Transportation Research Part C 92 (2018) 278–297

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Dynamic autonomous vehicle fleet operations: Optimization-based strategies to assign AVs to immediate traveler demand requests☆

Michael Hyland, Hani S. Mahmassani*

Northwestern University Transportation Center, 600 Foster Street, Evanston, IL 60208, United States

ARTICLE INFO

Keywords:
Autonomous vehicles
Mobility service
Fleet management
Assignment problem
Dynamic vehicle routing
Agent-based simulation

ABSTRACT

Motivated by the growth of ridesourcing services and the expected advent of fully-autonomous vehicles (AVs), this paper defines, models, and compares assignment strategies for a shared-use AV mobility service (SAMS). Specifically, the paper presents the *on-demand SAMS with no shared rides*, defined as a fleet of AVs, controlled by a central operator, that provides direct origin-to-destination service to travelers who request rides via a mobile application and expect to be picked up within a few minutes. The underlying operational problem associated with the on-demand SAMS with no shared rides is a sequential (i.e. dynamic or time-dependent) stochastic control problem. The AV fleet operator must assign AVs to open traveler requests in real-time as traveler requests enter the system dynamically and stochastically. As there is likely no optimal policy for this sequential stochastic control problem, this paper presents and compares six AV-traveler assignment strategies (i.e. control policies). An agent-based simulation tool is employed to model the dynamic system of AVs, travelers, and the *intelligent* SAMS fleet operator, as well as, to compare assignment strategies across various scenarios. The results show that optimization-based AV-traveler assignment strategies, strategies that allow en-route pickup AVs to be diverted to new traveler requests, and strategies that incorporate en-route drop-off AVs in the assignment problem, reduce fleet miles and decrease traveler wait times. The more-sophisticated AV-traveler assignment strategies significantly improve operational efficiency when fleet utilization is high (e.g. during the morning or evening peak); conversely, when fleet utilization is low, simply assigning traveler requests sequentially to the nearest idle AV is comparable to more-advanced strategies. Simulation results also indicate that the spatial distribution of traveler requests significantly impacts the empty fleet miles generated by the on-demand SAMS.

1. Introduction

The individually-owned and -operated vehicle has dominated passenger transportation for over sixty years in the United States (Mckenzie and Rapino, 2011). However, over the past decade, carsharing, ridesharing, and especially ridesourcing services have seen significant growth in the passenger transportation market (Clewlow and Mishra, 2017). Fully-autonomous vehicles (AVs) should accelerate the growth of these mobility services via eliminating the labor costs of human drivers and subsequently allowing shared-use AV mobility services (SAMSs) to compete with the personal (autonomous) vehicle in terms of cost, convenience, and service quality for nearly all trip purposes (Mahmassani, 2016a, 2016b). In addition to labor costs, AVs also eliminate the performance limitations of human drivers, including hours-of-service constraints and slow human reaction times.

☆ This article belongs to the Virtual Special Issue on “Operations & automation”.

* Corresponding author.
E-mail addresses: michaelhyland2013@u.northwestern.edu (M. Hyland), masmah@northwestern.edu (H.S. Mahmassani).

https://doi.org/10.1016/j.trc.2018.05.003
Received 31 July 2017; Received in revised form 1 April 2018; Accepted 2 May 2018
Available online 15 May 2018

This paper focuses on the real-time operation/control of an SAMS fleet, as such, the most-relevant operational-level advantage of AVs is their ability to safely and near-instantaneously receive and execute changes in vehicle plans (e.g. routes, schedules, and traveler assignments) coming from the fleet operator (i.e. a central computing system). From a fleet management perspective, the biggest advantage of AVs is their guaranteed compliance with these real-time plan changes, and more generally the fleet manager's operational policies. Although it is possible to require taxi and ridesourcing drivers to follow the fleet manager's operation policies and fleet operator's real-time instructions, the fact that taxi and ridesourcing services give drivers considerable autonomy suggests that driver compliance may be too difficult to mandate and/or ineffective in practice. With complete operational control, the fleet operator can maximize the profit of the fleet rather than allowing drivers to maximize their own profit. Moreover, the fleet operator can do so with near-perfect information about the location and state of AVs and open traveler requests, as well as the ability to near-instantaneously re-plan AV routes, schedules, and traveler assignments in real-time as new traveler requests enter the system.

Motivated by the cost and performance benefits of AVs, the potential for SAMSs to compete with the personal vehicle, the ability of SAMS fleet operators to completely control individual vehicles, and the importance of operational efficiency, this paper examines the underlying operational problem associated with an *on-demand SAMS with no shared rides*. The SAMS's characteristics are as follows:

- Travelers request rides dynamically via a mobile application; a request includes pickup and drop-off locations, both within a pre-defined geographical service region
- Travelers want to be served (i.e. picked up) immediately
- Travelers will always be served, if they are willing to wait, i.e. the fleet operator cannot reject traveler requests
- AVs transport travelers directly from their requested pickup location to their drop-off location, i.e. no en-route detours to pick up or drop off other travelers
- AVs in the fleet are functionally homogeneous
- The fleet size is fixed in the short term (i.e. one-day)
- The fleet operator has complete control over each AV
- The fleet operator seeks to minimize fleet miles and traveler wait times

Aside from the vehicles being driverless and the fleet operator having complete control over the AVs, this SAMS differs from existing ridesourcing services in that it has a fixed fleet size. The fleet size is assumed to be fixed as ridesourcing companies, technology companies, and car manufacturers have stated that they plan to provide mobility services with their own AV fleet, rather than sell individual AVs to travelers (Waymo, 2017; Wingfield, 2017).

This paper focuses on the underlying problem associated with operating an on-demand SAMS with no shared rides, which is a stochastic (future traveler requests are unknown) and highly-dynamic (travelers want to be picked up soon after requesting a ride) vehicle routing problem. This operational problem shares many features with existing dynamic vehicle routing problems in the academic literature, such as the taxi- and ambulance- dispatching problems, the personal rapid transit problem, and the dynamic truckload pickup and delivery problem. However, Section 2 outlines the uniqueness of the on-demand SAMS with no shared rides operational problem. The original contributions of this paper include:

- defining the *on-demand SAMS with no shared rides* and its underlying operational problem
- modeling the on-demand SAMS with no shared rides operational problem as a dynamic AV-traveler assignment problem
- presenting intelligent optimization-based AV-traveler assignment strategies (i.e. control policies, or solution algorithms) that consider the unique characteristics of the SAMS; e.g. the ability to safely and near-instantaneously re-assign en-route pickup AVs to new traveler requests
- testing these intelligent optimization-based strategies against simple, yet common, AV-traveler assignment strategies employed in the literature and in practice.

The remainder of the paper is structured as follows. Section 2 reviews the existing literature related to SAMSs and dynamic routing problems. Section 3 formally defines the on-demand SAMS with no shared rides operational problem and presents a mathematical formulation of the AV-traveler assignment problem. Section 4 presents six AV-traveler assignment strategies. Section 5 briefly describes the agent-based simulation framework employed to model the dynamic system of travelers, AVs, and the intelligent SAMS fleet operator. Section 6 lists the experiments designed to compare the six AV-traveler assignment strategies and presents the computational results. Finally, Section 7 concludes the paper.

2. Background

2.1. Shared-use AV mobility services

Over the past five years, researchers have begun to study SAMSs, such as shared-use automated taxi (aTaxi) services (Ford, 2012), shared-use autonomous vehicle (SAV) services with and without shared rides (Fagnant and Kockelman, 2015, 2014), and automated Mobility on Demand (aMoD) services (Spieser et al., 2014). This paper uses the SAMS acronym to be as general as possible when referring to shared-use AVs providing some form of mobility service.

Existing supply-side research on SAMSs generally involves an agent-based modeling framework with three components including a demand (i.e. traveler request) generator, an SAMS fleet operator/dispatcher, and some representation of the transportation network (Levin et al., 2017; Rigole, 2014).

The demand generator is relatively straightforward; the simulation model generates traveler requests with an origin location, destination location, and request time; the approach is rather well established in trucking fleet research (e.g. Regan et al., 1998). Researchers have calibrated their spatio-temporal traveler demand generators using synthetic travel demand from regions such as Austin (Chen et al., 2016; Fagnant et al., 2015; Fagnant and Kockelman, 2016; Levin et al., 2017), New Jersey (Zachariah et al., 2014; Zhu and Kornhauser, 2017), Lisbon (Martinez and Viegas, 2017; Viegas and Martinez, 2016), Berlin (Bischoff and Maciejewski, 2016) and Zurich (Boesch et al., 2016). Other researchers use taxi data (Burns et al., 2013) or the National Household Travel Survey information (Zhang et al., 2015a) to obtain spatial-temporal demand distributions.

In regards to the representation of the road network, seminal SAMS research models taxi stands but not the road network connecting the taxi stands (Ford, 2012; Zachariah et al., 2014). Other researchers use Manhattan grid networks (Fagnant and Kockelman, 2014) and Euclidean planes (Spieser et al., 2014) as abstractions of road networks. More-advanced network representations include quasi-dynamic grid-based (Zhang et al., 2015a, 2015b) and quasi-dynamic actual road (Fagnant et al., 2015; Fagnant and Kockelman, 2016; International Transport Forum, 2015; Martinez and Viegas, 2017) networks with time-dependent, but deterministic link travel times. Recent planning research focused on impact assessment employs dynamic traffic simulation software such as MATSIM (Bischoff and Maciejewski, 2016) and a cell-transmission simulation model (Levin et al., 2017) to model the AVs in a congestible road network. As the current study focuses on the SAMS fleet dispatching problem, the study employs a Manhattan grid network with static travel times.

The fleet dispatchers in most existing SAMS research use simplistic rules to assign AVs to travelers. Burns et al. (2013) assign travelers FCFS to the nearest idle or en-route drop-off AV. Zhang et al. (2015b) only consider idle AVs in their FCFS assignment strategy. Several researchers use a rule-based strategy that involves segmenting the service region into sub-regions, and assigning unassigned travelers (ordered randomly) to the closest idle AVs within their sub-region (Chen et al., 2016; Fagnant and Kockelman, 2014). If no idle AV is available in the traveler's sub-region, the traveler looks to the surrounding sub-regions. Boesch et al. (2016) employ a similar rule-based dispatching strategy. Fagnant et al. (2015) use a similar strategy in a road network setting and employ a modified-Djikstra's algorithm to determine the shortest network path between idle AVs and unassigned travelers. Bischoff and Maciejewski (2016) model a large-scale SAMS and use a slightly more-sophisticated assignment strategy that involves classifying the system state into two mutually exclusive categories. If there is an oversupply of AVs relative to unassigned traveler requests, travelers are assigned FCFS to the nearest idle AV. If there is an undersupply of AVs, when an AV becomes idle it is assigned to the nearest unassigned traveler request. The current paper employs optimization-based strategies to dynamically assign AVs to traveler requests, and additionally allows previously assigned travelers (en-route pickup AVs) to be reassigned (diverted) to other AVs (travelers) after new traveler requests enter the system.

The existing supply-side SAMS research focuses on the transportation planning and policy aspects of SAMSs, including the impacts of SAMSs on traffic congestion, vehicle miles traveled, and vehicle emissions. Other researchers provide estimates of the required fleet sizes for SAMSs (Boesch et al., 2016; Winter et al., 2016). Conversely, this paper focuses on the underlying operational problem associated with the on-demand SAMS with no shared rides. The models and solution algorithms presented in the current study aim to improve SAMS fleet efficiency, drive down operational costs, and improve service quality. There is certainly overlap between these two research aims, nevertheless they necessitate modeling frameworks that focus on different aspects of SAMS systems.

For example, the current study aims to illustrate that more-advanced optimization-based AV-traveler assignment strategies significantly outperform existing strategies in the literature. To prevent other factors such as network structure and AV refueling strategies from confounding the comparison of AV-traveler assignment strategies, this study employs a Manhattan grid-network and ignores AV refueling.

2.2. Dynamic routing problem

The dynamic routing problem (DRP) literature is broad and diverse. After seminal work on the dynamic dial-a-ride problem (D-DARP) nearly forty years ago (Psaraftis, 1980), researchers have been working on models and solution algorithms for various DRP applications. DRP applications include taxi services, ambulance and other emergency services, paratransit services, personal rapid transit services, and freight trucking services. DRP applications involve a vehicle fleet providing transportation service to demand requests that arrive dynamically and randomly, and require a fleet operator to assign vehicles to demand requests in real-time (Regan et al., 1995, 1996a; Powell, 1996).

Although all DRPs, by definition, are dynamic, the degree of dynamism varies considerably across applications. Lund et al. (1996) proposed a degree of dynamism (DoD) metric defined as the ratio of the number of dynamic demand requests (n_d) to the total number of requests (n_{tot}); ($DoD = n_d/n_{tot}$), wherein a dynamic request occurs while the vehicle fleet is providing service rather than before the start of the day. By definition, all the requests are dynamic in the on-demand SAMS presented in Section 1; therefore its $DoD = 1$.

The effective-DoD ($EDoD$) extends the DoD metric via considering the time individual demand requests become known to the operator, and the latest possible time the request could be received (Larsen et al., 2002). If r_i is the request time of demand i, and T is the length of the planning horizon, then $EDOD = \left(\sum_{i=1}^{n_d} \frac{r_i}{T}\right)/n_{tot}$. Larsen et al. (2002) also introduce the $EDoD$ with time windows ($EDoDTW$) that considers the gap between traveler i's latest allowable departure time (l_i) and her request time (r_i). The $EDoDTW$ is defined as follows: $EDoDTW = \frac{1}{n_{tot}} \sum_{i=1}^{n_d} \left(1 - \frac{l_i - r_i}{T}\right)$.

As the on-demand SAMS with no shared rides operational problem includes soft, rather than hard, time window constraints, l_i is not explicitly defined. However, $l_i - r_i$ is implicitly very small by the definition in Section 1; travelers want to be picked up shortly after requesting a ride. Larsen et al. (2002) classify DRPs as weakly dynamic (e.g. distribution of gas and oil to households; the DARP for the elderly and physically disabled), moderately dynamic (e.g. overnight mail services; appliance repair), or strongly dynamic (e.g. police, fire, and ambulance services; taxicab services). Other researchers introduce the notion of disclosure dates for demand requests and quantify the value of advanced information for the traveling salesman and repairman problems (Jaillet and Wagner, 2006). Additionally, van Lon et al. (2016) differentiate between dynamism and urgency and present metrics for both measures.

The dynamic truckload pickup and delivery problem (D-TLPDP) is moderately dynamic; however, much of the research on DRPs with no shared rides comes from the D-TLPDP. According to the classification presented by Berbeglia et al. (2010), the D-TLPDP is one of three types of dynamic pickup and delivery problems (D-PDP), along with the D-DARP and the dynamic vehicle routing problem with pickup and delivery (D-VRPPD). Unlike the D-DARP and the D-VRPPD, the D-TLPDP only allows one demand request in a vehicle at a time. As this paper analyzes an SAMS with no shared rides, the D-TLPDP is most relevant. The shared-ride SAMS operational problem is essentially a D-DARP. There are several recent DRP reviews in the literature (Pillac et al., 2013; Psaraftis et al., 2016) as well as a review on dynamic *and* stochastic routing problems (Ritzinger et al., 2015).

To solve the D-TLPDP, researchers commonly employ a rolling-horizon solution procedure that involves repeatedly re-solving a static mathematical programming problem (Fleischmann et al., 2004; Frantzeskakis and Powell, 1990; Yang et al., 1999, 2004). For the D-TLPDP, the two most-common static mathematical programming problems are the TLPDP and the assignment (or weighed bipartite matching) problem. The TLPDP model allows the fleet operator to sequence (or schedule) the pickup and drop-off of several open demand requests, for each vehicle; whereas, the assignment problem matches each vehicle to zero or one open demand request. The current paper employs the assignment problem for several reasons. First, the assignment problem is a linear programming problem, meaning *exact* solutions to relatively large problem instances can be obtained reasonably quickly; conversely, the D-TLPDP is an NP-hard integer programming problem. As the on-demand SAMS problem is highly-dynamic, it is beneficial that the assignment problem can be quickly re-solved in real-time. Second, the benefits of sequencing traveler requests are minimal for the on-demand SAMS with no shared rides problem because travelers want to be picked up within a few minutes of making a request. If a traveler is assigned to be the third pick up in an AV's schedule (i.e. queue), this traveler's wait time would be unreasonably high for an on-demand service. Two assignment strategies presented in the current paper implicitly allow an AV to sequence the drop-off of one traveler and the pickup and drop-off of a second traveler, without losing the computational efficiency of the assignment problem.

The taxi-dispatching problem with immediate demand requests is possibly the area of research most related to the on-demand SAMS with no shared rides operational problem. Seow et al. (2010) discuss the inefficiencies associated with assigning individual taxi requests FCFS to the nearest idle taxi. However, unlike the on-demand SAMS presented in this paper, Seow et al. (2010) model a taxi service that allows travelers to be rejected.

Maciejewski and Nagel (2013) present three different strategies for the taxi-dispatching problem with immediate requests. The first strategy assigns travelers FCFS to the nearest idle taxi; the second and third strategy consider idle and en-route drop-off AVs in the assignment. Maciejewski and Nagel (2013) explicitly consider the case wherein the demand rate of taxi requests temporarily outpaces the service rate of taxis. As a queue of unserved traveler requests forms, the fleet operator assigns multiple ordered requests to each taxi. In the third strategy, demand requests in the taxi queues can be re-assigned to other taxis as new information enters the system. Like the current study, Maciejewski et al. (2016) use the assignment problem framework to dispatch taxis to immediate traveler requests. However, unlike the current study, Maciejewski et al. (2016) do not allow en-route pickup taxis (assigned travelers) to be diverted (reassigned).

Like the on-demand SAMS with no shared rides, personal rapid transit (PRT) involves driverless vehicles/pods transporting travelers without ride-sharing. However, the PRT requires fixed guideways and pre-defined pickup and drop-off locations, i.e. stations (Lees-Miller, 2016). This design feature alters the underlying problem considerably.

The ambulance fleet management problem can be divided into an ambulance-dispatching problem and an ambulance relocation problem (Gendreau et al., 2001). Early research focuses on the ambulance relocation problem and employs simple heuristics, such as a nearest idle ambulance policy, for the ambulance-dispatching problem (Andersson and Värbrand, 2007; Gendreau et al., 2001). More-advanced methods use mathematical programming formulations that jointly consider the ambulance dispatch and relocation problem (Haghani and Yang, 2007). Other researchers employ approximate dynamic programming methods to solve the ambulance dispatching and relocation problem (Schmid, 2012). Lee (2012) examines the ambulance-dispatching problem in the presence of a disaster wherein the demand rate of calls outpaces the ambulance fleet's service rate.

The DRPs presented in this subsection share many attributes with the SAMS described in Section 1. However, the authors believe the *combination* of attributes associated with this SAMS operational problem is unique in the academic literature. The problem is significantly more dynamic than freight transportation problems in the literature and even the existing taxi-dispatching literature wherein previous research assumes *immediate* taxi requests should be picked up within 30 min (Wang et al., 2009). Moreover, the authors are unaware of taxi-dispatching strategies in the literature that allow assigned travelers (taxis en-route to pick up a traveler) to be reassigned to a different taxi (diverted to a different traveler request).

The ambulance-dispatching problem clearly matches the dynamism and urgency associated with the on-demand SAMS with no shared rides operational problem. However, relative to on-demand SAMS problem instances, ambulance-dispatching problem instances are typically smaller, have a less-diverse spatial distribution of demand destinations (i.e. the patients are either treated at their pickup location or transported to one of a few local hospitals), and involve significantly longer pickup and drop-off times in terms of mean and variance. These differences significantly impact the operational problem, applicable models, and solution algorithms. In fact, the current paper tests solution algorithms that involve assigning en-route drop-off AVs to new traveler requests. This is not advisable for an ambulance dispatcher because dropping patients off at the hospital is a time-consuming process that is also unpredictable. The ambulance dispatcher, unlike the SAMS operator, cannot assume the ambulance will be available less than a minute after arriving at the hospital.

3. Model

This section presents a formal description of the on-demand SAMS with no shared rides operational problem using the notation in the Appendix. The problem is a sequential stochastic control problem, wherein the SAMS fleet operator assigns AVs to dynamic traveler requests in real-time.

3.1. Problem statement

The on-demand SAMS with no shared rides operational problem is characterized by an AV fleet $V = \{1,2,...|V|\}$ serving traveler requests $R = \{1,2,...|R|\}$ that arrive dynamically and randomly over the course of a finite horizon $T = [0,|T|]$. The requests arrive randomly according to a Poisson process with spatial-temporal demand rate λ, which is unknown to the SAMS fleet operator. Each traveler's origin (o_i) and destination (d_i) can be located anywhere within a pre-defined service region. Each traveler's request time ($r_i \in T$) also denotes her earliest pickup time, her desired pickup time, and the time the SAMS fleet operator becomes aware of the request; time lags are omitted for clarity, but with no loss of generality.

At every time instant $t > r_i$, traveler $i \in R$ has a physical location $posR_i(t)$, state $\alpha_i(t)$ and elapsed wait time $w_i(t)$, wherein $w_i(t) = t - r_i$. At $t > r_i$, traveler $i \in R$ can be in one of four states $\alpha_i(t) \in \{1,2,3,4\}$ denoting unassigned, assigned, in-vehicle, and served, respectively. These four states correspond to four mutually exclusive subsets of travelers. Unassigned travelers $R_U(t) = \{R\ |t\rangle r_i, \alpha_i(t) = 1\}$ have made a request but have not yet been assigned to an AV. Assigned travelers $R_A(t) = \{R\ |t\rangle r_i, \alpha_i(t) = 2\}$ have been assigned to an AV but have not yet been picked up. In-vehicle travelers $R_{IV}(t) = \{R\ |t\rangle r_i, \alpha_i(t) = 3\}$ have been picked up and are en-route to their destination. Finally, served travelers $R_S(t) = \{R\ |t\rangle r_i, \alpha_i(t) = 4\}$ have been dropped off at their destination.

At any point in time $t \geqslant 0$, each AV $j \in V$ has a physical location $posV_j(t)$ and a state $\beta_j(t)$. AV $j \in V$ can be in one of three states $\beta_j(t) \in \{1,2,3\}$ differentiating between AVs that are idle, en-route to pick up a traveler, and en-route to drop off a traveler, respectively. These three states correspond to three mutually exclusive and collectively exhaustive subsets of AVs; idle AVs $V_I(t) = \{V|\beta_j(t) = 1\}$, en-route pickup AVs $V_P(t) = \{V|\beta_j(t) = 2\}$, and en-route drop-off AVs $V_D(t) = \{V|\beta_j(t) = 3\}$.

Fig. 1 shows an overview of the temporal nature of the problem. At time τ, the fleet operator has full information about the location and state of all travelers with request times $r_i \leqslant \tau$ and no information about future traveler requests $r_i > \tau$. At every interval h the SAMS fleet operator assigns and/or reassigns AVs to traveler requests; i.e. when $mod(t,h) = 0$. Additionally, in this study re-assignments are only made if the number of idle AVs is greater than or equal to one ($|V_I(\tau)| \geqslant 1$), and the number of unassigned travelers is greater than or equal to one ($|R_U(\tau)| \geqslant 1$). The inter-assignment time h is an adjustable parameter that the fleet manager can change depending on various external factors (e.g. ratio of traveler request rate to SAMS service rate) and internal priorities (importance of reducing traveler wait time vs. decreasing operational costs).

The SAMS fleet operator aims to efficiently assign AVs to traveler requests as they arrive randomly and dynamically; i.e. to minimize operational costs and customer service quality. This problem is a sequential stochastic optimization problem (Powell et al., 2012) because new information (i.e. traveler requests) enters the system dynamically, and the location and time of the new requests are unknown a priori.

Whereas static deterministic optimization problems involve optimizing over a set of decisions (denoted by $\boldsymbol{x}$ in objective function

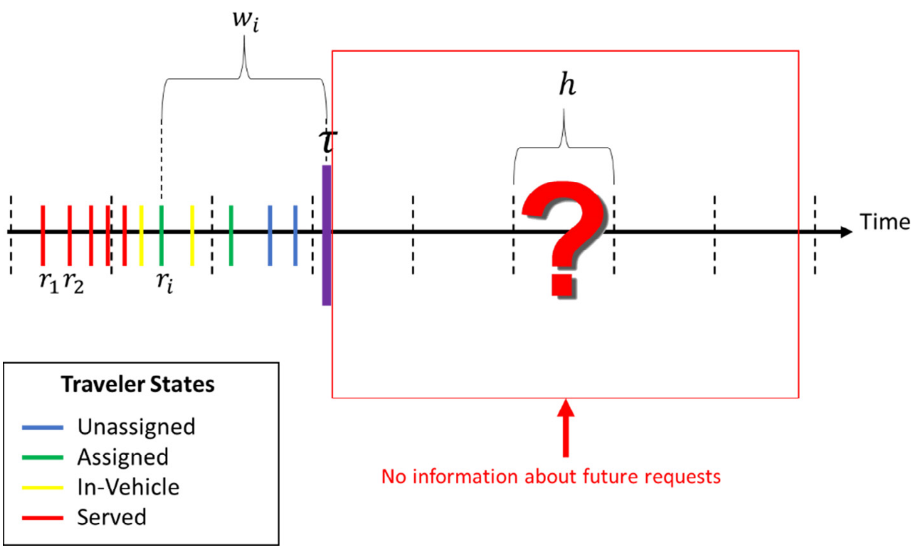

Fig. 1. Depiction of temporal aspects of on-demand SAMS with no shared rides operational problem.

$\min_{\boldsymbol{x}} c^t\boldsymbol{x}$), and sequential deterministic optimization problems involve optimizing over a set of time-dependent decisions (denoted by $\boldsymbol{x_t}$ in objective function $\min_{x_1,\ldots,x_T} \sum_{t=0}^{T} c_t\boldsymbol{x_t}$), for sequential stochastic problems, the goal is to find an optimal policy π from the set of all policies Π that optimizes the following problem: $\min_{\pi\in\Pi} E^{\pi} \sum_{t=0}^{T} C(S_t,X^{\pi}(S_t))$, wherein $C(\cdot)$ is a cost function; $S_{t+1} = S^M(S_t,x_t,W_{t+1})$; $S^M(\cdot)$ denotes the transition function; $X^{\pi}(S_t)$ is policy π 's decision function given state S_t; and W_{t+1} denotes the exogenous information that enters the systems between t and $t+1$. The objective function includes an expectation $E^{\pi}(\cdot)$ because W_t and therefore $S^M(S_t,x_t,W_{t+1})$ are random variables. In this study, the fleet operator needs to decide which AVs to assign to which open traveler requests. See Powell et al. (2012) for a detailed presentation of stochastic sequential optimization models in transportation.

The state of the system at time t (S_t) is defined by a tuple containing the location of open traveler requests $\{posR_i(t)|\ \forall\ i \in \{R_U,R_A\}\}$, the state of travelers $\{\alpha_i(t)|\ \forall\ i \in R\}$, the location of AVs $\{posV_j(t)|\ \forall\ j \in V\}$, the state of AVs $\{\beta_j(t)|\ \forall\ j \in V\}$, the elapsed wait time of open traveler requests $\{w_i(t)|\ \forall\ i \in \{R_U,R_A\}\}$, binary variables indicating if open traveler requests have previously been reassigned $\{b_i(t)|\ \forall\ i \in \{R_U,R_A\}\}$, and binary indicator variables indicating if AV $j \in V$ is en-route to pick up traveler $i \in R$ $\{y_{ij}(t)|\ \forall\ i \in \{R_U,R_A\}\ ,\forall\ j \in V\}$. Exogenous information (W_t) includes the arrival of new traveler requests with request time $r_i = t$. The decision variable $x_t = x_{ij}(t)$ denotes the assignment of AV $i \in R$ to open traveler request $j \in V$, at time t. A policy $\pi \in \Pi$, defines the assignment decision given the state of the system, $X^{\pi}(S_t) \to x_{ij}(t)$.

For most real-world sequential stochastic optimization problems with a large state space, such as the on-demand SAMS with no shared rides problem, there is unlikely to be an optimal policy due to the curses of dimensionality associated with sequential stochastic optimization problems (Powell, 2011). Hence, this study presents an extensive computational comparison of six different policies (i.e. AV-traveler assignment strategies).

This study focuses on the assignment of AVs to travelers and assumes the fleet operator has no information about the spatial-temporal demand distribution (λ) and therefore no information about the exogenous information (W_t). This assumption precludes the fleet operator from using probabilistic information about future demand requests when making decisions at the current time (τ). Hence, the fleet operator only considers the information that is known exactly at the current time (τ).

The Bellman Equation (see Eq. (1), where $V_t(s)$ denotes the value of being in state s at time t) represents a solution approach to sequential (deterministic or stochastic) optimization problems (Powell et al., 2012). Given the fleet operator has no information about the spatial-temporal demand distribution (λ) and exogenous information (W_t), $P(S_{t+1} = s'|S_t,x_t)$ is unknowable. Hence, this study compares six policies that differ in terms of how they solve the local optimization problem $\min_{x_t}\{C(S_t,x_t)\}$ at each time step t. These policies are referred to as AV-traveler assignment strategies because the decision made at each time step involves assigning AVs to open traveler requests. The next subsection provides the base formulation of the local optimization problem $\min_{x_t}\{C(S_t,x_t)\}$; it is referred to as the AV-traveler assignment problem.

$$\min_{x_t} \left\{ C(S_t,x_t) + \sum_{s' \in S} P(S_{t+1} = s'|S_t,x_t)V_{t+1}(s') \right\} \tag{1}$$

As mentioned previously, to handle the dynamic (i.e. sequential) aspect of the problem, the AV fleet operator repeatedly re-solves a local optimization problem $\min_{x_t}\{C(S_t,x_t)\}$ as new information enters the system. Two of the AV-traveler assignment strategies presented in Section 4 implicitly handle the problem's inherent stochasticity via allowing the fleet operator to alter decisions made at the current time (τ) in later time periods ($t > \tau$) as new information enters the system.

3.2. AV-Traveler assignment problem formulation

This subsection presents a mathematical formulation of the static local optimization problem $\min_{x_t}\{C(S_t,x_t)\}$, referred to as the AV-traveler assignment problem. Let $d_{ij}(\tau)$ denote the distance between the pickup location of traveler $i \in R'$ (o_i) and the current location of AV $j \in V'$ ($posV_j(\tau)$) at the current time τ. In this section, R' and V' denote the sets of travelers and AVs considered in the assignment problem, respectively. The sets of travelers and AVs considered in the assignment problem vary as a function of the assignment strategy. The elapsed wait time of traveler $i \in \{R_U \cup R_A\}$ at time τ is $w_i(\tau)$. Let x_{ij} equal 1 if AV j is assigned to traveler i, and 0 otherwise. The dependence of the variables on t is removed from the notation for simplicity and the fact that the local AV-traveler assignment problem is static.

The AV-traveler assignment formulation depends on whether, in the dynamic system at time τ, the number of travelers included in the assignment ($|R'|$) is less than or greater than the number of AVs in the assignment ($|V'|$). The mathematical programming formulation for the AV-traveler assignment problem when the number of travelers is greater than the number of AVs ($|R'| > |V'|$) is given in Eqs. (2)–(5).

$$\min_{x_{ij}} \sum_{i\in R'} \sum_{j\in V'} (d_{ij}x_{ij} - \gamma w_i x_{ij}) \tag{2}$$

s.t.

$$\sum_{j\in V'} x_{ij} \leqslant 1 \quad \forall\ i \in R' \tag{3}$$

$$\sum_{i\in R'} x_{ij} = 1 \quad \forall\ j \in V' \tag{4}$$

$$x_{ij} \geqslant 0 \quad \forall\, i \in R',\, \forall\, j \in V' \tag{5}$$

Eq. (3) ensures that each traveler is assigned to at most one AV. Eq. (4) ensures each AV is assigned to a traveler. Eq. (4) requires the decision variable, x_{ij}, to be non-negative. However, because the constraint matrix is totally unimodular, x_{ij} will only take on integer values. The objective function in Eq. (2) has two terms. The first term represents the total distance between travelers and the AVs that they are assigned to. The second term represents the elapsed wait time of the assigned travelers. The parameter γ weights the relative importance of assigning AVs to travelers that have been waiting a long time; γ also converts the time units associated with w_i to the distance units associated with d_{ij}. Given that all travelers are not assigned to an AV because $|R'| > |V'|$, the second term incentivizes the SAMS fleet operator to assign AVs to travelers with large elapsed wait times (w_i). Without this second term, the fleet operator will not assign AVs to travelers in the periphery of the service region, if they are all busy serving traveler requests in the core of the service region. The challenge of serving demand requests in the periphery is quite common and was recognized in one of the seminal DRP papers (Psaraftis, 1980). The objective function in Eq. (2) handles this challenge elegantly and effectively.

The mathematical programming formulation for the AV-traveler assignment problem changes slightly when the number of travelers is less than the number of AVs ($|R'| \leqslant |V'|$). This problem is formulated in Eqs. (6)–(9).

$$\min_{x_{ij}} \sum_{i \in R'} \sum_{j \in V'} d_{ij} x_{ij} \tag{6}$$

s.t.

$$\sum_{j \in V'} x_{ij} = 1 \quad \forall\, i \in R' \tag{7}$$

$$\sum_{i \in R'} x_{ij} \leqslant 1 \quad \forall\, j \in V' \tag{8}$$

$$x_{ij} \geqslant 0 \quad \forall\, i \in R',\forall\, j \in V' \tag{9}$$

Because the number of travelers is less than the number of AVs, and the constraint in Eq. (7) requires each traveler to be assigned to an AV, the second term in the objective function of Eq. (2) is no longer relevant. Every traveler, including those in the periphery will be assigned to an AV due to the constraint in Eq. (7). The objective in Eq. (6) is to minimize the overall distance between travelers and the AVs they are assigned. Eq. (8) ensures that each AV is assigned to at most one traveler. Finally, Eq. (9) requires the decision variable, x_{ij}, to be non-negative.

The solution algorithm determines whether to solve the AV-traveler assignment problem in Eqs. (2)–(5) or Eqs. (6)–(9) depending on the number of number of travelers ($|R'|$) and the number of AVs ($|V'|$). The formulations in Eqs. (2)–(5) and Eqs. (6)–(9) provide a baseline model for assigning AVs to traveler requests. Section 4 presents six unique assignment strategies to solve the stochastic sequential problem defined in Section 3.1. Four of the strategies employ the AV-traveler assignment problem; the formulation varies slightly, but significantly across the four strategies. For example, the simpler strategies only include unassigned travelers $i \in R' = \{R_U\}$ and idle AVs $j \in V' = \{V_I\}$. The advanced strategies include unassigned and assigned travelers $i \in R' = \{R_U \cup R_A\}$ and all AVs $j \in V' = \{V_I \cup V_P \cup V_D\}$.

4. AV-Traveler assignment strategies

This section presents six different AV-traveler assignment strategies (i.e. control policies). To solve the on-demand SAMS with no shared rides operational problem, the SAMS fleet operator repeatedly solves the local optimization problem (i.e. the AV-traveler assignment problem) based on the state of travelers and AVs at the current time τ.

The first two strategies are simplistic first-come, first-served (FCFS) strategies; whereas, strategies three through six employ the mathematical programming formulation presented in Section 3.2. The difference between the four strategies comes down to the AVs (V') and travelers (R') considered in the problem formulation.

Figs. 2–7 display a toy example. The left side of each figure displays the assignment of AVs to travelers at time τ, and the right side displays the updated assignment of AVs to travelers at time $\tau + h$. The solid lines indicate assignments made before time τ, and the dashed lines represent assignments or reassignments made between time τ and time $\tau + h$. The triangles represent traveler drop-off locations and the squares represent pickup locations, wherein the dashed-line squares represent new traveler requests that enter the system between time τ and time $\tau + h$.

4.1. Strategy 1

The first strategy assigns travelers FCFS to the longest idle AV (AV *1* was idling the longest when traveler *1* made a request, AV 2 had the second longest idle time, AV 3 the third longest idle time, etc.). Fig. 2 illustrates the inefficiency associated with this heuristic strategy. On the left side of Fig. 2, despite being the farthest AV from traveler *2*, AV *2* was assigned to traveler *2* only because AV *2* was idle longer than the other AVs in the fleet. Between τ and $\tau + h$, traveler *3* and traveler *4* enter the system sequentially and they are each assigned sequentially to the longest remaining idle AV.

The resultant assignment in Fig. 2 confirms the inefficient logic associated with the first assignment strategy. In this toy problem, given that the bottom edge of Figs. 2–7 is 3.5 miles, the total fleet mileage required to drop off traveler *1* and pick up travelers *2–4*

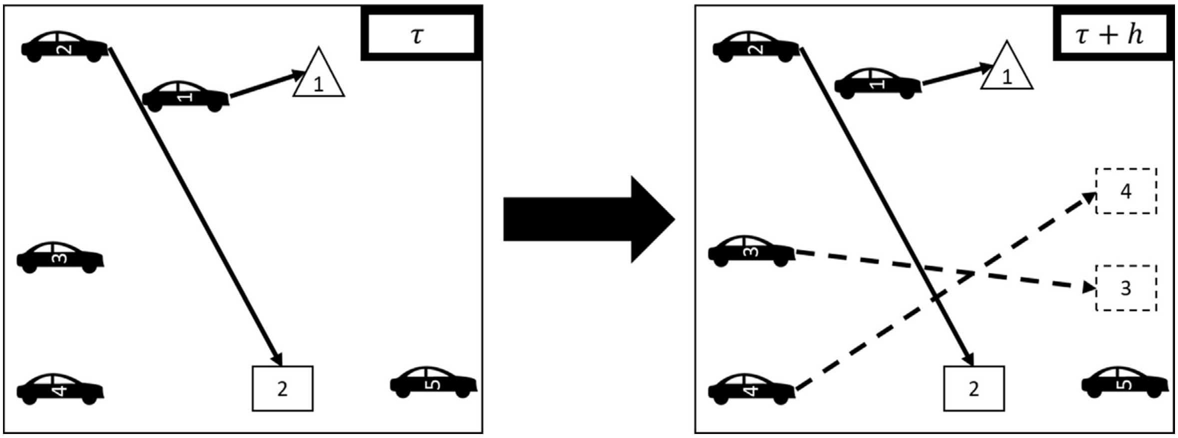

Fig. 2. AV-traveler assignment based on Strategy 1 (total fleet miles = 12.9).

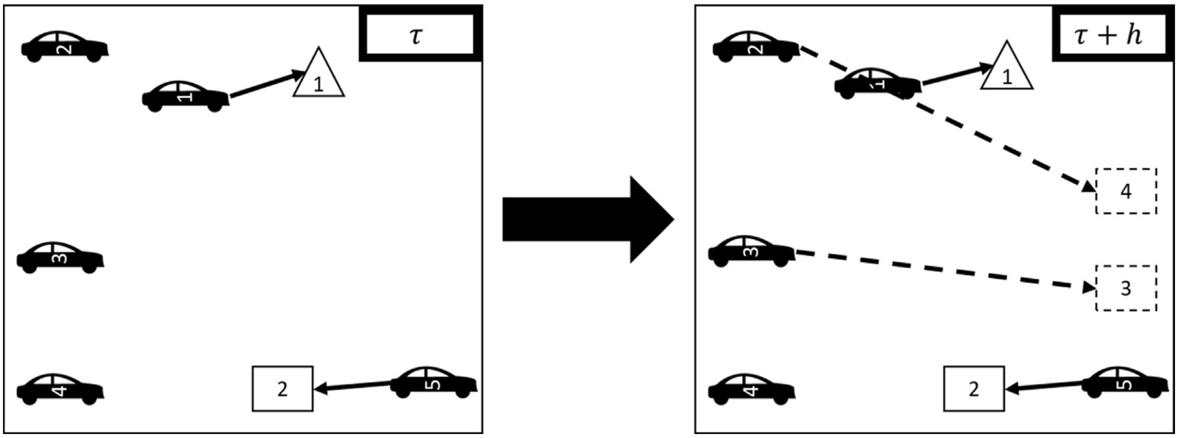

Fig. 3. AV-traveler assignment based on Strategy 2 (total fleet miles = 9.7).

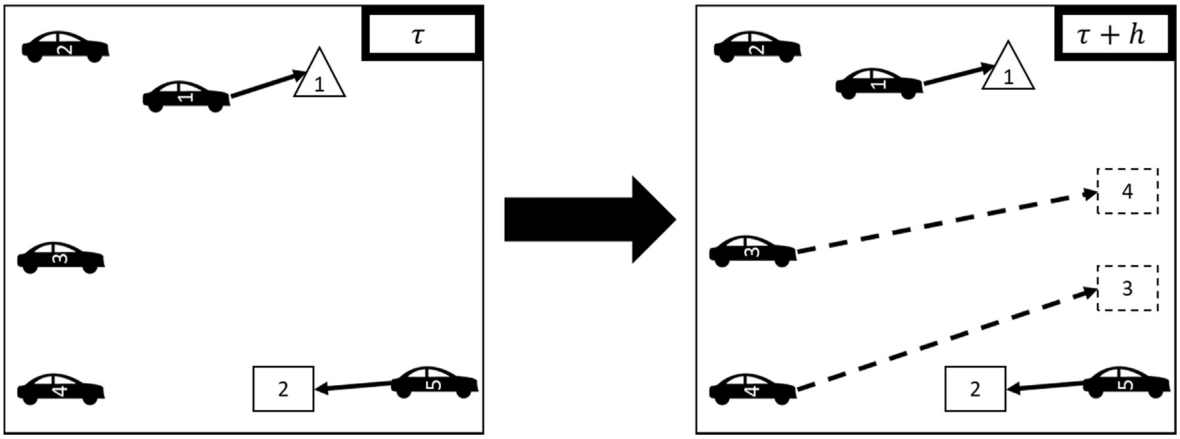

Fig. 4. AV-traveler assignment based on Strategy 3 (total fleet miles = 9.0).

under Strategy 1 is 12.9 miles.

4.2. Strategy 2

The second strategy assigns travelers FCFS to the nearest idle AV. Assigning travelers FCFS is still an inefficient strategy; however, assigning them to the nearest idle AV should improve the fleet's operational efficiency, relative to Strategy 1. For example, on the left

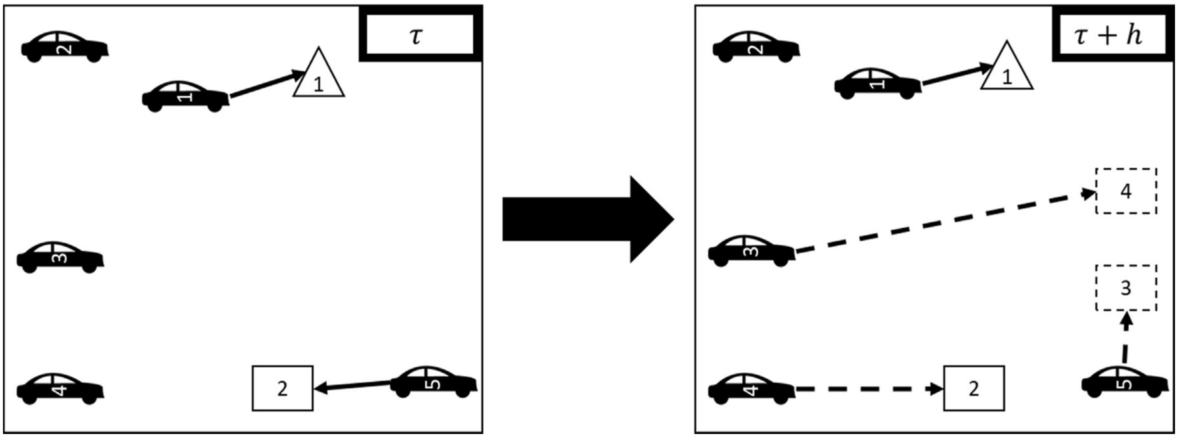

Fig. 5. AV-traveler assignment based on Strategy 4 (total fleet miles = 6.8).

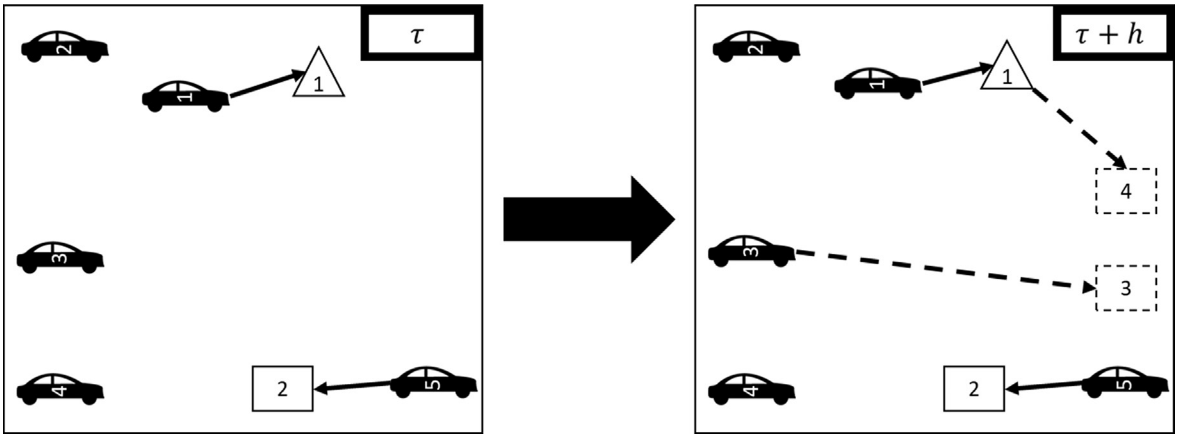

Fig. 6. AV-traveler assignment based on Strategy 5 (total fleet miles = 7.4).

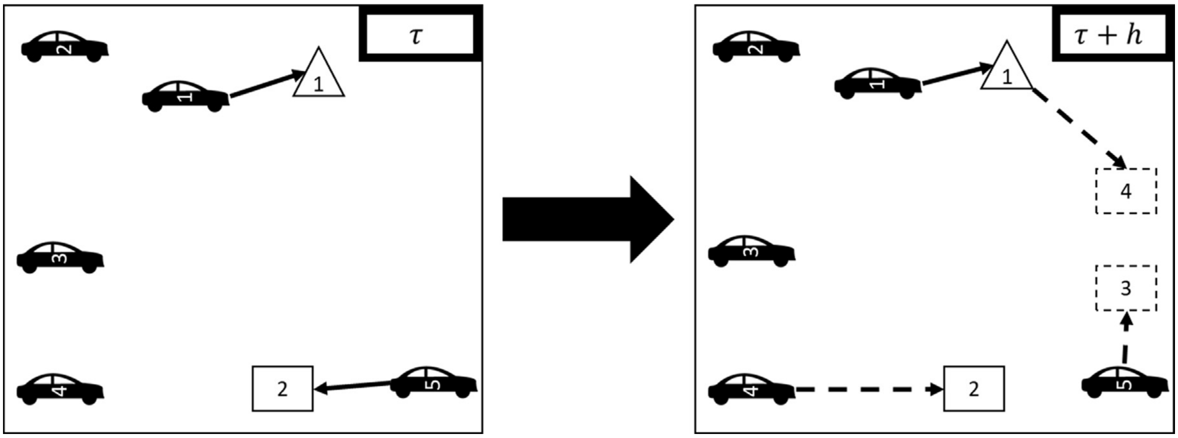

Fig. 7. AV-traveler assignment based on Strategy 6 (total fleet miles = 5.4).

side of Fig. 3, AV 5 is assigned to traveler 2, rather than the inefficient assignment of AV 2 to traveler 2 in Strategy 1.

On the right side of Fig. 3, traveler 3 is assigned to AV 3, the nearest idle AV. Then, after traveler 3 is assigned, traveler 4 is assigned to AV 2, the nearest remaining idle AV. In Strategy 2, the travelers are assigned sequentially, rather than simultaneously between τ and $\tau + h$. For the toy problem, the total fleet mileage associated with Strategy 2 is 9.7 miles, compared with 12.9 miles for Strategy 1. This is a significant improvement; however, assigning new sets of travelers simultaneously rather than FCFS or sequentially should further improve the fleet's efficiency.

4.3. Strategy 3

In the third strategy, only unassigned travelers ($i \in R' = \{R_U\}$) and idle AVs ($j \in V' = \{V_I\}$) are considered in the AV-traveler assignment problem. In this strategy, once an AV-traveler assignment is made, it is not altered.

Fig. 4 displays the results of the toy problem if Strategy 3 is employed. The assignment of AVs to travelers between τ and $\tau + h$ is more efficient in Fig. 4 than Fig. 3 because traveler *3* and traveler *4* are assigned simultaneously, rather than sequentially. Simultaneous assignment allows the optimization solver to find the best assignment across both traveler 3 and traveler 4. The total fleet mileage associated with Strategy 3 in this toy problem is 9.0 miles.

4.4. Strategy 4

In this strategy, unassigned and assigned travelers ($i \in R' = \{R_U \cup R_A\}$) as well as idle and en-route pickup AVs ($j \in V' = \{V_I \cup V_P\}$) are considered in the AV-traveler assignment problem. The inclusion of assigned travelers and en-route pickup AVs allows the reassignment of AVs to previously assigned travelers. In the dynamic fleet management literature, this reassignment of travelers is also known as vehicle diversion (Regan et al., 1995, 1996a, 1996b; Ichoua et al., 2006; Sheridan et al., 2013).

In the toy problem in Fig. 5, AV *5* is diverted from traveler *2* to traveler *3* between τ and $\tau + h$. Additionally, traveler 2 is reassigned to AV *4* and AV *3* is assigned to traveler *4*. The total fleet mileage associated with Strategy 4 is 6.8 miles. Hence, for this problem instance, allowing traveler reassignment (or vehicle diversion) significantly reduces fleet miles.

Allowing reassignment is often a very beneficial strategy as illustrated in Fig. 5. However, without adding constraints to the math program in Section 3.2, the strategy can result in unwanted outcomes, such as a previously assigned traveler no longer being assigned, and a traveler constantly being reassigned to different AVs.

To prevent previously assigned travelers from being completely unassigned, a constraint is added to the mathematical programming formulation of the AV-traveler assignment problem in Section 3.2. Let a_i equal 1 if traveler $i \in R_A$, and zero otherwise. The following constraint prevents a traveler from going from assigned to unassigned:

$$\sum_{j \in V'} x_{ij} = 1 \quad \forall\, i \in R_A \quad \text{or} \quad \sum_{j \in V'} x_{ij} - a_i \geqslant 0 \quad \forall\, i \in R' \tag{10}$$

To prevent travelers from constantly being reassigned, a constraint is added that only allows one reassignment per traveler. Let y_{ij} equal 1, if AV $j \in V_P$ is en-route to pick up traveler $i \in R_A$. Let b_i equal 1 if traveler $i \in R_A$ has previously been reassigned. The following constraint prevents a traveler from being reassigned more than once:

$$b_i(y_{ij} - x_{ij}) \leqslant 0 \quad \forall\, i \in R_A\,, \quad \forall\, j \in V_P \tag{11}$$

In addition to these constraints, a penalty is added to the objective function for reassigning AVs. Let δ denote the penalty for assigning a traveler to an en-route pickup AV $j \in V_P$. Let q_j equal 1, if AV j is en-route to pick up a traveler ($j \in V_P$). The objective of the AV-traveler assignment problem then becomes:

$$\min_{x_{ij}} \left(\sum_{i \in R'} \sum_{j \in V'} (d_{ij}x_{ij} - \gamma w_i x_{ij} + \delta(1 - y_{ij}) q_j x_{ij}) \right) \tag{12}$$

4.5. Strategy 5

In the fifth strategy, unassigned travelers ($i \in R' = \{R_U\}$) as well as idle and en-route drop-off AVs ($j \in V' = \{V_I \cup V_D\}$) are considered in the assignment problem; similar to the strategy employed in Maciejewski et al. (2016) for dispatching taxis. This strategy does not allow en-route pickup AVs to be diverted ($V_P \notin V'$), nor does it allow travelers to be reassigned ($R_A \notin R'$). However, this strategy essentially allows two-person schedules for AVs. That is, if a new traveler request $i' \in R_U$ enters the system, and an en-route drop-off AV $j' \in V_D$ can pick up traveler $i' \in R_U$ faster than all the other AVs, even after considering the remaining time/distance to drop off the traveler it is carrying $i'' \in R_{IV}$, then the en-route drop-off AV $j' \in V_D$ is assigned to the new traveler request $i' \in R_U$.

The right side of Fig. 6 is unique in that AV *1* is assigned to traveler *4* even though it was en-route to drop off traveler *1* at the time of assignment. For this problem instance, the total fleet mileage associated with Strategy 5 is 7.4 miles, an improvement over Strategy 3 (9.0 miles) but not quite as good as Strategy 4 (6.8 miles). It should be clear that combining Strategy 4 and Strategy 5, can further improve the fleet efficiency.

Determining d_{ij} for en-route drop-off AVs requires calculating two sets of distances. The distance between the current position of AV $j' \in V_D$ and the drop-off location of traveler $i'' \in R_{IV}$ in AV $j' \in V_D$: $dist(pos_{j'}, d_{i''})$ and the distance between the drop-off location of traveler $i'' \in R_{IV}$ and the pickup location of traveler $i' \in R_A$: $dist(d_{i''}, o_{i'})$. Therefore, $d_{ij} = dist(pos_{j'}, d_{i''}) + dist(d_{i''}, o_{i'})$.

As it takes time for a passenger to get out of an en-route drop-off AV ($j \in V_D$), and the fleet operator probably only wants to assign a traveler to an en-route drop-off AV ($j \in V_D$) if it is a much better option than an idle AV ($j \in V_I$), a penalty is added in the objective function for assigning an en-route drop-off AV to a traveler. Let p_j equal 1 if AVj is en-route to drop off a traveler ($j \in V_D$). Let φ denote the penalty of assigning a traveler to an en-route drop-off AV. The objective function in Eq. (2), changes to:

Table 1
Overview of AV-traveler assignment strategies.

Strategy	FCFS/Optimization	Travelers (R')	AVs (V')	Sequential/Simultaneous	Traveler Reassignment?	En-Route Drop-off AV
1	FCFS	R_U	V_I	Sequential	No	No
2	FCFS	R_U	V_I	Sequential	No	No
3	Optimization	R_U	V_I	Simultaneous	No	No
4	Optimization	$R_U \cup R_A$	$V_I \cup V_P$	Simultaneous	Yes	No
5	Optimization	R_U	$V_I \cup V_D$	Simultaneous	No	Yes
6	Optimization	$R_U \cup R_A$	$V_I \cup V_P \cup V_D$	Simultaneous	Yes	Yes

$$\min_{x_{ij}} \left(\sum_{i \in R'} \sum_{j \in V'} (d_{ij}x_{ij} - \gamma w_i x_{ij} + \varphi p_j x_{ij}) \right) \tag{13}$$

4.6. Strategy 6

In the sixth and final strategy, unassigned and assigned travelers ($i \in R' = \{R_U \cup R_A\}$) as well as all AVs (idle, en-route pickup, and en-route drop-off, $j \in V' = \{V_I \cup V_P \cup V_D\}$) are considered in the assignment problem. Strategy 6 combines the valuable additions in Strategy 4 (traveler reassignment and AV diversions) and Strategy 5 (inclusion of en-route drop-off AVs in AV-traveler assignment problem) to the base optimization-based assignment strategy, Strategy 3. The AV-traveler mathematical programming formulation associated with Strategy 6 includes the constraints in Eqs. (10) and (11) as well as the additional terms in Eqs. (12) and (13).

The right side of Fig. 7 shows the results of allowing AV diversions (AV *5* is diverted from traveler *2* to traveler *3*), traveler reassignment (traveler *2* is reassigned from AV *5* to AV *4*), and two-person schedules (AV *1* is assigned to pick up traveler 4, as it is en-route to drop off traveler *1*). For this problem instance, the total fleet mileage associated with Strategy 6 is 5.4 miles.

Table 1 distinguishes between the six AV-traveler assignment strategies. Section 6 compares these six AV-traveler assignment strategies on much larger problem instances. As the problem is dynamic and stochastic, it is not possible to guarantee any other strategy will perform the best. Hence, a variety of scenarios are presented in the computational results section, to empirically compare the six strategies.

5. Agent-Based simulation framework

This section briefly describes the agent-based simulation framework developed to model the dynamic system of AVs, travelers, and intelligent SAMS fleet operator. The simulation is coded in Python 3.5.1 (Python, 2017). The Gurobi 7.0.2 optimization solver (Gurobi, 2017) is embedded in the Python simulation model and is used to solve the AV-traveler assignment problem for Strategy 3 through Strategy 6. Gurobi is a state-of-the-art mathematical programming solver; however, the assignment problem is a relatively easy problem to solve.

5.1. Inputs

The analyst defines the area size, length of the analysis period (T), and spatial-temporal demand rate of traveler requests (λ). With this information, a random number generator is employed to generate traveler origin locations (o_i), destination locations (d_i), and request times (r_i) for each traveler $i \in R$.

In this paper, AVs travel on a Manhattan network (i.e. a grid-based network with omnipresent streets), which is an abstraction of an urban road network. The AVs travel at a constant speed (v) defined by the analyst. The locations of the AVs, and travelers in the AVs, are updated every time step (Δt). Smaller Δt values better represent continuous time; however, they increase the computer simulation time. The analyst must also set traveler drop-off times (c_d) and pickup times (c_p).

In a single simulation run, to model the SAMS fleet operator, it is necessary to choose an inter-assignment time (h), one of the AV-traveler assignment strategies presented in Section 4, the SAMS fleet size, and several parameters in the AV-traveler assignment problem. These parameters include the weight of elapsed wait time relative to travel distance (γ), the penalty for assigning a traveler to an en-route drop-off AV (φ), and the penalty for assigning a traveler to an en-route pickup AV (δ). Additionally, the analyst needs to provide the start location of each AV; this study assumes the AVs start in the middle of the service region.

5.2. Simulator

The agent-based simulation tool models the dynamic behavior of and interactions between travelers, AVs, and the SAMS fleet operator. The SAMS fleet operator is the only 'intelligent' agent in the simulation; it makes decisions to assign AVs to travelers.

Fig. 8 displays an overview of how the simulation model updates the positions and states of travelers. The simulation is time-driven; at each time step (Δt), the simulation updates the position and state of travelers. The simulation first updates the position of in-vehicle travelers (R_{IV}) by moving each traveler $i' \in R_{IV}$ one step ($\Delta t \times v$) towards her destination $d_{i'}$. After moving traveler i', the

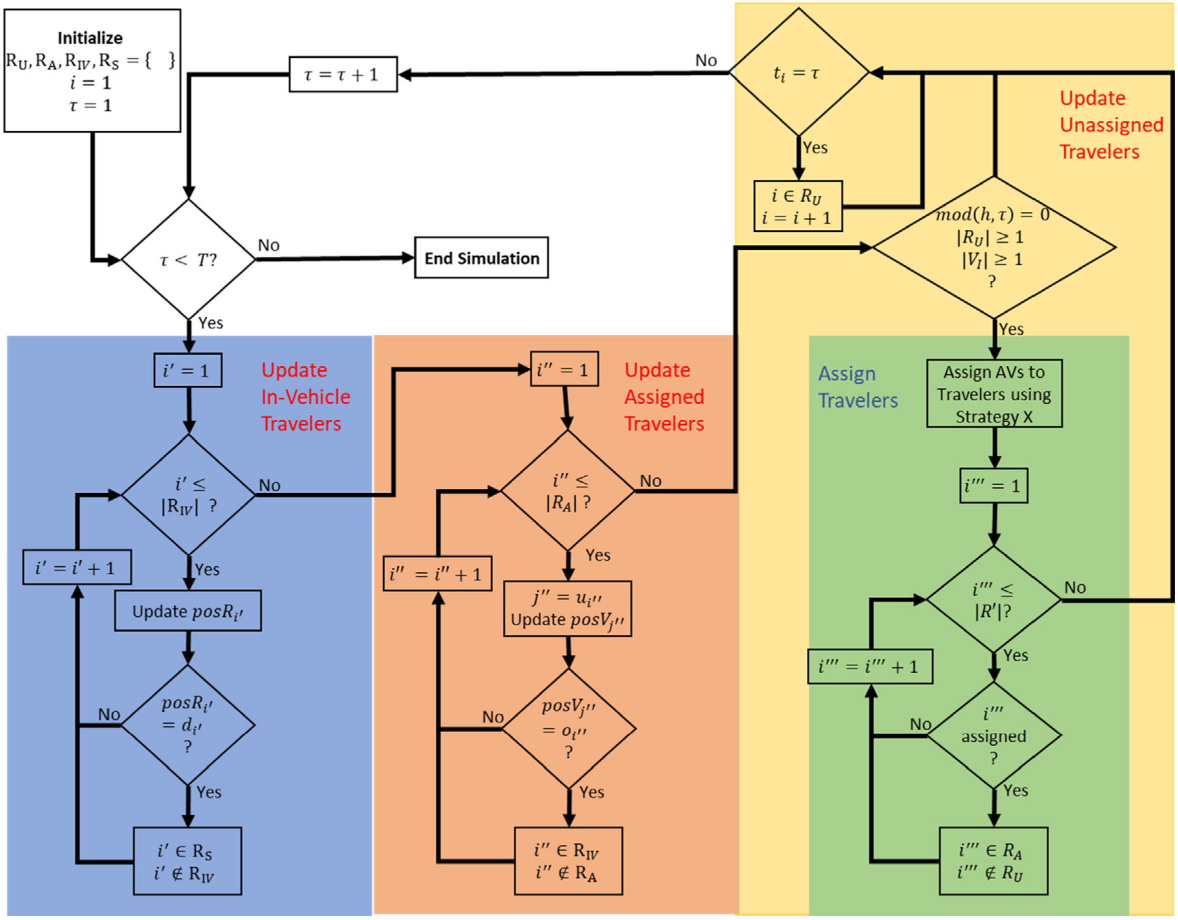

Fig. 8. Simulation flowchart of travelers states and positions.

simulation checks to see if she has reached her destination ($posR_{i'} = d_{i'}$). If she has reached her destination, she enters the set of served travelers $i' \in R_S$ and she is removed from the set of in-vehicle travelers $i' \notin R_{IV}$. Although not shown in Fig. 8, the simulation accounts for the drop-off time c_d.

The simulation goes through a similar process for all assigned travelers R_A. Let $u_{i''}$ denote the AV $j'' \in V_P$ assigned to traveler $i'' \in R_A$. The simulation moves AV $j'' \in V_P$ one step ($\Delta t \times v$) toward $o_{i''}$ the pickup point of traveler $i'' \in R_A$. If AV $j'' \in V_P$ has reached $o_{i''}$ ($posV_{j''} = o_{i''}$), traveler $i'' \in R_A$ enters the set of in-vehicle travelers $i'' \in R_{IV}$ and is removed from the set of assigned travelers $i'' \notin R_A$. Although not shown in Fig. 8, the simulation accounts for the pick-up time c_p.

Next the simulation checks to see if the SAMS fleet operator will assign or reassign AVs to travelers. If $mod(h,\tau) = 0$, and $|R_U| \geqslant 1$, and $|V_I| \geqslant 1$, then the simulation calls the mathematical programming solver (i.e. assignment module) to assign AVs to travelers. The solver employs one of the assignment strategies in Section 4 selected by the analyst. After the assignment module, the simulation loops through all the travelers considered in the assignment (R') and if the assignment module assigned traveler i''' to an AV, she enters or re-enters the set of assigned travelers $i''' \in R_A$.

Then the simulation checks to see if the request time of the next traveler request i is equal to the current time ($r_i = \tau$). If it is, then traveler i enters the set of unassigned traveler requests ($i \in R_U$) and the simulation looks at the next traveler ($i = i + 1$). Finally, when the request time of the next traveler request is greater than the current time $r_i > \tau$, the simulation advances to the next time instance $\tau = \tau + 1$. This process repeats until $\tau = T$.

Although not shown in Fig. 8, the simulation also updates the position and state of AVs at every time step.

5.3. Outputs

One major benefit of an agent-based microsimulation is that the model can output the complete history of each traveler and AV throughout the entire simulation period T at a pre-define level of temporal disaggregation (Δt). With this information, the analyst can obtain the individual, mean, median, standard deviation, etc. for metrics such as traveler wait time, traveler in-vehicle travel time, AV (empty and/or loaded) miles, and AV utilization rate. Another benefit of simulation methods is that simulation replications are independent and identically distributed. Therefore, the central limit theorem and law of large numbers hold.

Table 2
Parameter values for artificial demand scenarios.

Parameter	Symbol	Value	Units
Simulation Length	T	4	hours
Simulation Time Step	Δt	1	seconds
Traveler Demand Rate	λ	1000	travelers/hour
Area Size		(a) 16 (b) 64 (c) 256	miles^2
Spatial Demand Pattern		(1) Uniform (2) Clustered	NA
Trip Distance (mean)		(a-1) 2.8 (a2) 3.3 (b-1) 5.4 (b2) 6.7 (c-1) 10.7 (c2) 13.3	miles
Trip Distance (sd.)		(a-1) 1.2 (a-2) 1.1 (b-1) 2.6 (b-2) 2.1 (c-1) 5.2 (c-2) 4.3	miles
Inter-assignment Interval	h	10	seconds
Vehicle Speed	v	35	mph
Drop-off Time	c_d	15	seconds
Pickup Time	c_p	45	seconds
Weight of Elapsed Wait Time	γ	50	feet/s
En-route drop-off Assgn. Penalty	φ	750 (15)	feet (sec.)
En-route pickup Assgn. Penalty	δ	1500 (30)	feet (sec.)

6. Experiments and computational results

This section presents experiments to compare the efficiency of the AV-traveler assignment strategies across two metrics. The first metric is average traveler wait time and the second metric is the ratio of empty fleet miles to total fleet miles, wherein total fleet miles is the sum of empty and loaded fleet miles. The average traveler wait time metric aims to represent customer service quality. Minimizing wait times should increase the competitiveness of SAMSs with the personal (autonomous) vehicle.

The empty fleet miles metric aims to represent SAMS fleet operational costs. The number of loaded fleet miles is fixed given the transportation system is modeled on a Manhattan grid network with omni-present streets, and the SAMS does not allow shared rides. Minimizing empty fleet miles should decrease fuel costs, increase the life of the vehicle, and reduce maintenance costs. To compete with the personal vehicle, the SAMS provider can pass these cost savings on to customers.

This section presents two sets of computational experiments. In the first set of experiments, the AV-traveler assignment strategies are compared across hundreds of artificial demand scenarios. In the second set of experiments, the assignment strategies are tested on the Chicago taxi data, which represent a realistic spatio-temporal demand pattern.

The experiments aim to compare the six AV-traveler assignment strategies (across the two metrics) when the fleet size is small relative to the traveler demand rate. This is likely to happen during the morning and evening peak periods. With a fixed fleet size, efficiently assigning AVs to travelers can increase the number of traveler served, decrease average traveler wait times, and/or reduce operational costs.

6.1. Artificial demand

6.1.1. Experimental design

Table 2 displays the input parameter values for the simulation experiments with artificial demand. The four-hour period represents the morning or evening peak. The fixed fleet size would be most stressed during these two periods.

Traveler requests are generated based on the area size, traveler demand rate, and spatial demand pattern input parameters. Varying area size and the spatial demand pattern significantly impact trip distance, as shown in Table 2. The region sizes are varied to examine the impact of region size, and trip distance on the performance of the AV-traveler assignment strategies.

In addition to varying the area size and spatial demand pattern, this section presents the performance of the AV-traveler assignment strategies as a function of the fleet size. Given that SAMSs do not yet exist, and this paper aims to analyze the case where the fleet size is small relative to the demand rate, the analysis refrains from estimating and selecting a fleet size.

The combination of six AV-traveler assignment strategies, seven AV fleet sizes, three area sizes, and two spatial patterns requires 252 unique simulation experiments. Moreover, to produce statistically significant results, each of the 252 simulation experiments were replicated twenty times. A random number generator varies traveler origins, destinations, and request times across each set of twenty replications.

The simulations were run on a standard 64-bit desktop computer with 8 GB of RAM, and a 3.20 GHz processor. A single simulation experiment replication takes anywhere from twenty seconds to ten minutes to complete. Experiments with larger fleet sizes take longer to run, as do experiments with Strategy 4 and Strategy 6, and to a lesser extent Strategy 5.

Table 3
Average travel time (min) results for uniform demand scenarios.

Fleet Size	Strategy 1		Strategy 2		Strategy 3		Strategy 4		Strategy 5		Strategy 6	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Area Size = 16 mi^2												
130	52.4	*0.34*	43.4	*0.34*	10.4	*0.29*	8.8	*0.29*	7.5	*0.25*	6.1	*0.23*
140	45.7	*0.34*	33.8	*0.31*	4.4	*0.21*	3.2	*0.16*	3.2	*0.12*	2.4	*0.09*
150	37.3	*0.30*	25.6	*0.34*	2.5	*0.03*	1.7	*0.04*	1.7	*0.04*	1.5	*0.03*
160	30.1	*0.33*	18.2	*0.35*	1.9	*0.05*	1.2	*0.02*	1.2	*0.02*	1.2	*0.02*
170	23.8	*0.28*	9.1	*1.00*	1.2	*0.03*	1.0	*0.01*	1.0	*0.01*	1.0	*0.01*
175	20.8	*0.27*	4.1	*0.92*	1.1	*0.02*	0.9	*0.01*	1.0	*0.01*	0.9	*0.01*
200	9.0	*0.29*	0.8	*0.01*	0.8	*0.01*	0.8	*0.01*	0.8	*0.01*	0.8	*0.01*
Area Size = 64 mi^2												
230	55.7	*0.29*	47.6	*0.33*	10.5	*0.29*	8.7	*0.30*	8.8	*0.31*	6.8	*0.24*
240	52.3	*0.33*	43.4	*0.34*	7.0	*0.26*	5.4	*0.24*	5.8	*0.19*	4.5	*0.16*
250	49.0	*0.32*	38.3	*0.38*	4.7	*0.14*	3.7	*0.10*	4.1	*0.14*	3.2	*0.09*
255	47.3	*0.31*	36.0	*0.34*	4.3	*0.07*	3.3	*0.07*	3.5	*0.12*	2.9	*0.07*
280	37.6	*0.31*	25.0	*0.33*	3.3	*0.11*	2.0	*0.04*	2.1	*0.04*	2.0	*0.04*
305	28.7	*0.31*	12.6	*1.27*	1.9	*0.05*	1.5	*0.02*	1.7	*0.03*	1.7	*0.02*
330	21.2	*0.30*	2.6	*0.59*	1.5	*0.04*	1.3	*0.02*	1.6	*0.02*	1.6	*0.03*
Area Size = 256 mi^2												
390	64.7	*0.44*	58.6	*0.48*	22.6	*0.46*	20.8	*0.47*	22.0	*0.45*	18.7	*0.43*
400	63.0	*0.43*	56.6	*0.45*	20.4	*0.46*	18.6	*0.48*	20.0	*0.43*	16.8	*0.43*
410	61.1	*0.46*	54.5	*0.49*	18.3	*0.46*	16.5	*0.47*	17.9	*0.46*	14.9	*0.44*
415	60.4	*0.43*	53.5	*0.48*	17.3	*0.45*	15.4	*0.46*	16.9	*0.50*	14.0	*0.44*
440	56.2	*0.45*	48.8	*0.47*	12.7	*0.44*	10.9	*0.44*	12.1	*0.44*	10.1	*0.34*
465	52.1	*0.43*	44.0	*0.51*	9.0	*0.37*	7.4	*0.34*	8.9	*0.32*	7.5	*0.25*
490	48.4	*0.44*	38.9	*0.47*	7.3	*0.18*	5.8	*0.18*	6.7	*0.24*	6.0	*0.15*

6.1.2. Uniform demand results

Table 3 displays the mean and standard error (across twenty replications) of average traveler wait time for all the uniform synthetic demand scenarios. The standard error is quite low for each scenario because the variance is small across replications and the number of replications is reasonably high.

According to Table 3, Strategy 1 is always extremely inefficient. Additionally, Strategy 2, is significantly less efficient than even Strategy 3 when the fleet size is small relative to the demand rate. As fleet size increases, Strategy 2 performs much better than Strategy 1 and similar to the optimization-based strategies. When the fleet size is small relative to the demand rate, Strategy 6 outperforms all the other strategies. However, as the fleet size increases, Strategy 4 outperforms Strategy 6 in terms of average traveler wait time.

The finding that Strategy 4 slightly outperforms Strategy 6, in terms of average traveler wait time, when the fleet size is high relative to the demand rate is an important one. The results in Table 4 indicate that Strategy 6 always outperforms Strategy 4 in terms of empty fleet miles, independent of fleet size and demand rate. This suggests that there is an important trade-off an AV fleet manager may need to consider when choosing between AV-traveler assignment strategies, during the off-peak period. However, when the fleet size is small relative to the demand rate, the fleet manager should always use Strategy 6.

It is important to highlight the reason why the performance gap in average traveler wait times across assignment strategies is much greater for small fleets than large fleets. Given the demand rate of traveler requests is constant across all scenarios, as the fleet size increases, the number of idle vehicles in the service region at any moment in time increases with fleet size. When the fleet size is large, and there are many idle AVs in the service region at time t, less-sophisticated strategies that simply assign idle AVs to travelers work just fine. Conversely, when the fleet size is small, and there are very few idle AVs in the service region at time t, allowing en-route pickup AVs to be re-assigned, and allowing en-route drop-off AVs to be considered in the assignment problem, in addition to idle AVs, is highly beneficial.

Table 4 displays the average (across twenty replications) of the ratio of empty fleet miles to total fleet miles for all the uniform synthetic demand scenarios. The standard error is not presented because it is very small relative to the average value, similar to Table 3. Table 4 shows that Strategy 1 is terribly inefficient, independent of fleet size. Strategy 2 is also very inefficient when fleet size is low relative to the demand rate, but as the fleet size increases, Strategy 2 approaches the efficiency of the optimization-based strategies. The presentation of assignment strategies in Section 4 hints at why FCFS strategies are inefficient when the fleet size is small relative to the demand rate. If the number of idle AVs is small, and a new traveler request enters the system, one of the idle AVs is assigned to the new traveler request, even if the pickup location is very far away.

In terms of empty fleet miles, Strategy 6 unambiguously outperforms all the other strategies across all area sizes and fleet sizes. The size of the performance gap between Strategy 6 and slightly less efficient strategies 4 and 5 increases when the fleet size is small relative to the demand rate. This dominance across experiments suggests that if operational costs are the most important metric for SAMSs, then the fleet manager should employ Strategy 6.

Table 4
Ratio of empty fleet miles to total fleet miles results for uniform demand scenarios.

Fleet Size	Strategy 1	Strategy 2	Strategy 3	Strategy 4	Strategy 5	Strategy 6
Area Size = 16 mi²						
130	49.0%	43.6%	19.8%	18.2%	16.0%	14.5%
140	49.1%	43.3%	20.1%	19.0%	18.2%	16.7%
150	49.0%	42.9%	24.3%	21.5%	18.4%	16.8%
160	49.1%	42.0%	25.4%	19.2%	17.0%	16.0%
170	49.0%	37.5%	20.1%	17.0%	15.8%	15.2%
175	48.6%	27.2%	18.5%	16.2%	15.4%	14.8%
200	48.5%	15.0%	14.8%	14.0%	13.7%	13.4%
Area Size = 64 mi²						
230	50.2%	45.0%	18.3%	16.7%	16.1%	13.8%
240	50.2%	44.8%	18.5%	17.1%	16.6%	14.8%
250	50.3%	44.5%	19.9%	19.2%	17.4%	15.1%
255	50.2%	44.5%	21.3%	20.2%	17.2%	15.1%
280	50.2%	44.0%	24.0%	18.0%	15.7%	14.4%
305	50.2%	37.6%	18.2%	15.7%	14.6%	13.7%
330	49.6%	20.0%	16.2%	14.6%	13.9%	13.2%
Area Size = 256 mi²						
390	51.0%	46.6%	17.9%	16.5%	16.1%	13.1%
400	51.0%	46.5%	17.8%	16.5%	16.3%	13.2%
410	51.0%	46.4%	17.9%	16.5%	16.4%	13.3%
415	51.0%	46.3%	17.9%	16.5%	16.4%	13.3%
440	51.0%	45.9%	18.1%	16.8%	16.2%	13.9%
465	50.9%	45.8%	18.9%	18.1%	16.9%	14.6%
490	51.1%	45.5%	22.0%	20.5%	16.8%	14.6%

6.1.3. Clustered demand results

This section briefly presents the clustered demand results, which largely follow the same pattern as the uniform demand results. To generate clustered demand, four cluster centroids were placed in the four quadrants of the square service region. Travelers origins and destinations were each randomly assigned to one of the four quadrants. Their final origin and destination locations are determined by drawing from a random normal distribution centered on their quadrant's cluster centroid, with a standard deviation equal to 5% of the length of an edge of the square service region. In the generation of trip origins and destinations, if the origins and destinations are less than 0.8 miles apart, the traveler is assigned a new destination.

Table 5 displays both the average traveler wait time and empty fleet miles results for the clustered demand scenarios. The results for Strategy 1 and Strategy 2 are not included because Tables 3 and 4 show that these two strategies are significantly less efficient than the optimization-based strategies. Once again, the results indicate that Strategy 6 unambiguously outperforms the other strategies in terms of empty fleet miles across all scenarios. Similarly, when fleet size is small relative to the demand rate, Strategy 6 outperforms the other assignment strategies in terms of average traveler wait times. However, with large fleet sizes, Strategy 3 and Strategy 4 outperform Strategy 5 and Strategy 6 in terms of average traveler wait times.

Table 5
Computational results for the clustered artificial demand scenarios.

Fleet Size	Strategy 3	Strategy 4	Strategy 5	Strategy 6	Strategy 3	Strategy 4	Strategy 5	Strategy 6
	Average Traveler Wait Time (min)				Ratio of Empty Miles to Total Miles			
Area Size = 16 mi²								
130	16.3	15.8	13.8	13.4	11.5%	11.0%	8.2%	7.9%
140	9.4	8.9	7.8	7.4	11.6%	11.1%	8.8%	8.4%
150	3.9	3.4	3.3	3.0	12.2%	11.7%	10.4%	9.7%
160	2.1	1.5	1.8	1.7	16.6%	14.0%	10.5%	9.7%
170	1.5	0.9	1.3	1.3	17.1%	12.7%	10.0%	9.3%
175	1.1	0.8	1.2	1.2	15.2%	11.9%	9.8%	9.2%
200	0.6	0.6	1.0	0.9	10.4%	9.7%	9.0%	8.6%
Area Size = 64 mi²								
230	21.8	21.1	19.6	19.2	11.3%	10.7%	8.3%	8.0%
240	17.7	17.1	15.9	15.4	11.3%	10.7%	8.6%	8.2%
250	14.0	13.4	12.6	12.1	11.4%	10.8%	8.8%	8.4%
255	12.3	11.7	10.9	10.6	11.4%	10.8%	9.0%	8.6%
280	5.1	4.5	4.8	4.4	12.4%	11.9%	10.4%	9.4%
305	2.9	2.1	2.8	2.8	16.8%	13.2%	10.4%	9.4%
330	1.5	1.3	2.4	2.5	12.8%	11.3%	10.1%	9.2%

Comparing the results in Table 5 with those in Table 4, the ratio of empty fleet miles is significantly lower in the clustered demand case compared to the uniform demand case. This is unsurprising as AVs are likely to have shorter distances between their drop-off of one traveler and the pickup of the next traveler if traveler origins and destinations are clustered. Nevertheless, this is an important finding, as it suggests, the percentage of empty fleet miles generated by an on-demand SAMS with no shared rides is likely to be heavily impacted by the spatial demand distribution.

6.2. Chicago taxi demand

To test the AV-traveler assignment strategies under a more-realistic spatio-temporal demand distribution, the Chicago taxi data was used (Chicago Data Portal, 2017). The Chicago taxi dataset includes over 100 million taxi trips taken in the Chicago region between 2013 and 2016. A taxi-trip observation includes a traveler pickup location, drop-off location, and pickup time. In this study, the traveler's pickup time is treated as the traveler's request time. Random days from September 2014 were selected from the Chicago taxi data.

6.2.1. Experimental design

Most of the input parameters for the taxi demand scenarios were the same as those in Table 2. However, the taxi simulation is an entire day ($T = 24$ h), the traveler demand rate (λ) increased to 3000–4000 travelers per hour on average; however, the traveler request rate is time-dependent. On Friday and Saturday night the demand rate approaches 10,000–12,000 travelers per hour. The average trip distance for the taxi trips is surprisingly only around 5 miles; however, the standard deviation of trip distance is around 6 miles. The Chicago service region is approximately 625 mi^2. Lastly, the inter-assignment interval (h) was increased to 30 s.

6.2.2. Taxi demand results

Table 6 displays the computational results for the taxi demand including the average traveler wait time and the ratio of empty miles to total miles. In one set of experiments, the full taxi data was used. In several other sets of experiments, 50% of the taxi data was used. The results in Table 6 provide more-evidence to support the conclusions in the artificial demand section. Strategy 6 decidedly generates fewer empty miles than the other assignment strategies. Moreover, Strategy 6 outperforms the other strategies in terms of average traveler wait time when the fleet size is small relative to the demand rate. However, Strategy 4 outperforms Strategy 6 in terms of average traveler wait time when the fleet size is large relative to the demand rate.

6.2.3. Note on computational time

This subsection briefly discusses the computational time associated with solving single instances of the AV-traveler assignment problem, every interval h in the dynamic system. The computational time for the two FCFS strategies are miniscule. For a static problem instance with 500 AVs and 500 travelers, Strategy 1 and Strategy 2 take less than a half-second to assign AVs to travelers. For Strategy 3 through Strategy 6, although they scale quite well, computational time does increase with the number of AVs and the number of travelers. The issue is moot for Strategy 3 in most cases and Strategy 5 in many cases because the number of idle AVs (for Strategy 3) and the number of idle and en-route drop-off AVs without a next pickup (for Strategy 5), respectively, tend to be quite

Table 6
Experimental results for taxi demand.

Fleet Size	Strategy 3	Strategy 4	Strategy 5	Strategy 6	Strategy 3	Strategy 4	Strategy 5	Strategy 6
	Average Traveler Wait Time (min)				Percentage of Empty Miles			
Taxi Day 1 (100% Demand)								
700	5.7	4.9	8.9	7.3	22.7%	21.6%	22.3%	20.4%
750	2.7	2.1	5.6	5.2	23.3%	21.5%	22.6%	19.7%
800	2.2	1.9	4.6	4.6	23.6%	21.6%	21.7%	19.5%
Taxi Day 2 (50% Demand)								
275	23.2	21.2	24.5	19.6	25.6%	24.6%	25.6%	23.1%
300	11.9	10.6	14.2	10.3	27.1%	24.6%	26.4%	23.1%
325	6.6	5.6	8.1	6.1	26.9%	24.7%	26.0%	22.5%
350	3.9	3.2	5.8	5.2	26.6%	24.1%	25.3%	22.5%
375	2.8	2.5	5.1	5.0	26.0%	23.9%	24.7%	22.4%
400	2.8	2.4	5.0	4.8	26.5%	23.6%	24.8%	22.2%
Taxi Day 3 (50% Demand)								
275	8.3	7.0	10.6	6.8	24.8%	22.9%	23.7%	19.8%
300	4.2	3.2	5.5	4.9	26.2%	23.4%	23.2%	19.9%
325	2.7	2.2	4.5	4.4	26.7%	23.3%	23.8%	19.9%
Taxi Day 4 (50% Demand)								
275	9.4	8.3	11.8	8.5	25.3%	24.2%	24.5%	21.3%
300	4.5	3.1	6.3	4.7	27.0%	23.8%	23.7%	20.8%
325	2.6	2.2	4.1	3.7	25.6%	23.5%	23.2%	19.6%

small at every interval h.

However, as Strategy 6 includes all AVs as well as all unassigned and assigned traveler requests in the assignment problem at every interval h, computational time is a concern. Strategy 6 requires about one second to obtain the global optimal for a problem instance with 350 AVs and 250–450 traveler requests. However, Strategy 6 requires about five seconds to obtain the global optimal with 700 AVs and 600–800 traveler requests.

As fleet sizes increase to thousands or tens of thousands of AVs, and tens of thousands of traveler requests per hour, heuristic methods will likely be required to decrease the solution space and/or solve large instances of the assignment problem.

7. Conclusion

7.1. Summary

Fully-autonomous vehicles (AVs) promise to increase the competitiveness of shared-use mobility services via eliminating the costs and performance limitations of human drivers. To drive down operational costs and maximize service quality, it is critical that shared-use AV mobility services (SAMSs) are operated efficiently. As such, this paper examines an on-demand SAMS with no shared rides, defined as a fleet of AVs that provides direct origin-to-destination transportation service to travelers who request rides via a mobile phone application and expect to be picked up within a few minutes. The paper focuses on modeling and comparing solution algorithms for the on-demand SAMS with no shared rides operational problem.

The operational problem is highly-dynamic and stochastic as traveler requests arrive randomly, travelers want to be served immediately, and the SAMS fleet operator has no advanced information about the traveler requests. To solve the problem, the SAMS fleet operator re-solves an AV-traveler assignment problem in real-time as new requests enter the system. The paper compares six different AV-traveler assignment strategies.

The first two assignment strategies are simplistic FCFS assignment strategies; the last four strategies require a mathematical programming solver. The optimization-based strategies, particularly the strategies that involve reassigning (diverting) assigned travelers (en-route pickup AVs) to other AVs (traveler requests), significantly outperform the simplistic FCFS assignment strategies. The more-sophisticated optimization-based assignment strategies significantly reduce (empty) SAMS fleet miles and average traveler wait times when the fleet size is small relative to the demand rate. However, as fleet size increases, the simple assignment strategies are comparable to the more-advanced strategies.

SAMS fleet operators should strongly consider employing optimization-based strategies, strategies that allow en-route pickup AVs to be reassigned, and strategies that incorporate en-route drop-off AVs in the assignment problem. These strategies allow fleet operators to handle the urgency of new traveler requests and the stochasticity of future traveler requests in a computationally-efficient manner.

Another important finding presented in this paper is that the spatial distribution of traveler requests significantly impacts the percentage of total fleet miles that are empty. Clustered demands result in a lower percentage of empty miles than uniformly distributed demands.

7.2. Contribution and practical implications

This paper makes several contributions to the transportation research literature including defining the *on-demand SAMS with no shared rides* and its underlying operational problem; modeling the highly-dynamic and stochastic operational problem as a dynamic AV-traveler assignment problem; presenting intelligent optimization-based AV-traveler assignment strategies that consider the unique characteristics of the SAMS; and illustrating that the intelligent AV-traveler assignment strategies significantly outperform the assignment strategies employed in the SAMS literature.

The models and assignment strategies presented in this paper can be employed by SAMS fleet operators that plan to offer an on-demand service with no shared rides. Relative to other strategies in the literature, the more-efficient strategies presented in this paper allow SAMS fleet managers to either (1) keep their current fleet size and reduce empty fleet miles (i.e. reduce operational costs) and reduce traveler wait times (i.e. improve customer service quality), or (2) decrease their fleet size (i.e. reduce capital costs) and still reduce empty fleet miles (i.e. reduce operational costs) but keep traveler wait times constant.

7.3. Limitations

This study uses a Manhattan grid network to represent an urban road network. This was done intentionally to prevent network structure and congestion effects from confounding the comparison of assignment strategies. However, going forward, it is necessary to model real, congestible road networks. Similarly, the study assumes the SAMS fleet operator has no stochastic information about future traveler requests to focus on the comparison of assignment strategies. However, stochastic information could be obtained from historical data and used to reposition AVs to serve future traveler requests. Finally, the paper ignores AV refueling/recharging constraints to focus on the comparison of assignment strategies.

7.4. Future research

Modeling and optimizing SAMS fleets that serve passengers is a relatively new area of research that requires further study. To further improve fleet efficiency, researchers should consider testing proactive AV repositioning strategies. Repositioning, or rebalancing, requires stochastic information about future demand requests. Rather than using rule-base repositioning strategies (Chen et al., 2016; Fagnant and Kockelman, 2016; Levin et al., 2017), research can employ optimization-based strategies that jointly assign AVs to open traveler requests, and position AVs to serve future traveler requests.

In addition to improving solution approaches to the on-demand SAMS with no shared rides, there are many other potential SAMSs to model (Hyland and Mahmassani, 2017). Shared-ride services and advanced demand request services appear to be two of the most likely SAMSs to exist in the future after the advent of AVs. The authors of this paper are extending the modeling framework in this paper to incorporate shared rides. However, researchers should also consider modeling the shared ride problem using more complex yet scalable formulations (Alonso-Mora et al., 2017). Modeling advanced and immediate traveler demand requests, simultaneously, is also a challenging problem. The advanced demand requests have tight and strict time-windows; whereas, the immediate demand requests still want to be assigned to an AV immediately and picked up within a few minutes.

Other research areas include incorporating pricing into the dynamic fleet modeling framework (Chen and Kockelman, 2016; Figliozzi et al., 2007; Sayarshad and Chow, 2015) and allowing travelers to accept or reject price and wait time offers from SAMS fleet operators.

Finally, there needs to be a real focus on developing robust solution algorithms for the operational problems associated with SAMSs. Although trucking, taxi, and other vehicle fleets currently use real-time control algorithms to provide *decision support* to vehicle dispatchers and/or drivers, with AVs, control algorithms will need to *make decisions* rather than support decision makers. The difference is important and needs to be reflected in the solution algorithms.

Acknowledgment

Partial funding for this work is provided through an Eisenhower Fellowship awarded by the US Department of Transportation to the first author. Additional funding is provided by the Northwestern University Transportation Center. The authors remain responsible for all findings and opinions presented in the paper.

Appendix A

This appendix presents the notation used in this paper to model the on-demand SAMS with no shared rides operational problem.

Sets	
T	Set of discrete time-periods, indexed by $t \in T$
R	Set of travelers/riders, indexed by $i \in R$
$R_U(t)$	Set of unassigned travelers at time t; $R_U(t) \subseteq R$
$R_A(t)$	Set of travelers assigned to an AV at time t; $R_A(t) \subseteq R$
$R_{IV}(t)$	Set of in-vehicle travelers at time t; $R_{IV}(t) \subseteq R$
$R_S(t)$	Set of served travelers at time t; $R_S(t) \subseteq R$
V	Set of AVs, indexed by $j \in V$
$V_I(t)$	Set of idle AVs at time t; $V_I(t) \subseteq V$
$V_P(t)$	Set of en-route pickup AVs at time t; $V_P(t) \subseteq V$
$V_D(t)$	Set of en-route drop-off AVs at time t; $V_D(t) \subseteq V$
Simulation model parameters	
$\|T\|$	Length of analysis period
τ	Current system time
v	AV Speed
Δt	Simulation time step
λ	Spatial-temporal demand rate of traveler requests (travelers/hour)
c_d	Time required to drop off traveler
c_p	Time required to pick up traveler
Traveler Information	
o_i	Pickup location of traveler i
d_i	Drop-off location of traveler i
r_i	Request time of traveler i
$w_i(t)$	Elapsed wait time of traveler i at time t
$\alpha_i(t)$	State of traveler i at time t
$posR_i(t)$	Position of traveler i at time t
$u_i(t)$	The AV assigned to pick up traveler $i \in R_A(t)$ at time t
AV Information	
$\beta_j(t)$	State of AV j at time t
$posV_j(t)$	Position of AV j at time t

Parameters: AV-Traveler Assignment Problem	
$d_{ij}(t)$	Manhattan distance between AV j and traveler i at time t
$a_i(t)$	Binary variable equal to 1 if traveler i is assigned $i \in R_A$ at time t
$y_{ij}(t)$	Binary variable equal to 1 if AV $j \in V$ is en-route to pick up traveler $i \in R$ at time t
$b_i(t)$	Binary variable equal to 1 if traveler $i \in R$ has been reassigned prior to time t
$p_j(t)$	Binary variable equal to 1 if AV j is en-route to drop off any traveler ($j \in V_D$) at time t
$q_j(t)$	Binary variable equal to 1 if AV j is en-route to pick up a traveler ($j \in V_P$) at time t
Decision Variables: AV-Traveler Assignment Problem	
x_{ij}	Binary variable equal to 1 if AV j assigned to traveler i
Exogenous Parameters: On-demand SAMS Operational Problem	
γ	Relative weight of elapsed waiting time $w_i(t)$ in objective function
φ	Penalty for assigning a traveler to an en-route drop-off AV
δ	Penalty for assigning a traveler to an en-route pickup AV
h	Inter-assignment time
Sequential Stochastic Control Problem Variables	
Π	Set of policies, indexed by $\pi \in \Pi$
x_t	Decision made at time $t \in T$
S_t	State of the system at time $t \in T$, indexed by $s \in S_t$
$X^{\pi}(S_t)$	Decision function for policy $\pi \in \Pi$ that returns a decision x_t given state S_t
W_t	Stochastic information that arrives between $t-1$ and t
$S^M(\cdot)$	Transition Function
$C(\cdot)$	Cost Function
$\mathrm{V}(\cdot)$	Value Function

References

Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., Rus, D., 2017. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. Proc. Natl. Acad. Sci. USA 114, 462–467. http://dx.doi.org/10.1073/pnas.1611675114.

Andersson, T., Värbrand, P., 2007. Decision support tools for ambulance dispatch and relocation. J. Oper. Res. Soc. 58, 195–201. http://dx.doi.org/10.1057/palgrave.jors.2602174.

Berbeglia, G., Cordeau, J.-F., Laporte, G., 2010. Dynamic pickup and delivery problems. Eur. J. Oper. Res. 202, 8–15. http://dx.doi.org/10.1016/j.ejor.2009.04.024.

Bischoff, J., Maciejewski, M., 2016. Simulation of city-wide replacement of private cars with autonomous Taxis in Berlin. In: Procedia Computer Science. http://doi.org/10.1016/j.procs.2016.04.121.

Boesch, P.M., Ciari, F., Axhausen, K.W., 2016. Autonomous vehicle fleet sizes required to serve different levels of demand. Transp. Res. Rec. J. Transp. Res. Board 2542, 111–119. http://dx.doi.org/10.3141/2542-13.

Burns, L., Jordan, W., Scarborough, B., 2013. Transforming Personal Mobility. New York.

Chen, T.D., Kockelman, K.M., 2016. Management of a shared autonomous electric vehicle fleet: implications of pricing schemes. Transp. Res. Rec. J. Transp. Res. Board 2572, 37–46. http://dx.doi.org/10.3141/2572-05.

Chen, T.D., Kockelman, K.M., Hanna, J.P., 2016. Operations of a shared, autonomous, electric vehicle fleet: Implications of vehicle & charging infrastructure decisions. Transp. Res. Part A Policy Pract. 94, 243–254. http://dx.doi.org/10.1016/j.tra.2016.08.020.

Chicago Data Portal, 2017. Taxi Trips.

Clewlow, R., Mishra, G.S., 2017. Disruptive Transportation: The Adoption, Utilization, and Impacts of Ride-Hailing in the United States.

Fagnant, D.J., Kockelman, K.M., 2016. Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in Austin, Texas. Transportation (Amst). 1–16. http://dx.doi.org/10.1007/s11116-016-9729-z.

Fagnant, D.J., Kockelman, K.M., 2015. Dynamic ride-sharing and optimal fleet sizing for a system of shared autonomous vehicles. In: Transportation Research Board 94th Annual Meeting.

Fagnant, D.J., Kockelman, K.M., 2014. The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. Transp. Res. Part C Emerg. Technol. 40, 1–13. http://dx.doi.org/10.1016/j.trc.2013.12.001.

Fagnant, D.J., Kockelman, K.M., Bansal, P., 2015. Operations of Shared Autonomous Vehicle Fleet for Austin, Texas, Market. Transp. Res. Rec. J. Transp. Res. Board.

Figliozzi, M.A., Mahmassani, H.S., Jaillet, P., 2007. Pricing in dynamic vehicle routing problems. Transp. Sci. 41, 302–318.

Fleischmann, B., Gnutzmann, S., Sandvoß, E., 2004. Dynamic vehicle routing based on online traffic information. Transp. Sci. 38, 420–433.

Ford, H., 2012. Shared Autonomous Taxis: Implementing an Efficient Alternative to Automobile Dependency. Princeton University.

Frantzeskakis, L.F., Powell, W.B., 1990. A successive linear approximation procedure for stochastic, dynamic vehicle allocation problems. Transp. Sci. 24, 40–57. http://dx.doi.org/10.1287/trsc.24.1.40.

Gendreau, M., Laporte, G., Semet, F., 2001. A dynamic model and parallel tabu search heuristic for real-time ambulance relocation. Parallel Comput. 27, 1641–1653. http://dx.doi.org/10.1016/S0167-8191(01)00103-X.

Gurobi, 2017. Gurobi Optimizer [WWW Document]. URL < http://www.gurobi.com/products/gurobi-optimizer >.

Haghani, A., Yang, S., 2007. Real-Time Emergency Response Fleet Deployment: Concepts, Systems, Simulation & Case Studies, in: Dynamic Fleet Management. Springer US, Boston, MA, pp. 133–162. http://doi.org/10.1007/978-0-387-71722-7_7.

Hyland, M.F., Mahmassani, H.S., 2017. Taxonomy of shared autonomous vehicle fleet management problems to inform future transportation mobility. Transp. Res. Rec. J. Transp. Res. Board 2653, 26–34. http://dx.doi.org/10.3141/2653-04.

Ichoua, S., Gendreau, M., Potvin, J.-Y., 2006. Exploiting knowledge about future demands for real-time vehicle dispatching. Transp. Sci. 40, 211–225. http://dx.doi.org/10.1287/trsc.1050.0114.

International Transport Forum, 2015. Urban Mobility System Upgrade: How shared self-driving cars could change city traffic.

Jaillet, P., Wagner, M.R., 2006. Online routing problems: value of advanced information as improved competitive ratios. Transp. Sci. 40, 200–210.

Seow, Kiam Tian, Dang, Nam Hai, Lee, Der-Horng, 2010. A collaborative multiagent taxi-dispatch system. IEEE Trans. Autom. Sci. Eng. 7, 607–616. http://dx.doi.org/10.1109/TASE.2009.2028577.

Larsen, A., Madsen, O.B.G., Solomon, M.M., 2002. Partially dynamic vehicle routing-models and algorithms. J. Oper. Res. Soc. 53, 637–646.

Lee, S., 2012. The role of centrality in ambulance dispatching. Decis. Support Syst. 54, 282–291. http://dx.doi.org/10.1016/j.dss.2012.05.036.

Lees-Miller, J.D., 2016. Minimising average passenger waiting time in personal rapid transit systems. Ann. Oper. Res. 236, 405–424. http://dx.doi.org/10.1007/s10479-013-1492-3.

Levin, M.W., Kockelman, K.M., Boyles, S.D., Li, T., 2017. A general framework for modeling shared autonomous vehicles with dynamic network-loading and dynamic ride-sharing application. Comput. Environ. Urban Syst. 64, 373–383. http://dx.doi.org/10.1016/j.compenvurbsys.2017.04.006.

Lund, K., Madsen, O.B.G., Rygaard, J.M., 1996. Vehicle routing with varying degree of dynamism.

Maciejewski, M., Bischoff, J., Nagel, K., 2016. An assignment-based approach to efficient real-time city-scale taxi dispatching. IEEE Intell. Syst. 31, 68–77. http://dx.doi.org/10.1109/MIS.2016.2.

Maciejewski, M., Nagel, K., 2013. Simulation and dynamic optimization of taxi services in MATSim.

Mahmassani, H.S., 2016a. 50th anniversary invited article—autonomous vehicles and connected vehicle systems: flow and operations considerations. Transport. Sci. 50 (4), 1140–1162.

Mahmassani, H.S., 2016b. Technological Innovation and the Future of Urban Personal Travel. In: Schofer, Mahmassani (Eds.), MOBILITY 2050: A Vision for Transportation Infrastructure. Northwestern University Transportation Center, Evanston, IL, USA (Chapter 4).

Martinez, L.M., Viegas, J.M., 2017. Assessing the impacts of deploying a shared self-driving urban mobility system: an agent-based model applied to the city of Lisbon. Portugal. Int. J. Transp. Sci. Technol. 6, 13–27. http://dx.doi.org/10.1016/j.ijtst.2017.05.005.

Mckenzie, B., Rapino, M., 2011. Commuting in the United States: 2009.

Pillac, V., Gendreau, M., Guéret, C., Medaglia, A.L., 2013. A review of dynamic vehicle routing problems. Eur. J. Oper. Res. 225, 1–11. http://dx.doi.org/10.1016/j.ejor.2012.08.015.

Powell, W.B., 2011. Approximate Dynamic Programming, second. ed. Hoboken.

Powell, W.B., 1996. A Stochastic formulation of the dynamic assignment problem, with an application to truckload motor carriers. Transp. Sci. 30, 195–219. http://dx.doi.org/10.1287/trsc.30.3.195.

Powell, W.B., Simao, H.P., Bouzaiene-Ayari, B., 2012. Approximate dynamic programming in transportation and logistics: a unified framework. EURO J. Transp. Logist. 1, 237–284. http://dx.doi.org/10.1007/s13676-012-0015-8.

Psaraftis, H.N., 1980. A dynamic programming solution to the single vehicle many-to-many immediate request dial-a-ride problem. Transp. Sci. 14, 130–154. http://dx.doi.org/10.1287/trsc.14.2.130.

Psaraftis, H.N., Wen, M., Kontovas, C.A., 2016. Dynamic vehicle routing problems: three decades and counting. Networks 67, 3–31. http://dx.doi.org/10.1002/net.21628.

Python, 2017. Python 3.5.1.

Regan, A.C., Mahmassani, H.S., Jaillet, P., 1995. Improving efficiency of commercial vehicle operations using real-time information; potential uses and assignment strategies. Transp. Res. Rec. 188–198.

Regan, A., Mahmassani, H., Jaillet, P., 1996a. Dynamic decision making for commercial fleet operations using real-time information. Transp. Res. Rec. J. Transp. Res. Board 1537, 91–97. http://dx.doi.org/10.3141/1537-13.

Regan, A.C., Mahmassani, H.S., Jaillet, P., 1996b. Dynamic dispatching strategies under real-time information for carrier fleet management. In: Transportation and Traffic Theory. Proceedings of the 13th International Symposium on Transportation and Traffic Theory, Elsevier.

Regan, A., Mahmassani, H., Jaillet, P., 1998. Evaluation of dynamic fleet management systems: simulation framework. Transp. Res. Rec. J. Transp. Res. Board 1645, 176–184. http://dx.doi.org/10.3141/1645-22.

Rigole, P.-J., 2014. Study of a Shared Autonomous Vehicles Based Mobility Solution in Stockholm.

Ritzinger, U., Puchinger, J., Hartl, R.F., 2015. A survey on dynamic and stochastic vehicle routing problems. Int. J. Prod. Res. 54, 215–231. http://dx.doi.org/10.1080/00207543.2015.1043403.

Sayarshad, H.R., Chow, J.Y.J., 2015. A scalable non-myopic dynamic dial-a-ride and pricing problem. Transp. Res. Part B Methodol. 81, 539–554. http://dx.doi.org/10.1016/j.trb.2015.06.008.

Schmid, V., 2012. Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. Eur. J. Oper. Res. 219, 611–621. http://dx.doi.org/10.1016/j.ejor.2011.10.043.

Sheridan, P.K., Gluck, E., Guan, Q., Pickles, T., Balcıoğlu, B., Benhabib, B., 2013. The dynamic nearest neighbor policy for the multi-vehicle pick-up and delivery problem. Transp. Res. Part A Policy Pract. 49, 178–194. http://doi.org/10.1016/j.tra.2013.01.032.

Spieser, K., Treleaven, K., Zhang, R., Frazzoli, E., Pavone, M., 2014. Toward a systematic approach to the design and evaluation of automated mobility-on-demand systems: a case study in Singapore. In: Meyer, G., Beiker, S. (Eds.), Road Vehicle Automation, Lecture Notes in Mobility. Springer International Publishing, pp. 229–245. http://doi.org/10.1007/978-3-319-05990-7.

van Lon, R.R.S., Ferrante, E., Turgut, A.E., Wenseleers, T., Berghe, G. Vanden, Holvoet, T., 2016. Measures of dynamism and urgency in logistics. Eur. J. Oper. Res. 253, 614–624. http://dx.doi.org/10.1016/J.EJOR.2016.03.021.

Viegas, J., Martinez, L., 2016. Shared Mobility: Innovation for Liveable Cities.

Wang, H., Lee, D.-H., Cheu, R., 2009. PDPTW based taxi dispatch modeling for booking service. In: 2009 Fifth International Conference on Natural Computation. IEEE, pp. 242–247. http://doi.org/10.1109/ICNC.2009.676.

Waymo, 2017. Waymo's fully self-driving vehicles are here [WWW Document]. Medium. URL < https://medium.com/waymo/with-waymo-in-the-drivers-seat-fully-self-driving-vehicles-can-transform-the-way-we-get-around-75e9622e829a > (accessed 11.7.17).

Wingfield, N., 2017. Automakers race to get ahead of silicon valley on car-sharing. New York Times.

Winter, K., Cats, O., Correia, G.H. de A., van Arem, B., 2016. Designing an automated demand-responsive transport system. Transp. Res. Rec. J. Transp. Res. Board 2542, 75–83. http://doi.org/10.3141/2542-09.

Yang, J., Jaillet, P., Mahmassani, H., 1999. On-line algorithms for truck fleet assignment and scheduling under real-time information. Transp. Res. Rec. J. Transp. Res. Board 1667, 107–113.

Yang, J., Jaillet, P., Mahmassani, H., 2004. Real-time multivehicle truckload pickup and delivery problems. Transp. Sci. 38, 135–148.

Zachariah, J., Gao, J., Kornhauser, A., Mufti, T., 2014. Uncongested mobility for all: a proposal for an area wide autonomous taxi system in New Jersey. In: Transportation Research Board 93rd Annual Meeting.

Zhang, W., Guhathakurta, S., Fang, J., Zhang, G., 2015a. The performance and benefits of a shared autonomous vehicles based dynamic ridesharing system: an agent-based simulation approach. In: Transportation Research Board 94th Annual Meeting.

Zhang, W., Guhathakurta, S., Fang, J., Zhang, G., 2015b. Exploring the impact of shared autonomous vehicles on urban parking demand: an agent-based simulation approach. Sustain. Cities Soc. 19, 34–45. http://dx.doi.org/10.1016/j.scs.2015.07.006.

Zhu, S., Kornhauser, A., 2017. The interplay between fleet size, level-of-service and empty vehicle repositioning strategies in large-scale, shared-ride autonomous taxi mobility-on-demand scenarios. In: Transportation Research Board 96th Annual Meeting. Transportation Research Board, Washington DC.