Optimising Parcel Deliveries in London Using Dual-Mode Routing

Nguyen, T.B.T., Bektas, T., Cherrett, T., McLeod, F., Allen, J., Bates, O., Piotrowska, M., Piecyk, M., Friday, A. and Wise, S.

Optimising Parcel Deliveries in London Using Dual-Mode Routing

Thu Ba T. Nguyễn[1][*], Tolga Bektaş[1], Tom J. Cherrett[2], Fraser N. McLeod[2], Julian Allen[3]
Oliver Bates[4], Marzena Piotrowska[3], Maja Piecyk[3], Adrian Friday[4], Sarah Wise[5]

[1]*Southampton Business School, Centre for Operational Research, Management Science and Information Systems (CORMSIS), University of Southampton, UK*
[2]*Faculty of Engineering and Environment, University of Southampton, UK*
[3]*Faculty of Architecture and the Built Environment, University of Westminster, UK*
[4]*School of Computing and Communications, Lancaster University, UK*
[5]*Faculty of the Built Environment, University College London, UK*

Abstract

Last-mile delivery operations are complex, and the conventional way of using a single mode of delivery (e.g. driving) is not necessarily an efficient strategy. This paper describes a two-level parcel distribution model that combines walking and driving for a single driver. The model aims to minimise the total travelling time by scheduling a vehicle's routing and the driver's walking sequence when making deliveries, taking decisions on parking locations into consideration. The model is a variant of the Clustered Travelling Salesman Problem with Time Windows, in which the sequence of visits within each cluster is required to form a closed tour. When applied to a case study of an actual vehicle round from a parcel carrier operating in London, savings of over 20% in the total operation time were returned over the current situation where 104 transactions from 99 consignees were being delivered to 57 stopping points.

Keywords: last-mile delivery, clustered travelling salesman problem; time windows.

Introduction

Business-to-consumer (B2C) and Consumer-to-all-parties (C2X) parcel deliveries account for almost two-thirds of UK parcel volume with forecasts suggesting that these sectors will grow by up to 5.5% per annum (Royal Mail, 2016). With consumers demanding ever faster, more reliable and convenient delivery services, the UK market has seen an increase in the proportion of parcels sent for next-day delivery (IMRG and Metapack, 2016) with 43% of consumers reporting a negative experience with

---

[*]Corresponding author, Email addresses: `thuba.nguyen@soton.ac.uk`

home delivery (Metapack, 2015). Such demands put great pressure on parcel logistics providers who operate in a very competitive market and for low profit margins. Vehicle rounds are organised out of local depots with loads partitioned into geographical areas with round orders dictated by pre-planned premium deliveries. In dense urban areas, the operational reality is that for more than 50% of the daily round time, the vehicle can be stationary by the roadside whilst the driver is out delivering to consignees, walking up to 9 km per day on foot (Piotrowska et al. (2017); and Allen et al. (2017a)). For this reason, many multi-drop parcel carriers allocate drivers to specific geographical patches so that they develop detailed local route knowledge including parking, access restrictions, traffic congestion hot spots along with a rapport with their clients. This means that the last-mile optimisation challenge facing schedulers in this domain can be considerable with many companies leaving the route order entirely to the driver's discretion as they believe that the combination of driving and walking cannot be accurately replicated by traditional routing and scheduling software (Allen et al. (2017b)).

One of the main issues, seldom considered in current routing and scheduling or catered for within vehicle routing packages, is that parking directly outside the delivery address may not be possible which can lead to an over-estimation of driving distance as a result of being unable to model 'final-approach' walking. For dense urban environments, a dual-mode (driving and walking) routing model is warranted which would require a walking network to be defined alongside a road network with associated parking locations and waiting time limits. This is inherently difficult to construct as the detailed network knowledge is held by the driver, acquired through time and experience. This is made more complex by issues such as: (1) Dynamic collection requests received during the round which the driver has to satisfy (dynamic or on-line vehicle routing problems Toth and Vigo (2014); (2) Time constraints placed by consignees for premium 'same-day' deliveries or on drivers in terms of drivers hours regulations; (3) Traffic and road network conditions which may be known in advance or may change randomly due to incidents and congestion; (4) Access restrictions which could be temporary relating to infrastructure maintenance or vehicle specific as a result of new emission zone legislation .

While there is a rich and ever-growing literature base on vehicle routing and scheduling applications and solution methods (Toth and Vigo, 2014), there still remains a significant gap between theory (and the software packages based on this theory) and its practical application to the parcel distribution sector. This paper attempts to contribute to this gap by investigating how walking can be integrated as part of the last-mile carrier activity. The distribution model studied here arises as part of actual delivery operations of a parcel carrier operating in London for a single driver. We pose the problem as a variant of the Clustered Travelling Salesman Problem with Time Windows, where driving takes place between the clusters and walking is used within the clusters, hence the use of a *dual-mode* routing structure. There is an additional condition that the sequence of deliveries within each cluster must form a closed tour. A mixed-integer programming formulation is described that provides a schedule for the driving and walking to minimise the total operation time such that deliveries are made to customers within prescribed time windows.

The remainder of this paper is structured as follows. After a brief review of the literature in the next section, we present the problem and the formulation. The subsequent section describes the case study and the application of the model. Conclusions are provided in the last section.

Literature Review

Within the routing and scheduling literature, the studies that are most relevant to our setting are those that find routes on clustered (or generalised) networks, where the set of nodes is assumed to be partitioned. There are two main ways in which such problems have been described. The first is to use generalised routing models, including the generalised travelling salesman problem (GTSP) and the generalised vehicle routing problem (GVRP), in which only one node from each cluster is required to be visited. However, these problems ignore the intra-cluster routing decisions. The second way is to use clustered routing models, including the clustered travelling salesman problem (Clustered-TSP) and the clustered vehicle routing problem (Clustered-VRP), which require all nodes in the graph to be visited, and where, upon entering a cluster, all nodes in that cluster must be visited before moving on to another cluster. The general problem assumes that the nodes within a cluster are visited successively, but the point of entry to and exit from a cluster will not necessarily be the same node. In other words, intra-cluster routes are paths and not closed tours.

Generalised routing models

Earlier studies on generalised routing Laporte et al. (1987), Noon and Bean (1991), Fischetti et al. (1995), Laporte et al. (1996), Fischetti et al. (1997) have focused on solving the generalised travelling salesman problem (GTSP). Laporte et al. (1987) described a branch and bound algorithm for solving the GTSP where the cost of travel between any two nodes is asymmetric. Noon and Bean (1991) presented a Lagrangian relaxation algorithm to compute a lower bound on the total cost of an optimal solution and heuristically determined an upper bound by identifying and removing arcs and nodes which were not in an optimal solution. Fischetti et al. (1995) introduced a general theorem to derive classes of facet-inducing inequalities for GTSP relevant to subtour elimination and comb constraints. Laporte et al. (1996) described how a wide variety of combinatorial problems could be modelled as the GTSP. Fischetti et al. (1997) described a branch and cut algorithm for the symmetric GTSP.

The relevant studies on the GVRP include those by Ghiani and Improta (2000), Pop et al. (2009), Baldacci et al. (2010), Bektaş et al. (2011), Moccia et al. (2012). Ghiani and Improta (2000) described an efficient transformation of the GVRP into a capacitated arc routing problem, limiting the transformation to symmetric cases. Pop et al. (2009) presented an application of an ant colony algorithm to solve the GVRP. Baldacci et al. (2010) have presented a wide variety of applications of the GVRP as modelling tools for combinatorial optimisation problems. These included the travelling salesman problem (TSP) with profits and several vehicle routing problem (VRP) extensions. Bektaş

et al. (2011) presented and compared four integer linear programming formulations for the GVRP, and described branch-and-cut algorithms for two of the formulations. The only work on the GVRP with time windows that we are aware of is the one by Moccia et al. (2012), who described an incremental neighbourhood tabu search heuristic for the problem.

Clustered routing models

Algorithms for the Clustered-TSP include the Lagrangian relaxation of Jongens and Volgenant (1985) for the symmetric variant of the problem, the tabu search of Laporte et al. (1996), the $\frac{5}{3}$-approximation algorithm of Anily et al. (1999), the two-level genetic algorithm of Ding et al. (2007), the simulated annealing algorithm of Schneider et al. (2010), and the sequential constructive crossover 2-opt search and a local search mechanism of Ahmed (2014). The Clustered-VRP is a generalisation of the Clustered-TSP, for which two integer linear programming formulations have been described by Pop et al. (2012), and solved using local search and a hybrid genetic algorithm by Vidal et al. (2015). We are not aware of extensions of the clustered routing models that take into account time window restrictions.

One other relevant problem to our setting is the truck-and-trailer routing problem (TTRP), which does not assume that the input graph is clustered, but treat the cluster formation as part of the decision problem. In this problem, a truck coupled with a trailer is used to perform deliveries but some customers may not be accessible using the trailer. In such cases, the trailer can be decoupled and left at a parking place, but needs to be collected at a later point in the round. The problem also addresses the decision process related to the number of times the trailer should be parked and where the best sites would be. Solution algorithms that have been described for the TTRP include a tabu search heuristic (Scheuerer, 2006) and a simulated annealing heuristic (Lin et al., 2010). The TTRP with time windows has also been investigated, for which the existing methods include a branch-and-price algorithm that uses a path-based formulation of the problem (Parragh and Cordeau, 2017), a simulated annealing algorithm (Lin et al., 2011) and a memetic heuristic (Mirmohammadsadeghi and Ahmed, 2015). Our problem can be seen as a special case of the TTRP where the clusters are fixed, with the 'truck' corresponding to the driver and the 'trailer' being the van.

Contribution of the study

In an attempt to better replicate the realities of last-mile delivery where servicing groups of customers on-foot makes up a significant part of the driver's daily activity, we make the following contributions in this paper:

1. Based on actual parcel delivery operations, we introduce and formally define a two-level clustered routing problem for a single driver that consists of one driving route and a set of intra-cluster walking routes, and which incorporates time window restrictions dictated by the customers.

4

2. We present a mixed integer programming formulation for the problem that integrates the decisions concerning routing of one driver, determining a parking location in each cluster, and the walking sequences within each cluster. The model is initially stated as a nonlinear integer programming formulation, which is then linearised.

3. We test the model on real data collected from parcel carrier rounds operating in central London and present results that yield the potential savings in overall operation time as well as effects of introducing additional time window restrictions on the resulting solutions.

In contrast to the generalised routing models described above, an explicit consideration of intra-cluster routes and the walking time required makes our problem different to the generalised routing models described above. Furthermore, the fact that intra-cluster visits should be closed tours, as opposed to paths, differentiates our problem from the existing clustered routing models.

Problem Definition and Modelling

Let $G = (V, A)$ be a complete graph, where $V = \{0, 1, 2, \ldots\}$ is the set of nodes and $A = \{(i, j) : i \in V, j \in V, i \neq j\}$ is the set of arcs. Set V is partitioned into a number of clusters, with each cluster V_p identified with an index p from within a set K of indices, with $p = 0$ showing the (singleton) cluster that the depot belongs to. For each arc $(i, j) \in A$, we assume a known driving time $\alpha_{ij} > 0$, and walking time $\beta_{ij} > 0$. Each customer node (or a delivery point) $i \in V \setminus \{0\}$ is associated a hard time window $[a_i, b_i]$ within which the delivery for that node must be made.

The routing problem is defined for a single driver with a van that carries all the parcels of customers that have been assigned to the driver. The choice of the clusters is down to the driver or the dispatcher, which we assume to have also been fixed prior to driver leaving the depot. In each cluster $p \in K \setminus \{0\}$, we assume that the driver will either park at or near a delivery point, which we refer to as a parking location, the selection of which will be part of the decision problem. However, irrespective of the choice of the parking location, we assume a constant walking time ω_p from the parking location to the first delivery point within the cluster, and a constant time ρ_p for parking and unloading parcels from the vehicle. Finally, we assume that the driver spends ζ units of time for handing over parcels at each consignee, which we assume to be constant and identical for each delivery point.

The problem consists of jointly (i) selecting one parking location in each cluster V_p, for all $p \in K$, (ii) finding a driving route that starts and ends at the depot and spans the parking locations, and (iii) determining a set of walking routes for each cluster V_p with $|V_p| \geq 2$ that start and end at the parking node of that cluster. The aim is to minimise a weighted combination of the driving and walking time, weighted by parameters $\gamma \geq 0$ and $\theta \geq 0$ such that $\gamma + \theta = 1$, and to ensure that the deliveries at each customer are made within the respective time windows.

Figure 1 shows an example solution to a problem instance with five clusters, including that of the depot, four of which contain a total of eight delivery points shown by C_1–C_8.
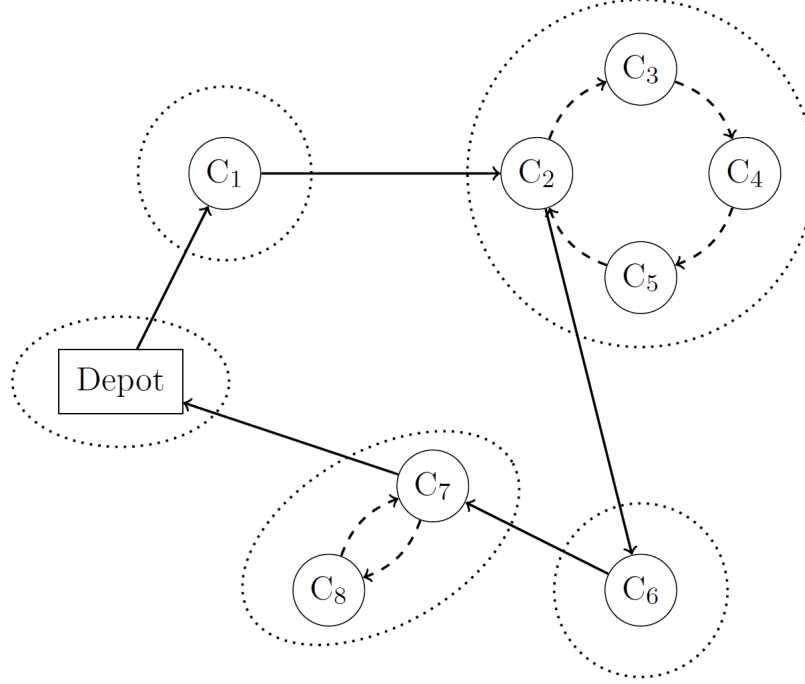
Figure 1: An example solution to a problem instance with eight customers and five clusters

The thick lines in Figure 1 from the depot show the driving sequence where the vehicle parks at (or nearby) customers C_1, C_2, C_6 and C_7, in the given order, before returning to the depot. The dashed-lines show the walking trajectory between consignees' delivery points within each cluster. The driver must visit all consignee delivery points in any cluster and return to the parking location before driving the vehicle to any subsequent cluster. This, for example, can be seen in the cluster containing customers C_2–C_5, in which the vehicle is parked at location C_2, following which the driver visits the customers in the order shown, before returning to C_2 to continue on to the subsequent cluster on the route.

Integer programming formulation

An integer programming formulation for the problem above uses a binary variable x_{ij} for each arc $\{(i,\ j) \in A : i \in V_p, j \in V_l, p, l \in K, p \neq l\}$ that equals 1 if the arc is traversed in the driving mode, and 0 otherwise. Similarly, a binary variable y_{ij} for arc $\{(i,\ j) \in A : i, j \in V_p, p \in K, i \neq j, |V_p| \geq 2\}$ is equal to 1 if the arc is traversed using the walking mode, and is equal to 0 otherwise. We also define, for each cluster $p \in K$, a non-negative continuous variable t_p that shows the arrival time into the cluster, $\widehat{t_i}$ that shows the arrival time at a node $i \in V$ and s_p as the total time spent in a customer that depends on the order in which nodes are visited in the cluster. The problem can then

be modelled as the following integer linear programming formulation that we denote by P1.

$$\text{Minimise} \qquad \gamma \sum_{i \in V_p, p \in K} \sum_{j \in V_l, l \in K \setminus \{p\}} \alpha_{ij} x_{ij} + \theta \sum_{p \in K : |V_p| \geq 2} \sum_{(i,j) \in A : i, j \in V_p} \beta_{ij} y_{ij} \tag{1}$$

subject to

$$\sum_{i \in V_p} \sum_{j \in V \setminus V_p} x_{ij} = 1 \qquad\qquad p \in K \tag{2}$$

$$\sum_{i \in V \setminus V_p} \sum_{j \in V_p} x_{ij} = 1 \qquad\qquad p \in K \tag{3}$$

$$\sum_{i \in V \setminus V_p} x_{ij} = \sum_{i \in V \setminus V_p} x_{ji} \qquad\qquad j \in V_p,\, p \in K \tag{4}$$

$$s_p = \omega_p + \rho_p + \zeta \qquad\qquad p \in V,\, |V_p| = 1 \tag{5}$$

$$s_p = \omega_p + \rho_p + |V_p|\zeta + \sum_{i \in V_p} \sum_{j \in V_p \setminus \{i\}} \beta_{ij} y_{ij} \qquad\qquad p \in K,\, |V_p| \geq 2 \tag{6}$$

$$t_p - t_l + M \sum_{i \in V_p} \sum_{j \in V_l} x_{ij} + s_p \leq M - \sum_{i \in V_p} \sum_{j \in V_l} \alpha_{ij} x_{ij} \qquad\qquad p \in K,\, l \in K \setminus \{0\},\, p \neq l \tag{7}$$

$$t_p - t_l + M \sum_{i \in V_l} \sum_{j \in V_p} x_{ij} - s_l \leq M - \sum_{i \in V_l} \sum_{j \in V_p} \alpha_{ij} x_{ij} \qquad\qquad p \in K,\, l \in K \setminus \{0\},\, p \neq l \tag{8}$$

$$\sum_{i \in V_p \setminus \{j\}} y_{ij} = 1,\, j \in V_p \qquad\qquad p \in K \setminus \{0\},\, |V_p| \geq 2 \tag{9}$$

$$\sum_{j \in V_p \setminus \{i\}} y_{ij} = 1,\, i \in V_p \qquad\qquad p \in K \setminus \{0\},\, |V_p| \geq 2 \tag{10}$$

$$\sum_{i \in V_p} y_{ij} = \sum_{i \in V_p} y_{ji} \qquad\qquad j \in V_p \setminus \{i\},$$

$$\qquad\qquad p \in K \setminus \{0\},\, |V_p| \geq 2 \tag{11}$$

$$\widehat{t}_i \geq t_p \qquad\qquad i \in V_p,\, p \in K \tag{12}$$

$$\widehat{t}_i \leq t_p + s_p \qquad\qquad i \in V_p,\, p \in K \tag{13}$$

$$a_i \leq \widehat{t}_i \leq b_i \qquad\qquad i \in V \tag{14}$$

$$\left(1 - \sum_{h \in V \setminus V_p} x_{hi}\right)\left(\widehat{t}_i - \widehat{t}_j + M y_{ij}\right) \leq M - (\beta_{ij} + \zeta) \qquad\qquad i, j \in V_p,\, i \neq j,$$

$$\qquad\qquad p \in K \setminus \{0\},\, |V_p| \geq 2 \tag{15}$$

$$x_{ij} \in \{0,1\} \qquad\qquad i \in V_p,\, j \in V_l,\, p, l \in K,\, p \neq l \tag{16}$$

$$y_{ij} \in \{0,1\} \qquad\qquad i, j \in V_p,\, p \in K,\, i \neq j,\, |V_p| \geq 2 \tag{17}$$

$$t_p,\, s_p \in \mathbb{R}^+ \qquad\qquad p \in K \tag{18}$$

$$\widehat{t}_i \in \mathbb{R}^+ \qquad\qquad i \in V. \tag{19}$$

The objective function (1) minimises the total weighted travel time which includes driving time between clusters and the time spent in each cluster. Constraints (2) and (3) ensure that only one visit is made to each cluster in the driving mode. Constraint (4) demands that the arrival into and departure from a cluster are made from the same node. Constraints (5) and (6) model the total time spent in each cluster, and differentiate between singletons and those with at least two delivery points. In the former case, the total time includes the parking and unloading time, the walking time from the parking location to the delivery point, and the service time at the delivery point. In the latter case, the total dwell time within a cluster will depend on the sequence of visits made within that cluster, and includes the walking time, parking and unloading time, the walking time between the delivery points, and the total service times of all the nodes of that cluster. Constraints (7) and (8) eliminate subtours between clusters. In particular, they dictate that the arrival time at a cluster l is no earlier than the sum of the time of arrival at the immediately preceding cluster p, the driving time between the two clusters l and p, and the total time s_p spent at cluster p.

Constraints (9)–(11) model the walking tours within each cluster. In particular, each delivery point within a cluster is visited only once. Constraint (12) ensures that the arrival time at a delivery point in a cluster can be no earlier than arrival time into the parking location of that cluster. Similarly, constraint (13) models the restriction that the arrival time to a delivery point in a cluster cannot be later than the departure time from that cluster. Constraint (14) ensures that the service start time to each delivery point must be within the prescribed time window. Constraint (15) eliminates subtours within each cluster. Constraints (16) and (17) imposes the restrictions on the variables relating to the driving and walking decisions. Similarly, constraints (18) and (19) model the non-negativity restrictions for the remaining variables.

To see the validity of the above formulation for the problem under consideration, it suffices to observe that the model represented by constraints (1)–(4), (7), (8), (14), (16) and (18) is that of a TSP with time windows (TSPTW) which we refer to as P1a. Similarly, the model represented by constraints (1), (5), (6), (9)–(15), (17) and (19) that we will name as P1b, also corresponds a TSPTW, for each cluster $p \in K$ with $|V_p| \geq 2$.

We now show the validity of the former model.

Lemma 1 *The model P1a is valid formulation for the TSPTW.*

Proof. Constraints (2)–(4) are the assignment restrictions. Constraints (7) and (8) prohibit the formation of subtours as follows. For a given pair $(p \in K, l \in K \setminus \{0, l\})$ of clusters, let $X(p, l) = \sum_{i \in V_p} \sum_{j \in V_l} x_{ij}$, $X(l, p) = \sum_{i \in V_l} \sum_{j \in V_p} x_{ij}$, $\alpha_{pl} = \sum_{i \in V_p} \sum_{j \in V_l} \alpha_{ij}$, and $\alpha_{lp} = \sum_{i \in V_l} \sum_{j \in V_p} \alpha_{ij}$. For the given pair (p, l) of clusters, constraint (7) can be written as:

$$t_p - t_l + MX(p, l) + s_p \leq M - \alpha_{pl}X(p, l) \tag{20}$$
$$t_l - t_p + MX(l, p) + s_l \leq M - \alpha_{lp}X(l, p), \tag{21}$$

8

and constraint (8) can be written as:

$$t_p - t_l + MX(l,p) - s_l \leq M - \alpha_{lp}X(l,p) \tag{22}$$
$$t_l - t_p + MX(p,l) - s_p \leq M - \alpha_{pl}X(p,l). \tag{23}$$

We consider the following two cases:

1. If the driver goes from the cluster p to the cluster l, then $X(p,l) = 1$ and $X(l,p) = 0$. Therefore, constraints (20) and (23), respectively, describe $t_l \geq t_p + s_p + \alpha_{pl}$ and $t_l \leq t_p + s_p + \alpha_{pl}$, which implies a tight constraint $t_l = t_p + s_p + \alpha_{pl}$. In this case, constraints (21) and (22), respectively become $t_l - t_p + s_l \leq M$ and $t_p - t_l - s_l \leq M$, which are always satisfied with a sufficiently large M.

2. If the driver goes from cluster l to cluster p, then $X(p,l) = 0$ and $X(l,p) = 1$. Therefore, constraint (20) and constraint (23), respectively, imply $t_p - t_l + s_p \leq M$ and $t_l - t_p - s_p \leq M$, which are always satisfied with a sufficiently large M. In this case, constraints (21) and (22) collectively imply a tight constraint $t_p = t_l + s_l + \alpha_{lp}$.

It is now easy to see that constraints (7) and (8) will be infeasible for a given subtour. $\square$

The proof shows that constraints (7) and (8) not only eliminate subtours but also calculate the arrival times in an exact manner. We now show the validity of formulation P1b.

Lemma 2 *The model P1b is valid formulation for the TSPTW where the 'depot' is the parking location chosen by model P1a.*

Proof. Consider a cluster $p \in K$. Constraints (5) and (6) will ensure that all nodes within the cluster will be visited exactly once. As to ensuring a feasible tour, consider the two possible cases below for each node $i \in V_p$:

1. If i is a parking location, then $\sum_{h \in V \backslash V_p} x_{hi} = 1$, in which case constraint (15) reads $M - (\beta_{ij} + \zeta) \geq 0$, for all $i, j \in V_p$, $i \neq j$, $p \in K \backslash \{0\}$, $|V_p| \geq 2$, which is satisfied with a significant large M.

2. If $i \in V$ is not a parking location, then $\sum_{h \in V \backslash V_p} x_{hi} = 0$, in which case constraint (15) reads $\widehat{t_i} - \widehat{t_j} + My_{ij} \leq M - (\beta_{ij} + \zeta)$, for all $i, j \in V_p$, $i \neq j$, $p \in K \backslash \{0\}$, $|V_p| \geq 2$, which is the well-known subtour elimination constraint by Miller et al. (1960). $\square$

It is worth pointing out that, although constraints (15) are written using the weaker subtour elimination constraints, they will always hold tight at optimality to guarantee that $\widehat{t_j} = \widehat{t_i} + \beta_{ij} + \zeta$ when $y_{ij} = 1$ for a given pair $i, j \in V_p$, $i \neq j$ of nodes in a cluster $p \in K$. In particular, any solution where $\widehat{t_j} > \widehat{t_i} + \beta_{ij} + \zeta$ holds for a given pair $i, j \in V_p$, $i \neq j$ of nodes in a cluster $p \in K$ would violate constraints (12) and (13) given that s_p is the minimum possible time needed to service the nodes within the cluster.

Linearised model

Formulation P1 is nonlinear due to constraints (15). In this section, we provide a linearisation of this constraint to allow the formulation to be solved by a mixed integer linear programming solver. To this end, let

$$z_i = \widehat{t_i} \sum_{h \in V \backslash V_p} x_{hi} \qquad i \in V_p,\ p \in K \backslash \{0\},\ |V_p| \geq 2 \tag{24}$$

$$z_{ij}^1 = \widehat{t_j} \sum_{h \in V \backslash V_p} x_{hi} \qquad i,\ j \in V_p,\ i \neq j,\ p \in K \backslash \{0\},\ |V_p| \geq 2 \tag{25}$$

$$z_{ij}^2 = y_{ij} \sum_{h \in V \backslash V_p} x_{hi} \qquad i,\ j \in V_p,\ i \neq j,\ p \in K \backslash \{0\},\ |V_p| \geq 2. \tag{26}$$

We now present the linearisation.

Lemma 3 *Constraint (15) can be rewritten in a linearised form as the following set of constraints*:

$$\widehat{t_i} - \widehat{t_j} + M y_{ij} - z_i + z_{ij}^1 - M z_{ij}^2 \leq M - (\beta_{ij} + \zeta) \qquad i \neq j,\ i,\ j \in V_p,\ p \in K \backslash \{0\},\ |V_p| \geq 2 \tag{27}$$

$$a_i \sum_{h \in V \backslash V_p} x_{hi} \leq z_i \leq b_i \sum_{h \in V \backslash V_p} x_{hi} \qquad i \in V_p,\ p \in K \backslash \{0\},\ |V_p| \geq 2 \tag{28}$$

$$z_i \leq \widehat{t_i} - a_i \left(1 - \sum_{h \in V \backslash V_p} x_{hi} \right) \qquad i \in V_p,\ p \in K \backslash \{0\},\ |V_p| \geq 2 \tag{29}$$

$$z_i \geq \widehat{t_i} - b_i \left(1 - \sum_{h \in V \backslash V_p} x_{hi} \right) \qquad |V_p| \geq 2,\ i \in V_p,\ p \in K \backslash \{0\} \tag{30}$$

$$a_j \sum_{h \in V \backslash V_p} x_{hi} \leq z_{ij}^1 \leq b_j \sum_{h \in V \backslash V_p} x_{hi} \qquad i \neq j,\ i \in V_p, j \in V_p,\ p \in V \backslash \{0\},\ |V_p| \geq 2 \tag{31}$$

$$z_{ij}^1 \leq \widehat{t_j} - a_j \left(1 - \sum_{h \in V \backslash V_p} x_{hi} \right) \qquad i \neq j,\ i \in V_p,\ j \in V_p,\ p \in V \backslash \{0\},\ |V_p| \geq 2 \tag{32}$$

$$z_{ij}^1 \geq \widehat{t_j} - b_j \left(1 - \sum_{h \in V \backslash V_p} x_{hi} \right) \qquad i \neq j,\ i \in V_p,\ p \in V \backslash \{0\},\ |V_p| \geq 2 \tag{33}$$

$$0 \leq z_{ij}^2 \leq \sum_{h \in V \backslash V_p} x_{hi} \qquad i \neq j,\ i \in V_p,\ j \in V_p,\ p \in K \backslash \{0\},\ |V_p| \geq 2 \tag{34}$$

$$z_{ij}^2 \leq y_{ij} \qquad i,\ j \in V_p,\ i \neq j,\ p \in K \backslash \{0\},\ |V_p| \geq 2 \tag{35}$$

$$z_{ij}^2 \geq y_{ij} + \sum_{h \in V \backslash V_p} x_{hi} - 1 \qquad i \in V_p,\ p \in K \backslash \{0\},\ |V_p| \geq 2. \tag{36}$$

Proof. We first consider the nonlinear term $z_i = \widehat{t_i} \sum\limits_{h \in V \setminus V_p} x_{hi}$ where x_{hi} is binary and $\widehat{t_i}$ is restricted to be within the interval $[a_i, b_i]$. Let $u = \sum\limits_{h \in V \setminus V_p} x_{hi}$. Constraints (2) and (3) imply that u is also a binary variable. Therefore, the product z_i can be written as $z_i = \widehat{t_i} \times u$ and can be represented as the following linear set of inequalities:

$$u \times a_i \leq z_i \leq u \times b_i \tag{37}$$
$$z_i \leq \widehat{t_i} - a_i (1 - u) \tag{38}$$
$$z_i \geq \widehat{t_i} - b_i (1 - u). \tag{39}$$

To see that (37)–(39) linearise the nonlinear term, it suffices to observe that if $u = 0$ then $z_i = 0$, otherwise the inequalities read $a_i \leq z_i \leq b_i$, $z_i \leq \widehat{t_i}$ and $z_i \geq \widehat{t_i}$, and collectively imply $z_i = \hat{t_i}$. Constraints (31)–(33), and (34)–(36) linearise constraints (25) and (26), respectively, in the same way. $\square$

The linearised problem represented by constraints (1)– (14), (16)–(19), and (27)– (36) now is solvable with any commercial MILP optimiser.

Case Study

In this section, we report on numerical experiments on the application of the proposed model and the insights of the clustering and routing has revealed, using data from a carrier operating in London during October 2016. The model is implemented in IBM ILOG CPLEX Optimization Studio V12.70 on Desktop Intel Core i5, Processor 3.3GHz running under the Windows 10 operating system. A computational time limit of 10 hours has been imposed on the solution time of the model for each run.

Input data description

The input data set used here forms a part of a larger set of data collected, full details of which are described in Allen et al. (2017a). In particular, the input data set corresponds to a selected round where one driver make 104 deliveries (transactions) to 99 consignees from 57 stopping locations (which is thereafter named as delivery locations) on the 27th October 2016. The delivery locations are used to determine the driving and walking times using Google maps.

The carrier offers three different types of time windows, namely before 9am, before 10am and before 12pm, as are commonly used in practical deliveries. In our input data set, only six out of the 57 delivery locations have a time window attached to them, namely one customer has requested a delivery before 9am, two customers before 10am and the remaining three before 12pm.

Figure 2 shows the actual route used by the driver along with the clustering chosen. Figure 2 shows the actual routing used by the driver along with the 49 clusters chosen, most of which are singletons.
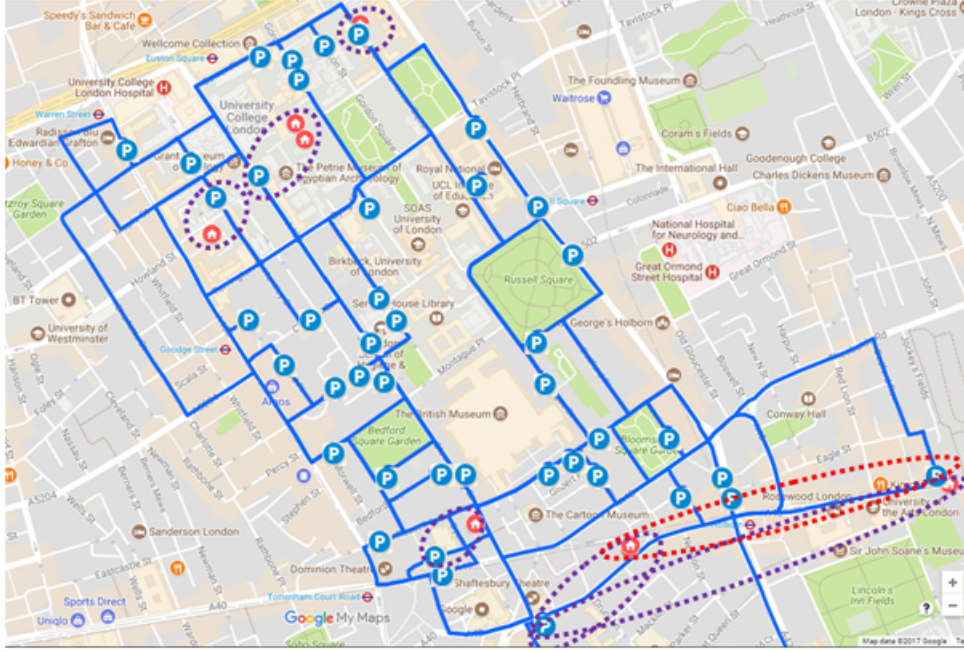

Figure 2: The actual clustering & routing strategy.

An analysis of the actual operations revealed that 70% of delivery points lied within a minute walk from where the driver chose as the parking location. As the driver needs time to park the vehicle and unload parcel/s, the analysis suggested that parking plus unloading time was about a minute for all clusters. The model was therefore implemented using a one minute walking time allowance ω_p and an additional minute for parking and unloading time ρ_p for each cluster $p \in K \setminus \{0\}$. Finally, consignee service time ζ at each delivery point was also assumed to be a minute. The value of big-M is set to the latest time that the driver needs to return to the depot, which, on the basis of a maximum of nine working hours a day for each driver, translates into a value equal to 540 minutes. The weights γ, θ in the objective function are set equal to 1 and 0, respectively.

A number of statistics on the actual operations are shown in Table 1, under the column 'Actual', where we report the total driving time (T_d), total walking time (T_w), total walking allowance (T_a), total parking, unloading and service time (T_p), total (dwell) time spent in the clusters $(T_u = T_w + T_a + T_p)$, Total operating time $(T_o = T_d + T_u)$ and finally the maximum time spent in any cluster (T_{max}) extracted from the data collected on the day.

Using inputs described above, we then undertook a validation of the parameters used in the model by fixing the sequence of visits of the delivery points as that chosen by the driver on the day and report the same statistics under column 'Model' of Table 1, and comparison results under column six of the table showing percentage differences. As the comparisons show, the model overestimates the driving time by 17.5% and underestimates the total time spent in the clusters by 17.9%. The

discrepancies are due to the estimated times by the online map source as well as possible 'short-cuts' that the driver might have taken on the day. However, the model was able to output the same total operation time, and identified nearly the same value for the maximum time spent in a cluster. The latter two results give us sufficient indication that the parameters of the model can be considered to be accurate enough, which will be used in the rest of the experiments.

Table 1: Summary of the case study's results

	Performance (in minutes)				Comparison (%)		
	Actual	Model	Scenario A	Scenario B	Model vs. Actual	Scenario A vs. Model	Scenario B vs. Scenario A
T_d	240	282	130	115	17.5	−53.9	−11.5
T_w	65	48	82	105	–	–	–
T_a	48	48	48	34	–	–	–
T_p	121	96	96	68	–	–	–
T_u	234	192	226	207	−17.9	17.7	-8.4
T_o	474	474	356	322	0	−24.9	−9.6
T_{max}	28	29	29	23	3.4	0	−20.6

Figure 3: Routing optimisation using the existing clusters

Optimisation results

In this section, we use the model to optimise the routing under two scenarios, namely Scenario A where we leave the 49 clusters chosen by the driver intact and only optimise the route, and Scenario

B where we propose a new design comprising 35 clusters[1], mainly chosen through geographical proximity of the delivery points, and optimise the route on these new clusters. The results of the two scenarios are also shown in Table 1.

Figure 3 shows the resulting optimal solution under Scenario A, which, as also indicated by the results in the table, leads to a significant reduction (53.9%) on the total driving time, but with an increase of 17.7% the total dwell time. The total operation time however reduced by 24.9%, with no increase in the maximum dwell time spent in any cluster.

As for Scenario B, the optimal routes identified and the corresponding statistics are shown in Figure 4 and Table 1, respectively. The new route indicates savings of 9.6%, 11.5%, 8.4% and 20.6% as compared to Scenario A on T_o, T_d, T_u and T_{max}, respectively.
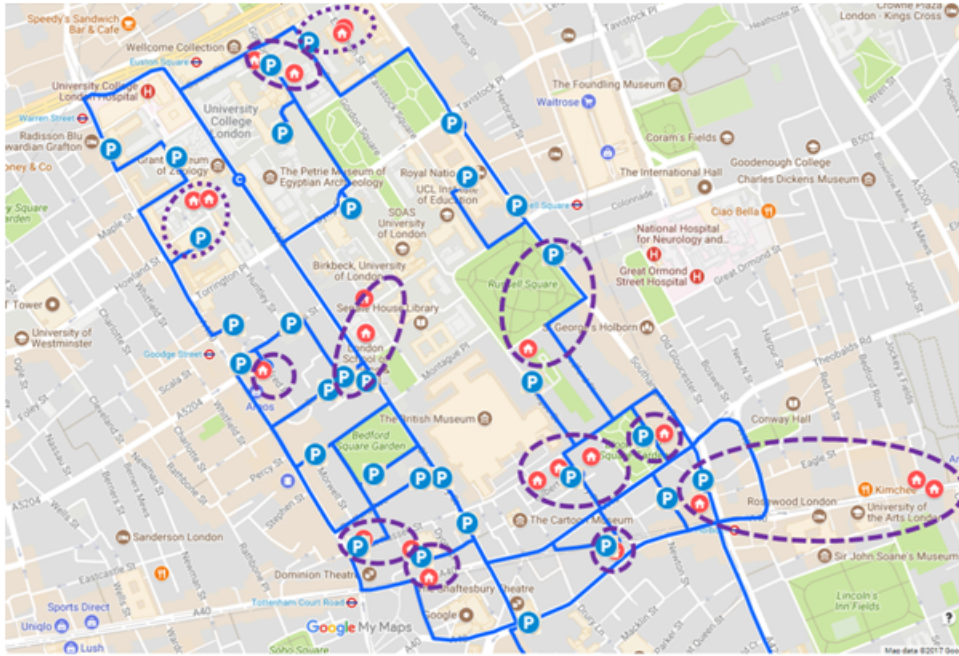

Figure 4: Routing optimisation of a proposed strategy

The average dwell times in Scenarios A and B are 15 minutes, and 10.5 minutes, respectively, in comparison to the 13.5 minutes assumed in the actual operations as prescribed by the model. In addition, whereas the driver spent 1.6 hours on foot (6.5 km) in the actual operation, this number changes to 2 hours (7.3 km) and 2.3 hours (10 km) in Scenarios A and B, respectively. The proposed cluster design implies that the reduction in the overall operation time comes at the expense of the driver needing to spend more time on foot than with the existing clusters. Moreover, only one cluster of the proposed clustering strategy exceeds a maximum of 20 minutes parking time, while 2 clusters of the real operation and the optimal routing with the existing clusters do (Figure 5).

Finally, we look at implication of the workload within each cluster in Scenarios A and B, as shown in Figure 6. The results indicate that with the proposed clustering, the majority of the proposed

---

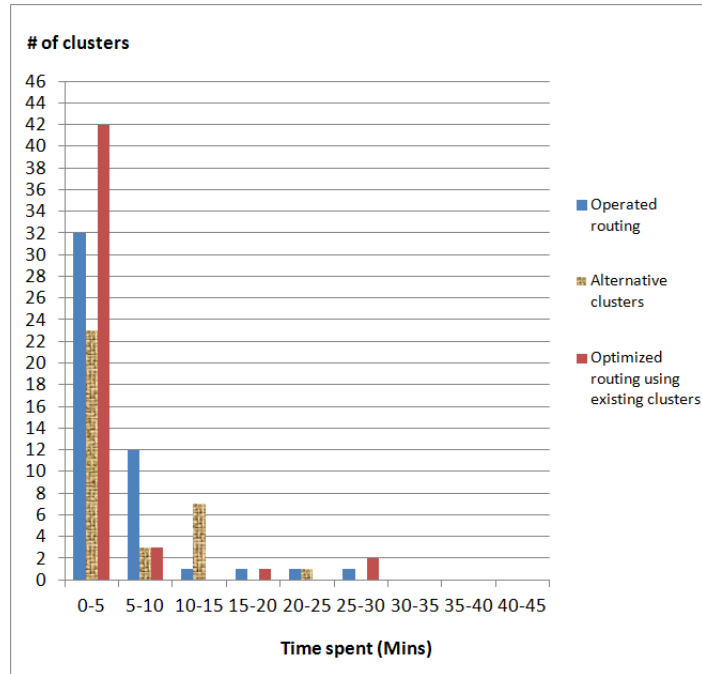[1]We elaborate further on the choice of clustering in the conclusions.

Figure 5: Number of clusters having the time spent within chosen ranges

clusters have work-loads within the maximum of 90 kilograms, except one cluster with a work-load of 100 kilograms.
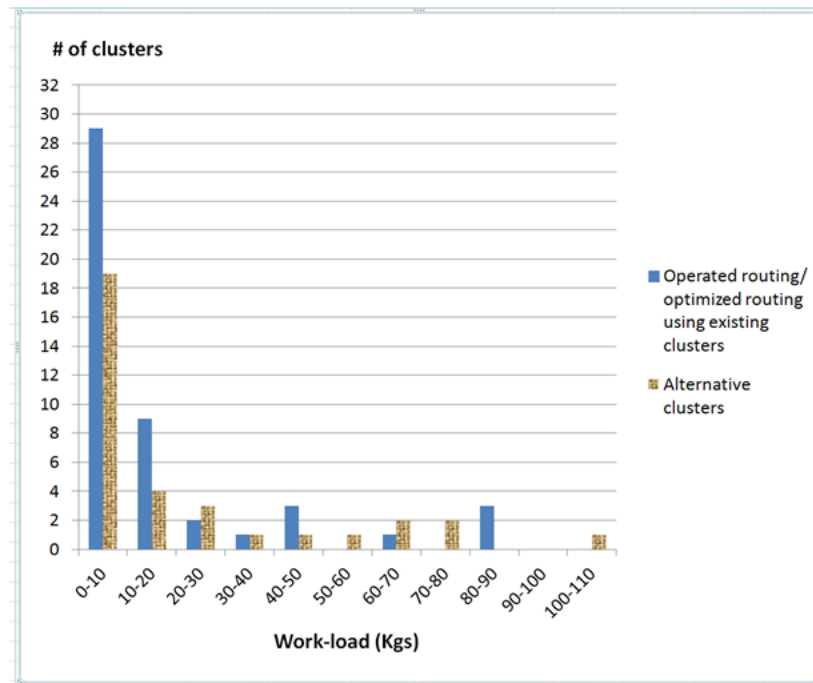

Figure 6: Number of clusters having the work-load within chosen ranges

The impact of timed services

A key issue impacting on the efficiency of parcel logistics is the proportion of consignees who request specific timed services, and the impact of the increase of this proportion in line with the grown in same-day delivery requests. To shed some light into this issue, we present the results of additional experiments in this section by increasing the number of customers with time-sensitive deliveries. To this end, we first randomly choose $\sigma\%$ of the total number $|V|-1$ of nodes to which time windows are to be added, where $\sigma \in \{10, 15, 20\}$ and randomly assign the time windows to these nodes. There are four ways in which the assignment can be performed: one way is to (approximately) equally distribute the three types of time windows (before 9am, before 10am and before 12pm) amongst the chosen nodes, and the other ways are to uniformly apply one of the three time windows to the chosen nodes. In this way, we generate five instances for each configuration which results in a total of $5 \times 3 \times 4 = 60$ instances. Table 2 presents the results of the experiments.

As the table shows, increasing the number of time-sensitive deliveries results in a 43.3% of the instances being infeasible, indicating that the time window constraints would have to be violated in order for a feasible route to be identified. In this case, the computational performance of the model also deteriorates. Under a time limit of 10 hours, we identify feasible, but not optimal routes for only 21.7% of the instances. In this case, the optimality gaps range from an average value of 2.97% to a maximum value of 17.8%.

For instances that are feasible (or optimal), however, the addition of time-window restrictions increase the total driving time only marginally, if at all, indicating that those locations with the new time window restrictions do not change the optimal routes, particularly as the total dwell time either remains unchanged or is marginally increased. The maximum dwell time remains the same in all generated instances, showing the robustness of the design with respect to changes in the time window constraints.

Conclusions

We have described a practical distribution problem that arises in last-mile deliveries and presented a mathematical programming formulation that, when applied to the problem, results in optimised solutions that offer reductions ing operations time, and particularly driving time. The latter has implications in terms of reduction of traffic and energy requirements within urban areas, as well as greenhouse gas emissions.

The model operates on the basis of a fixed set of clusters within urban zones, which is one aspect that limits its applicability. One extension of this work would be to optimising the cluster formation along with the routing and scheduling decisions, which would give rise to a much more difficult optimisation problem (e.g., as in Parragh and Cordeau, 2017), and where the difference between the two modes of walking and driving would have to be explicitly represented. In particular, the

Table 2: Summary of different percentage of delivery points with time window

Type of additional time windows			Equally distributed			Before 9AM			Before 10AM			Before 12PM		
	σ		10%	15%	20%	10%	15%	20%	10%	15%	20%	10%	15%	20%
T_d (minutes)	Instance	1	127.0	116.0	-	-	-	-	127.0	115.0	125.0	115.0	115.0	115.0
		2	119.0	115.0	-	-	115.0	-	121.0	115.0	-	116.0	115.0	-
		3	118.0	116.0	-	-	-	-	119.0	116.0	-	115.0	116.0	115.0
		4	115.0	115.0	-	-	116.0	-	123.0	115.0	139.0	115.0	115.0	118.0
		5	119.0	115.0	136	-	-	-	132.0	120.0	132.0	116.0	115.0	116.0
	Average		119.6	115.4	136	-	115.5	-	124.4	116.2	132.0	115.4	115.2	116.0
T_u (minutes)	Instance	1	204.0	207.0	-	-	-	-	206.0	208.0	206.0	204.0	208.0	206.0
		2	207.0	206.0	-	-	206.0	-	206.0	220.0	-	204.0	207.0	-
		3	206.0	208.0	-	-	-	-	207.0	207.0	-	204.0	207.0	204.0
		4	207.0	208.0	-	-	206.0	-	206.0	207.0	206.0	204.0	207.0	204.0
		5	205.0	207.0	206	-	-	-	206.0	206.0	-	204.0	207.0	204.0
	Average		205.8	207.2	206	-	206.0	-	206.2	209.6	206.0	204.0	207.2	204.5
T_{max} (minutes)	Instance	1	23.0	23.0	-	-	-	-	23.0	23.0	23.0	23.0	23.0	23.0
		2	23.0	35.0	-	-	23.0	-	23.0	35.0	-	23.0	23.0	-
		3	23.0	23.0	-	-	-	-	23.0	23.0	-	23.0	23.0	23.0
		4	23.0	23.0	23	-	23.0	-	23.0	23.0	23.0	23.0	23.0	23.0
		5	23.0	23.0	23	-	-	-	23.0	23.0	-	23.0	23.0	23.0
	Average		23.0	25.4	23	-	23.0	-	23.0	25.4	23.0	23.0	23.0	23.0
Run time (seconds)	Instance	1	36,307.0	432.0	-	-	-	-	36,050.0	117.0	36,021.0	18.0	10.0	89.0
		2	36,014.0	33.0	-	-	40.0	-	36,037.0	209.0	-	260.0	4.0	-
		3	36,048.0	122.0	-	-	-	-	36,038.0	397.0	-	24.0	36.0	11,660.0
		4	16.0	22.0	36,049.0	-	129.0	-	36,037.0	18.0	36,049.0	957.0	44.0	36,327.0
		5	36,024.0	15.0	36,049.0	-	-	-	36,052.0	36,033.0	-	6,462.0	31.0	35,452.0
	Average		28,881.8	124.8		-	84.5	-	36,042.8	7,354.8	36,035.0	1,544.2	25.0	20,882.0
Optimality gap (%)	Instance	1	9.4	0	-	-	-	-	10.1	-	8.7	0	0	0
		2	2.5	0	-	-	0	-	4.2	-	-	0	0	-
		3	1.7	0	-	-	-	-	3.3	-	-	0	0	0
		4	-	0	16.1	-	0	-	6.8	-	17.8	0	0	2.54
		5	2.4	0	16.1	-	-	-	12.3	3.3	-	0	0	0
	Average		3.2	0	16.1	-	0	-	7.4	0.7	13.3	0	0	0.6

Notes: - :   *No feasible solution found within the time limit*

17

underlying graphs and other data (e.g., travel time) for the two modes exhibit substantial differences, which would need to be taken into account in such an extended model.

Although the model was able to identify optimal solutions to the case study, the difficulties in solving the model when there are additional time-window restrictions necessitate the adaptation or extension of the existing heuristics to the problem at hand, but this would once again have to be done with an explicit consideration of the dual-mode nature of the underlying routing strategy.

Acknowlegement

References

Ahmed, Z. H. (2014). The ordered clustered travelling salesman problem A hybrid genetic algorithm. *The Scientific World Journal.* Article ID 258207.

Allen, J., Bektas, T., Cherrett, T., Friday, A., Mcleod, F., Piecyk, Majaand Piotrowska, M., and Zaltz Austwick, M. (2017a). Enabling the freight traffic controller for collaborative multi-drop urban logistics: practical and theoretical challenges. page 17. Transportation Research Board 96th Annual Meeting (TRB 2017), Washington, US, 17-04295, 08 - 12 Jan 2017.

Allen, J., Piecyk, M., Piotrowska, M., Mcleod, F.and Cherrett, T., Nguyen, T., Bektas, T., Bates, O.and Friday, A., Wise, S., and Austwick, M. (2017b). Understanding the impact of e-commerce on last-mile light goods vehicle activity in urban areas: the case of london. *Transportation Research Part D: Transport and Environment.*

Anily, S., Bramel, J., and Hertz, A. (1999). A 5/3-approximation algorithm for the clustered traveling salesman tour and path problems. *Operations Research Letters*, 24(1):29–35.

Baldacci, R., Bartolini, E., and Laporte, G. (2010). Some applications of the generalized vehicle routing problem. *Journal of the Operational Research Society*, 61(7):1072–1077.

Bektaş, T., Erdogan, G., and Ropke, S. (2011). Formulations and branch-and-cut algorithms for the generalized vehicle routing problem. *Transportation Science*, 45(3):299–316.

Ding, C., Cheng, Y., and He, M. (2007). Two-level genetic algorithm for clustered traveling salesman problem with application in large-scale TSPs. *Tsinghua Science & Technology*, 12(4):459–465.

Fischetti, M., Gonzalez, J. J. S., and Toth, P. (1995). The symmetric generalized traveling salesman polytope. *Networks*, 26(2):113–123.

Fischetti, M., Gonzalez, J. J. S., and Toth, P. (1997). A branch-and-cut algorithm for the symmetric generalized traveling salesman problem. *Operations Research*, 45(3):378–394.

Ghiani, G. and Improta, G. (2000). An efficient transformation of the generalized vehicle routing problem. *European Journal of Operational Research*, 122(1):11–17.

IMRG and Metapack (2016). Imrg metapack uk delivery index report, metapack. *https://www.imrg.org/data-and-reports/imrg-metapack-delivery-indexes/*.

Jongens, K. and Volgenant, T. (1985). The symmetric clustered traveling salesman problem. *European Journal of Operational Research*, 19(1):68–75.

Laporte, G., Asef-Vaziri, A., and Sriskandarajah, C. (1996). Some applications of the generalized travelling salesman problem. *Journal of the Operational Research Society*, 47(12):1461–1467.

Laporte, G., Mercure, H., and Nobert, Y. (1987). Generalized travelling salesman problem through n sets of nodes: the asymmetrical case. *Discrete Applied Mathematics*, 18(2):185–197.

Lin, S.-W., Yu, V. F., and Chou, S.-Y. (2010). A note on the truck and trailer routing problem. *Expert Systems with Applications*, 37(1):899–903.

Lin, S. W., Yu, V. F., and Lu, C. C. (2011). A simulated annealing heuristic for the truck and trailer routing problem with time windows. *Expert Systems with Applications*, 38(12):15244–15252.

Metapack (2015). Imrg metapack uk delivery index report, metapack.

Miller, C. E., Tucker, A. W., and Zemlin, R. A. (1960). Integer programming formulation of traveling salesman problems. *Journal of the Association for Computing Machinery*, 7(4):326–329.

Mirmohammadsadeghi, S. and Ahmed, S. (2015). Memetic heuristic approach for solving truck and trailer routing problems with stochastic demands and time windows. *Networks and Spatial Economics*, 15(4):1093–1115.

Moccia, L., Cordeau, J. F., and Laporte, G. (2012). An incremental tabu search heuristic for the generalized vehicle routing problem with time windows. *Journal of the Operational Research Society*, 63(2):232–244.

Noon, C. E. and Bean, J. C. (1991). A Lagrangian based approach for the asymmetric generalized traveling salesman problem. *Operations Research*, 39(4):623–632.

Parragh, S. N. and Cordeau, J. F. (2017). Branch-and-price and adaptive large neighborhood search for the truck and trailer routing problem with time windows. *Computers & Operations Research*, 83:28–44.

Piotrowska, M., Piecyk, M., Allen, J., McLeod, F., Austwick, M., Bates, O., Friday, A., Wise, S., Bektas, T., Nguyen, T., Ghali, K., and Cherrett, T. (2017). Analysing the performance of last-mile parcel delivery operations in london. *22nd Annual Conference of The Chartered Institute of Logistics & Transport, Logistics Research Network (LRN), Southampton Solent University, GB.*

Pop, P. C., Kara, I., and Marc, A. H. (2012). New mathematical models of the generalized vehicle routing problem and extensions. *Applied Mathematical Modelling*, 36(1):97–107.

Pop, P. C., Pintea, C., Zelina, I., and Dumitrescu, D. (2009). Solving the generalized vehicle routing problem with an ACS-based algorithm. *AIP Conference Proceedings*, 1117(1):157–162.

Royal Mail, p. l. c. (2016). Market overview. *www.royalmailgroup.com/ar2014-15/overview/marketoverview.html. Accessed July 25, 2016.*

Scheuerer, S. (2006). A tabu search heuristic for the truck and trailer routing problem. *Computers & Operations Research*, 33(4):894–909.

Schneider, J. J., Bukur, T., and Krause, A. (2010). Traveling salesman problem with clustering. *Journal of Statistical Physics*, 141(5):767–784.

Toth, P. and Vigo, D. (2014). *Vehicle Routing: Problems, Methods, and Applications, Second Edition.* Number 18 in MOS-SIAM Series on Optimization. SIAM.

Vidal, T., Battarra, M., Subramanian, A., and Erdogan, G. (2015). Hybrid metaheuristics for the clustered vehicle routing problem. *Computers & Operations Research*, 58:87–99.