Chapter 2
Introduction to Tour Planning: Vehicle Routing and Related Problems

In the literature, tour planning problems (or routing problems) are a vital research area. As illustrated in Fig. 2.1, the number of journal publications on Vehicle Routing Problems (VRPs, see Sect. 2.3.2) has steadily increased over the years. This increased scientific interest can be explained by several factors. For example, progress in computational resources has opened new possibilities for modeling more complex routing problems such as dynamic routing problems (see Sect. 2.5.3.2). Furthermore, new arising real-world applications provide inspiration for developing new approaches for coordinating complex transportation processes.

In this chapter, a general introduction to tour planning problems as well as aspects of specific types of routing problems are given. Note that this chapter mainly deals with routing problems which comprise characteristics of short-haul routing problems, i.e., routing problems in which transportation activities are performed in a bounded service area without transshipment during a planning horizon which usually has a length of up to one working day (see Sect. 1.2 and cf. Ghiani et al. 2004, p. 247 et seq.). First, the *general task and definitions* of tour planning problems are introduced. Afterwards, the *representation of requests and general types of routing problems*, i.e., node-based and edge-based routing problems, are described. Since the considered RDOPG applications can be modeled as variants of node-based routing problems, different *node-based routing problems* known in the literature and characteristics of these, including solution complexity issues and existing solution methods, are presented. Furthermore, since the considered RDOPG applications can be modeled as a variant of the well-known Vehicle Routing Problem, *extensions to the Vehicle Routing Problem* which can be found in the literature, e.g., specific problem extensions or other objective functions, are presented. Afterwards, different types of *information revelation in routing problems* are discussed by describing appropriate classifications known in the literature. Moreover, a new classification with regard to the type of information revelation and coordination of the transportation process is proposed in order to describe specific attributes which are characteristic of RDOPG applications.

Since RDOPG applications utilize a centralized coordination of the transportation process, characteristics of *dynamic routing problems which utilize a centralized*

F. Ferrucci, *Pro-active Dynamic Vehicle Routing*, Contributions to Management Science,
DOI 10.1007/978-3-642-33472-6_2,

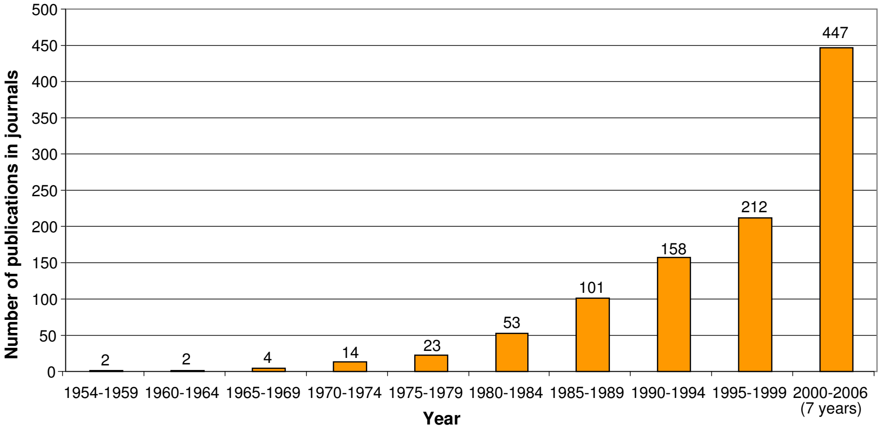


Fig. 2.1 Number of journal publications with regard to VRPs (cf. Eksioglu et al. 2009)

coordination are presented in the following section. In addition to typical objectives considered in dynamic routing problems and a description of different types of possible dynamic events, a measure known in the literature for determining the difficulty of solving dynamic routing problems is presented. Furthermore, dynamic real-world routing applications are described and classified according to the presented measure. Moreover, two measures for evaluating the performance of solution approaches for dynamic routing problems are presented. Finally, a *general classification scheme for routing problems* known in the literature is described and extended according to characteristics which are specific for RDOPG applications.

2.1 General Task and Definitions

In general, the tour planning process can be defined as follows: A set of requests needs to be assigned to a set of resources, for example vehicles that execute the transportation process. Since resources always represent vehicles in the RDOPG applications considered in this book, the terms vehicle and resource are used synonymously. For each vehicle, the assigned requests are required to be ordered in a specific sequence in which the requests will be serviced by the vehicle afterwards. Performing the assignment and sequencing tasks results in a *tour* for each vehicle. Both of these tasks can either be consecutively or simultaneously performed. The step in a tour which consists of traveling from one request to the next is called a *route*. The set of all vehicle tours forms the *tour plan* which is also denoted as the *solution* of the considered routing problem. Note that problem-specific constraints, e.g., time window restrictions of requests, need to be fulfilled by the tour plan in order to obtain a feasible solution. Figure 2.2 illustrates the described tour planning process using five pending requests and two vehicles which start and end their tour at a depot D.

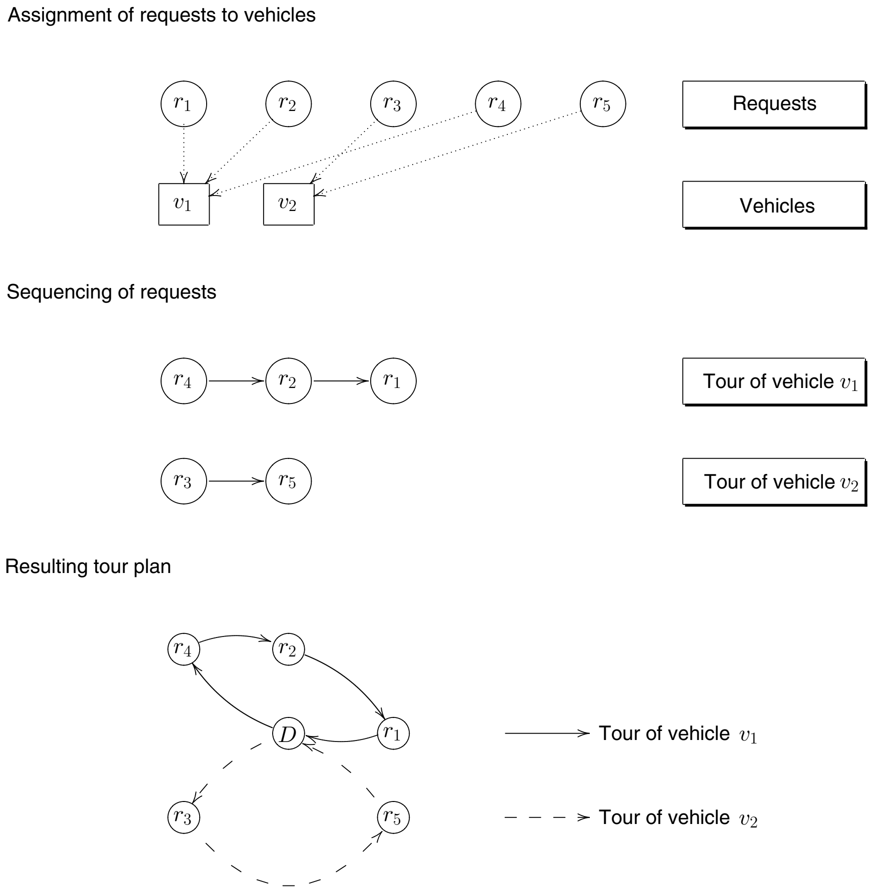


Fig. 2.2 Visualization of the tasks in a routing problem

The generation of a tour plan is connected to a specific *aim* which is pursued, e.g., minimizing the total travel distance. Mathematically, this aim is defined by the *objective function* which assigns a specific *objective function value* to each solution. By using this objective function value, different solutions can be compared and ordered according to their *solution quality*. The goal is to find the *optimal solution*, i.e., the one with the best objective function value among all feasible solutions. In what follows, we assume that the aim is always to minimize a given objective function value, i.e., an optimal solution has the minimum objective function value among all feasible solutions. Since routing problems usually belong to the class of strongly non-deterministic polynomial time hard (NP-hard) combinatorial optimization problems (cf. Lenstra and Rinnooy Kan 1981), it is often not possible to determine an optimal solution in reasonable time due to complexity reasons. Hence, the aim is to determine a good solution whose objective function value is as close as possible to the objective function value of an optimal solution. Details on the complexity of routing problems are described in Sect. 2.3.4. Furthermore, different objectives which are pursued in routing problems are presented in Sect. 2.4.12.

2.2 Representation of Requests and General Types of Routing Problems

In routing problems, requests are represented by *relevant locations* and are modeled in a *graph*. This graph consists of a set of vertices (or nodes) defined by V and of a set of arcs A or edges E. Specifically, if the graph is directed, it is defined by $G = (V, A)$. If it is undirected, it is defined by $G = (V, E)$. In the directed case, direction-dependent travel costs can be assigned so that the costs for traveling from location $i \in V$ to location $j \in V$ represented by c_{ij} can be different from c_{ji} whereas in the undirected case, it holds that $c_{ij} = c_{ji}\ \forall i, j \in V$.

Routing problems can be distinguished according to the type of requests that need to be serviced. Specifically, a distinction is made between *arc-based* routing problems and *node-based* routing problems. In arc-based routing problems, customer requests are represented by arcs or edges, depending on whether the graph is directed or undirected. A characteristic of these types of problems is that the vehicle is at a different position before and after servicing a request. In the Chinese Postman Problem introduced by Kwan (1962), each road is assumed to be undirected so that each edge of G needs to be visited at least once. Moreover, if G also contains directed roads, the problem is called the Mixed Chinese Postman Problem. Another variant is the Rural Postman Problem in which only a given subset of edges $E' \subset E$ needs to be serviced. For more information on arc-based routing problems, see Eiselt et al. (1995) and Golden et al. (2008, p. 37 et seq.). In contrast to arc-based routing problems, node-based routing problems arise if customers are located at vertices $v \in V$ so that vehicles remain at the same position while servicing a customer request. In addition to pure node-based and pure arc-based routing problems, mixed routing problems where requests are located at arcs as well as nodes exist.

Since the RDOPG applications considered in this book belong to the class of node-based routing problems, we consider node-based routing problems in what follows.

2.3 Node-Based Routing Problems

In this section, classic node-based routing problems are introduced in their asymmetric variant so that different costs for both travel directions can be modeled. First, the *Traveling Salesman Problem (TSP)* and the *Vehicle Routing Problem (VRP and CVRP)* are described. Afterwards, routing problems in which goods or persons have to be transported between customer locations are presented. These problems consist of the *General Pickup and Delivery Problem (GPDP) and related variants* known as the Pickup and Delivery Problem and the Dial-A-Ride Problem. Furthermore, the *complexity of node-based routing problems* as well as *solution methods for node-based routing problems* are addressed.

2.3.1 The Traveling Salesman Problem (TSP)

The first routing problem which was considered in the literature is the Traveling Salesman Problem (TSP, cf. Dantzig et al. 1954). In the TSP, one salesman needs to service a given set of $\mathcal{R}$ customer requests $R = \{1, \dots, \mathcal{R}\}$, also denoted as cities. In this routing problem, the set of requests is equal to the set of relevant locations V in the digraph $G = (V, A)$ which are defined by $V = \{1, \dots, \mathcal{R}\}$. The salesman is originally located at one of the customer requests and needs to service the requests by visiting each customer location exactly once and finally returning to its origin. Traveling from i to j, $(i, j) \in A = V \times V$ produces costs denoted by c_{ij}. For example, these costs can represent travel distances (see Sect. 2.4.12). The sought optimal solution is a cost-minimal tour plan that represents the sequence in which the requests are to be visited. Specifically, the solution is given by an assignment of the binary variables $X = \{x_{ij}\}$ $(i, j \in V,\ i \neq j)$ which are 1 if the salesman directly travels from location i to j and 0 otherwise. The TSP is constituted by the following mathematical model (cf. Laporte 1992b):

$$\min z = \sum_{i \in V} \sum_{j \in V} c_{ij} \cdot x_{ij} \quad \text{s.t.} \tag{2.1}$$

$$\forall i \in V\colon \quad \sum_{j \in V} x_{ij} = 1 \tag{2.2}$$

$$\forall i \in V\colon \quad \sum_{j \in V} x_{ji} = 1 \tag{2.3}$$

$$\forall S \subset V,\ 2 \leq |S| \leq \mathcal{R} - 2\colon \quad \sum_{i \in S} \sum_{j \in \{V \setminus S\}} x_{ij} \geq 1 \tag{2.4}$$

The objective function value of a solution is determined by the objective function (2.1). Constraints (2.2) and (2.3) ensure that each location is visited and left exactly once. Constraints (2.4) eliminate subtours in the solution (see Fig. 2.3). Note that this subtour formulation results in $2^{\mathcal{R}} - 2\mathcal{R} - 2$ subtour elimination constraints which cannot be reasonably handled by using Mixed Integer Programming (MIP) solvers even for small problem instances (cf. Laporte 1992b). Besides this subtour elimination formulation, other subtour elimination constraint formulations are possible. For example, Miller et al. (1960) propose a formulation which only contains $O(\mathcal{R}^2)$ restrictions but at the expense of additional variables. Moreover, it may produce weaker lower bounds than the previous formulation (cf. Desrochers and Laporte 1991). Other subtour elimination formulations can be found for example in Laporte (1992b) and in Domschke and Scholl (2010, p. 100).

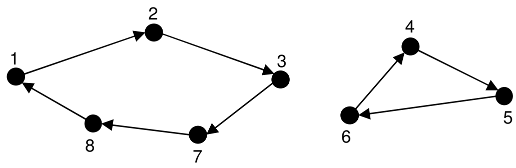


Fig. 2.3 Infeasible TSP solution due to a subtour in the solution

2.3.2 The Vehicle Routing Problem (VRP and CVRP)

The Vehicle Routing Problem (VRP) was introduced by Dantzig and Ramser (1959). In its basic version, a set of requests has to be serviced by a fleet of homogeneous vehicles. Each vehicle begins and ends its tour at the same location called the *depot*. The VRP is used to model two different types of real-world business cases:

1. Goods that are loaded onto the vehicles at the depot are delivered to the customers.
2. Goods that are located at customer locations are collected and transported to the depot.

The uncapacitated VRP in which capacity constraints of the vehicles can be neglected is equal to a multiple-salesman variant of the previously introduced TSP (m-TSP, cf. Laporte and Nobert 1980 and Bektas 2006). In the VRP, $R = \{1, \dots, \mathcal{R}\}$ defines the set of $\mathcal{R}$ customer requests. The set of nodes is defined by $V = R \cup \{0\}$, i.e., customer requests and the depot denoted by node 0. Furthermore, the set K represents the set of $\mathcal{K}$ uncapacitated vehicles $K = \{1, \dots, \mathcal{K}\}$. As is the case in the TSP, travel costs on the arcs are defined by c_{ij} with $i, j \in V$. Since each request location is visited exactly once by a vehicle and since the utilized vehicles are homogeneous, it is not required to specify which tour belongs to which vehicle in the solution. Hence, it is possible to model the VRP using only a *two-index vehicle flow formulation* that uses $O(\mathcal{R}^2)$ binary variables as is the case in the TSP (cf. Toth and Vigo 2002, p. 30). In this formulation, the solution consists of a tour plan represented by a set of binary variables $X = \{x_{ij}\}$ $(i, j \in V,\ i \neq j)$ equal to 1 if and only if a vehicle directly travels from location i to location j. The VRP can then be defined by the following mathematical model (cf. Laporte and Nobert 1980 and Bektas 2006):

$$\min z = \sum_{i \in V} \sum_{j \in V} c_{ij} \cdot x_{ij} \quad \text{s.t.} \tag{2.5}$$

$$\forall i \in R: \quad \sum_{j \in \{V \setminus i\}} x_{ij} = 1 \tag{2.6}$$

$$\forall i \in R: \quad \sum_{j \in \{V \setminus i\}} x_{ji} = 1 \tag{2.7}$$

$$\sum_{i \in R} x_{0i} = \mathcal{K} \tag{2.8}$$

$$\sum_{i \in R} x_{i0} = \mathcal{K} \tag{2.9}$$

$$\forall S \subset R,\ 2 \leq |S| \leq \mathcal{R} - 2\colon \quad \sum_{i \in S} \sum_{j \in \{V \setminus S\}} x_{ij} \geq 1 \tag{2.10}$$

The objective function is given by (2.5). Constraints (2.6) and (2.7) ensure that each customer request is visited and left exactly once. These constraints also determine that exactly one vehicle services each request. Constraints (2.8) and (2.9) make sure that each vehicle leaves the depot and returns to it exactly once. Finally, constraints (2.10) eliminate subtours in the vehicle tours.

The Capacitated VRP The Capacitated Vehicle Routing Problem (CVRP) is closely related to the uncapacitated VRP. In addition, each request $i \in R$ has a demand d_i and all vehicles have the same loading capacity denoted by C. The CVRP can be modeled using the previously introduced formulation of the VRP and replacing constraints (2.10) by

$$\forall S \subset R,\ 2 \leq |S| \leq \mathcal{R} - 2\colon \quad \sum_{i \in S} \sum_{j \in \{V \setminus S\}} x_{ij} \geq r(S) \tag{2.11}$$

Constraints (2.11) are known as *capacity-cut constraints* (CCCs) and they ensure both the elimination of subtours as well as vehicle capacity limits. The value $r(S)$ defines the minimum number of vehicles required to service all requests of set S. It can be calculated by solving a bin-packing problem for determining the minimum number of required bins, each with capacity C, for loading the items of set S (cf. Toth and Vigo 2002, p. 7) or by using a trivial lower bound given by $r(S) = \lceil \frac{\sum_{i \in S} d_i}{C} \rceil$. Note that setting $C = \infty$ leads to constraints (2.10).

Despite requiring only the same amount of binary variables as in a TSP, the described two-indexed formulation has certain drawbacks. Specifically, since no tour-specific attributes can be integrated, e.g., individual vehicle arrival times at customer requests, it does not allow to model more complex variants of the VRP. In order to overcome this limitation, a *three-index vehicle flow formulation* is utilized (cf. Toth and Vigo 2002, p. 15). In this formulation, let $X = \{x_{ijk}\}$ be the set of binary variables that are 1 if and only if vehicle $k \in K$ directly travels from location i to location j $(i, j \in V,\ i \neq j)$. Furthermore, each binary variable of the set $Y = \{y_{ik}\}$ is 1 if and only if vehicle $k \in K$ visits location $i \in V$. Using this notation, the CVRP can be modeled as follows:

$$\min z = \sum_{i \in V} \sum_{j \in V} \sum_{k \in K} c_{ij} \cdot x_{ijk} \quad \text{s.t.} \tag{2.12}$$

$$\forall i \in R\colon \quad \sum_{k \in K} y_{ik} = 1 \tag{2.13}$$

$$\sum_{k \in K} y_{0k} = \mathcal{K} \tag{2.14}$$

$$\forall k \in K, i \in V: \quad \sum_{j \in V} x_{ijk} = \sum_{j \in V} x_{jik} = y_{ik} \tag{2.15}$$

$$\forall k \in K: \quad \sum_{i \in V} d_i \cdot y_{ik} \le C \tag{2.16}$$

$$\forall S \subseteq R, h \in S, k \in K: \quad \sum_{i \in S} \sum_{j \notin S} x_{ijk} \ge y_{hk} \tag{2.17}$$

Besides the objective function defined in (2.12), constraints (2.13) ensure that each request is serviced by exactly one vehicle while constraints (2.14) make sure that the depot belongs to the tour of each vehicle. Constraints (2.15) ensure the connectedness of the vehicle tours and by constraints (2.16) vehicle capacity restrictions are taken into account. Subtours are eliminated by constraints (2.17).

2.3.3 The General Pickup and Delivery Problem (GPDP) and Related Variants

In this section, we describe the General Pickup and Delivery Problem as well as two variants denoted as the Pickup and Delivery Problem and the Dial-A-Ride Problem. In contrast to the previously introduced routing problems, these problems consider applications in which goods or passengers are picked up at specific locations and need to be transported to other locations on the same tour. Hence, the vehicles are empty at the start and at the end of their tour.

2.3.3.1 The General Pickup and Delivery Problem

The General Pickup and Delivery Problem (GPDP) was described by Savelsbergh and Sol (1995). In the GPDP, each customer request $i \in R$ consists of a total cargo of weight $\bar{q}_i$ which needs to be transported from a set of pickup locations R_i^+ to a set of delivery locations R_i^- which are individual for each customer request. Each location $j \in \{R_i^+ \cup R_i^-\}$ has a weight q_j assigned. Positive weights assigned to pickup locations indicate the weight of goods that have to be picked up whereas negative weights at delivery locations denote the quantity of goods that need to be delivered. Moreover, it holds that $\bar{q}_i = \sum_{j \in R_i^+} q_j = -\sum_{j \in R_i^-} q_j$. The set $R^+ = \bigcup_{i \in R} R_i^+$ comprises all pickup locations and set $R^- = \bigcup_{i \in R} R_i^-$ is composed of all delivery locations of all requests. Additionally, set $L = R^+ \cup R^-$ unifies all request locations. Requests are serviced by a *heterogeneous* vehicle fleet $K = \{1, \ldots, \mathcal{K}\}$. Specifically, each vehicle $k \in K$ has an individual capacity Q_k, as well as an individual start location k^+ and end location k^-. Furthermore, we define $K^+ = \bigcup_{k \in K} k^+$, $K^- = \bigcup_{k \in K} k^-$, and $W = K^+ \cup K^-$. Hence, in the corresponding digraph G, the set of nodes V is defined as $V = L \cup W$. For all $i, j \in V$, d_{ij} denotes the travel distance, t_{ij} the travel time, and c_{ij} the costs for traveling from location i to j. Note that service times required at request locations can be easily integrated into the travel times, hence they are not considered explicitly.

The model of the GPDP is formulated using four types of variables. Each binary decision variable z_{ik} $(i \in R, k \in K)$ is equal to 1 if request i is assigned to vehicle k, 0 otherwise. Each binary decision variable x_{ijk} $((i, j) \in (L \times L) \cup \{(k^+, j)|k \in K, j \in L\} \cup \{(j, k^-)|j \in L, k \in K\},\ i \neq j)$ is 1 if vehicle k directly travels from location i to j, 0 otherwise. Each non-negative variable D_i $(i \in V)$ specifies the departure time at location i. Finally, each non-negative variable y_i $(i \in V)$ denotes the current load of the corresponding vehicle when it arrives at location i. The GPDP can be modeled as follows:

$$\min z = f(x) \quad \text{s.t.} \tag{2.18}$$

$$\forall i \in R: \quad \sum_{k \in K} z_{ik} = 1 \tag{2.19}$$

$$\forall i \in R, l \in \left\{R_i^+ \cup R_i^-\right\}, k \in K: \quad \sum_{j \in V} x_{ljk} = \sum_{j \in V} x_{jlk} = z_{ik} \tag{2.20}$$

$$\forall k \in K: \quad \sum_{j \in \{L \cup \{k^-\}\}} x_{k^+jk} = 1 \tag{2.21}$$

$$\forall k \in K: \quad \sum_{i \in \{L \cup \{k^+\}\}} x_{ik^-k} = 1 \tag{2.22}$$

$$\forall k \in K: \quad D_{k^+} = 0 \tag{2.23}$$

$$\forall i \in R, p \in R_i^+, q \in R_i^-: \quad D_p \leq D_q \tag{2.24}$$

$$\forall i, j \in V, k \in K: \quad x_{ijk} = 1 \Rightarrow D_i + t_{ij} \leq D_j \tag{2.25}$$

$$\forall k \in K: \quad y_{k^+} = 0 \tag{2.26}$$

$$\forall i \in R, l \in \left\{R_i^+ \cup R_i^-\right\}: \quad y_l \leq \sum_{k \in K} Q_k \cdot z_{ik} \tag{2.27}$$

$$\forall i, j \in V, k \in K: \quad x_{ijk} = 1 \Rightarrow y_j = y_i + q_i \tag{2.28}$$

The objective function is defined by (2.18). The authors state that depending on the considered application, different functions can be used for $f(x)$, for example travel distance, travel time, or customer inconvenience. For details, we refer to Savelsbergh and Sol (1995) and Sect. 2.4.12. Constraints (2.19) make sure that each customer request is assigned to exactly one vehicle. Constraints (2.20) ensure that a vehicle only visits and leaves customer request locations which are belonging to requests assigned to the vehicle exactly once. Constraints (2.21) and (2.22) define that each vehicle starts and ends its tour at the specified location. Constraints (2.23) enforce that all vehicles start their tour at time point 0. Constraints (2.24) form the request precedence constraints. Specifically, all pickup locations of a request have to be serviced before its delivery locations are allowed to be serviced. Constraints (2.25) set the correct departure times at visited locations. Note that the less or equal formulation allows the integration of time windows defined by $[e_i, l_i],\ i \in L$ and $D_i = \max(A_i, e_i)$ with A_i denoting the arrival time of the servicing vehicle at request location i. Hence, vehicles are allowed to wait at request locations if they arrive there before e_i. Constraints (2.26), (2.27), and (2.28) together form capacity constraints.

In what follows, two special cases of the GPDP are presented. Note that the previously described TSP and VRP also represent special cases of the GPDP.

2.3.3.2 The Pickup and Delivery Problem (PDP)

The Pickup and Delivery Problem (PDP) is a special case of the GPDP. Specifically, in the PDP, each request $i \in R$ has exactly one pickup and one delivery location and only one depot exists at which all vehicles start and end their tour. The PDP can be formulated by using the model of the GPDP and setting its parameters to $|W| = 1$ and $\forall i \in R$: $|R_i^+| = |R_i^-| = 1$ so that all vehicles have the same start and end position and each request consists of exactly one pickup and one delivery location.

2.3.3.3 The Dial-A-Ride Problem (DARP)

In Dial-A-Ride Problems (DARP, cf., e.g., Cordeau and Laporte 2003, 2007), applications which deal with the transportation of passengers are considered. The DARP can be formulated by using the mathematical model of the GPDP and defining its parameters as $|W| = 1$, $\forall i \in R$: $|R_i^+| = |R_i^-| = 1$, and $\forall i \in R$: $\bar{q}_i = 1$ since each request is assumed to consist of exactly one passenger. Common real-world applications of the DARP are taxi cab related problems or scheduled transportation of patients to medical services. Since in the DARP people are transported instead of goods, the objective function often includes specific passenger convenience attributes. For example, the ride time of individual passengers can be restricted or longer ride times penalized. According to Larsen (2000), the DARP exists in three different versions:

- **Many-to-One (MTO):** In this variant of the DARP, only one destination to which passengers are transported is considered. This case represents for example a real-life application where different customers from a small residential area are collected and afterwards transported to the same destination.
- **Many-to-Few (MTF):** In this case, the number of destinations is still relatively small compared to the number of origins. In real-life applications, this case represents for example picking up customers from a residential area and transporting them to different stops at a large shopping mall.
- **Many-to-Many (MTM):** This variant of the DARP can be considered as a taxi cab system in which multiple individual passengers with different origins and destinations have to be serviced.

More information on the DARP and its dynamic variants can be found in Sects. 2.6.5.1 and 2.6.5.3.

2.3.4 Complexity of Node-Based Routing Problems

All the previously described node-based routing problems represent *combinatorial optimization problems*. Specifically, since the corresponding tasks of assigning and sequencing requests on vehicles can only be performed in a finite number of different *combinations*, the number of existing solutions is also finite. Hence, one method for determining an optimal solution consists of just evaluating all feasible solutions. However, as we illustrate in this section, the computational time required to solve routing problems by using complete enumeration is often too high. Combinatorial problems can be distinguished by their complexity in order to estimate the required computational time for solving them. According to Garey and Johnson (1979, p. 6), a function $f(n)$ is $O(g(n))$ if and only if there exists a constant c so that it holds that $|f(n)| \leq c \cdot |g(n)|$ for all values of $n \geq 0$ with n denoting the length of the input, e.g., the number of customer requests in a routing problem. According to the authors and this notation, an algorithm is called a *polynomial time algorithm* if its computational time complexity function is $O(p(n))$ for some polynomial function p. The authors state that all algorithms with no polynomial function bounding their time complexity are called *exponential time algorithms*. Hence, the group of exponential time algorithms covers all algorithms that have at least an exponential time complexity (cf. Gass 2003, p. 304). Using this classification, it can be decided whether an algorithm can be used to solve the considered problem with reasonable computational effort. Specifically, in Table 2.1, the computational time complexity for algorithms with a polynomial, exponential, and factorial time complexity are illustrated. In this table, it is assumed that each computational operation consumes the same amount of time which is defined as one microsecond, i.e., 10^{-6} seconds. As can be derived from this table, complete enumeration cannot be used to solve a given problem which belongs to the exponential classification in reasonable time if larger problem instances are considered.

In order to link the results illustrated above to the time complexity of algorithms for solving routing problems, we consider the required computational time for finding an optimal solution of a TSP instance with n customer requests. Since the number of feasible solutions amounts to $(n-1)!$, the number of feasible solutions grows in a factorial manner with n. Therefore, even TSP instances with a small size cannot be practically solved by enumerating all possible solutions in a reasonable amount of time (see again Table 2.1). Although more sophisticated solution methods have been developed for solving TSPs, the TSP still belongs to the class of strongly NP-hard combinatorial optimization problems (cf., e.g., Garey and Johnson 1979, pp. 211–214, Laporte 1992b, and Korte and Vygen 2012, p. 405). In computational complexity theory, the problem class P comprises all problems which can be solved in polynomial time using a deterministic Turing machine while those of

Table 2.1 Comparison of several polynomial and exponential time complexity functions (cf. Garey and Johnson 1979, p. 7)

Time complexity function	Problem size n					
	10	20	30	40	50	60
n	0.00001 seconds	0.00002 seconds	0.00003 seconds	0.00004 seconds	0.00005 seconds	0.00006 seconds
n^2	0.0001 seconds	0.0004 seconds	0.0009 seconds	0.0016 seconds	0.0025 seconds	0.0036 seconds
n^3	0.001 seconds	0.008 seconds	0.027 seconds	0.064 seconds	0.125 seconds	0.216 seconds
n^5	0.1 seconds	3.2 seconds	24.3 seconds	1.7 minutes	5.2 minutes	13.0 minutes
2^n	0.001 seconds	1.0 seconds	17.9 minutes	12.7 days	35.7 years	366 centuries
3^n	0.059 seconds	58 minutes	6.5 years	3,855 centuries	$2 \cdot 10^8$ centuries	$1.3 \cdot 10^{13}$ centuries
$n!$	3.6288 seconds	771 centuries	$8.4 \cdot 10^{16}$ centuries	$2.6 \cdot 10^{32}$ centuries	$9.6 \cdot 10^{48}$ centuries	$2.6 \cdot 10^{66}$ centuries

the problem class NP can only be solved in polynomial time by a non-deterministic Turing machine (cf., e.g., Garey and Johnson 1978; 1979, pp. 27 and 31). Moreover, a problem is denoted as NP-hard if all problems of the problem class NP are *reducible* to the considered problem. According to Lenstra and Rinnooy Kan (1979), a problem A is defined as reducible to problem B if for each instance of problem A, an instance of problem B can be generated in polynomial time so that solving the instance of problem B will also solve the instance of problem A. Hence, the reducibility of problem A to problem B implies that problem A represents a special case of problem B. Therefore, problem B is at least as hard as problem A. Furthermore, a problem is NP-complete if it is NP-hard and also belongs to problem class NP. For more information on computational complexity theory, see, e.g., Garey and Johnson (1978; 1979, pp. 23–33) and Lenstra and Rinnooy Kan (1979).

An NP-hard combinatorial optimization problem is called strongly (or unary) NP-hard if it remains NP-hard even when all input numbers are encoded in unary, cf., e.g., Garey and Johnson (1978; 1979, p. 120) and Lenstra and Rinnooy Kan (1979). Otherwise, NP-hard combinatorial problems are called weakly (or binary) NP-hard if a pseudo-polynomial time algorithm for solving the problem is known: "A pseudo-polynomial time algorithm can be useful even when there is no natural bound on the input numbers we expect. It will display 'exponential behavior' only when confronted with instances containing 'exponentially large' numbers, and instances of this sort might be rare for the application we are interested in. If so, this type of algorithm might serve our purposes almost as well as a polynomial time algorithm" (Garey and Johnson 1979, p. 91).

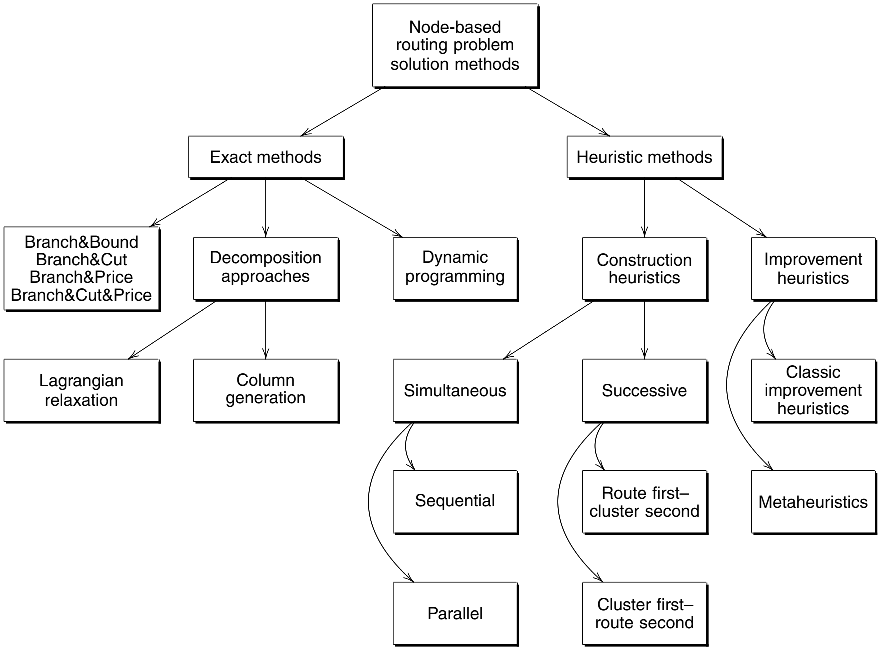


Fig. 2.4 Classification of different solution methods for node-based routing problems (in accordance with Richter 2005, p. 42 and Aberle 2009, p. 543)

Since the TSP is a strongly NP-hard combinatorial optimization problem as mentioned above, all currently known algorithms for solving TSPs to optimality have at least an exponential computational time complexity in the worst case. Note that since all of the previously described routing problems, such as the VRP and the GPDP, are extensions of the TSP, these problems are also strongly NP-hard combinatorial problems. Nevertheless, by utilizing specific characteristics of these routing problems or by accepting close-to-optimum solutions, various solution techniques which are applicable for solving practical-sized problem instances in reasonable time have been developed for the considered node-based routing problems. An overview of various solution techniques is presented in the next section.

2.3.5 Solution Methods for Node-Based Routing Problems

In order to solve node-based routing problems, various exact and heuristic solution methods have been developed (cf., e.g., Laporte 1992a, Toth and Vigo 2002, Richter 2005, Cordeau et al. 2007b, and El-Sherbeny 2010). Figure 2.4 gives an overview of several techniques developed for solving node-based routing problems exactly and heuristically.

2.3.5.1 Exact Solution Methods

Despite their exponential computational time complexity in the worst case, exact solution methods determine an optimal solution of a routing problem in finite time. Using problem-specific characteristics, they can be used to derive an optimal solution for certain routing problems within reasonable time in the average case. Moreover, they usually provide information about the solution quality of the currently best known solution during the solution process by calculating the current *gap*. This gap denotes the maximum percentage deviation from the optimal objective function value. Specifically, this gap is determined by calculating the deviation of the currently best known solution's objective function value from the value of a calculated lower bound of the considered problem (cf. Toth and Vigo 2002, p. 159).

One class of exact solution methods are *Branch-and-Bound* (B& B) and related approaches which were first introduced by Land and Doig (1960) and extended by Dakin (1965) for solving MIP problems using *tree-search* approaches. In general, first a relaxed version of the MIP problem, e.g., its linear relaxation, is solved in the *root node* of the *search tree*. If all variables in the solution of the relaxed problem are integer, the procedure stops. Otherwise, an integer variable x_k is selected whose value x_{k0} is continuous in the current relaxed solution. Using x_k, two new problems are constructed in the *branching step* by adding the restriction $x_k \leq \lfloor x_{k0} \rfloor$ and $x_k \geq \lceil x_{k0} \rceil$, respectively, to the corresponding relaxed problem, thus creating two new nodes, i.e., subproblems, in the search tree. By solving these relaxed versions of the original problem again, lower bounds for the optimal objective function value of the original problem existing in the corresponding subtree of the current node can be calculated. In doing so, it is known with certainty that no better solution can be found in the subtree belonging to this node if the lower bound of the current node is higher than the objective function value of the currently best known solution of the original problem. Hence, this subtree can be removed from further consideration—this activity is called *bounding*. The aim of B&B approaches is to find an optimal solution by examining as few nodes as possible in the search tree by calculating tight lower bounds in each node so that subtrees of the search tree which do not lead to an optimal solution can be early bounded (cf. Gass 2003, p. 313). One of the first B&B related approaches for the TSP is the approach of Little et al. (1963). In Branch-and-Cut (B& C) approaches, additional cutting planes are subsequently integrated in order to strengthen the relaxed problem solved in each node in order to calculate better lower bounds. In Branch-and-Price (B&P, cf. Barnhart et al. 1998), column generation is applied in each node with the same aim of achieving better lower bounds. For example, Savelsbergh and Sol (1998) propose a B&P approach for solving a real-life GPDP. Furthermore, both techniques can be combined to Branch-and-Cut-and-Price (B&C&P) approaches (cf., e.g., Fukasawa et al. 2006).

Lagrangian relaxation and column generation based approaches belong to the class of decomposition approaches. *Lagrangian relaxation* (cf. Fisher 2004) is often used as a sub-procedure in order to determine good lower bounds for B&B and related approaches. Specifically, by removing certain restrictions from the original problem, a relaxed problem which can be solved significantly faster than the original problem is created. Due to the removed restrictions, an optimal solution of the relaxed problem is likely to be infeasible for the original problem. Nevertheless, it represents a lower bound of the optimal solution's objective function value of the original problem. In order to reduce such infeasibilities, a violation of the removed restrictions is penalized in the objective function of the relaxed problem. The challenge then is to find good Lagrange multipliers by which these penalties are integrated into the objective function of the relaxed problem so that infeasibilities with regard to the original problem are reduced as much as possible. In so doing, the calculated lower bounds are tightened, i.e., they are closer to the objective function value of an optimal solution of the original problem thereby gaining a higher quality. In order to find such good Lagrange multipliers, subgradient methods are frequently applied. One approach for computing tight lower bounds for the Symmetric Traveling Salesman Problem (STSP) using Lagrangian relaxation is proposed by Held and Karp (1970). In this approach, the problem is relaxed to a Minimum Spanning Tree (MST) problem which can be solved in polynomial time (cf. Sedgewick 1983, p. 413). Since the solution of an MST often disregards tour connectivity characteristics of a TSP solution, this Lagrangian relaxation-based approach is used as a sub-procedure of a B&B solution approach proposed in Held and Karp (1971).

Column generation (cf. Desaulniers et al. 2005) is another solution technique for reducing the average computational effort which is required to compute an optimal solution of combinatorial optimization problems. Starting with a small subset of decision variables, duality costs of the generated optimal solution of this reduced problem are examined in order to add further promising decision variables which lead to better solutions. Column generation has been successfully applied to VRPs, e.g., by Krumke et al. (2002) and Westphal and Krumke (2008). In these approaches, binary decision variables are used to select tours out of a given set of available tours. Specifically, set partitioning is utilized in order to construct feasible tour plans out of the currently available tours. New tours, i.e., columns, are subsequently generated by extending the set of available tours with tours comprising more customer requests. Since in the practical application considered in both approaches optimal solutions often consist of tours with a low number of requests, optimal solutions are often achieved within acceptable time limits. Note that this is one example of efficiently utilizing the given problem structure in developing practically applicable solution methods. More information about the mentioned approaches can be found in Sect. 4.6 of the literature review.

Dynamic programming (cf. Bellman 1954, 2003) can be applied to combinatorial problems which can be modeled as *multi-stage decision processes*. The current state of such a process is determined by a set of quantities denoted as *state variables* (cf. Bellman 1954, p. 503). The current stage is advanced by making *decisions* which transform the state variables. According to Bellman (1954, p. 503), a "sequence of decisions [is] called a policy, and a policy which is most advantageous according to some preassigned criterion [is] called an optimal policy" which represents an optimal solution of the considered problem. Moreover, the author states the principle of optimality as follows, "An optimal policy has the property that whatever the initial state and initial decisions are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decisions" (Bellman 1954, p. 504). According to Lew and Mauch, this also means that optimal policies have optimal subpolicies, "That this observation is valid follows from the observation that, if a policy has a subpolicy that is not optimal, then replacement of the subpolicy by an optimal subpolicy would improve the original policy" (Lew and Mauch 2007, p. 5). Hence, the optimal policy is determined by making optimal decisions in the current stage by using the optimal decision of previous stages with regard to the considered objective function. This procedure is also known in the literature as the *optimal substructure* property. In each stage, the size of the considered subproblems is increased so that an optimal solution of the original problem is finally calculated. One issue is that in NP-hard routing problems "the cardinality of the state space is usually exponential in the problem size" (Desrochers et al. 1988). If problem-specific characteristics can be used to limit the set of feasible subproblems early, i.e., removing states which cannot be extended to a feasible solution of the original problem, these characteristics can be utilized to reduce the number of states required to be solved for calculating an optimal solution of the original problem. For instance, the approach of Dumas et al. (1995) for solving a TSP with hard time windows uses different feasibility tests in order to detect infeasible states due to time window violations. At each stage, the set of currently considered requests is increased until finally an optimal policy for all requests of the original problem can be derived using the calculated optimal policies for the previously considered subproblems. If tight time windows are present in a problem instance, the approach is able to significantly reduce the search space and hence also the computational time required to solve the problem instance to optimality.

2.3.5.2 Heuristic Solution Methods

In contrast to exact solution approaches, heuristic solution approaches aim at finding good solutions quickly, often without providing information about the gap to optimality. Heuristic solution methods for routing problems can be divided into *construction heuristics* and *improvement heuristics*. Construction heuristics generate a tour plan from scratch while improvement heuristics improve a given solution.

Construction heuristics can be classified into simultaneous and successive approaches. In *simultaneous* approaches, the assignment and sequencing of requests is performed at the same time. Such approaches can be further distinguished according to how the tours are constructed. Specifically, if the tours are constructed one after another, a construction heuristic is called *sequential*. If the vehicles' tours are constructed at the same time so that a pending request can be inserted in different tours, the construction heuristic is called *parallel*. Apart from simultaneous construction heuristics, *successive* construction heuristics exist in which the assignment and sequencing tasks are consecutively performed. For example, if a *route first–cluster second* heuristic is applied to a VRP, a TSP is solved first resulting in a so-called giant tour which determines the sequence of all requests. In a second step, restrictions such as capacity constraints are used to divide the giant tour into individual vehicle tours. In *cluster first–route second* heuristics, first the assignment of requests to vehicles is performed by considering existing constraints. Afterwards, a TSP is solved for each vehicle tour. Further information and examples of construction heuristics can be found among others in Bräysy and Gendreau (2005a, pp. 6–10).

Besides construction heuristics, improvement heuristics are used to modify an existing solution in order to improve its solution quality. Improvement heuristics can be categorized into *classic improvement heuristics* and *metaheuristics*. What both types of heuristics have in common is that a given current solution is modified using one or more *neighborhood operators*. The set of solutions that can currently be generated by using a specific neighborhood operator is called the *neighborhood* of the current solution. After evaluation of the neighborhood, one solution out of the neighborhood is selected as the new current solution. Using this solution as the new current solution, the described neighborhood search process starts all over again.

Classic improvement heuristics (cf. Bräysy and Gendreau 2005a) modify a given tour plan until a local optimum is reached, i.e., until there exists no solution in the neighborhood of the current solution with a better objective function value than the current solution. Besides classic improvement heuristics, various metaheuristic concepts for routing problems have been developed. The basic idea is to embed one or more problem-dependent neighborhood operators into a problem-independent framework which controls the application of these neighborhood operators. The purpose of these frameworks is to overcome local optima in order to find better solutions in other areas of the solution space. In Hoos and Stützle (2005), a large variety of known metaheuristics is described where the most popular are: Variable Neighborhood Descent (VND), Simulated Annealing (SA), Tabu Search (TS), Greedy Randomized Adaptive Search Procedure (GRASP), Ant Colony Optimization (ACO), and Memetic Algorithms (MAs). Furthermore, metaheuristic solution approaches for routing problems are also based on Genetic Algorithms (GAs, cf. Mitchell 1998) and Variable Neighborhood Search (VNS, cf. Hansen and Mladenovic 2001). Another sophisticated solution technique is Large Neighborhood Search (LNS) which has for example been applied to routing problems by Bent and van Hentenryck (2004a). Ropke and Pisinger (2006) utilize LNS in a modified version called Adaptive Large Neighborhood Search (ALNS). Further information on metaheuristics for routing problems can be found in Toth and Vigo (2002), Cordeau et al. (2005), Bräysy and Gendreau (2005b), Gendreau et al. (2006), and Gendreau et al. (2008). The Tabu Search metaheuristic developed and utilized in the real-time control approaches will be presented in Chap. 7 of this book.

The solution of a routing problem can be regarded as a graph in which customer requests and the depot represent the nodes of the graph and in which the arcs represent the routes, i.e., the vehicle travel activities between the relevant locations. In doing so, neighborhood operators which are applied in improvement heuristics can be subdivided into *arc-exchanging* and *node-exchanging* operators. In arc-exchanging operators, two or more arcs which, depending on the operator, belong to the same vehicle tour or to different tours, are exchanged so that one or more customer requests are indirectly moved to a different position in the tour plan. In contrast, node-exchanging operators directly move customer requests to another position in the same tour or in another tour. The following neighborhood operators are frequently applied in improvement heuristics for VRPs:

- **2-opt and 3-opt operator:** These operators introduced by Lin (1965) exchange two or three arcs within the same tour.
- **Or-opt operator:** This operator proposed by Or (1976) exchanges three arcs in the same tour in order to move a sequence of customer requests to another position on the same tour while preserving the sequence in which the exchanged customers are serviced.
- **2-opt* operator:** This operator by Potvin and Rousseau (1995) exchanges the latter part of two different tours while preserving the sequence of the requests in the exchanged tour parts.
- **Relocate, exchange, and cross operator:** Savelsbergh (1992) introduces these three node-exchanging operators. The relocate operator moves one customer to another tour while the exchange operator swaps two customers in two different tours preserving their positions in the tours. The cross operator works similar to 2-opt*.
- **CROSS-exchange operator:** Badeau et al. (1997) propose the CROSS-exchange operator. It considers two tours in which a part of each tour is exchanged. Both exchanged tour parts can have a different number of customers. Since one part can also consist of zero customers, it is possible to move one or more customers from one tour to another thereby generalizing other neighborhood operators such as the previously mentioned relocate, exchange, Or-Opt, and 2-opt* operator.

The neighborhood operators which are utilized in the proposed Tabu Search metaheuristic will be presented in Sect. 7.2.

2.4 Extensions to the Vehicle Routing Problem

As mentioned above, since RDOPG applications can be modeled as a variant of the VRP, the remainder of this chapter mainly deals with characteristics and aspects of VRPs. In this section, several extensions and special cases of the classic VRP and other objective functions are presented. For further variants and explanations we refer to Toth and Vigo (2002, p. 177), Baldacci et al. (2010), and El-Sherbeny (2010).

2.4.1 Time Windows

In VRPs with time windows, a time window $[r_i^{\mathrm{e}}, r_i^{\mathrm{l}}]$ is assigned to each customer request $i \in R$. We denote the point in time when the servicing vehicle arrives at request i by y_i^{s}. In general, two types of time windows exist. In case of *hard time windows*, service of a request is required to start before the end of the time window, i.e., $\forall i \in R$: $y_i^{\mathrm{s}} \leq r_i^{\mathrm{l}}$. Tour plans which do not fulfill these hard time window restrictions are infeasible. If a vehicle arrives at the location of request i before r_i^{e}, it is usually allowed to wait there. This problem variant is called the Vehicle Routing Problem with Time Windows (VRPTW, cf. Toth and Vigo 2002, p. 8). The second variant is the Vehicle Routing Problem with *Soft Time Windows* (VRPSTW, cf. Toth and Vigo 2002, p. 179). In the VRPSTW, service of requests is allowed to be performed after the end of the time window but at the expense of additional *penalty costs*. These penalty costs may either be *fixed*, i.e., occurring for each late serviced request regardless of the length of the lateness, *variable*, i.e., depending on the length of the lateness, or a combination of both. More information on fixed and variable penalty costs are given in Sect. 2.4.12.

Another variant is the Vehicle Routing Problem with Multiple Time Windows in which each request i can be serviced during different time windows (cf. Toth and Vigo 2002, p. 179). A further variant of the VRPTW is the Vehicle Routing Problem with Time Deadlines (VRPTD, cf. Thangiah et al. 1993), where each request i can be serviced at any point in time up to a given latest time point denoted by r_i^{l}. This variant can be modeled using a VRPTW and setting $r_i^{\mathrm{e}} = 0$ for all requests $i \in R$.

2.4.2 Backhauls

In the Vehicle Routing Problem with Backhauls (VRPB), both mentioned real-world business cases of VRPs dealing with delivering and collecting goods at customer locations are combined. For this purpose, the set of requests is split into the ones requiring *linehauling* (delivering goods) and those that require *backhauling* (collecting goods). In VRPB problems, usually all linehaul customer requests must be serviced before the backhaul requests (cf. Toth and Vigo 2002, p. 9, Parragh et al. 2008).

2.4.3 Simultaneous Pickup and Delivery

The Vehicle Routing Problem with Simultaneous Pickup and Delivery (VRPSPD) is an extension of the VRPB. Specifically, in this problem, backhaul requests can already be collected by a vehicle before it has delivered all linehaul requests (cf. Toth and Vigo 2002, p. 10).

2.4.4 Multi-Depot Problems

In the Multi-Depot Vehicle Routing Problem (MDVRP, cf. Golden et al. 2008, p. 164), dispatchable vehicles are located at different depots. There may be special restrictions for assigning requests to vehicles. Specifically, certain requests may require service by a vehicle from a specific depot, for example if goods to be delivered are only available at certain depots.

2.4.5 Open Routing Problems

In the Open Vehicle Routing Problem (OVRP), vehicles do not need to return to their depot after completing their assigned tour. This problem is for example considered in the work of Tarantilis et al. (2004).

2.4.6 Vehicle Scheduling

While in VRPs decisions only comprise the assignment and sequencing of requests to vehicles, in Vehicle Routing and Scheduling Problems (VRSP) scheduling decisions of vehicles are additionally considered (cf. Bodin and Golden 1981, Solomon 1987, and Mitrovic-Minic et al. 2004). Specifically, vehicles leave the depot and customer locations only at specified time points. Vehicle scheduling decisions are useful for different purposes. For example, in order to consider law-regulated driving time restrictions (cf. Brandão and Mercer 1997 and Sect. 2.4.10), vehicles can be instructed to wait for a certain amount of time at a specified request location. Furthermore, this allows the implementation of waiting strategies (see Sect. 4.3). Note that the integration of scheduling significantly increases the decision space so that the problem complexity is increased.

2.4.7 Multiple Compartments

In the VRP variant denoted as the Multiple Compartment Vehicle Routing Problem (MCVRP, cf. Muyldermans and Pang 2010), vehicles have different storage areas for goods. This is important for transporting goods on the same vehicle which must not be mixed, e.g., different liquids.

2.4.8 Multiple Trips

In Multiple Trip Vehicle Routing Problems, vehicles can return to the depot during the planning horizon and set for another trip afterwards. Such intermediate returns to the depot are useful if distribution problems are considered in which the load of all requests assigned to a vehicle's tour exceeds its capacity. Hence, the vehicle can replenish its capacity and continue to service further requests. Furthermore, vehicle scheduling decisions may become relevant, e.g., for determining when the vehicle sets for its next trip. For example, Brandão and Mercer (1997) consider a Multiple Trip Vehicle Routing and Scheduling Problem (MTVRSP) in which legal time breaks and maximum legal driving times per day are additionally considered.

2.4.9 Time-Dependent Travel Times

VRPs with time-dependent travel times are denoted as Time-Dependent Vehicle Routing Problems (TDVRP). A characteristic of these problems is that the same route requires a different amount of time depending on the current time of the day (cf. Malandraki and Daskin 1992, Ichoua et al. 2003). Specifically, the travel time between two relevant locations $i, j \in V$ is a function $t_{ij}(\tau)$ which depends on the current time τ at which the travel activity starts at location i. In order to model time-dependent travel times, the time during the day is often divided into m time intervals $T' = \{T_1, T_2, \ldots, T_m\}$. Within each of these intervals $t \in T'$, the travel times are assumed to be constant. Note that when modeling time-dependent travel times it is reasonable to consider the first in-first out (FIFO) property, i.e., a later departure at location i does not lead to an earlier arrival at location j (cf. Ichoua et al. 2003 and Fleischmann et al. 2004a). Depending on the type of problem, time-dependent travel times can be known in advance with certainty, with some uncertainty, or even completely unknown (see also Sect. 2.5.2.1). Further approaches which consider time-dependent travel times can be found in the survey of Flatberg et al. (2005). If time-dependent travel times are present, scheduling decisions can become important in order to reduce travel times.

2.4.10 Legal Driving Time Regulations

In order to comply with legal regulations, different VRP approaches consider legal driving time regulations (cf. Brandão and Mercer 1997 and Ferrucci 2006, pp. 38–42). Depending on the country, these regulations consist of a maximum driving time per work shift or legal time breaks required after a certain driving time has been completed.

2.4.11 Heterogeneous Fleet

In the Heterogeneous Vehicle Routing Problem (HVRP), the available vehicles are divided into different classes. Differences between vehicle classes consist for example of different fixed and variable operational costs as well as individual loading capacities (cf. Gendreau et al. 1999b and Tarantilis and Kiranoudis 2007) and different travel speeds (cf. Ferrucci 2006, pp. 36–37).

2.4.12 Objective Functions in Vehicle Routing Problems

As mentioned in Sect. 2.3.3, different objective functions are pursued in VRPs according to the considered application. In what follows, we present different types of objective functions. Note that these individual objectives can also be combined by utilizing a weighted sum of different objective functions.

2.4.12.1 Minimization of Travel-Dependent Parameters

Besides the minimization of operative transportation costs denoted by c_{ij} as has been mentioned for example in the description of the TSP and the VRP, the minimization of travel distances d_{ij} or travel times t_{ij} is often considered $(i, j \in V)$.

2.4.12.2 Minimization of the Number of Utilized Vehicles

This objective function aims at minimizing the number of vehicles which are utilized to service the requests. This aim is often combined with the minimization of the total travel distance (cf., e.g., Homberger and Gehring 2005) or the total travel time (cf., e.g., Hvattum et al. 2006).

2.4.12.3 Minimization of the Sum of Tour Durations

This objective pursues the minimization of the total time utilized by all vehicles in order to service the requests (cf. Savelsbergh 1992). If time windows are present, vehicle scheduling aspects become relevant.

2.4.12.4 Minimization of the Completion Time

This objective function aims at minimizing the time point at which the last vehicle returns to the depot after all requests have been serviced. Note that if in the previously described objective all vehicles have to depart from the depot directly at the beginning of the execution of the transportation process, this and the previous objective become equal (cf. Savelsbergh 1992).

2.4.12.5 Minimization of Lateness Costs

In VRPSTWs, costs occurring for late serviced requests can be classified into fixed and variable lateness costs. We denote the length of the lateness that has occurred at request i by $r_i^{\mathrm{lt}} = \max(0, y_i^{\mathrm{s}} - r_i^{\mathrm{l}})$ where, as mentioned above, y_i^{s} describes the start time of service at request i. Furthermore, we define $F(x)$ as a function representing the amount of the penalty costs that occur at a lateness of length x. By using the objective function $\min z = \sum_{i \in R} F(r_i^{\mathrm{lt}})$, fixed penalty costs of amount P can be modeled by defining $F(x) = P \cdot \Theta(x)$ with $\Theta(x) = 1$ if $x > 0$ and 0 otherwise. Accordingly, variable penalty costs can be modeled by defining $F(x)$ as a monotonously increasing function, e.g., a linear or a quadratic function. As mentioned above, both types of penalty costs can also be combined.

2.4.12.6 Minimization of the Number of Unserviced Customers

This objective function is related to the previously presented one, but instead of soft time windows, hard time windows exist. Hence, late service of a customer is not allowed. Instead, if it holds that $y_i^{\mathrm{s}} > r_i^{\mathrm{l}}$, request i is removed from the system and is counted as an unserviced customer. Hence, in order to minimize the number of unserviced customers, it is the primary objective to service as many customers as possible within their time windows. In general, this objective can be modeled by fixed penalty costs which have been introduced in the previous section. However, it is advisable to provide more information to the solution method about how urgent a customer is to the solution method, i.e., how far away customer request i is still from its last allowed time point for start of service denoted by r_i^{l}. The utilization of such additional information has also been addressed in van de Klundert and Wormer (2010). In their work, the authors state that just using fixed penalty costs in order to fulfill maximum request response times defined by SLAs (which is related to the considered objective in this book, see also Sect. 1.1) is not advisable. By integrating appropriate additional information into the objective function, the solution method can precisely discriminate between different solutions in which all customer requests are feasibly serviced within their time windows. Therefore, the solution method is able to identify solutions in which requests are currently scheduled for service with larger buffer times before the end of their time windows. This is also advantageous in the considered RDOPG applications since it increases the possibilities for integrating future arriving requests into the tour plan within the maximum allowed response time so that no penalty costs occur (see also Sect. 5.3.2.4).

2.4.12.7 Minimization of Customer Inconvenience and Request Response Time

As mentioned in Chap. 1, the considered RDOPG applications aim at servicing requests so that customer inconvenience is minimized and a maximum request response time whose violation causes penalty costs is maintained if possible. While a maximum allowed request response time can be represented by using fixed lateness costs, customer inconvenience can be modeled by using variable lateness costs (see Sect. 2.4.12.5) and defining $r_i^{\mathrm{e}} = r_i^{\mathrm{l}} = r_i^{\mathrm{a}}$ with r_i^{a} denoting the point in time when customer request i arrives in the system. Depending on the utilized customer inconvenience function in the objective function $\min z = \sum_{i \in R} F(r_i^{\mathrm{lt}})$, different objectives can be achieved. Specifically, if a linear customer inconvenience function for $F(x)$ is assumed, the pursued aim is equal to the minimization of the average request response time. In contrast, by utilizing an over-linear customer inconvenience function such as a quadratic function for $F(x)$, more fairness with regard to the request response times of individual requests can be attained since the longer a customer waits, the quicker the inconvenience grows. More information on the evaluated customer inconvenience functions and the utilized objective function in the considered RDOPG applications will be presented in Sect. 5.3.2.4. Different effects on the solution quality resulting from the application of these customer inconvenience functions will be analyzed in the computational experiments in Chap. 8.

2.5 Information Revelation in Routing Problems

In all of the previously described routing problems, the solution is obtained by solving the corresponding model on basis of *problem data* which is given for the parameters of the model. In what follows, we define that this data is derived from the *relevant information* of the corresponding routing problem. Depending on the considered routing problem, the problem data is either constant or changes during the transportation process of the considered routing problem. Moreover, we define that the problem data changes during the execution of the transportation process if during its execution different parameter values are present in the model. Depending on the type of relevant information, such changes in the problem data may be known in advance with certainty or can contain uncertainty. Moreover, if changes in the relevant information are not known with certainty, stochastic knowledge about expected changes may be available. This is explained in detail in Sect. 2.5.3.

Basically, the problem data of a routing problem consists of the following types of information:

- **Customer requests:** Information about customer requests comprises the request location, the demand, start and end times of time windows, and fixed as well as variable penalty costs.
- **Vehicles:** In order to efficiently assign the requests to vehicles according to the given objective, information about the available fleet size and properties of the vehicles is required.
- **Travel times and travel distances:** This type of information is provided by the digraph G for all connections between the considered relevant locations (see Sect. 2.2).

In this section, we describe differences in the *revelation of relevant information in vehicle routing problems*, i.e., how and when changes in the relevant information occur in the considered routing problem during the execution of its transportation process. Furthermore, we present *approaches in the literature for characterizing relevant information and distinguishing between static and dynamic routing problems*.

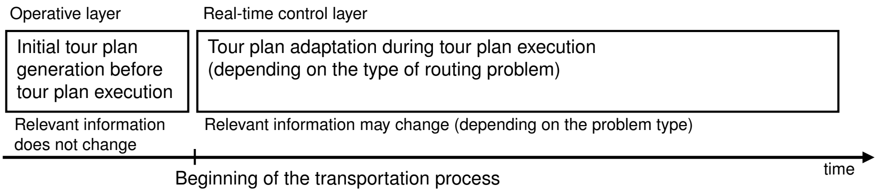


Fig. 2.5 General process of tour plan generation and adaptation before and during the execution of the transportation process

Based on the results of these approaches, *a unified approach for classifying routing problems with regard to characteristics and processing of relevant information*, which are both important for RDOPG applications, in routing problems is proposed. Using this approach, different types of VRPs known in the literature are presented in detail.

2.5.1 Revelation of Relevant Information in Vehicle Routing Problems

With regard to information revelation in VRPs, a common practice is to distinguish between the relevant information which is known *before* the execution of the transportation process and the relevant information which only becomes known *during* the execution of the transportation process (see Fig. 2.5).

Before the execution of the transportation process, an initial tour plan is generated based on the problem data given by the relevant information which is known at that time point. With regard to the layers of transportation processes (see Sect. 1.2), this initial tour plan generation corresponds to the operative layer. It is assumed that the time available for generating the initial tour plan is not restricted and that the relevant information does not change during this generation. After the initial tour plan generation is complete, the execution of the transportation process starts. Specifically, the vehicles leave the depot in order to service customer requests according to the initial tour plan. An important factor is that, depending on the considered routing problem, the existing relevant information can change during the execution of the transportation process. If such changes are caused by events whose time of occurrence during the execution of the transportation process cannot be precisely predicted in advance, we denote them as *dynamic events*—one example of a dynamic event is the arrival of a new request. Otherwise we state that *incomplete relevant information* is present. Specifically, although the relevant information does not change by itself, some parts of it are unknown during the initial tour plan generation process and are only revealed later. For example, the actual customer demand is only disclosed when the vehicle reaches the corresponding customer location. Routing problems with changing relevant information and appropriate tour plan adaptation techniques are described in detail in Sect. 2.5.3.2. In the next section, we deal with characteristics of relevant information as well as with differences between static and dynamic routing problems.

2.5.2 Approaches in the Literature for Characterizing Relevant Information and Distinguishing Between Static and Dynamic Routing Problems

In what follows, differences in routing problems according to characteristics of relevant information are presented. Moreover, approaches published in the literature for distinguishing between *static* and *dynamic* routing problems are described.

2.5.2.1 Characteristics of Relevant Information

In Psaraftis (1995), the author introduces four different criteria for classifying relevant information in routing problems:

- **Evolution of information:** In static routing problems, the relevant information remains constant while in dynamic routing problems the relevant information changes during the execution of the transportation process. More information about differences in static and dynamic routing problems is given in the next section.
- **Quality of information (about future events):** Information available about future events is of one of four types; future events can either be *known with certainty (deterministic)*, *known with uncertainty (forecasted)*, *follow prescribed probability distributions (probabilistic information)*, or *unknown (no information)*.
- **Availability of information:** Changes which occur in the relevant information during the transportation process can either be *locally* or *globally* available. One example of a local information availability is if a driver is informed of the actual amount of goods that the current customer requires only when the corresponding customer location is reached and there is no possibility of informing other vehicles or a central dispatching center about this change. In contrast, in routing problems where such changes are globally available, they are recorded in a central system.
- **Processing of information:** If a *centralized processing* of relevant information is applied, all changes in the relevant information are recorded and handled by a central dispatching center. In a *decentralized processing*, changes in the relevant information are autonomously handled by the vehicles.

2.5.2.2 Static and Dynamic Routing Problems

In this section, we describe different approaches proposed in the literature which classify static and dynamic routing problems. In what follows, we present the twelve aspects by Psaraftis (1988) in which static and dynamic routing problems can be distinguished from each other:

1. **Time dimension is essential:** In a static routing problem the current time dimension during the execution of the transportation process is not important since, as stated above, static routing problems only consider the generation of the initial tour plan before the execution of the transportation process and the relevant information does not change during its execution. In dynamic routing problems, time is essential for several reasons. First, since changes in the relevant information often occur unexpectedly during the execution of the transportation process, depending on the current time, vehicles have a different position and status which needs to be considered in order to efficiently adapt the tour plan according to occurring changes in the relevant information. Furthermore, during the adaptation of the tour plan according to the new system situation, time elapses in the transportation process. Hence, further changes in the system situation that occur during this adaptation need to be carefully considered (see also Sects. 2.5.3.2 and 4.7.4).
2. **Problem may be open-ended:** In contrast to static routing problems, in some dynamic routing problems the length of the planning horizon is not determined in advance.
3. **Future information may be imprecise or unknown:** As mentioned above, in static routing problems all relevant information is known in advance with certainty. In dynamic routing problems, future events are only known with uncertainty or are not known at all before the execution of the transportation process.
4. **Near-term events are more important:** In static applications where all relevant information is known with certainty, no drawbacks exist in directly committing vehicles to decisions which are only taking place in the far future. For example, a vehicle can be sent to an isolated customer request without negative consequences, even if it needs to wait there for a long time until this request can be serviced. In contrast, in dynamic routing problems, postponing far future decisions can be advantageous in order to make better decisions later when more relevant information is available, for example due to additional requests which have arrived intermediately.
5. **Information update mechanisms are essential:** Since in static problems the relevant information does not change, no tour plan adaptations are required during the transportation process. In dynamic routing problems, it is necessary to integrate occurring relevant information updates into the real-time control approach. Specifically, ongoing processes, e.g., requests currently being serviced or vehicle movements being performed, have to be considered in tour plan adaptations. Moreover, in order to inform vehicles about new decisions made during an adaptation, the existence of a bidirectional communication between the vehicles and the dispatching center is important.
6. **Resequencing and reassignment decisions may be warranted:** In dynamic routing problems, changes in the relevant information may render the currently executed tour plan infeasible so that it needs to be adapted. During this adaptation, existing requests may be required to be resequenced or reassigned to

other vehicles. Moreover, newly arrived requests are required to be integrated in order to comply with the new system situation.

7. **Faster computation times are necessary:** As mentioned in Sect. 2.5.1, in static routing problems, a large amount of computational time is assumed to be available for generating the initial tour plan with the demanded solution quality. In contrast, the available time in dynamic problems is limited so that a quick adaptation of the tour plan according to occurred changes in the relevant information is required (see also aspect 1). Hence, adaptations have to be carried out by sophisticated solution methods which provide a good solution quality in a short amount of time.
8. **Indefinite deferment mechanisms are essential:** In dynamic problems, it may happen that the service of a particular request is theoretically indefinitely postponed. Specifically, this happens if a request possesses an unfavorable geographical location and new requests keep arriving close to the current vehicles' positions. Hence, mechanisms which prohibit an indefinite deferment of individual pending requests are required. For example, time window constraints or a nonlinear objective function which generates higher penalty costs for excessive waiting times can be utilized for this purpose.
9. **Objective function may be different:** In static routing problems, the considered objective often pursues the minimization of operative transportation costs, travel distances, or travel times. In addition, different objectives are considered in specific dynamic routing problems. Specifically, in real-world dynamic routing applications it can be observed that the higher the number of dynamic events, the more important customer-oriented objectives become (see also Sect. 2.6.5).
10. **Time constraints may be different:** In dynamic routing problems, the unexpected occurrence of new requests may lead to situations in which it becomes very difficult or impossible to service all pending requests while fulfilling given hard time window constraints. Hence, in dynamic routing problems, time constraints are often soft so that service of requests is also allowed after the end of the corresponding time windows, but at additional penalty costs. Using hard time window constraints in dynamic routing problems allows for modeling business applications in which customers are treated as unserviced if their time windows are not met (see also Sect. 2.4.12.6).
11. **Flexibility to vary vehicle fleet size is lower:** In static problems it is often possible to utilize an arbitrary number of vehicles in order to meet request time window constraints. In contrast, in dynamic routing problems, a small available fleet size often results in a lower flexibility for efficiently integrating requests which dynamically arrive during the execution of the transportation process. This may lead to the violation of time window constraints and lower quality of service.
12. **Queueing considerations may become important:** In dynamic routing problems with a high number of dynamically arriving requests (also called heavy traffic scenarios), situations may occur in which the arrival rate of new requests exceeds a certain amount beyond which vehicles cannot service pending

requests without significant delays. Psaraftis states that in this case, classic solution approaches for routing problems result in a poor system performance so that the utilization of methods from the area of queueing theory is recommendable. Bertsimas and van Ryzin evaluate the efficiency of different request dispatching strategies in light and heavy traffic scenarios by using queueing theory insights in their approaches. Their results will be discussed in the literature review in Sect. 4.5.1.

Besides Psaraftis' twelve aspects, Bianchi (2000) proposes another approach to distinguish between static and dynamic routing problems. In this approach, a routing problem (or, in general, an optimization problem) is classified as dynamic if it is uncertain how parts of the relevant information are revealed or updated during the execution of the transportation process. Hence, impacts of occurring changes in the problem data are not always known with certainty. As mentioned above, if the tour plan currently in execution is required to be adapted, this has to be carried out while time in the system elapses so that the solution process is concurrently executed with occurring changes in the relevant information. Since it is not sufficient to only determine an initial solution once, only a strategy or policy, i.e., a solution method, which is used to solve the problem during the execution of the transportation process can be specified in advance. If this is not the case, a routing problem is deemed not dynamic according to this classification. Therefore, routing problems in which one of the following two characteristics is present are not dynamic according to the author:

- The problem data changes over time, but these changes are known with certainty in the relevant information.
- The relevant information is constant, but some parts of it are unknown at the beginning of the execution of the transportation process and are only revealed during its execution, for example when a vehicle reaches a customer location. Moreover, for all of these uncertain parts, prescribed probabilistic information is given as stochastic knowledge in advance. Hence, an a priori solution which is only changed in minor ways during the execution of the transportation process can be determined in advance (see Sect. 2.5.3.2.1).

Another approach for characterizing static and dynamic routing problems is proposed in the work of Ghiani et al. (2003). The authors state that a routing problem is static if all types of its relevant information are independent of the current time during the execution of the transportation process. If this is not the case, then the routing problem is dynamic. Moreover, if all changes in the relevant information which occur during the execution of the transportation process are known in advance with certainty, the routing problem is called *deterministic*, otherwise *stochastic*. Consequently, four types of routing problems arise in this characterization approach:

- In static deterministic routing problems, no changes in the relevant information occur during the execution of the transportation process so that the problem data is constant. Hence, only the generation of the initial tour plan before the execution of the transportation process is considered in problems of this type. Problems of this type are also denoted as static routing problems which are presented in Sect. 2.5.3.1.1.

- In static stochastic routing problems, some parts of the relevant information are only known with uncertainty before the execution of the transportation process. This uncertain relevant information often belongs to customer attributes, e.g., demand of customers or whether customers are required to be serviced at all. Characteristic for these types of routing problems is that the uncertain relevant information is given by random variables that follow prescribed probability distributions which makes the problem stochastic. Moreover, the prescribed probability distributions are often available as stochastic knowledge before the execution of the transportation process. Since the values of these random variables do not change during the execution of the transportation process, the problem is characterized as static. The actual values of the uncertain relevant information are revealed during the execution of the transportation process. For example, when a vehicle reaches a customer location, its actual demand is revealed. Hence, the problem data changes in this type of problem, but these changes can be considered in the initial tour plan generation due to the available stochastic knowledge. Routing problems of this type are known in the literature as variants of the Stochastic Vehicle Routing Problem which is described in Sect. 2.5.3.2.1.
- In dynamic deterministic routing problems, the relevant information remains constant, but the problem data changes over time during the execution of the transportation process. Characteristic of such routing problems is that the point in time when these changes in the problem data happen as well as their effects are known in advance with certainty. For example, routing problems in which travel times change during the planning horizon according to a function which is known in advance belong to this problem type. Routing problems of this type are described in Sect. 2.5.3.1.2.
- In dynamic stochastic routing problems, the relevant information changes at unknown time points during the execution of the transportation process when dynamic events occur. Moreover, no knowledge about these dynamic events is available. Such problems are described in Sect. 2.5.3.2.2.

Pillac et al. (2010) propose an approach similar to the previous one. Specifically, static routing problems are classified as described in Ghiani et al. (2003). In dynamic routing problems, relevant information changes during the execution of the transportation process and these changes always inhibit uncertainty. Specifically, dynamic deterministic routing problems are characterized by the fact that no information about future changes in relevant information is available. Hence, this group of routing problems is equal to the dynamic stochastic problems in the aforementioned classification. In contrast, this classification regards dynamic stochastic routing problems as real-time routing problems in which some stochastic knowledge about future events is available. Problems of this type are presented in Sect. 2.5.3.2.2.

2.5.2.3 Dynamism of the Problem, Model, and Application in Routing Problems

In Powell et al. (1995), the authors do not only deal with the type of changes of the problem data, which they denote as the *problem*, during the transportation process, but they also categorize the approach utilized to solve the given problem. Specifically, besides a *model* which is used to represent the problem in the solution approach, the *application* defines the way in which the approach solves the model during the transportation process. Hence, in a given routing application, the problem is predetermined while the model and the application are approach-dependent.

In the proposed categorization, the authors distinguish each of these three items according to the type of *dynamism* with which each of them exists in a given routing problem and the applied solution approach:

- **Dynamism of the problem:** The problem is dynamic if one or more of its parameters change over time during the execution of the transportation process according to functions of time which may be known or unknown before its execution. Two types of dynamic problems are distinguished by the authors: Problems with *dynamic data* and problems with *time-dependent data*. In the first type, the relevant information and hence the problem changes unexpectedly, i.e., the mentioned functions are unknown. For instance, this includes dynamically arriving requests, unexpected changes in traffic conditions, or unexpected changes of vehicles or driver statuses. In dynamic problems with time-dependent data, changes in the problem occur deterministically according to functions which are known in the relevant information before the execution of the transportation process.
- **Dynamism of the model:** The model represents the problem in a specific solution approach. It is dynamic if it explicitly considers changes in the problem during the execution of the transportation process whereas a static model does not consider any occurring changes in the problem. Furthermore, models are subdivided into deterministic and stochastic models so that four types of models exist. Specifically, a *deterministic static model* can be used to model static routing problems in which the relevant information does not change and also the problem is static. A *stochastic static model* represents a routing problem in which uncertain parts in the relevant information are considered by generating a robust tour plan with regard to expected changes in the problem that occur according to available probability distributions. By using a *deterministic dynamic model*, changes which occur in the problem over time are considered but the model does not consider the dynamic structure of the problem. Specifically, when solving the model, the possible occurrence of further changes in the problem is not considered so that only the currently deterministically known parts of the relevant information are considered, even in situations where stochastic knowledge is available. In contrast, in *stochastic dynamic models*, decisions are made on the basis of relevant information known with certainty and do additionally consider available stochastic knowledge about expected future events.
- **Dynamism of the application:** The application defines how the solution approach is applied in order to solve the model. The application is *dynamic* if the

related model is repeatedly solved thereby considering changes in the problem during the execution of the transportation process. In contrast, the application of an approach is *static* if it does not adapt the solution during the transportation process even when the problem has changed.

In order to illustrate examples of different solution approaches according to the utilized model and the application, the authors discuss a routing problem in which a vehicle has to be guided from an origin to a destination location in a road network. The objective is to find the route with the minimum expected travel time. It is assumed that the traffic conditions change over time with uncertainty. Hence, the considered problem is dynamic. Moreover, it is assumed that stochastic knowledge is available in advance by prescribed probability distributions describing the expected changes in the travel time. The authors present different approaches for using a specific model and a specific application for this given dynamic problem:

- **Deterministic static model and static application:** In such an approach, the route is calculated once by using one particular static representation of the problem in the model, for example by using average expected travel times for the roads in the road network. Since the application is static, occurring changes in the problem during the execution of the transportation process are not considered so that no adaptation of the solution is performed. In dynamic routing problems, approaches of this type are likely to result in infeasible tour plans during the execution of the transportation process (see Sect. 2.5.3.2). Therefore, the utilization of this type of approach is only advisable for static routing problems (see Sect. 2.5.3.1.1).
- **Deterministic dynamic model and dynamic application:** In approaches of this type, the model is repeatedly solved during the execution of the transportation process due to the dynamic application. The model considers changes that have already occurred in the problem, but the available stochastic knowledge about future expected changes is not considered. In this book, such approaches are denoted as *reactive real-time control approaches* and *deterministic real-time control approaches* (see Sect. 2.5.3.2.2).
- **Stochastic dynamic model and static application:** These approaches consider changes that have already occurred in the problem as well as future expected changes. However, since the application is static, only one initial solution is computed which is not adapted when changes in the problem occur during the execution of the transportation process. Hence, the aim is to generate an initial tour plan which is robust, i.e., which is likely to represent a good solution during the execution of the transportation process even when the problem changes according to the available prescribed probability distributions. In dynamic routing problems, such a priori optimization-based approaches are denoted as Stochastic Vehicle Routing Problems (see Sect. 2.5.3.2.1).
- **Stochastic dynamic model and dynamic application:** Comparable to the previously described approach type, the model considers changes in the problem that have occurred as well as expected future changes. Moreover, the model is repeatedly solved during the execution of the routing problem so that occurred

and expected changes in the traffic conditions are considered. In this book, approaches of this type are denoted as *pro-active real-time control approaches* (see Sect. 2.5.3.2.2).

2.5.3 A Unified Approach for Classifying Routing Problems with Regard to Characteristics and Processing of Relevant Information

In the previous section, different approaches in the literature for characterizing relevant information and different types of routing problems were described. Based on the results of these approaches, in particular those by Powell et al. (1995) and Psaraftis (1995), we present a unified approach for classifying routing problems according to RDOPG-specific attributes in this section. This approach gives a structured overview about the type of relevant information, i.e., constant or changing, and considers the availability and processing of occurring changes in the relevant information in detail. Moreover, in case of changing relevant information, it describes the type of stochastic knowledge available, if any.

The proposed classification is illustrated in Fig. 2.6. The resulting items of this classification represent individual classes of routing problems. According to the main distinction concerning whether the relevant information is constant or comprises uncertain changes during the transportation process, the type of problem data denoted as the problem is identified. If changing relevant information exists, the availability of resulting changes in the problem is presented. Moreover, the proposed classification considers in which way the problem is processed in order to coordinate the transportation process. Finally, depending on the quality of information, i.e., the type of stochastic knowledge which is available about expected future events, appropriate variants of VRPs known in the literature are described. In what follows, characteristics for each of these classes and general solution approaches are presented.

2.5.3.1 Routing Problems with Constant Relevant Information

In routing problems with *constant relevant information*, all relevant information is exactly known before the execution of the transportation process and does not change during its execution (see Fig. 2.7). In such routing problems, the *evolution of information* is *static* and the constant relevant information is *globally available*. Moreover, a *centralized processing* of this relevant information is used. In doing so, a *centralized coordination* of the transportation process is applied. Since only the initial tour plan generation process is considered in such routing problems, solution approaches make often use of a *static deterministic model* and a *static application*. Since the relevant information is constant and no future changes occur, the *quality of information about future events* is *deterministically known*.

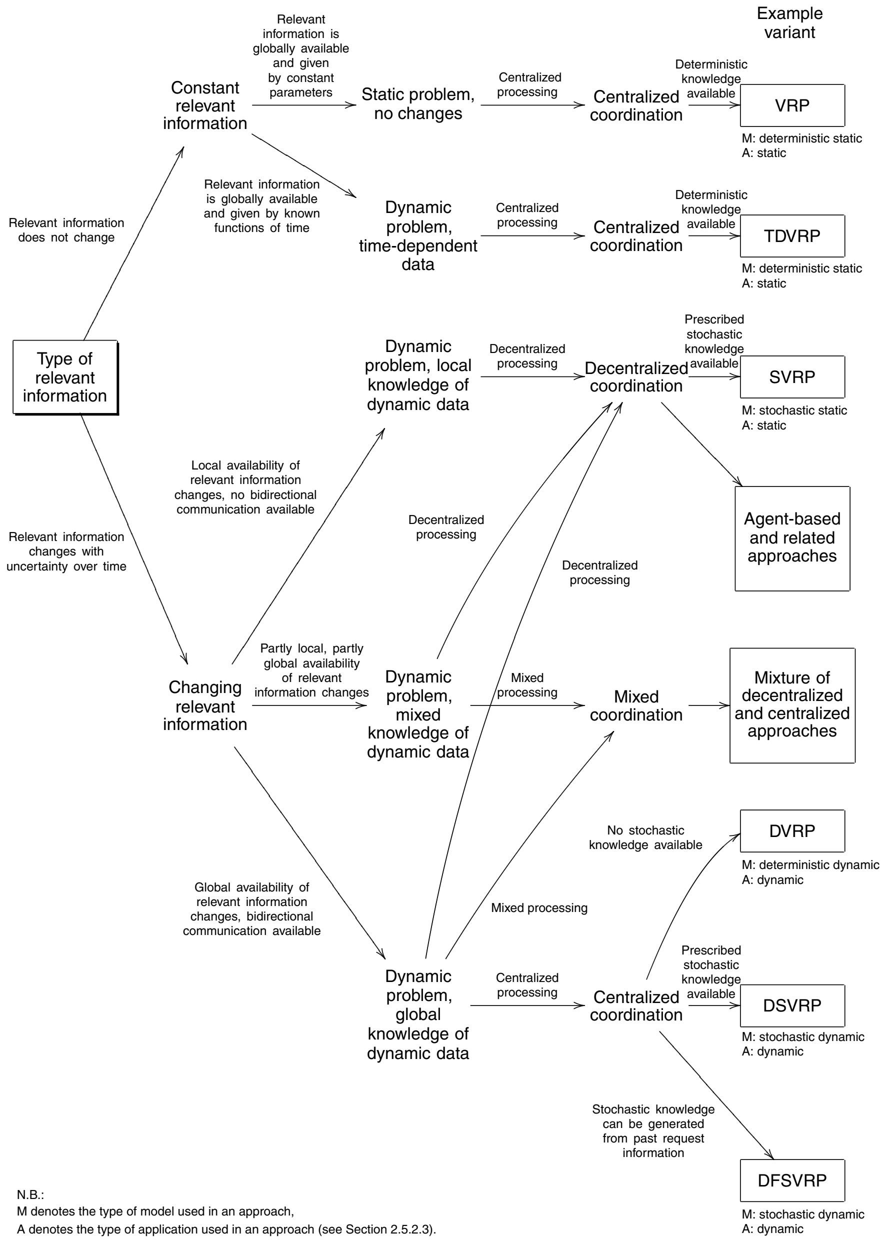


Fig. 2.6 A new classification of routing problems according to properties and processing of relevant information

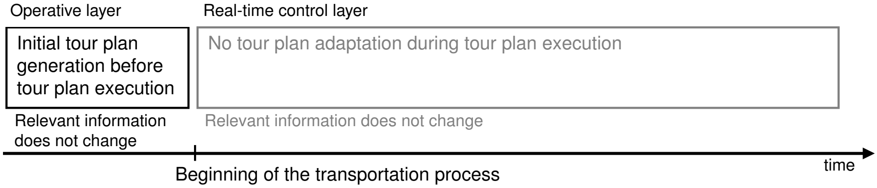


Fig. 2.7 The transportation process of routing problems with constant relevant information

As illustrated in the classification in Fig. 2.6, two types of routing problems with constant relevant information exist. In Powell et al. (1995), they are distinguished according to the dynamism of the problem, i.e., whether the problem is static or dynamic.

2.5.3.1.1 Static Problem: Static VRPs

In routing problems in which the constant relevant information is given by *constant parameters*, the problem data also remains constant during the planning horizon so that a *static problem*, in which *no changes* occur, is present. One example of this routing problem class is the CVRP which was described in Sect. 2.3.2.

2.5.3.1.2 Dynamic Problem: Time-Dependent VRPs

In routing problems of this class, the existing problem data changes during the execution of the transportation process. However, since these changes occur according to known functions of time which are available in the relevant information before the execution of the transportation process and since the relevant information does not change, all problem data changes are deterministically known. Therefore, a *dynamic problem with time-dependent data* is present. An example routing problem is the Vehicle Routing Problem with Time-Dependent Travel Times (TDVRP, see Sect. 2.4.9).

2.5.3.2 Routing Problems with Changing Relevant Information

In addition to the previously discussed routing problems with constant relevant information, there are routing problems in which the relevant information changes during the execution of the transportation process. What is important is that these changes cannot be predicted with certainty. Hence, a *dynamic problem with dynamic data* in which changes in the problem data cannot be predicted with certainty is present. According to Psaraftis, the *evolution of information* is *dynamic* and the *quality of information about future events* is not *deterministic* and depends on the type of available stochastic knowledge.

As mentioned in Sect. 2.5, uncertain changes in the relevant information are brought about by one of the following two circumstances:

1. **Occurrence of dynamic events:** Since the time point of occurrence and the impact of dynamic events cannot be predicted with certainty, dynamic events may lead to an outdated tour plan with regard to the system situation generated by the dynamic event. For example, a dynamic event by which a new request arrives in the system requires an integration of this request into the tour plan. Moreover, dynamic events may render the tour plan infeasible, e.g., if a vehicle breaks down, its assigned requests need to be reassigned to other vehicles. An overview about different types of dynamic events is presented in Sect. 2.6.2.
2. **Presence of incomplete relevant information:** In addition to routing problems in which the relevant information is changed by dynamic events, there are routing problems in which parts of the relevant information are unknown during the initial tour plan generation process. In general, these unknown parts are random variables whose actual values are revealed during the execution of the transportation process. They follow prescribed probability distributions which are usually available in the initial tour plan generation process.

What both circumstances have in common is that the problem data changes with uncertainty during the execution of the transportation process so that specific *adaptations* of the tour plan may become necessary. Therefore, the dimension of the *availability of changes in the relevant information* is important since the resulting availability of changes in the problem mainly determines the types of tour plan coordinations applicable during the execution of the transportation process. In our classification, there are three different values for this dimension:

- **Local availability of relevant information changes:** Local availability of relevant information changes is present if only one particular vehicle is informed about changes that have occurred and there is no possibility of informing a central dispatching center due to the lack of bidirectional communication equipment (see Sect. 2.6.3) or since such a central dispatching center does not exist in the considered routing problem. Due to the *local knowledge of dynamic data*, occurring changes cannot be centrally processed so that a *decentralized processing* of changes and therefore a **decentralized coordination** of the tour plan is inevitable. Hence, the vehicle driver who is informed about the changes is in charge of processing appropriate adaptations of the vehicle's tour. In VRPs, these types of routing problems are represented by variants of the Stochastic Vehicle Routing Problem presented in Sect. 2.5.3.2.1.
- **Global availability of relevant information changes:** In such scenarios, a centralized dispatching center is available and bidirectional communication devices in the vehicles allow communication between the dispatching center and the vehicles (see Sect. 2.6.3). In this case, a *global knowledge of dynamic data* exists in the dispatching center either by centrally collecting changes (e.g., customers directly call the dispatching center) or by vehicles transmitting locally revealed changes to the central dispatching center. This allows for different types of co-

ordination of the transportation process which, in what follows, are explained by the example of adapting the tour plan in case of newly arriving requests:

- **Decentralized coordination:** In this case, some or all of the vehicles are informed about the new requests and autonomously decide which new requests will be serviced by each vehicle; this results in a *decentralized processing* of changes in the relevant information. In our classification, such routing problems are denoted as *agent-based/related approaches*. This type of approaches is not discussed any further in this book. For more information, we refer for example to the work of Fischer et al. (1996), Kohout and Erol (1999), Weiss (2000), Perugini et al. (2003), Cubillos et al. (2007), Bloos et al. (2009), Hülsmann et al. (2009), and Schönberger (2010).
- **Centralized coordination:** By making use of a *centralized processing* of changes in the relevant information, the dispatching center centrally adapts the tour plan and transmits resulting changes to the vehicles involved. In the considered example, the new requests are centrally integrated into the tour plan and the vehicles to which these new requests are assigned are informed after the tour plan adaptation is complete. Routing problems which belong to this class are variants of the Dynamic Vehicle Routing Problem. They are described in detail in Sect. 2.5.3.2.2.
- **Mixed coordination:** In systems applying a mixed tour plan coordination, some types of changes in the relevant information are *centrally processed*, e.g., newly arriving requests, while other changes are *locally processed*, e.g., rerouting of a vehicle in case of traffic congestions. Such approaches can be realized by utilizing a combination of the methodologies for decentralized and centralized tour plan coordination.

• **Partly local, partly global availability of relevant information changes:** This scenario is present if specific types of changes in the relevant information are globally available and other types are only locally available. In such scenarios, tour plan adaptations can be realized either by using a **decentralized coordination** or a **mixed coordination**.

In what follows, we present variants of Stochastic Vehicle Routing Problems which utilize a decentralized coordination and variants of dynamic VRPs using a centralized coordination of the transportation process. Furthermore, the type of stochastic knowledge which is available in the individual routing problems is described.

2.5.3.2.1 Decentralized Coordination: Stochastic Vehicle Routing Problems

In Stochastic Vehicle Routing Problems (SVRP), the relevant information available during the initial tour plan generation process is incomplete so that some parts are unknown (as mentioned in aspect 2 in Sect. 2.5.3.2). In SVRPs, such unknown relevant information can comprise uncertain servicing of customers, uncertain customer demands, stochastic travel times, or stochastic service times (cf. Cordeau et al. 2007a). As mentioned above, an important aspect is that these unknown parts are

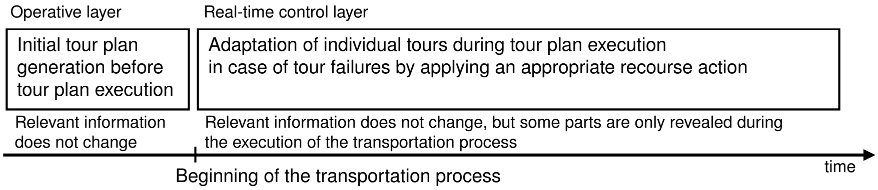


Fig. 2.8 The transportation process in Stochastic Vehicle Routing Problems

represented by random variables. Their values are unknown in the initial tour plan generation but do not change during the execution of the transportation process. They are revealed to individual vehicles during the transportation process which is discussed in detail below. Moreover, these random variables follow prescribed probability distributions which are known and can be utilized as stochastic knowledge during the initial tour plan generation. With regard to the classification of Psaraftis (1995), the *quality of information about future events follows prescribed probability distributions (probabilistic information)*. Hence, the unknown parts of the relevant information are given by distributions known to be correct with certainty (cf. Richter 2005, p. 65).

The transportation process in SVRPs is illustrated in Fig. 2.8. A decentralized coordination is applied during the execution of the transportation process since no technical equipment for a bidirectional communication is available. The initial tour plan is centrally generated irrespective of this lack of technical equipment since all vehicles are assumed to be located at the depot before the execution of the transportation process. During its execution, the individual vehicles can be required to adapt their tour when uncertain relevant information is revealed. Note that such an adaptation is usually only performed if changes in the relevant information make a tour infeasible; this is denoted as a *tour failure*. For example, this happens when the demand of a customer exceeds the currently available capacity of the assigned vehicle.

In case of a tour failure, adaptations denoted as *recourse actions* or *corrective actions* are required in order to re-establish a feasible tour (cf. Gendreau et al. 1996). Due to the decentralized coordination, recourse actions are limited to local adaptations of the vehicle's tour. Moreover, since often no computational equipment is assumed to be available in vehicles, local adaptations are often carried out by utilizing predefined *policy-based operations* consisting of specific rules which can be directly applied by the drivers. Due to these restrictions, recourse actions often result in suboptimal tour plan adaptations which generate high additional costs and therefore lead to a significantly reduced solution quality. Since the minimization of operational costs is usually pursued in SVRPs, the reduction of costs generated by recourse actions during the execution of the transportation process is desired. This is achieved by including the available stochastic knowledge in the initial tour plan generation process which is carried out by sophisticated solution methods. Due to the fact that only policy-based tour plan adaptations are performed during the execution of the transportation process, SVRPs are also denoted as *a priori optimization problems* (cf. Powell et al. 1995).

In order to integrate the available stochastic knowledge into the initial tour plan generation process, SVRPs are often formulated as *stochastic programs*. According to Dror (1989), stochastic programming techniques can be distinguished into *wait and see* as well as *here and now* approaches. While in the first type of approaches a solution is calculated *after* the disclosure of some uncertain relevant information, approaches of the second type compute a solution *before* the uncertain relevant information is revealed. Since the a priori solution of an SVRP is generated before unknown information is revealed during the execution of the transportation process, SVRPs belong to the second type of approaches. For SVRPs, two main solution techniques exist in the literature (cf. Gendreau et al. 1996):

- **Chance Constrained Program (CCP):** In this type of solution technique, the aim is to generate the initial tour plan so that the *chance* of a tour failure during the execution of the transportation process is below a defined threshold. A drawback is that CCPs neglect the costs of a recourse action that are generated when a tour failure nonetheless occurs during the execution of the transportation process.
- **Stochastic Program with Recourse (SPR):** This type of solution technique pursues the generation of an a priori tour plan P^{a} with minimum total expected costs $c_1(P^{\mathrm{a}}) + c_2(P^{\mathrm{a}})$. In this formulation, $c_1(P^{\mathrm{a}})$ describes the costs which occur in the tour plan according to the relevant information known with certainty. Moreover, utilizing the available stochastic knowledge, $c_2(P^{\mathrm{a}})$ represents the estimated costs of recourse actions which are expected to be required during the execution of the transportation process. In general, SPRs require more complex solution methods than CCPs but the considered objective function has a higher practical relevance since it explicitly considers expected additional costs.

As mentioned above, different recourse actions can be performed in case of a tour failure. The best choice for a recourse action depends on the way in which the uncertain relevant information is revealed during the execution of the transportation process. For example, according to Gendreau et al. (1996), a customer's demand can become known when service of the previous customer has been completed or when the vehicle has arrived at the corresponding customer request location. Moreover, applicable recourse actions also depend on the type of uncertainty in the relevant information which is present in the considered SVRP. In the literature, the following types of uncertainty are conducted in SVRPs (cf. Gendreau et al. 1996, Toth and Vigo 2002, p. 332, and Cordeau et al. 2007b, p. 45 et seq.):

- **Vehicle Routing Problem with Stochastic Demands (VRPSD):** In VRPSDs, the demands of customer requests are unknown. Therefore, infeasible tour plans can occur due to capacity constraints of vehicles. This SVRP variant was first studied by Tillman (1969) and Golden and Stewart Jr. (1978) and is by far the most studied variant of the SVRP. More studies on the VRPSD can be found among others in the work of Dror (1989), Secomandi (2000), Laporte et al. (2002), Novoa et al. (2006), Christiansen and Lysgaard (2007), Mendoza et al. (2009), and Secomandi and Margot (2009).
- **Vehicle Routing Problem with Stochastic Customers (VRPSC):** Each customer request $i \in R$ has an occurrence probability p_i and therefore requires no

service with a probability of $1 - p_i$. If more customers than originally expected demand service, infeasible tour plans can arise due to capacity constraints of vehicles or time window violations. This variant has been first studied by Jézéquel (1985) and Jaillet (1987). Note that the set of potential customer requests is known in advance. This is in contrast to Dynamic Vehicle Routing Problems considered in Sect. 2.5.3.2.2 in which new customer requests can theoretically occur anywhere in the considered service region as well as at unknown time points.

- **Vehicle Routing Problem with Stochastic Travel Times (VRPSTT):** In this SVRP variant, travel times between the relevant locations are given by random variables which can lead to unfavorable solutions. For example, in Lambert et al. (1993) a VRPSTT is considered in which a late service of requests results in penalty costs.
- **Vehicle Routing Problem with Stochastic Service Times (VRPSST):** In this variant of the SVRP, the time required for servicing customer requests is uncertain. If a vehicle is required to spend more time at a customer request location than planned, postponed service of consecutive requests can occur which can lead to violations of time window constraints and penalty costs. The VRPSST is considered among others in Laporte et al. (1992), Hadjiconstantinou and Roberts (2002), and Verweij et al. (2003).

Moreover, combinations of the described types of uncertainty in the relevant information are possible. A frequently studied variant is the Vehicle Routing Problem with Stochastic Customers and Demands (VRPSCD) which is also known as the Probabilistic Vehicle Routing Problem (PVRP, cf. Larsen 2000, p. 28). It was first studied by Jézéquel (1985) and Jaillet (1987). Another SVRP variant which considers two types of uncertainty denoted as the Stochastic Vehicle Routing Problem with Stochastic Travel Times and Stochastic Service Times is analyzed in the work of Laporte et al. (1992) and Kenyon and Morton (2003).

As mentioned above, the VRPSD is the most studied SVRP variant in the literature. Hence, various policy-based recourse strategies have been developed for the VRPSD. In what follows, some of them are described. One of the first recourse strategies proposed for the VRPSD lets a vehicle return to the depot whenever the request assigned next in a vehicle's tour cannot be serviced since the currently remaining vehicle capacity is too low. Laporte and Louveaux (1990) and Dror et al. (1993) describe other recourse strategies in which preventive breaks are performed at strategic positions in the tour. For example, when a vehicle is near the depot, it delivers the currently loaded goods there. Another policy-based recourse strategy is presented by Novoa et al. (2006). In this strategy, more flexibility is integrated by allowing vehicles to postpone the next assigned request and to service other requests first before returning to the depot if it is likely that those other requests can be serviced with the currently remaining capacity. Further recourse strategies for the VRPSD are discussed in Dror (1989), Gendreau et al. (1996), and Cordeau et al. (2007b).

Besides policy-based recourse actions, more complex recourse strategies are based on a reoptimization, i.e., adaptation, of the remaining tour of each vehicle when a tour failure occurs. As an advantage, this can improve the system performance. However, such recourse strategies complicate the estimation of costs for expected recourse actions during the initial tour plan generation (cf. Gendreau et al. 1996). In the work of Dror (1989), the author introduces a formulation based on a Markov Decision Process (MDP, cf. Puterman 2005) for solving a single-vehicle VRPSD. In this approach, the solution is reoptimized whenever the vehicle reaches the next request where its demand is revealed. At each customer request, the decision is made whether to service the request now or to service it later. In the work of Dror (2005), the author states that reoptimization-based approaches represent the most promising approaches for solving VRPSDs. According to Secomandi and Margot (2009), recourse actions should be divided into reactive actions which are only taken when a tour failure occurs and pro-active actions in which decisions are taken by anticipating the likelihood of a tour failure. Furthermore, they state that the opportunities given by technological advances are especially beneficial to reoptimization-based approaches.

2.5.3.2.2 Centralized Coordination: Dynamic Vehicle Routing Problems

The application of a centralized coordination of the transportation process during its execution enables the utilization of *real-time control approaches*. Real-time control approaches are able to monitor the execution of the transportation process and perform tour plan adaptations simultaneous to its execution (cf. Bock 2010). Figure 2.9 illustrates the coordination of the transportation process in dynamic vehicle routing problems. As can be seen, the tour plan currently in execution is required to be adapted during its execution. Such adaptation activities belong to the real-time control layer of the transportation process (see Sect. 1.2). The utilization of real-time control approaches allows the handling of all types of dynamic events which are described in Sect. 2.6.2.

An important factor is the decision *when* an adaptation of the tour plan shall be performed. Specifically, it is necessary to define how the complete dynamic problem is divided into subproblems which are solved during the planning horizon (cf. Bianchi 2000). For example, the tour plan can be adapted whenever a new dynamic event arrives in the system. This is denoted as single event optimization (SEO, cf. Bianchi 2000). Another possibility is to adapt the transportation process in specific time intervals (cf. Bock 2010). Further characteristics of real-time control approaches will be discussed in Sects. 4.7.4 and 5.1. More information on the implemented real-time control approach will be given in Sect. 5.2.

In the literature, different variants of dynamic vehicle routing problems centrally coordinated by real-time control approaches can be found. They can be distinguished according to the type of available stochastic knowledge, i.e., the quality of information about future events (see Sect. 2.5.2.1):

1. **Stochastic knowledge is unavailable:** If no stochastic knowledge is available, real-time control approaches can only react after new dynamic events have arrived in the system during the execution of the transportation process. In the

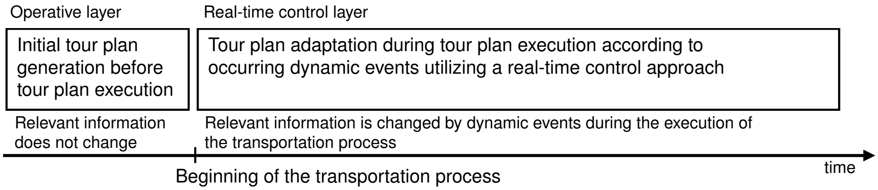


Fig. 2.9 The transportation process in dynamic vehicle routing problems with a centralized coordination

literature, such routing problems are classified as variants of the Dynamic Vehicle Routing Problem (DVRP, cf. Psaraftis 1988).

2. **Stochastic knowledge is probabilistic:** In this case, stochastic knowledge is available and *follows prescribed probability distributions* which usually describe all uncertain elements of the relevant information. Since occurring dynamic events change the relevant information according to the known probability distributions, the utilization of this type of stochastic knowledge is likely to provide significant improvements on the solution quality. In the literature, these approaches are often denoted as variants of the Dynamic Stochastic Vehicle Routing Problem (DSVRP, cf. Hvattum et al. 2006 and Ichoua et al. 2006).
3. **Stochastic knowledge is forecasted:** In many practical applications, stochastic knowledge is not directly available. However, if historical data is available, forecasted stochastic knowledge can be generated by analyzing this data. For example, by applying sophisticated methods on request arrivals of elapsed days including arrival time and location, stochastic knowledge can be generated out of this historical data. Such historical data is denoted as *past request information* in what follows. Since the stochastic knowledge is forecasted, in this book, we denote such problem scenarios as variants of a *Dynamic Forecasted Stochastic Vehicle Routing Problem* (DFSVRP) to which also the considered RDOPG applications belong.

   Note that in routing problems which forecast stochastic knowledge it is assumed that common characteristics observed in the past provide good predictions of future events. Hence, there exist situations in which future dynamic events may possess significantly different attributes as predicted. Since the stochastic knowledge is *known with uncertainty* only, in the worst case, the utilization of forecasted stochastic knowledge can lead to an inferior system performance compared to applying a deterministic real-time control approach which only considers information known with certainty. Moreover, it is advantageous if the request data of the considered routing application possesses specific characteristics. We denote them as *structural quality* of a request data set in what follows. Hence, in practice, it is advantageous to analyze existing real-world request data according to their suitability with regard to generating stochastic knowledge which can be advantageously applied. We will discuss this further in the computational experiments in Chap. 8.

With regard to Powell et al. (1995), in the first case, a *deterministic dynamic model* is utilized. Specifically, it is dynamic because it is updated during the execution of the transportation process but since it only considers changes which have already occurred, it is deterministic. In the latter two cases, the utilized model is a *stochastic dynamic model* since additional stochastic knowledge about expected future events is utilized.

As mentioned in Sect. 2.5.3.2.1, tour plan adaptations required by unexpected changes in the relevant information may lead to significant changes in the currently executed tour plan. Although the discussed real-time control approaches are able to perform a centralized adaptation of the tour plan by elaborated solution methods, dynamic events can nonetheless cause high additional costs. With regard to RDOPG applications in which new customer requests arrive during the execution of the transportation process, the arrival of many unexpected customer requests may require the dispatching of additional vehicles. If no additional vehicles are available (see aspect 11 of Psaraftis' approach in Sect. 2.5.2.2), vehicle tours are likely to require significant adaptations. This negatively influences the execution of the transportation process but also results in higher travel distances and longer request response times which often leads to a lower solution quality. Therefore, the aim is to coordinate the transportation process by real-time control approaches in a way that provides an increased level of *tour plan flexibility*. In doing so, tour plans provide more possibilities for adapting the transportation process according to changes in the relevant information that have occurred without causing significant distortions. This increased flexibility can also be regarded as a technique for increasing the robustness of tour plans towards uncertain dynamic events.

Depending on the availability of stochastic knowledge and the pursued objective, two types of approaches are useful in providing an increased level of tour plan flexibility:

- **Flexibility-increasing approaches without using stochastic knowledge:** Tour plan flexibility can be increased by least-commitment strategies such as *waiting strategies*. By postponing decisions to a later point in time, it is likely that better decisions with regard to travel distance can be made since more relevant information is available then. In what follows, such real-time control approaches are denoted as *deterministic real-time control approaches*.
- **Flexibility-increasing approaches using stochastic knowledge:** Stochastic knowledge offers additional possibilities for realizing the aforementioned waiting strategies. Moreover, waiting strategies can be extended to *relocation strategies* by actively repositioning idle vehicles to subregions in which future requests are likely to arrive soon. Furthermore, *request assignment strategies* can be used for efficiently coordinating utilized vehicles. Specifically, by estimating future vehicle positions resulting from specific request assignments, expected impacts, e.g., request response times, can be calculated for expected future requests. This information can be directly utilized for evaluating the solution quality of possible adaptations during the transportation process. Such real-time control approaches are denoted as *pro-active real-time control approaches* in this book.

In theory, stochastic knowledge is available for all types of dynamic events that have been presented in Sect. 2.6.2. However, in real-world applications, some dynamic events can be predicted better than others. For example, since in many real-world applications recurring arrival patterns of requests can be observed, the arrival of new requests has been very often considered in approaches in the literature. Besides, dynamism arising from changes in travel times is nowadays also well predictable due to floating car data (FCD, cf. Ehmke et al. 2009, 2010). Furthermore, traffic jams are automatically detectable by traffic sensors which are commonly installed at highways and major roads nowadays (cf. Ghiani et al. 2003 and Attanasio et al. 2007). In contrast to both these sources of dynamism, predicting the location and time of vehicle breakdowns is rather impossible. Hence, in the literature, most research deals with dynamic events arising from new request arrivals or changing travel times.

Real-time control approaches which do not provide an increased tour plan flexibility only react towards newly arriving dynamic events. Hence, we denote such approaches as *reactive real-time control approaches*. In the literature review, we will mainly focus on approaches which provide an increased flexibility in the transportation process. Specifically, details on approaches utilizing waiting strategies will be discussed in Sect. 4.3.1. Moreover, relocation strategies will be presented in Sect. 4.3.2 and request assignment strategies will be described in Sect. 4.3.3.

2.6 Dynamic Routing Problems Which Utilize a Centralized Coordination

In the previous section, different variants of routing problems and their characteristics were presented. In accordance with the considered RDOPG applications, this section deals with specific dynamic routing problems in detail, in particular those with a centralized coordination of the transportation process utilizing real-time control approaches (see Sect. 2.5.3.2.2). First, *typical objectives in dynamic routing problems* are described. Afterwards, *dynamic events in dynamic routing problems* by which the relevant information is changed are presented. In addition, *technologies utilized in real-time control approaches* are introduced. In order to classify dynamic routing problems according to their difficulty level with regard to finding good solutions, a measure denoted in the literature as the *degree of dynamism* is described. Based on this measure, practical dynamic routing problems are introduced using a *three-echelon classification of dynamic routing problems*. Finally, two measures for *evaluating the performance of solution approaches for dynamic routing problems* are presented.

2.6.1 Typical Objectives in Dynamic Routing Problems

In general, three different objectives (or combinations of these objectives) are pursued in dynamic routing problems (cf. Larsen et al. 2007):

- **Minimization of distribution costs:** The minimization of costs which are generated during the execution of the transportation process (see Sects. 2.4.12.1 and 2.4.12.2) is the most widely used objective in static routing problems. Furthermore, this objective plays an important role in specific dynamic routing problems, in particular those in which the transportation process is less time-critical and only a small part of the relevant information changes during the execution of the transportation process.
- **Maximization of service quality:** If a high quality of customer service is pursued, dynamically arriving customer requests are required to be quickly serviced. As described in Chap. 1, high request response times lead to customer inconvenience and may impose significantly high costs on the service company, for example due to penalty costs incurred by SLA contracts. This type of objective is considered in the RDOPG applications which are investigated in this book.
- **Maximization of system throughput:** This objective is applied in dynamic routing problems in which the service of as many customers as possible within a given planning horizon is pursued. In such applications, it may happen that the service of an isolated customer request is postponed for a long time in favor of servicing a larger amount of other requests which are closely located to each other. Hence, as mentioned by Psaraftis in aspect 8 of his approach in Sect. 2.5.2.2, it is reasonable to integrate methods which ensure that no customer is postponed indefinitely.

Note that the presented objectives cannot be fulfilled to their full extent at the same time. For example, tour plans which are generated according to the objective of maximizing quality of customer service usually produce additional travel activities during the day which makes them suboptimal from a distribution costs perspective. This has also been addressed by Richter (2005, p. 200 et seq. and p. 224). In Sect. 2.6.5, we present various real-world routing problems and their pursued objective.

2.6.2 Dynamic Events in Dynamic Routing Problems

In dynamic routing problems, dynamic events change the relevant information during the execution of the transportation process. In general, the following four types of dynamic events are considered in the literature:

- **New requests:** Each newly arriving request has a location in the service area. Depending on the considered routing problem, a new request comprises additional attributes such as demands, time windows, lateness penalties, and service times. With the arrival of a new request, an adaptation of the tour plan in order to integrate this request is required. In some applications, it may be allowed to reject new requests, often at additional penalty costs. One approach in which this applies is the approach of Yang et al. (2004). Such a rejection may be beneficial if, at request arrival, it can already be determined that a request cannot be feasibly serviced (e.g., due to violations of hard time window constraints) or if the costs of servicing the request are higher than the costs of its rejection. However, in applications which aim at maximizing quality of service, for example measured by the number of serviced customers or the minimization of request response times, all arriving requests usually need to be serviced. This is considered among others in Bent and van Hentenryck (2004b) and Kleywegt et al. (2009). According to Pillac et al. (2010), the arrival of new requests is the most studied type of dynamic events in the literature.

- **Changes in request attributes:** In some dynamic routing problems, attributes of known requests change or, similar to SVRPs, are only revealed when the vehicle reaches the customer location. For example, if the demand of a request turns out to be different from what was expected, the capacity of the assigned vehicle may not be sufficient. In this case, the request needs to be reassigned to another vehicle. This type of dynamism is conducted in Wu et al. (2004) and Hvattum et al. (2007).
- **Traffic congestions and road blockages:** This type of dynamic events causes either increased travel times or a complete blockage of affected roads. The length of such impacts may be known or unknown. Since the service of requests can be delayed by such dynamic events, re-routing of vehicles to non-congested roads or the reassignment of requests to other vehicles may be advantageous. This type of dynamic event is considered in Fleischmann et al. (2004b), Ferrucci (2006, p. 44), Attanasio et al. (2007), Barceló et al. (2007), Haghani and Yang (2007), Cortés et al. (2008), and Bock (2010).
- **Vehicle disturbances:** This results in a decreased efficiency or a complete failure of a vehicle due to a breakdown or a driver drop-out. As a consequence, pending requests may be required to be reassigned to other vehicles. Moreover, if the affected vehicle is transporting requests when a complete failure occurs, its loaded requests need to be picked up by other vehicles at the affected vehicle's current location. This type of dynamic events is studied in Ferrucci (2006, pp. 44–45), Li et al. (2009a, 2009b), Bock (2010), and Mu et al. (2010).

2.6.3 Technologies Utilized in Real-Time Control Approaches

Approximately 20 years ago, technical limitations often prohibited the application of real-time control approaches in transportation processes. In addition to new developments in computational technology, advances in communication and positioning technology which have been achieved in the last two decades also support the practicability of real-time control approaches.

In this section, we describe technologies which are used in on-board units of vehicles and those utilized in computer systems in dispatching centers in order to enable the coordination of the transportation process in real-time. After describing these technologies separately, we present an exemplary implementation of a system establishing a bidirectional communication between the vehicles and the dispatching center. Moreover, we present an exemplary architecture of a real-time control approach which makes use of the previously described bidirectional communication system. In doing so, it is illustrated how the presented technologies can be jointly used in order to efficiently realize real-time control approaches in real-world dynamic routing problems.

2.6.3.1 Recent Advances in On-Board Technologies

In order to reasonably adapt the tour plan during its execution, the current vehicle positions are required to be provided to the dispatching center. In particular, if positions of traveling vehicles are available in real-time at the dispatching center, a newly arriving request which is close to a vehicle can be directly serviced by this vehicle. In doing so, the required detour and the request response time can be reduced, compared to returning to this request later. For this purpose, on-board systems are utilized.

On-board systems allow the transmission of different types of vehicle data, such as its current position, to the dispatching center in real-time during the execution of the transportation process. The automatic determination of the current vehicles' positions and velocities became possible with the introduction of the Global Positioning System (GPS) in 1995. It consists of 24 satellites which are arranged in 6 orbital planes with 4 satellites each (cf. Kaplan and Hegarty 2006, p. 3). These satellites broadcast the current time as well as their current position which can be received and further used by GPS-capable devices. By utilizing the transmitted data of individual GPS satellites, GPS devices can determine their current position with an accuracy of a few meters using the GPS operation mode called Standard Positioning Service (SPS) which provides sufficient precision for routing applications. SPS is free of charge for commercial and non-commercial use and is utilized by millions of GPS devices worldwide (cf. Kaplan and Hegarty 2006, pp. 3–4). For more information on GPS and other satellite-assisted positioning systems such as GALILEO, see Kaplan and Hegarty (2006).

Besides GPS-capable on-board vehicle units, a data transmission technology is required for establishing an electronic communication between the vehicles and the dispatching center. In the last ten years, significant technological advances have emerged in this area. According to Larsen (2000, p. 12 et seq.), in the year 2000, dedicated radio based communication systems were favorable compared to mobile phone communication systems which had low setup costs but high operational costs. Over the years, alternative tariff models have made electronic communication via smartphones much more economical. In the year 2011, mobile phone contracts providing Internet access at a flat rate are available for less than 30 euros per month in Germany (cf. Telefónica Germany GmbH & Co. OHG 2011). Moreover, data transfer rates in mobile phone networks has evolved from GSM

(Global System for Mobile Communications, before Groupe Spécial Mobile) to GRPS/EDGE (General Packet Radio Service and Enhanced Data Rates for GSM Evolution) to UMTS (Universal Mobile Telecommunications System). UMTS provides data transfer rates with HSDPA (High Speed Downlink Packet Access) and HSUPA (High Speed Uplink Packet Access). These are comparable to fixed line Internet connections (cf. Halonen 2003, p. xxvi). The use of mobile Internet-based data transmission technologies in combination with low priced but powerful GPS-enabled smartphones and other mobile devices allows for the continuous automated transmission of current vehicles' positions to the central dispatching center in real-time at reasonable costs.

Moreover, vehicle on-board units can be used to provide information about current activities to the dispatching center during the execution of the transportation process. For example, when a service worker is at a customer request, information about the progress of the service activity can be efficiently transmitted. If unexpected difficulties occur during a service activity, the dispatching center can identify delays in scheduled tour plan activities early and can perform appropriate adaptations in time. This is especially advantageous in applications pursuing customer service quality aspects in which service times required at customer locations are uncertain. Furthermore, on-board units can provide a request management system which displays changes in the vehicle's tour to the driver and allows the driver to communicate with the dispatching center, e.g., confirm changes in the tour. Moreover, request management systems can be combined with integrated navigation systems which automatically update driving route instructions in case of a tour plan adaptation or when traffic conditions change. Note that on-board units can exist in different variants, for example installed units, GPS-capable smartphones, or tablet computers with a mobile Internet connection.

2.6.3.2 Recent Advances in Dispatching Center Technologies

Besides the previously mentioned on-board technologies which allow for the tracking and monitoring of vehicle activities in real-time, there has been an advancement in the technologies of dispatching centers in recent years. Specifically, the coordination of vehicles in real-time has become possible due to the development of sophisticated real-time control approaches which make use of elaborated solution methods providing a high solution quality within a short amount of computational time (see Sect. 4.7.4 and Chap. 5). Hence, tour plan adaptations can be carried out quickly after dynamic events have occurred in the system. This allows real-time control approaches to coordinate the transportation process under real-world restrictions.

In order to efficiently handle occurring dynamic events in a real-time control approach, an electronic transmission of unexpected changes in the relevant information (see Sect. 2.5) is advisable. For dynamically arriving customer requests, this can be realized by electronically recording new customer requests in a call center (see also Fig. 1.1). Besides day- and time-dependent travel time profiles, different information sources like traffic flow sensors automatically provide updated

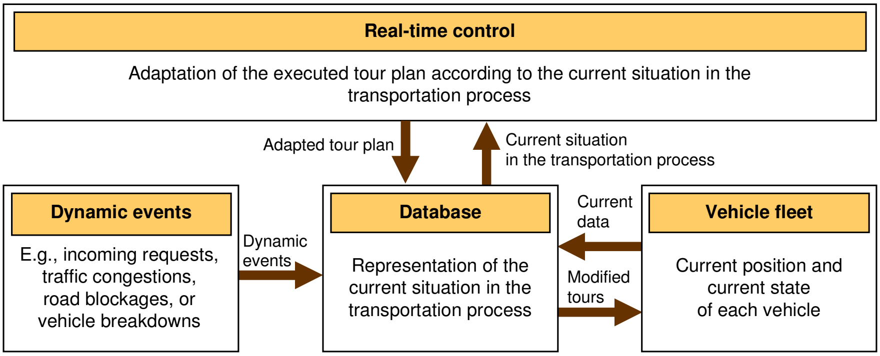


Fig. 2.10 Example architecture of a real-time control approach (cf. Bock 2010)

travel times on roads and highways. In order to quickly calculate travel distances and travel times between considered relevant locations and vehicles, the utilization of efficient shortest path approaches in combination with a Geographic Information System (GIS) which provides road network data is advantageous. Apart from commercially available road networks, OpenStreetMap is a comprehensive community-based road network data project to which users can make contributions (cf. Haklay and Weber 2008, OpenStreetMap 2011). For more information about existing approaches in the literature which make use of GIS systems, see also Sect. 4.7.5 in the literature review.

2.6.3.3 Interplay of Utilized Technologies and Information Flow in Real-Time Control Approaches

The joint use of the described technologies allows for the efficient application of real-time control approaches. In Fig. 2.10, an exemplary architecture for a real-time control approach is presented (cf. Bock 2010). In this architecture, the current system situation is stored in a central database which is supplied with new information by two systems. The first system transmits changes in the relevant information arising from dynamic events (see Sect. 2.6.2) while the second one establishes the bidirectional communication between the dispatching center and the vehicles. It provides current information about the vehicles to the database, e.g., their positions and the current activities of the service workers. Moreover, by using this system, vehicles are informed when changes in their tour have been realized by the real-time control. The real-time control adapts the current tour plan by transforming the current system situation from the database into the considered problem data. After the tour plan adaptation has been completed, the generated tour plan is stored in the database and changes in the tours are transmitted to the vehicles.

In order to realize the bidirectional communication system, Larsen (2000, p. 14) describes a communication architecture based on an Internet data connection. As illustrated in Fig. 2.11, in this architecture, vehicles are equipped with the previously

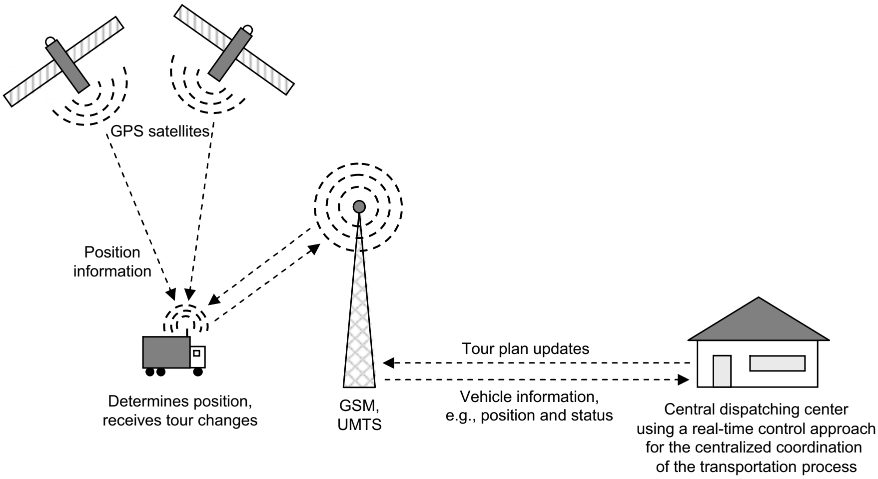


Fig. 2.11 Example architecture for a bidirectional communication system between the vehicles and the central dispatching center (cf. Larsen 2000, p. 14)

described GPS-enabled on-board units. Using this data connection, position and status are sent from the vehicles to the central dispatching center and changes in the tour plan are transmitted to the vehicles.

To sum up, it can be stated that using a real-time control approach by utilizing the described systems can significantly support dispatchers by providing decision support for tour plan adaptations. Specifically, by assessing various tour plan adaptation possibilities in a short amount of time and by enabling vehicle en-route diversion activities, the solution quality can be significantly increased (see also Sect. 8.5). Detailed information about real-time control approaches will be given in the literature review in Sect. 4.7.4 and in Chap. 5. Moreover, further research work on the utilization of communication technologies in dynamic routing problems will be described in the literature review in Sect. 4.7.5.

2.6.4 The Degree of Dynamism

Each dynamic routing problem comprises a specific difficulty in coordinating the corresponding transportation process with regard to the considered objective. We denote this as the *difficulty level* in what follows. This difficulty level depends on many factors. For example, the type of dynamic events considered in a dynamic routing problem and the considered objective are crucial attributes. Moreover, various instances of the same dynamic routing problem can also have a significantly different difficulty level. Since a higher difficulty level makes it more complex to attain a highly qualitative coordination of the transportation process, this imposes higher requirements on the solution method utilized for adapting the tour plan. Note

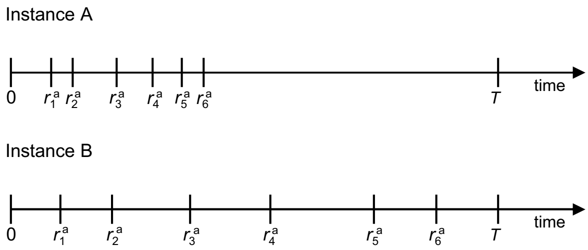


Fig. 2.12 Different arrival times of dynamic requests in two instances of a dynamic routing problem with the same degree of dynamism (cf. Larsen 2000, p. 57)

that awareness of this difficulty level is beneficial. For example, this information can be used for assessing the performance of a solution method when applied to different problem instances comprising individual difficulty levels.

In the literature, an approach for determining such differences in difficulty levels arising in dynamic routing problems is the *degree of dynamism (dod)*. This measure considers dynamic routing problems in which dynamism arises by newly arriving requests which is, as mentioned in Sect. 2.6.2, the most studied type of dynamic events in the literature. The degree of dynamism exists in three versions. Its basic version for dynamic routing problems without time windows was introduced by Lund et al. (1996). For defining the *dod*, it is assumed that the execution of the transportation process starts at time point 0 and new requests can arrive up to time point T. Furthermore, n_{stat} denotes the number of requests which are known during the initial tour plan generation process while n_{dyn} defines the number of dynamically arriving requests. With $n_{\text{all}} = n_{\text{stat}} + n_{\text{dyn}}$, the *dod* is defined as:

$$dod = \frac{n_{\text{dyn}}}{n_{\text{all}}}$$

The *dod* ranges from 0 to 1 where instances with a higher difficulty level have a higher *dod*. Specifically, if all requests are known in the initial tour plan generation process, it holds that $dod = 0$ indicating a static routing problem. In contrast, if all requests just arrive during the execution of the transportation process, it holds that $dod = 1$ so that a completely dynamic routing problem is present. This points to the fact that an instance of a dynamic routing problem in which only a small fraction of all requests dynamically arrives can be considered to be less challenging and easier to solve than an instance of the same dynamic routing problem in which all customer requests dynamically arrive. In what follows, we use the term *problems which are easier to solve* which refers to problems that make it easier to attain a solution with a good solution quality in reasonable time. Clearly, the opposite is true for problems which are more difficult to solve.

Despite its simplicity, other characteristics which also have a significant impact on the difficulty level are not considered in the *dod*. For example, it does not consider the time points at which dynamic requests arrive in the system. As illustrated in Fig. 2.12, two instances with significantly different request arrival properties can inhibit the same *dod*. Specifically, if the minimization of travel distance is the pursued objective, instance A is easier to solve than instance B since all dynamic requests arrive early in instance A (cf. Larsen 2000, p. 57). Hence, after the arrival of the last request, the remaining routing problem is a static problem. Since at this point in time it is likely that many other requests are still pending in instance A, a tour plan with a high solution quality can be generated according to this large amount of still changeable decisions. In contrast, in instance B, additional changes in the relevant information are also revealed late. Hence, at this point in time it is likely that the majority of requests have already been serviced.

In order to consider the mentioned differences in request arrival times, Larsen (2000, p. 58) extends the degree of dynamism to the *effective degree of dynamism* denoted as *edod*. In the *edod*, each dynamic request $i \in \{1, \ldots, n_{\text{dyn}}\}$ arrives at time point r_i^{a} in the system. The *edod* is then defined as follows:

$$edod = \frac{\sum_{i=1}^{n_{\text{dyn}}} (\frac{r_i^{\text{a}}}{T})}{n_{\text{all}}}$$

The effective degree of dynamism represents the ratio of the arrival times of dynamic requests to the latest possible time point T at which they could arrive. In accordance with the previously mentioned objective of minimizing travel distances, if dynamic requests arrive early, the *edod* has a low value while later request arrivals result in higher *edod* values. Similar to the *dod*, the *edod* is a measure between 0 and 1 in which larger values indicate problem instances which are more difficult to solve.

If time windows are considered in dynamic routing problems, it is reasonable to consider them in the calculation of the degree of dynamism since tight hard time windows limit dispatching possibilities making it difficult to attain good solutions. In case of soft time windows, this leads to less possibilities for servicing requests without penalty costs compared to problem instances with relaxed request time window constraints. If the objective of distance minimization is considered, requests can be fulfilled in more distance-optimized ways. Therefore, such problem instances can be considered to be easier to solve than those with tight time windows. In order to consider such time window characteristics of dynamic routing problems, Larsen (2000, p. 59) proposes a third version of the degree of dynamism denoted as the *effective degree of dynamism with time windows* ($edod_{\text{tw}}$). It is defined as

$$edod_{\text{tw}} = \frac{1}{n_{\text{all}}} \sum_{i=1}^{n_{\text{dyn}}} \left(\frac{T - (r_i^{\text{l}} - r_i^{\text{a}})}{T} \right) = \frac{1}{n_{\text{all}}} \sum_{i=1}^{n_{\text{dyn}}} \left(1 - \frac{r_i^{\text{re}}}{T} \right)$$

where $r_i^{\text{re}} = r_i^{\text{l}} - r_i^{\text{a}}$ denotes the available *reaction time* of request i and represents the time period from the arrival time of request i to the end of its time window. Hence, the existence of larger reaction times provides more possibilities for dispatching and servicing pending requests which leads to a low $edod_{\text{tw}}$ and indicates relaxed system situations. In Fig. 2.13, the reaction times of two instances of dynamic routing problems, each comprising two dynamic requests, are illustrated. As can be seen in this figure, the reaction times are larger in instance A than in instance B making instance A easier to solve.

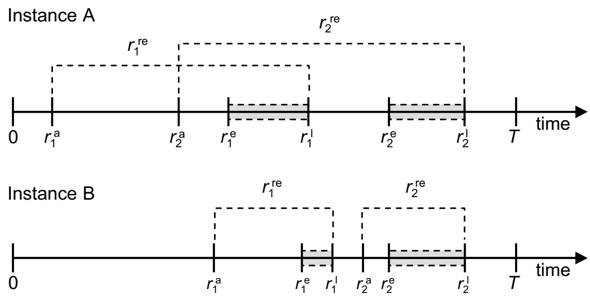


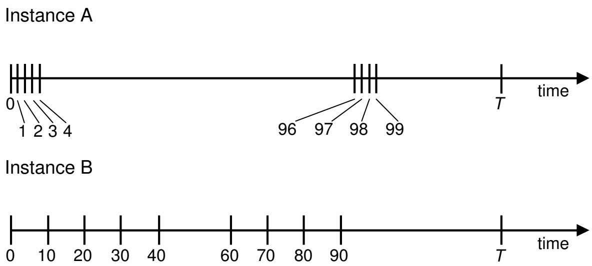
Fig. 2.13 The effective degree of dynamism with time windows (cf. Larsen 2000, p. 60)

Note that in dynamic routing problems in which the minimization of customer inconvenience is pursued, it holds that $r_i^{l} = r_i^{e} = r_i^{a}\ \forall i \in \{1, \ldots, n_{\text{dyn}}\}$ since customer inconvenience represented by penalty costs directly starts upon request arrival in the system. In such applications, the reaction time r_i^{re} is 0 for all dynamically arriving requests. Since almost all customer requests are dynamic in RDOPG applications considered in this book, these applications have an $edod_{\text{tw}}$ close to 1 making them difficult to solve with a high solution quality.

Further Extensions to the Degree of Dynamism The described *dod* and its variants provide efficient measures for determining the difficulty level of routing problems in which the minimization of travel distance or comparable objective functions is pursued. With regard to RDOPG applications in which the minimization of customer inconvenience is pursued, further extensions are possible out of which two are described in what follows. For example, considering Fig. 2.12, a different difficulty level for the two instances illustrated arises. Under this objective, instance A can be considered to be more difficult to solve than instance B since in instance A, the dispatcher has to deal with a large amount of newly arriving requests in a short amount of time. This can be complex, especially if customer inconvenience depends on request response times, good and fast tour plan adaptations are required. Hence, with regard to RDOPG applications, it is advantageous to additionally evaluate other request arrival characteristics. Specifically, the distribution of request inter-arrival times is another important aspect. As illustrated in Fig. 2.14, two instances with the same *edod* can have a significantly different request arrival distribution. If the minimization of customer inconvenience is pursued, instance A is more difficult to solve than instance B since, as mentioned above, the dispatcher is required to make good decisions for a large amount of requests at the same time. Moreover, in dynamic VRPs, the fleet size is often restricted (see aspect 11 of Psaraftis' approach presented in Sect. 2.5.2.2). Hence, a large number of requests shortly arriving one after the other causes longer request response times and hence higher customer inconvenience values compared to evenly distributed request arrivals.

Another possible extension which is relevant for RDOPG applications is mentioned by Pillac et al. (2010). The authors state that the geographical distribution of newly arriving requests and resulting travel times between requests should be considered since they significantly determine the complexity of dynamic routing problems in which the minimization of the request response times is pursued. Related to

Fig. 2.14 Two instances with the same effective degree of dynamism of 0.4 and significantly different inter-arrival times



this, we will propose a new measure for characterizing differences in dynamic routing problems with regard to the applicability of stochastic knowledge in pro-active real-time control approaches in Sect. 8.9.

2.6.5 Three-Echelon Classification of Dynamic Routing Problems

In Larsen (2000, p. 60 et seq.) and Larsen et al. (2007) a two-dimensional classification for distinguishing real-world dynamic routing problems is proposed according to their degree of dynamism and the pursued objective. In this classification, dynamic routing problems are, according to the percentage of dynamically arriving requests, subdivided into *weakly*, *moderately*, and *strongly* dynamic systems. Hence, these echelons represent routing problems with a low, medium, and high degree of dynamism. Moreover, the pursued objective ranges from the minimization of distribution costs to the maximization of service quality. As mentioned in Sect. 2.6.1, these two objectives usually contradict each other so that both objectives cannot be completely fulfilled at the same time. Nevertheless, there exist dynamic routing problems in which both objectives are considered with a similar priority.

The proposed classification is illustrated in Fig. 2.15. As can be seen, there is a strong correlation between the degree of dynamism and the pursued objective in real-world dynamic routing problems, i.e., the higher the degree of dynamism, the more important service quality aspects become in the considered routing problem. In what follows, we describe characteristics of each echelon and present corresponding real-world applications (cf. Larsen 2000 and Ghiani et al. 2003).

2.6.5.1 Echelon I: Weakly Dynamic Systems

Weakly dynamic systems are characterized by only a small number of customers dynamically arriving during the execution of the transportation process. According to the authors, the majority, i.e., at least 80 % of all customer requests are already known during the initial tour plan generation process (cf. Larsen et al. 2007). Hence, routing problems in this echelon have a *dod* < 0.2. In what follows, we present typical real-world routing problems which belong to this echelon:

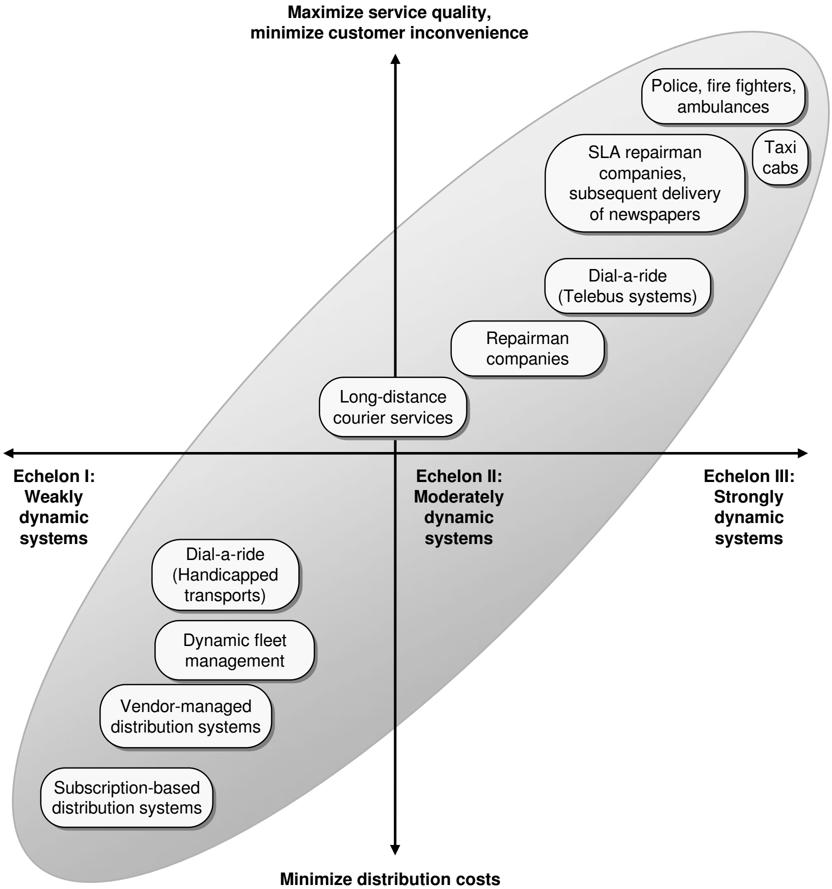


Fig. 2.15 Three-echelon classification of dynamic routing problems (cf. Larsen 2000, p. 63, Larsen et al. 2007)

- **Subscription-based distribution systems:** In general, customers subscribe to this type of service so that customer requests and their demands are often known in advance. The authors mention that less than 5 % of all customers request an additional service or make changes in their pre-ordered demands. In such systems, the tour plan is usually only adapted in longer time intervals, e.g., on a monthly basis, in order to save coordination effort and increase efficiency by utilizing driver learning curves. In order to service the few spontaneous customer requests, some slack time is included in the tour plan. One real-world example of subscription-based distribution systems is the delivery of groceries to private households on a regular basis.
- **Vendor-managed distribution systems:** In such distribution systems, companies manage the inventory of their customers and automatically replenish their inventory when necessary. In order to determine which customers are required to

be serviced during the next planning horizon which is usually the next day, the distribution company estimates customers' inventory levels using specific methods. Hence, the set of pending requests can be well estimated during the initial tour plan generation process. However, since the actual consumption rates of customers may vary from what has been predicted, a small percentage of customers additionally require service. Moreover, new customers may demand urgent service. Besides newly arriving requests, another source of dynamism is given by the uncertain demand of customers. Hence, capacity limitations of vehicles may require additional but usually few tour plan adaptations during the day. A real-world application belonging to such distribution systems is the distribution of heating oil to private households (cf. Larsen 2000, pp. 16–17). In this application, the oil consumption is estimated by the degree-days method which estimates consumption rates based on measured outdoor temperatures. The author states that about 20 % of all customers are dynamic in this real-world problem.

- **Dynamic fleet management:** In many international freight forwarding companies, cost aspects often have a higher priority than urgency of requests. Due to the long-haul character of these transportation processes, customers are usually required to negotiate transportation requests some time in advance with the transportation company. Nevertheless, tracking vehicles in real-time and allowing the re-routing of traveling vehicles, which still have available loading capacities, can help to integrate dynamically arriving requests. In doing so, significant cost savings can be achieved on long-distance transportation activities.
- **Dial-a-ride for handicapped people:** Specialized companies offer dial-a-ride services for transporting handicapped people between their home and their desired destination, usually for medical appointments. The arising transportation processes can be modeled as dynamic variants of the Dial-A-Ride Problem (DARP, see Sect. 2.3.3.3). Since customers usually book transportation services in advance, only few customers request a service on the same day. Hence, it is possible to generate cost-efficient tour plans which also provide a high customer service quality by meeting the customers' desired appointment times and minimizing their ride times. For more information on related routing problems, see, e.g., Cordeau and Laporte (2003, 2007) and Cordeau et al. (2007a).

2.6.5.2 Echelon II: Moderately Dynamic Systems

This echelon consists of real-world applications in which the percentage of dynamically arriving customer requests during the execution of the transportation process is noticeable. Nevertheless, there is a considerable proportion of requests still known before the execution of the transportation process (cf. Larsen et al. 2007). The following real-world applications belong to the class of moderately dynamic systems:

- **Long-distance courier services and (express) parcel services:** This routing application occurs in companies which offer services for transporting mail and parcels. Besides regular customers, there are customers who demand transportation services on short notice so that the vehicle tours are required to be replanned each day. Furthermore, some companies accept late-in requests. In this case, customers may demand goods to be picked up at their location after the vehicles have already started to collect goods so that their tours require real-time adaptations. In order to realize long-distance transportation services, different logistical structures can be used. For example, shipments are collected at the customer locations and are transported to a hub. From this hub, the shipments are transported to individual destination hubs from which they are transported to their final destinations. For more information on such logistical structures, see, e.g., Bock (2004, p. 48 et seq.) and Crainic and Kim (2007).

Besides standard transportation services, several companies offer next-day express services which guarantee the delivery of shipments to national and international destinations by a specified time point the next day, e.g., 9:30 am, 10:30 am, or 12:00 am. Moreover, very urgent freights can also be delivered on the same day (cf. DHL Express Germany GmbH 2010). In order to efficiently coordinate such time-critical transportation processes, the utilization of sophisticated real-time control approaches is advantageous.

- **Repairman companies with scheduled customers:** Specialized companies provide repair services and maintenance for their customers, e.g., for maintaining automated teller machines (cf. Larsen 2000, p. 15). A characteristic of this type of repairman problems is that most requests are known in advance since they arise from maintenance contracts in which regular service is specified in known time intervals. In addition to customer requests known in advance, some urgent customer requests which require an immediate service can occur during the planning horizon. Other repairman companies mainly dealing with urgent requests which are not known in advance are described in the next echelon.

2.6.5.3 Echelon III: Strongly Dynamic Systems

This echelon comprises real-world routing problems in which only few or no requests are known before the execution of the transportation process. In these applications, the *dod* amounts to at least 80 % (cf. Larsen et al. 2002). Furthermore, arriving customer requests often require an urgent service. Due to primarily customer-oriented objectives which often consider customer inconvenience depending on the request response time, one possibility for improving the solution quality is the utilization of stochastic knowledge in order to utilize available vehicles more efficiently. More information about this will be given in the literature review in Sect. 4.3. The following real-world routing problems belong to the echelon of strongly dynamic systems:

- **Telebus dial-a-ride systems:** Larsen (2000, p. 64) presents a variant of a DARP which can be found in telebus systems. In such systems, buses only operate on demand, i.e., if customers request for transportation. For example, telebus systems are used on bus lines, either all day or at times where there is usually a very low demand. According to Borndörfer et al. (1997), telebuses can also be rented for handicapped people who cannot use public transportation systems. Similar to taxi cab services, most customer requests dynamically arrive during the day. However, since telebuses are usually required to be ordered some time in advance, the urgency is lower than in taxi applications described later in this section.

- **SLA repairman companies:** Besides the previously mentioned repairman companies in which most of the requests are known in advance, there exist repairman companies which mainly deal with urgent problems arising in case of severe failures. As described in Chap. 1, example applications comprise professional repairman companies which quickly resolve severe problems faced by their customers, e.g., manufacturing companies which are part of a supply chain. In such applications, repairman companies often need to comply with the regulations of SLA contracts negotiated with their customers. Failing to fulfill the SLA results in high penalty costs so that a quick reaction to dynamically arriving customer requests is of high importance. Bertsimas and van Ryzin (1991) introduce the Dynamic Traveling Repairman Problem (DTRP) and evaluate solution strategies which are described in detail in Sect. 4.5.1.
- **Subsequent delivery of newspapers:** As introduced in Chap. 1, the subsequent delivery of newspapers occurs when a subscriber has not received his newspaper in the morning as expected. In this case, a quick subsequent delivery of the missing newspaper is required in order to minimize customer inconvenience and maintain customer loyalty.
- **Taxi cab services:** Taxi drivers are often connected to a central call center where new customer requests are recorded and dispatched. By using appropriate technical equipment (see Sect. 2.6.3), individual taxi cabs can be located in real-time and pending customer requests are assigned to them (cf. Horn 2002a, 2002b). Since people frequently demand instant transportation, quality of service is also measured in this application by the resulting response time. These processes are modeled as a Dynamic Dial-A-Ride Problem (cf. Cordeau and Laporte 2007) since a customer request consists of two locations instead of one. Moreover, the service time of a request, i.e., the time needed for transporting a customer comprises some uncertainty since travel times can only be estimated.
- **Emergency services:** For emergency services such as police, fire fighters, and ambulances, typically no requests are known in advance. The solution quality of an emergency system is often measured by the achieved request response time. According to Larsen et al. (2007), typical ambulance emergency systems agree on a certain level of service quality so that, for instance, 90 % of the requests should be serviced within 5 minutes and the remaining requests within a maximum of 8 minutes. Due to the high significance of emergency services, there is a lot of ongoing research on this topic which has also been the case for the last decades. For example, Larson and Odoni (1981) propose the hypercube queuing model which has been widely used, especially in the coordination of police patrols. In Brotcorne et al. (2003), the authors study approaches for ambulance dispatching developed during the past three decades. More information on relocation strategies which have also been applied to this type of applications will be given in the literature review in Sect. 4.3.2.

As can be derived from this description, RDOPG applications, i.e., SLA repairmen companies and the subsequent delivery of newspapers (see Sect. 1.1) belong to the echelon of strongly dynamic systems.

2.6.6 *Evaluating the Performance of Solution Approaches for Dynamic Routing Problems*

In what follows, we describe two measures known in the literature for evaluating the performance of solution approaches for dynamic routing problems.

2.6.6.1 Competitive Analysis

Sleator and Tarjan (1985) introduced a general framework denoted as *competitive analysis*. In dynamic routing problems, this framework can be used to define the *competitive ratio* of a solution approach A denoted as CR_A (cf. Larsen et al. 2007). Given an instance I of the considered dynamic routing problem, $z_A(I)$ defines the objective function value of the solution generated by A applied to this instance. Furthermore, $z^*(I^{\mathrm{O}})$ represents the objective function value of the solution generated by an optimal solution approach when applied to the corresponding *offline instance* I^{O}, i.e., the corresponding static instance of I in which the optimal solution approach has access to all relevant information before the execution of the transportation process. Furthermore, the set M comprises all theoretically possible instances of the considered dynamic problem. The competitive ratio of solution approach A is defined as:

$$CR_A = \sup_{I \in M} \frac{z_A(I)}{z^*(I^{\mathrm{O}})}$$

By definition, CR_A represents the worst-case ratio of solution approach A compared to an optimal offline solution approach. Therefore, the competitive ratio is a measure for evaluating the worst-case performance of A on all instances of a given dynamic routing problem. One issue with applying the competitive ratio is that up to now many dynamic routing problems can only be empirically evaluated by simulation-based methods. Hence, for such problems, it is not possible to evaluate all theoretically possible instances $I \in M$. Nevertheless, there exist specific dynamic routing problems to which the competitive ratio has been successfully applied (cf. Bertsimas and van Ryzin 1991, 1993a, 1993b, de Paepe 2002, and Angelelli et al. 2007). In these problems, specific assumptions, e.g., about request arrivals and their distribution, are made. In case of dynamic events, plan adaptations are carried out by strategy-oriented approaches which allow the competitive ratio to be applied since the effects of the applied adaptation methods on the solution quality can be analytically studied. Strategy-oriented approaches will be described in the literature review in Sect. 4.5.1.

2.6.6.2 The Value of Information

In order to evaluate the performance of solution approaches on dynamic routing problems which cannot be evaluated by the competitive ratio previously introduced, Mitrovic-Minic et al. (2004) propose a measure denoted as the *value of information*. It evaluates the worth of having complete knowledge of all relevant information before the execution of the transportation process. Given a solution approach H (often a heuristic approach) and an instance I of the considered dynamic routing problem, $z_H(I)$ represents the objective function value of the solution generated by H for problem instance I. Furthermore, $z_H(I^{\text{O}})$ represents the objective function value of the solution which has been generated by H on the corresponding offline instance of I. The value of information of solution approach H on instance I is defined as:

$$V_H(I) = \frac{z_H(I^{\text{O}}) - z_H(I)}{z_H(I^{\text{O}})}$$

Following this definition, the value of information represents the maximum possible gain in solution quality when H is applied on I if all information is known in advance. Hence, a larger value of $V_H(I)$ indicates instances in which it is especially beneficial when more relevant information is known before the execution of the transportation process, for example by utilizing stochastic knowledge about future expected events. Based on the value of information, we will describe in Sect. 8.2.6 a measure which is used in our computational results for evaluating differences in the performance of the real-time control approaches developed in this book.

Although the value of information gives an upper bound on improvements in solution quality which are attainable if more information, e.g., suitable stochastic knowledge, is known in advance, it does not consider specific characteristics which are present in the application or in test instances. Moreover, note that the estimation of this measure is also solution approach-dependent. In this book, we will propose new measures which take into account specific characteristics present in the considered routing problem. Hence, these measures allow for estimating the improvements which are additionally attainable just by analyzing the request data without requiring empirical assessment of the attained solution quality in the online and offline case and therefore allows for an approach-independent estimation. Moreover, these measures allow for identifying specific characteristics of request data sets which contribute to an efficient utilization of stochastic knowledge thereby assessing possible improvements and supporting estimations more precisely. These measures will be described in Sects. 8.8.2 and 8.9.

2.7 A General Classification Scheme for Routing Problems

In order to capture all attributes of routing problems, several authors have developed a general classification scheme (cf., e.g., Bodin and Golden 1981, Stumpf 1998, Richter 2005, and Eksioglu et al. 2009). In this section, a variant of this scheme

Table 2.2 General classification scheme for routing problems including RDOPG-specific attributes (in accordance with, e.g., Bodin and Golden 1981, Stumpf 1998, Richter 2005, and Eksioglu et al. 2009)

Attribute	Possible values
Requests	
Type	Pickup only
	Delivery only
	Pickup and delivery on the same tour (with backhauls, see Sect. 2.4.2)
	Combined requests (Pickup and delivery)
Location	At nodes
	At arcs/edges
	Mixed
Geographical distribution	Distributed with a pattern
	Randomly distributed
Degree of dynamism	Low
	Moderate
	High
Percentage of regular customers	None
	Few
	Medium
	High
	All
Time restrictions	No time windows
	One-sided time window
	Two-sided time window
	Fixed time points for service
	Hard time windows
	Soft time windows
	Multiple time windows
Divisibility	Non-divisible requests
	Arbitrarily divisible requests
	Restricted divisible requests (e.g., minimum amount per part)
Transshipment	No transshipment
	Transshipment possible
Request size	Full-truckload (FT)
	Less-than-truckload (LTL, e.g., piece goods)
Types of transportation objects	Goods
	Services
	Persons

Table 2.2 (Continued)

Attribute	Possible values
Size of service area	Urban
	Rural
	Regional
	National
	International
Sequence of requests	Arbitrary
	Predecessor-/Successor-restrictions
Selection of requests	Service of all requests
	Possibility to select requests to be serviced, i.e., requests can be rejected
Frequency of request occurrence	Once
	Semi-periodic
	Periodic
Fleet	
Size	One vehicle
	Multiple vehicles, restricted fleet size
	Multiple vehicles, unlimited fleet size
Consistency	Homogeneous (one type of vehicles)
	Heterogeneous (different types of vehicles)
Capacity	No capacity constraints
	One-dimensional capacity constraints
	Multi-dimensional capacity constraints
Location	One depot
	Multiple depots
	Freely distributed
Number of tours per vehicle and planning horizon	One tour
	Multiple tours
Time restrictions	No restrictions
	Maximum duration of tour
Loading restrictions	Arbitrary loading
	Restricted set of goods can be loaded
Number of drivers per vehicle	(Number)
Driver to vehicle assignment	Fixed
	Free
Vehicle reliability	Reliable
	Vehicle slowdowns possible
	Vehicle breakdowns possible

Table 2.2 (Continued)

Attribute	Possible values
Network	
Type of network	Road network (real distances)
	Coordinate system (Euclidean distances)
Data	Symmetric (undirected graph)
	Asymmetric (directed graph)
	Mixed
Travel times	Constant
	Time-dependent known
	Dynamic with uncertainty
Planning horizon	
Length	One period
	Multiple periods
	Open-ended
Type	Once
	Continuous
	Repeating
Tour plan	
Type	Daily changing tour plan
	Fixed tour plan
	Fixed tour plan with slack time for flexibility
Type of tours	Open tours (without return to depot)
	Closed tours (with return to depot)
Restrictions	Maximum tour duration
	Maximum tour length
	Worker-dependent restrictions (e.g., maximum working time, driving time restrictions)
Relevant information	
Type of relevant information	Constant
	Changing
Type of problem	Static
	Dynamic, changes deterministically known
	Dynamic, changes comprise uncertainty
Availability of changes	Global
	Local
	Mixed
Types of considered dynamic events	New requests
	Changes in request attributes

Table 2.2 (Continued)

Attribute	Possible values
	Traffic congestions and road blockages
	Vehicle disturbances
Type of data	Real-world data
	Designed data
	Both real-world and designed data
Stochastic knowledge	
Type of stochastic knowledge	None
	Prescribed probability distributions are available
	Forecasted by analyzing historical data, e.g., past request information
Increasing tour plan flexibility	
Methods used for increasing tour plan flexibility	None
	Waiting strategies
	Relocation strategies
	Request assignment strategies
Objective	
Considered objective	Minimization of travel length
	Minimization of travel time
	Minimization of operational costs
	Minimization of the number of utilized vehicles
	Minimization of the number of unserviced customers
	Minimization of response time
	Minimization of customer inconvenience
Applied approach	
Processing of changes in the relevant information	Centralized
	Decentralized
	Mixed
Type of applied methods	Exact
	Heuristic
	Simulation
	Real-time control methods
Type of model	Deterministic static
	Stochastic static
	Deterministic dynamic
	Stochastic dynamic
Type of application	Static
	Dynamic

extended by attributes which are relevant for the considered RDOPG applications is presented. The resulting classification scheme is illustrated in Table 2.2.

The RDOPG applications considered in this book will be characterized according to this general classification scheme in Sect. 3.2.3.

2.8 Summary

In this chapter, we discussed different characteristics and variants of routing problems. First, an introduction to classic static routing problems was given. Since RDOPG applications can be modeled as a DVRP variant in which new requests dynamically arrive, variants of the VRP and the DVRP made up the main part of this chapter. Specifically, first static VRPs which differ from each other in terms of problem attributes, such as considered type of requests, number of depots, available vehicles, and pursued objective function, were described. After this description, dynamic routing problems in which the relevant information changes during the execution of the transportation process were discussed and classifications known in the literature which deal with various differences between static and dynamic routing problems were described. With regard to attributes specific for RDOPG applications, a unified approach for classifying routing problems based on insights from known classifications in the literature was introduced. In this unified approach, the type of relevant information and corresponding changes that occur in the problem data during the execution of the transportation process as well as the availability and processing of these changes were considered. Moreover, depending on the type of available stochastic knowledge, different variants of VRPs known in the literature were described. After this description, in accordance with the considered RDOPG applications, dynamic routing problems in which a centralized coordination of the transportation process is performed were considered in detail. This included typical objectives pursued in dynamic routing problems and a description of dynamic events considered in dynamic routing problems. Moreover, a measure denoted as the degree of dynamism which has been proposed by Lund et al. (1996) and Larsen (2000) was described. This measure is used for determining the difficulty level of solving individual problem instances of dynamic routing problems. Based on this measure, a three-echelon classification proposed by Larsen (2000) was introduced which includes the characterization of various real-world dynamic routing applications. In order to evaluate the performance of dynamic routing approaches, two measures known in the literature were presented, in particular the competitive ratio and the value of information. Finally, a general classification scheme for characterizing routing problems was presented and extended with regard to additional characteristics which are relevant for RDOPG applications.

http://www.springer.com/978-3-642-33471-9

Pro-active Dynamic Vehicle Routing
Real-Time Control and Request-Forecasting Approaches
to Improve Customer Service
Ferrucci, F.
2013, LII, 319 p., Hardcover
ISBN: 978-3-642-33471-9
A product of Physica-Verlag Heidelberg