrt{n} \right\rceil \), respectively. These correspond to the mean values of the ranges used in the former work, making TS_vrpmd a deterministic approach.

6 COMPUTATIONAL EXPERIMENTS

Algorithms SAV, SAV_md and TS_vrpmd were implemented in Borland Delphi 7, and the experiments were performed in a microcomputer Intel Core2 2.4 GHz with 2 GB RAM. The purpose of our tests is to assess the relative gains of SAV_md to SAV and of TS_vrpmd to SAV_md in terms of number of served nodes, and the impact of multiple deliverymen solutions on the number of vehicles and total distance. For this end, 108 examples with 100 demand nodes were generated from benchmark instances of sets C1, C2, R1, R2, RC1 and RC2, proposed in Solomon (1987) for the VRP with time windows. Since the time windows are the only factors that differentiate instances in the same set, only one instance from each set can be used as seed. For all generated examples, the original \( x \) and \( y \) coordinates of each node in the seed instance are maintained and the maximum crew size in each vehicle is set to 3.

Six scenarios prescribe how service times, demand, fleet size and vehicles’ capacity are modified in order to generate the examples. In scenario 1, the original data of the seed instance is maintained, with the exception of the fleet size which is set to 18 vehicles for all cases. This value corresponds to the largest trivial lower bound on the number of vehicles (the trivial bound of a particular instance is given by \( \sum_{i=2}^{D} q_i / Q \)) computed for all seed instances.
The following four scenarios are similar to scenario 1 except for one factor. In scenario 2 the demand in each node is doubled. In scenario 3 service times in each node \(i (i = 2, \ldots, n)\) are computed by the function

\[
st_i = \min \left\{ \phi \times q_i, T - 2 \times t_{v1} \right\}
\]

where \(q_i\), as before, is the product demand of node \(i\) and \(\phi\) is the number of products delivered per time unit in each node (in the experiments we used \(\phi = 2\)). As aforementioned, \(T\) is the maximum return time for all vehicles to the depot, and \(t_{v1}\) is the travel time between the depot and node \(i\) (i.e., \(d_1/v\); in the experiments we used \(v = 1\)). Note that \(st_i\) in (12) is proportional to the demand \(q_i\) of node \(i\) when the first term in the right-hand-side of (12) is the minimum. The second term in (12) is to ensure that this service time does not turn the instance infeasible regarding the maximum route time \(T\). In the experiments, the value of \(st_i\) in (12) corresponds to the service time in node \(i\) with one deliveryman. For the sake of simplicity, for the cases with two and three deliverymen we simply divided \(st_i\) by 2 and 3, respectively, so that \(st_i = st_i/l\) for \(l = 1, 2, 3\). As pointed out in Pureza et al. (2012), in practice service times may not be linearly dependent on the crew size.

In scenario 4 the fleet size is reduced to a third of its original value (that is, to 6 vehicles) while in scenario 5, the vehicle’s capacity is reduced by 15%. Finally, the sixth scenario combines the prescribed modification of the previous four scenarios. Table 3 summarizes the above discussion.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>(st_i)</th>
<th>(q_i)</th>
<th>(F)</th>
<th>(Q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>○</td>
</tr>
<tr>
<td>2</td>
<td>○</td>
<td>○ × 2</td>
<td>●</td>
<td>○</td>
</tr>
<tr>
<td>3</td>
<td>(12)</td>
<td>○</td>
<td>●</td>
<td>○</td>
</tr>
<tr>
<td>4</td>
<td>○</td>
<td>○</td>
<td>● × 1/3</td>
<td>○</td>
</tr>
<tr>
<td>5</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>○ × 0.85</td>
</tr>
<tr>
<td>6</td>
<td>(12)</td>
<td>○ × 2</td>
<td>● × 1/3</td>
<td>○ × 0.85</td>
</tr>
</tbody>
</table>

○ As prescribed by the seed instance. ● 18 vehicles.

The last necessary piece of data for generating our test examples consists of the maximum route time \(T\). We use 3 different values (210, 270 and 330 t.u.) for each scenario, totaling 108 (6 seed instances × 6 scenarios × 3 maximum route time values) examples. It should be noted that for all instances, each direct route with one deliveryman is feasible for \(T = 210\).

6.1 Results for \(SAV_{md}\)

Table 4 presents the average results obtained by \(SAV_{md}\) for the six instances in each scenario and maximum route time \(T\). Columns SERV, VEH, DLVM and DIST and TIME show, respectively, the number of served nodes, the required number of vehicles, the required number of deliverymen, the average total distance traveled by the vehicles, and the average route time expressed as...
the percent deviation from SAV corresponding results. Furthermore, columns SERV and VEH depict, in parenthesis, the percentage of served nodes and used vehicles obtained by the proposed algorithms regarding the total number of nodes and fleet size (small percentages of served nodes and large percentages of used vehicles characterize difficult instances). Runtimes are not presented since they are very short, typically below 1 second. As expected, SAV_md produce better solutions than SAV for all examples, considering the lexicographic order (number of served nodes, number of vehicles, number of deliverymen, total distance) employed in model (1-11).

<table>
<thead>
<tr>
<th>Scenario</th>
<th>T (t.u.)</th>
<th>SERV (%)</th>
<th>VEH (%)</th>
<th>DLVM (%)</th>
<th>DIST (%)</th>
<th>TIME (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>210</td>
<td>62.1 (87.2)</td>
<td>0.0 (72.2)</td>
<td>57.4</td>
<td>21.5</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td>270</td>
<td>44.0 (94.5)</td>
<td>-2.1 (63.0)</td>
<td>49.1</td>
<td>18.7</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>330</td>
<td>28.8 (98.3)</td>
<td>0.0 (60.2)</td>
<td>40.8</td>
<td>29.8</td>
<td>0.6</td>
</tr>
<tr>
<td>2</td>
<td>210</td>
<td>62.1 (87.2)</td>
<td>0.0 (83.3)</td>
<td>57.4</td>
<td>21.5</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td>270</td>
<td>44.0 (94.5)</td>
<td>-2.1 (77.8)</td>
<td>48.2</td>
<td>17.9</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>330</td>
<td>26.9 (97.3)</td>
<td>0.0 (75.0)</td>
<td>38.0</td>
<td>30.6</td>
<td>0.7</td>
</tr>
<tr>
<td>3</td>
<td>210</td>
<td>8.0 (83.0)</td>
<td>0.0 (100.0)</td>
<td>14.8</td>
<td>5.9</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>270</td>
<td>4.8 (98.3)</td>
<td>-5.9 (92.6)</td>
<td>8.3</td>
<td>3.8</td>
<td>-2.7</td>
</tr>
<tr>
<td></td>
<td>330</td>
<td>0.0 (100.0)</td>
<td>-8.9 (75.9)</td>
<td>0.0</td>
<td>-3.5</td>
<td>-6.4</td>
</tr>
<tr>
<td>4</td>
<td>210</td>
<td>56.5 (50.2)</td>
<td>0.0 (100.0)</td>
<td>66.7</td>
<td>26.3</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>270</td>
<td>69.1 (65.2)</td>
<td>0.0 (100.0)</td>
<td>69.5</td>
<td>4.6</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>330</td>
<td>47.2 (73.2)</td>
<td>0.0 (100.0)</td>
<td>63.9</td>
<td>23.8</td>
<td>-1.0</td>
</tr>
<tr>
<td>5</td>
<td>210</td>
<td>62.1 (87.2)</td>
<td>-1.5 (73.2)</td>
<td>57.4</td>
<td>21.1</td>
<td>7.1</td>
</tr>
<tr>
<td></td>
<td>270</td>
<td>44.0 (94.5)</td>
<td>-2.1 (65.7)</td>
<td>49.1</td>
<td>18.7</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>330</td>
<td>28.8 (98.3)</td>
<td>0.0 (63.0)</td>
<td>40.8</td>
<td>29.8</td>
<td>0.6</td>
</tr>
<tr>
<td>6</td>
<td>210</td>
<td>69.7 (33.8)</td>
<td>0.0 (100.0)</td>
<td>133.4</td>
<td>20.5</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>270</td>
<td>41.1 (34.3)</td>
<td>0.0 (100.0)</td>
<td>88.9</td>
<td>25.9</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>330</td>
<td>29.8 (37.0)</td>
<td>0.0 (100.0)</td>
<td>58.4</td>
<td>9.7</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Scenario 1 is the test setting for which T is the main limiting factor for route construction. This may be verified by the increase in the percentage of served nodes (reaching 98.3%) along with the decrease of the number of required vehicles (from 72.2% to 60.2%) as T gets larger. The average deviation in the number of served nodes reaches its peak (62.1%) with T = 210. Even though SAV_md uses the same average number of vehicles as SAV, the total distance is considerably larger (21.5%), and the number of deliverymen represents an increase of 57.4% (1.8 deliverymen per vehicle). When T = 330, gains in served nodes go down to 28.8%. This can be explained by the fact that although SAV_md significantly improves the number of served nodes, the increase in T also allows a large improvement with the application of SAV (73.8% to 83.6% of served nodes).

When the demand of each node is doubled (scenario 2), serving all requests become just a little more difficult because most of the increase in the demand is satisfied by using vehicles that are
idle in scenario 1. The relative performance of SAV_{md} is almost identical to what is observed in scenario 1, reaching its peak with 62.1% of additional served nodes for T = 210.

In scenario 3, service times are a function of the node demand and this seems to have a different effect on the algorithm performance. The increase in the percentage of served nodes is smaller than for the other two scenarios when T = 210 (83.0% for SAV_{md}) but reaches 100% with T = 330 for both algorithms. These results along with the fact that all vehicles are used when T = 210 and a quarter of them become idle when T = 330 indicate that service times consume more of the route time available in this scenario than in the previous two. The average deviation in number of served nodes reaches its peak (8.0%) with T = 210, requiring an increase of 14.8% in the number of deliverymen. It should be noted that when T = 330, the same amount of deliverymen is used in both algorithms even though SAV_{md} requires 8.9% less vehicles. This means that the number of extra deliverymen assigned is equal to the number of drivers that became unnecessary with the reduction of the number of vehicles.

The reduction of the fleet size F in 1/3 (scenario 4) makes this parameter a strong limiting factor for route construction. The percentage of served nodes is smaller than in the previous scenarios for all T, reaching a maximum of 73.2% for SAV_{md} with T = 330. All vehicles are used in each T. The average gain in number of served nodes of SAV_{md} relative to SAV reaches its peak (69.1%) with T = 270. Note that even though only 65.2% of nodes are being served, SAV_{md} uses 69.5% more deliverymen (1.7 deliverymen per vehicle) than SAV, and these numbers remain high for T = 210 and 330.

Results for scenario 5 (reduction of the vehicle’s capacity Q in 15%) are very similar to what is found for scenario 1. The increase in the percentage of served nodes (reaching up 98.3%) along with the decrease of the number of required vehicles (from 73.2% to 63.0%) shows that T (rather than Q) is the main limiting factor. Note that the decrease in capacity is compensated by using vehicles that are idle in scenario 1.

Scenario 6 is the most restrictive of all test environments (as mentioned before, it combines all modifications of the four previous scenarios). This is reflected on the smaller percentage of served nodes and larger percentage of used vehicles for all T. As one can see, the combination of the modifications do not contribute additively to the gains of SAV_{md} relative to SAV (compare, for instance, the average deviation in number of served nodes with T = 270 with the corresponding route time in scenario 4) and to the extra costs in number of deliverymen and distance. On the other hand, the use of extra deliverymen substantially improves the service level, providing the largest gain in number of served nodes found in the experiments (69.7% when T = 210).

6.2 Results for TS_{vprd}

Table 5 shows the average results obtained by TS_{vprd} for the six instances in each scenario and maximum route time T. Columns SERV, VEH, DLVM and DIST and TIME refer to the same performance criteria as in the previous table and are given as the percent deviation from SAV_{md} corresponding average results. Average times (in seconds) required to find the best solutions are also presented (column CPU).
As expected, TS$_{vrpmd}$ improves SA$_V$md solutions for all examples considering the lexicographic order ⟨number of served nodes, number of vehicles, number of deliverymen, total distance⟩ employed in model (1-11). Largest gains in scenarios 1, 2 and 5 are mostly concentrated in traveled distance which may suggest that the other performance criteria in SA$_V$md solutions have reasonable quality in these settings.

Results for scenario 3 reinforce the observation presented in subsection 6.1 that service times as a function of the node demand consume more of the route time than service times provided by the seed instances. TS$_{vrpmd}$ considerably improves the service levels (20.5% of served nodes) when $T = 210$. It should be emphasized the role of the insertion heuristic (Fig. 4) when short maximum route times limits the effectiveness of route merging type algorithms. The route reduction algorithm is particularly beneficial when $F$ is small (scenario 4) and $T$ is short; the average gain in number of served nodes reaches 27.5%. Note however that such improvements require an increase greater than 40% in number of deliverymen for both scenarios. Finally, the largest gains in service levels obtained in scenario 6 highlight the positive impact of more sophisticated algorithms when treating difficult settings.

The analysis of the preceding sections allows us to list the studied scenarios in decreasing order of difficulty (regarding the model’s objective function) as: the combined modifications (scenario 6), fleet reduction in 1/3 (scenario 4), doubled demand (scenario 2), reduction of the vehicles’ capacity in 15% (scenario 5), standard setting (scenario 1) and service times as a function of the demand (scenario 3). The largest gains obtained by SA$_V$md relative to SAV (and by TS$_{vrpmd}$...
relative to $\text{SAV}_{\text{md}}$ as well) with these sets of instances were usually obtained for the most difficult environments, that is, when $T$ is short and the available fleet is small.

7 CONCLUSIONS AND PERSPECTIVES OF FUTURE RESEARCH

This paper addressed a variant of the vehicle routing problem, which despite representing several real distribution systems, has not yet received enough attention from the literature and commercial routing systems. The variant allows the use of multiple deliverymen in each route as means to reduce service times and facilitate service to a larger number of requests within regular working hours. For this end, we developed an extended version of the classic savings algorithm which iteratively tests and selects increments in the number of deliverymen in each route. We also presented the results with an adaptation of a tabu search approach for a related problem. Results obtained with six sets of generated 100-node examples based on Solomon (1987)'s instances has shown that the proposed extension ($\text{SAV}_{\text{md}}$) can provide relevant gains relative to the solutions with one deliveryman. Considering the sets all together, $\text{SAV}_{\text{md}}$ produced average increments in the number of served nodes of 40% with 1.2% less vehicles. In contrast, such gains often imply in larger route times and total traveled distances (18%) and a substantial increase (52%) in number of deliverymen. The application of the tabu search approach ($\text{TS}_{\text{VPRMd}}$) improves the solutions of $\text{SAV}_{\text{md}}$ in number of served nodes, required fleet, total distance and route time in 15.6%, 1.7%, 1.93% and 4.5%, respectively, with an increase of 10.7% in the required crew.

We conclude that the assignment of multiple deliverymen can provide significant benefits to companies for which customer service is a priority issue. An important development for this research thus consists in applying the heuristics to real world instances and performing a qualitative analysis of the solutions, considering what is actually adopted in practice. Solving the model with an exact method can also give more insight on the performance of $\text{SAV}_{\text{md}}$ and $\text{TS}_{\text{VPRMd}}$, even though the difficulties in deriving tight lower bounds on the objective value of vehicle routing problems will most likely imply in a slow convergence rate if partial enumeration methods are employed (Cordeau et al., 2002). An interesting perspective for future research is to modify the model and proposed methodologies in order to include features such as heterogeneous fleet, multiple depots, and mixed pickup and delivery of goods. The last feature in particular represents more realistically many distribution systems in the beverage industry (among others), since empty bottles are very often picked up in each demand point and returned to the depot upon the last visit completion. In addition, a study of how service times vary with the cluster demand and configuration, the crew size, and the delivery strategy is essential for supporting VRPMD decisions in practice. Finally, due to its simplicity and short runtimes, $\text{SAV}_{\text{md}}$ could be used as the basis for an algorithm suitable for commercial routing systems.

ACKNOWLEDGMENTS

The authors thank the two anonymous referees for their valuable comments and suggestions, and Bebidas Ipiranga for their collaboration. This study was partially supported by CAPES (DS) and CNPq (grant 303001/2009-7).
REFERENCES


