A survey on demand-responsive public bus systems

Pieter Vansteenwegen[a,*], Lissa Melis[b], Dilay Aktaş[a], Bryan Galarza Montenegro[b], Fábio Sartori Vieira[a], Kenneth Sörensen[b]

[a]*KU Leuven, Mobility Research Center – CIB, Celestijnenlaan 300, 3001, Leuven, Belgium*
[b]*ANT/OR – University of Antwerp Operations Research Group, Prinsstraat 13, 2000, Antwerp, Belgium*

Abstract

When demand for transportation is low or highly variable, traditional public bus services tend to loose their efficiency and typically frustrate (potential) passengers. In the literature, a large number of demand-responsive systems, that promise improved flexibility, have therefore been developed. At present, however, a comprehensive survey of these systems is lacking. In this paper, we fill this gap by presenting a unifying framework that classifies all demand-responsive public bus systems. The classification is mainly based on three degrees of responsiveness: dynamic online, dynamic offline, and static. For each system we discuss, among others, the optimization technique presented, whether realistic data is considered and the size of the instances used for testing. Different tables are included to structure and summarize the information of all papers.

1. Introduction

On-demand, *dial-a-ride*, *demand-adaptive*, *demand-responsive*, *flexible*, *flex-route*, *hybrid*, and *variable-type* are just a few of the names given to different types of modern bus systems that do not operate using fixed routes and timetables. Such systems offer a more flexible approach to public bus planning that takes into account the individual demand for transportation. Such systems — which we will call *demand-responsive public bus systems* (DR-PBS) — require additional information on where individual passengers are, where they would like to go, and at what time they wish to travel. This information can be gained in different ways, e.g., by asking passengers to make an explicit request for transportation using a mobile device, or by predicting it based on historical or real-time data.

A large number of DR-PBS has been recently proposed, that differ in the way the routes and/or timetables of the buses may change when the system receives a request for transportation. The sheer number of proposed systems, often differing only in a few details, make it difficult to draw conclusions from the state-of-the-art. Indeed, a comprehensive survey of all demand-responsive public bus systems is currently lacking. In this paper, the aim is to fill this gap by presenting a unifying framework or taxonomy that classifies all demand-responsive public bus systems, and to present an overview of recent studies guided by this classification. The aim of this paper is to (1) provide a comprehensive literature review on demand-responsive public bus system, and (2) develop a *taxonomy* of all recently proposed systems. We thereby focus on the properties of the (optimization) models that are built to support the planning of such systems (decision variables, objective(s), constraints), and only briefly discuss the solution methods developed.

Demand-responsive public bus systems are most useful in situations where conventional public bus systems do not perform well: when demand for transportation is either low (e.g., in a rural area) or has a large variance (e.g., there is a large difference between peak and off-peak hours). In these situations, buses

*Corresponding author
Email addresses: pieter.vansteenwegen@kuleuven.be (Pieter Vansteenwegen), lissa.melis@uantwerpen.be (Lissa Melis), dilay.aktas@kuleuven.be (Dilay Aktaş), bryan.galarzamontenegro@uantwerpen.be (Bryan Galarza Montenegro), fabio.vieira@kuleuven.be (Fábio Sartori Vieira), kenneth.sorensen@uantwerpen.be (Kenneth Sörensen)

that run according to fixed routes and timetables will tend to run either almost empty or filled to the brim. Both situations are undesirable: empty buses are inefficient for the bus company, whereas overfull buses are frustrating for the passenger. These systems take into account actual demand in an individual and short-term manner and adapt bus routes and/or timetables in much shorter time frames than traditional public bus systems.

It is incorrect to say that traditional public bus systems do not take into account demand for transportation, as both routes and timetables will have been determined based on historical demand data and may change throughout the day (e.g., running with a higher frequency during peak hours). However, the relationship between demand and supply in traditional public bus systems is a long-term and collective one. Short-term, individual user requests for transportation do not influence either the routes or the timetables of the buses in a conventional public bus system, except on the rare occasions when bus system operators manually intervene to mitigate problems (vehicle breakdown, road blocks, heavy congestion, etc.) or to better handle a predicted demand surge (e.g., a sports match).

We limit our review to *public* bus systems, i.e., systems that are open to the public. Obviously, these systems might be operated by private companies. However, systems operated by a private company to facilitate commuting for a select group of people (e.g., a company's own employees) are considered out of scope. Furthermore, we use the term public *bus* instead of public *transport* since we exclude train, tram, bike sharing, or other modes of public transport from this study. Also systems typically used individually, such as taxi services, including Lyft or Uber, are not considered in this paper. Finally, some papers focus on deciding, either in real-time or not, on the level of demand that justifies a switch from a conventional system to a demand-responsive system or the other way around, e.g., Chien et al. (2001) and Kim and Schonfeld (2012). If those papers are mentioned below, they are classified based on the demand-responsive system they propose to operate.

1.1. Structured literature review methodology

In order to gather all relevant papers on this topic, the Web of Science is searched for articles and book chapters published in operational research or transportation journals from the year 2000 or later. The search terms are constructed such that a term for "transportation" (Filter 1) and a term for "demand-responsive" (Filter 2) should be included in the title *and* the abstract, and a term for "solution method" (Filter 3) is mentioned in the abstract. Moreover, to limit the studies to the relevant domain, a list of forbidden terms (Filter 4) is used to eliminate irrelevant papers on topics that use similar terminology in a completely different context.

The search terms, per filter, are:

- F1: **Title and abstract contain at least one of:** Transportation, Transit, Transport, Connector, Mobility, Bus; AND
- F2: **Title and abstract contain at least one of:** Demand responsive, Customized, Ride sharing, Dynamic, Feeder, Flexible, Flex-route, On demand, Demand adaptive, Dial a ride; AND
- F3: **Abstract contains at least one of:** Optimization, Optimize, Model, Simulation, Simulate, Planning, Assignment, Design; BUT
- F4: **Title does not contain any of:** Dynamics, Kinematics, Spintronics, Polymer, Composite, Perovskite, Tetracene, Concrete, Ion, Quantum, Chemical, Thermal, Transistor, Production, Manufacturing, Flexible flowshop, Water transport, Heat transport, Air mobility

This search resulted in more than 300 papers, of which around half were considered out of scope for this survey after a manual check. In this phase, papers were excluded mainly because they described individual transport (taxi) systems or systems that are not really demand-responsive as discussed in the previous paragraph. This resulted in a total of 151 papers that are included in this survey.

1.2. Overview

The rest of this paper is organized as follows. In Section 2.1, the concepts that will be used in the classification are defined. The main classification of the demand-responsive public bus systems (DR-PBS) is based on the degree of responsiveness: the *dynamic online*, *dynamic offline*, and *static* public bus systems are presented in Section 3, Section 4, and Section 5, respectively. In those sections, the DR-PBS are further classified using a unified structure and conclusions on the respective demand-responsive systems are given. Lastly, in Section 6, summarizing conclusions on DR-PBS are presented.

2. Definitions and classification

Before presenting the classification used in this paper, a number of concepts need to be defined. This section also briefly introduces a number of well-know demand-responsive systems of which variants appear throughout the classification.

2.1. Definitions

In DR-PBS, pickups and drop-offs of passengers can occur on a *stop-based* basis, using a predefined set of potential stops; or on a *door-to-door* basis, where any location within a predefined area can be used for picking up or dropping off passengers.

Some systems are *many-to-many*, in which passengers can be picked up from any origin and dropped off at any destination, or *many-to-one* where the pickup and drop-off locations include a common origin or destination. The latter systems are also known as *feeder lines*. The term *service* is used for different departures of a single bus operated on a line or in an area. This means that if a frequency of three per hour is offered, three services are offered per hour.

Regarding the communication between passengers and operators, the system is considered to be *on-demand* if passengers are required to make *a request* in order for a service to be available for them. In this case, the minimum time required between the time of the request and the preferred departure time is called the *lead time* and it determines the computation time available for planning the service. For some systems, this planning may include *bus stop assignment*. This means that the operator assigns the pickup and/or drop-off location for each passenger.

An on-demand DR-PBS operates with *zero lead time*, if last minute bookings are possible and requests may come in real-time. If the operator allows passengers to make bookings in advance as well as last minute, then the system operates with *mixed lead times*. However, in order for a service to be *demand-responsive*, it does not necessarily need to be *on-demand*. For instance, an operator could register or accurately predict the number of passengers present at each stop and modify the next service based on that information. This is a demand-responsive system, which is not on-demand.

The demand-responsive system is considered *semi-flexible* if a standard bus route and timetable is predetermined for the different services during the planning horizon, but buses can deviate from these standards based on the demand. On the contrary, the system is *fully flexible* if the routes and timetables are determined from scratch based on actual demand information.

The operator's *planning horizon* is defined as the period for which the demand-responsive operation is planned with optimized stops, routes, scheduled departures or arrivals, and fleet assignment. This will typically be a whole day or a few hours during a peak or off-peak period, for instance.

The objective considers the *passengers' perspective* if it minimizes the passenger travel time (PTT) or the *operator's perspective* if it minimizes operational costs. Both objectives may be simultaneously optimized in a *multi-objective* manner. Another approach is to model one perspective as an objective, while the other is forced to attain a certain level through the constraints imposed.

2.2. Classification

In this paper, the DR-PBS will be mainly classified based on the degree of responsiveness, i.e., which demand-responsive changes are still possible during the planning phase. This degree of responsiveness is also directly related to the calculation time available for optimizing a system and, therefore, it also determines how challenging this optimization will be. The optimization of demand-responsive systems can occur *statically* or *dynamically*:

- Dynamic Optimization
  A basic schedule is made for the planning horizon, but changes to this schedule are made based on actual demand data. A dynamic operation can be classified as online or offline, according to the moment that changes can be made:
  - **Dynamic Online Optimization**: The basic plan for all services can change during operations, even for services which are already running.
  - **Dynamic Offline Optimization**: The basic plan for some services can change during operations, however, changes are only possible for services which have not started yet.
- **Static Optimization**
  The planning is completed before the start of the planning horizon, i.e., while the services are not running yet. A schedule is made for all services during the entire planning horizon and this schedule is not subject to any changes during the operations. The difference with fixed or conventional services is that static demand-responsive services offer a different service every planning horizon. E.g., the service of a given day may be based on the measured or predicted demand of that particular day.

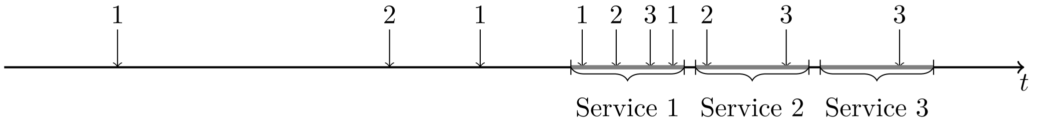
Since this degree of responsiveness is crucial for this survey but not always easy to grasp, an example is discussed now. Consider an on-demand DR-PBS with a planning horizon of one day. If the demand-responsive system updates the planning of stops, routes, scheduled departures, arrivals, etc. during the day of operation, the system is considered as *dynamic*. With short or zero lead time, the planning is required to be updated even for the services that are currently running in order to accommodate last-minute bookings. For instance, an ongoing service might be required to wait longer at a stop or to make a detour to serve a new passenger request that just came in. Therefore, these *dynamic online* systems offer the most flexible operation with the shortest lead times. With somewhat longer lead times, such as one hour, the operator offers to serve the passengers as long as the request is received one hour before the desired pickup time. To accommodate these newly incoming requests, the planning can be kept as fixed for the services that are already running and it will be updated only for the services that have not started yet. For example, the departure time of one of the next services is advanced and an additional pickup is added to the route. In this case, this is considered as *dynamic offline*. On the contrary, if the system collects passenger requests a day in advance in order to optimize its operations, passengers face a day of lead time and the system is considered as *static*. This means that the stops, routes, scheduled departures, arrivals, and fleet assignment do not change during the day of operation, but are executed exactly as planned before the day starts. The different degrees of responsiveness are further illustrated in Fig. 1.

The next three sections will discuss systems with dynamic online optimization (Section 3), dynamic offline optimization (Section 4), and static optimization (Section 5), respectively. In each of these sections, papers are grouped first considering whether the system is many-to-many or many-to-one, then whether it is fully flexible or semi-flexible, and lastly based on the objective (passenger's perspective, operator's perspective, or multi-objective). For each system we discuss the optimization technique presented and how close the system is to implementation. Some systems are (almost) implemented in practice, while most are more theoretical. We also discuss if realistic data is considered and the size of the instances used for testing. This information is also summarized in the tables in the Appendix.

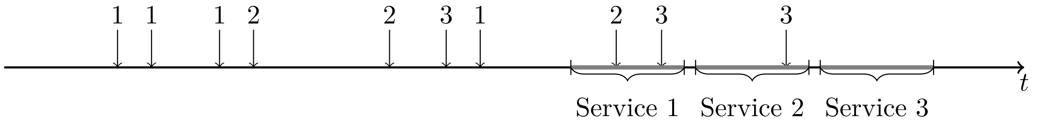
2.3. Well-known systems

In this section, four well-know demand-responsive systems are briefly introduced. Because different variants of these systems have been developed, they appear across our classification, and we therefore discuss them in this section.

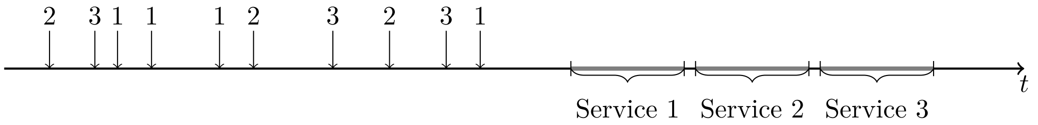
The well-known **dial-a-ride (DAR)** problem (Wilson et al., 1971) is the first example of such a system. In the DAR problem, passengers make requests with desired pickup and drop-off locations and the aim is



(a) Dynamic *Online* Optimization: requests are received after the buses start, even for services that have already departed



(b) Dynamic *Offline* Optimization: requests are received after the buses start, but only for services that have not yet departed



(c) Static Optimization: all requests are received before the first service starts

Figure 1: Different types of system according to degree of responsiveness; arrows indicate when a request for a service with a certain number (1, 2, 3) appear.

to accommodate as many passengers as possible with the least operating cost. This is probably the most general system considered in this survey paper and it is typically classified as a fully flexible many-to-many door-to-door system. Ho et al. (2018) discuss a set of recent studies on the DAR problem.

The **Demand-Responsive Connector (DRC)** is a more specific demand-responsive system connecting a residence area with low demand to a transportation hub. It is typically operated as a fully flexible many-to-one system. Next to having a low demand spread over a large area, this demand is mostly concentrated in peak hours and in one direction. DRC typically operates a door-to-door system and variations of the problem determine optimal routes (Ceder (2013), Lee and Savelsbergh (2017)), optimal zone sizes (Li and Quadrifoglio (2009), Li and Quadrifoglio (2011)) and passenger costs (Zheng et al. (2018)).

Quadrifoglio and Dessouky (2004) introduce a system called **Mobility Allowance Shuttle Transit Service (MAST)** where the vehicles have a basic route from which they can deviate in case of sufficient slack time. This system is discussed in detail by Errico et al. (2013). The authors present a unifying framework for these semi-flexible many-to-many systems using the demand-responsive characteristics: route deviation, point deviation, demand-responsive connector, request stops, flexible route segments, and zone routes, defined in the book of Koffman (2004).

Finally, the **Customized Bus (CB)** was operated in practice in 22 Chinese cities, before it was thoroughly and scientifically analyzed the first time by Liu and Ceder (2015). In most cases, it is operated as a many-to-many door-to-door semi-flexible system. Typically, an origin area is connected to a destination area with an express service and mostly dedicated lanes, but the routes inside both areas are flexible. Moreover, the CB users are actively involved in the planning activities, using interactive and integrated information platforms. Depending on the degree of responsiveness of those design decisions, many CB papers present static, dynamic offline and dynamic online variants, which will be considered in the next sections.

3. Dynamic online demand-responsive public bus systems (DON-PBS)

Demand responsive public bus systems with the highest levels of responsiveness and flexibility employ a dynamic and online optimization strategy. This generally means that requests can come in during the planning horizon as last minute bookings, with little to zero lead time. However, it is usually still possible to book a trip in advance as well. Changes to the bus routes and schedules can be triggered by online requests, but also, e.g., by an unexpected crowding on a bus or a traffic jam. Dynamic online demand-responsive public bus systems (DON-PBS) can be reactive or proactive. Reactive systems process demand when it occurs, i.e., they only take into account the currently known demand. Proactive systems not only process known demand, but also anticipate future unknown demand. Research is mainly done on reactive DON-PBS, therefore the papers mentioned in this section investigate reactive DON-PBS, unless stated otherwise.

3.1. Many-to-many

When discussing DON-PBS, we first consider systems of the type *many-to-many*, where passengers can travel from any bus stop or location to any other bus stop or location. In Section 3.2, *many-to-one* systems will be considered where all passengers travel to (or from) the same destination (departure stop).

3.1.1. Fully flexible routes and timetables

As for the other demand responsive bus systems, the fully flexible systems are discussed first, i.e., systems where no standard route or timetable is available. Semi-flexible systems, where a basic service is modified based on the demand, are discussed in Section 3.1.2.

3.1.1.1 Passengers' perspective

Jokinen et al. (2011) consider systems with a set of potential stops that customers can choose from as their pickup and drop-off points. Demand is high and requests can be made online, with zero lead time. A set of small buses is then operated to pickup customers from these stops and bring them to their destination stop without a transfer. To define the routes of the vehicles, with a limited capacity, in response to these online requests, the authors use a simple greedy heuristic. The initial plan of a request cannot change anymore after insertion, however, note that the routes of the vehicles still change dynamically when new requests keep coming in. Compared to private car usage, the DON-PBS offers a similar service for a lower price to be paid by the user. Compared to a taxi service, it shows more resiliency to changing demand. Mean travel times of the taxi service rise exponentially while those of the DON-PBS rise in an almost linear way with increasing demand. Vallée et al. (2017) present another high demand DON-PBS adopting the stop-based approach and an optimization from the passengers' perspective, but with mixed lead times. The algorithm presents several transport options within the requested time window, that a passenger can choose from.It is possible for a request to be rejected when it cannot be inserted in a feasible way. The authors aim to maximize the number of served passengers for a given fleet size, by either minimizing the total travel time or maximizing the slack times. The slack time is the difference between the latest arrival time of a passenger at a stop and the actual arrival time when the bus drops the passenger at the arrival stop. It is concluded that using the travel time objective allows to insert the maximum number of requests. The method first uses a quick insertion heuristic to answer a potential customer whether or not it would be feasible to serve his request. In a second phase the authors perform an adaptive large neighborhood search to further improve the solution until a next dynamic request comes in. In contrast to the previous work, the initial plan for a request can still be adapted, but only requests that are not picked up yet at the present time, can be removed and re-added. The instances are based on real-life data from the DR-PBS service *Padam*, implemented in Paris (France) and Bristol (UK). The authors also perform a comparison between the dynamic and the static version of the DR-PBS. In the static variant, the same requests are all known in advance and a time limit of 10 minutes is implemented as a stop criterion. Counter-intuitively, better results are found for the dynamic variant. A reason for this finding can be found in the fact that the static optimization problem is rather complicated to solve in only 10 minutes. The authors conclude that the difference between the dynamic and static variant rises with the fleet size. The larger the number of vehicles, the larger the difference in performance in favor of the static variant. Another stop-based mixed lead time approach can be found in Melis and Sörensen (2021). In contrast with the previous study, the authors perform a theoretical study on a grid which represents an urban context. The total user ride time is optimized, which is the time the passengers spend on the bus. Every request has a time window, and the fleet size and capacity of the vehicles is fixed. A large neighborhood search heuristic is used, with some embedded local search. They allow itineraries to change even if a passenger is already on the bus. Of course, when a passenger is already picked up, he has to be dropped off by the same bus as well and the time window constraints of all passengers need to be maintained. Also, the algorithm includes bus stop assignment. Every request has *a set* of bus stops for departure and arrival, all within walking distance of the actual origin and arrival locations. In the analysis, the authors examine the influence on the solution quality of the real-time requests, compared to an entirely static system. In contrast to the findings of Vallée et al. (2017), it is concluded that the more requests are known in advance in the static system, the lower the total user ride time compared to the dynamic system with real-time requests.

Besides a stop-based approach, there are also DON-PBS adopting a door-to-door service with zero lead time. Navidi et al. (2018) perform a simulation of a high demand door-to-door DON-PBS and compare the results with a conventional stop-based PT system. The capacity of the DON-PBS-vehicles (4 passengers) is chosen to be substantially lower than the buses used for the fixed transport lines (75 passengers). The fleet size of the on-demand system is adapted in such a manner that the reject rate is zero. The authors minimize travel time and the ratio of the scheduled travel time on the direct travel time. Their study is performed in a theoretical grid-based and star-shaped network, but also in a network based on a real world suburban city. It is concluded that the DON-PBS performs better for both low and high demand areas in terms of passenger travel time, but passengers do have to wait longer for a vehicle to pick them up compared to conventional PT. The routing algorithm used is based on the work of Ronald et al. (2013). The algorithm uses an insertion algorithm that checks every position in every vehicle and finally inserts where the objective function increases the least. Hence, the initial plan for a request cannot change after insertion. A grid-based environment is used and contrarily to Navidi et al. (2018), request rejection is possible. Especially for long distance, but also for random requests, the waiting time increases substantially and the percentage of passenger sharing a ride is high. On the other hand, empty kilometers increase for the many-to-one and short distance trips. The door-to-door approach is also simulated by Archetti et al. (2018), Narayan et al. (2017) and Alonso-Mora et al. (2017). Their work all focuses on high demand transit systems, e.g., in an urban context. Both Archetti et al. (2018) and Narayan et al. (2017) simulate different transport modes, e.g., fixed PT or private cars, in co-existence with a dynamic online system to analyze choice behavior and possible travel times. For both systems the fleet size of the DON-PBS is given. In Archetti et al. (2018) (and in the more extended version of their work in Archetti et al. (2016)) a user chooses between a fixed public bus and an on-demand system in a grid-based theoretical network and takes the one with the shortest travel time. Travel time includes walking time. If none of the above can satisfy the request, the user takes a private car. In the on-demand system, a request is inserted in a greedy manner where the travel time increases the least. When inserting a request, the actual pickup time of any passenger cannot exceed the desired pickup time plus a maximum flexibility value. Narayan et al. (2017) also run a scenario with fixed PT, private car and shared on-demand transport, but also one with a private on-demand variant in a high demand network based on an American city. Another difference with the previous study is that monetary values of time for using the different transport modes are included. The simulation maximizes a utility function consisting of these monetary values and the utility of travel. The authors do not mention any constraints on the time windows of requests or on the capacity of the vehicles. The main conclusions of both studies are that a DON-PBS system would be able to compete with private car usage in terms of costs and travel time, if the fleet size and the number of requests is high enough. Narayan et al. (2017) also conclude that a shared on-demand transport system is preferred over a private taxi system when the cost of flexible transport increases compared to the one of the fixed PT. Alonso-Mora et al. (2017) use a unique approach within this research topic. A reactive anytime optimal algorithm is used, by performing integer linear programming (ILP). Every few seconds, the algorithm efficiently assigns a batch of requests to a fleet of autonomous vehicles with a certain capacity in a greedy manner and then refine the solution over time. A cost function is minimized containing the travel delay (drop-off time minus request time minus direct driving time) and a large constant for unassigned requests. Request time is the time when the request is sent. By choosing this, instead of pickup time, the waiting time is included in the objective. After the assignment stage, the single-vehicle DARP is used to compute the schedule (routes and timetable) for the assigned requests in each vehicle. For vehicles with small capacities, the optimal schedule is found, otherwise they propose to use tabu search, simulated annealing or Lin-Kernighan as a heuristic approach. The New York (USA) taxi dataset is used and consequently the authors handle extremely high demand. In addition, a first step towards a proactive DON-PBS is made by relocating empty vehicles to places where there is currently a high demand, with a high request rejection rate. It is concluded that 98% of all taxi rides currently served with 13.000 taxis, can be served with only 3000 taxis of capacity four.

A heuristic for a medium scale door-to-door, many-to-many, fully flexible DON-PBS can be found in Santos and Xavier (2013). The heuristic maximizes the number of requests served and the number of shared rides. Constraints are set on the vehicle capacity, request time window and each vehicle has a time window of operation. The authors solve the dynamic online problem by solving the static version of the problem (with the requests currently known) with a GRASP-based heuristic every short period of time. Once a feasible solution is found, the algorithm performs local search. Other recent dynamic ride-sharing algorithms can be found in, among others, Pelzer et al. (2015), Najmi et al. (2017), Li et al. (2018a) and Smet (2019). The main difference between ride-sharing (using private vehicles and car owners as drivers) and taxi/bus-sharing (using operator owned vehicles) is the fact that in the first the car owner's destination always has to be the last scheduled stop in the route. Otherwise (s)he would have to hand over his (or her) private vehicle to the other passengers scheduled in the route. As this is no longer PT and outside the scope of this survey, literature on ride-sharing using driver owned vehicles is kept to a minimum.

Liu et al. (2019a) formally define a bus rides-haring problem with a large-scale bus ride-sharing service in operation. Real-life taxi data from Shanghai (China) is used to model the problem. The objective function is to maximize the ride-sharing success rate and to minimize passenger travel distance. The Customized Bus (see Section 2.3) routes are set according to pickup and drop-off locations and based on the scheduled time for the trip. The capacity of each bus limits the clustering of passengers on the same trip. The efficiency of the system is identified by comparison with the taxi ride-sharing service in similar conditions in relation with the time and price for the service.

Finally, a comparison between DON-PBS simulation approaches can be found in Ronald et al. (2017). Passenger inconvenience is minimized by minimizing the travel time with a DR-PBS divided by the time a passenger would need to traverse the same distance directly. It is concluded that over- or underestimation of the predicted travel output (vehicles kilometers, proportion of shared rides, empty kilometers etc.) can occur, because the value of the results statistically differs according to the simulation method used.

An overview of many-to-many fully flexible DON-PBS optimizing from the passengers' perspective can be found in Table 1. Most authors generally focus on minimizing passenger travel time, whether or not in relation to the direct travel time, or on maximizing the service rate. Ronald et al. (2017) is not included in the table because the authors analyze simulation methods for DON-PBS in general without specifying any constraints.

3.1.1.2 Operator's perspective

Many-to-many fully flexible DON-PBS are also frequently optimized from the operator's perspective, mostly to make sure the system is profitable. A stop-based approach with zero lead time is adopted by Tsubouchi et al. (2010) and Bertelle et al. (2009). The first build a heuristic based on efficient vehicle choosing and online insertion. Their algorithm is more thoroughly explained in Tsubouchi et al. (2009). The authors focus on reducing vehicle distance and computation time. The travel times of the users are constrained with a maximum ride time and each passenger needs to be transported within a time window. The fleet size and capacity are given. Actual field tests of a DON-PBS are carried out in different Japanese cities, but the number of requests stayed rather low. The authors conclude that the field tests show that the heuristic can handle large scale problems and customers perceived the system as practical and efficient. The work of Bertelle et al. (2009) stands in contrast with the previous because this work investigates a theoretical decentralized DON-PBS. The system emerges from the present fleet of vehicles. The authors minimize the additional vehicle travel time when adding a request to the current solution. Once a request comes in, every vehicle calculates the additional time it would take to serve this request, plus a potential penalty cost and sends the answer to the rest of the fleet. To make sure waiting times are reasonable, vehicles are attached to zones. The penalization refers to a vehicle leaving its zone. Then, every driver ranks the vehicles and the vehicle which is ranked first the most, wins the new request. The authors also take into account current traffic by putting time-dependent weights on the travel times between stops. The authors conclude that their algorithm works better with a limited number of passengers having long distance requests, instead of

Table 1: An overview of many-to-many fully flexible DON-PBS

		Constraints on ...					
Reference	Objective	C	F	VTT	PT	DT	PTT
Jokinen et al. (2011)	min PTT	x		x	x		
Vallée et al. (2017)	max P	x	x			x	x
Melis and Sörensen (2021)	min total user ride time	x	x		x	x	x
Navidi et al. (2018)	min PTT + PTT/DTT	x					
Ronald et al. (2013)	min PTT + PTT/DTT	x	x				
Archetti et al. (2016)	min PTT	x	x		x		
Narayan et al. (2017)	min PTT		x				
Alonso-Mora et al. (2017)	min PTT - DTT and max P	x	x		x	x	
Santos and Xavier (2013)	max P and shared rides	x	x	x		x	x
Liu et al. (2019a)	min PTT + max P	x			x	x	
Tsubouchi et al. (2009, 2010)	min VTD	x	x	x	x	x	x
Bertelle et al. (2009)	min VTT	x	x				
Bischoff et al. (2017, 2018), Viergutz and Schmidt (2019), Leich and Bischoff (2019)	min VTT	x	x	x	x		x
Wang et al. (2019)	min VTD for pickup	x	x		x		
Simonetto et al. (2019), Pandey et al. (2019)	min VTT	x	x		x	x	
Ma et al. (2013)	min VTD	x	x		x	x	x
Kawamura and Mukai (2009)	min VTD		x				
Bruni et al. (2014)	min costs	x	x	x	x	x	x
Horn (2002b,a)	min VTT	x	x		x	x	x
Liyanage and Dia (2020)	min PTT and costs	x	x				
Jäger et al. (2018)	min PTT and costs	x	x				
Winter et al. (2018)	min PPT and VTT	x			x		
Gomes et al. (2014, 2015)	min PPT and costs, and max P	x	x		x		
Hyland and Mahmassani (2020)	min PPT and VTT	x	x				x
Atasoy et al. (2015b,a), Ikeda et al. (2015)	max profit or max CS or max profit and CS	x	x				x
Ronald et al. (2015)	min PTT/DTT or min VTD		x				
Dessouky et al. (2003)	min costs and environmental impact	x	x		x		
Van Engelen et al. (2018)	min PPT and VTT, and max P	x	x	x	x	x	x

(P = The number of passengers served, C = Capacity, F = Fleet size, VTT = Vehicle travel time, VTD = Vehicle travel distance, PT = Pickup time, DT = Drop-off time, PTT = Passenger travel time, CS = Customer surplus, DTT = Direct travel time)

a large number of passengers having short distance requests.

A comparison between a DON-PBS with autonomous vehicles and conventional, fixed PT is found in Viergutz and Schmidt (2019), Bischoff et al. (2018) and Leich and Bischoff (2019). The insertion algorithm used for the DON-PBS in these simulation studies can be found in Bischoff et al. (2017). It focuses on minimizing the total taxi workload, meaning the total vehicle driving time needed to handle all requests. Therefore the algorithm inserts a request, where the detour is the smallest. There are constraints on both the maximum passenger travel time and waiting time, but each vehicle also has a time window of operation. The fleet size and capacity of the vehicles are given. In this work, the authors conclude that the DON-PBS saves 15-20% vehicle kilometers compared to a non-shared taxi system. Viergutz and Schmidt (2019) perform a simulation study in a rural town. As a consequence, the number of requests is relatively low. Even though the DON-PBS performs better from a passenger, ecological and societal perspective, it is concluded that a DON-PBS would not be beneficial in a rural context, caused by the high operational costs. For rural areas, a DON-PBS which starts from a standard route with possible deviations or extra stops on demand is proposed in Section 3.1.2. Bischoff et al. (2018) perform the same study in an urban context, with a data set for the city of Cottbus (Germany). The authors also compare a stop-based and door-to-door DON-PBS with conventional PT. It is found that the fleet size necessary to fulfill all constraints for a door-to-door service needs to be 1.33 times as large as the fleet size needed for a stop-based approach. It is also concluded that both are cheaper to operate than a fixed-line PT system. In addition, travel times for passengers are lower, with the lowest travel times for the door-to-door approach. Lastly, Leich and Bischoff (2019) do the simulation using a real-life Berlin (Germany) data set, more specifically in a Berlin suburb. In the service area, the existing bus lines are replaced with a DON-PBS. In this study only a door-to-door approach is used. The results are less spectacular compared to the previous study, but still in favor for the DON-PBS. In contrast to Viergutz and Schmidt (2019) and Bischoff et al. (2018) the authors vary, next to the fleet size, also the capacity of the taxi vehicles used in the DON-PBS and conclude that 150 vehicles with a capacity of 4 people is enough to serve an area with 24.000 inhabitants.

Wang et al. (2019) switch between a stop-based and door-to-door approach depending on the time of day (peak or off-peak hours) and preferences of the users. The travel times between two locations vary depending on traffic and congestion. When assigning requests to vehicles, the algorithm minimizes the empty-vehicle travel distance for pickups, by using the Hungarian method, an algorithm to solve an assignment problem in polynomial time (Kuhn, 1955). Autonomous vehicles are simulated in a theoretical urban context, but the capacity of the vehicles is only two. A time dependent system can provide decent quality of service, while needing less energy and traveled vehicle distance.

Simonetto et al. (2019) improve the work of Alonso-Mora et al. (2017) (mentioned in paragraph 3.1.1.1) and solve a linear assignment problem for a door-to-door DON-PBS with automated vehicles. The same New York taxi data set is used. The objective is slightly different compared to the one of Alonso-Mora et al. (2017): the route duration of the vehicles is minimized. Because the authors use a one to one assignment instead of a one to multiple assignment, there are less possibilities to consider and the algorithm needs four times less computation time, while the quality of the solution remains. The work also makes a first move towards investigating competition among two DON-PBS operators, with a 75% and 25% market share respectively. The objective function of their request assignment procedure is adapted to impose these market shares and find that especially vehicle travel time, but also passenger waiting time increases. Pandey et al. (2019) picked up the idea of competition and examine optimization-based approaches to model cooperation and competition between multiple ride-/taxi-sharing companies. The authors use the same assignment problem as Simonetto et al. (2019) to assign requests to vehicles. It is found that competition, where the passenger chooses the best offer between different ride-sharing firms, lowers the quality of the service, but that cooperation counteracts this decline. Pandey et al. (2019) prove they can still solve the problem very close to optimality, even when, as in reality, noise occurs, e.g. a company sends discounted prices to attract more requests.

Ma et al. (2013) minimize the total travel distance of a large-scale door-to-door taxi-sharing DON-PBS with a given fleet size. A network based on the city of Beijing (China), is used. Each passenger has two time windows, one for the pickup and one for the arrival. To limit the passenger waiting time, a passenger needs to be picked up five minutes after the request was send. The algorithm first compiles a taxi candidate list that would be able to satisfy a new request by taking into account the spatio-temporal location of taxis near only the origin and, both the origin and drop-off point/time window of the request. It is concluded that by taking into account both the origin and drop-off, the computation time halves as less vehicles need to be considered. The authors find that the travel distance is reduced with 13% compared to a non-sharing alternative. A DON-PBS with electrical vehicles is optimized by Kawamura and Mukai (2009). The algorithm minimizes vehicle travel distance and the battery usage by using an insertion algorithm combined with a genetic algorithm, that takes into account the need for battery recharging. The authors only test their algorithm in a low demand setting and do not implement capacity constraints. It is found that the system reduces carbon emissions with 80% and the cost with 60% compared to traditional systems.

As mentioned before, research is mainly done on reactive DON-PBS, but a proactive DON-PBS system, stop-based and with mixed lead times can be found in Bruni et al. (2014). Probabilistic information about future demand is used to build robust routes in a low demand theoretical context. When the algorithm builds routes based on the requests known beforehand, it already takes into account the probability of future requests to make the routes more steady. A tabu search heuristic is used with a frequency matrix to solve each scenario as a subproblem. In a second stage, the scenario solutions are merged with routes based on the static requests by looking at similarities. The authors minimize the routing, vehicle usage and recourse cost with constraints on the passenger travel times through time windows and a constraint on the total route duration. Also, each vehicle has a certain capacity. The results of their heuristic are compared with CPLEX for instances with only 4-6 requests and found a 6.6% gap on average. It is found that the cost decreases when the percentage of real-time requests decreases. Also, the advantage of taking into account the uncertainties rises when the percentage of real-time requests is relatively lower.Also Horn (2002b) anticipate for future demand and simulate several transport modes in a high demand context, among which a door-to-door shared taxi system and a stop-based on-demand bus system. Also multi-leg journeys are possible, consisting of different transport modes (see paragraph 3.1.2.3). The goal of this research is to analyze theLITRES-2 modeling system. Vehicle travel time is minimized and the vehicle occupancy rate is maximized. Passenger travel time is constrained by a time window. A local search heuristic is used after every request insertion and a more general improvement procedure that is periodically applied. The heuristic is explained in Horn (2002a). In this work, each request asks for a specific transport mode. Two request insertion strategies are tested: one inserts the origin in a greedy manner and then inserts the arrival given the found origin insertion, and the second checks all origin-destination combinations (exhaustive search). The authors compare both insertion ways and conclude that the exhaustive search works best for shared transport services like the ones mentioned above. After insertion, the original route is expanded and a local search procedure looks for requests that are now close-by the expanded route for a possible route-transfer. Also within route order changes of requests are allowed. The simulator also introduces trip cancellations, vehicle breakdowns and delays, causing trips to be rescheduled. The authors also compare the results for a static or dynamic DR-PBS to the results obtained with only a static or dynamic private taxi system. The authors find 91% and 33% longer ride times for a dynamic private taxi system compared to a S-PBS and DON-PBS, respectively.

An overview of the literature optimizing many-to-many fully flexible DON-PBS from the operator's perspective can be found in Table 1. Most literature focuses on minimizing the vehicle travel time or distance, which has a direct impact on the operating costs. To make sure the passenger travel times do not increase too much, all authors place constraints on either the pickup time, delivery time and/or the maximum passenger travel time.

3.1.1.3 Multi-objective

Liyanage and Dia (2020) and Jäger et al. (2018) both simulate a stop-based, zero lead time DON-PBS. Both works do another comparison with an existing fixed bus service. The first optimizes the passenger travel time and operational costs in the inner city of Melbourne (Australia). The results show benefits for both passengers and operator. Waiting times are reduced 78 to 95% depending on the time of day and vehicle utilization increases. Also the trip completion rate increases from 67% for the traditional PT to 85% for the DON-PBS. The total passenger-kilometers traveled stay the same for both systems of the comparison.

Jäger et al. (2018) simulate a DON-PBS with autonomous vehicles on a Singapore-based network. A possible vehicle directly rejects a request when this request would exceed the capacity, otherwise the algorithm tries to insert the requested stops in their route. A central dispatch calculates a cost function based on expected travel time (including waiting time) of the current request and existing passengers in the route, and energy consumption. Then, the best vehicle assignment is determined. It is found that the energy consumption improves compared to a fixed, static system, but no passenger service improvement is found.

Winter et al. (2018) simulate a high demand stop-based, mixed lead time, DON-PBS and minimize operational and passenger travel time. The latter includes both in-vehicle time and the waiting time at the pickup stop. The authors keep increasing the fleet size until all passengers can be served within a predefined average and individual passenger waiting time. The authors compare the automated DON-PBS with an existing bus system in the city of Arnhem (The Netherlands) and find that the operational costs of the DON-PBS lie in the same range as the fixed bus system, while the average passenger waiting time of the DON-PBS is only four minutes. It is also found that economies of scale play an important role in DON-PBS. Higher passenger demand decreases operation costs and travel times.

Gomes et al. (2014) also adopt a stop-based approach with mixed lead times and start by solving the static part of the problem. For the dynamic part, when a new request arrives, the algorithm solves the static scenario again with a GRASP-based approach, including an improvement phase. Hard constraints are set on the pickup time window and soft constraints on the delivery time window. Request rejection is possible.Their algorithm is tested on instances with a maximum of 500 requests and optimize the solution both from the passengers' as from the operators perspective. It is found that computation time is strongly dependent on the number of requests and not on the number of possible stops. Also, it is concluded that the objective function value increases 45% when going from an entirely static scenario to one where there are 90% real-time and 10% static requests. Later, their algorithm is integrated in a simulation study of the DR-PBS system in Porto (Portugal). A fixed night-time bus lineis replaced with an existing DON-PBS, called *Gato*, and considers the operating costs (Gomes et al., 2015). A request for Gato must be made 1h30 in advance. Due to this lack of flexibility, demand is low and the high cost of operation of the Gato-system makes the DON-PBS financially unsustainable. The authors try to improve the existing Gato-service by simulating an improved version. In this version, a request can be made at least 15 minutes in advance and the ticket price is halved. From the operator perspective, a mini-bus is used instead of a normal-sized one.Due to the increased flexibility the demand rises and operating costs decrease due to the use of another vehicle. However, even though average profit increases 25% compared to the original Gato-service, it remains negative.

Hyland and Mahmassani (2020) adopt a door-to-door zero lead time multi-objective approach. The method minimizes the vehicle travel time and passenger travel time. Constraints are placed on the maximum detour distance. Their algorithm assigns new requests to vehicles in a greedy way, but only considers vehicles which are empty or currently going to a drop-off point. Once a request is assigned, the assignment is not changed anymore. The authors simulate a high-demand system with a fixed fully automated fleet size, but each vehicle only has a capacity of two people. The New York taxi data set, also used by, among others, (Alonso-Mora et al., 2017) is once again chosen as an input. It is found that allowing shared rides significantly improves operation efficiency, even when the maximum detour distance is relatively low. The waiting time, system cost and fleet kilometers decrease, especially when overall demand is high.

Atasoy et al. (2015b), Ikeda et al. (2015) and Atasoy et al. (2015a) investigate the Flexible Mobility On Demand concept with mixed lead times. A request can be for an immediate ride, or can be sent in advance. A user can choose in real-time between a taxi, shared taxi or on-demand fixed line bus, all operated by the same company, but offered at different prices. The (shared-)taxi service is a door-to-door service and is scheduled with the DARP insertion heuristic of Jung et al. (2013). The operator wants to maximize profit, maximize consumer surplus or both. A constraint is set on the passenger travel time and request rejects are possible. The fleet of vehicles is of a fixed size, but can switch roles during the day, which has a positive influence on the objective function value (both profit and consumer surplus). The system is simulated in the Tokyo suburbs (Japan). The authors find that optimizing the operator's profit causes profit to increase with 74% compared to the scenario where they optimize consumer surplus. In the multi-objective scenario this increase is 54%.

Ronald et al. (2015) compare a fully flexible many-to-many DON-PBS to a semi-flexible system in a small network of two adjacent towns. (More on semi-flexible many-to-many DON-PBS in Section 3.1.2). The focus of their work is on the fully flexible scheme. A set of potential bus stops and a real-time scheduling approach are used. Actually, the authors do not consider a multi-objective optimization, but they compare two different objective functions separately: minimize passenger travel time compared to a direct distance travel time and minimize driving distance. It is concluded that the objective chosen is less important. The semi-flexible system where this DON-PBS is compared to, exists. There are three to six services a day, each starting from the same first stop at fixed times. The rest of the route is flexible (but stop-based) and demand-responsive. Customers have to make a reservation at least 10 minutes before the service is scheduled to depart from the first stop (depot) or by arriving at the first stop in person to begin their trip. The authors conclude that this semi-flexible system performs better from the operator's point of view, compared to the fully flexible one. When demand is high, the fully flexible DON-PBS results in extensively high vehicle distance kilometers compared to the semi-flexible system. However, from the passengers' point of view, the fully flexible system is quicker en delivers a more individual experience.

Dessouky et al. (2003) model a paratransit service considering a real-time scheduling heuristic in order to maximize the number of served requests considering economic and environmental costs. The methodology combines routing and scheduling of several fleet composition scenarios. Two types of requests are possible: immediate and in advance. In advance are requests scheduled until the beginning of the shift, as in the DOFF-PBS, and immediate request are included in last-minute schedules. Results are based on paratransit service data for Los Angeles (USA) and the focus of this paper is on the environmental impact.

Lastly, Van Engelen et al. (2018) solve a stop-based DON-PBS-problem **proactively** and develop an online greedy insertion algorithm with demand forecasts and empty vehicle rerouting that can handle relatively big instances. Their algorithm is tested in the port area of Ijmuiden (The Netherlands). The algorithm minimizes the passenger travel time, but also the number of rejected requests and the overall travel time of the vehicles. Each request has a time window and a maximum cost of serving the request is maintained. Insertion happens where the objective function value increases the least. It is concluded that the empty vehicle rerouting and demand forecasts drastically reduce the number of rejected requests, but also decrease the travel time and total distance driven. In addition, after testing different fleet sizes and capacities, it is also concluded that the DON-PBS performs better with a large fleet of small vehicles instead of with a small fleet of large capacity vehicles.

Table 1 gives an overview of the literature optimizing many-to-many fully flexible DON-PBS from a multi-objective perspective. Because both perspectives are included in the objective, typically only a only few extra constraints are present.

3.1.1.4 Conclusion many-to-many fully flexible DON-PBS

Already quite some research is done on many-to-many fully flexible DON-PBS. Although most literature deals with demand in a reactive manner. A proactive approach is only found in Bruni et al. (2014), Horn (2002a,b) and Van Engelen et al. (2018). Also, there are only few papers mentioning they take into account real-time traffic and congestion. For the pickup and drop-off locations, both the door-to-door and the stop-based approach are commonly used. However, a comparison concludes that a stop-based approach can save costs by a lower required fleet size (Viergutz and Schmidt, 2019; Bischoff et al., 2018). Only one paper is found on many-to-many fully flexible DON-PBS that includes bus stop assignment (Melis and Sörensen, 2021), even though this has proven to yield better results in static many-to-many fully flexible DR-PBS (Melis and Sörensen, 2020; Czioska et al., 2019).

Most optimization solution methods used, are based on a greedy insertion, whether or not in combination with an efficient vehicle choosing algorithm. Only few more advanced meta-heuristic approaches are used, but most of these are only tested on low to medium sized instances (local search, (Horn, 2002a), GRASP and local search (Santos and Xavier, 2013; Gomes et al., 2014, 2015), (A)LNS (Vallée et al., 2017; Simonetto et al., 2019; Melis and Sörensen, 2021), TS (Bruni et al., 2014), GA (Kawamura and Mukai, 2009)). In general, the instances used are mostly based on real cities in high demand areas. For these scenarios, it is found that a large fleet size of minibuses works best for a fully flexible many-to-many DON-PBS. The capacity of these minibuses differs among papers, with an average of ten seats per vehicle and a median of eight. More recent literature imagines these minibuses to be electric and autonomous, eliminating the drivers' cost. However, besides from Kawamura and Mukai (2009), authors ignore the need to recharge during the day.

A lot of work is done in testing whether or not fully flexible many-to-many DON-PBS would yield better results compared to conventional, fixed PT systems or compared to private taxis. For the first comparison, results are not entirely unanimous, however benefits for the DON-PBS are always found. Some literature finds the travel time to be halved compared to conventional PT (Navidi et al., 2018; Viergutz and Schmidt, 2019; Bischoff et al., 2018). On the other hand, Jäger et al. (2018) find travel times to increase 17% compared to conventional PT. A reason for this finding might be the extremely high demand in the city of Singapore and the well established, high frequency fixed PT in the city. Most authors find lower waiting times for the DON-PBS (Viergutz and Schmidt, 2019; Bischoff et al., 2018; Liyanage and Dia, 2020). Only Navidi et al. (2018) find waiting times to be higher compared to the conventional PT, however this can be explained by the fact that a door-to-door approach is used instead of a stop-based one. A door-to-door approach causes detours and consequently waiting times to be slightly higher (Viergutz and Schmidt, 2019; Bischoff et al., 2018). Furthermore, vehicle utilization more than doubles in the DON-PBS, but the fleet size necessary to serve the same requests as in the conventional PT increases (Jäger et al., 2018; Liyanage and Dia, 2020). Lastly, energy consumption, CO_2-emissions and pollutants are lower in the DON-PBS compared to conventional PT (Jäger et al., 2018; Viergutz and Schmidt, 2019; Liyanage and Dia, 2020). For the comparison with a private taxi system, it is clear a DON-PBS performs better and that economies of scale have a positive impact on the DON-PBS and not on a private taxi system. When the demand (density) rises, the cost per trip with a DON-PBS decreases and a lower price can be asked. This in turn will invoke an increase in demand, starting a positive spiral (Jokinen et al., 2011; Hyland and Mahmassani, 2020; Horn, 2002a). For both the DON-PBS and the private taxi system travel times rise when demand increases and the fleet size is fixed, however the DON-PBS shows more resiliency (Jokinen et al., 2011). In a high demand area, the total kilometers driven with the DON-PBS are 13-20% lower compared to the kilometers driven with a private taxi fleet (Bischoff et al., 2017; Ma et al., 2013). Also, the fleet size necessary to serve every request with the DON-PBS is a lot lower (Alonso-Mora et al., 2017; Jokinen et al., 2011).

Besides from direct comparisons, some authors analyze among other transportation systems, a fixed PT, DON-PBS and a private taxi system in co-existence. Horn (2002b) simulate several transport systems in co-existence, but instead of focusing on choice behavior the work analyses the LITRES-2 modeling system and focuses on integrating theses different transport systems in multi-modal journeys. Leich and Bischoff (2019) examine a DON-PBS that can be combined with a fixed PT-trip and compare these results with a base case where there is only the fixed PT. It is concluded that the DON-PBS add-on slightly lowers the waiting time and overall travel time. Archetti et al. (2018) simulate a DON-PBS, private car and a fixed PT, but only allow the use of a private car when the other two transport systems are unavailable. Consequently the authors find that most trips are served by the DON-PBS. On the other hand, Narayan et al. (2017) find private cars to have the highest modal shares, followed by the DON-PBS or a private taxi system and lastly the fixed PT has the lowest modal share. Lastly, Atasoy et al. (2015a), Atasoy et al. (2015b) and Ikeda et al. (2015) examine a DON-PBS, private taxi system and an on-demand fixed PT system, where all systems are served by the same vehicle fleet. The modal shares depend on the objective function chosen, the costs of the services and the perceived utility of travel.

Lastly, it can be concluded that allowing real-time requests, instead of imposing all requests to be sent in advance deteriorates the solution quality, independent on the optimization perspective (Gomes et al., 2014; Melis and Sörensen, 2021; Bruni et al., 2014; Horn, 2002a). Of course allowing real-time requests increases the flexibility for passengers and it is found that imposing large lead times increases the cancellation rate which also increases the costs of operation (Gomes et al., 2015).

3.1.2. Semi-flexible routes and/or timetables

In this section, many-to-many bus systems with semi-flexible routes and/or timetables are considered. In these systems, a standard line or timetable is available and deviations are considered dynamically.

3.1.2.1 *Passenger perspective*

Cortés and Jayakrishnan (2002) use a stop-based, mixed lead time approach and include bus stop assignment for the pickup points. In their simulation, a set of hub regions is used and vehicles with fixed capacity are assigned to such a region. Within the hubs the routes are fully flexible, but every vehicle is also assigned to drive to a neighboring hub on a fixed route. The authors allow passengers to make one transfer and minimize the passenger travel time. The system is proposed as a better alternative to the classical two-transfer DR-PBS where the middle part of the route is carried out by a high-speed transport system. This design concept improves the travel time and quality of service. Hickman and Blume (2001) investigate the before-mentioned classical two-transfer DON-PBS with a door-to-door zero lead time approach. A fully flexible many-to-many DON-PBS is integrated with fixed lines and two transfers are allowed. To eliminate some of the different routing alternatives, a geographical circle is drawn around the origin and destination of each request and identify the fixed PT stops lying in these regions. Next, the method checks whether or not there is a fixed line going from the origin region to the destination region within the passengers' requested time window. If this is not the case, the trip is served entirely by the fully flexible DON-PBS, otherwise the integrated trip is proposed to the passenger. The second part is a vehicle trip scheduling procedure, which the authors propose to solve with an existing insertion heuristic or a clustering and column generation technique. Their algorithm is tested in a real-life network of Houston (USA). It is found that time savings are especially present for shorter trips and that the size of the geographical circle has a major influence as the number of possible trips grows substantially when the circles are expanded.

Another semi-flexible many-to-many DON-PBS is presented by Pei et al. (2019c). The authors start from a fixed bus line, and minimize passenger travel time, including waiting time, by deleting unnecessary stops and possibly shortening the line. A simulation study is performed in MATLAB to compare this flexible system with a fully-fixed bus line in the city of Guangzhou (China). The authors conclude that there is a 10% reduction in total travel time by implementing the flexible system. However, this reduction disappears when demand increases up to 40 passengers per hour or more, making the system more attractive in a low-demand area.

Table 2 gives an overview of the many-to-many semi-flexible DON-PBS optimizing from the passengers' perspective. None of the two papers places constraints on the vehicle travel time and both papers minimize the classical passenger travel time objective.

3.1.2.2 *Operator perspective*

As mentioned in paragraph 3.1.1.2, Horn (2002b) simulates several transport modes. One of these is an on-demand timetabled service between stops. The routing is flexible. These timetables can be linked to the stops itself, but also to zones, implicating that the bus can be at a certain time at any stop within this zone. When a passenger wants to go from stop 1 to 2, he only knows his travel will take place between the earliest possible departure of a vehicle at stop 1 and the latest possible arrival of a vehicle at stop 2. Consequently, passengers are given only little assurances regarding their waiting and travel time. As was mentioned, multi-modal journeys are also possible. Scheduling these journeys is explained in Horn (2004). In this case, part of the route can be with fixed PT, while the other part is with a DON-PBS The journey planner uses a search tree to find all possibilities as long as they are feasible and as long as the branch has potential to find better solutions. Then the best journey is proposed to the passenger.

The literature mentioned in this paragraph is already incorporated in Table 1.

3.1.2.3 *Multi-objective*

Pei et al. (2019a) adopt another semi-flexible many-to-many DON-PBS with a stop-based zero lead time approach, but they optimize in a multi-objective matter. The total system income is maximized, which they define as the income minus the operating costs minus the passenger time costs. The authors propose a system with fixed A-level bus stops, that are guaranteed to be served, and a set of B-level bus stops, only served on demand. There is also a constraint on the maximum number of B-level bus stops served between two A-level bus stops, on the maximum ride time and on the maximum detour. Also, each vehicle has a given capacity. The solution methods used are either a TS heuristic or the enumeration method. The first is chosen when the number of B-level bus stops in between two A-level bus stops exceeds 12, otherwise, when using the enumeration method, the computation time increases too much. Their system is tested in an off-peak, low demand setting. It is concluded the income of such a system increases compared to a classical bus line. A similar optimization problem can be found in Quadrifoglio et al. (2007). The problem also has a fixed set of A-level bus stops that form a standard route, however the bus can deviate within a certain deviation area. There are no B-level bus stops. Thus, requests can be door-to-door. The algorithm minimizes a weighted average of the vehicle driven kilometers, the average user ride time and average waiting time per customer by using an insertion heuristic. There are no time window constraints but the route has a certain amount of incorporated slack time to make the deviations. The algorithm is tested in a real-life low-demand line in Los Angeles (USA). The results of their heuristic are compared with CPLEX for the static version of the problem and find only small optimality gaps ranging from 0 to 16%.

Inturri et al. (2018) also start from a fixed route and take into account possible flexible route segments served in case of demand. Their simulation model focuses on the distribution of a given number of vehicles with a given capacity over the fixed and flexible routes.A constraint is placed on the maximum waiting time. If the waiting time of a customer exceeds this threshold, the request is declared "unsatisfied". It is found that many vehicles with a small capacity induce lower passenger travel times, while from the operators perspective a few large buses are more cost-efficient. Also, a clever assignment of vehicles to routes helps to avoid empty kilometers and reduce operating costs.

An overview of the many-to-many semi-flexible DON-PBS optimized with a multi-objective approach can be found in Table 2.

Table 2: An overview of many-to-many semi-flexible DON-PBS

		Constraints on ...					
Reference	Objective	C	F	VTT	PT	DT	PTT
Cortés and Jayakrishnan (2002)	min PTT	x	x		x		
Hickman and Blume (2001)	min PTT	x			x	x	x
Pei et al. (2019c)	min PTT		x				
Pei et al. (2019a)	max income - costs of operation and costs of PTT	x	x	x			
Quadrifoglio et al. (2007)	min VTD and PTT		x				
Inturri et al. (2018)	min costs of operation and PTT	x	x		x		

(C = Capacity, F = Fleet size, VTT = Vehicle travel time, VTD = Vehicle travel distance, PT = Pickup time, DT = Drop-off time, PTT = Passenger travel time)

3.1.2.4 Conclusion many-to-many semi-flexible DON-PBS

Literature on many-to-many semi-flexible DON-PBS is rather scarce and really diverse. On one side there are systems combining fully flexible DON-PBS, mostly at the beginning and ending of the journey, with in between a fixed route segment or fixed PT line. On the other hand, there are systems starting from a standard route with the possibility to deviate from it, using a deviation area, potential (B-level) bus stops or flexible route segments. Two solution methods are mostly used: insertion heuristics (Cortés and Jayakrishnan, 2002; Quadrifoglio et al., 2007) and the enumeration method (Hickman and Blume, 2001; Pei et al., 2019a). In contrast to fully flexible many-to-many DON-PBS, in this section we see more studies performed in a low demand area and in real-life networks. Only Cortés and Jayakrishnan (2002) perform a theoretical study in a high demand context and this work is also the only one including bus stop assignment for the pickup. Another difference with the previous section, is the higher allowance of transfers. Besides from Horn (2002b) none of the literature found includes traffic conditions.

When comparing semi-flexible to fully flexible DON-PBS, the latter of course gives more flexibility to the passengers and results in lower passenger travel times (Ronald et al., 2015). From the operator's perspective, on the one hand a semi-flexible DON-PBS is proposed as a better alternative for low demand areas (Viergutz and Schmidt, 2019), but on the other hand in a high demand area a fully flexible system can cause vehicle kilometers to increase substantially making a semi-flexible service an interesting alternative as well (Ronald et al., 2015). However, the cost of vehicle kilometers decreases so much when using autonomous electrical vehicles, that the last argument can probably be ignored.

3.2. Many-to-one or feeder lines

In this section, many-to-one DON-PBS systems are considered. In such systems requests are made to (or from) the same location.

3.2.1. Fully flexible routes and timetables

We start with discussing the fully flexible feeder lines. In these DON-PBS systems routes are build from scratch, but they all begin or end in the same location.

3.2.1.1 Passenger perspective

Perera et al. (2018a) do a simulation of a stop-based, zero lead time, fully flexible feeder line with an underlying heuristic algorithm, on university grounds. The number of requests is low and they all have a common destination, just outside the university area. In the simulation, the authors primarily focus on scenario generation, the execution of the heuristic algorithm and on the visualization of the solution. A hybrid genetic algorithm is used with local search to minimize passenger travel time, described in Perera et al. (2018b). Constraints are placed on the vehicle capacity, the maximum vehicle travel time and the maximum passenger travel time. It is assumed that the number of vehicles is large enough to serve all passengers. The problem is solved periodically, meaning that each short period of time the static problem is solved in a fast way.

3.2.1.2 Operator perspective

In contrast to the previous two mentioned works, Perera et al. (2017) solve the first mile problem with a fully flexible DON-PBS on a door-to-door basis, optimizing from the operator's perspective. The authors propose a greedy heuristic algorithm, which they compare with CPLEX-results of the static problem. Each request has a pickup time window and all requests have a common destination, the nearest rapid transit node. The fleet size is fixed and homogeneous with fixed capacity, but the number of vehicles is presumed large enough to serve all requests. The objective is to minimize the total vehicle miles traveled. Requests arrive in real-time but are only periodically addressed every 2 minutes. From an instance scale of 132 requests the exact method needs 39 hours to come up with a solution, while the heuristic method only needs 168 milliseconds on average. Performance-wise the heuristic finds solutions with 18% higher objective function values compared to the exact method.

3.2.1.3 Multi-objective

Both Yu et al. (2015) and Li et al. (2018b) tackle the last mile problem with a fully flexible feeder bus, stop-based and with zero lead time. The first solve the problem with an adaptive TS heuristic using different local search strategies and neighborhoods, while minimizing the passenger travel time, including walking time, and operating costs. All demand for a certain destination area in the graph is aggregated and only one bus stop per destination area is chosen. Every local search iteration first determines the set of bus stops to be visited before building a route with the selected bus stops. For relatively small instances with only 20 possible stops, their results are compared with results obtained with the enumeration method and find only a 1% gap in solution quality, while the computation time is drastically lower. It is also concluded that the adaptive search procedure outperforms the non-adaptive one of Lownes and Machemehl (2010). The authors do not go into much detail about how they cope with real-time demand, but rather state that they want to find a heuristic that is able to cope with real-time demand in a fast way. The same statement is true for Li et al. (2018b). A genetic algorithm is developed and the algorithm minimizes both passenger access time to a pickup location and the operation costs. Besides from the routes, the algorithm also optimizes pickup locations (bus stop assignment). When comparing their algorithm with a CPLEX solver, a 10-30% gap is found. Also, their results created with real-time passenger information, are compared to results obtained with historical data and find a saving in operating costs and reduced walking distances for the passengers.

Koh et al. (2018) also adopt a stop-based, zero lead time approach, but instead of optimizing one feeder line, the authors optimize a set of high demand feeder lines. Every passenger has a maximum waiting and travel time and some degree of bus stop assignment is used.No multi-objective optimization is performed but two different objective functions are tried. In a first scenario the objective is to minimize passenger travel time and the fleet size and bus capacity is seen as an input variable. An existing insertion heuristic made for the static multi-vehicle DARP is modified to incorporate dynamic demand.The model is tested in a high demand residential town in Singapore, and wants to replace four fixed feeder lines where transfers are often necessary, with a set of dynamic lines without transfers. The same fleet size and capacity of the vehicles used in the fixed feeder lines, is adopted. In this case a decline in average travel time of 30% is observed and a decrease in waiting time of 5%. In a second scenario, the authors minimize the fleet size and capacity while maintaining the travel time constraints. In this scenario, it is found that the fleet size can be reduced with 12.5% and the capacity can be reduced from 90 to 30 passengers per vehicle. Even though waiting times increase slightly, the average travel times in this scenario are still lower compared to the fixed feeder lines with transfers.

Shen et al. (2017) also use a stop-based approach and allow mixed lead times for a DRC (see Section 2.3).First the pre-booked requests are scheduled (stage 1), then the real-time requests (stage 2), if not rejected due to fixed vehicle capacities. The method minimizes the total travel time of both the vehicles and passengers.The optimization algorithm is based on changing the order of stops in the routes in stage 1. The simulation is done in a district of the city Nanjing (China), picking up passengers near their homes to drop them off at a common metro station. The demand is relatively low. Their results show that the mixed lead time feeder line would be able to serve medium-sized demand with low trip density. A door-to-door approach with mixed lead times is investigated by Wang et al. (2020). The operator's profit is maximized and the passengers' travel time is minimized.Constraints are set on the vehicle capacity and each request has a time window. The static scenario is firstly solved with the requests known beforehand using a genetic algorithm. Afterwards the algorithm starts adding the real-time requests, if not rejected. The research is also done in a low demand area with one transfer station. The authors compare the performance of this mixed lead time on-demand feeder line to a feeder line with a fixed time schedule and find that the total system utility increases 38.1%. The fleet size required is equally large in case of mixed lead times or in a static feeder line system, where all requests are known in advance, while the routes of the vehicles vary.

3.2.1.4 Conclusion many-to-one fully flexible DON-PBS

The fully flexible many-to-one dynamic online systems are summarized in Table 3. Literature on fully flexible feeder line DON-PBS is not that extensive. Authors mostly focus on low demand areas, bringing people from close to their homes to a transport hub. Only Koh et al. (2018) perform an optimization of multiple feeder lines in a high demand area. It is noteworthy that the capacity of the vehicles used is significantly larger compared to the ones used in a many-to-many fully flexible system. On average authors presume 65 seats per vehicle (median is 45). Except for the theoretical study of Wang et al. (2020), all systems are optimized in a real-life based network.

In terms of solution methods more metaheuristic approaches are present such as GAs (Wang et al., 2020; Li et al., 2018b,a; Perera et al., 2018b,a) and adaptive TS (Yu et al., 2015). Most authors also compare their results with the results of an exact solution approach using CPLEX or an enumeration method. This is possible because of the small scale of the instances. The only work found optimizing a large scale feeder line problem uses once again a quick insertion-based heuristic (Koh et al., 2018).

Table 3: An overview of fully flexible many-to-one DON-PBS

		Constraints on ...					
Reference	Objective	C	F	VTT	PT	DT	PTT
Perera et al. (2018b,a)	min PTT	x		x		x	
Perera et al. (2017)	min VTT	x	x		x		
Yu et al. (2015)	min PTT and costs					x	
Li et al. (2018b)	min access time pickup and costs	x		x			
Koh et al. (2018)	min PTT or min C and fleet size	(x)	(x)		x		x
Shen et al. (2017)	min PTT and VTT	x	x	x		x	
Wang et al. (2019)	min PTT and max profit	x	x		x	x	x

(C = Capacity, F = Fleet size, VTT = Vehicle travel time, VTD = Vehicle travel distance, PT = Pickup time, DT = Drop-off time, PTT = Passenger travel time)

3.2.2. Semi-flexible routes and/or timetables

Besides from feeder lines build from scratch, there are also feeder lines starting from a standard route or timetable but deviating from this schedule.

Only one dynamic feeder line optimization using semi-flexible lines is found in Pratelli et al. (2018), considering the passenger perspective. Instead of building routes from scratch, the authors start from a standard route and deviate from it. This means that a part of the stops in the route is fixed, while others are only served while deviating from the route if a request is made. Each route has a minimum and maximum number of deviating stops. The objective function focuses on the passengers' perspective, including the minimization of (1) the extra waiting time for passengers who want to get on the bus after the deviations from the standard route, (2) the time elapsed on board during deviations and (3) the extra time on foot of users who have to walk to a fixed bus stop because their deviation-bus stop is decided to be not served. Each term is multiplied by a weight coefficient because the value of each time component is perceived differently. The fleet size and capacity is fixed. CPLEX is used to solve the problem in an exact manner. The stop-based semi-flexible feeder line system, with different numbers of deviation stops, is compared to an existing fully flexible door-to-door service in Florence (Italy). In the existing system most demand is in advance, but real-time demand is dealt with as well. It is found that waiting times are lower in the semi-flexible system, but on the other hand there is a larger share of lost demand.

The semi-flexible many-to-one dynamic online systems are summarized in Table 4.

Table 4: An overview of semi-flexible many-to-one DON-PBS

		Constraints on ...					
Reference	Objective	C	F	VTT	PT	DT	PTT
Pratelli et al. (2018)	min PTT	x	x				

(C = Capacity, F = Fleet size, VTT = Vehicle travel time, VTD = Vehicle travel distance, PT = Pickup time, DT = Drop-off time, PTT = Passenger travel time)

4. Dynamic offline demand-responsive public bus systems

Dynamic offline demand responsive public bus systems (DOFF-PBS) only allow changes to an already defined schedule for some of the services, as long as these changing services have not started yet. Once a certain service of the system has started, no more changes to this service are allowed. These changes can be motivated by incoming data from passengers or (un)expected circumstances. This strategy is only possible if the solution method for the optimization problem is fast enough to optimize (part of) the planning again, based on new last minute data. The same structure is used as in the previous section to classify the DOFF-PBS.

4.1. Many-to-many

In this section, systems of the type many-to-many are considered, where passengers can travel from any bus stop or origin location to any other bus stop or destination. Later, in Section 4.2, systems of the type many-to-one are discussed.

4.1.1. Fully flexible routes and timetables

First, the fully flexible systems are discussed, i.e., systems where no standard route or timetable is available. Semi-flexible systems, which start from a standard route or timetable, are discussed in Section 4.1.2.

4.1.1.1 Passengers' perspective

Optimizing fully flexible DOFF-PBS of the many-to-many type from the passengers' perspective is essential in making them competitive with private transport. Hadas and Ceder (2008) deal with a request-based bus system, where passengers state their desired pickup and drop-off locations. The system works with predefined bus stops and takes transfers into consideration. Transfers to other bus lines do not take place at designated bus stops, but rather at any point along a road segment. The objective of this bus system is to minimize the total travel time of the passengers, by maximizing the encounter probability of buses at a transfer point. The travel time of the buses and their encounter probability are estimated based on real-time information. Based on these estimations, the route and deployment tactics are updated for the next group of operating buses. To solve this problem, a multi-agent system approach is used. This multi-agent system incorporates simulation based models that describe the encounter probability of the buses, and an optimization model based on distributed dynamic programming. A simulation model is employed for the validation and evaluation of the optimization process. An instance with 14 bus stops and three transfer areas is presented. This results in an improved system, with a reduction of around 10% of the average travel time and an increase of around 300% of direct transfers. In Fatnassi et al. (2015), shared goods as well as passengers are transported on a stop-based, on-demand bus system. The bus system periodically receives requests, which consist of pickup and delivery locations for either a single passenger, a groups of passengers, or a set of parcels of goods. In each period, the planning is determined and the requests are assumed to be known before the service starts. The goal is to minimize the waiting time of requests as well as empty vehicle moves. This system makes use of electric vehicles with limited battery capacity, i.e. the range of the vehicles is constrained and the vehicles need to be recharged. The system is optimized with a simulation based approach, where both reactive and proactive strategies are used to predict the demand. The assignment of vehicles to passengers or goods is optimized with mathematical programming. The model is tested on an instance based on the passenger rapid transit network in Corby town in Northampton (UK). Demand is modeled as a Poisson process during the morning peak hours, i.e., 1900 passenger requests over the span of two hours. The experiments show that the proposed solution approach provides good results in terms of solution quality and resolution time.

Masoud et al. (2017) present a stop-based ride-sharing service that, besides private cars, also considers timetabled transport services, such as local buses or trains. The transit services can enter the ride-sharing model as inflexible drivers. Masoud et al. (2017) minimize the total travel cost of the passengers. There are vehicle capacity constraints and a maximum number of allowed transfers. The system is subject to time window constraints for passenger journeys. A dynamic programming model is used to solve the multi-hop matching problem between passengers to drivers. The algorithm was previously developed in Masoud and Jayakrishnan (2017). The system in Masoud et al. (2017) is tested on the red metro line in Los Angeles (USA) during peak demand.

4.1.1.2 Operators' perspective

Bus systems can also be optimized from the operator's perspective. In these cases, the objective is to make such systems as profitable or as least costly as possible for the operator. Zhao et al. (2018) study a ride-sharing system that picks up and drops off passengers at their desired location. However, the passengers can be reassigned to a different pickup and/or delivery location to increase the efficiency of the system. The pickup location and time, as well as the drop-off location and time are stated by the passengers as a request for transportation. The requests are received in real-time, before the vehicles start operating, in order to optimize the vehicle routes and dispatching plans. The goal of this ride-sharing service is to minimize the operational costs, which consist of fixed vehicle dispatching costs, vehicle routing costs and inconvenience costs for picking up or dropping off a passenger. The bus system is subject to vehicle capacity constraints, fleet size constraints, and time window constraints. The solution approach is based on Lagrangian relaxation. The system is tested on three theoretical grids, with 100, 400 and 900 nodes respectively, and varying demand levels with 20 to 40 passenger requests. In Rigas et al. (2018), a fleet of electrical vehicles transports passengers from origin bus stops to destination bus stops. These pickup and drop-off locations are stated by the passengers, as well as their desired departure times. Most requests are assumed to be known beforehand and requests in the future are predicted using an algorithm in order to increase the vehicle utilisation. However, once a vehicle task is planned it cannot be changed. The goal of this system is to maximize the number of accepted requests in order to increase the operators' profit. The electric vehicles have a limited range and their batteries need to be swapped or charged at charging stations. This system is optimized with an incremental mixed integer programming (MIP) algorithm for medium to large problems, with less than 1000 requests, and a greedy heuristic algorithm for very large problems with up to 3000 requests. Furthermore, a tabu search algorithm is implemented as well, in order to further improve the quality of the solutions. Real-world data of locations of shared vehicle pickup and drop-off stations (Washington DC (USA)), with high demand levels, is used to test this system. Lotfi et al. (2019) study a door-to-door ride-sharing system with transfer services. Passengers need to make a request for transportation between their pickup and drop-off locations, their earliest departure time, and their latest arrival time. Furthermore, passengers state a desired maximum trip duration, and their willingness to pay a fare for the ride-sharing and/or transfer services. The system's objective is to maximize the utilization of the vehicles and minimize the vehicles' route distances in order to maximize the operators' profit. The system is subject to restrictions, such as the vehicle capacity, the fleet size and a maximum passenger ride time. The system is optimized with a metaheuristic, a modified version of a column generation algorithm. The system is tested on a real-life, moderate-sized, network in Dallas (USA), with up to 70 passenger requests.

In Huang et al. (2020a), the optimization problem of the demand-responsive Customized Bus (see Section 2.3) is studied. This is a door-to-door service, where passengers can make requests for transportation until one hour before operation. The solution method consists of two phases: a dynamic phase and a static phase. In the dynamic phase, new incoming passenger requests are dynamically inserted into an existing solution in an interactive manner. This is done by a branch-and-bound algorithm with a "cluster-first-route-second" scheme to reduce the search space and the runtime. The objective of the dynamic phase is to maximize the operators' profit. In the static phase, the service network is optimized statically based on the overall demand and the objective is to minimize the operating costs. The system is subject to vehicle capacity constraints and to time window constraints. The system is tested on the Sioux Falls network (Bar-Gera, 2001), with up to 30 passenger requests. The system is also used on a real-life case study in Nanjing (China), with up to 200 passenger requests.

4.1.1.3 Multi-objective

Bus systems can also be optimized from both the passengers' and the operators' perspective. Winter et al. (2016) study a stop-based system of shareable rides in automated vehicles. Passengers make requests stating their pickup and drop-off location, as well as their desired departure time. Requests are received in a real-time manner, however once a vehicle is dispatched, its route and timetables cannot be changed. The objective is to minimize the operational and travel costs. The system is optimized through an event-based simulation model with an integrated iterative assignment procedure. The system was implemented in a pilot project in the Netherlands, with different demand patterns, vehicle capacities and other operational factors. Demand levels are high: between 1740 and 1953 requests in a working day. In Amirgholy and Gonzales (2016), a door-to-door bus system is studied with the goal of finding a management strategy that stimulates users to adapt their requests to be more uniform over time. Analytical expressions for the expected operation costs and the passenger costs are provided. Different cases, where different variables are constrained, are studied to obtain the best management strategy. The optimal solutions are given as closed-form expressions, with the help of the analytical objective functions. Numerical examples with up to 150 passenger requests are used to test the system. Guo et al. (2018) deal with a Customized Bus design problem that aims to determine which optional bus stops to visit and in which sequence, for each bus, based on real-time passenger requests. The objective is to minimize both the user cost and the operation cost. The optimization problem is modeled as a multi-vehicle routing problem and it is solved with both a branch-and-cut algorithm as well as a genetic algorithm. The system is tested on an instance with 620 passenger requests within one hour. The instance is created from real life historical smart card data from Beijing, China. Ji-Yang et al. (2020) study a stop-based bus system with passenger requests. The passengers state their desired origin and destination stops, together with their desired arrival time at their destination. The goal is to minimize vehicle operating times, passenger waiting times before boarding, and the difference in actual arrival time and desired arrival time. The operation of this bus system is optimized through the use of a constructive heuristic algorithm. The heuristic is tested on a small theoretical network, with 18 bus stops and 90 passenger requests.

4.1.1.4 Conclusion on fully flexible many-to-many DOFF-PBS

A substantial number of studies is dedicated to fully flexible many-to-many DOFF-PBS. Actually, most DOFF-PBS studied in the literature belong to this category. Both door-to-door and stop-based systems are used, however, the majority of these systems work with bus stops. Most studies optimize the system taking both the needs of the operators and the passengers into account. In this case, the fleet size, passenger travel times and vehicle operating costs are often optimized. With the exception of Amirgholy and Gonzales (2016), the vehicle capacity is limited in the multi-objective DOFF-PBS. The fleet size is often a constraint as well in these systems. Furthermore, Winter et al. (2016) limit the vehicle travel times and Ji-Yang et al. (2020) have restrictions on the drop-off times of the passengers.

When the system is optimized from the operators' perspective, the fleet size is often minimized in order to decrease operational costs. These studies also aim to satisfy as many passenger requests as possible to increase the operator's revenue. All of these systems take limited vehicle capacities into account. Outside of Huang et al. (2020a), the fleet size of these systems is also limited. The passengers' service quality is taken into account by imposing restrictions on the pickup and drop-off times of the passengers. Lotfi et al. (2019) limit the allowable passenger travel time as well.

Not many authors optimize the system solely from the passengers' perspective. Hadas and Ceder (2008) minimize the total travel time of the passengers, while Fatnassi et al. (2015) aim to minimizing waiting times of requests by reducing empty moves of the vehicles. Hadas and Ceder (2008) take the operators' needs into account by limiting the fleet size, while Fatnassi et al. (2015) and Masoud et al. (2017) limit the vehicle mileage and capacity. Masoud et al. (2017) consider different modes of transportation and have restrictions on the pickup and drop-off times of the passengers.

The majority of the studies about fully flexible many-to-many DOFF-PBS use heuristics to optimize the system. Most often, these studies optimize their systems on large, realistic instances based on real networks and demand levels. Amirgholy and Gonzales (2016) construct closed-form mathematical expressions for the objective functions. Winter et al. (2016) use primarily simulation models to determine the best planning for their system. Some studies, such as Huang et al. (2020a), use exact methods like mathematical programming and branch-and-price algorithms. However, these solution methods are tested on relatively small networks and/or low demand levels. Masoud et al. (2017) make use of dynamic programming. An overview of the fully flexible many-to-many DOFF-PBS is given in Table 5.

Table 5: An overview of many-to-many fully flexible DOFF-PBS

		Constraints on ...						
Reference	Objective	C	F	VTT	VTD	PT	DT	PTT
Hadas and Ceder (2008)	min PTT		x					x
Fatnassi et al. (2015)	min PWT	x			x			
Masoud et al. (2017)	max PTC	x				x	x	
Zhao et al. (2018)	min operational costs	x	x			x	x	
Rigas et al. (2018)	max P		x		x	x	x	
Lotfi et al. (2019)	max P	x	x			x	x	x
Huang et al. (2020a)	max operators profit	x				x	x	
Winter et al. (2016)	min PTT + F	x	x	x				
Amirgholy and Gonzales (2016)	min PTT + F + VTT		x					
Guo et al. (2018)	min VTD + PTT + PWT	x	x		x			
Ji-Yang et al. (2020)	min VTT + PWT + arrival delay	x	x				x	

(P = The number of passengers served, C = Vehicle capacity, F = Fleet size, VTT = Vehicle travel time, VTD = Vehicle travel distance, PT = Pickup time, DT = Drop-off time, PTT = Passenger travel time, PWT = Passenger waiting time, PTC= Passenger transportation costs)

4.1.2. Semi-flexible routes and/or timetables

In this section, bus systems of the many-to-many type with semi-flexible routes and/or timetables are discussed. This means that standard routes and/or timetables are determined beforehand, but these are modified dynamically.

4.1.2.1 Passengers' perspective

Horn (2004) studies a system that combines different modes of transportation to construct a sequence of journey-legs. The system works with requests, in which passengers state their origin and destination, together with their desired arrival and departure times. These requests are assumed to be received in advance at any time before the operation starts. New incoming requests allow the system to adjust its current planning, however, once the vehicles are in operation, their planning cannot be changed. Passengers are encouraged to combine walking, "smart shuttle" buses, "roving" buses, taxis and fixed transportation services to get to their destination. Smart shuttles offer a fixed timetabled service with flexible routing, which is a form of semi-flexible transport. Roving buses are a fully flexible service. The taxis provide a door-to-door service, while the other modes provide a stop-based service. The objective is to minimize the walking costs, the waiting costs in transfers and at the origin locations, and the travel costs of the passengers. The system is optimized by a journey planning procedure based on a branch-and-bound algorithm. To reduce the run time, specialized bounding and reduction techniques are applied. A simulation study of transport options for the Gold Coast, an urban region in Queensland (Australia), is used to test the system. Qiu et al. (2014)

study a many-to-many MAST system (see Section 2.3), in which passengers make curb-to-curb requests that may or may not be accepted. Accepted curb-to-curb stops are labeled as "temporary stations", which can be utilised by rejected requests for their pickup or drop-off locations. The objective is to minimize a weighted sum of the walking time, the waiting time and the on-board travel time of the passengers. Buses follow a predefined route, from which they can deviate in order to serve nearby requests. Buses are not allowed to deviate more than a certain distance from the mandatory route and there is a preset schedule at the mandatory bus stops. Requests arrive in real-time, before the buses start operating, and their route is optimized with an insertion heuristic. The system is tested on a real-life network in Los Angeles (USA), with up to 70 passenger requests per hour.

Sheu (2002) implements a real-time demand-responsive bus dispatching control methodology for a DOFF-PBS. The dispatching system considers a short-term forecasting of demand to initially plan the routes and service frequency, and a fuzzy clustering to identify variance in passengers demand and identify the best suitable service strategy for each service dispatched. With the gathered information, it is possible to determine the most suitable service among four options: regular, express, short-turn, or zonal services. While the regular service serves all the stops in a line, express only stops at a few stops. Short-turn only serves part of the route and a zonal service skips part of the route in between origin and destination. The passenger waiting time is used to select the most suitable service for the fuzzy clustering algorithm. Synthetic data of a bus line in Taipei (China) is used to assess the performance of the proposed model.

4.1.2.2 Operators' perspective

Crainic et al. (2005) propose a stop-based, semi-flexible transportation system. In this system, flex routes are created for passengers that make a request for transportation, while the buses still serve mandatory stops at fixed schedules. The objective of this system is to maximize the profit of the operator. The profit is defined as the difference between passenger fares, and the cost of operation. Requests are received in a real-time manner and can be rejected if they make the tour infeasible or not profitable. Once the vehicles start operating, the planning cannot be changed. The constraints of this problem are time window constraints, defined by the desired pickup and drop-off times of the passengers, and the fixed schedules at the mandatory stops. The optimization model of this system is solved with three different heuristics: a memory-enhanced greedy randomized algorithm, a multi-start constructive algorithm, and a tabu search algorithm. The algorithms are tested with several numerical examples with varying demand levels. Similarly to Crainic et al. (2005), Pei et al. (2019b) also study a system with mandatory and optional stops. However, in Pei et al. (2019b), the optional stops are only visited if there is a minimum willingness to pay for an extra fare. Passengers make requests for transportation with their origin and destination bus stop and their willingness to pay a certain transportation fare. The goal of the service is to maximize the profit, defined as the revenue from the fares minus the operational costs. A tabu search algorithm is employed to determine the routing of the vehicles and to decide which optional stops are visited. The system is tested on a small numerical example, as well as on a large network based on a real-life bus line in Guangzhou (China).

4.1.2.3 Multi-objective

Fu et al. (2003) study a stop-based semi-flexible bus system, in which vehicles operate in pairs. The lead vehicle provides an all-stop local service, while the following vehicle is allowed to skip some stops depending on the demand for transportation. The schedule and route of the pair of vehicles are determined right before the first vehicle is dispatched, after which the planning cannot be changed. The goal is to minimize a weighted sum of the on-board travel time of all passengers, the waiting time of all passengers and the vehicle travel time. The bus system is modeled as a non-linear programming problem (NLP). Since only one vehicle is considered for each optimization cycle, the NLP is solved using complete enumeration. Fu et al. (2003) test the system on a real-life network in Waterloo (Canada), with high demand levels between 700 and 1400 passengers per hour. Quadrifoglio et al. (2008) study a many-to-many MAST system, in which one vehicle visits a set of mandatory stops a certain number of times. Vehicles are allowed to deviate from the main route within a given maximum distance in order to serve door-to-door requests nearby. Passengers make a request, before the vehicle starts its operation, by stating their pickup and drop-off location, together with their ready time for pickup. The departure times of the vehicle at the stops in the main route are predefined and fixed. The objective is to minimize a weighted sum of three different factors: the total distance driven by the vehicle, the total ride time of all customers and the total waiting time of all customers. This system is modelled as an MIP and solved with a branch-and-cut algorithm. The system is tested on a network based on a real-life network in Los Angeles (USA) with up to 17 passenger requests.

Crainic et al. (2012) examine a stop-based semi-flexible S-PBS. The routes include compulsory stops and efficiently select additional stops from a set of predefined locations. The goal is to provide sufficient time between mandatory stops to serve all requests that arrive, but with the shortest travel time possible. Since this system is a mix of traditional transit and a demand responsive system, its scheduling encompasses two planning processes. Initially, a master scheduling partially defines routes and time windows. At operation time, the actual schedule of each service is built to include optional stops. Gkiotsalitis et al. (2019) propose a rule-based method to generate line alternatives for a given service, such as short-turning and merging lines. These flexibility options are efficient to adapt lines to demand variation. The objective function is to reduce passengers' waiting time and operational costs using a demand estimation for the next 6 hours. The problem is solved with a genetic algorithm using General Transit Feed Specifications for a bus network in The Hague (The Netherlands).

Kim and Schonfeld (2015) consider a bus system that integrates both a fixed route and flexible-route service in multiple dissimilar regions and periods. Both services operated with a fixed timetable, which means that the service with flexible routes is a semi-flexible service. A region is subdivided into zones, and each zone is dedicated to either a fixed route service or a flexible route service. In each period, requests are received dynamically until the buses are dispatched. The flexible route service is a door-to-door service, while the fixed route service works with bus stops. The "system welfare" is maximized with elastic demand relations. The system is subject to vehicle capacity constraints, upper and lower bounds for the fleet size, a maximum headway for the buses, and the fixed schedules of the fixed route service. The system is optimized with a genetic algorithm and applied on a theoretical case study with up to 90 passenger requests per hour per mile^2. Chen and Nie (2017) also study a transportation system where a semi-flexible service is operated in parallel with a fixed service. In this system, both services operate in the same area. The demand responsive service can deviate from its route to provide a door-to-door service. The system makes use of so-called "walking zones". These zones are used to limit the distance each passenger is allowed to walk to a bus stop and determine which passengers are eligible to use the demand-responsive service. Furthermore, transfers are only allowed on the fixed route service. Passengers make requests stating their origin and destination, together with their desired departure time. The objective is to minimize both the operator cost and the user cost. A mathematical model is formulated as a MIP problem and it is solved using a genetic algorithm and a continuous relaxation method. The solution methods are commercially available as MATLAB functions. The system is tested with numerical examples of a small and a mid-size city in the USA, with a diameter of 10km and 20km respectively, and varying demand levels up to 90 passengers per hour per km^2.

4.1.2.4 Conclusion on semi-flexible many-to-many DOFF-PBS

The number of papers on semi-flexible many-to-many DOFF-PBS is comparable to the number of papers on fully flexible systems. None of these systems offer a pure door-to-door service: each system works either solely with bus stops or it offers a door-to-door service alongside a stop-based service. Quadrifoglio et al. (2008), Kim and Schonfeld (2015) and Chen and Nie (2017) work both with bus stops and door-to-door requests. Semi-flexible many-to-many DOFF-PBS are more focused on satisfying both the passengers' and the operators' needs, with the majority of these studies optimizing both perspectives. Fu et al. (2003) and Quadrifoglio et al. (2008) minimize the waiting and travel time of the passengers, together with the travel time or travel distance of the vehicles. Crainic et al. (2012) maximizes the number of accepted requests. Kim and Schonfeld (2015) and Chen and Nie (2017) minimize the user ride times of the passengers, together with either the overall operational costs of the operator or the operators' profit. Kim and Schonfeld (2015) consider the demand elasticity of the operators' revenue, while Chen and Nie (2017) only consider the operational costs. Kim and Schonfeld (2015) restrict the vehicle capacity and the fleet size, and have constraints related to the pick-off and drop-off times. Chen and Nie (2017) only consider restrictions related the pickup time, but take a maximum passenger walking time into account.

Horn (2004) and Qiu et al. (2014) minimize the waiting and walking time of the passengers, as well as their in-vehicle travel times. All of the systems focusing on the passengers' needs have constraints related to pickup and drop-off times. Horn (2004) limit the vehicle capacity, and the fleet size is only considered in Horn (2004). Qiu et al. (2014) limit the vehicle travel time.

Crainic et al. (2005) and Pei et al. (2019b) optimize their systems from the operators' perspective. Both papers aim to maximize the operators' profit. On one hand, Pei et al. (2019b) take the vehicle capacity and mileage into consideration. On the other hand, Crainic et al. (2005) impose pickup and drop-off restrictions to their system.

Heuristics are most often used to solve the optimization problem of the semi-flexible many-to-many DOFF-PBS. However, Fu et al. (2003) use complete enumeration, Horn (2004) use a branch-and-bound algorithm and Quadrifoglio et al. (2008) use a branch-and-cut algorithm. Most often, the semi-flexible many-to-many DOFF-PBS are optimized on large, realistic instances based on real networks and demand levels. An overview of the semi-flexible many-to-many DOFF-PBS is given in Table 6.

Table 6: An overview of many-to-many semi-flexible DOFF-PBS

		Constraints on ...					
Reference	Objective	C	F	VTT	PT	DT	PWD
Horn (2004)	min PTT + PWT + PWL	x	x		x	x	
Qiu et al. (2014)	min PTT + PWT + PWL			x			
Crainic et al. (2005)	max operators' profit				x	x	
Pei et al. (2019b)	max operators' profit	x		x			
Fu et al. (2003)	min PWT+PTT+VTT		x	x			
Quadrifoglio et al. (2008)	min PWT+PTT+VTD			x	x		
Crainic et al. (2012)	max P	x	x	x	x	x	
Gkiotsalitis et al. (2019)	min PTT + operation costs	x					
Kim and Schonfeld (2015)	max CS + PS	x	x		x	x	
Chen and Nie (2017)	min CS + Operation costs				x		x

(P = The number of passengers served, C = Capacity, F = Fleet size, VTT = Vehicle travel time, VTD = Vehicle travel distance, PT = Pickup time, DT = Drop-off time, PTT = Passenger travel time, PWT = Passenger waiting time, PWL= Passenger walking time, PWD = Passenger walking distance, PTC= Passenger travel cost, CS = Consumer surplus, PS = Producer surplus)

4.2. Many-to-one

In this section, transport systems of the type *many-to-one* and *one-to-many* are discussed. In these services, passengers with different origins or destinations are either dropped off or picked up at the same location. These systems are commonly known as feeder systems.

4.2.1. Fully flexible routes and timetables

First, transportation systems with fully flexible timetables and routes are discussed. Semi-flexible systems are discussed in Section 4.2.2.

4.2.1.1 Passengers' perspective

Li and Quadrifoglio (2010) present analytical and simulation models to help service providers choose between a fixed feeder service and a DRC (see Section 2.3), depending on operational circumstances. The objective is to minimize the travel time, walking time and waiting time of the passengers. The DRC provides a door-to-terminal service for passengers that make a request for transportation. Passengers are able to notify their presence by means of a phone or internet booking service. Immediately before the beginning of each trip, waiting customers are scheduled and the route for the trip in the service area is constructed. There is no planned idle time in between trips. The routes of the buses and the passenger schedules are constructed with an insertion heuristic. Li and Quadrifoglio (2010) give numerical examples with varying demand levels between 24 and 48 passengers/mile^2/h. Sun et al. (2018a) study a stop-based feeder system that takes transfers between the feeder bus and a shuttle with fixed timetables into account. Passengers make transportation requests in real-time through a cell phone app. The app considers real-time traffic conditions to provide an optimal plan for the passengers. However, no modifications are possible to buses that are already dispatched. The objective is to minimize the total travel time for all passengers, i.e., the in-vehicle time and the transfer waiting time. The optimization problem of this feeder service is divided into sub-problems and solved with a genetic algorithm. The system is tested on an instance with a real distribution of passenger demand, aggregated from cellular data of a network in Nanjing (China), with 30 passenger requests within half an hour.

Sun et al. (2019a) present a feeder system that feeds passengers to a single destination from several depots in different locations. The authors introduce "pickup locations", where some passengers can be picked up collectively, while other passengers are picked up at their doorstep. The objective is a weighted sum of the passenger walking time to the "pickup locations" and the in-vehicle ride time. The optimization problem is solved in two stages. In stage 1, a genetic algorithm is used to select passengers and allocate them to vehicles to be picked up at their doorstep. In stage 2, the passengers that are not served yet are allocated to the "pickup locations", based on a greedy algorithm. The system is tested on a real life network in Chongqing (China), with 40 passenger requests within 40 minutes. Unlike Sun et al. (2019a), Wei et al. (2020) study a stop-based feeder system, without a door-to-door service. This system assigns passengers to a bus stop based on their real-time locations during peak hours. The goal is to minimize passenger walking times, passenger in-vehicle ride times, and passenger waiting times at the pickup locations. A genetic algorithm is used to assign demand points to buses and routes, and to determine the frequency of departure. A greedy algorithm, based on Dijkstra's algorithm, is employed to find the optimal walking time and riding time of the passengers. This system is tested on a test instance in Chongqing (China) with 779 passenger requests within one hour.

4.2.1.2 Operators' perspective

Marković et al. (2019) study a one-to-many transportation system that transports passengers from a train station to the doorstep of their home. In this system, available vehicles are only dispatched when the number of boarded passengers reaches or exceeds a certain threshold. The objective of this system is to maximize the system's profit, i.e., the revenue from fares minus the operational costs. A simulation model is employed to sample the passengers' arrivals at the terminal. An insertion algorithm provides an initial solution for the underlying traveling salesperson problem, thereafter a 2-opt procedure improves the solution. The system is tested on numerical examples, which include realistic cost estimates for Washington metropolitan area (USA), with demand levels up to 120 passengers in a time-span of four hours.

4.2.1.3 Multi-objective

In Sun et al. (2018b), a flexible stop-based feeder system is presented. The objective of this system is to minimize the operational costs related to the vehicle mileage and to minimize passenger ride times. The system is optimized with an improved version of the bat algorithm, a population-based heuristic. The system is tested on an instance with a demand level of 42 passengers in a two hour time-span. The network is based on a real life network in Nanjing (China). Quadrifoglio and Li (2009) present an analytical model to help service providers choose between a fixed feeder service and a DRC. They consider two scenarios: each service is operated in a different region and both services operate in the same region simultaneously. The objective is to minimize the operational costs of the vehicles, together with the travel time, walking time and waiting time of the passengers. Passengers using the DRC are able to notify their presence by means of a phone or internet booking service. Immediately before the beginning of each trip, waiting customers are scheduled and the route for the trip in the service area is constructed. The routes of the buses and the passenger schedules are constructed with an insertion heuristic. The system is tested on numerical examples with varying demand levels with up to 90 passengers/mile^2/h.

Dou and Meng (2019) present a stop-based feeder system that focuses on transfers between the feeder buses and the trains at the terminal. The planning is optimized for a short planning horizon, e.g., 30 minutes. The goal is to minimize transfer waiting times, as well as failure costs of non-served passengers and operational costs. The system is modeled as a non-linear integer problem. The model is solved with both a hybrid artificial bee colony algorithm, and with a brute search algorithm. The system is then tested on a theoretical network, with 300 passenger requests within 30 minutes. Liu et al. (2019b) study a stop-based feeder system that operates in a bi-modal context with bike sharing. The system has fixed routes and timetables during the morning rush hour, but switches to a demand-responsive system with dynamic frequencies during the evening hour. The passengers using the demand-responsive system need to state their estimated arrival time at their closest bus stop. The objective is to minimize the average passenger waiting time and maximize the operators profit. The system is optimized with both a genetic algorithm and a particle swarm algorithm. A feeder bus planning problem of a metro station and three nearby communities in Chengdu (China) are used to test the system. In Lee et al. (2019), a door-to-door feeder system that works with small autonomous vehicles is presented. In this system, the feeder buses can be relocated to serve multiple lines that feed different train stations. The goal of this service is to minimize the total distance traveled by each vehicle, as well as the in-vehicle travel time of the passengers. The operation of the feeder system is optimized with the use of a Simulated Annealing algorithm. The system is tested on a hypothetical network with four train stations and 317 passenger requests. Huang et al. (2020b) study a door-to-door feeder service. The objective function is a weighted sum of operating costs and passenger costs. The authors provide closed-form analytical expressions for the cost function that makes it possible to find the optimal scheduling of the vehicles. The routing is optimized with a nearest neighbor strategy. The system is tested on a hypothetical area with different demand densities up to 150 passengers per hour per km^2.

4.2.1.4 Conclusion on fully flexible many-to-one DOFF-PBS

There are several fully flexible many-to-one DOFF-PBS in the state-of-the-art. About half of these systems work with bus stops, while the other half works with a door-to-door service. Sun et al. (2019a) work both with bus stops and door-to-door requests. The optimization models of Sun et al. (2019a), Wei et al. (2020) take bus stop assignment into consideration. Most fully flexible many-to-one DOFF-PBS optimize their system from both the passengers' perspective as well as from the operators' perspective. All the multi-objective DOFF-PBS minimize the operational costs of the operator. Dou and Meng (2019) minimizes also transfer waiting times and aim to satisfy as many requests as possible. The other multi-objective DOFF-PBS also minimize the passengers' travel times. Lee et al. (2019) include the vehicle mileage in their objective function as well as the passengers in-vehicle travel time, and also limit the fleet size. Liu et al. (2019b) minimizes the passenger waiting time as well as the operators' profit. Huang et al. (2020b) also take the passengers' waiting time into account and Huang et al. (2020b) penalizes schedule deviations. With the exception of Dou and Meng (2019), all of the multi-objective DOFF-PBS work with a limited vehicle capacity. The fleet size is also always considered. Sun et al. (2018b) impose constraints related to the travel time of the passengers and to the pickup times. Liu et al. (2019b), Lee et al. (2019) and Sun et al. (2018b) limit the travel time of their passengers as well.

Only Marković et al. (2019) optimize their system solely from the operators' perspective, by maximizing the operators' profit. They take limited vehicle capacities and fleet sizes into account. More authors optimize the system from the passengers' perspective. The passenger travel time is always minimized in these systems. Sun et al. (2019a) also minimizes the passenger walking time, and Wei et al. (2020) minimizes the waiting and walking time of the passengers as well. The vehicle capacity and vehicle mileage is always restricted. Sun et al. (2018a) and Sun et al. (2019a) both have restrictions on the pickup time of the passengers. All are optimized with the use of heuristics and most are tested on realistic networks, with moderate to high demand levels. An overview of the many-to-one fully flexible DOFF-PBS is given in Table 7.

Table 7: An overview of many-to-one fully flexible DOFF-PBS

		Constraints on ...						
Reference	Objective	C	F	VTD	PT	DT	PTT	PWT
Li and Quadrifoglio (2010)	min PTT + PWT + PWL				x	x		
Sun et al. (2018a)	min PTT	x		x	x			
Sun et al. (2019a)	min PTT + PWL	x		x	x			
Wei et al. (2020)	min PTT + PWT + PWL	x		x				
Marković et al. (2019)	max operators' profit	x	x					
Quadrifoglio and Li (2009)	min OC + PTT + PWT + PWL				x	x		
Sun et al. (2018b)	min PTT + OC	x	x		x		x	
Dou and Meng (2019)	max P + OC + min transfer waiting times		x					
Liu et al. (2019b)	max operators' profit + min PWT	x	x				x	
Lee et al. (2019)	min VTD+ PTT	x	x				x	
Huang et al. (2020b)	min PTT + PWT + OC + schedule deviations	x	x					

(P = The number of passengers served, C = Capacity, F = Fleet size, VTT = Vehicle travel time, VTD = Vehicle travel distance, PT = Pickup time, DT = Drop-off time, PTT = Passenger travel time, PWT = Passenger waiting time, PWL = Passenger walking time, OC= Operational costs)

4.2.2. Semi-flexible routes and/or timetables

In this section, DOFF-PBS feeder systems with semi-flexible timetables or routes are discussed.

4.2.2.1 Passengers' perspective

The work presented in Kim and Schonfeld (2014) combines coordinated transfers with the integration of conventional and flexible feeder systems. The system is designed to be optimized during a short planning horizon, typically 30 minutes. The system considers several regions where passengers are picked up from, and a single terminal. In each region, either a conventional feeder system or a semi-flexible system operates. The semi-flexible system offers a door-to-door service and has a preset schedule to make timed transfers, however, it works with a flex route policy to pick up passengers. The model minimizes the transfer cost of the passengersThe vehicle capacity, the headway and the fleet size of this system are optimized with a probabilistic optimization model. The model is solved with a genetic algorithm. Numerical examples, with different stochastically sampled demand densities of up to 13 trips per minute, are used to test the system. Qiu et al. (2015b) study a feeder system that is already implemented in a suburban area of Zhengzhou City (China). The feeder system has two operation policies and they propose a third one. In the flag-stop policy, the buses pick up or drop off passengers upon request alongside a predefined road. In the fixed route policy, the buses can deviate from the fixed route to offer a door-to-door service to passengers. However, passengers can be rejected. In the new policy, the dynamic-station policy, passengers that are rejected by the deviation service can make use of accepted curb-to-curb stops for their pickup and drop-off. The goal is to minimize the expected in-vehicle travel time, walking time and waiting time of the passengers. The system is optimized with a simulation model with demand levels of up to 50 passenger requests per hour. In this simulation, the routing is optimized with a constructive heuristic and the schedule is chosen to ensure smooth transfers at the terminal.

4.2.2.2 Multi-objective

In Lu et al. (2016), feeder buses in a stop-based feeder service can temporarily deviate from their current route to serve demand at the requested locations. When multiple feeder buses are operating in the target service area, the system provides an optimal plan to locate the nearest one to respond to the demands. The objective is to minimize the total bus travel time, with the aim of reducing travel time costs of both the passengers and the operators. The system is optimized every two hours with new incoming requests, however, the number of accepted requests is limited. The optimization problem of this system can be transformed into a traveling salesperson problem and it is solved with a three-stage genetic algorithm. The system is used for a case study in Jinan (China), with 39 passenger requests.

Kim and Schonfeld (2013) study a feeder system, in which one terminal is fed by six regions nearby. In each region, either a fixed transit system or a flexible system operates. The fixed transit system works with fixed routes and timetables. The flexible system is a door-to-door service and works with a flexible route, however, the schedules of the buses have a preset headway. The objective of the system is to minimize the operators' cost, the user in-vehicle cost and the user waiting cost. The transit system is optimized with a hybrid heuristic that combines analytical optimization and a genetic algorithm. They test the system on numerical examples with demand levels between five to 80 trips/mile^2/h. Qiu et al. (2015a) present a model to determine the upper bound of the demand for implementing a MAST flex-route policy instead of a fixed route policy. The MAST service is used as a feeder service, where the base route consists of a set of mandatory stops located at high-demand sites with preset schedules. The buses can deviate from this route within a given distance in order to serve passengers nearby. The objective is to minimize the vehicle operation costs and the transit cost per passenger. The system is optimized with an insertion heuristic and it is tested on a real life case in Salt Lake city (USA), with demand levels of up to 95 passenger requests per hour.

4.2.2.3 Conclusion on semi-flexible many-to-one DOFF-PBS

Research on many-to-one semi-flexible DOFF-PBS is limited (See Table 8). Kim and Schonfeld (2014), Qiu et al. (2015b) optimize DOFF-PBS from the passengers' perspective. Kim and Schonfeld (2014), Qiu et al. (2015b)work with both bus stops and door-to-door requests. Kim and Schonfeld (2014) minimize costs related to delays caused by transfers and slack time. Qiu et al. (2015b) minimize the passengers' travel time, waiting and walking time. Kim and Schonfeld (2014) have restrictions on the pickup and drop-off times, while Qiu et al. (2015b) take a limited fleet size into consideration along with drop-off time constraints. Lu et al. (2016) optimize stop-based systems, taking the operators' and the passengers' need into account. They minimize the vehicle travel time, and by proxy the passengers' travel times. Kim and Schonfeld (2013) and Qiu et al. (2015a) minimize the operators' cost together with the passengers' costs. All of the many-to-one semi-flexible DOFF-PBS are optimized with a heuristic.

5. Static demand-responsive public bus systems

Compared to the systems considered in Section 3 and Section 4, a Static demand-responsive Public Bus System (S-PBS) does not allow changes to individual services just before or during their operation. S-PBS operations are scheduled before the system starts running. Therefore, requests for the S-PBS should arrive before a certain deadline, the day before or some hours before the services start running, or the system should rely on accurate data or predictions available beforehand, in order to optimize the performance. Once the services start running, no more changes are possible. This implies an S-PBS requires no real-time

Table 8: An overview of many-to-one semi-flexible DOFF-PBS

		Constraints on ...					
Reference	Objective	C	F	VTT	PT	DT	PWL
Kim and Schonfeld (2014)	min delay costs				x	x	
Qiu et al. (2015b)	min PTT+ PWT + PWL		x			x	
Lu et al. (2016)	min VTT		x				
Kim and Schonfeld (2013)	min OC + PTT + PWT	x	x				
Qiu et al. (2015a)	min OC + PTT + PWT + PWL			x			

(P = The number of passengers served, C = Capacity, F = Fleet size, VTT = Vehicle travel time, VTD = Vehicle travel distance, PT = Pickup time, DT = Drop-off time, PTT = Passenger travel time, PWT = Passenger waiting time, PWL = Passenger walking time)

data and the planning is much less complicated and often does not differ from the planning of a traditional bus system. However, typically less computation time is available compared to for traditional systems.

It should be noted that many of these Static PBS can easily be transformed into a Dynamic Offline PBS, e.g., Stiglic et al. (2018), Melachrinoudis et al. (2007), and Galarza Montenegro et al. (2021). This can be done, for instance, by introducing the rolling horizon principle, where the next planning horizon considered overlaps with the part of the planning horizon which has already been planned. As a result, some not-yet-departed services in the earlier parts of the planning horizon can still be modified. This requires fast optimization approaches and a way to communicate (modified) departure times to passengers last-minute. Furthermore, part of the literature focuses on identifying under which conditions it is suitable to switch from a traditional bus system to a S-PBS (Zhang et al. (2017), Diana et al. (2009), Li and Quadrifoglio (2009), Kim and Schonfeld (2012), Papanikolaou and Basbas (2020), Lakatos et al. (2020)). All mentioned papers identify and compare costs and passenger travel times for both systems.

5.1. Many-to-many

In this section, *many-to-many* S-PBS are considered. In Section 5.2, we will discuss many-to-one S-PBS services, usually called *feeder* systems.

5.1.1. Fully flexible routes and timetables

First we discuss papers on fully flexible systems without predefined routes or timetables. The most common type of S-PBS known in this category is the Dial-a-Ride service (DAR), introduced in Section 2.3. It operates as a door-to-door service in low demand areas with requests received in advance. Passengers must typically request a trip the day before. Sometimes late requests can still be included, but only if they can be served within the scheduled route of the service.

5.1.1.1 Passengers' perspective

Only one fully flexible many-to-many S-PBS, optimizing from the passengers' perspective is identified: Melis and Sörensen (2020) introduce the on-demand bus routing problem. Given are a fleet of buses with fixed capacity, a set of bus stops in a grid and a set of requests. Each request has a time window within which the transportation needs to take place and a set of close-by bus stops for both departure and arrival. The algorithm decides on both the routing of the buses and assigns stops to requests. The authors use a large neighbourhood search heuristic with embedded local search and optimize the total user ride time, which is the time passengers spend on the bus. By assigning bus stops, within walking distance, to requests, instead of letting passengers choose their bus stops, the objective function value decreases 24% for an instance-scale of 1500 requests. The authors end by comparing on-demand bus routing to a traditional public transport network and find significant improvements in user ride time in favor of on-demand bus routing.

5.1.1.2 Operators' perspective

Stiglic et al. (2018) study a ride-sharing system that integrates public transit. The ride-sharing system serves as a feeder system that connects less densely populated areas to public transit. The public transit system extends the reach of ride-sharing and reduces the detours. The objective of the system is to maximize the number of matched passengers and drivers, and to minimize the extra trip duration that results from matching a rider to a driver. The system works with a limited vehicle capacity of 2 persons per vehicle. Furthermore, the walking distance of the passengers and the trip time of the passengers and drivers must not exceed a certain limit. The passengers also have time windows in which they need to be picked up and dropped off. The matching system is optimized with a branch-and-bound algorithm. The system is tested on a real case in San Francisco (USA) with up to 2000 participants. Luo et al. (2019) study a Dial-a-Ride (DAR) problem, introduced in Section 2.3, that offers a stop-based service for patient transportation. The system takes the lunch breaks of drivers into consideration, as well as their working hours. The seating configuration for patients and availability of drivers in the vehicles is also taken into consideration. Passengers make requests stating their pickup and drop-off bus stop, together with their desired departure time. These requests are received before the operation starts and the optimal operation is determined for the next two hours. The profit of the operator is maximized by increasing the number of satisfied passenger requests, and by decreasing the total distance traveled by the vehicles. To optimize this system, a two-phase branch-and-price-and-cut algorithm is utilized. In order to test the system, instances with 16 to 23 passenger requests are generated with the use of real world data.

Guo et al. (2019) develop a Customized Bus (CB) system, also introduced in Section 2.3, and use an exact model with time windows similar to the DAR problem. The solution includes intermediate stops where passengers can transfer to other lines and systems. The objective function is to minimize the total cost of the service and to maximize the revenue from the passengers served by the CB. Compared to a conventional bus system, the main gains are a reduction in total cost, travel time, route length, and number of vehicles. A case study in Beijing (China) compares the branch-and-cut results with a genetic algorithm (GA) and a tabu search (TS) algorithm for an instance with 1228 requests. CB problems are often categorized as semi-flexible services, if passengers can request a trip on a predefined route (e.g., Zhang et al. (2017) and Cao and Wang (2017) discussed in Section 5.1.2). However, this here is an example of a CB system with time windows and no predefined routes, i.e., a fully flexible service.

Another variant of the door-to-door DAR problem, proposed by Garaix et al. (2010), models an optimal route design considering the concept of alternative paths. In this approach, the road network is represented as a weighted multigraph with several attributes such as travel time and travel costs. The objective function of the mathematical model is to minimize costs guaranteeing an acceptable quality of service for the users and respecting the pickup time and a time window for the drop-off. An insertion heuristic is developed to solve the problem. Tong et al. (2017) implement an on-demand shuttle similar to a Customized Bus (CB). They aim to optimize the load rate in the vehicles to reach profitability and to optimize routing and scheduling in order to satisfy users' constraints. However, in the CB commuters are considered relatively fixed and updates mostly occur only with a monthly frequency. In the system considered here, however, the authors propose more regular optimization to update or generate new routes when the number of bookings for non-served demand meets a threshold. A Lagrangian relaxation-based algorithm is designed to identify potential stops based on passengers walking time and the assignment of buses to a service.

5.1.1.3 Multi-objective

Bakas et al. (2016) address a static DAR with time windows. In the static DAR, all requests are known beforehand, and users indicate a pickup time window and location. The objective is to maximize the number of passengers served and the quality of the service. A secondary objective is to minimize the operator costs. The obtained results are tested with 2500 customers and compared with a similar fixed-route system. To maximize the number of passengers included, the algorithm uses a best-insertion routine. An experimental setup in a medium-sized city in Greece for two time slots (peak and off-peak) is discussed. Nourbakhsh and Ouyang (2012) propose a flexible-route transit system, in which each bus offers a door-to-door service to passengers across a predetermined area. The system works based on requests, in which passengers state their desired origin and destination location and desired time window. The requests are assumed to be known before the operation starts. The optimal planning is then determined for the planning horizon. The objective is to minimize both operator and user costs. The system's operating performance is expressed as closed-form analytical functions of a few key design variables, such as the size of the area where passengers are picked up or dropped off. The optimal design of the system is then obtained by solving a simple constrained nonlinear optimization problem with a steepest decent method. The system is tested with some numerical examples, with demand densities between 1 and 500 passengers per hour per km^2. Chevrier et al. (2012) solve a door-to-door DAR problem as a multi-objective optimization problem. The system makes use of requests that are known before the start of the operation. The goal of this DAR problem is to simultaneously minimize the number of vehicles that are used, minimize the journey duration of the passengers, and minimize the delay of delivery. The DAR problem is optimized through the use of a multi-objective population-based heuristic that is hybridized with a local search procedure. Two sets of numerical examples, with 100 and 1000 requests, are performed in order to test the system.

5.1.1.4 Conclusion on fully flexible many-to-many S-PBS

S-PBS for fully flexible services generally consider route design or service scheduling for DAR or taxi-like services. The literature on fully flexible many-to-many static demand-responsive PBS is somewhat scarce. This is somewhat surprising, given the fact that such systems do appear in practice, but can partially explained by the fact that most of these highly flexible systems will tend to be dynamic, which means that the static model is used mainly for comparative purposes or as a first step in the development of more realistic systems.

An overview of all contributions is presented in Table 9. In this category, Melis and Sörensen (2020) are the only to focus on the passenger perspective. All other papers consider the operator perspective or use a multi-objective approach. Except for Bakas et al. (2016), all contributions include a constraint on the vehicle capacity. The vehicle travel time is only considered in Stiglic et al. (2018) and the vehicle travel distance is considered in Melis and Sörensen (2020) and Guo et al. (2019). Only Stiglic et al. (2018) and Luo et al. (2019) propose an exact algorithm to optimize the performance of the S-PBS.

5.1.2. Semi-flexible routes and/or timetables

A stop-based service in this category is again the Customized Bus (CB). In this mode of operation, passengers submit their requests through an online platform. If the request matches with an already existing CB line, passengers can buy a travel seat. Unsuccessful requests are stored to improve the system in the longer term (Liu and Ceder, 2015).

5.1.2.1 Passengers' perspective

Hrnčíř et al. (2015) present a stop-based ride-sharing service that, besides private cars, also considers timetabled transport services, such as local buses or trains. The transit services can enter the ride-sharing model as inflexible drivers. Hrnčíř et al. (2015) minimize the total travel cost of the passengers. A passenger is able to use different modes of transport during his or her journey: walking, trains, and coaches. They use a multi-agent algorithm to optimize the system. This solution method is tested on a large and complex public transport network in the UK, with realistic travel demand. Zheng et al. (2019) work with a door-to-door system where passengers state their pickup and drop-off locations. These requests are received before a certain deadline after which no more requests are accepted. The system has a certain number of stops that need to be visited at predetermined times. Furthermore, along with a door-to-door service, "meeting points" are introduced, i.e., locations where passengers can be picked up or dropped off collectively. Only a limited number of these meeting points are determined and assigned to passengers. The optimization model

Table 9: An overview of fully flexible many-to-many S-PBS

		Constraints on ...							
Reference	Objective	C	F	VTT	VTD	PT	DT	PTT	PWL
Melis and Sörensen (2020)	min total user ride time	x	x		x	x	x		
Stiglic et al. (2018)	max P + min VTT	x	x	x		x	x	x	x
Luo et al. (2019)	max P + min VTD	x				x	x		
Guo et al. (2019)	min operational costs + max P	x	x		x	x	x		
Garaix et al. (2010)	min operational costs	x				x	x		
Tong et al. (2017)	min operational costs	x				x	x		
Bakas et al. (2016)	max P + min operational costs		x			x			
Nourbakhsh and Ouyang (2012)	min VTD + F + PTT	x							
Chevrier et al. (2012)	min F + PTT + arrival delay	x				x	x		

(P = The number of passengers served, C = Vehicle capacity, F = Fleet size, VTT = Vehicle travel time, VTD = Vehicle travel distance, PT = Pickup time, DT = Drop-off time, PTT = Passenger travel time, PWL = Passenger walking time)

maximizes the number of requests that are satisfied and minimizes the total trip time of the passengers. The optimization model is restricted by predefined departure times at the bus stops and by a maximum walking distance for passengers that are assigned to a meeting point. The optimization problem is solved with a heuristic that incorporates characteristics of a genetic algorithm and local search. Simulation experiments are performed based on a real-life flex-route, with demand levels between 5 and 25 passengers per trip.

Zhang et al. (2020) aim at reducing the passengers' travel time by deciding on the distance between the stops along a 3 km route. Constraints limit the number of vehicles in order to limit the operational costs, considering a passenger density varying until 50 passengers/km along the route. More stops reduce the average speed of the buses, increasing the travel time. However, fewer stops increase walking distance for passengers. Passengers have different time perceptions and the authors try to find the optimal stop spacing considering it. Cao and Wang (2017) aim to reduce passengers' in-vehicle time, waiting time and delay penalty for a CB. They propose a technique to assign passengers to buses. First, the shortest path for each CB line is determined by a branch and prune algorithm. Then passengers are assigned in order to minimize the travel time.

5.1.2.2 Operators' perspective

Considering the operators' perspective, a CB optimizes bus routes with the lines' total profit as objective (Lyu et al., 2019). Here, three steps are performed consecutively. First, demand is clustered to identify potential users for the CB system and travel patterns are created. The second step selects the locations where the bus will stop. In the third step, both routing and timetabling of the lines is done, but part of the route is fixed beforehand. The system periodically plans new CB lines or re-plans existing lines. This routing and timetable can vary from day to day, based on the requests for that day. This service combines a door-to-door service when collecting passengers in the origin area, and a stop-based service when bringing passengers to their destination. The objective is to maximize the total profit of all buses, calculating the total revenue and subtracting the operational costs. Data used for the prediction of demand are taxi trajectories in Nanjing (China). Results present a moderate increase in travel time, however, passengers save significantly in travel fares.

Fu (2002) design a hybrid semi-flexible system merging a traditional fixed-route transit system and a demand-responsive paratransit operation. The model optimizes services with mandatory stops, including the possibility to deviate for door-to-door paratransit requests. The benefits are that the operator does not require an additional service and includes these requests in the regular service. From the users' perspective of the traditional system, this system increases travel time, but the system explicitly limits the additional time for these passengers. It should be noted that many more paratransit systems are developed, offering mobility solutions for people with disabilities. However, most of these systems are used individually, and therefore only a few of these systems appear in this survey.

5.1.2.3 Conclusion on semi-flexible many-to-many S-PBS

S-PBS for semi-flexible services are typically designed to improve a CB system or to develop a hybrid system, with an initial fixed route and the possibility of including deviations. Only a limited number of papers on semi-flexible many-to-many S-PBS is available. An overview is presented in Table 10.

Table 10: An overview of semi-flexible many-to-many S-PBS

		Constraints on ...						
Reference	Objective	C	F	VTT	VTD	PT	DT	PTT
Hrnčíř et al. (2015)	min PTC					x	x	
Zheng et al. (2019)	max P + min PTT					x	x	
Zhang et al. (2020)	min PTT	x		x	x			
Cao and Wang (2017)	min PTT	x	x			x		
Lyu et al. (2019)	max profit	x						
Fu (2002)	max P		x	x		x	x	x

(P = The number of passengers served, C = Vehicle capacity, F = Fleet size, VTT = Vehicle travel time, VTD = Vehicle travel distance, PT = Pickup time, DT = Drop-off time, PTT = Passenger travel time)

5.2. Many-to-one

Many-to-one or feeder services generally operate in low demand areas. The purpose is to gather passengers and bring them to a transit hub or city center. Almasi et al. (2015) include a literature review on the design and implementation of feeder bus lines and the integration with high demand services like train or metro stations. Sets of traditional feeder bus lines are considered combined with one or more rail stations. The authors further present a detailed list of the cost for users and operators, and social costs. Ceder (2013) examines different operational strategies for feeder shuttle transit. These strategies can use fixed or flexible routes, fixed or flexible timetables, single or bi-directional lines, short-turns and short-cuts. These concepts, when applied to feeder S-PBS, increase operational efficiency and require the use of intelligent transportation technologies.

5.2.1. Fully flexible routes and timetables

The problem statement typically is: given an area, and considering a time slot, a shuttle bus performs a route collecting passengers and bringing them to a terminal. In the other direction, passengers go from this terminal to their destination. Both door-to-door and stop-based systems are considered. A fully flexible feeder system occurs mostly as a demand-responsive connector (DRC), introduced in Section 2.3. It is one of the most used flexible transit systems currently in use (Koffman, 2004).

5.2.1.1 *Passengers' perspective*

To the best of our knowledge, no papers on fully flexible feeder lines optimize the passengers' perspective.

5.2.1.2 *Operators' perspective*

Pan et al. (2015) study a stop-based demand-responsive feeder system that serves irregularly shaped and gated communities. In this system, passengers make a request for transportation stating their origin location. The goal of this research is to determine an appropriate service area and routing plans for this feeder system. Designated pickup locations for passengers are determined as well. Furthermore, the scheduling of the feeder buses must be in coordination with the timetable of the urban rail transit that is connected to the feeder system. Passenger requests are received before a small number of buses start to operate. The optimization problem for this system has a bi-level objective. In the upper level, the number of satisfied requests need to be maximized. In the lower level, the operators' cost is minimized. The fleet size and the vehicle capacity are limited. Furthermore, passengers have maximum travel time and waiting time at their origin bus stop.The system is optimized with the use of a constructive heuristic inspired by the gravity model. The system is tested on a network of a district in Jinan (China), with 130 passenger requests.

Lee and Savelsbergh (2017) consider the DRC to minimize operator costs, considering the time window for pickups and drop-offs. The authors consider the train frequency at the station and use these time windows as parameters of the operation, allowing the S-PBS to select the best operation period for servicing passengers. The authors discuss a heuristic to solve DRC routing and scheduling problems. The objective function is to minimize operator costs with the constraint of serving all passengers. Small instances are used to validate the heuristic by comparing the results with the results obtained by an exact mathematical model.

Chien et al. (2001) develop an exact method to optimize both a conventional system during peak hours and a demand-responsive system during off-peak hours. The model optimizes vehicle size and zone area for the 10-hour operation of the demand-responsive system. The authors also identify a switching point between a conventional system and a S-PBS in order to optimize costs considering demand variation, different vehicle sizes and service headways. Sun et al. (2019b) design an ant colony optimization (ACO) heuristic for a demand-responsive system to optimize a feeder line towards a rail station. The objective is to minimize total travel time for all buses, considering a boarding time window and an expected ride time. A case study in Nanjing (China) compares the proposed algorithm with a particle swarm optimization and genetic algorithm to prove its validity.

5.2.1.3 *Multi-objective*

Melachrinoudis et al. (2007) study a DAR that transports passengers between their doorstep and a Health and Recovery Center (HRC). Passengers can make requests stating either their pickup location and their desired arrival time at the HRC, or their drop-off location and desired departure time from the HRC. The requests can be received up until the day before. The authors minimize a convex combination of total vehicle transportation costs and total clients' inconvenience time. The optimization model of this system is solved with a branch-and-bound algorithm, as well as a tabu search algorithm. Real world data of two busy care centers is used to test the system. Li and Quadrifoglio (2009) compare a fixed-route and a demand-responsive system operating a door-to-door feeder service in a large residential area. The purpose is to optimize the zone size for the feeder lines. The comparison aims to minimize total costs, composed of user and operator costs, and maximize transit service performance in the zone design problem. Furthermore, Li and Quadrifoglio (2011) apply the same optimal zone design optimization for a two-vehicle system, which means that instead of optimizing the number of zones, there are now two vehicles operating. Considering a fixed zone served by a conventional service during peak hour and switching to a flexible door-to-door service during low demand periods, Kim and Schonfeld (2012) analyze this variable-type system based on demand variation. The objective function is to minimize the total cost, including operator and user costs. In a variation of a demand-responsive system with the terminal inside the serving area, Wang et al. (2018) optimize how a large area should be divided in zones for each cycle of service. Each zone is served by a feeder demand-responsive system line assigned to a selected terminal within the area. The objective function of the mathematical model optimizes zone division minimizing operational costs. A case study in Calgary (Italy) identifies the impact of the vehicle travel time on costs, headway and service area. Chandra and Quadrifoglio (2013) consider a DRC while focusing on the cycle length. The objective is to find the optimal cycle length so vehicles have enough time to serve all requests received during each cycle. The authors compare five routing methods (Nearest-neighbor, approximate TSP, No-backtrack, Random and insertion heuristic) to identify the optimal cycle length based on the size of the area and the demand.

Kim et al. (2019) study a flexible door-to-door feeder system, where the goal is to minimize the average cost per passenger. The cost per passenger is defined as the operating cost of the vehicles, the in-vehicle time of the passengers and the waiting time of the passengers. The authors develop complex analytical cost functions with the service zone size and the headway of the buses as variables. Service zones contain passengers that need to be picked up and transported to the terminal. The minimum cost is found by solving a sixth order polynomial objective function using Newton's method. The system is tested on a theoretical instance, with 54 passenger requests per hour.

In order to obtain a generic feeder system for interurban areas, Papanikolaou and Basbas (2020) model the S-PBS for a rectangular area with a fixed low demand rate. The objective is to minimize operator and user costs with smaller vehicles in a door-to-door service. For low-demands, the S-PBS tries to find the optimal trade-off of costs by changing the service headway. The case study is a 20 thousands inhabitants small city in Greece, near Thessaloniki. Demand varies during the day from 40 to 100 pax/hr/direction. The conclusion is that it is worthwhile to invest in the S-PBS for off-peak hours. For low demand, the S-PBS remains the most preferable option, but as demand reaches its peak, larger traditional bus services become more efficient.

Another aspect to consider in the many-to-one system is the pattern of the network. The road network layout of cities tends to follow widely different patterns. Most American cities and new metropolises like Beijing (China) have a square or rectangular grid street pattern. Meanwhile, older European cities have a circular or radial street pattern. This variation has a significant impact on feeder systems since transfers in a grid network occur in any crossing of lines, in a radial street pattern these typically take place in a central transfer station. While there are many studies of S-PBS in grid pattern cities, only a few focus on radial street patterns. Shi and Gao (2020) propose a S-PBS with a radial route structure, where buses operate with high speed radial lines and a circular door-to-door feeder service. The objective function is to minimize operational costs and users' travel, waiting and transfer times. It makes metro and rapid transit more competitive by generating extra demand from the fast-moving feeders. A genetic algorithm obtains a flexible service during low-demand operation for a door-to-door service while considering operator and passenger costs. To complement the high-level public transportation system typical for intercity transportation networks, Uchimura et al. (2002) organize a door-to-door S-PBS as a feeder system at the lower level. A genetic algorithm optimizes routing and scheduling of the fully flexible operation of this DAR type service.

5.2.1.4 Conclusion on fully flexible many-to-one S-PBS

All papers modelling a fully flexible feeder S-PBS operate a door-to-door service, which leaves a research gap in the study of stop-based systems. Models typically includes at least a vehicle capacity or fleet size constraint, but typically not many other constraints are considered. Exact solution approaches are only presented in Chien et al. (2001), Melachrinoudis et al. (2007), and Wang et al. (2018). No papers where found that optimize the passengers' perspective only. All papers are summarized in Table 11.

5.2.2. Semi-flexible routes and/or timetables

Compared to the fully flexible S-PBS considered before, systems in this section start from a standard route and/or timetable which can be modified to better serve the demand.

Table 11: An overview of fully flexible many-to-one S-PBS

		Constraints on ...						
Reference	Objective	C	F	VTD	PT	DT	PTT	PWL
Chien et al. (2001)	min operational costs		x					
Pan et al. (2015)	max P, then OC	x	x	x			x	x
Lee and Savelsbergh (2017)	min operational costs		x		x	x		
Sun et al. (2019b)	min VTT							
Uchimura et al. (2002)	min VTD + onboard distance	x			x			
Melachrinoudis et al. (2007)	min OC + passenger inconvenience time	x			x	x		
Li and Quadrifoglio (2009)	min PTT + operational costs		x					
Li and Quadrifoglio (2011)	min PTT + operational costs		x					
Kim and Schonfeld (2012)	min PTT + operational costs	x	x					
Kim et al. (2019)	min PTT + PWT + OC	x						
Chandra and Quadrifoglio (2013)	optimal cycle	x	x					
Wang et al. (2018)	optimal zone	x		x				
Papanikolaou and Basbas (2020)	min PTT + operational costs	x	x					
Shi and Gao (2020)	min PTT + operational costs	x		x				

(P = The number of passengers served, C = Vehicle capacity, F = Fleet size, VTT = Vehicle travel time, VTD = Vehicle travel distance, PT = Pickup time, DT = Drop-off time, PTT = Passenger travel time, PWL = Passenger walking time)

5.2.2.1 Passengers' perspective

Zheng et al. (2018) compare two possible modifications considered in semi-flexible systems: selecting from a limited set of route deviations or designing new partial routes for a door-to-door service. Both are more constrained than fully flexible services since they operate along a well-defined path along a set of stops with the possibility to deviate from the original route. However, the problem description remains the same as for the DRC: a given area with a specific size connected to terminal stations where all passengers want to go. The objective of the mathematical model is the users' cost, comparing both systems to expected and unexpected demand variation and setting the best suitable policy according to operation's limitations.

Galarza Montenegro et al. (2021) study a semi-flexible stop-based feeder service. Passengers make a request for transportation, before the buses start operating, in which they state a desired arrival time a the destination and their origin location. There are two types of bus stops present: optional and mandatory bus stops. Mandatory bus stops are visited by each bus in the system and serve as a safety-net for passengers that did not make a formal request for transportation. Optional bus stops are bus stops that are only visited when there is a request for transportation nearby. The objective is to minimize a weighted sum of the walking time of the passengers, the travel time of the buses and the absolute difference in the actual arrival time and the desired arrival time of each passenger. The system is optimized with the use of a large neighborhood search heuristic. The system is tested on a theoretical network with up to 67 bus stops and demand varying between 12 and 510 passenger requests.

5.2.2.2 Operators' perspective

Lakatos et al. (2020) analyze the hypothesis of replacing a traditional bus by a S-PBS in small rural villages with dead-end lines and smaller vehicles. This network configuration can be considered a many-to-one system because it is a connection from an urban centre to surrounding villages, and deciding not to visit some of the villages reduces significantly the travel time. The objective function is to make the service more cost efficient, offering the same service level for the users. The paper presents a case study in Hungary to determine when it is feasible to operate the traditional bus system or the S-PBS and to find an operation with the same costs but an increased frequency.

Mehran et al. (2020) explore the possibility to replace traditional bus services with S-PBS in low-demand areas. In this S-PBS passengers need to book their trips in advance, choosing a pickup and drop-off location. The model implements a stop-based semi-flexible system minimizing operational costs. The case study is a low-demand bus route in a medium-sized city in Canada. The proposed semi-flexible system enable transit agencies to estimate operating costs and compare this system with a traditional transit service.

5.2.2.3 Conclusion on semi flexible many-to-one S-PBS

The few semi-flexible many-to-one static systems are summarized in Table 12.

5.3. S-PBS systems without optimizations

Some papers evaluate an S-PBS or compare it to another system, without trying to optimize the system itself. In this section, these paper are briefly mentioned. Mukai and Watanabe (2009) propose a decision-making policy to select a fixed, semi-flexible or fully flexible on-demand system, based on the expected demand and profit. Narayan et al. (2020) propose a multi-modal route choice for passengers with scenarios with fixed and demand-responsive systems. Results suggest that users should combine both services in order to lower travel time and costs.

In order to evaluate mode selection by passengers, Zhang et al. (2017) analyze a CB with a multiple travel options in a corridor. It focuses on long distance commutes and analyzes different modes such as private vehicle, park and ride system, taxi, metro and conventional bus. Travel time, monetary cost and travel discomfort were used to support passengers' modal choice. The CB appears to have a high impact on long-distance trips among passengers that prioritize comfort and reliability. They also conclude that the introduction of the CB had no impact on the demand for the conventional bus service. In another system

Table 12: An overview of semi-flexible many-to-one S-PBS

		Constraints on ...						
Reference	Objective	C	F	VTT	VTD	PT	DT	PTT
Zheng et al. (2018)	min PTC		x			x	x	
Galarza Montenegro et al. (2021)	min VTT + PWL + DT difference	x	x			x	x	
Lakatos et al. (2020)	min operational costs		x		x			
Mehran et al. (2020)	min operational costs		x			x	x	

C = Vehicle capacity, F = Fleet size, VTT = Vehicle travel time, VTD = Vehicle travel distance, PT = Pickup time, DT = Drop-off time, PTT = Passenger travel time, PTC= Passenger transportation costs, PWL =passenger walking time)

selection model, Zhang et al. (2018) compare the performance of a park-and-ride system and an S-PBS for a corridor connecting a residence area to a transfer station. Even with the current comparison between traditional bus services and demand-responsive systems, Franco et al. (2020) state that the traditional transport model is not suited to model new mobility services. Since there are only a few of these new mobility services in operation, sufficient data is lacking to model, analyse and optimize these systems. They present alternatives to obtain OD matrices in order to design appropriate new S-PBS systems.

Drakoulis et al. (2018) study demand behaviour for demand-responsive transportation systems by implementing a gamification layer of a simulated service. This game aims to motivate public transport users to behave correctly in a demand-responsive service. The implemented system is similar to a DAR problem with time windows with a fixed fleet size. The optimization of the service considers both users' and operator's perspective and a case study is set in a medium-sized city in Greece. On a more strategic level, Diana et al. (2006) considers a demand responsive system to determine the number of vehicles needed in order to guarantee a predetermined level of quality for the users. The S-PBS operates as a fully flexible many-to-many system. Requests are considered with an origin, destination, and a pickup time. A maximum waiting time for passengers determines the time window allowed for the passengers. A maximum riding time from the initial pickup until drop-off must be guaranteed as well. In order to serve all requests within a specific area, the model estimates the optimal number of vehicles for variations of the duration of waiting time and maximum riding time.

6. Conclusions

This survey clearly shows that between the extremes of conventional public bus systems with fixed routes and timetables, and individual on-demand systems such as taxi or Lyft, many types of demand-responsive public bus systems (DR-PBS) can be (and have been) developed in the last decades. We structure and classify over a hundred papers based on three different degrees of responsiveness: (1) For the *dynamic online DR-PBS*, the planning of each service can be modified during the planning horizon, even if that particular service has already started; (2) For the *dynamic offline DR-PBS*, the planning of each service can still be modified during the planning horizon, as long as that particular service has not started yet; (3) For *static DR-PBS*, the planning is optimized for the entire planning horizon before the services start running.

A distinction is also made between systems optimized from the passenger's perspective, the operator's perspective or both perspectives. Moreover, different concepts are defined that allow further classification and comparison of different DR-PBS: many-to-many or many-to-one, fully-flexible or semi-flexible, door-to-door or stop-based, based on lead time, etc. For each paper, the optimization approach, the instances that are used and the possible implementation in practice, are discussed. We want to repeat here that many traditional systems or feeder systems are also designed while taking the (potential) demand into account.

Sometimes these systems are even called *on-demand*. However, the papers related to these systems are only included in this survey as they are actually demand-responsive. Just designating the system as *on-demand* and then operating it with a fixed route and timetable during the next weeks or months does not make it demand-responsive, at least not according to the definitions put forward in this paper.

In our paper, we have attempted to draw some conclusions for each of the large problem classes. Tables are included to structure and summarize the information, hopefully making clear in which areas research is still lacking and which types of systems have been explorered less than others. We refrain from repeating each of these sub-conclusions here, but refer the reader to each of the specific subsections.

From our survey, it is clear that systems with a higher degree of responsiveness require more (real-time) data, are more complex to operate and require much faster optimization techniques. This also explains the relatively low number of exact techniques implemented, and even the apparent preference for relatively simple heuristics. When looking at the size of the instances considered, expressed by the number of requests and/or vehicles, it is remarkable that the many-to-one systems typically deal with smaller instances compared to the many-to-many systems. This is probably due to the low demand areas that are often served with many-to-one feeder lines. Furthermore, as expected, mostly small vehicles are considered in demand-responsive systems: many with at most 10 passengers or less and only a few systems with buses with 40 seats or more.

As a general conclusion, we must note that there seems to be a considerable gap between theory and practice, since a large number of different systems have been proposed in the literature, of which only a small fraction is also implemented in practice. A notable exception is the "Customized Bus" system, that was discussed extensively in this paper, and that was first implemented in practice and only then studied in the scientific literature. Even though most of the systems discussed in the literature seem to offer some advantages over more traditional public transport systems, there are no doubt considerable hurdles that need to be taken before such systems can be implemented in practice and research should focus on identifying (ways to overcome) them.

Finally, research in this domain organically happens in a bottom-up manner, with authors proposing systems they believe (or know) to be realistic and interesting to study. As a result, a large number of systems have been proposed, with an almost equally large number of names, that often differ only in a few details. Additionally, we must mention that some papers fail to mention some information we consider as basic to describe a demand-responsive system. As a result, it is hard to see the forest for the trees in the literature on demand-responsive public bus systems. We hope this paper will be able to partially remedy this situation and encourage and help researchers to clearly mention aspects such as the lead time for requests, the fleet size and vehicle capacity, the number of requests, the planning horizon, whether the system is stop-based or not, and the degree of responsiveness.

7. Acknowledgments

This project was supported by the FWO (Research Foundation Flanders), project G.0759.19N.

References

Almasi, M.H., Sadollah, A., Mounes, S.M., Karim, M.R., 2015. Optimization of a transit services model with a feeder bus and rail system using metaheuristic algorithms. Journal of Computing in Civil Engineering 29, 04014090.

Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., Rus, D., 2017. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. Proceedings of the National Academy of Sciences 114, 462–467.

Amirgholy, M., Gonzales, E.J., 2016. Demand responsive transit systems with time-dependent demand: User equilibrium, system optimum, and management strategy. Transportation Research Part B: Methodological 92, 234–252.

Archetti, C., Speranza, M.G., Weyland, D., 2016. On-demand public transportation, in: Working paper no.6/2016. Department of Economics and Manegement, University of Brescia, Italy.

Archetti, C., Speranza, M.G., Weyland, D., 2018. A simulation study of an on-demand transportation system. International Transactions in Operational Research 25, 1137–1161.

Atasoy, B., Ikeda, T., Ben-Akiva, M.E., 2015a. Optimizing a flexible mobility on demand system. Transportation Research Record 2563, 76–85.

Atasoy, B., Ikeda, T., Song, X., Ben-Akiva, M.E., 2015b. The concept and impact analysis of a flexible mobility on demand system. Transportation Research Part C: Emerging Technologies 56, 373–392.

Bakas, I., Drakoulis, R., Floudas, N., Lytrivis, P., Amditis, A., 2016. A flexible transportation service for the optimization of a fixed-route public transport network. Transportation Research Procedia 14, 1689–1698.

Bar-Gera, H., 2001. The traffic assignment problem. URL: `https://www.bgu.ac.il/~bargera/tntp/`.

Bertelle, C., Nabaa, M., Olivier, D., Tranouez, P., 2009. A decentralised approach for the transportation on demand problem, in: From System Complexity to Emergent Properties. Springer, pp. 281–289.

Bischoff, J., Führer, K., Maciejewski, M., 2018. Impact assessment of autonomous drt systems. Transportation Research Procedia , 1–10.

Bischoff, J., Maciejewski, M., Nagel, K., 2017. City-wide shared taxis: A simulation study in berlin, in: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), IEEE. pp. 275–280.

Bruni, M., Guerriero, F., Beraldi, P., 2014. Designing robust routes for demand-responsive transport systems. Transportation research part E: logistics and transportation review 70, 1–16.

Cao, Y., Wang, J., 2017. An optimization method of passenger assignment for customized bus. Mathematical Problems in Engineering 2017, Article ID 7914753.

Ceder, A., 2013. Integrated smart feeder/shuttle transit service: simulation of new routing strategies. Journal of Advanced Transportation 47, 595–618.

Chandra, S., Quadrifoglio, L., 2013. A model for estimating the optimal cycle length of demand responsive feeder transit services. Transportation Research Part B: Methodological 51, 1–16.

Chen, P.W., Nie, Y.M., 2017. Analysis of an idealized system of demand adaptive paired-line hybrid transit. Transportation Research Part B: Methodological 102, 38–54.

Chevrier, R., Liefooghe, A., Jourdan, L., Dhaenens, C., 2012. Solving a dial-a-ride problem with a hybrid evolutionary multi-objective approach: Application to demand responsive transport. Applied Soft Computing Journal 12, 1247–1258.

Chien, S.I., Spasovic, L.N., Elefsiniotis, S.S., Chhonkar, R.S., 2001. Evaluation of feeder bus systems with probabilistic time-varying demands and nonadditive time costs. Transportation research record 1760, 47–55.

Cortés, C.E., Jayakrishnan, R., 2002. Design and operational concepts of high-coverage point-to-point transit system. Transportation Research Record 1783, 178–187.

Crainic, T.G., Errico, F., Malucelli, F., Nonato, M., 2012. Designing the master schedule for demand-adaptive transit systems. Annals of Operations Research 194, 151–166.

Crainic, T.G., Malucelli, F., Nonato, M., Guertin, F., 2005. Meta-heuristics for a class of demand-responsive transit systems. INFORMS Journal on Computing 17, 10–24.

Czioska, P., Kutadinata, R., Trifunović, A., Winter, S., Sester, M., Friedrich, B., 2019. Real-world meeting points for shared demand-responsive transportation systems. Public Transport 11, 341–377.

Dessouky, M., Rahimi, M., Weidner, M., 2003. Jointly optimizing cost, service, and environmental performance in demand-responsive transit scheduling. Transportation Research Part D: Transport and Environment 8, 433–465.

Diana, M., Dessouky, M.M., Xia, N., 2006. A model for the fleet sizing of demand responsive transportation services with time windows. Transportation Research Part B: Methodological 40, 651–666.

Diana, M., Quadrifoglio, L., Pronello, C., 2009. A methodology for comparing distances traveled by performance-equivalent fixed-route and demand responsive transit services. Transportation planning and technology 32, 377–399.

Dou, X., Meng, Q., 2019. Feeder Bus Timetable Design and Vehicle Size Setting in Peak Hour Demand Conditions. Transportation Research Record 2673, 321–332.

Drakoulis, R., Bellotti, F., Bakas, I., Berta, R., Paranthaman, P.K., Dange, G.R., Lytrivis, P., Pagle, K., De Gloria, A., Amditis, A., 2018. A gamified flexible transportation service for on-demand public transport. IEEE Transactions on Intelligent Transportation Systems 19, 921–933.

Errico, F., Crainic, T.G., Malucelli, F., Nonato, M., 2013. A survey on planning semi-flexible transit systems: Methodological issues and a unifying framework. Transportation Research Part C: Emerging Technologies 36, 324–338.

Fatnassi, E., Chaouachi, J., Klibi, W., 2015. Planning and operating a shared goods and passengers on-demand rapid transit system for sustainable city-logistics. Transportation Research Part B: Methodological 81, 440–460.

Franco, P., Johnston, R., McCormick, E., 2020. Demand responsive transport: Generation of activity patterns from mobile phone network data to support the operation of new mobility services. Transportation Research Part A: Policy and Practice 131, 244–266.

Fu, L., 2002. Planning and design of flex-route transit services. Transportation Research Record 1791, 59–66.

Fu, L., Liu, Q., Calamai, P., 2003. Real-time optimization model for dynamic scheduling of transit operations. Transportation Research Record 1857, 48–55.

Galarza Montenegro, B.D., Sörensen, K., Vansteenwegen, P., 2021. A large neighborhood search algorithm to optimize a demand-responsive feeder service. Transportation Research Part C: Emerging Technologies (Accepted) 127, 103102.

Garaix, T., Artigues, C., Feillet, D., Josselin, D., 2010. Vehicle routing problems with alternative paths: An application to on-demand transportation. European Journal of Operational Research 204, 62–75.

Gkiotsalitis, K., Wu, Z., Cats, O., 2019. A cost-minimization model for bus fleet allocation featuring the tactical generation of short-turning and interlining options. Transportation Research Part C: Emerging Technologies 98, 14–36.

Gomes, R., de Sousa, J.P., Dias, T.G., 2014. A grasp-based approach for demand responsive transportation. International Journal of Transportation 2, 21–32.

Gomes, R., de Sousa, J.P., Dias, T.G., 2015. Sustainable demand responsive transportation systems in a context of austerity: The case of a portuguese city. Research in Transportation Economics 51, 94–103.

Guo, R., Guan, W., Zhang, W., 2018. Route Design Problem of Customized Buses: Mixed Integer Programming Model and Case Study. Journal of Transportation Engineering, Part A: Systems 144, Article ID 04018069.

Guo, R., Guan, W., Zhang, W., Meng, F., Zhang, Z., 2019. Customized bus routing problem with time window restrictions:

model and case study. Transportmetrica A: Transport Science 15, 1804–1824.
Hadas, Y., Ceder, A., 2008. Multiagent approach for public transit system based on flexible routes. Transportation Research Record , 89–96.
Hickman, M.D., Blume, K.L., 2001. An investigation of integrated transit service. Technical Report. Southwest University Transportation Center, Texas Transportation Institute.
Ho, S.C., Szeto, W., Kuo, Y.H., Leung, J.M., Petering, M., Tou, T.W., 2018. A survey of dial-a-ride problems: Literature review and recent developments. Transportation Research Part B: Methodological 111, 395–421.
Horn, M.E., 2002a. Fleet scheduling and dispatching for demand-responsive passenger services. Transportation Research Part C: Emerging Technologies 10, 35–63.
Horn, M.E., 2002b. Multi-modal and demand-responsive passenger transport systems: a modelling framework with embedded control systems. Transportation Research Part A: Policy and Practice 36, 167–188.
Horn, M.E., 2004. Procedures for planning multi-leg journeys with fixed-route and demand-responsive passenger transport services. Transportation Research Part C: Emerging Technologies 12, 33–55.
Hrnčíř, J., Rovatsos, M., Jakob, M., 2015. Ridesharing on timetabled transport services: A multiagent planning approach. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations 19, 89–105.
Huang, D., Gu, Y., Wang, S., Liu, Z., Zhang, W., 2020a. A two-phase optimization model for the demand-responsive customized bus network design. Transportation Research Part C: Emerging Technologies 111, 1–21.
Huang, D., Tong, W., Wang, L., Yang, X., 2020b. An analytical model for the many-to-one demand responsive transit systems. Sustainability (Switzerland) 12, Article ID 298.
Hyland, M., Mahmassani, H.S., 2020. Operational benefits and challenges of shared-ride automated mobility-on-demand services. Transportation Research Part A: Policy and Practice 134, 251–270.
Ikeda, T., Fujita, T., Ben-Akiva, M.E., 2015. Mobility on demand for improving business profits and user satisfaction. FUJITSU Sci. Tech. J 51, 21–26.
Inturri, G., Giuffrida, N., Ignaccolo, M., Le Pira, M., Pluchino, A., Rapisarda, A., 2018. Testing demand responsive shared transport services via agent-based simulations, in: New Trends in Emerging Complex Real Life Problems. Springer, pp. 313–320.
Jäger, B., Brickwedde, C., Lienkamp, M., 2018. Multi-agent simulation of a demand-responsive transit system operated by autonomous vehicles. Transportation Research Record 2672, 764–774.
Ji-Yang, S., Jian-Ling, H., Yan-Yan, C., Pan-Yi, W., Jian-Lin, J., 2020. Flexible Bus Route Optimization for Multitarget Stations. Mathematical Problems in Engineering 2020, Article ID 7183465.
Jokinen, J.P., Sihvola, T., Hyytiä, E., Sulonen, R., 2011. Why urban mass demand responsive transport?, in: 2011 IEEE Forum on Integrated and Sustainable Transportation Systems, IEEE. pp. 317–322.
Jung, J., Jayakrishnan, R., Park, J.Y., 2013. Design and modeling of real-time shared-taxi dispatch algorithms, in: Proc. Transportation Research Board 92nd Annual Meeting.
Kawamura, K., Mukai, N., 2009. Optimization of transport plan for on-demand bus system using electrical vehicles, in: International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Springer. pp. 656–663.
Kim, M., Schonfeld, P., 2012. Conventional, flexible, and variable-type bus services. Journal of Transportation Engineering 138, 263–273.
Kim, M., Schonfeld, P., 2014. Integration of conventional and flexible bus services with timed transfers. Transportation Research Part B: Methodological 68, 76–97.
Kim, M.E., Levy, J., Schonfeld, P., 2019. Optimal zone sizes and headways for flexible-route bus services. Transportation Research Part B: Methodological 130, 67–81.
Kim, M.E., Schonfeld, P., 2013. Integrating bus services with mixed fleets. Transportation Research Part B: Methodological 55, 227–244.
Kim, M.E., Schonfeld, P., 2015. Maximizing net benefits for conventional and flexible bus services. Transportation Research Part A: Policy and Practice 80, 116–133.
Koffman, 2004. Operational experiences with flexible transit services. Washington, D.C: Transportation Research Board.
Koh, K., Ng, C., Pan, D., Mak, K.S., 2018. Dynamic bus routing: A study on the viability of on-demand high-capacity ridesharing as an alternative to fixed-route buses in singapore, in: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), IEEE. pp. 34–40.
Kuhn, H.W., 1955. The hungarian method for the assignment problem. Naval research logistics quarterly 2, 83–97.
Lakatos, A., Tóth, J., Mándoki, P., 2020. Demand responsive transport service of 'dead-end villages' in interurban traffic. Sustainability 12, Article ID 3820.
Lee, A., Savelsbergh, M., 2017. An extended demand responsive connector. EURO Journal on Transportation and Logistics 6, 25–50.
Lee, Y.J., Meskar, M., Nickkar, A., Sahebi, S., 2019. Development of an Algorithm for Optimal Demand Responsive Relocatable Feeder Transit Networks Serving Multiple Trains and Stations. Urban Rail Transit 5, 186–201.
Leich, G., Bischoff, J., 2019. Should autonomous shared taxis replace buses? a simulation study. Transportation Research Procedia 41, 450–460.
Li, X., Hu, S., Fan, W., Deng, K., 2018a. Modeling an enhanced ridesharing system with meet points and time windows. PloS one 13, Article ID e0195927.
Li, X., Quadrifoglio, L., 2009. Optimal zone design for feeder transit services. Transportation Research Record 2111, 100–108.
Li, X., Quadrifoglio, L., 2010. Feeder transit services: Choosing between fixed and demand responsive policy. Transportation Research Part C: Emerging Technologies 18, 770–780.
Li, X., Quadrifoglio, L., 2011. 2-vehicle zone optimal design for feeder transit services. Public Transport 3, 89–104.

Li, X., Wei, M., Hu, J., Yuan, Y., Jiang, H., 2018b. An agent-based model for dispatching real-time demand-responsive feeder bus. Mathematical Problems in Engineering 2018, Article ID 6925764.

Liu, K., Zhang, J., Yang, Q., 2019a. Bus pooling: A large-scale bus ridesharing service. IEEE Access 7, 74248–74262.

Liu, L., Sun, L., Chen, Y., Ma, X., 2019b. Optimizing fleet size and scheduling of feeder transit services considering the influence of bike-sharing systems. Journal of Cleaner Production 236, 117550.

Liu, T., Ceder, A.A., 2015. Analysis of a new public-transport-service concept: Customized bus in china. Transport Policy 39, 63–76.

Liyanage, S., Dia, H., 2020. An agent-based simulation approach for evaluating the performance of on-demand bus services. Sustainability 12, Article ID 4117.

Lotfi, S., Abdelghany, K., Hashemi, H., 2019. Modeling Framework and Decomposition Scheme for On-Demand Mobility Services with Ridesharing and Transfer. Computer-Aided Civil and Infrastructure Engineering 34, 21–37.

Lownes, N.E., Machemehl, R.B., 2010. Exact and heuristic methods for public transit circulator design. Transportation Research Part B: Methodological 44, 309–318.

Lu, X., Yu, J., Yang, X., Pan, S., Zou, N., 2016. Flexible feeder transit route design to enhance service accessibility in urban area. Journal of Advanced Transportation 50, 507–521.

Luo, Z., Liu, M., Lim, A., 2019. A two-phase branch-and-price-and-cut for a dial-a-ride problem in patient transportation. Transportation Science 53, 113–130.

Lyu, Y., Chow, C.Y., Lee, V.C., Ng, J.K., Li, Y., Zeng, J., 2019. Cb-planner: A bus line planning framework for customized bus systems. Transportation Research Part C: Emerging Technologies 101, 233–253.

Ma, S., Zheng, Y., Wolfson, O., 2013. T-share: A large-scale dynamic taxi ridesharing service, in: 2013 IEEE 29th International Conference on Data Engineering (ICDE), IEEE. pp. 410–421.

Marković, N., Kim, M.E., Kim, E., Milinković, S., 2019. A threshold policy for dispatching vehicles in demand-responsive transit systems. Promet - Traffic - Traffico 31, 387–395.

Masoud, N., Jayakrishnan, R., 2017. A real-time algorithm to solve the peer-to-peer ride-matching problem in a flexible ridesharing system. Transportation Research Part B: Methodological 106, 218–236.

Masoud, N., Nam, D., Yu, J., Jayakrishnan, R., 2017. Promoting peer-to-peer ridesharing services as transit system feeders. Transportation Research Record 2650, 74–83.

Mehran, B., Yang, Y., Mishra, S., 2020. Analytical models for comparing operational costs of regular bus and semi-flexible transit services. Public Transport 12, 147–169.

Melachrinoudis, E., Ilhan, A.B., Min, H., 2007. A dial-a-ride problem for client transportation in a health-care organization. Computers and Operations Research 34, 742–759.

Melis, L., Sörensen, K., 2020. The on-demand bus routing problem: A large neighborhood search heuristic for a dial-a-ride problem with bus station assignment. Working Papers. University of Antwerp, Faculty of Business and Economics.

Melis, L., Sörensen, K., 2021. The on-demand bus routing problem: The cost of real-time requests, in: Submitted to CASPT21.

Mukai, N., Watanabe, T., 2009. Simulation analysis of decision-making policy for on-demand transport systems. Applied Intelligence 31, 225.

Najmi, A., Rey, D., Rashidi, T.H., 2017. Novel dynamic formulations for real-time ride-sharing systems. Transportation research part E: logistics and transportation review 108, 122–140.

Narayan, J., Cats, O., van Oort, N., Hoogendoorn, S., 2017. Performance assessment of fixed and flexible public transport in a multi agent simulation framework. Transportation Research Procedia 27, 109–116.

Narayan, J., Cats, O., van Oort, N., Hoogendoorn, S., 2020. Integrated route choice and assignment model for fixed and flexible public transport systems. Transportation Research Part C: Emerging Technologies 115, Article ID 102631.

Navidi, Z., Ronald, N., Winter, S., 2018. Comparison between ad-hoc demand responsive and conventional transit: a simulation study. Public Transport 10, 147–167.

Nourbakhsh, S.M., Ouyang, Y., 2012. A structured flexible transit system for low demand areas. Transportation Research Part B: Methodological 46, 204–216.

Pan, S., Yu, J., Yang, X., Liu, Y., Zou, N., 2015. Designing a Flexible Feeder Transit System Serving Irregularly Shaped and Gated Communities: Determining Service Area and Feeder Route Planning. Journal of Urban Planning and Development 141, Article ID 04014028.

Pandey, V., Monteil, J., Gambella, C., Simonetto, A., 2019. On the needs for maas platforms to handle competition in ridesharing mobility. Transportation Research Part C: Emerging Technologies 108, 269–288.

Papanikolaou, A., Basbas, S., 2020. Analytical models for comparing demand responsive transport with bus services in low demand interurban areas. Transportation Letters , 1–8.

Pei, M., Lin, P., Du, J., Li, X., 2019a. Operational design for a real-time flexible transit system considering passenger demand and willingness to pay. IEEE Access 7, 180305–180315.

Pei, M., Lin, P., Liu, R., Ma, Y., 2019b. Flexible transit routing model considering passengers' willingness to pay. IET Intelligent Transport Systems 13, 841–850.

Pei, M., Lin, P., Ou, J., 2019c. Real-time optimal scheduling model for transit system with flexible bus line length. Transportation Research Record 2673, 800–810.

Pelzer, D., Xiao, J., Zehe, D., Lees, M.H., Knoll, A.C., Aydt, H., 2015. A partition-based match making algorithm for dynamic ridesharing. IEEE Transactions on Intelligent Transportation Systems 16, 2587–2598.

Perera, T., Gamage, C.N., Prakash, A., Srikanthan, T., 2018a. A simulation framework for a real-time demand responsive public transit system, in: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), IEEE. pp. 608–613.

Perera, T., Prakash, A., Gamage, C.N., Srikanthan, T., 2018b. Hybrid genetic algorithm for an on-demand first mile transit

system using electric vehicles, in: International Conference on Computational Science, Springer. pp. 98–113.
Perera, T., Prakash, A., Srikanthan, T., 2017. A scalable heuristic algorithm for demand responsive transportation for first mile transit, in: 2017 IEEE 21st International Conference on Intelligent Engineering Systems (INES), IEEE. pp. 000157–000162.
Pratelli, A., Lupi, M., Farina, A., Pratelli, C., 2018. Comparing route deviation bus operation with respect to dial-a-ride service for a low-demand residential area. DATA ANALYTICS 2018 , Article ID 151.
Qiu, F., Li, W., Haghani, A., 2015a. An exploration of the demand limit for flex-route as feeder transit services: a case study in salt lake city. Public Transport 7, 259–576.
Qiu, F., Li, W., Zhang, J., 2014. A dynamic station strategy to improve the performance of flex-route transit services. Transportation Research Part C: Emerging Technologies 48, 229–240.
Qiu, F., Shen, J., Zhang, X., An, C., 2015b. Demi-flexible operating policies to promote the performance of public transit in low-demand areas. Transportation Research Part A: Policy and Practice 80, 215–230.
Quadrifoglio, L., Dessouky, M.M., 2004. Mobility allowance shuttle transit (mast) services: formulation and simulation comparison with conventional fixed route bus services, in: Modelling, simulation, and optimization Á Proceedings of the 4th IASTED international conference. Kauai, HI, USA, 17Á19 August. Calgary: Acta Press, 6pp.
Quadrifoglio, L., Dessouky, M.M., Ordóñez, F., 2008. Mobility allowance shuttle transit (mast) services: Mip formulation and strengthening with logic constraints. European Journal of Operational Research 185, 481–494.
Quadrifoglio, L., Dessouky, M.M., Palmer, K., 2007. An insertion heuristic for scheduling mobility allowance shuttle transit (mast) services. Journal of Scheduling 10, 25–40.
Quadrifoglio, L., Li, X., 2009. A methodology to derive the critical demand density for designing and operating feeder transit services. Transportation Research Part B: Methodological 43, 922–935.
Rigas, E.S., Ramchurn, S.D., Bassiliades, N., 2018. Algorithms for electric vehicle scheduling in large-scale mobility-on-demand schemes. Artificial Intelligence 262, 248–278.
Ronald, N., Thompson, R., Haasz, J., Winter, S., 2013. Determining the viability of a demand-responsive transport system under varying demand scenarios, in: Proceedings of the Sixth ACM SIGSPATIAL International Workshop on Computational Transportation Science, pp. 7–12.
Ronald, N., Thompson, R., Winter, S., 2017. Simulating ad-hoc demand-responsive transportation: a comparison of three approaches. Transportation Planning and Technology 40, 340–358.
Ronald, N., Thompson, R.G., Winter, S., 2015. Comparison of constrained and ad hoc demand-responsive transportation systems. Transportation Research Record 2563, 44–51.
Santos, D.O., Xavier, E.C., 2013. Dynamic taxi and ridesharing: A framework and heuristics for the optimization problem, in: Twenty-Third International Joint Conference on Artificial Intelligence.
Shen, J., Yang, S., Gao, X., Qiu, F., 2017. Vehicle routing and scheduling of demand-responsive connector with on-demand stations. Advances in Mechanical Engineering 9, Article ID 1687814017706433.
Sheu, J.B., 2002. A fuzzy clustering-based approach to automatic freeway incident detection and characterization. Fuzzy Sets and Systems 128, 377–388.
Shi, H., Gao, M., 2020. Analysis of a flexible transit network in a radial street pattern. Journal of Advanced Transportation 2020, Article ID 5379218.
Simonetto, A., Monteil, J., Gambella, C., 2019. Real-time city-scale ridesharing via linear assignment problems. Transportation Research Part C: Emerging Technologies 101, 208–232.
Smet, P., 2019. Ride sharing with flexible participants: a metaheuristic approach for large-scale problems. International Transactions in Operational Research 28, 91–118.
Stiglic, M., Agatz, N., Savelsbergh, M., Gradisar, M., 2018. Enhancing urban mobility: Integrating ride-sharing and public transit. Computers Operations Research 90, 12–21.
Sun, B., Wei, M., Wu, W., 2019a. An optimization model for demand-responsive feeder transit services based on ride-sharing car. Information 10, Article ID 370.
Sun, B., Wei, M., Yang, C., Ceder, A., 2019b. Solving demand-responsive feeder transit service design with fuzzy travel demand: A collaborative ant colony algorithm approach. Journal of Intelligent & Fuzzy Systems 37, 3555–3563.
Sun, B., Wei, M., Yang, C., Xu, Z., Wang, H., 2018a. Personalised and coordinated demand-responsive feeder transit service design: A genetic algorithms approach. Future Internet 10, Article ID 61.
Sun, B., Wei, M., Zhu, S., 2018b. Optimal design of demand-responsive feeder transit services with passengers' multiple time windows and satisfaction. Future Internet 10, Article ID 30.
Tong, L.C., Zhou, L., Liu, J., Zhou, X., 2017. Customized bus service design for jointly optimizing passenger-to-vehicle assignment and vehicle routing. Transportation Research Part C: Emerging Technologies 85, 451–475.
Tsubouchi, K., Hiekata, K., Yamato, H., 2009. Scheduling algorithm for on-demand bus system, in: 2009 Sixth International Conference on Information Technology: New Generations, IEEE. pp. 189–194.
Tsubouchi, K., Yamato, H., Hiekata, K., 2010. Innovative on-demand bus system in japan. IET Intelligent Transport Systems 4, 270–279.
Uchimura, K., Takahashi, H., Saitoh, T., 2002. Demand responsive services in hierarchical public transportation system. IEEE Transactions on Vehicular Technology 51, 760–766.
Vallée, S., Oulamara, A., Cherif-Khettaf, W.R., 2017. Maximizing the number of served requests in an online shared transport system by solving a dynamic darp, in: International Conference on Computational Logistics, Springer. pp. 64–78.
Van Engelen, M., Cats, O., Post, H., Aardal, K., 2018. Enhancing flexible transport services with demand-anticipatory insertion heuristics. Transportation Research Part E: Logistics and Transportation Review 110, 110–121.
Viergutz, K., Schmidt, C., 2019. Demand responsive-vs. conventional public transportation: A matsim study about the rural town of colditz, germany. Procedia Computer Science 151, 69–76.

Wang, L., Wirasinghe, S., Kattan, L., Saidi, S., 2018. Optimization of demand-responsive transit systems using zonal strategy. International Journal of Urban Sciences 22, 366–381.

Wang, S., Correia, G.H.d.A., Lin, H.X., 2019. Exploring the performance of different on-demand transit services provided by a fleet of shared automated vehicles: An agent-based model. Journal of Advanced Transportation 2019, Article ID 7878042.

Wang, Z., Yu, J., Hao, W., Tang, J., Zeng, Q., Ma, C., Yu, R., 2020. Two-step coordinated optimization model of mixed demand responsive feeder transit. Journal of Transportation Engineering, Part A: Systems 146, Article ID 04019082.

Wei, M., Liu, T., Sun, B., Jing, B., 2020. Optimal Integrated Model for Feeder Transit Route Design and Frequency-Setting Problem with Stop Selection. Journal of Advanced Transportation 2020, Article ID 6517248.

Wilson, N.H., Sussman, J.M., Wong, H.K., Higonnet, T., 1971. Scheduling algorithms for a dial-a-ride system. Massachusetts Institute of Technology. Urban Systems Laboratory.

Winter, K., Cats, O., Correia, G., van Arem, B., 2018. Performance analysis and fleet requirements of automated demand-responsive transport systems as an urban public transport service. International journal of transportation science and technology 7, 151–167.

Winter, K., Cats, O., De Almeida Correia, G.H., Van Arem, B., 2016. Designing an automated demand-responsive transport system: Fleet size and performance analysis for a campus-train station service. Transportation Research Record 2542, 75–83.

Yu, Y., Machemehl, R.B., Xie, C., 2015. Demand-responsive transit circulator service network design. Transportation Research Part E: Logistics and Transportation Review 76, 160–175.

Zhang, J., Wang, D.Z., Meng, M., 2017. Analyzing customized bus service on a multimodal travel corridor: An analytical modeling approach. Journal of Transportation Engineering, Part A: Systems 143, Article ID 04017057.

Zhang, J., Wang, D.Z., Meng, M., 2018. Which service is better on a linear travel corridor: Park & ride or on-demand public bus? Transportation Research Part A: Policy and Practice 118, 803–818.

Zhang, J., Wang, D.Z., Meng, M., 2020. Optimization of bus stop spacing for on-demand public bus service. Transportation Letters 12, 329–339.

Zhao, M., Yin, J., An, S., Wang, J., Feng, D., 2018. Ridesharing Problem with Flexible Pickup and Delivery Locations for App-Based Transportation Service: Mathematical Modeling and Decomposition Methods. Journal of Advanced Transportation 2018, Article ID 6430950.

Zheng, Y., Li, W., Qiu, F., 2018. A methodology for choosing between route deviation and point deviation policies for flexible transit services. Journal of Advanced Transportation 2018, Article ID 6292410.

Zheng, Y., Li, W., Qiu, F., Wei, H., 2019. The benefits of introducing meeting points into flex-route transit services. Transportation Research Part C: Emerging Technologies 106, 98–112.

8. Appendix: Tables on solution methods and test instances

Table 13: Solution methods according to different demand levels for DON-PBS

1. Exact methods	
	1.1 Mathematical Programming (e.g. CPLEX)
L	Bruni et al. (2014), Li et al. (2018b,a), Pratelli et al. (2018)
	1.2 Enumeration Method
L	Hickman and Blume (2001), Pei et al. (2019a), Shen et al. (2017), Yu et al. (2015)
	1.3 Hungarian Method
H	Wang et al. (2019)
2. (Meta)heuristic Methods	
	2.1 Greedy insertion
L	Ronald et al. (2015), Shen et al. (2017), Perera et al. (2017)
M	Atasoy et al. (2015b,a), Ikeda et al. (2015)
H	Archetti et al. (2018) , Bischoff et al. (2018, 2017) , Jokinen et al. (2011), Leich and Bischoff (2019), Koh et al. (2018), Narayan et al. (2017), Navidi et al. (2018), Ronald et al. (2013), Viergutz and Schmidt (2019)
	2.1 Efficient vehicle assignment with greedy insertion
L	Bertelle et al. (2009)
M	Tsubouchi et al. (2009, 2010)
H	Cortés and Jayakrishnan (2002), Hyland and Mahmassani (2020), Jäger et al. (2018), Ma et al. (2013), Van Engelen et al. (2018), Winter et al. (2018)
	2.3 Metaheuristics based on local search
L	Bruni et al. (2014), Yu et al. (2015)
H	Horn (2002a,b)
	2.4 Metaheuristics based on construction
H	Vallée et al. (2017)
	2.5 Metaheuristics based on population search
L	Kawamura and Mukai (2009) , Wang et al. (2020)
M	Li et al. (2018b,a)
	2.6 Hybrid metaheuristics
L	Perera et al. (2018b,a) , Gomes et al. (2014, 2015)
M	Santos and Xavier (2013)
H	Melis and Sörensen (2021)
	2.7 Matheuristic
H	Alonso-Mora et al. (2017), Simonetto et al. (2019), Pandey et al. (2019)
	2.8 Simulation without using heuristics
L	Inturri et al. (2018), Pei et al. (2019c)
H	Liyanage and Dia (2020)

(L = Low demand level (e.g. rural area), M = Medium demand level (e.g. suburban area), H = High demand level (e.g. urban area))

Table 14: Solution methods according to different demand levels for DOFF-PBS

1. Exact methods	
	1.1 Mathematical programming
L	Quadrifoglio et al. (2008), Huang et al. (2020a)
M	Horn (2004)
	1.2 Dynamic programming
L	Hadas and Ceder (2008)
H	Masoud et al. (2017)
	1.3 Enumeration Method
H	Fu et al. (2003)
	1.4 Differentiation techniques
H	Amirgholy and Gonzales (2016)
2. (Meta)heuristic Methods	
	2.1 Insertion algorithms
M	Quadrifoglio and Li (2009), Li and Quadrifoglio (2010), Qiu et al. (2014), Qiu et al. (2015a), Marković et al. (2019)
H	Huang et al. (2020a)
	2.2 Metaheuristics based on local search
H	Crainic et al. (2005), Rigas et al. (2018), Pei et al. (2019b), Lee et al. (2019), Crainic et al. (2012)
	2.3 Metaheuristics based on construction
L	Qiu et al. (2015b)
M	Ji-Yang et al. (2020)
H	Crainic et al. (2005), Huang et al. (2020b)
	2.4 Metaheuristics based on population search
L	Lu et al. (2016)
M	Kim and Schonfeld (2015), Chen and Nie (2017), Sun et al. (2018a), Sun et al. (2018b)
H	Kim and Schonfeld (2014), Guo et al. (2018), Dou and Meng (2019), Gkiotsalitis et al. (2019)
	2.5 Hybrid heuristics
M	Sun et al. (2019a)
H	Kim and Schonfeld (2013), Liu et al. (2019b), Wei et al. (2020),
	2.6 Matheuristics
L	Lotfi et al. (2019)
M	Chen and Nie (2017)
H	Zhao et al. (2018)
	2.7 Simulation without the use of heuristics
M	Winter et al. (2016)
H	Fatnassi et al. (2015)

(L = Low demand level (e.g. rural area), M = Medium demand level (e.g. suburban area), H = High demand level (e.g. urban area))

Table 15: Solution methods according to different demand levels for S-PBS

1. Exact methods	
	1.1 Mathematical Programming (e.g. CPLEX)
L	Li and Quadrifoglio (2009), Li and Quadrifoglio (2011), Luo et al. (2019), Lakatos et al. (2020), Mehran et al. (2020), Bakas et al. (2016)
M	Cao and Wang (2017), Fu (2002), Chandra and Quadrifoglio (2013), Zheng et al. (2018)
H	Wang et al. (2018), Stiglic et al. (2018), Tong et al. (2017)
	1.2 Dynamic programming
L	Garaix et al. (2010), Zhang et al. (2020)
	1.3 Numerical analysis
L	Papanikolaou and Basbas (2020)
M	Kim and Schonfeld (2012), Shi and Gao (2020)
2. (Meta)heuristic Methods	
	2.1 Metaheuristics based on local search
M	Nourbakhsh and Ouyang (2012), Lee and Savelsbergh (2017)
H	Melachrinoudis et al. (2007), Guo et al. (2019)
	2.2 Metaheuristics based on population search
H	Guo et al. (2019), Sun et al. (2019b), Uchimura et al. (2002)
	2.3 Metaheuristics based on construction
M	Pan et al. (2015)
H	Hrnčíř et al. (2015), Galarza Montenegro et al. (2021)
	2.4 Hybrid metaheuristics
M	Zheng et al. (2019)
H	Chevrier et al. (2012), Lyu et al. (2019), Czioska et al. (2019) Melis and Sörensen (2020)
	2.5 Simulation without using heuristics
L	Chien et al. (2001)
	2.6 Numerical Approximation
M	Kim et al. (2019)

(L = Low demand level (e.g. rural area), M = Medium demand level (e.g. suburban area), H = High demand level (e.g. urban area))

Table 16: Instance scale of DON-PBS literature

Reference	S/D	PH	C	R	V	S	N	Location
Many-to-many fully flexible DON-PBS								
Jokinen et al. (2011)	S	10	14	5000 - 50000	236	n.m.	T	
Vallée et al. (2017)	S	Day	8	148 - 2200	5 - 75	213 - 286	B	Paris (France), Bristol (UK)
Melis and Sörensen (2021)	S	4	8	2000	0 - 600	121	T	
Navidi et al. (2018)	D	Day	4	720 - 10800	3 - 31	d.s.a.	T+B	Belgrave (Australia)
Ronald et al. (2013)	D	5	4	7500	8 - 18	d.s.a.	T	
Archetti et al. (2018)	D	1	8	500 - 5000	200 - 500	d.s.a.	T	
Narayan et al. (2017)	D	Day	∞	84110	1000 - 3000	d.s.a.	B	Sioux Falls (USA)
Alonso-Mora et al. (2017)	D	Week	10	3 million	1000 - 3000	d.s.a.	B	New York (USA)
Santos and Xavier (2013)	D	Day	4	540 - 744	1 - 200	d.s.a.	B	São Paulo (Brazil)
Tsubouchi et al. (2010)	S	Day	n.m.	40 - 170	3 - 10	209 - 384	B	Kashiwa, Sakai, Moriyama (Japan)
Bertelle et al. (2009)	S	n.m.	4	47 - 315	4 - 8	50	T	
Bischoff et al. (2017)	D	Day	2 - 4	27336	4212	d.s.a.	B	Berlin (Germany)
Viergutz and Schmidt (2019)	S+D	Day	6 - 14	500	5 - 10	n.m.	B	Colditz (Germany)
Bischoff et al. (2018)	S+D	Day	8	21346	200 - 600	400	B	Cottbus (Germany)
Leich and Bischoff (2019)	D	Day	1 - 20	18664	120 - 1000	d.s.a.	B	Berlin (Germany)
Wang et al. (2019)	S+D	Day	2	110000	2000 - 4500	n.m.	T	
Simonetto et al. (2019)	D	Day	4 - 10	382779 - 460700	150 - 3000	d.s.a.	B	New York (USA)
Pandey et al. (2019)	D	Day	4	382779 - 460700	500	d.s.a	B	New York (USA)
Ma et al. (2013)	D	0.5	n.m.	n.m.	3000	d.s.a.	B	Beijing (China)
Kawamura and Mukai (2009)	D	n.m.	∞	n.m.	2	d.s.a.	B	(Japan)
Bruni et al. (2014)	S	Day	1 - 8	24 - 144	10	n.m.	T	
Horn (2002b)	S+D	Day	mb	14000	n.m.	n.m.	T	
Liyanage and Dia (2020)	S	Day	7	13550	n.m.	n.m.	B	Melbourne (Australia)
Jäger et al. (2018)	S	Day	6	2.3 million	43000	4923	B	(Singapore)
Winter et al. (2018)	S	4	2 - 40	11697	400 - 600	d.s.a.	B	Arnhem (the Netherlands)
Gomes et al. (2015)	S	4	27	32	1	n.m.	B	Porto (Portugal)
Hyland and Mahmassani (2020)	D	5	2	3620 - 7240	650 - 10000	d.s.a.	B	New York (USA)
Atasoy et al. (2015b)	D	Day	mb	5000	60	d.s.a.	B	Tokyo(Japan)
Ikeda et al. (2015)	D	Day	mb	4600	25 - 150	n.m.	B	Tokyo(Japan)
Atasoy et al. (2015a)	D	Day	mb	5000	60	n.m.	B	Tokyo(Japan)
Ronald et al. (2015)	S	Day	∞	10 - 50	1	n.m.	B	Tokyo(Japan)
Van Engelen et al. (2018)	S	3	5	2000	100	36	B	Ijmouden (the Netherlands)
Many-to-many semi-flexible DON-PBS								
Cortés and Jayakrishnan (2002)	S	3	7	n.m.	2000	n.m.	T	
Hickman and Blume (2001)	D	Day	n.m.	3500	n.m.	d.s.a	B	Houston (USA)
Pei et al. (2019c)	S	1	n.m.	120	4	9	B	Guangzhou (China)
Pei et al. (2019a)	S	Day	75	702	1	162	B	Guangzhou (China)
Quadrifoglio et al. (2007)	D	1	∞	2-25	n.m.	d.s.a	B	LA (USA)
Inturri et al. (2018)	n.m.	n.m.	n.m.	n.m.	0 - 30	d.s.a	B	Ragusa (Italy)
Many-to-one fully flexible DON-PBS								
Perera et al. (2018b)	S	n.m.	8	300	45	1 - 25	T	
Perera et al. (2018a)	S	n.m.	n.m.	10 - 15	2 - 4	5	B	Nanyang Technological University (Singapore)
Perera et al. (2017)	D	n.m.	12	48-144	4-12	d.s.a.	B	Nanyang Technological University (Singapore)
Yu et al. (2015)	S	mp	∞	0 - 170	n.m.	50 - 118	B	Austin (Texas, USA)
Li et al. (2018b)	S	n.m	200	125 - 1750	∞	16	B	Chongqing (China)
Koh et al. (2018)	S	4	30-90	3700	24	n.m.	B	(Singapore)
Shen et al. (2017)	S	1	36-60	41	n.m.	11	B	Nanjing (China)
Wang et al. (2020)	D	1	15	226	4- 16	d.s.a	T	
Many-to-one semi-flexible DON-PBS								
Pratelli et al. (2018)	S	Day	n.m.	38 - 147	n.m.	n.m.	B	Campi Bisenzio (Italy)

(S/D = Stop-based approach or door-to-door approach, PH = Planning horizon (in hours), C = Capacity vehicles, R = Number of requests, V = Number of vehicles, S = Number of stops, N = Type of network used, T = Theoretical, B = Network based on real city, n.m. = not mentioned in the paper, d.s.a.= does not apply to this work, mb = minibus-size, mp = morning peak hour)

Table 17: Instance scale of DOFF-PBS literature

Reference	S/D	PH	C	R	V	S	N	Location
Many-to-many fully flexible DOFF-PBS								
Hadas and Ceder (2008)	S	n.m.	d.s.a.	n.m.	d.s.a.	14	T	
Fatnassi et al. (2015)	S	2h	n.m.	1900	80-400	19	B	Corby (England)
Masoud et al. (2017)	S	3h	4	1000	1000	1000	B	Los Angeles (USA)
Zhao et al. (2018)	S	40 min	2	20-40	30	100-900	T	
Rigas et al. (2018)	S	11h	d.s.a.	400-3000	5-35	8	B	Washington D.C. (USA)
Lotfi et al. (2019)	D	mp	4	4-70	2-25	d.s.a.	B	Dallas (USA)
Huang et al. (2020a)	D	2.5h	15	100-200	30	d.s.a.	B	Nanjing (China)
Winter et al. (2016)	S	1 day	10	1740-1953	d.s.a.	2	B	Wageningen (The Netherlands)
Amirgholy and Gonzales (2016)	D	3h	d.s.a.	150	5	d.s.a.	T	
Guo et al. (2018)	S	1h	30-40	620	12	24	B	Beijing (China)
Ji-Yang et al. (2020)	S	n.m.	7	90	18	18	T	
Many-to-many semi-flexible DOFF-PBS								
Horn (2004)	S+D	24h	n.m.	40000	120	n.m.	B	Queensland (Australia)
Qiu et al. (2014)	S+D	1h	d.s.a.	10-70	n.m.	6	B	Los Angeles (USA)
Crainic et al. (2005)	S	n.m.	d.s.a.	125-250	1	>10	T	
Pei et al. (2019b)	S	17h	75	697	d.s.a.	162	B	Guangzhou (China)
Fu et al. (2003)	S	5h	d.s.a.	3500-7000	n.m.	14	B	Waterloo (Canada)
Quadrifoglio et al. (2008)	S+D	n.m.	d.s.a.	3-17	1	10-30	B	Los Angeles (USA)
Kim and Schonfeld (2015)	S+D	24h	20-25	158400	166-268	d.s.a.	T	
Chen and Nie (2017)	S+D	1h	d.s.a.	20000	d.s.a.	n.m.	B	Chicago (USA)
Crainic et al. (2012)	S+D	n.m.	n.m.	300 - 500	n.m.	20 - 50	T	
Gkiotsalitis et al. (2019)	S	6h	d.s.a.	n.m.	220	n.m.	B	The Hague (NL)
Many-to-one fully flexible DOFF-PBS								
Li and Quadrifoglio (2010)	D	1h	d.s.a.	24-48	n.m.	d.s.a.	T	
Sun et al. (2018a)	S	30 min	10	30	3	22	B	Nanjing (China)
Sun et al. (2019a)	D	40 min	10	34	3	3	B	Chongqing (China)
Wei et al. (2020)	S	1h	35	779	23	42	B	Chongqing (China)
Marković et al. (2019)	D	4h	10	20-120	3-20	d.s.a.	B	Washington D.C. (USA)
Quadrifoglio and Li (2009)	D	1h	d.s.a.	0-90	n.m.	d.s.a.	T	
Sun et al. (2018b)	S	2h	12	42	3	15	B	Nanjing (China)
Dou and Meng (2019)	S	30 min	40-120	300	10	10	T	
Liu et al. (2019b)	S	mp +ep	15	418	5-14	4	B	Chengdu (China)
Lee et al. (2019)	D	n.m.	8	317	64	4	T	
Huang et al. (2020b)	D	n.m.	30	50-1500	10-50	d.s.a.	T	
Many-to-one semi-flexible DOFF-PBS								
Kim and Schonfeld (2014)	S+D	30 min	10-16	150	19	5	T	
Qiu et al. (2015b)	S+D	1h	d.s.a.	26-50	1	2	B	Zhengzhou (China)
Lu et al. (2016)	S	n.m.	d.s.a.	39	d.s.a.	20	B	Jinan (China)
Kim and Schonfeld (2013)	D	4-8h	19	38-1200	5-41	d.s.a.	T	
Qiu et al. (2015a)	S+D	1h	d.s.a.	20-65	n.m.	4	B	Salt Lake City (USA)

(S/D = Stop-based approach or door-to-door approach, PH = Planning horizon (in hours), C = Capacity vehicles, R = Number of requests, V = Number of vehicles, S = Number of stops, N = Type of network used, T = Theoretical, B = Network based on real city, n.m. = not mentioned in the paper, d.s.a. = does not apply to this work, mb = minibus-size, mp = morning peak hour, ep=evening peak hour)

Table 18: Instance scale of S-PBS literature

Reference	S/D	PH	C	R	V	S	N	Location
Many-to-many fully flexible S-PBS								
Melis and Sörensen (2020)	S	4	8-16	2000	0-400	121	T	
Stiglic et al. (2018)	S+D	n.m.	2	500-1000	500-1000	49	B	San Francisco (USA)
Luo et al. (2019)	S	2h	n.m.	16-23	3-21	52-74	B	(Hong Kong)
Guo et al. (2019)	S	per	50	≈ 1000	4	36	B	Beijing
Garaix et al. (2010)	S	serv	6	25-100	3-10	d.s.a.	T	
Nourbakhsh and Ouyang (2012)	D	1h	mb	100-5000	80-307	3-35	T	
Chevrier et al. (2012)	D	1h	n.m.	100-1000	n.m.	63	T	
Bakas et al. (2016)	D	per	6	2500	4	d.s.a.	T	
Tong et al. (2017)	S	1week	10	20	3	24	T	
Many-to-many semi-flexible S-PBS								
Hrnčíř et al. (2015)	S	24h	d.s.a.	2 million	d.s.a.	n.m.	B	Yorkshire (UK)
Zheng et al. (2019)	S+D	per	d.s.a.	5-25	d.s.a.	3	B	Los Angeles (USA)
Zhang et al. (2020)	S	per	40	100	d.s.a.	d.s.a.	B	Beijing
Cao and Wang (2017)	S	per	40	d.s.a.	3	4	B	Harbin
Lyu et al. (2019)	D	per	10	n.m.	n.m.	d.s.a.	B	Nanjing
Fu (2002)	D	serv	20	10	1	n.m.	T	
Many-to-one fully flexible S-PBS								
Pan et al. (2015)	S	n.m.	20	130	5	21	B	Jinan (China)
Lee and Savelsbergh (2017)	S	per	5-10	50 - 200	1 - 2	3 - 5	T	
Chien et al. (2001)	D	per	n.m.	≤ 30	n.m.	d.s.a.	T	
Sun et al. (2019b)								
Melachrinoudis et al. (2007)	D	20 days	10-15	58	n.m.	d.s.a.	B	Danvers (USA)
Li and Quadrifoglio (2009)	D	per	n.m.	d.s.a.	d.s.a.	d.s.a.	T	
Li and Quadrifoglio (2011)	D	per	n.m.	d.s.a.	d.s.a.	d.s.a.	T	
Kim and Schonfeld (2012)	D	per	40	d.s.a.	3 - 10	d.s.a.	T	
Wang et al. (2018)	D	per	n.m.	d.s.a.	d.s.a.	d.s.a.	B	Calgary
Chandra and Quadrifoglio (2013)	D	serv	n.m.	d.s.a.	1	d.s.a.	T	
Papanikolaou and Basbas (2020)	S	per	16-22	40 - 100	n.m.	n.m.	B	Lagkadas
Kim et al. (2019)	D	1h	45	54	n.m.	d.s.a.	T	
Shi and Gao (2020)	S	serv	n.m.	1-1000	n.m.	d.s.a.	T	
Uchimura et al. (2002)	D	15min	10-20	n.m.	d.s.a.	d.s.a.	B	Seattle's district
Many-to-one semi-flexible S-PBS								
Zheng et al. (2018)	D	per	n.m.	n.m.	1	n.m.	T	
Galarza Montenegro et al. (2021)	S	3h	10-50	12-510	2-26	9-67	T	
Lakatos et al. (2020)	S	per	8	n.m.	1	n.m.	B	Hungary
Mehran et al. (2020)	S	serv	n.m.	≈ 30	1	26	B	Regina

(S/D = Stop-based approach or door-to-door approach, PH = Planning horizon (in hours), C = Capacity vehicles, R = Number of requests, V = Number of vehicles, S = Number of stops, N = Type of network used, T = Theoretical, B = Network based on real city, n.m. = not mentioned in the paper, d.s.a.= does not apply to this work, serv = one service period, per = daily operation period)