Chapter 2

The *Vehicle Routing Problem*: origins and variants

Rardin (1998) states that the organizing of a collection of customer locations, jobs, cities, or points, into sequences and routes are among the most common discrete optimization problems. The first of the two review chapters focus on the origins and the mathematical formulation of the VRP and its variants.

2.1 The origins of the basic VRP

2.1.1 The Traveling Salesman Problem (TSP)

The simplest, and probably most famous of routing problems known to researchers is the TSP that seeks a minimum-total-length route visiting every one of N points in a given set $V = \{1, 2, \ldots, N\}$ exactly once across an arc set A. The distance between all point combinations in A, (i, j), where $(i, j) \in V | i \neq j$, is known. In the notation introduced by Rardin (1998), the symbol '$\triangleq$' denotes *defined to be*. With the decision variable x_{ij} defined as:

$$x_{ij} \triangleq \begin{cases} 1 & \text{if a salesman travels from node } i \text{ to node } j \text{, where } i, j = \{1, 2, \ldots, N\} \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

we formulate the problem as

$$\min z = \sum_{(i,j) \in A} c_{ij} x_{ij} \tag{2.2}$$

subject to

$$\sum_{i=1}^{N} x_{ij} = 1 \qquad \forall j \in \{2,\ldots,N\} \tag{2.3}$$

$$\sum_{j=1}^{N} x_{ij} = 1 \qquad \forall i \in \{2,\ldots,N\} \tag{2.4}$$

$$\sum_{i\in S}\sum_{j\in S} x_{ij} \leq |S| - 1 \qquad \forall S \subset V \tag{2.5}$$

$$x_{ij} \in \{0,1\} \qquad \forall i,j \in \{2,\ldots,N\} \tag{2.6}$$

The objective of the problem minimizes the total distance traveled in (2.2). Each node must be visited exactly once according to (2.3) and (2.4), also referred to as *degree constraints*. Subtours are eliminated through the introduction of (2.5). The $|S|$ denotes the number of elements in the subset S. Schrage (2002) states that there are of the order 2^n constraints of type (2.5), as opposed to the alternative in (2.7)

$$u_j \geq u_i + 1 - (1 - x_{ij})n \qquad \forall j \in \{2,\ldots,N\} | j \neq i \tag{2.7}$$

of which there are of the order $N-1$ constraints. Only a few of the former type constraints will be binding in the optimum. Padberg and Rinaldi (1987) therefor propose an efficient and effective iterative process of adding violated constraints of type (2.5) as needed.

Although a number of TSP variations exist, our interest is in the variant where multiple salesmen are routed simultaneously.

2.1.2 The Multiple Traveling Salesman Problem (MTSP)

The MTSP is similar to the notoriously difficult TSP that seeks an optimal tour of N cities, visiting each city exactly once with no sub-tours. In the MTSP, the N cities must be partitioned into M tours, with each tour resulting in a TSP for one salesperson. The MTSP is more difficult than the TSP because it requires determining which cites to assign to each salesperson, as well as the optimal ordering of the cities within each salesperson's tour (Carter and Ragsdale, 2005; Kara and Bektas, 2005). Consider a complete directed graph $G = (V, A)$ where V is the set of N nodes (or cities to be visited), A is the set of arcs and $C = (c_{ij})$ is the cost (distance) matrix associated with each arc $(i, j) \in A$. The cost matrix can be symmetric, asymmetric, or Euclidean. The latter refers to the straight-line distance measured between the two geographically dispersed nodes. There are M salesmen based at the depot, denoted as node 1. The single depot MTSP consists of finding tours for the M salesmen subject to each salesman starting and ending at the depot, each node is located in exactly one tour, and the number of nodes visited by a salesman lies within a predetermined time (or distance) interval. The objective is to minimize the cost of visiting all the nodes. We define the decision variable, x_{ij}, in (2.1). For any salesman, u_i denotes the number of nodes visited on that salesman's route up to node i, with corresponding parameters K and L denoting the minimum and maximum number of nodes visited by any one salesman, respectively. We can therefor state that $1 \leq u_i \leq L$ when $i \geq 2$, and when $x_{i1} = 1$, then $K \leq u_i \leq L$. The following Integer Linear Program (ILP) formulation is proposed by Kara and Bektas (2005).

$$\min z = \sum_{(i,j)\in A} c_{ij}x_{ij} \tag{2.8}$$

subject to

$$\sum_{j=2}^{N} x_{1j} = M \tag{2.9}$$

$$\sum_{i=2}^{N} x_{i1} = M \tag{2.10}$$

$$\sum_{i=1}^{N} x_{ij} = 1 \qquad \forall j \in \{2,\ldots,N\} \tag{2.11}$$

$$\sum_{j=1}^{N} x_{ij} = 1 \qquad \forall i \in \{2,\ldots,N\} \tag{2.12}$$

$$u_i + (L-2)x_{1i} - x_{i1} \leq L-1 \qquad \forall i \in \{2,\ldots,N\} \tag{2.13}$$

$$u_i + x_{1i} + (2-K)x_{i1} \geq 2 \qquad \forall i \in \{2,\ldots,N\} \tag{2.14}$$

$$x_{1i} + x_{i1} \leq 1 \qquad \forall i \in \{2,\ldots,N\} \tag{2.15}$$

$$u_i - u_j + Lx_{ij} + (L-2)x_{ji} \leq L-1 \qquad \forall i,j \in \{2,\ldots,N\} | i \neq j \tag{2.16}$$

$$x_{ij} \in \{0,1\} \qquad \forall i,j \in \{2,\ldots,N\} \tag{2.17}$$

The objective in (2.8) minimizes the total cost of traveling to all nodes, while constraints (2.9) and (2.10) ensures that all M salesmen are allocated routes. Degree constraints are imposed by (2.11) and (2.12). The MTSP-specific constraints (2.13) and (2.14) are referred to as *bounding constraints* and Kara and Bektas (2005) introduce these as the upper and lower bound constraints on the number of nodes visited by each salesman. The value of u_i is initialized to 1 if and only if node i is the first node on the tour of any salesman. Inequality (2.15) forbids a salesman to only visit a single node on its tour. The formation of subtours between all nodes in $V \setminus \{1\}$ (all nodes except the depot) are eliminated by (2.16) as it ensures that $u_j = u_i + 1$ if and only if $x_{ij} = 1$. They are also referred to as Subtour Elimination Constraints (SEC).

Next we consider a variant where each of the M salespeople has a predefined, yet similar, capacity. An analogy is having salespeople traveling with samples in their vehicles. Not only do their cars have limited space for the samples, but each customer visited may require a different number of the samples. As a variant of the MTSP it is referred to as the Capacitated Multiple Traveling Salesman Problem (CMTSP), but in the context of this thesis the vehicular related name, Vehicle Routing Problem (VRP), is preferred.

2.1.3 The Vehicle Routing Problem (VRP)

The distribution problem in which vehicles based at a central facility (depot) are required to visit — during a given time period — geographically dispersed customers in order to fulfill known customer requirements are referred to as the VRP (Christofides, 1985). The main objective of the VRP is to minimize the distribution costs for individual carriers, and can be described as the problem of assigning optimal delivery or collection routes from a depot to a number of geographically distributed customers, subject to constraints (**?**). The most basic version of the VRP have also been called *vehicle scheduling*, *truck dispatching*, or simply the *delivery problem*. A number of different formulations appear in the authoritative work of Christofides (1985). The basic problem can be defined with $G = (V, A)$ being a directed graph where $V = \{v_1, \ldots, v_N\}$ is a set of vertices representing N customers, and with v_1 representing the depot where M identical vehicles, each with capacity Q, are located (**?**). $E = \{(v_i, v_j)|v_i, v_j \in V, i \neq j\}$ is the edge set connecting the vertices. Each vertex, except for the depot $(V\backslash\{v_1\})$, has a non-negative demand q_i and a non-negative service time s_i. A matrix $C = (c_{ij})$ is defined on A. In some contexts, c_{ij} can be interpreted as travel cost, travel time, or travel distance for any of the identical vehicles. Hence, the terms cost, time, and distance are used interchangeably, although t_{ij} denotes the travel time between nodes i and j in the formulation provided below. The basic VRP is to route the vehicles one route per vehicle, each starting and finishing at the depot, so that all customers are supplied with their demands and the total travel cost is minimized. Although Christofides (1985) presents three different formulations from the early 1980s, the following mathematical formulation of the VRP is adapted from Bodin et al. (1983) and Filipec et al. (1998). During this period little changes were made to the formulation of the problem. The decision variable, x_{ij}^k is defined as

$$x_{ij}^{k} \triangleq \begin{cases} 1 & \text{if vehicle } k \text{ travels from node } i \text{ to } j, \text{ where} \\ & i,j \in \{1,2,\ldots,N\} | i \neq j, \text{ and } k \in \{1,2,\ldots,K\} \\ 0 & \text{otherwise} \end{cases} \tag{2.18}$$

$$\min z = \sum_{i=0}^{N} \sum_{\substack{j=0 \\ j\neq i}}^{N} \sum_{k=1}^{K} c_{ij} x_{ij}^{k} \tag{2.19}$$

subject to

$$\sum_{i=0}^{N} \sum_{k=1}^{K} x_{ij}^{k} = 1 \qquad \forall j \in \{1,\ldots,N\} \tag{2.20}$$

$$\sum_{j=0}^{N} \sum_{k=1}^{K} x_{ij}^{k} = 1 \qquad \forall i \in \{1,\ldots,N\} \tag{2.21}$$

$$\sum_{i=0}^{N} x_{ip}^{k} - \sum_{j=0}^{N} x_{pj}^{k} = 0 \qquad \forall p \in \{1,\ldots,N\}, k \in \{1,\ldots,K\} \tag{2.22}$$

$$\sum_{j=0}^{N} q_j \left(\sum_{i=0}^{N} x_{ij}^{k} \right) \leq Q \qquad \forall k \in \{1,\ldots,K\} \tag{2.23}$$

$$\sum_{i=0}^{N} \sum_{j=0}^{N} t_{ij} x_{ij}^{k} \leq D \qquad \forall k \in \{1,\ldots,K\} \tag{2.24}$$

$$\sum_{j=1}^{N} x_{0j}^{k} \leq 1 \qquad \forall k \in \{1,\ldots,K\} \tag{2.25}$$

$$\sum_{i=1}^{N} x_{i0}^{k} \leq 1 \qquad \forall k \in \{1,\ldots,K\} \tag{2.26}$$

$$x_{ij}^{k} \in \{0,1\} \qquad \forall i,j \in \{1,\ldots,N\}, k \in \{1,\ldots,K\} \tag{2.27}$$

The degree constraints are represented by (2.20) and (2.21). Route continuity is enforced by (2.22) as once a vehicle arrived at a node, it must also leave that node. No one vehicle can service customer demands that exceeds the vehicle capacity in (2.23). A maximum route length is limited by (2.24). Constraints (2.25) and (2.26) ensures that each vehicle is scheduled no more than once.

2.2 Variants of the VRP

The basic VRP makes a number of assumptions, including utilizing a homogeneous fleet, a single depot, one route per vehicle, etc. These assumptions can be eliminated by introducing additional constraints to the problem. This implies increasing the complexity of the problem, and, by restriction, classifies the extended problem as an *np*-hard problem. It should be noted that most of these additional constraints are often implemented in isolation, without integration, due to the increased complexity of solving such problems. In the next few sections, these variants are introduced in isolation, before proposing an integrated formulation in Section 2.3.

2.2.1 The concept of time windows

A *time window* can be described as a window of opportunity for deliveries. It is an extension of the VRP that has been researched extensively (Ibaraki et al., 2005; Taillard, 1999; Taillard et al., 1997; Tan et al., 2001c). A time window is the period of time during which deliveries can be made to a specific customer i, and has three main characteristics:

- Earliest allowed arrival time, e_i, also referred to as the *opening time*
- Latest allowed arrival time, l_i, also referred to as the *closing time*
- Whether the time window is considered *soft* or *hard*

Consider the example, illustrated in Figure 2.1, where customer i requests delivery between 07:30 and 17:00. To distinguish between the actual and the specified times of arrival, the

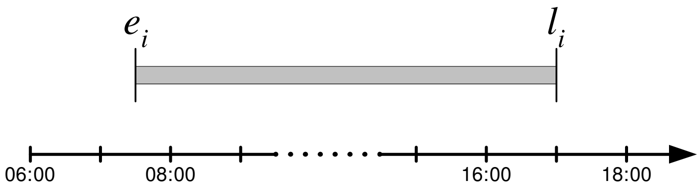


Figure 2.1: Double sided hard time window

variable a_i denotes the actual time of arrival at node i. Should the actual arrival time at node i, denoted by a_i, be earlier than the earliest allowed arrival at the node, e_i, then the vehicle will incur a waiting time, w_i, which can be calculated as $w_i = \max\{0, e_i - a_i\}$. The introduction of time windows to the basic VRP sees the introduction of three new constraints.

$$a_0 = w_0 = s_0 = 0 \tag{2.28}$$

$$\sum_{k=1}^{K} \sum_{i=0;i\neq j}^{N} x_{ij}^k (a_i + w_i + s_i + t_{ij}) \leq a_j \qquad \forall j \in \{1,2,\ldots,N\} \tag{2.29}$$

$$e_i \leq (a_i + w_i) \leq l_i \qquad \forall i \in \{1,2,\ldots,N\} \tag{2.30}$$

Constraint (2.28) assumes that vehicles are ready and loaded by the time the depot opens, which is indicated as time 0 (zero). Constraint (2.29) calculates the actual arrival time, while (2.30) ensures that each customer i is serviced within its time window.

When both an earliest and latest allowed arrival is stipulated, the time window is referred to as *double sided*. If no arrivals are allowed outside of the given parameters, the time window is said to be *hard*, as is the case in Figure 2.1. When delivery is allowed outside the specified time window, the time window is said to be *soft*, and customer i may penalize lateness at a cost of α_i (Koskosidis et al., 1992). Customer i may specify a maximum lateness, L_i^{max}. The example illustrated in Figure 2.2 sees customer i specifying a time window between 07:30 and 15:30. The customer will, however, allow late deliveries until 17:00. A hard time window

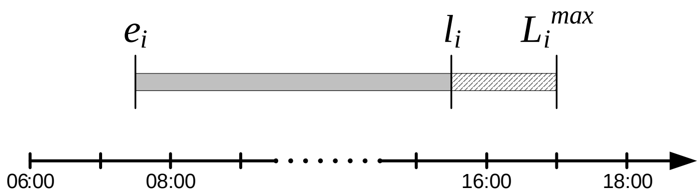


Figure 2.2: Soft time window

is therefor a special type of soft time window where $L_i^{max} = 0$. Should a vehicle arrive after the latest allowed arrival time, l_i, but prior to the maximum lateness, L_i^{max}, the lateness at node i, L_i, can be calculated as $L_i = \max\{0, a_i - l_i\} | a_i \leq L_i^{max}$. The lateness is penalized by introducing a penalty term to the VRP objective function (2.19), resulting in(2.31).

$$\min z = \sum_{i=0}^{N} \sum_{j=0,j\neq i}^{N} \sum_{k=1}^{K} c_{ij} x_{ij}^k + \sum_{i=1}^{N} \alpha_i \times \max\{0, L_i\} \tag{2.31}$$

The time window for the depot, node 0, can be specified. The case illustrated in Figure 2.3 sees the depot specifying operating hours (time window) from 06:00 to 18:00, while the first customer on the route, customer 1, specifies a time window between 07:00 and 09:00, and the last customer, customer N, requests delivery between 15:00 and 17:00.

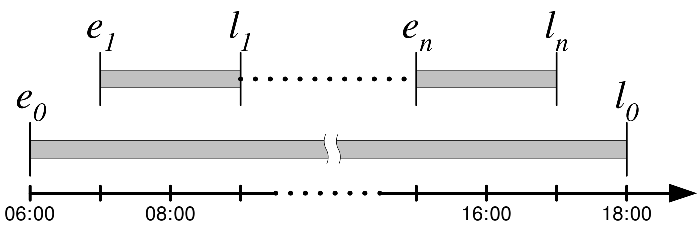


Figure 2.3: Time window for the depot, node 0

Should a customer specify multiple time windows, an indexing symbol, a, is introduced as superscript to the earliest and latest allowed arrival times, respectively, where $a \in \{1, 2, \ldots, A\}$ in which A indicates the maximum number of time windows allowed for each customer. Consider the example where customer n requests delivery either between 06:30 and 09:00, or between 16:00 and 17:30 as illustrated in Figure 2.4. This example is

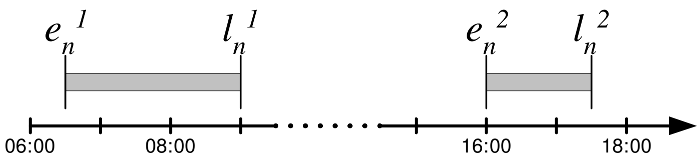


Figure 2.4: Multiple time windows

typical of residents requesting home shopping deliveries outside business hours. The formulation changes with the introduction of the decision variable

$$\psi_i^a \triangleq \begin{cases} 1 & \text{if the } a^{\text{th}} \text{ time window of customer } i \text{ is used, where } i \in \{1, 2, \ldots, N\}, \\ & a \in \{1, 2, \ldots, A\} \\ 0 & \text{otherwise.} \end{cases}$$

To ensure that the decision variable is appropriately enforced in the formulation, we change constraint (2.30) to distinguish between different time windows, as proposed in (2.32)

$$e_i^a - (1 - \psi_i^a)\, M \leq (a_i + w_i) \leq l_i^a + (1 - \psi_i^a)\, M \quad \forall i \in \{1, 2, \ldots, n\}, a \in \{1, 2, \ldots, A\} \tag{2.32}$$

where M is a sufficiently large number, typically greater than the scheduling horizon. An enforcement of a single time window for each customer is required, and is subsequently introduced in (2.33).

$$\sum_{a=1}^{A} \psi_i^a = 1 \qquad \forall i \in \{1, 2, \ldots, N\} \tag{2.33}$$

2.2.2 Capacity constraints and vehicle characteristics

Gendreau et al. (1999) propose a solution methodology for cases where the fleet is heterogeneous, that is, where the fleet is composed of vehicles with different capacities and costs. Their objective is to determine what the optimal fleet composition should be, and is referred to as either a Heterogeneous Fleet Vehicle Routing Problem (HVRP) or a Fleet Size and Mix Vehicle Routing Problem (FSMVRP). Liu and Shen (1999b) adds time windows in their problem application and refer to the problem as a Fleet Size and Mix Vehicle Routing Problem with Time Windows (FSMVRPTW). In yet another paper, Liu and Shen (1999a) refers to the heterogeneous fleet variant as the Vehicle Routing Problem with Multiple Vehicle Types and Time Windows (VRPMVTTW). Taillard (1999) formulates the Vehicle Routing Problem with a Heterogeneous fleet of vehicles (VRPHE) where the number of vehicles of type t in the fleet is limited; the objective being to optimize the utilization of the given fleet. Salhi and Rand (1993) incorporate vehicle routing into the vehicle composition problem, and refer to it as the Vehicle Fleet Mix problem (VFM).

The implication of a heterogeneous fleet on the standard VRP is that T type of vehicles are introduced, with $t \in \{1, 2, \ldots, T\}$. The vehicle capacity parameter p is changed. The new parameter, p_t, represents the capacity of vehicles of type t, resulting in each vehicle k having a unique capacity, p_k. The use of one vehicle of type t implies a fixed cost f_t. A unique fixed cost, f_k, is introduced for each vehicle k, based on its vehicle type. The objective function changes to

$$\min z = \sum_{i=0}^{n} \sum_{\substack{j=0 \\ j \neq i}}^{n} \sum_{k=1}^{K} c_{ij} x_{ij}^{k} + \sum_{k=1}^{K} \sum_{j=1}^{n} f_k x_{0j}^{k} \tag{2.34}$$

while (2.23) changes to indicate the new capacity parameter

$$\sum_{i=1}^{n} q_i \left(\sum_{j=0}^{n} x_{ij}^{k} \right) \leq p_k \qquad \forall k = \{1, 2, \ldots, K\} \tag{2.35}$$

Taillard (1999) introduces a variable c_{ijt} to represent the cost of traveling between nodes i and j, using a vehicle of type t. It is possible to introduce the variable portion of the vehicle cost into the objective function proposed in (2.34). The introduction will lead to (2.36)

$$\min \sum_{i=0}^{n} \sum_{\substack{j=0 \\ j \neq i}}^{n} \sum_{k=1}^{K} \sum_{t=1}^{T} c_{ijt} x_{ij}^{k} \xi_t^{k} + \sum_{k=1}^{K} \sum_{j=1}^{n} f_k x_{0j}^{k} \tag{2.36}$$

where

$$\xi_t^k \triangleq \begin{cases} 1 & \text{if vehicle } k \text{ is of type } t \text{, where } k = \{1,2,\ldots,K\} \text{, and } t = \{1,2,\ldots,T\} \\ 0 & \text{otherwise} \end{cases}$$

2.2.3 Uncertainty in vehicle routing

The statements in Section 2.1.3 do not adequately describe a variety of practical VRP situations where one or several parameters are uncertain. Powell (2003) confirms that research into routing and scheduling algorithms, which explicitly captures the uncertainty of future decisions made now, is extremely young. Laporte et al. (1992), Lambert et al. (1993), and Ong et al. (1997) provide examples including vehicles collecting random quantities at various customers; and customers being visited on a random basis. A vehicle incurs a penalty proportional to the duration of its route in excess of a predetermined constant B — typical of applications where drivers are paid overtime for work done after normal hours. Laporte et al. (1992) propose an attractive and relatively simple chance constrained model (from a computational point of view). However, as the expected cost related to excess route duration needs to be taken into account, this thesis reverts to proposing a stochastic programming model with recourse.

First stage decisions made are the number of vehicles required, as well as their respective routes. Once the random travel time and service time variables are realized in the second stage, penalties are incurred for the excess duration. The following variables are defined.

$$x_{ij}^k \triangleq \begin{cases} 1 & \text{if vehicle } k \text{ travels from node } i \text{ to } j \text{, where} \\ & i,j = \{1,2,\ldots,n\} | i \neq j \text{, and } k = \{1,2,\ldots,K\} \\ 0 & \text{otherwise} \end{cases}$$

$$z_i^k \triangleq \begin{cases} 1 & \text{if node } i \text{ is visited by vehicle } k \text{, where } i = \{1,\ldots,n\}, k = \{1,\ldots,m\} \\ 0 & \text{otherwise} \end{cases}$$

$\tilde{\xi} \triangleq$ a vector of random variables corresponding to travel and service times. Each realization r of $\tilde{\xi}$, denoted by ξ^r, is referred to as a *state of the world* (Kall and Wallace, 1994)

$\Xi \triangleq$ the finite support of $\tilde{\xi}$ such that $\Xi = \left\{1,2,\ldots,\xi^r,\ldots,\xi^R\right\}$ where R is the total number of states in the problem *world*

$y^k(\tilde{\xi}) \triangleq$ the excess duration of route k as a function of the realization of $\tilde{\xi}$

$c_{ij}^k \triangleq$ the travel cost from node i to j with vehicle k, where $i, j = \{1, \ldots, n\}, k = \{1, \ldots, K\}$

$t_{ij}^k(\tilde{\xi}) \triangleq$ the travel time from node i to j with vehicle k, where $i, j = \{1, \ldots, n\}, k = \{1, \ldots, K\}$ expressed as a function of the realization of $\tilde{\xi}$

$\tau_i^k(\tilde{\xi}) \triangleq$ the service time at node i with vehicle k, where $i = \{1, \ldots, n\}, k = \{1, \ldots, K\}$, expressed as function of the realization of $\tilde{\xi}$

$\beta^k \triangleq$ the positive unit penalty cost for excess duration traveled by vehicle k, where $k = \{1, \ldots, m\}$

$f^k \triangleq$ the fixed cost of vehicle k, where $k = \{1, \ldots, K\}$

$B^k \triangleq$ the maximum time for route k over which a penalty is incurred, where $k = \{1, \ldots, K\}$

The model is then

$$\min z = \sum_{k=1}^{K} f^k z_0^k + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{k=1}^{K} c_{ij}^k x_{ij}^k + E_{\tilde{\xi}} \left(\sum_{k=1}^{K} \beta^k y^k(\tilde{\xi}) \right) \tag{2.37}$$

subject to

$$\sum_{k=1}^{K} z_i^k = 1 \qquad \forall i \in \{1, \ldots, n\} \tag{2.38}$$

$$\sum_{j=1}^{n} \left(x_{0j}^k + x_{j0}^k \right) = 2z_0^k \qquad k \in \{1, \ldots, K\} \tag{2.39}$$

$$\sum_{j=1}^{n} \left(x_{ij}^k + x_{ji}^k \right) = 2z_i^k \qquad \forall i \in \{1, \ldots, n\}, k \in \{1, \ldots, K\} \tag{2.40}$$

$$\sum_{i \in S} \sum_{\substack{j \in S \\ j \neq i}} x_{ij}^k \leq |S| - 1 \qquad S \subset V, 3 \leq |S| \leq n-3, k = \{1, \ldots, K\} \tag{2.41}$$

$$B^k - \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} t_{ij}^k(\tilde{\xi}) x_{ij}^k - \frac{1}{2} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(\tau_i^k(\tilde{\xi}) + \tau_j^k(\tilde{\xi}) \right) x_{ij}^k + y^k(\tilde{\xi}) \geq 0 \tag{2.42}$$

$$\forall k \in \{1, \ldots, K\}, \tilde{\xi} \in \Xi$$

$$x_{ij}^k \in \{0, 1\} \qquad \forall i, j \in \{1, \ldots, n\}, k \in \{1, \ldots, K\} \tag{2.43}$$

$$z_i^k \in \{0, 1\} \qquad \forall i \in \{1, \ldots, n\}, k \in \{1, \ldots, K\} \tag{2.44}$$

$$y^k\left(\tilde{\xi}\right) \geq 0 \qquad \forall k \in \{1, \ldots, K\}, \tilde{\xi} \in \Xi \tag{2.45}$$

The objective function minimizes total cost in (2.37) that includes fixed vehicle costs, travel costs, as well as the expected penalty costs as a result of exceeded route duration. All vehicles must be routed according to (2.38), while (2.39) calculates the number of routed vehicles. Degree constraints are introduced in (2.40). Subtours are eliminated through (2.41) where the reader may infer that $n > 6$. Constraint (2.42) combined with (2.45) implies a penalty to be calculated for vehicle k, but only if the total route length including service times exceed B^k.

2.2.4 Time-dependent travel time

Although unpredictable events such as accidents and vehicle breakdowns render travel times as stochastic, the candidate postulates that the subtle, yet partially predictable event of congestion during peak hours of the day requires more attention. The assumption is made that by addressing the time-dependent nature of travel times, a modeling approach that is a stronger approximation of the actual real-world conditions of vehicle routing and scheduling than by catering for stochastic travel times, will be achieved.

Hill and Benton (1992) review the two main approaches in estimating travel distance between two nodes i and j, denoted by d_{ij}, namely Minkowski distance and Pythagorean distance. The former is presented in (2.46).

$$d_{ij} = [|x_i - x_j|^\omega + |y_i - y_j|^\omega]^{\frac{1}{\omega}} \tag{2.46}$$

When ω is 2, the Minowski distance, denoted by d_{ij}, is the Pythagorean distance. When ω is 1, the Minowski distance is the *city-block* right-angled distance. In (2.46) the coordinate pair (x_i, y_i) of each node i is required. A similar approach can be followed if only latitude and longitude data is available, i.e. from a Geographical Information System (GIS) database. The problem, however, is that researchers often reduce vehicle travel speed to an approximate speed, denoted by r_c, and simply apply the scalar transformation of distance in (2.47) to find the travel time between the two nodes,

$$t_{ij} = \frac{d_{ij}}{r_c} \tag{2.47}$$

without cognisance of an acceleration stage to get onto the road, the cruising stage, and the deceleration stage at the destination node (Assad, 1988). If the three stages were to be acknowledged, d_c denotes the distance required for the vehicle to reach its cruising speed, and α denotes the acceleration, a more appropriate way of calculating the travel time is given in (2.48).

$$t_{ij} = \begin{cases} 2\left(\frac{d_{ij}}{\alpha}\right)^{\frac{1}{2}} & \text{if } d_{ij} \leq 2d_c \\ \frac{d_{ij}}{r_c} + \frac{r_c}{\alpha} & \text{if } d_{ij} > 2d_c \end{cases} \tag{2.48}$$

In most metropolitan areas, travel times are much longer during the start and end of workday *rush hours*, especially on main arterial routes. If one were to inflate all route times equally during peak periods, one would be able to route and schedule vehicles without taking time-dependent travel times into consideration, and not compromise optimality of routes. However, road networks are unevenly congested, i.e. traveling from A to B during the morning rush hour traffic might be more congested than when traveling from B to A at the same time.

Malandraki and Daskin (1992) state that the travel time is not only a function of the distance, but should take the time of day into account as well. Ichoua et al. (2003) state that research on time-dependent problems started towards the end of the 1950s with references to the *time-dependent shortest path problem*, the *time-dependent path choice problem*, and the Time Dependent Traveling Salesman Problem (TDTSP). Of the earliest research found on the Time Dependent Vehicle Routing Problem (TDVRP) is Hill et al. (1988), followed by Hill and Benton (1992). In their papers customer *nodes* were assigned time-dependent piecewise constant speeds — these speeds reflect the traveling speed *surrounding* the nodes. The edge travel time between two nodes were derived as the average speed of the two nodes concerned. At the time Hill and Benton (1992) attribute the lack of time-dependent travel time research to:

- Immense efforts to estimate travel time parameters
- Prohibitive data storage requirements
- Inefficient solution algorithms

Malandraki and Daskin (1992) formulate an elegant variant of the Vehicle Routing Problem with Time Windows (VRPTW) with the introduction of piecewise constant travel times on the edges. Approaches to accommodate time-dependent travel times mentioned so far all allow *passing*: the event where one vehicle my *pass* another vehicle on the same edge although it started later than the vehicle it passed, but in a different time period with shorter traveling time.

Ahn and Shin (1991) use similar notation as used in the introduction of the VRPTW, and also introduce:

$\tau_{ij}(x) \triangleq$ travel time from node i to node j via arc $(i,j) \in A$, given that the trip starts from node i at time x

$s_i \triangleq$ the constant service time at node i

$t_i \triangleq$ the time at which service begins at node i

$A_{ij}(t_i) \triangleq$ arrival time at node j through arc $(i,j) \in A$ given t_i, that is $A_{ij}(y) = y + s_i + \tau_{ij}(t_i + s_i)$

$d_i \triangleq$ the *effective* latest service start time at node i that allows us to maintain the feasibility of a current route

Each customer i is to be serviced within its time window $[e_i, l_i]$. The internode travel time $\tau_{ij}(\cdot)$ and the arrival time $A_{ij}(\cdot)$ are functions of the departure time representing time-dependent congestion levels. In this thesis multiple links are not considered. The non-passing property can be expressed as:

> *For any two nodes i and j, and any two service start times x and y at node i such that $x < y$, $A_{ij}(x) < A_{ij}(y)$ must hold, that is, earlier departure from node i guarantees earlier arrival at node j.*

Raw travel time data in the form of a step function is not appropriate for use in the routing of vehicles, as it only provides average travel time data for specific time periods. In such data sources, let:

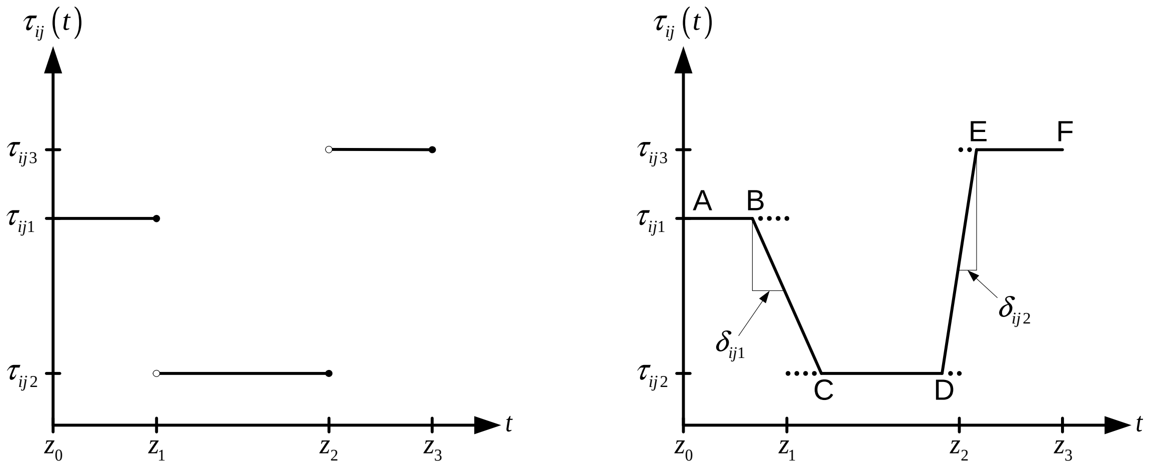
$\tau_{ijk} \triangleq$ the shortest travel time from node i to node j if the start time at node i is in time slot Z_k, where $i, j \in A$, and $k \in \{1, 2, \ldots, K\}$,

where the day (planning horizon) is divided into time slots such that

$$Z_k = [z_{k-1}, z_k] \qquad \forall k \in \{1, 2, \ldots, K\},$$

where the interval $[z_0, z_K]$ reflects the full day, or planning horizon under consideration. Figure 2.5 is used for illustrative purposes. The travel time, being a function of the time of day, is not continuous in the point z_k and may lead to passing if travel time decrease for the $k + 1^{\text{th}}$ segment. To obtain a smoothed travel time function, let:

$\tau_{ij}(t) \triangleq$ the travel time from node i to node j given that the travel started at time t from node i



(a) Step function (b) Smoothed function (Fleischmann et al., 2004)

Figure 2.5: Travel time function

A parameter δ_{ijk} is introduced for each breakpoint z_k, where $k \in \{1, 2, \ldots, K\}$, between two consecutive time slots Z_{k-1} and Z_k. The values of $\delta_{ij0} = \delta_{ijK} = 0$. The jump between two consecutive travel times segments Z_{k-1} and Z_k is linearized in the interval $[z_k - \delta_{ijk}, z_k + \delta_{ijk}]$ provided the parameter δ_{ijk} and determining the slope

$$s_{ijk} = \frac{\tau_{ij,k+1} - \tau_{ijk}}{2\delta_{ijk}} \tag{2.49}$$

The travel time function, as illustrated by Figure 2.5(b), is expressed as

$$\tau_{ij}(t) = \begin{cases} \tau_{ijk} & \text{for} \quad z_{k-1} + \delta_{ij,k-1} \leq t \leq z_k - \delta_{ijk} \\ \tau_{ijk} + (t - z_k + \delta_{ijk})\, s_{ijk} & \text{for} \quad z_k - \delta_{ijk} < t < z_k + \delta_{ijk} \end{cases} \tag{2.50}$$

The travel time function holds for all $k \in \{1, 2, \ldots, K\}$. Fleischmann et al. (2004) prove that if $\delta_{ijk} > 0$ for all intermediate breakpoints and the slope $s_{ijk} > -1$, that the arrival time function

$$A_{ij}(t) = t + \tau_{ij}(t) \tag{2.51}$$

is continuous and monotonic[1], i.e. adheres to the non-passing property. The papers by Ichoua et al. (2003) and Potvin et al. (2006) also refer to the non-passing property as the First-In-First-Out (FIFO) property. As $A_{ij}(\cdot)$ is a strictly increasing function, it possesses

[1]There is a designated sequence such that successive members are either consistently increasing or decreasing with no oscillation in relative value, i.e. each member of a monotone increasing sequence is greater than or equal to the preceding member; each member of a monotone decreasing sequence is less than or equal to the preceding member.

the inverse function $A_{ij}^{-1}(\cdot)$. $A_{ij}^{-1}(x)$ is interpreted as the departure time at node i so that node j can be reached at time x. Let $(i_0, i_1, i_2, \ldots, i_m, i_0)$ denote a partially constructed feasible route with m customer nodes where i_0 denotes the depot. The partial route could be simplified for illustration purposes to $(0, 1, 2, \ldots, m, 0)$.

In the presence of the non-passing property, the effective latest service start time at node i on the partial feasible route, denoted by d_i, could then be given by the backward recursive relation given in (2.52).

$$d_i = \begin{cases} \min\left\{l_i, A_{i0}^{-1}(l_0)\right\} & \text{for} \quad i = m \\ \min\left\{l_i, A_{i,i+1}^{-1}(d_{i+1})\right\} & \text{for} \quad 0 \leq i \leq m-1 \end{cases} \tag{2.52}$$

The actual service start time for each node i can be determined by the forward recursion given in (2.53).

$$t_i = \begin{cases} \max\left\{e_i, A_{01}(t_0)\right\} & \text{for} \quad i = 1 \\ \max\left\{e_i, A_{i-1,i}(t_{i-1})\right\} & \text{for} \quad 2 \leq i \leq m \end{cases} \tag{2.53}$$

The computation of both d_i and t_i is fairly elementary. The advantage is only apparent when route improvements are made, and subsequent feasibility check routines are eased.

The formulation used in this thesis refers to both travel and service times as *uncertain* and dependent on the realization of uncertain events. A principle distinction, however, is made between *stochastic* service times and *time-dependent* travel times. The implications of such a distinction will become apparent in the calculations and feasibility checks when solution algorithms are developed in later chapters, as only time-dependent travel time is considered. In the majority of applications, demand is assumed to be known at the time of establishing the actual route.

2.2.5 Multiple scheduling

It is often not viable to assume that each vehicle will only complete a single route. *Multiple scheduling* is concerned with the case where a vehicle could complete deliveries on a scheduled route, return to the depot where its capacity is renewed, after which a second, or consecutive trip is executed with the renewed capacity. Taillard et al. (1996) refer to this type of problem as the Vehicle Routing Problem with Multiple use of vehicles (VRPM). Butt and Ryan (1999) consider the Multiple Tour Maximum Collection Problem (MTMCP) and assumes that the routes are constrained in such a way that all of the customers cannot be visited. Their approach aims to maximize the number of customers serviced. Brandão and Mercer (1997) introduce the Multi-Trip Vehicle Routing Problem (MTVRP) and address the combination of multiple trips with time windows. The special case of multiple scheduling where only trips are considered is referred to as *Double Scheduling*.

This thesis considers a vehicle that starts and ends its tour at the depot. A *tour* consists of one or more *routes*, each starting and ending at the depot. The same vehicle can only be used for two or more routes if the routes do not overlap. As opposed to (2.28) multiple routes require a service time to be specified for the depot. Consider the example illustrated in Figure 2.6. The depot has a time window from 06:00 to 18:00. A vehicle fills its capacity

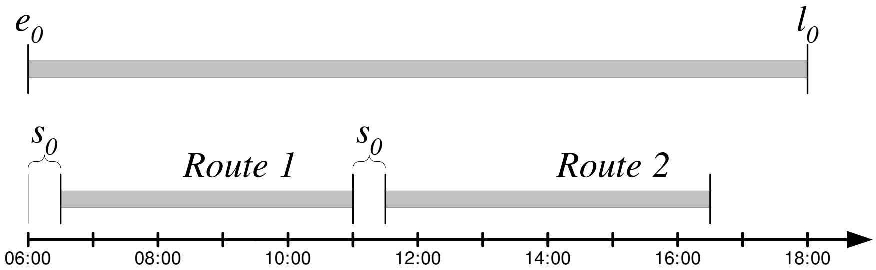


Figure 2.6: Double scheduling

at the depot for a time period of $s_0 = 0.5$ hours. It leaves the depot at 06:30, services the first route, and returns to the depot at 11:00, where its capacity is renewed. A second route, of five hours, is serviced before the vehicle returns to the depot.

Taillard et al. (1996) state that the multiple scheduling type of problem has received very little attention in literature. This thesis proposes a way to deal with multiple routes. The proposed solution involves a time verification process. If a vehicle arrives back at the depot at time a_m, and the service time is specified as s_0, then the vehicle is considered for an additional route on its current tour if, after the capacity has been renewed, the depot's time window is still open. The case is presented in (2.54).

$$a_m + s_0 \leq l_0 \tag{2.54}$$

The mathematical formulation of the VRPM requires a redefinition of the decision variables, as well as the constraints. The VRPM is addressed in the next section where the complete problem is defined and formulated.

2.3 The integrated problem at hand

An extended variant of the VRP, where multiple soft time windows, a heterogeneous fleet, and multiple scheduling are considered in an environment with uncertain travel and service times, is presented. Due to the complexity associated when concatenating elements from various variant acronyms, we revert to using a simple reference, the Thesis Problem (TP). To formulate the complex problem, we will redefine some of the variables and parameters used earlier, and introduce a few additional variables. We define the following basic parameters.

$N \triangleq$ total number of customers to be serviced

$q_i \triangleq$ deterministic demand for customer i, where $i = \{1, 2, \ldots, N\}$

$K \triangleq$ total number of vehicles available

$$z_i^k \triangleq \begin{cases} 1 & \text{if node } i \text{ is visited by vehicle } k \text{, where } i = \{1, \ldots, N\}, k = \{1, \ldots, K\} \\ 0 & \text{otherwise} \end{cases}$$

$\tilde{\xi} \triangleq$ a vector of uncertain variables corresponding to travel and service times. Each realization γ of $\tilde{\xi}$, denoted by ξ^γ, is referred to as a *state of the world* (Kall and Wallace, 1994)

$\Xi \triangleq$ the finite support of $\tilde{\xi}$ such that $\Xi = \left\{1, 2, \ldots, \xi^\gamma, \ldots, \xi^\Gamma\right\}$ where Γ is the total number of states in the problem *world*

$t_{ij}^k\left(\tilde{\xi}\right) \triangleq$ the travel time from node i to j with vehicle k, where $i, j = \{1, \ldots, N\}, k = \{1, \ldots, K\}$ expressed as a function of the realization of $\tilde{\xi}$

$\tau_i^k\left(\tilde{\xi}\right) \triangleq$ the service time at node i with vehicle k, where $i = \{1, \ldots, N\}, k = \{1, \ldots, K\}$, expressed as function of the realization of $\tilde{\xi}$

To expand the formulation and to include a heterogeneous fleet, we let:

$T \triangleq$ number of different types of vehicles available

$c_{ijt} \triangleq$ travel cost if a vehicle of type t travels from customer i to customer j, where $t = \{1, 2, \ldots, T\}$, and $i, j = \{0, 1, 2, \ldots, N\}$

$p_t \triangleq$ capacity of a vehicle of type t, where $t = \{1, 2, \ldots, T\}$

$f_t \triangleq$ fixed cost of a vehicle of type t, where $t = \{1, 2, \ldots, T\}$

$$\phi_t^k \triangleq \begin{cases} 1 & \text{if vehicle } k \text{ is of type } t \text{, where } k = \{1, 2, \ldots, K\} \text{, and} \\ & t = \{1, 2, \ldots, T\} \\ 0 & \text{otherwise} \end{cases}$$

Multiple soft windows will be addressed by introducing the following parameters:

$A_i \triangleq$ number of time windows for customer i, where $i = \{0, 1, 2, \ldots, N\}$

$a_i\left(\tilde{\xi}\right) \triangleq$ the actual arrival time at customer i, where $i = \{0, 1, 2, \ldots, N\}$, expressed as a function of the realization of $\tilde{\xi}$

$e_i^a \triangleq$ earliest allowed arrival time for customer i's a^{th} time window, where $i = \{0, 1, 2, \ldots, N\}$ and $a = \{1, 2, \ldots, A_i\}$

$l_i^a \triangleq$ latest allowed arrival time for customer i's a^{th} time window, where $i = \{0, 1, 2, \ldots, N\}$ and $a = \{1, 2, \ldots, A_i\}$

$L_i^{\max} \triangleq$ maximum lateness allowed by customer i, where $i = \{0, 1, 2, \ldots, N\}$

$\alpha_i \triangleq$ lateness penalty at customer i in cost per time unit, where $i = \{0, 1, 2, \ldots, N\}$

$\lambda_i\left(\tilde{\xi}\right) \triangleq$ actual lateness at customer i, where $i = \{0, 1, 2, \ldots, N\}$, expressed as a function of the realization of $\tilde{\xi}$

$w_i\left(\tilde{\xi}\right) \triangleq$ waiting time at customer i, where $i = \{0, 1, 2, \ldots, N\}$, expressed as a function of the realization of $\tilde{\xi}$

To ensure that multiple scheduling is considered, we let:

$R^k \triangleq$ number of routes scheduled for vehicle k, where $k = \{1, 2, \ldots, K\}$

$Q \triangleq$ maximum number for routes allowed for any one vehicle

$M^k \triangleq$ maximum tour time (all routes) allowed for vehicle k, where $k = \{1, 2, \ldots, K\}$

$d^{kr}\left(\tilde{\xi}\right) \triangleq$ vehicle k's departure time from the depot as it embarks on servicing its r^{th} route, where $k = \{1, 2, \ldots, K\}$ and $r = \{1, 2, \ldots, R_k\}$, expressed as a function of the realization of $\tilde{\xi}$

$g^{kr}\left(\tilde{\xi}\right) \triangleq$ vehicle k's return time at the depot after servicing its r^{th} route, where $k = \{1, 2, \ldots, K\}$ and $r = \{1, 2, \ldots, R^k\}$, expressed as a function of the realization of $\tilde{\xi}$

$\delta^k\left(\tilde{\xi}\right) \triangleq$ the amount by which vehicle k exceed its allowable tour time, where $k = \{1, 2, \ldots, K\}$, expressed as a function of the realization of $\tilde{\xi}$

$\beta^k \triangleq$ the positive unit penalty cost for vehicle k when exceeding its allowable tour time, where $k = \{1, \ldots, K\}$

With the notation established the decision variables for the TP are defined as:

$$x_{ij}^{kr} \triangleq \begin{cases} 1 & \text{if vehicle } k \text{ travels from customer } i \text{ to customer } j \text{ on its } r^{\text{th}} \text{ route,} \\ & \text{where } i,j = \{1,2,\ldots,N\}, k = \{1,2,\ldots,K\}, r = \{1,2,\ldots,R_k\} \\ 0 & \text{otherwise} \end{cases}$$

$$\psi_i^a \triangleq \begin{cases} 1 & \text{if the } a^{\text{th}} \text{ time window of customer } i \text{ is used, where } i \in \{1,2,\ldots,N\}, \\ & a \in \{1,2,\ldots,A\} \\ 0 & \text{otherwise.} \end{cases}$$

The mathematical formulation of the TP is provided.

$$\min z = \sum_{i=0}^{N} \sum_{\substack{j=0 \\ j \neq i}}^{N} \sum_{k=1}^{K} \sum_{t=1}^{T} \sum_{r=1}^{R^k} c_{ijt} x_{ij}^{kr} \phi_t^k + \sum_{j=1}^{N} \sum_{k=1}^{K} \sum_{r=1}^{R^k} \frac{f^k x_{0j}^{kr}}{R^k} + E_{\tilde{\xi}} \left[\sum_{i=1}^{N} \alpha_i \lambda_i \left(\tilde{\xi} \right) + \sum_{k=1}^{K} \beta^k \delta^k \left(\tilde{\xi} \right) \right] \quad (2.55)$$

subject to

$$\sum_{j=1}^{N} \sum_{r=1}^{Q} x_{0j}^{kr} = R^k \qquad \forall k \in \{1,2,\ldots,K\} \quad (2.56)$$

$$\sum_{j=1}^{N} \sum_{r=1}^{Q} x_{j0}^{kr} = R^k \qquad \forall k \in \{1,2,\ldots,K\} \quad (2.57)$$

$$\sum_{\substack{i=1 \\ i \neq j}}^{N} \sum_{k=1}^{K} \sum_{r=1}^{R^k} x_{ij}^{kr} = 1 \qquad \forall j \in \{1,2,\ldots,N\} \quad (2.58)$$

$$\sum_{\substack{j=1\\j\neq i}}^{N}\sum_{k=1}^{K}\sum_{r=1}^{R^k} x_{ij}^{kr} = 1 \qquad \forall i \in \{1,2,\ldots,N\} \tag{2.59}$$

$$\sum_{q=1}^{N} q_i \sum_{\substack{j=0\\j\neq i}}^{N} x_{ij}^{kr} \leq p^k \qquad \forall k \in \{1,2,\ldots,K\},\; r = \{1,2,\ldots,R_k\} \tag{2.60}$$

$$e_i^a - (1-\psi_i^a)\,M \leq a_i\left(\tilde{\xi}\right) + w_i\left(\tilde{\xi}\right) \qquad \forall i \in \{1,2,\ldots,N\},\; \forall a \in \{1,2,\ldots,A_i\} \tag{2.61}$$

$$L_i^{\max} + (1-\psi_i^a)\,M \geq a_i\left(\tilde{\xi}\right) + w_i\left(\tilde{\xi}\right) \qquad \forall i \in \{1,2,\ldots,N\},\; \forall a \in \{1,2,\ldots,A_i\} \tag{2.62}$$

$$\sum_{a=1}^{A_i} \psi_i^a = 1 \qquad \forall i \in \{1,2,\ldots,N\} \tag{2.63}$$

$$\max\left\{0, e_j - \left(d^{kr}\left(\tilde{\xi}\right) + t_{0j}\right)\sum_{k=1}^{K}\sum_{r=1}^{R_k} x_{0j}^{kr}\right\} = w_j\left(\tilde{\xi}\right) \qquad \forall j \in \{1,2,\ldots,N\} \tag{2.64}$$

$$\max\left\{0, \left(a_i\left(\tilde{\xi}\right) - l_i^a\right)\right\} = \lambda_i^a\left(\tilde{\xi}\right) \qquad \forall i \in \{1,2,\ldots,N\},\; \forall a \in \{1,2,\ldots,A_i\} \tag{2.65}$$

$$d^{k1} \geq e_0 + s_0 \qquad \forall k \in \{1,2,\ldots,K\} \tag{2.66}$$

$$\sum_{k=1}^{K}\sum_{r=1}^{R^k} x_{0j}^{kr}\left(d^{kr}\left(\tilde{\xi}\right) + t_{0j}\right) \leq a_j\left(\tilde{\xi}\right) \qquad \forall j \in \{1,2,\ldots,N\} \tag{2.67}$$

$$\sum_{\substack{i=1\\i\neq j}}^{N}\sum_{k=1}^{K}\sum_{r=1}^{R^k} x_{ij}^{kr}\left(a_i\left(\tilde{\xi}\right) + w_i\left(\tilde{\xi}\right) + \tau_i^k\left(\tilde{\xi}\right) + t_{ij}^k\left(\tilde{\xi}\right)\right) \leq a_j\left(\tilde{\xi}\right) \qquad \forall j \in \{1,2,\ldots,N\} \tag{2.68}$$

$$\sum_{i=1}^{N} x_{i0}^{kr}\left(a_i\left(\tilde{\xi}\right) + \tau_i^k\left(\tilde{\xi}\right) + w_i\left(\tilde{\xi}\right) + t_{i0}\right) \leq g^{kr}\left(\tilde{\xi}\right) \qquad \forall k \in \{1,2,\ldots,K\},\; r \in \{1,2,\ldots,R^k\} \tag{2.69}$$

$$g^{k,r-1}\left(\tilde{\xi}\right) + s_0 = d^{kr}\left(\tilde{\xi}\right) \qquad \forall k \in \{1,2,\ldots,K\},\; r \in \{2,3,\ldots,R_k\} \tag{2.70}$$

$$g^{kr}\left(\tilde{\xi}\right) + s_0 \leq l_0 \qquad \forall k \in \{1,2,\ldots,K\}, \quad r \in \{2,3,\ldots,R_{k-1}\} \tag{2.71}$$

$$g^{kR^k}\left(\tilde{\xi}\right) \leq M_k + \delta^k\left(\tilde{\xi}\right) \qquad \forall k \in \{1,2,\ldots,K\} \tag{2.72}$$

$$R^k \leq Q \qquad \forall k \in \{1,2,\ldots,K\} \tag{2.73}$$

$$\sum_{r=R^{k+1}}^{Q}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N} x_{ij}^{kr} = 0 \qquad \forall k \in \{1,2,\ldots,K\} \tag{2.74}$$

$$x_{ij}^{kr} \in \{0,1\} \qquad \forall i,j \in \{1,2,\ldots,N\}, \quad k \in \{1,2,\ldots,K\}, \quad r \in \{1,2,\ldots,R^k\} \tag{2.75}$$

$$\psi_i^a \in \{0,1\} \qquad \forall i \in \{1,2,\ldots,N\}, \quad \forall a \in \{1,2,\ldots,A_i\} \tag{2.76}$$

The objective function in (2.55) minimizes a combination of deterministic and stochastic cost components. The first expression represents the total variable traveling cost, followed by the total fixed fleet cost. The third expression represents the expected lateness penalties and constitutes firstly the lateness at each customer, and secondly the lateness for each vehicle.

The combination of (2.56) and (2.57) calculates the total number of routes and ensures that the same number of routes that starts for each vehicle, also finishes. Each customer is visited exactly once according to the constraint combination (2.58) and (2.59). Vehicular capacity is enforced through (2.60) by ensuring that the sum of the demands of all customers assigned to a specific route of a given vehicle do not exceed the vehicle's capacity, which may either by represented as weight or volumetric capacity, or both if additional constraints are added.

Constraints (2.61) and (2.62) ensure that the multiple soft time windows are adhered to where the parameter M represents a sufficiently large number, as discussed when multiple soft time windows were introduced. Actual arrival times and waiting times at any given customer is a function of the stochastic travel and service times of all customers preceding that specific customer, hence the stochastic notation. As each customer is visited only once, (2.63) ensures that only one time window for each customer is considered. The waiting time and lateness at each customer, both expressed as a stochastic variable, are determined in (2.64) and (2.65), respectively.

The departure time for each vehicle's first route is determined by (2.66), while the actual arrival time at the first customer on each route is determined by (2.67). Arrival times for subsequent customers are determined by (2.68).

The return time for each route is determined by (2.69). Consecutive route start times is determined by (2.70) by taking the service time of the depot into account where vehicles' capacities are renewed as proposed in (2.54). Constraint (2.71) enforces all routes to finish within the operating hours of the depot, while (2.72) determines the lateness for each vehicle when exceeding its allowed tour time. Each vehicle may not execute more than a predetermined number of routes as provided for in (2.73). Should it be determined in equations (2.56) and (2.57) that the required number of routes is less than the preset limit Q, then all *allowed* routes not required are eliminated through the introduction of (2.74). Binary decision variables are provided for with the introduction of (2.75) and (2.76).

2.4 Conclusion

This chapter deals with the background of the VRP, as well as the integration of multiple variants into a single problem instance — each contributing to the already complex nature of the problem. Although the model formulation is the first step in describing the problem comprehensively, only very small instances of the problem is currently solvable to optimality.

The following chapter introduces the complexity of the problem at hand, and reviews solution approaches for solving the problem. Exact, heuristic, as well as metaheuristic solution algorithms are considered.