On Modeling Stochastic Dynamic Vehicle Routing Problems

Marlin W. Ulmer *[1], Justin C. Goodson[2], Dirk C. Mattfeld[1], and Barrett W. Thomas[3]

[1]Technische Universität Braunschweig, Carl-Friedrich-Gauß-Fakultät, Mühlenpfordtstraße 23, 38106 Braunschweig, Germany

[2]Saint Louis University, Richard A. Chaifetz School of Business, 3674 Lindell Blvd, St. Louis, MO 63108, United States

[3]University of Iowa, Tippie College of Business, 108 John Pappajohn Business Building, Iowa City, IA 52242, United States

Abstract

Operations research requires models that unambiguously define problems and support the generation and presentation of solution methodology. In the field of dynamic routing, capturing the joint evolution of complex sequential routing decisions and stochastic information is challenging, leading to a situation where rigorous methods have outpaced rigorous models and thus making it difficult for researchers to engage in rigorous science. We provide a modeling framework that strongly connects application with method and that leverages the rich body of route-based planning and optimization. As a generalization of conventional Markov decision processes (MDPs), route-based MDPs augment the state space, action space, and reward structure to include routing information. Accordingly, route-based MDPs make it conceptually easier to connect dynamic routing problems with the route-based methods typically used to solve them – construct and revise routes as new information is learned. We anticipate route-based MDPs will facilitate more scientific rigor in dynamic routing studies, provide researchers with a common modeling language, allow for better inquiry, and improve classification and description of solution methods.

Keywords: stochastic dynamic vehicle routing, modeling framework, literature review, Markov decision process

*Corresponding Author, Email: m.ulmer@tu-braunschweig.de

1 Introduction

Stochastic dynamic vehicle routing problems (SDVRPs) are problems in which a set of geographically dispersed customers is visited by one or more travelers or vehicles. The problems are dynamic because information stochastically changes over the operating horizon and there exist opportunities to make decisions in response to the revealed information. For example, a ride-sharing service may dynamically assign vehicles to customer requests, a delivery truck may adjust its route in response to changing traffic conditions, or appointments may be rescheduled on-the-fly to compensate for uncertain service times. While stochastic dynamic vehicle routing has a 30-year history in the operations research literature, the majority of SDVRP work is recent. For example, more than half of the papers in our literature review have been published since 2010. Further, the 2013-14 most cited paper in the *European Journal of Operational Research* reviewed stochastic dynamic vehicle routing (Pillac et al., 2013), indicating high interest in the topic among researchers. This trend in SDVRP research is likely to accelerate as opportunities afforded by data availability and computing power increase. Such developments facilitate the integration of models for uncertainty with optimization methods that use predictions to prescribe anticipatory decisions. Savelsbergh and Van Woensel (2016) suggest taking advantage of these opportunities "may necessitate new views, paradigms, and models for decision support." Further, Speranza (2018) indicates a major obstacle for researchers addressing SDVRPs is "the importance of appropriately modeling dynamic events and simultaneously incorporating information about the uncertainty of future events." Moreover, in the context of general dynamic optimization, Powell (2019) notes that "modeling sequential decision problems is often the most difficult task," a sentiment we find to be particularly salient for SDVRPs.

The Dual Role of a Mathematical Model

Suitable optimization models are the key to addressing the challenges and opportunities surrounding SDVRPs. In operations research, mathematical modeling plays a dual role as the intermediary between real-world applications and solution methods. The first role of a model is to capture the essence of the decision-making task in the form of objective, constraints, and decision variables. Modeling begins with a description, but should go further: "Mathematical models have many ad-

vantages over a verbal description of the problem. One advantage is that a mathematical model describes a problem much more concisely. This tends to make the overall structure of the problem more comprehensible ... [and] facilitates dealing with the problem in its entirety and considering all its interrelationships simultaneously" (Hillier and Lieberman, 2001, p. 11). The second role of a model is to inform the design of solution methods: "... a mathematical model forms a bridge to the use of high-powered mathematical techniques and computers to analyze the problem" (Hillier and Lieberman, 2001, p. 12). A classic example of synergy between model and solution method is set partitioning and column generation (see Desaulniers et al. 2006). Though multiple integer programming formulations may accurately represent an application and yield equivalent optimal solutions, models with large numbers of decision variables facilitate decomposition methods that offload difficult portions of the optimization to a subproblem and rely on general branch-and-bound techniques to solve a master problem. When properly positioned between problem and solution approach, SDVRP models will allow researchers to connect applications with methods, unambiguously defining problems as well as driving solution development.

No Common Models in the SDVRP literature

Despite the undebated importance of modeling among operations researchers, the extant SDVRP literature experiences difficulty identifying models that bridge application with solution method. A significant challenge in SDVRP research is transformation of the problem description into something unambiguous and convenient for analysis. In particular, SDVRP papers often struggle to formally express the sequential decision-making, intricate uncertainties, and plan-based characteristics of many stochastic dynamic vehicle routing applications.

While modeling problems is a precondition for publishing in many areas of operations research, the vast majority of SDVRP papers do not provide a model of the stochastic and dynamic problem. Rather, most papers seek to describe the problem using models for deterministic counterparts of the stochastic and dynamic problem. Considering the history of routing research, namely its role in advancing discrete optimization – especially integer programming and metaheuristics – it is not surprising that many routing researchers seek to model SDVRPs via extensions of deterministic modeling paradigms. However, such notation is often inadequate to model problems where information and route planning evolve over time.

The result is a disparate body of literature, research studies claiming to address similar applications, but which are difficult to compare and contrast because the ambiguities of problem description are not resolved by a mathematical model. Further, the Markov decision process (MDP) models used by some authors are often dissimilar to the route-based solution methods employed in their work and in much of the extant literature. In these MDPs, the decision variable, or policy, typically directs the decision maker in the moment, but provides little guidance on future actions. Because route-based methods outline decisions now and potential decisions in the future, state-of-the-art SDVRP solution techniques and conventional MDP models are often incongruent and disconnected (see, for example, Thomas (2007); Ulmer et al. (2018); Klapp et al. (2018a)).

A Modeling Framework Designed for SDVRPs

Recognizing that rigorous SDVRP methods have outpaced rigorous SDVRP models, we propose in this paper a framework that models the evolution of information and route planning while simultaneously connecting SDVRP applications and their solution approaches. We construct *route-based MDPs* by redefining the conventional MDP action space to operate on sets of planned routes, or *route plans*. We define a route plan as a potential course of action to follow for future decision-making, e.g., a path through a set of realized service requests or a customer sequence coupled with various quantities required to implement heuristic decision-making procedures. The modification leads to a generalization of the conventional MDP state to include route plans and to a definition of the current-period reward or cost to be the marginal change in value associated with a route plan update. Though conventional MDPs fill the first role of a model, our route-based MDPs offer improvements by also connecting with state-of-the-art methods, the second role of a model, thereby providing the framework necessary for improved analysis of important and challenging problems.

In addition to strengthening the role of models in routing research, we anticipate route-based MDPs will facilitate more scientific rigor in SDVRP studies. As a unified modeling framework, route-based MDPs will allow researchers to express their problems in a common language, thereby facilitating inquiry by better identifying contrasts with existing work and opportunities to derive new insights from the extant literature. Additionally, as a standard model for SDVRPs, we further anticipate route-based MDPs will facilitate better classification and description of solution methods – how plans are formulated, generated, evaluated, impacted by information, etc.

To develop the framework, we completed the following steps:

1. **Literature Review:** We analyze existing literature on stochastic and dynamic vehicle routing for two reasons. First, we justify the need for a modeling framework by highlighting the lack of modeling in most papers. Second, we identify properties of applications and methodology in the literature to derive structural components of a suitable modeling framework. Namely, we highlight that *route plans* are a dominating feature in many papers.

2. **Framework:** We develop our framework in the context of an existing modeling approach. Specifically, we generalize, replace, or extend existing elements of the conventional MDP to the needs of *route-based* problems and methodology. We also prove that the functionality of our proposed framework is equivalent to that of the conventional MDP.

3. **Proof of Concept:** Finally, we present justification that our modeling framework is tailored to the challenges of modeling dynamic and stochastic routing problems. In addition to the empirical evidence of the literature review, as a proof of concept, we consider a specific dynamic and stochastic vehicle routing problem that is based on a complex business problem. For this dynamic pickup and delivery problem, the conventional MDP is incomprehensible and lacks connections to both the business problem and the applied methodology. We show that our modeling framework intuitively establishes these connections.

The Goal of This Paper

We emphasize that the focus of the paper is not computational. Indeed, as we show, the literature is replete with sophisticated SDVRP procedures. Rather, we propose a modeling framework that connects these existing methods to the problems they seek to solve. We acknowledge that our route-based MDP increases model dimensionality relative to the conventional MDP. However, for most SDVRPs of practical interest, methods to identify an optimal policy are computationally intractable, whether applied to a conventional or route-based MDP. Thus, whether addressed in the conventional or route-based framework, these SDVRPs require the heuristic solution procedures that dominate the literature. Further, these procedures typically operate on the added dimensionality inherent to route-based models. Thus, rather than overcomplicating, the added dimensions of

the route-based MDP model often lay the groundwork for expressing solution methods in terms of model components. Our work fills the gap between current methods and the conventional model, allowing researchers to seamlessly define problems and frame solution approaches.

We make four contributions:

- Our paper is the first to present an MDP modeling framework based on traditional routing plans, connecting the problem model to both the routing application and to state-of-the-art solution methods.
- In contrast to existing SDVRP literature surveys, which largely classify SDVRP research via problem characteristics (e.g., the degree of stochasticity and operational constraints), our review compares and contrasts modeling techniques and solution methods.
- We prove the equivalence of conventional and route-based MDPs, thus establishing a theoretical basis for combining route-based optimization with dynamic and stochastic modeling.
- Using SDVRPs from the literature, we demonstrate how route-based MDP formulations more closely align problem models with state-of-the-art solution approaches.

The remainder of the paper is outlined as follows. In Section 2, we graphically illustrate the difference between conventional and route-based MDP models. In Section 3, we survey the SDVRP literature. In Section 4 and Section 5, we formalize conventional and route-based MDP models, respectively, providing examples throughout. In Section 6, we present a route-based MDP model for a challenging SDVRP from the literature with the aim of demonstrating the generality of the proposed model. We conclude in Section 7 with a list of takeaways.

2 Motivating Example

To prepare the reader for our literature review and modeling framework, we illustrate the concepts of route plans and sequential decision models via the *vehicle routing problem with stochastic service requests* (VRPSSR), a problem considered frequently in the SDVRP literature (Gendreau et al., 1999; Bent and Van Hentenryck, 2004; Mitrović-Minić and Laporte, 2004; Thomas and White III, 2004; Hvattum et al., 2006; Ichoua et al., 2006; Thomas, 2007; Ghiani et al., 2009,

2012; Ferrucci et al., 2013; Ulmer et al., 2018). In Section 2.1, we illustrate how route plans often inform VRPSSR decision-making. In Section 2.2, we show how a conventional MDP models the stochasticity and dynamism of the problem, but lacks route-based intuition. Finally, in Section 2.3, we highlight the potential for a route-based MDP model to bridge application with method. In subsequent sections, we continue use of the VRPSSR as an illustrative example to formalize conventional and route-based MDP models.

Our VRPSSR variant is characterized by the need to dynamically route one uncapacitated vehicle to meet service calls arriving randomly over a working day of duration T and from a set of potential customers. We denote the known customer locations by the set $\mathcal{N} = \{0, 1, \ldots, N\}$, where 0 represents a depot and the remaining locations represent customers. Although the location of each customer in $\mathcal{N}$ is known, whether or not a customer requests service is uncertain. The known travel time between two locations n and n' in $\mathcal{N}$ is denoted $d(n, n')$. Each customer served accrues a unit reward and the objective is to maximize total expected reward, the expected number of serviced customers.

2.1 Decision State and Route Plan

Figure 1 illustrates a VRPSSR decision state and route plan. We consider a situation with nine known customers, depicted as 1 through 9 in each of the two panels of Figure 1. The left-most panel shows a state of the problem at a point in time at which a decision must be made. At time $t = 20$, the vehicle has just serviced Customer 4 at its current location, Customers 8 and 9 have not requested service, Customers 2, 3, 5, 6, and 7 have requested service but have not yet been visited, and Customer 1 has already received service. The state comprises all information necessary to make a decision – time, vehicle location, and customer statuses. Given the information in the current state, we must route the vehicle to another location.

A common decision-making technique in the VRPSSR literature is to direct immediate vehicle movement by constructing a route plan, in this case a sequence of visits comprising the customer to visit next as well as potential future visits. In the center panel of Figure 1, a route plan commits the vehicle to next visit Customer 2, a movement we label as $x_k = 2$ and denote by the solid line connecting the current vehicle location to Customer 2. Denoted by dashed lines, the route plan suggests future visits to Customers 7, 6, and 5, followed by a return to the depot. Though the route

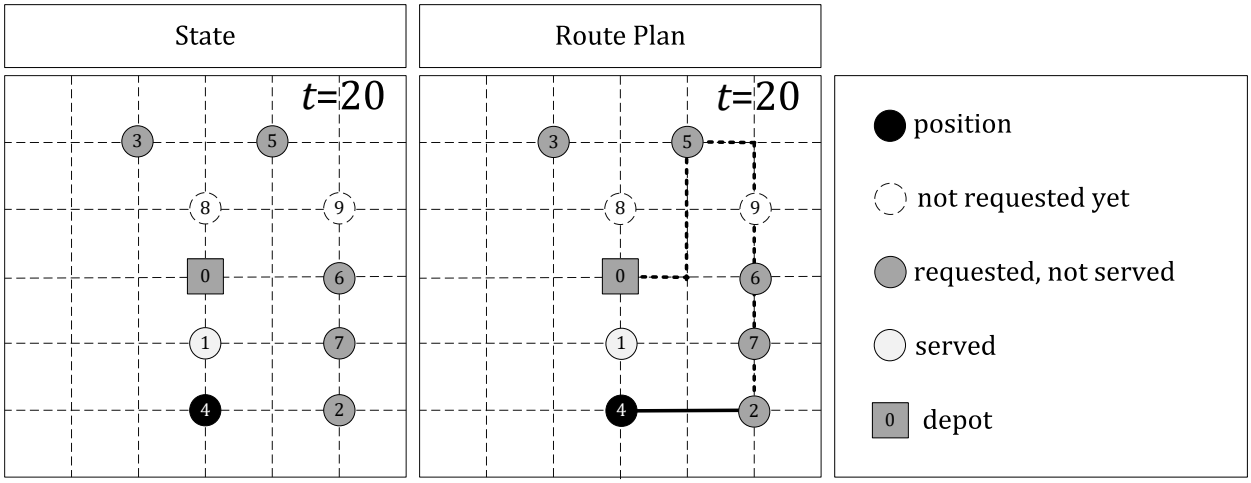

Figure 1: VRPSSR Decision State and Route Plan

plan suggests a sequence of actions, due to the stochastic and dynamic nature of the VRPSSR, only the first action is implemented. As is common in the VRPSSR literature, at subsequent decision states, the route plan may be updated in response to new information.

A model for the VRPSSR should incorporate the concepts of sequential states and route plans. We first consider an MDP, the standard framework to model sequences of states, decisions, and realizations of exogenous information. We demonstrate that a conventional MDP model captures the dynamism and stochasticity of the problem, but is disjoint from state-of-the-art solution methods.

2.2 Conventional MDP

Figure 2 illustrates a conventional MDP model for the VRPSSR. The left-most panel displays the same decision state s_k given in the left panel of Figure 1. With the information given in the state, we make a decision, which in the conventional MDP for the VRPSSR directs only immediate vehicle movement. As in Figure 1, we choose to travel to Customer 2. We assume the vehicle traverses a Manhattan-style grid where each edge requires 10 time units. We illustrate arrival of the vehicle to Customer 2 at time 40 in the right-hand-side panel of Figure 2. At this time, we also observe any new requests, the random information ω_{k+1}, occurring between the departure from Customer 4 at time 20 and the arrival to Customer 2 at time 40. At this point, we have a new state s_{k+1} with the vehicle located at Customer 2, Customer 4 has now been serviced, and Customer 8 has just requested service but has not yet been visited. Customers 3, 5, 6, and 7 have also requested

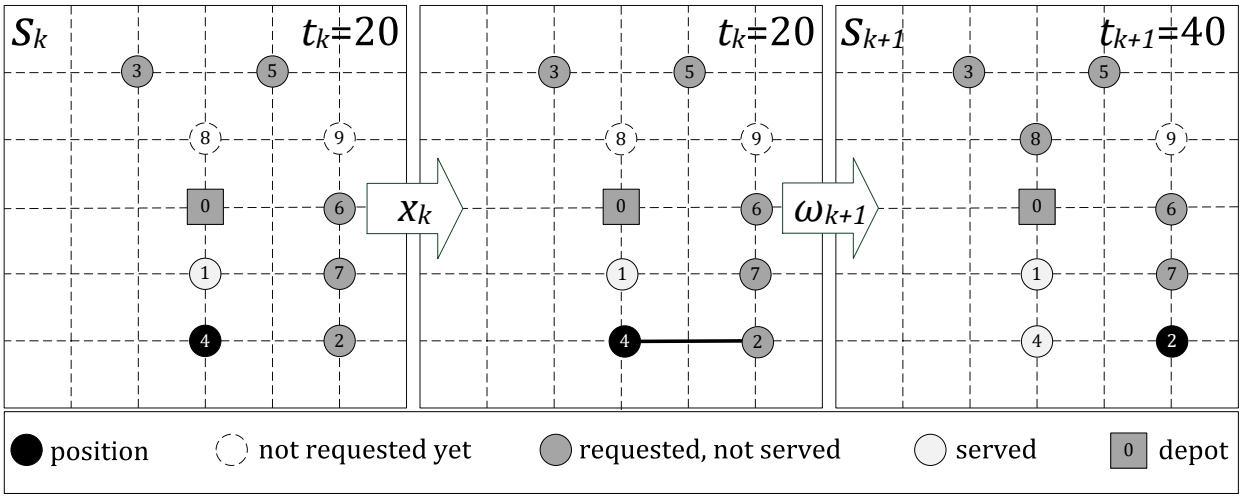

Figure 2: Conventional MDP for the VRPSSR

service but have not yet been visited and Customer 1 has already received service.

2.3 Route-based MDP

Figure 2 reveals a gap between the problem formulation and the methods most practitioners – as well as almost all approaches in the literature – employ to solve the VRPSSR. Notably, Figure 2 does not present a route through known customers, even a planned route that might later be modified. With this motivation, our route-based MDP merges the concept of route plans with the conventional MDP model.

Figure 3 illustrates a route-based MDP model for the VRPSSR. The three panels of Figure 3 are analogous to those of Figure 2, but incorporate the notion of a route plan. The left-most panel of Figure 3 shows the same situation at time 20 as displayed in Figure 2. However, it also illustrates a planned route we associate with the state and action of the route-based MDP model. In a current state s_k, the route plan is typically the route plan selected in the previous decision epoch, less the portion that was implemented. The route plan in this example proposes travel from the vehicle's current location at Customer 4 to Customer 3, then to Customer 5, and back to the depot. The decision to determine a new route plan, which may be the previous plan, depends on the methodology and is independent of the model. In this example, we assume the route plan is updated to the route plan presented in Figure 1, thus committing the vehicle to visit Customer 2,

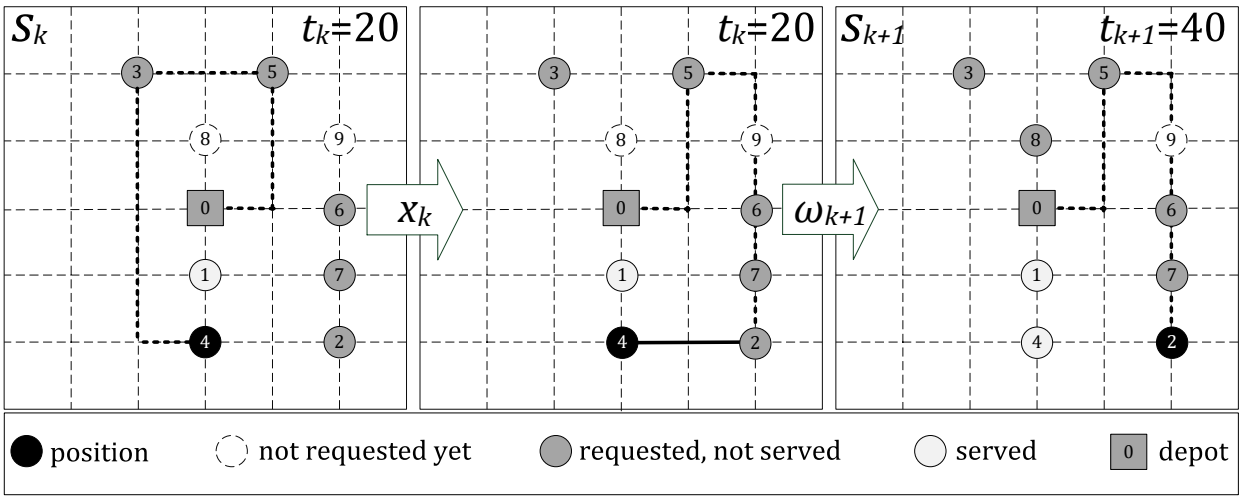

Figure 3: Route-Based MDP for the VRPSSR

an action we denote by the solid line from the vehicle's current location to Customer 2. While Customer 9 has not yet requested service, we note the updated planned route puts the vehicle in a position to serve Customer 9 if the opportunity arises. Finally, the panel on the right-hand side of Figure 3 shows the state of the route-based MDP at time 40 when the vehicle arrives at Customer 2. At this point, Customer 8 has requested service and the planned route is updated to reflect the vehicle's location.

In both Figures 2 and 3, the immediate action, that of traveling to Customer 2, is the same. However, Figure 3 combines the action with a route plan, thus providing a stronger connection to the solution methods predominate in the literature. Further, as demonstrated by the route plan and Customer 9 in Figure 3, route plans can be an indicator of the potential impact of future realizations of exogenous information, thus enhancing synergies between model and method.

3 Literature Review

Building on the illustrative example of Section 2, in this section we more formally examine the SDVRP literature. Our review is not exhaustive. In particular, we focus on papers from the reviews of Psaraftis et al. (2016) and Ulmer (2017), as well as recent publications, where the routing problems are both dynamic and stochastic. For readers interested in other perspectives on dynamic vehicle routing, we recommend Pillac et al. (2013), Ritzinger et al. (2016), and Demir et al. (2019).

Focusing largely on modeling choices and solution methods, we show many SDVRP research papers lack dynamic and stochastic models and the overwhelming majority of methods rely on route plans, thus highlighting the need for a framework connecting state-of-the-art routing methods with optimization models.

Table 1 summarizes our analysis. A checkmark in the "Model" column indicates the paper presents a scientific model of the problem, one capturing the dynamic and stochastic aspects of the SDVRP the paper addresses. The "Solution Method" column identifies the predominant methodology employed in the paper – *reoptimization* (RO), *policy function approximation* (PFA), *lookahead algorithm* (LA), and *value function approximation* (VFA). Our methodology types mirror the four policy classes of Powell (2011), ranging from myopic procedures (RO) to methods that implicitly (PFA) and explicitly (LA and VFA) anticipate uncertainties and dynamics. A checkmark in the "Route Plan" column indicates the methodology employs route plans. Below we discuss each dimension of Table 1.

Table 1: Stochastic Dynamic Vehicle Routing Literature

Paper	**Model**	**Solution Method**	**Route Plan**
Psaraftis (1980)	-	RO	✓
Bertsimas and Van Ryzin (1991)	-	PFA	✓
Papastavrou (1996)	-	PFA	✓
Tassiulas (1996)	-	PFA	✓
Savelsbergh and Sol (1998)	-	PFA	✓
Gendreau et al. (1999)	-	RO	✓
Swihart and Papastavrou (1999)	-	PFA	✓
Ichoua et al. (2000)	-	RO	✓
Secomandi (2000)	✓	VFA, LA	✓
Secomandi (2001)	✓	LA	✓
Larsen et al. (2002)	-	PFA	✓
Mitrović-Minić and Laporte (2004)	-	PFA	✓
Thomas and White III (2004)	✓	LA	✓
Bent and Van Hentenryck (2004)	-	LA	✓
Van Hemert and La Poutré (2004)	-	PFA	✓
Ichoua et al. (2006)	✓	PFA	✓
Gendreau et al. (2006)	-	RO	✓
Hvattum et al. (2006)	-	LA	✓
Chen and Xu (2006)	-	RO	✓
Huisman and Wagelmans (2006)	-	LA	✓
Fabri and Recht (2006)	-	RO	✓
Thomas (2007)	✓	PFA	✓

Bent and Van Hentenryck (2007)	-	LA	✓
Hvattum et al. (2007)	-	LA	✓
Sáez et al. (2008)	✓	LA	✓
Pureza and Laporte (2008)	-	PFA	✓
Novoa and Storer (2009)	✓	LA	✓
Secomandi and Margot (2009)	✓	LA	-
Ghiani et al. (2009)	-	LA	✓
Angelelli et al. (2009)	-	PFA	✓
Cortés et al. (2009)	✓	LA	✓
Maxwell et al. (2010)	✓	VFA	-
Mes et al. (2010)	-	VFA	✓
Beaudry et al. (2010)	-	RO	✓
Meisel (2011)	✓	VFA	✓
Schmid (2012)	✓	VFA	-
Ghiani et al. (2012)	✓	LA	✓
Azi et al. (2012)	-	LA	✓
Berbeglia et al. (2012)	-	RO	✓
Sheridan et al. (2013)	-	PFA	-
Ferrucci et al. (2013)	-	LA	✓
Tirado et al. (2013)	-	LA	✓
Goodson et al. (2013)	✓	LA	✓
Coelho et al. (2014)	✓	LA	✓
Ferrucci and Bock (2014)	-	RO	✓
Ehmke and Campbell (2014)	-	LA	✓
Schilde et al. (2014)	-	LA	✓
Toriello et al. (2014)	✓	VFA	-
Mavrovouniotis and Yang (2015)	-	RO	✓
de Armas and Melián-Batista (2015)	-	RO	✓
Wang and Kopfer (2015)	-	PFA	✓
Sarasola et al. (2015)	-	LA	✓
Schyns (2015)	-	RO	✓
Goodson et al. (2016)	✓	LA	✓
Angelelli et al. (2016)	✓	LA	✓
Kuo et al. (2016)	-	RO	✓
Tirado and Hvattum (2016)	-	LA	✓
Ferrucci and Bock (2016)	-	LA	✓
Tirado and Hvattum (2017)	-	LA	✓
Ng et al. (2017)	-	RO	✓
Zhang et al. (2018)	✓	LA	✓
Ulmer et al. (2018)	✓	VFA	✓
Srour et al. (2018)	-	LA	✓
Klapp et al. (2018a)	✓	LA	✓
Klapp et al. (2018b)	✓	LA	✓
Pillac et al. (2018)	-	RO	✓
Ulmer et al. (2019)	✓	LA	✓

Voccia et al. (2019)	✓	LA	✓

3.1 Modeling

A striking finding of our analysis is the number of papers – 44 of 68 – without a scientific model capturing the stochastic and dynamic elements of the problem. Among such papers, it is common to find a detailed problem description. However, the mathematical formalism almost always contained in static routing research is not present.

Though papers with scientific models differ in their terminologies – e.g., from Markov decision processes (Maxwell et al., 2010), to model predictive control (Sáez et al., 2008), to dynamic programming (Ichoua et al., 2006) – there are common themes. For example, all papers with a checkmark in the "Model" column utilize the concepts of states, decisions, and transitions to formalize the sequence of decision making and information evolution. States often describe vehicle positions and customers to serve, decisions typically depict vehicle movement, and transitions update the state following decision selection and a realization of uncertainty, such as demand or travel time. Further, these papers present solutions as policies, functions mapping states to decisions. The common modeling themes across the literature are most often cast formally as MDPs, in part motivating our embrace of MDPs as a base model for SDVRP optimization.

3.2 Solution Methods

Another notable finding of our analysis is that 63 of 68 papers utilize route plans in their solution methodology. Below, we discuss the four primary techniques in these papers guiding the use of route plans to identify SDVRP solutions. We also discuss examples of each solution technique. The four main types of methodology are:

- *Reoptimization* (RO): The route plan is determined with respect to the information currently available to the decision maker.
- *Policy function approximation* (PFA): The route plan is determined by a rule mimicking effective decision making in practice.

- *Lookahead algorithm* (LA): Such procedures sample stochastic information and derive route plans based on the sampled scenarios.

- *Value function approximation* (VFA): Such techniques learn the value of a route plan decision via repeated simulations and training.

Decision-making in early SDVRP research relies on RO methods, rolling horizon procedures to select route plans at each epoch via a static routing problem. Typically solved via math programming or metaheuristic techniques, the static problem often encompasses known information and gives little or no consideration to future uncertainties. For example, the pioneering work of Gendreau et al. (1999) employs a tabu search procedure whenever a service request is made, at each decision epoch seeking a travel-time-minimizing route plan through known customer locations, but ignoring potential future requests. Even though such methods do not anticipate future uncertainties and dynamics in route plan development, RO remains common in recent literature, particularly in rich routing problems with large numbers of operational constraints. In these problems, identifying feasible route plans can be challenging, thus anticipating future uncertainties often becomes secondary to managing known information. For example, Schyns (2015) considers the problem of dynamically routing refueling trucks at an airport. The fleet is heterogeneous in skills and capacities, split deliveries are permitted, and aircraft service time windows change stochastically over the planning horizon. Similar to the method of Gendreau et al. (1999), the approach of Schyns (2015) is executed whenever an uncertainty is realized and seeks to minimize response times given known time window values, but without considering future time window uncertainties. As we illustrate in subsequent sections, our route-based MDP connects RO methods with an optimization model.

PFA methods represent first steps away from myopic selection of route plans toward anticipation of future uncertainties. PFAs often consist of decision rules designed to mimic "common sense" route plans. For example, consider the work of Mitrović-Minić and Laporte (2004), which develops methods for a dynamic pickup and delivery problem where customer requests arrive randomly over a service horizon. As part of their solution method, Mitrović-Minić and Laporte (2004) design waiting strategies, rules dictating where and how long a vehicle should delay travel. Though the waiting strategies do not explicitly incorporate when and where origin-destination requests might be realized, the strategies do design route plans passing through various geographic

zones, implicitly considering service to known requests as well as to future requests in certain service areas and at particular times. Similarly, Van Hemert and La Poutré (2004) favor route plans with flexibility to incorporate future requests, but do not explicitly consider unrealized customers in decision-making. Our route-based MDP model provides a framework to join PFA-style rules with route plans.

Leveraging recent increases in computing power, LA methods give explicit consideration to stochasticity and dynamics, often drawing on samples of future uncertainties to generate and evaluate route plans. For instance, the route-based solution methodology of Ghiani et al. (2009) estimates time window violations across a route plan by simulating future customer requests. Explicit anticipation of when and where service requests might be realized favors route plans less likely to visit customers outside of desired service windows. A particularly popular type of LA, MSAs (Bent and Van Hentenryck, 2004) generate a set of scenarios representing future uncertainties, develop a route plan for each scenario, then employ a consensus function to select a distinguished route plan to guide decision-making in the current period. In Section 6, we discuss in-depth how the LA method of Schilde et al. (2014) for a dynamic dial-a-ride problem can be strongly connected to its application via a route-based MDP model.

Similar to LAs, VFAs explicitly anticipate future events, but do so via an approximation of the dynamic programming value function that considers both the immediate and future impact of a decision made now. In particular, VFAs typically make a functional approximation of Bellman's reward-to-go. For example, for a SDVRP with stochastic customer requests, Ulmer et al. (2018) approximate the future value of a routing decision via a non-parametric function of current time and slack, the latter a measure of a route plan's capacity to accommodate service requests not yet realized. In the analysis of Section 7, we connect the value functions of a conventional MDP with those of a route-based MDP, thus establishing an analytical basis for future research in route-based VFA methods.

3.3 Route Plans

In addition to a high frequency of route-based solution methodologies, our analysis of the SDVRP literature observes a broad range of what qualifies as a route plan. For example, in the SDVRP with stochastic customers addressed by Gendreau et al. (1999), a route plan is a path through a

set of realized service requests, a fundamental concept stemming from the static routing literature. In contrast, for the traveling purchaser problem studied by Angelelli et al. (2016), the authors incorporate into a route plan not only a customer sequence, but various quantities required to characterize supply and demand uncertainties as well as to implement potential heuristic decision-making procedures. Our route-based MDP model defines a route plan quite generally as a potential course of action to follow for future decision-making. We demonstrate the flexibility in our concept of a route plan via a basic example in Section 5 and a more complex depiction in Section 6.

While many papers employ route plans, they differ in how plans are generated. The literature highlights three techniques for updating route plans. The first method recalculates route plans entirely, as in Klapp et al. (2018a), which solves a probabilistic routing problem in each decision state to identify a new route plan. Stemming from evolutionary solution procedures, the second method maintains a pool of feasible routes throughout the decision process, as in Gendreau et al. (1999) and Bent and Van Hentenryck (2004), which identify new route plans via tabu and consensus criteria, respectively. The third method generates a new route plan from the incumbent, for example via insertion as in Ulmer et al. (2018) and Schilde et al. (2014).

Though the majority of SDVRP papers in our analysis utilize route plans in their solution methodologies, it is instructive to examine works that do not. For example, to dynamically dispatch and relocate ambulances, Maxwell et al. (2010); Schmid (2012) develop a VFA based on state-space aggregation, but without explicitly referencing potential vehicle movements beyond those available in the current period. Additionally, Secomandi and Margot (2009) employ standard backward dynamic programming techniques over a subset of states, heuristically selected via a lookahead mechanism, neither method requiring consultation of a route plan to dynamically direct a vehicle in response to stochastic customer demands. In nearest-neighbor fashion, Sheridan et al. (2013) restrict attention to immediate vehicle movements for a dynamic pickup and delivery problem. In a traveling salesman problem with stochastic arc costs, Toriello et al. (2014) estimate costs-to-go via VFA, operating directly on the conventional MDP.

More generally, it is notable that most papers in Table 1 lacking a checkmark in the "Route Plan" column contain a checkmark in the "Model" column and vice-versa. We feel this trend reflects the rich history of research focusing on deterministic routing problems. Thus, it is natural researchers would seek to connect existing technology with emerging dynamic problems. Recog-

nizing models need to keep pace with the latest advances, our work provides researchers with a rigorous model connecting SDVRP applications to state-of-the-art routing methods.

4 Conventional MDP Model

In this section, we review the conventional MDP model and give an example for the VRPSSR to prepare our route-based MDP. The notation presented in this section is used as the basis for the route-based MDP formulation in the following section. More detailed introductions to MDPs can be found in Puterman (2005) and Powell (2011).

4.1 Model

Conventional MDP models are characterized by six model elements:

Decision Epochs Points in time at which decisions are made.

State A tuple containing all information necessary to define the feasible actions at a particular decision epoch, the reward for choosing an action, and the resulting transition.

Actions Feasible decisions available for a particular state, also referred to as decisions.

Reward The quantity earned by choosing a particular action in a particular state.

Transition A function describing how the system evolves given a chosen action in a particular state. Transitions may involve exogenous information, notably the realization of random variables.

Objective A function of the rewards, typically maximization of the expected sum of rewards across all decision epochs.

Throughout the paper we refer to rewards. However, for minimization problems, rewards would more naturally be referred to as costs and the typical objective would minimize the expected sum of costs across all decision epochs.

We assume a finite horizon in which decisions are made at epochs $0, 1, \ldots, K$, where K may be a random variable. At the k^{th} decision epoch, the system occupies state s_k in state space $\mathcal{S}$.

A state contains the minimum amount of information necessary to determine the actions, current-period rewards, and transition probabilities if known. At epoch k and given state s_k, the decision maker chooses an action x_k from the set of feasible decisions $\mathcal{X}(s_k)$. Choosing action x_k when in state s_k induces a state transition from state s_k to state s_{k+1} in decision epoch $k+1$. The transition is random and determined by selected decision x_k and the set of random variables Ω_{k+1} representing the random information arriving between decision epochs k and $k+1$. We denote a realization of Ω_{k+1} as ω_{k+1}. As discussed in Powell (2011), the transition can be split into two parts – a transition from pre-decision state s_k to post-decision state s_k^x and a transition from s_k^x to pre-decision state s_{k+1}. The deterministic transition for the pre- to post-decision state is given by the function $S^X(s_k, x_k)$ and the random transition to the next pre-decision state is given by the function $S^\Omega(s_k^x, \omega_{k+1})$. Thus, $s_{k+1} = S(s_k, x_k, \omega_{k+1}) = S^\Omega(S^X(s_k, x_k), \omega_{k+1})$. These transitions are seen in Figures 2 and 3 in which the left-most panels illustrate pre-decision states, the middle panels post-decision states, and the right-most panels subsequent pre-decision states.

Let $\hat{R}_{k+1}(s_k, x_k, \omega_{k+1})$ be the random reward earned at decision epoch k when selecting decision x_k in state s_k and observing random information ω_{k+1}. Because ω_{k+1} may not be realized when selecting decision x_k, we define the reward in decision epoch k as the expected reward $R_k(s_k, x_k) = \mathbb{E}[\hat{R}_k(s_k, x_k, \Omega_{k+1}) \mid s_k, x_k]$, where $\mathbb{E}[\cdot]$ denotes the expectation operator (in this case with respect to Ω_{k+1}).

Let π be a function mapping a state to an action and Π the set of all such functions. Then, the objective is to maximize the total expected reward, conditional on initial state s_0, given as $\max_{\pi \in \Pi} \mathbb{E}\left[\sum_{k=0}^{K} R_k\left(s_k, x^\pi\left(s_k\right)\right) \mid s_0\right]$, where $x^\pi(s_k)$ is the action prescribed by policy π for state s_k. Discounting can be incorporated into the reward function, but we omit discount factors to simplify notation.

4.2 Conventional MDP Model for the VRPSSR

In this section, we formalize the content of Section 2.2 by presenting a conventional MDP model of the VRPSSR. Serving as an example of a classically-formulated SDVRP, the model facilitates comparison of the conventional and route-based formulations. Importantly, the example demonstrates a correct model for the VRPSSR does not require ongoing knowledge of a planned route. Additional examples of conventional MDP formulations for SDVRPs can be found in Thomas and

White III (2004), Ichoua et al. (2006), Secomandi and Margot (2009), Goodson et al. (2016), and Ulmer et al. (2018).

Decision Epochs

A decision epoch begins when the vehicle arrives at a location and observes new customer requests.

States

The state of the system at decision epoch k is the tuple $s_k = (n_k, t_k, z_k)$, where n_k in $\mathcal{N}$ is the vehicle's current location, t_k in the range $[0, T]$ is the time of arrival to n_k, and $z_k = (z_k(1), z_k(2), \ldots, z_k(N))$ is an N-dimensional vector describing the status of each customer at epoch k. For each customer n, $z_k(n)$ takes on a value in the set $\{0, 1, 2\}$:

$$z_k(n) = \begin{cases} 0, & \text{if customer } n \text{ has not requested service by time } t_k, \\ 1, & \text{if customer } n \text{ has made a request but has not been serviced by time } t_k, \\ 2, & \text{if customer } n \text{ has been serviced by time } t_k. \end{cases}$$

In initial state $s_0 = (0, 0, z_0)$, the vehicle resides at the depot at time 0 and $z_0(n)$ is 0 or 1 for each customer n in $\mathcal{N} \setminus \{0\}$. At final decision epoch K the process occupies a terminal state s_K in the set $\{(0, t_K, z_K) : t_K \in [0, T], z_K \in \{0, 1, 2\}^N\}$, where the vehicle has returned to the depot by time T and each customer's status is 0, 1, or 2. The state space is the set $\mathcal{S} = \mathcal{N} \times [0, T] \times \{0, 1, 2\}^N$.

Actions

An action at epoch k is an assignment of the vehicle to a location in $\mathcal{N}$. When the process occupies state s_k, the set of feasible actions is

$$\mathcal{X}(s_k) = \Big\{ x_k \in \{n \in \mathcal{N} : z_k(n) = 1\} \cup \{0\} : \tag{1}$$

$$t_k + d(n_k, x_k) + d(x_k, 0) \leq T \Big\}. \tag{2}$$

Condition (1) requires assignment to customers who have requested service or to the depot. Condition (2) restricts movement to locations from which the depot can be reached by time T.

Rewards

When the process occupies state s_k and decision x_k in $\mathcal{X}(s_k)$ is taken, a unit reward is accrued if a decision x_k moves the vehicle to a location where service has been requested: $R_k(s_k, x_k) = 1$ if $z_k(x_k) = 1$ and $R_k(s_k, x_k) = 0$ otherwise.

Transition

Following selection of action x_k from state s_k, the process transitions to a new state. We describe the transition in two parts, first to a post-decision state and then to a new pre-decision state. Transition to the post-decision state reflects the new vehicle location resulting from the selected action, the time at which the vehicle will arrive at the location, and a status update for the new location. We denote the post-decision state as $s_k^x = (n_k^x, t_k^x, z_k^x)$, where vehicle location is the selected decision $n_k^x = x_k$, time of arrival to n_k^x is $t_k^x = t_k + d(n_k, n_k^x)$, the status for location n_k^x is set to $z_k^x(n_k^x) = 2$ if n_k^x is a customer, and all other customer statuses remain the same.

We denote the next pre-decision state as $s_{k+1} = (n_{k+1}, t_{k+1}, z_{k+1})$. The transition from s_k^x to s_{k+1} marks the arrival of the vehicle to location $n_{k+1} = n_k^x$ at time $t_{k+1} = t_k^x$ as well as observation of any service requests arriving after time t_k and at or before time t_{k+1}. Denote the set of customers requesting service by $\omega_{k+1} \subseteq \{n \in \mathcal{N} \setminus \{0\} : z_k^x(n) = 0\}$, a subset of the status-0 customers in post-decision state s_k^x. Customer statuses are updated accordingly: $z_{k+1}(n) = 1$ for all n in ω_{k+1}. All other customer statuses remain the same.

Objective

The objective is to maximize the expected sum of rewards across the decision epochs.

5 Route-Based MDP Model

In this section, we formally present our route-based MDP model. The route-based MDP model includes the same model elements (decision epochs, states, rewards, transitions, and objective) as the conventional MDP model, but redefines the feasible action space $\mathcal{X}(s_k)$ to include planned routes. Definition 1 formalizes a space of route plans as a set coupled with an evaluation function:

Definition 1 (Route Plans and Rewards)**.** A space of route plans Θ is a general set on which is defined a real-valued function $R^\theta : \Theta \to \mathbb{R}$.

Though Definition 1 is broad, route plans typically describe a sequence of potential future actions, such as customer visits, and reward functions typically correlate with the objective. While the components of route plans are problem-specific, they may simply contain a sequence of customers for each vehicle, but might also include additional information such as assumed arrival times or planned amounts of goods to be delivered. The route-based model carries the route plan in the state and reformulates the current-period reward to capture the difference in plan values rather than the immediate return for selecting a decision. Despite these differences, as we show in the Appendix, the route-based MDP model is equivalent to the conventional MDP model.

5.1 Model

We first define the actions associated with the route-based MDP model, followed by descriptions of states and transitions as well as the reward function. We note that the decision epochs and objective are unchanged from the conventional MDP.

Actions and Route Plans

We denote an action by the pair $(x_k, \theta_{k+1}) \in \mathcal{X}(s_k) \times \Theta_{k+1}(s_k)$, where x_k and $\mathcal{X}(s_k)$ are defined by the conventional MDP, θ_{k+1} is a route plan, and $\Theta_{k+1}(s_k)$ is the set of route plans given current state s_k. The action and route plans are indexed differently, k versus $k+1$, to reflect that the action x_k will be executed in the current epoch k and that the route plan includes planned actions beginning in epoch $k+1$. It is possible that $\Theta_{k+1}(s_k) = \emptyset$ or that $\emptyset \in \Theta_{k+1}(s_k)$. In these cases, selection of the action $(x_k, \emptyset)$ proceeds as in the conventional MDP.

The example of Figure 3 illustrates the concept of an action-plan pair. In the example, we choose a route plan to be a sequence of feasible actions, an ordering of customers who have requested service but have not yet been visited. Beginning in state s_k with current route plan $\theta_k = (3, 5)$, the selected action $(x_k, \theta_{k+1}) = (2, (7, 6, 5))$ next visits Customer 2 and updates the route plan to visit Customers 7, 6, and 5. We provide additional examples in the next section.

States and Transitions

We augment the conventional MDP state variable as the pair (s_k, θ_k) to carry both the original state and the route plan selected by action (x_{k-1}, θ_k) in the previous decision epoch. In the initial state, we assume route plan θ_0 is empty and has value $R^\theta(\emptyset) = 0$. Figure 3 illustrates the state variable in a route-based MDP. In addition to depicting vehicle location, arrival time, and customer statuses, the left panel of Figure 3 shows the current route plan $\theta_k = (3, 5)$ visiting Customers 3 and 5. Similarly, the right panel of Figure 3 gives the updated route plan $\theta_{k+1} = (7, 6, 5)$ paired with state s_{k+1}.

Analogous to the conventional MDP model, upon selection of an action and route plan, the state variable transitions to an augmented post-decision state $S^X(s_k, (x_k, \theta_{k+1})) = (s_k^x, \theta_{k+1})$. Upon observation of random information ω_{k+1}, the post-decision state transitions to the next pre-decision state with the post-decision plan becoming the route plan of the new pre-decision state resulting in pre-decision state (s_{k+1}, θ_{k+1}).

Marginal Rewards

We connect route plan selection with decision-making by redefining the conventional MDP reward function as a marginal reward accounting for the difference in value between a previous route plan and a new route plan as well as immediate contributions by the current-period action. Let R^θ map a route plan to a real number. We can think of R^θ as the value of a route plan θ. Though R^θ may be freely defined, it is typically aligned with the problem and chosen structure of a route plan. We define the route-based MDP reward function as

$$R_k^\Delta\big((s_k, \theta_k), (x_k, \theta_{k+1})\big) = R_k(s_k, x_k) + R^\theta(\theta_{k+1}) - R^\theta(\theta_k). \tag{3}$$

We illustrate the marginal reward function via Figure 3, where $\theta_k = (3, 5)$, $\theta_{k+1} = (7, 6, 5)$, and $x_k = 2$. Letting $R^\theta(\cdot)$ be the number of customers included in a route plan, $R^\theta(\theta_k) = 2$ and $R^\theta(\theta_{k+1}) = 3$. Then, noting $R_k(s_k, x_k) = 1$, $R_k^\Delta((s_k, \theta_k), (x_k, \theta_{k+1})) = 1 + 3 - 2 = 2$.

While it is possible to include route plans in action selection but with no value by choosing $R^\theta(\cdot) = 0$ (in which case the route-based MDP behaves as a conventional MDP), defining the reward function as a marginal difference highlights the long-term impact of route plan changes.

In particular, as the above example demonstrates, the marginal reward function shifts rewards that may potentially be accumulated later in the horizon into an earlier period. Thus, connecting solution approaches to a route-based model may naturally encourage anticipatory decision-making, even when methods focus primarily on current-period marginal rewards.

5.2 Route-Based MDP Model for the VRPSSR

In this section, we formalize the content of Section 2.3 by presenting a route-based MDP model for the VRPSSR, demonstrating differences between our formulation and the conventional MDP. To fully characterize the model, we need only specify the structure of route plans and how route plans are valued.

We define a route plan as a sequence of customers, each of which has requested service but has not yet been visited. Let $\mathcal{N}'(s_k, x_k) = \{n \in \mathcal{N} \setminus \{x_k\} : z_k(n) = 1\}$ be the set of status-1 customers remaining in state s_k after selecting action x_k. Then, given a current state s_k, we define the space of routing plans $\Theta_{k+1}(s_k)$ as the set of feasible ordered subsets of $\mathcal{N}'(s_k, x_k)$. An ordered subset is feasible if the customers in the subset can be visited in order, beginning at location x_k, and ending at the depot, without violating duration limit T. We define the value of a plan θ_k as the number of status-1 customers on the route: $R^\theta(\theta_k) = |\theta_k|$, where $|\cdot|$ is the cardinality operator.

6 A Route-based Model for a Dynamic Dial-A-Ride Problem

In previous sections of the paper, we use the VRPSSR as an illustrative example to motivate and explain route-based MDP models. In particular, we show a route-based MDP model for the VRPSSR is more intuitive and makes a stronger bridge between application and methods than does a conventional MDP model. To further highlight the advantages of route-based models, we present in this section a more realistic and more complex problem from the literature, the dynamic dial-a-ride problem (DDARP) addressed in Schilde et al. (2014). Similar to the VRPSSR, the DDARP seeks to service customer requests arriving randomly over a given time horizon, but with two added complexities often encountered in ride-sharing ventures. One, each request must be picked up and dropped off within a certain time frame. Two, travel times are stochastic, becoming known only after a vehicle enters a path between two locations. Consequently, and often the case in

passenger transportation, the objective function considers passenger ride times between origins and destinations. Though the DDARP can be modeled as a conventional MDP, with decisions directing immediate vehicle movement as illustrated for the VRPSSR, such a model obscures the relationship between pickup and delivery decisions. Further, modeling decisions as the next location to visit explicitly captures ride-time violations only in the current period cost, hiding the impact of future violations. Thus, conventional MDP modeling techniques are not connected to an intuitive solution method, one explicitly linking present and future decisions via route plans, nor do conventional modeling approaches connect with the solution method of Schilde et al. (2014). In this section, we show how a route-based MDP model overcomes the challenges faced by the conventional MDP model.

Following the notation presented in Schilde et al. (2014), we associate with a stochastic service request by customer n a pickup location, a drop-off destination, and a time window $[e_n, l_n]$ where e_n is the earliest a customer needs pickup and l_n is the latest time by which the customer desires to arrive at the destination. The objective is to minimize the expected sum of tardiness (drop-off time for a customer n exceeds l_n), earliness (pickup time for a customer n occurs before e_n), and ride-time violations (the length of time a customer spends in the vehicle exceeds 40 minutes) accrued over service of all customer requests.

Though we build on the route-based MDP for the VRPSSR, our choice of route plan for the DDARP requires more information than a sequence of pickups and drop-offs. In particular, because our choice of route evaluation function estimates the total tardiness, earliness, and ride-time violations for a given route, we carry in a route plan additional information about planned arrival times, planned pickup times, and time windows necessary to make the estimate. These choices allow for an unambiguous mapping from route plans to rewards and facilitate a route-based solution method. In the sections that follow, we present a route-based MDP model for the DDARP and discuss how the model connects with the solution method of Schilde et al. (2014).

6.1 Route-Based MDP

We begin a route-based MDP model by defining conventional MDP model elements. A decision epoch occurs when the vehicle arrives at a location. The current state s_k at decision epoch k includes the time of arrival, the vehicle's current location, passengers on the vehicle, the times at

which the passengers were picked up, the pickup and delivery locations of these in-process requests as well as those of any outstanding requests, and the associated time windows. An action x_k provides service at the vehicle's current location and moves the vehicle to a location in the set $\mathcal{X}(s_k)$ of origins and destinations of outstanding requests as well as destinations of in-process requests. Cost $R(s_k, x_k)$ is the total tardiness, earliness, and ride-time violation incurred at the customer at the vehicle's current location. Transition to a subsequent state begins with the observation of ω_{k+1}, the travel time required to reach the next location followed by the realization of new customer requests upon arrival to the next location, the event marking the start of decision epoch $k + 1$.

Structured similar to Figures 2 and 3, Figure 4 illustrates conventional and route-based MDP model elements for the DDARP in a single graphic. The left pane illustrates the current state. At time $t_k = 20$ the vehicle arrives to drop-off location D1 for Customer 1, picked up at time 0 and with a late time window of 15. In-process requests include Customer 2, currently a passenger on the vehicle, with drop-off location D2 and a closing time window of 50. Outstanding requests include customer 3 and the just-made request from Customer 4. Respectively, the outstanding requests have pickup and delivery locations P3, D3, P4, and D4 denoted on the grid and time windows of $[20, 80]$ and $[40, 90]$. Available actions include drop-off of Customer 1 plus movement to drop-off Customer 2 or to pickup Customers 3 or 4. The cost associated with any of these actions is is $R(s_k, \cdot) = 20 - 15 = 5$ time units of tardiness dropping off Customer 1. The center pane illustrates the action servicing Customer D1 and moving the vehicle to P3. The right pane shows a travel time realization of 10 time units from D1 to P3 and does not indicate any new customer requests. We use the remaining portions of Figure 4 to illustrate route plans.

The space of routing plans $\Theta_k(s_{k-1})$ in state s_{k-1} is the set of feasible ordered subsets of destinations associated with in-process requests (excluding the vehicle's current location) as well as pickup and drop-off locations accompanying outstanding requests. We represent such a subset by $(\theta_{k_1}, \ldots, \theta_{k_m})$ and label it feasible if pickups are sequenced ahead of deliveries for each customer and if the first element of the sequence is selected action x_{k-1}. Additionally, to facilitate route plan evaluation, we associate with a sequence of pickups and deliveries four pieces of information. First, denote by $a(\theta_{k_i})$ the planned arrival time to location θ_{k_i} (a number that might be a calculated expectation, an estimate via simulation, or a quantity obtained via some other method). Second,

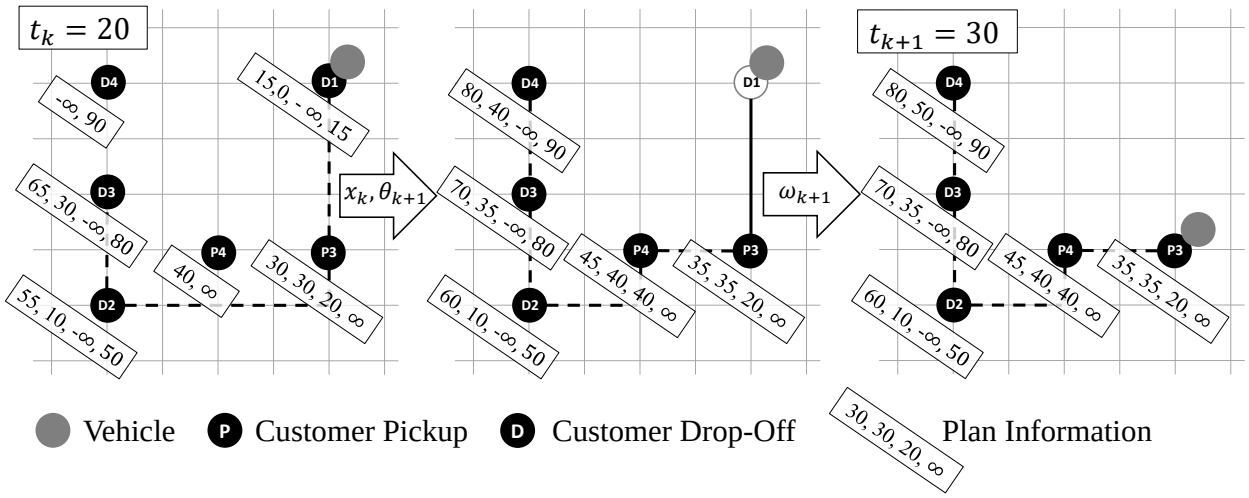

Figure 4: Route-Based MDP Example for the DDARP

denote by $p(\theta_{k_i})$ the pickup time for the service request associated with location θ_{k_i}. If θ_{k_i} is a pickup location, then $p(\theta_{k_i})$ is a planned quantity (again obtained via some measure). If θ_{k_i} is a drop-off location, then $p(\theta_{k_i})$ is set to the pickup time of the customer's origin. Third, denote by $e(\theta_{k_i})$ the earliest time service may begin at location θ_{k_i}, set to early time window $e_{\theta_{k_i}}$ if θ_{k_i} is a pickup location and to $-\infty$ if θ_{k_i} is a delivery location. Fourth, denote by $l(\theta_{k_i})$ the latest time at which service is desired at location θ_{k_i}, set to late time window $l_{\theta_{k_i}}$ if θ_{k_i} is a destination and to ∞ otherwise. Then, a route plan is the sequence of five-tuples

$$\theta_k = \Big(\big(\theta_{k_1}, a(\theta_{k_1}), p(\theta_{k_1}), e(\theta_{k_1}), l(\theta_{k_1})\big), \ldots, \big(\theta_{k_m}, a(\theta_{k_m}), p(\theta_{k_m}), e(\theta_{k_m}), l(\theta_{k_m})\big) \Big). \quad (4)$$

We calculate the reward $R^\theta(\theta_k)$ associated with a route plan θ_k as the sum of projected earliness (i), tardiness (ii), and ride time violations (iii):

$$\begin{aligned} R^\theta(\theta_k) \;=\; \sum_{i=1}^{m} \big(& \underbrace{\max\left(e(\theta_{k_i}) - a(\theta_{k_i}), 0\right)}_{(i)} + \underbrace{\max\left(a(\theta_{k_i}) - l(\theta_{k_i}), 0\right)}_{(ii)} \\ & + \underbrace{\max\left(a(\theta_{k_i}) - p(\theta_{k_i}) - 40, 0\right)}_{(iii)} \big). \end{aligned}$$

Defining initial and terminal routing plans θ_0 and θ_{K+1} as empty, and noting $R^\theta(\emptyset) = 0$, our modeling choices satisfy the equivalency conditions in the appendix. More specific, Condition 1

and thus Proposition 1 holds, thereby establishing a formal relationship between the conventional DDARP model and our route-based MDP model.

Figure 4 illustrates route plan evolution and evaluation. The left pane of Figure 4 shows a current route plan θ_k with a sequence of locations $(\text{D}1, \text{P}3, \text{D}2, \text{D}3)$, each location tagged with the four-tuple of associated data. Because the planned arrival time of 55 to D2 is five time units beyond the closing time window of 50, and because the projected ride time of Customer 2 is five time units greater than 40, $R^{\theta}(\theta_k) = 5 + 5 = 10$. Following action selection of $x_k = \text{P}3$, the center pane of Figure 4 depicts updated route plan θ_{k+1} with a sequence of locations $(\text{P}3, \text{P}4, \text{D}2, \text{D}3, \text{D}4)$ and associated data. The updated route plan adds Customer 4 to the location sequence and adjusts planned arrival and pickup times to account for the five time units of tardiness at location D1. Because the planned arrival time of 60 to $D2$ is 10 time units past the latest drop-off time of 50 and because the projected ride time of Customer 2 is 10 time units beyond the limit, $R^{\theta}(\theta_{k+1}) = 10 + 10 = 20$. Thus, the marginal reward is $R_k^{\Delta}((s_k, \theta_k), (x_k, \theta_{k+1})) = 5 + 20 - 10 = 15$. The right pane of Figure 4 shows the arrival of the vehicle to location P3 at time 30, five time units before the planned arrival time of 35. Though an earliness penalty will be incurred, the arrival time to P3 allows a subsequent route plan update to move projected arrival times ahead by five time units, an update that would yield a negative marginal reward in decision epoch $k + 1$.

6.2 Solution Method

The route-based MDP model we propose for the DDARP aligns with the solution methodology of Schilde et al. (2014). In particular, rather than explicitly considering conventional MDP actions at each epoch, Schilde et al. (2014) work instead with routes that address not only current-period decisions, but which seek to anticipate potential vehicle movement in future scenarios.

At decision epoch k in state s_k, the method of Schilde et al. (2014) begins by inserting new requests realized in ω_k into a route plan θ_k. As in equation (4), each element of the plan is a customer location with an associated four-tuple of data. Then, Schilde et al. (2014) seek to improve the route plan via a variable neighborhood search (VNS) scheme. The VNS evaluates candidate route plans via reward function R^{θ} and by sampling travel times, returning both an action for the current period as well as an updated route plan. Thus, the VNS concurrently searches the space of feasible actions $\mathcal{X}(s_k)$ and the space of route plans $\Theta(s_k)$. Because Schilde et al. (2014) seek to

maximize cost savings achieved by x_k and θ_{k+1} relative to θ_k, the VNS procedure may be viewed as a method to minimize marginal reward $R^{\Delta}((s_k, \theta_k), (x_k, \theta_{k+1}))$.

Connecting the solution approach of Schilde et al. (2014) to the route-based SDVRP via the marginal reward, as well as operating on routes, are strong links between model and method. Similar connections to a route-based MDP model might be made with the works of Bent and Van Hentenryck (2004), Coelho et al. (2014), Gendreau et al. (1999), Ghiani et al. (2012), Klapp et al. (2018b), Powell et al. (2000), Thomas (2007), and Voccia et al. (2019). In each of these papers, the SDVRP solution methods operate on routes, at each epoch simultaneously considering current-period actions and updated routing plans.

7 Takeaways

We present a route-based MDP model for SDVRPs, demonstrating equivalence to conventional MDPs under an easily satisfied condition. Importantly, our modeling approach provides a direct connection between a dynamic and stochastic optimization model and the route-based solution methods found in much of the SDVRP literature. Moreover, the model-method connection we establish provides routing researchers with a framework to clearly describe their problems and related solution approaches.

As a postscript to the formal presentation and explanation of route-based MDPs, we conclude with a list of takeaways, representing the most common questions we have encountered in response to our work. We hope this informal question-and-answer format will serve as a guide to those new to SDVRPs as well as be clarifying to the seasoned SDVRP researcher.

- **Why is a mathematical model required? Are a problem description and a method enough?** Problem descriptions and methods are necessary for scientific study, but they are not sufficient. Without a mathematical model, it is difficult and sometimes impossible to reconstruct research, thus making validation and extension of methods challenging. In contrast to some widely studied problems (e.g., the traveling salesman problem), which are so well known among researchers that a full model formulation may be unnecessary in some papers, SDVRPs are often new and complex, thus warranting detailed mathematical models. In short, rigorous methods should be coupled with rigorous models.

- **If route-based MDPs have larger state- and action-spaces than conventional MDPs, then what is the motivation to use them?** Though in the broad field of operations research we often begin with a model and then develop solution methods, within the subfield of stochastic dynamic vehicle routing, we find a host of solution methods lacking models. While conventional MDPs are capable of filling this void (any model is better than no model at all), route-based MDPs often make a stronger appeal to intuition, connecting methodology more directly to how we often think about dynamic routing applications.

- **Can all stochastic and dynamic routing problems be modeled as a route-based MDP, or is the route-based MDP a model for a specific problem?** Because a route-based MDP is a generalization of an underlying conventional MDP, any conventional MDP can be modeled as a route-based MDP. Practically speaking, route-based MDPs are general enough to model a host of SDVRPs as well as more general dynamic and stochastic optimization problems, particularly those problems where sequential decision-making might benefit from operating on plans rather than only on conventional MDP model elements.

- **Can route-based MDPs be used to model problems outside of dynamic routing?** Though we have focused our discussion on dynamic routing problems, other problem domains should be amenable to plan-based MDPs. In particular, whenever decisions now strongly influence the future, either via future rewards or by shaping the space of future feasible decisions, making tentative plans in the present can be helpful.

- **Route-based MDPs seem tied to route-based solution methods. Should a model be independent of the solution method?** Not necessarily. So long as a model unambiguously defines the problem, customization leaning toward a particular method may strengthen the link between model and method. For example, consider set partitioning and column generation. Though multiple integer programming formulations may accurately represent an application and yield equivalent optimal solutions, models with large numbers of decision variables facilitate decomposition methods that offload difficult portions of the optimization to a subproblem and rely on general branch-and-bound techniques to solve a master problem. Similarly, a route-based MDP can facilitate connections to the rich body of routing heuristics developed by the transportation community.

- **Do the added dimensions of a route-based MDP relative to a conventional MDP just make it more difficult to find an optimal policy?** For most SDVRPs of practical interest, methods to identify an optimal policy are computationally intractable, whether applied to a conventional or route-based MDP. Thus, SDVRPs typically require the heuristic solution procedures that dominate the literature. Further, these procedures typically operate on the added dimensionality inherent to route-based models. Thus, rather than overcomplicating, the added dimensions of the route-based MDP model often lay the groundwork for expressing solution methods in terms of model components.

- **What is the computational evidence that route-based MDPs work better than conventional MDPs?** Solution methods can be gauged in computational terms (e.g., solution quality and CPU seconds), but models only influence computation to the extent that they inform solution method design. Thus, a more appropriate question to ask is whether or not route-based MDPs can aid the design of solution approaches more readily than conventional MDPs. The evidence points strongly toward the affirmative. Route-based solution methods dominate a body of literature that in many cases lacks models entirely and in other cases provides models only loosely connected to the methods. Route-based MDPs provide a model that connects strongly with existing state-of-the-art methods. This portends toward stronger method development in the future.

- **If I already have a conventional MDP model, how can route-based MDPs help me?** Conventional and route-based MDPs are equivalent models, but route-based MDPs may be advantageous when exploring route-based solution methods. In this case, route-based MDPs provide a stronger narrative to connect application and method. As a result, we anticipate improved flow in the write-up of SDVRP research as well as in implementations of models and solution methods.

- **Is the literature review complete?** The aim of our literature review is to demonstrate the need for modeling and the widespread use of route plans in solution methods. Though we do not perform an exhaustive review of the large body of SDVRP studies, our review is representative, pulling dynamic and stochastic routing research from two recent review papers. For further study, we refer to the reviews by Ritzinger et al. (2016); Ulmer (2017).

- **Can you point to instances of route-based MDPs in the literature?** Yes, Ulmer et al. (2017) address dynamic pickup and delivery in the context of restaurant services. In this problem, the conventional MDP action, which typically directs immediate movement of a vehicle, is non-intuitive. The route-based MDP captures a more natural representation of the problem by modeling sequences of pickups and deliveries. Also, Ulmer (2020) examine a same-day delivery problem. In both papers, an action is a set of route plans, one for each vehicle in a fleet. Other works employing route-based MDPs include Ulmer et al. (2019) and Ulmer and Thomas (2019).

Acknowledgements

Justin Goodson wishes to express appreciation for the support of Saint Louis University's Center for Supply Chain Excellence.

References

Angelelli, E., N. Bianchessi, R. Mansini, and M. G. Speranza (2009). Short term strategies for a dynamic multi-period routing problem. *Transportation Research Part C: Emerging Technologies 17*(2), 106–119.

Angelelli, E., R. Mansini, and M. Vindigni (2016). The stochastic and dynamic traveling purchaser problem. *Transportation Science 50*(2), 642–658.

Azi, N., M. Gendreau, and J.-Y. Potvin (2012). A dynamic vehicle routing problem with multiple delivery routes. *Annals of Operations Research 199*(1), 103–112.

Beaudry, A., G. Laporte, T. Melo, and S. Nickel (2010). Dynamic transportation of patients in hospitals. *OR Spectrum 32*(1), 77–107.

Bent, R. W. and P. Van Hentenryck (2004). Scenario-based planning for partially dynamic vehicle routing with stochastic customers. *Operations Research 52*(6), 977–987.

Bent, R. W. and P. Van Hentenryck (2007). Waiting and relocation strategies in online stochastic vehicle routing. In *IJCAI*, pp. 1816–1821.

Berbeglia, G., J.-F. Cordeau, and G. Laporte (2012). A hybrid tabu search and constraint programming algorithm for the dynamic dial-a-ride problem. *INFORMS Journal on Computing 24*(3), 343–355.

Bertsimas, D. J. and G. Van Ryzin (1991). A stochastic and dynamic vehicle routing problem in the Euclidean plane. *Operations Research 39*(4), 601–615.

Chen, Z.-L. and H. Xu (2006). Dynamic column generation for dynamic vehicle routing with time windows. *Transportation Science 40*(1), 74–88.

Coelho, L. C., J.-F. Cordeau, and G. Laporte (2014). Heuristics for dynamic and stochastic inventory-routing. *Computers & Operations Research 52*, 55–67.

Cortés, C. E., D. Sáez, A. Núñez, and D. Muñoz-Carpintero (2009). Hybrid adaptive predictive control for a dynamic pickup and delivery problem. *Transportation Science 43*(1), 27–42.

de Armas, J. and B. Melián-Batista (2015). Variable neighborhood search for a dynamic rich vehicle routing problem with time windows. *Computers & Industrial Engineering 85*(Supplement C), 120 – 131.

Demir, E., K. Huckle, A. Syntetos, A. Lahy, and M. Wilson (2019). Vehicle routing problem: Past and future. In P. Wells (Ed.), *Contemporary Operations and Logistics*, Chapter 7, pp. 97–117. Cham, Switzerland: Palgrave Macmillan.

Desaulniers, G., J. Desrosiers, and M. M. Solomon (2006). *Column generation*, Volume 5. Springer Science & Business Media.

Ehmke, J. F. and A. M. Campbell (2014). Customer acceptance mechanisms for home deliveries in metropolitan areas. *European Journal of Operational Research 233*(1), 193–207.

Fabri, A. and P. Recht (2006). On dynamic pickup and delivery vehicle routing with several time windows and waiting times. *Transportation Research Part B: Methodological 40*(4), 335 – 350.

Ferrucci, F. and S. Bock (2014). Real-time control of express pickup and delivery processes in a dynamic environment. *Transportation Research Part B: Methodological 63*, 1 – 14.

Ferrucci, F. and S. Bock (2016). Pro-active real-time routing in applications with multiple request patterns. *European Journal of Operational Research 253*(2), 356 – 371.

Ferrucci, F., S. Bock, and M. Gendreau (2013). A pro-active real-time control approach for dynamic vehicle routing problems dealing with the delivery of urgent goods. *European Journal of Operational Research 225*(1), 130–141.

Gendreau, M., F. Guertin, J.-Y. Potvin, and R. Séguin (2006). Neighborhood search heuristics for a dynamic vehicle dispatching problem with pick-ups and deliveries. *Transportation Research Part C: Emerging Technologies 14*(3), 157–174.

Gendreau, M., F. Guertin, J.-Y. Potvin, and E. Taillard (1999). Parallel tabu search for real-time vehicle routing and dispatching. *Transportation Science 33*(4), 381–390.

Ghiani, G., E. Manni, A. Quaranta, and C. Triki (2009). Anticipatory algorithms for same-day courier dispatching. *Transportation Research Part E: Logistics and Transportation Review 45*(1), 96–106.

Ghiani, G., E. Manni, and B. W. Thomas (2012). A comparison of anticipatory algorithms for the dynamic and stochastic traveling salesman problem. *Transportation Science 46*(3), 374–387.

Goodson, J. C., J. W. Ohlmann, and B. W. Thomas (2013). Rollout policies for dynamic solutions to the multivehicle routing problem with stochastic demand and duration limits. *Operations Research 61*(1), 138–154.

Goodson, J. C., B. W. Thomas, and J. W. Ohlmann (2016). Restocking-based rollout policies for the vehicle routing problem with stochastic demand and duration limits. *Transportation Science 50*(2), 591–607.

Hillier, F. and G. Lieberman (2001). *Introduction to Operations Research* (7th ed.). McGraw Hill.

Huisman, D. and A. P. Wagelmans (2006). A solution approach for dynamic vehicle and crew scheduling. *European Journal of Operational Research 172*(2), 453 – 471.

Hvattum, L. M., A. Løkketangen, and G. Laporte (2006). Solving a dynamic and stochastic vehicle routing problem with a sample scenario hedging heuristic. *Transportation Science 40*(4), 421–438.

Hvattum, L. M., A. Løkketangen, and G. Laporte (2007). A branch-and-regret heuristic for stochastic and dynamic vehicle routing problems. *Networks 49*(4), 330–340.

Ichoua, S., M. Gendreau, and J.-Y. Potvin (2000). Diversion issues in real-time vehicle dispatching. *Transportation Science 34*(4), 426–438.

Ichoua, S., M. Gendreau, and J.-Y. Potvin (2006). Exploiting knowledge about future demands for real-time vehicle dispatching. *Transportation Science 40*(2), 211–225.

Klapp, M. A., A. L. Erera, and A. Toriello (2018a). The dynamic dispatch waves problem for same-day delivery. *European Journal of Operational Research 271*(2), 519 – 534.

Klapp, M. A., A. L. Erera, and A. Toriello (2018b). The one-dimensional dynamic dispatch waves problem. *Transportation Science 52*(2), 402–415.

Kuo, R., B. Wibowo, and F. Zulvia (2016). Application of a fuzzy ant colony system to solve the dynamic vehicle routing problem with uncertain service time. *Applied Mathematical Modelling 40*(23), 9990 – 10001.

Larsen, A., O. B. G. Madsen, and M. M. Solomon (2002). Partially dynamic vehicle routing-models and algorithms. *Journal of the Operational Research Society 53*(6), 637–646.

Mavrovouniotis, M. and S. Yang (2015). Ant algorithms with immigrants schemes for the dynamic vehicle routing problem. *Information Sciences 294*(Supplement C), 456 – 477. Innovative Applications of Artificial Neural Networks in Engineering.

Maxwell, M. S., M. Restrepo, S. G. Henderson, and H. Topaloglu (2010). Approximate dynamic programming for ambulance redeployment. *INFORMS Journal on Computing 22*(2), 266–281.

Meisel, S. (2011). *Anticipatory Optimization for Dynamic Decision Making*, Volume 51 of *Operations Research/Computer Science Interfaces Series*. Springer.

Mes, M., M. van der Heijden, and P. Schuur (2010). Look-ahead strategies for dynamic pickup and delivery problems. *OR Spectrum 32*(2), 395–421.

Mitrović-Minić, S. and G. Laporte (2004). Waiting strategies for the dynamic pickup and delivery problem with time windows. *Transportation Research Part B: Methodological 38*(7), 635–655.

Ng, K., C. Lee, S. Zhang, K. Wu, and W. Ho (2017). A multiple colonies artificial bee colony algorithm for a capacitated vehicle routing problem and re-routing strategies under time-dependent traffic congestion. *Computers & Industrial Engineering 109*(Supplement C), 151 – 168.

Novoa, C. and R. Storer (2009). An approximate dynamic programming approach for the vehicle routing problem with stochastic demands. *European Journal of Operational Research 196*(2), 509–515.

Papastavrou, J. D. (1996). A stochastic and dynamic routing policy using branching processes with state dependent immigration. *European Journal of Operational Research 95*(1), 167–177.

Pillac, V., M. Gendreau, C. Guéret, and A. L. Medaglia (2013). A review of dynamic vehicle routing problems. *European Journal of Operational Research 225*(1), 1–11.

Pillac, V., C. Guéret, and A. L. Medaglia (2018). *A Fast Reoptimization Approach for the Dynamic Technician Routing and Scheduling Problem*, pp. 347–367. Cham: Springer International Publishing.

Powell, W. (2019). A unified framework for stochastic optimization. *European Journal of Operational Research 275*, 795–821.

Powell, W. B. (2011). *Approximate Dynamic Programming: Solving the Curses of Dimensionality* (Second ed.). Wiley Series in Probability and Statistics. John Wiley & Sons, Inc.

Powell, W. B., M. T. Towns, and A. Marar (2000). On the value of optimal myopic solutions for dynamic routing and scheduling problems in the presence of user noncompliance. *Transportation Science 34*(1), 67–85.

Psaraftis, H. N. (1980). A dynamic programming solution to the single vehicle many-to-many immediate request dial-a-ride problem. *Transportation Science 14*(2), 130–154.

Psaraftis, H. N., M. Wen, and C. A. Kontovas (2016). Dynamic vehicle routing problems: Three decades and counting. *Networks 67*(1), 3–31.

Pureza, V. and G. Laporte (2008). Waiting and buffering strategies for the dynamic pickup and delivery problem with time windows. *INFOR: Information Systems and Operational Research 46*(3), 165–176.

Puterman, M. L. (2005). *Markov decision processes: discrete stochastic dynamic programming*. Wiley Series in Probability and Statistics. Hoboken, New Jersey: John Wiley & Sons, Inc.

Ritzinger, U., J. Puchinger, and R. F. Hartl (2016). A survey on dynamic and stochastic vehicle routing problems. *International Journal of Production Research 54*(1), 215–231.

Sáez, D., C. E. Cortés, and A. Núñez (2008). Hybrid adaptive predictive control for the multi-vehicle dynamic pick-up and delivery problem based on genetic algorithms and fuzzy clustering. *Computers & Operations Research 35*(11), 3412–3438.

Sarasola, B., K. F. Doerner, V. Schmid, and E. Alba (2015). Variable neighborhood search for the stochastic and dynamic vehicle routing problem. *Annals of Operations Research*, 1–37.

Savelsbergh, M. and M. Sol (1998). Drive: Dynamic routing of independent vehicles. *Operations Research 46*(4), 474–490.

Savelsbergh, M. and T. Van Woensel (2016). 50th anniversary invited article—city logistics: Challenges and opportunities. *Transportation Science 50*(2), 579–590.

Schilde, M., K. F. Doerner, and R. F. Hartl (2014). Integrating stochastic time-dependent travel speed in solution methods for the dynamic dial-a-ride problem. *European Journal of Operational Research 238*(1), 18–30.

Schmid, V. (2012). Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. *European Journal of Operational Research 219*(3), 611–621.

Schyns, M. (2015). An ant colony system for responsive dynamic vehicle routing. *European Journal of Operational Research 245*(3), 704 – 718.

Secomandi, N. (2000). Comparing neuro-dynamic programming algorithms for the vehicle routing problem with stochastic demands. *Computers & Operations Research 27*(11), 1201–1225.

Secomandi, N. (2001). A rollout policy for the vehicle routing problem with stochastic demands. *Operations Research 49*(5), 796–802.

Secomandi, N. and F. Margot (2009). Reoptimization approaches for the vehicle-routing problem with stochastic demands. *Operations Research 57*(1), 214–230.

Sheridan, P. K., E. Gluck, Q. Guan, T. Pickles, B. Balcıog, B. Benhabib, et al. (2013). The dynamic nearest neighbor policy for the multi-vehicle pick-up and delivery problem. *Transportation Research Part A: Policy and Practice 49*, 178–194.

Speranza, M. G. (2018). Trends in transportation and logistics. *European Journal of Operational Research 264*(3), 830–836.

Srour, F. J., N. Agatz, and J. Oppen (2018). Strategies for handling temporal uncertainty in pickup and delivery problems with time windows. *Transportation Science 52*(1), 3–19.

Swihart, M. R. and J. D. Papastavrou (1999). A stochastic and dynamic model for the single-vehicle pick-up and delivery problem. *European Journal of Operational Research 114*(3), 447–464.

Tassiulas, L. (1996). Adaptive routing on the plane. *Operations Research 44*(5), 823–832.

Thomas, B. W. (2007). Waiting strategies for anticipating service requests from known customer locations. *Transportation Science 41*(3), 319–331.

Thomas, B. W. and C. C. White III (2004). Anticipatory route selection. *Transportation Science 38*(4), 473–487.

Tirado, G. and L. M. Hvattum (2016). Determining departure times in dynamic and stochastic maritime routing and scheduling problems. *Flexible Services and Manufacturing Journal*, 1–19.

Tirado, G. and L. M. Hvattum (2017, Jun). Improved solutions to dynamic and stochastic maritime pick-up and delivery problems using local search. *Annals of Operations Research 253*(2), 825–843.

Tirado, G., L. M. Hvattum, K. Fagerholt, and J.-F. Cordeau (2013). Heuristics for dynamic and stochastic routing in industrial shipping. *Computers & Operations Research 40*(1), 253–263.

Toriello, A., W. B. Haskell, and M. Poremba (2014). A dynamic traveling salesman problem with stochastic arc costs. *Operations Research 62*(5), 1107–1125.

Ulmer, M. W. (2017). *Approximate Dynamic Programming for Dynamic Vehicle Routing*, Volume 61 of *Operations Research/Computer Science Interfaces Series*. Springer.

Ulmer, M. W. (2020). Dynamic pricing and routing for same-day delivery. *Transportation Science*.

Ulmer, M. W., J. C. Goodson, D. C. Mattfeld, and M. Hennig (2019). Offline-online approximate dynamic programming for dynamic vehicle routing with stochastic requests. *Transportation Science 53*(1), 185–202.

Ulmer, M. W., D. C. Mattfeld, and F. Köster (2018). Budgeting time for dynamic vehicle routing with stochastic customer requests. *Transportation Science 52*(1), 20–37.

Ulmer, M. W. and B. W. Thomas (2019). Meso-parametric value function approximation for dynamic customer acceptances in delivery routing. *European Journal of Operational Research*.

Ulmer, M. W., B. W. Thomas, A. M. Campbell, and N. Woyak (2017). The restaurant meal delivery problem: Dynamic pick-up and delivery with deadlines and random ready times. *Working Paper*.

Ulmer, M. W., B. W. Thomas, and D. C. Mattfeld (2019). Preemptive depot returns for dynamic same-day delivery. *EURO journal on Transportation and Logistics 8*(4), 327–361.

Van Hemert, J. I. and J. A. La Poutré (2004). Dynamic routing problems with fruitful regions: Models and evolutionary computation. In *Parallel Problem Solving from Nature-PPSN VIII*, pp. 692–701. Springer.

Voccia, S. A., A. M. Campbell, and B. W. Thomas (2019). The same-day delivery problem for online purchases. *Transportation Science 53*(1), 167–184.

Wang, X. and H. Kopfer (2015, Dec). Rolling horizon planning for a dynamic collaborative routing problem with full-truckload pickup and delivery requests. *Flexible Services and Manufacturing Journal 27*(4), 509–533.

Zhang, S., J. W. Ohlmann, and B. W. Thomas (2018). Dynamic orienteering on a network of queues. *Transportation Science 52*(3), 691–706.

Appendix: Model Equivalence

In this appendix, we make a formal connection between the conventional MDP model of Section 4 and the route-based MDP model of Section 5. Notably, for a given problem instance, we show the optimal value of a state for a suitably defined route-based MDP is related to the optimal value of a state in the conventional model, and that from an initial state the optimal values are the same. The result means that optimal policy values of conventional and route-based MDPs are the same and that we can easily derive an optimal policy for the conventional MDP model from the optimal policy of a route-based MDP.

To introduce the result, we first discuss the Bellman Equation. Often referred to as a value function or as the optimality equation, the Bellman Equation is a way of expressing the maximum reward that may be accumulated from a given state s_k onward to a terminal state. The optimality of the Bellman equation follows from MDP model assumptions (see Puterman (2005) for a detailed presentation, properties, and derivation). For a conventional MDP, the optimal value of a state is

$$V(s_k) = \max_{x_k \in \mathcal{X}(s_k)} \left\{ R_k(s_k, x_k) + \mathbb{E}\left[V(s_{k+1}) \middle| s_k^x \right] \right\}, \tag{5}$$

where the first term is the current-period reward and the second term, often referred to as the reward-to-go, is the value of the post-decision state. As is common, for all terminal post-decision states s_K^x we assume a null value for the reward-to-go:

$$\mathbb{E}\left[V(s_{K+1}) | s_K^x \right] = 0. \tag{6}$$

Similarly, in the route-based case, the Bellman Equation to calculate the optimal value of a

state is

$$V^{\theta}(s_k, \theta_k) = \max_{(x_k, \theta_{k+1}) \in \mathcal{X}(s_k) \times \Theta_{k+1}(s_k)} \left\{ R_k^{\Delta}((s_k, \theta_k), (x_k, \theta_{k+1})) + \mathbb{E}\left[V^{\theta}(s_{k+1}, \theta_{k+1}) \middle| s_k^{x_k} \right] \right\} \quad (7)$$

and assume for all s_K^x

$$\mathbb{E}\left[V^{\theta}(s_{K+1}, \theta_{K+1}) | s_K^x \right] = 0. \quad (8)$$

While similar in form to the conventional Bellman Equation, the route-based Bellman Equation differs in two important ways. One, the first term of Equation (7) pulls reward into the current period that is captured in the second term of Equation (5). Two, the second term of Equation (7) is the value of the route-based post-decision state, but following the definition of R^{Δ}, it represents the expected change in marginal rewards rather than an accumulation of rewards as in Equation (5).

Though Equations (5) and (7) have different interpretations, Proposition 1 provides a relation between the conventional and route-based value functions. The proposition requires the values of initial and terminal route plans be zero. Condition 1 formalizes the requirement:

Condition 1. *The value of the initial route plan θ_0 and the terminal route plan θ_{K+1} are zero:* $R^{\theta}(\theta_0) = R^{\theta}(\theta_{K+1}) = 0$.

Proposition 1 states the value functions of the two formulations follow the simple relation $V^{\theta}(s_k, \theta_k) = V(s_k) - R^{\theta}(\theta_k)$. Thus, with Condition 1, an optimal policy for each problem achieves the same value from the initial state and the models are equivalent. The intuition behind the relation follows from recognizing the optimal value of a state in the route-based MDP results not only from current-period rewards but also from route plan values. Thus, relative to the conventional MDP, the route-based MDP accumulates value earlier in the horizon.

For an instance of the VRPSSR, Figure 5 depicts how rewards might accrue while servicing six customer requests in the conventional and route-based MDP models. In the conventional model, we observe unit increases at each decision epoch as a result of visiting a status-1 customer. The route-based model achieves the same value of six achieved by the conventional model, but in contrast we observe up-and-down value fluctuations across decision epochs. A rise in value indicates an increase in the marginal number of customers served by an updated routing plan whereas a decrease in value signifies a decline in the marginal number of planned visits.

We now formally state and prove the proposition.

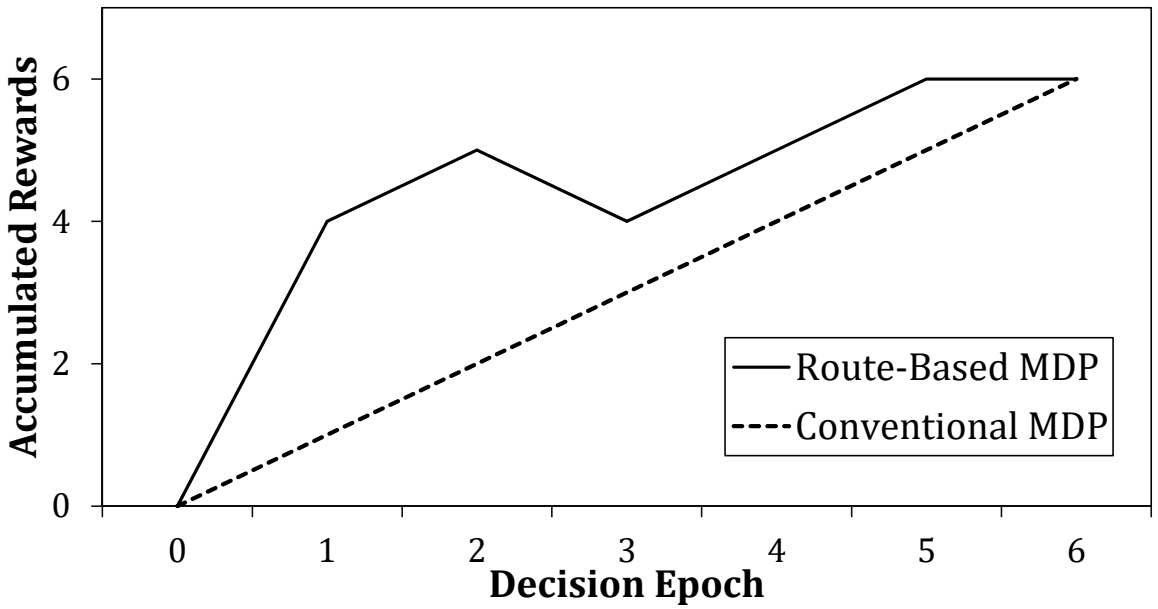

Figure 5: Accumulated Reward Over Time for Route-Plan versus Conventional MDP

Proposition 1 (Value Function Relation)**.** *If Condition 1 is satisfied, then* $V^{\theta}(s_k, \theta_k) = V(s_k) - R^{\theta}(\theta_k)$ *for all* s_k *in* $\mathcal{S}$*, for all* θ_k *in* $\Theta(s_k)$*, and for* $k = 0, 1, \ldots, K$.

Proof. The proof is by induction. We first show the result for terminal period K:

$$
\begin{aligned}
V^{\theta}(s_K, \theta_K) &= \max_{(x_K, \theta_{K+1}) \in \mathcal{X}(s_K) \times \Theta(s_K)} \left\{ R_K^{\Delta}\left((s_K, \theta_K), (x_K, \theta_{K+1})\right) \right\} \\
&= \max_{(x_K, \theta_{K+1}) \in \mathcal{X}(s_K) \times \Theta(s_K)} \left\{ R_K(s_K, x_K) + R^{\theta}\left(\theta_{K+1}\right) - R^{\theta}\left(\theta_K\right) \right\} \qquad (9) \\
&= -R^{\theta}\left(\theta_K\right) + \max_{(x_K, \theta_{K+1}) \in \mathcal{X}(s_K) \times \Theta(s_K)} \left\{ R_K\left(s_K, x_K\right) \right\} \qquad (10) \\
&= -R^{\theta}\left(\theta_K\right) + \max_{x_K \in \mathcal{X}(s_k)} \left\{ R_K\left(s_K, x_K\right) \right\} \qquad (11) \\
&= V(s_K) - R^{\theta}\left(\theta_K\right). \qquad (12)
\end{aligned}
$$

The first equality follows from the definition of the route-based Bellman Equation and the assumption of Equation (8). Equation (9) follows from the definition of the marginal reward function for the route-based MDP. Equation (10) follows from Condition 1 and the fact that $R^{\theta}\left(\theta_K\right)$ is constant

over the maximization. The equality in Equation (11) is valid because the selection of a plan θ_{K+1} does not impact $R_K(s_K, x_K)$. Equation (12) follows from Equations (5) and (6).

We assume the result holds for periods $k+1, k+2, \ldots, K-1$. Then, for period k:

$$
\begin{aligned}
V^{\theta}(s_k, \theta_k) &= \max_{(x_k, \theta_{k+1}) \in \mathcal{X}(s_k) \times \Theta(s_k)} \left\{ R_k^{\Delta}\left((s_k, \theta_k), (x_k, \theta_{k+1})\right) + \mathbb{E}\left[V^{\theta}\left(s_{k+1}, \theta_{k+1}\right) \middle| s_k^{x_k} \right] \right\} \\
&= \max_{(x_k, \theta_{k+1}) \in \mathcal{X}(s_k) \times \Theta(s_k)} \Big\{ R_k(s_k, x_k) + R^{\theta}\left(\theta_{k+1}\right) - R^{\theta}\left(\theta_k\right) \\
&\qquad + \mathbb{E}\left[V(s_{k+1}) - R^{\theta}(\theta_{k+1}) \middle| s_k^{x_k} \right] \Big\} \qquad (13) \\
&= \max_{(x_k, \theta_{k+1}) \in \mathcal{X}(s_k) \times \Theta(s_k)} \Big\{ R_k\left(s_k, x_k\right) + R^{\theta}\left(\theta_{k+1}\right) - R^{\theta}\left(\theta_k\right) \\
&\qquad - R^{\theta}\left(\theta_{k+1}\right) + \mathbb{E}\left[V(s_{k+1}) | s_k^{x_k} \right] \Big\} \qquad (14) \\
&= -R^{\theta}(\theta_k) + \max_{(x_k, \theta_{k+1}) \in \mathcal{X}(s_k) \times \Theta(s_k)} \left\{ R_k(s_k, x_k) + \mathbb{E}\left[V(s_{k+1}) \middle| s_k^{x_k} \right] \right\} \qquad (15) \\
&= -R^{\theta}(\theta_k) + \max_{x_k \in \mathcal{X}(s_k)} \left\{ R_k(s_k, x_k) + \mathbb{E}\left[V(s_{k+1}) \middle| s_k^{x_k} \right] \right\} \qquad (16) \\
&= V(s_k) - R^{\theta}(\theta_k). \qquad (17)
\end{aligned}
$$

The first equality follows from the definition of the Bellman Equation for the route-based MDP. Equation (13) follows from the definition of the marginal reward for the route-based MDP and from the induction hypothesis. Equation (14) acknowledges the value of route plan θ_{k+1} can be calculated separately from the post-decision value of s_{k+1}. Equation (15) recognizes the value of θ_k depends on neither x_k nor θ_{k+1}. The equality in Equation (16) recognizes that neither $R_k(s_k, x_k)$ nor $\mathbb{E}\left[V(s_{k+1}) \middle| s_k^{x_k} \right]$ are affected by θ_{k+1}. Finally, Equation (17) follows from the definition of the period-k value function in the conventional MDP. □