Modeling and Solving the Clustered Capacitated Vehicle Routing Problem

Christopher Expósito Izquierdo∗ Андрé Rossi† Marc Sevaux†

Abstract
The Clustered Capacitated Vehicle Routing Problem is to determine the routes with minimum length used by a fleet of delivery vehicles in order to fulfill the demand of customers organized into given clusters.

The main difficulty of this problem is that when a vehicle serves a customer belonging to a given cluster, it has to serve all the customers of that cluster before it can enter another cluster or go back to the depot. A two-level solution approach based on the decomposition of the problem is proposed in this paper. This approach allows to use different optimization techniques with the goal of tackling the arising subproblems. Real size instances are used to measure the suitability of the proposed algorithms.

Keywords: Capacitated Vehicle Routing Problem, Logistics, Customer Clustering

1 Introduction
The business of the parcel delivery and courier services companies can be seen as a well-defined chain of steps. These companies answer to the pickup requests of parcels and carry them towards their end points indicated by the customers. The parcels are picked up as soon as possible, in the same or the following day. All the collected parcels are brought to a central depot dedicated to manage the distribution process. In the depot, the parcels are sorted and packed into containers, which ease the delivery tasks (Werners and Wülfing [17]). Finally, the containers loaded with the parcels are delivered to the final customers through the available means of transportation at the company. For this purpose, most of the delivery companies have only used strategies based on their experience or their acquired practical knowledge concerning the logistics field. However, the optimization techniques stemming from the Operations Research (OR) have been shown to be a useful tool in helping to ensure a suitable performance along the delivery process stages. Several practical examples of their application can be found in the works of Wasner and Zäpfel [16] and Chang and Yen [2].

In real-world environments, the number of final customers for many parcel delivery and courier companies can be extremely large in some cases (hundreds or thousands). This fact motivates the design of strategies aimed at organizing the customers according to some particular characteristic (Cao and Glover [1]). In practice, this organization is usually determined by the relative position of the customers. For instance, the most widely extended criterion is based on the postal codes (or ZIP codes), in such a way that, all the customers with a similar postal code are included into the same cluster. The main advantage of these organizational strategies is the great reduction of the complexity when managing the logistics issues derived from the distribution tasks at the company. That is, all the customers belonging to the same cluster must be served by a single delivery vehicle. In the following and without loss of generality, it is assumed that the delivery vehicles are trucks.

The previous discussion gives rise to the definition of a new challenging variant of the well-known Vehicle Routing Problem (VRP) with clustering constraints. The VRP is one relevant problem within the OR due to its many application fields. Some of them are transportation (Park and Kim [14]) and distribution (Mingyong and Erbao [13]). The reader is referred to the book of Golden et al. [5] for an exhaustive review on the VRP.

The Clustered Capacitated Vehicle Routing Problem (CCVRP) is a generalization of the VRP where a fleet composed of a unlimited number of identical delivery trucks has to fulfill the service demand of a set of n customers organized in clusters. All the delivery trucks are initially located at a central depot (node 0). Each customer \( i \in \{1, ..., n\} \) has a positive demand, \( d_i \). In addition, we are also given a partition of \( m \) clusters, in such a way that, for each customer \( i \in \{1, ..., n\} \), its cluster is denoted by \( r_i \in \{1, ..., m\} \). The set of customers belonging to a given cluster \( r \in \{1, ..., m\} \) is denoted by \( C_r \). Also, \( c_{ij} \) denotes the euclidean distance between the nodes \( i \) and \( j \), \( \forall i, j \in \{0, ..., n\} \). It is worth mentioning that all the distances satisfy the triangle inequality. Furthermore, the maximum load capacity of the trucks is denoted by \( k \). It is assumed that the total demand of each cluster does not exceed the maximum load capacity of the avail-
Figure 1: Example of solution for the CCVRP. It is composed of three routes aimed at serving a set of customers organized in 8 clusters.

able trucks, that is, $\sum_{i \in C_k} d_i \leq k, \forall r \in \{1, ..., m\}$. This means that a single delivery truck can fulfill the demand of several clusters whenever its load capacity allows it.

The goal of the CCVRP is to define the set of routes with minimum travel distance used by the fleet of delivery trucks in order to fulfill the demand of the final customers. The feasible solutions for the CCVRP must satisfy the clustering constraints: when a truck visits a customer in a cluster, it has to serve all the customers of that cluster before it can enter another cluster or return to the depot.

The CCVRP can be seen as an extension of the classic VRP where the points to visit are actually composed of a set of final customers. Therefore, its complexity is NP-hard by reduction to the classic VRP (Laporte [9]).

An illustrative example of solution for the CCVRP is depicted in the Figure 1. In this case, there is a set of customers (represented as small squares) organized into $m = 8$ clusters. The customers are served by means of three routes starting and ending at the central depot (represented as a square). The first route serves the clusters 1 and 2, the second one serves the clusters 3, 4 and 5 and the third one serves the remaining clusters, that is, the clusters 6, 7 and 8.

In general terms, several logistics issues should be properly addressed in order to tackle the CCVRP from a practical point of view:

- Determining the clusters with final customers to be served. This problem is usually known by parcel delivery and courier services companies as Districting Problem.

- Assigning the existing clusters to one delivery truck of the company.

- Determining the sequence of clusters visited by each delivery truck.

- Defining the visiting order of the customers belonging to each cluster.

This work pursues to be a first approach to a noteworthy problem addressed by parcel delivery companies worldwide. The main contributions of the work are the introduction of the CCVRP. In addition, an optimization approach is developed in order to provide a suitable solution for the problem. Its performance is tested by means of a set of computational experiments.

The remainder of this paper is structured as follows. Firstly, Section 2 describes an approximate two-level solution approach aimed at solving the CCVRP. Afterwards, Section 3 presents a set of computational experiments dedicated to assess the performance of the proposed solution approach. Finally, Section 4 discusses the main concluding remarks extracted from the work and adds suggestions for further research.

2 Solution Approach

This section is devoted to providing an approximate solution approach for solving the CCVRP. However, a brief discussion concerning the aforementioned logistics issues to consider when solving this problem is firstly carried out.

In most of the cases, the service demand for a given parcel delivery company does not suffer great fluctuations over time. In this regard, the number of customers remains constant over the same regions or is only subject to slight changes. This means that the selection of the clusters aimed at organizing the customers is a long-term problem and, therefore, it should be addressed only after an outstanding change in the demand distribution has happened. Several approaches have been published within the field of data clustering with the goal of defining suitable clusters for a given set of points. See the works of Xu and Xu [18] and He et al. [7]. Therefore, the Districting Problem is out of the scope of the present work.

In order to address the CCVRP, this work proposes an approximate two-level solution approach. At the high-level, the number and composition of the routes dedicated to serve the final customers are defined. Furthermore, the visiting order of the customers belonging to each cluster is defined at the low-level. It is easy to check that solving both subproblems allows to obtain a complete feasible solution for the CCVRP.

Several integration strategies have been proposed within the structured modeling area (Geoffrion [4]). In this work, the subproblems of the CCVRP are related according to a functional integration, in such a way that, they are sequentially solved while they exchange data. At a first step, the high-level is
solved and the obtained solution is provided in order to solve the low-level problems associated with the clusters. No feedback loop is considered in this case.

2.1 High-Level Routing Problem

As briefly stated in the introduction of this work, the CCVRP can be understood as the classic VRP where the points to visit are defined as clusters composed of customers. In this regard, the High-Level Routing Problem (HRP) is defined as the problem of finding the number and composition of the routes starting and ending at the central depot used by the fleet of delivery trucks in order to fulfill the service demand of the clusters. The pursued objective function for this problem is to minimize the total travel distance required by the trucks when visiting the clusters.

Using the existing optimization techniques aimed at solving the classic VRP allows to obtain feasible solutions for the HRP. However, according to the definition of the VRP, it is required to represent the available clusters as points. In this work, we have decided to define a virtual center associated with each cluster. The goal of these virtual centers is to represent the positional characteristics of the clusters at hand. These virtual centers and the proper estimation of the amount of time required to serve the corresponding clusters allow to use the optimization techniques for the VRP. The coordinates of the virtual centers are used in order to compute the distances between the points to visit in the VRP.

The first step to be taken before solving the HRP is to define the virtual center for each cluster. In this case, the physical concept of center of mass or barycenter is adapted for this purpose. The center of mass for a set of objects defines an individual point found at that position where the weighted relative position of the distributed mass is equal to zero. In the context of the CCVRP, the virtual center \( (x_r, y_r) \) for each cluster \( r \in \{1, ..., n\} \) can be defined as follows:

\[
x_r = \frac{\sum_{i \in C_r} d_{ir} \cdot x_i}{\sum_{i \in C_r} d_i}
\]

\[
y_r = \frac{\sum_{i \in C_r} d_{ir} \cdot y_i}{\sum_{i \in C_r} d_i}
\]

A solution for the HRP derived from the instance of the CCVRP described in the introduction of the present work is shown in Figure 2. As can be seen, the virtual center of each cluster is depicted as a circle.

Multitude of optimization techniques have been proposed in the related literature with the goal of solving the VRP. With the goal of obtaining a high-quality solution for the HRP, we have used in this work the original implementation (provided by the authors) of the Record-to-Record algorithm (RTR) developed in the papers of Li et al. [10, 12]. This algorithm is a deterministic version of the well-known Simulated Annealing and has demonstrated to be highly competitive when solving several variants of the classic VRP. For instance, the work of Li et al. [11] presents the application of this algorithm over the Open VRP, where the available vehicles do not have to come back to the depot after visiting the last point assigned to them.

2.2 Low-Level Routing Problem

Traditionally, the visiting order of the final customers belonging to a given zone is defined by the driver of the delivery truck assigned to it. In this way, once the regions to be served by each delivery truck are known, confidence is put on the corresponding driver with the goal of performing a milk-run over the final customers belonging to those regions. This is derived from the fact that the driver is usually familiar with the assigned regions or has some kind of knowledge about the customers to serve. Unfortunately, this practice usually leads to non-robust or suboptimal performances at the intra-cluster level. In addition, it establishes a great dependence on the truck drivers.

The Low-Level Routing Problem (LRP) is defined as the problem of finding a visiting order for the final customers belonging to a given cluster, in such a way that, the total travel distance is minimized. Each cluster has to be visited once and, therefore, this scenario can be interpreted as the problem of finding the hamiltonian path with minimum distance within the cluster at hand. This problem appears in the graph theory modeled by means of the so-called Shortest Hamiltonian Path.
Problem (SHPP), whenever the pursued objective function is to minimize the travel distance.

In the context of the CCVRP, there is one SHPP associated with each cluster of the instance at hand. In this case, the SHPPs are solved sequentially according to the visiting order defined by the routes, determined by means of the solution for the HRP previously computed. Due to the fact that the routes are undirected by definition, the first SHPP to be tackled for each route is the one related to the cluster with the lowest identifier. For instance, in the previous example shown in the Figure 2, the SHPPs derived from the clusters 3, 4 and 5 are solved sequentially.

The visiting order of the clusters set by the solution of the HRP determines a dependence between them. In this regard, it is interesting to find that hamiltonian path for each cluster that allows to be at the shortest possible distance to the first customer served in the next cluster of that route. For this purpose, the SHPP between two given nodes is tackled for each cluster. That is, the set of nodes involved in each SHPP are those representing the final customers belonging to the corresponding cluster, whereas the last visited node from the previous cluster (or the depot if it is the first SHPP to be solved) is set as the starting node of the hamiltonian path to find. Similarly, the virtual center of the next cluster (or the depot if it is the last SHPP to be solved) is set as the final node in the hamiltonian path.

In order to provide an illustrative example concerning the above discussion, Figure 3 shows the process performed when finding the shortest hamiltonian path associated with cluster 3 from the previous example depicted in Figure 2. For illustrative purposes we can assume that an optimization technique to solve the SHPP is available. In this case, the depot and the virtual center of the cluster 4 are set as the starting and ending nodes, respectively. Similarly, the SHPP associated with cluster 4 can be addressed, setting the last visited customer in cluster 3 and the virtual center of cluster 5 as the starting and the ending nodes, respectively. The strategy is also carried out with the goal of finding the shortest hamiltonian path within cluster 5. In this last scenario, the virtual center of cluster 4 and the depot are set as the starting and the ending nodes of the hamiltonian path, respectively.

The last issue required to provide a complete solution for the LRP is to develop some method aimed at solving the SHPP. Several optimization techniques have been implemented in this work with the goal of solving the SHPP between two known nodes:

- **MILP formulation.** The SHPP is modeled by means of a MILP based on a classic formulation for the Travelling Salesman Problem (TSP) and uses the classic edge formulation. This formulation is proposed in the work of Sevaux and Sørensen [15].

  - **Christofides algorithm.** This algorithm is developed in the work of Christofides [3]. It is an exact approach based on a decision-tree search with a set lower bounds dedicated to speed up the searching process.

  - **Lin-Kernighan heuristic.** The TSP is closely related to the SHPP between two known nodes due to the fact that the latter can be seen as the problem of finding a tour visiting all the points in a complete graph setting as fixed the existing edge between the starting and ending nodes of the hamiltonian path. The Lin-Kernighan heuristic is one of the most competitive solution algorithms aimed at solving the TSP. In this work, we have used the source code available on the website\(^1\) for free research use.

3 Computational Experiments

This section is aimed at assessing the performance of the solution approach introduced in Section 2 of the present work. The algorithm has been implemented using the C++ language and the compiler g++ 4.7.2. All the computational experiments have been carried out on a PC equipped with Ubuntu 12.10, an Intel Core 2 Duo processor at 3.16 GHz and 4 GB of RAM. Additionally, the MILP described in the Section 2 was implemented using CPLEX 12.4.

The benchmark suite we have used in order to perform the computational experiments was pro-

\(^1\)http://www.akira.ruc.dk/~held/research/LKH/
posed in the work of Golden et al. [6]. It is composed of 20 instances with information about the demand of the customers and the load capacity of the delivery trucks, however, they do not have any information about clusters. For this purpose, we have developed a simple strategy aimed at generating clusters for the customers. The first cluster is built by selecting one customer at random and, at each step, the nearest customer is added to it. Another cluster is built whenever it is not possible to add any other customer without exceeding the maximum allowed demand of the clusters, denoted by $\alpha$ and where $\alpha = \rho \ast k$. The value of the parameter $\rho \in (0..1]$ is set by the user. All the instances used in this paper are available in [8].

The computational results obtained by means of the solution approach when solving the modified instances belonging to the benchmark suite are shown in Table 1. In this case, we set $\rho = 0.1$, in such a way that, each delivery truck can fulfill the demand of at most 10 clusters. The first three columns show the index, the number of points (customers and depot) and the load capacity of the trucks in the instance to solve. The next columns show the objective function value and the computational time (measured in seconds) required by the solution approach when the Christofides algorithm, the MILP and the Lin-Kernighan heuristic are used when solving the LRP, respectively.

The best-known solutions are obtained by means of the different exact approaches used for solving the LRP, that is, the algorithm of Christofides and the MILP. Only slight gaps are presented when the CCVRP is solved using the Lin-Kernighan heuristic at the low-level (see the results for the instance 3). On the other hand, the computational times required by different algorithms present outstanding differences. Firstly, the MILP and the Lin-Kernighan heuristic have similar performances, but the algorithm of Christofides requires large computational times (more than 1000 seconds) when solving several instances (instances 5 and 11). This fact derives from the large number of nodes implicitly generated by the algorithm.

We have carried out additional experiments for which the value of the parameter $\rho$ is changed from 0.1 up to 1.0. Despite the fact that the comparisons are not provided in this paper (they will be presented at the conference), we can conclude that when the number of customers within a cluster is small (8 or 10), using the algorithm proposed by Christofides is inefficient as it requires more than one day in some cases. A similar behavior is presented by the MILP. The reason can be found on the large number of constraints to generate in order to remove the subtours. However, it is possible to set a reasonable maximum computational time (for example, 60 seconds) in this case and obtain high-quality solutions for the CCVRP. Finally, it is worth highlighting that using the Lin-Kernighan heuristic for solving the LRP allows to obtain the best-known solutions for most of the instances within a few seconds.

4 Conclusions and Further Research

This work introduces the CCVRP, a new logistic problem for parcel delivery and courier services companies where the demand of a large number of customers organized in clusters have to be fulfilled. This problem presents the clustering constraints, in such a way that, the delivery trucks have to serve all the customers belonging to the same cluster in a row.

An approximate two-level solution approach is proposed with the goal of solving the CCVRP. It is based on a decomposition of the CCVRP into two general subproblems. The first one pursues to define the number and composition of the routes aimed at serving the clusters and the latter is aimed at determining the visiting order of the customers within each cluster. This approach allows to use specific optimization techniques for both subproblems. For this purpose, several methods have been proposed.

The computational experiments have allowed to check that using the adaptation of the Lin-Kernighan heuristic for the LRP is highly competitive in a wide range of scenarios. Similarly, exact methods require large computational times in order to obtain high-quality solutions for the CCVRP and, therefore, they can be dismissed in real environments.

Acknowledgments

This work has been partially funded by the University of South-Brittany – Programme de mobilité entrante. Christopher Expósito Izquierdo thanks Ca-
jaCanarias the financial support he receives through his post-graduate grant.

References

<table>
<thead>
<tr>
<th>Instance</th>
<th>$n+1$</th>
<th>$k$</th>
<th>Christofides $f(x)$</th>
<th>$t$ (s)</th>
<th>MILP $f(x)$</th>
<th>$t$ (s)</th>
<th>LK Heuristic $f(x)$</th>
<th>$t$ (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>240</td>
<td>550</td>
<td>5862.5892</td>
<td>1.71</td>
<td>5862.5892</td>
<td>2.84</td>
<td>5862.5892</td>
<td>2.52</td>
</tr>
<tr>
<td>2</td>
<td>320</td>
<td>700</td>
<td>10058.3980</td>
<td>2.30</td>
<td>10058.3980</td>
<td>3.82</td>
<td>10058.3980</td>
<td>3.51</td>
</tr>
<tr>
<td>3</td>
<td>400</td>
<td>900</td>
<td>13778.6046</td>
<td>4.86</td>
<td>13778.6046</td>
<td>4.47</td>
<td>13778.6046</td>
<td>4.24</td>
</tr>
<tr>
<td>4</td>
<td>480</td>
<td>1000</td>
<td>19765.6887</td>
<td>105.93</td>
<td>19765.6887</td>
<td>3.66</td>
<td>19765.6894</td>
<td>3.36</td>
</tr>
<tr>
<td>5</td>
<td>500</td>
<td>900</td>
<td>10137.8822</td>
<td>1393.47</td>
<td>10137.8822</td>
<td>0.82</td>
<td>10146.4773</td>
<td>0.80</td>
</tr>
<tr>
<td>6</td>
<td>280</td>
<td>900</td>
<td>12051.4151</td>
<td>1.44</td>
<td>12051.4151</td>
<td>2.30</td>
<td>12051.4151</td>
<td>1.79</td>
</tr>
<tr>
<td>7</td>
<td>440</td>
<td>900</td>
<td>14131.8829</td>
<td>3.79</td>
<td>14131.8829</td>
<td>2.35</td>
<td>14131.8833</td>
<td>4.30</td>
</tr>
<tr>
<td>8</td>
<td>500</td>
<td>1000</td>
<td>721.7685</td>
<td>1.00</td>
<td>721.7685</td>
<td>1.49</td>
<td>721.7685</td>
<td>1.41</td>
</tr>
<tr>
<td>9</td>
<td>322</td>
<td>1000</td>
<td>925.1231</td>
<td>2.23</td>
<td>925.1231</td>
<td>3.51</td>
<td>925.1231</td>
<td>1.80</td>
</tr>
<tr>
<td>10</td>
<td>399</td>
<td>1000</td>
<td>1164.5885</td>
<td>1963.00</td>
<td>1164.5885</td>
<td>2.56</td>
<td>1164.5885</td>
<td>1.53</td>
</tr>
<tr>
<td>11</td>
<td>483</td>
<td>1000</td>
<td>1395.9766</td>
<td>1.66</td>
<td>1395.9766</td>
<td>10.90</td>
<td>1395.9766</td>
<td>2.07</td>
</tr>
<tr>
<td>12</td>
<td>252</td>
<td>1000</td>
<td>1068.4437</td>
<td>4.48</td>
<td>1068.4437</td>
<td>4.23</td>
<td>1068.4437</td>
<td>3.89</td>
</tr>
<tr>
<td>13</td>
<td>320</td>
<td>1000</td>
<td>1535.0922</td>
<td>3.11</td>
<td>1535.0922</td>
<td>5.38</td>
<td>1535.0922</td>
<td>4.66</td>
</tr>
<tr>
<td>14</td>
<td>396</td>
<td>1000</td>
<td>1754.2680</td>
<td>7.57</td>
<td>1754.2680</td>
<td>5.39</td>
<td>1754.2680</td>
<td>5.03</td>
</tr>
<tr>
<td>15</td>
<td>480</td>
<td>1000</td>
<td>2120.7177</td>
<td>12.55</td>
<td>2120.7177</td>
<td>6.56</td>
<td>2120.7177</td>
<td>5.98</td>
</tr>
<tr>
<td>16</td>
<td>240</td>
<td>200</td>
<td>890.8079</td>
<td>2.70</td>
<td>890.8079</td>
<td>4.81</td>
<td>890.8079</td>
<td>4.16</td>
</tr>
<tr>
<td>17</td>
<td>300</td>
<td>200</td>
<td>1111.2364</td>
<td>4.63</td>
<td>1111.2364</td>
<td>7.13</td>
<td>1111.2367</td>
<td>6.91</td>
</tr>
<tr>
<td>18</td>
<td>360</td>
<td>200</td>
<td>1522.8257</td>
<td>5.04</td>
<td>1522.8257</td>
<td>7.51</td>
<td>1522.8257</td>
<td>8.55</td>
</tr>
</tbody>
</table>

Table 1: Computational results of the two-level approach using the RTR algorithm for solving the HRP and the several proposed algorithms for solving the LRP over the modified instances of [6] with $\rho = 0.1$


