Capacity planning for demand-responsive multimodal transit

Bernardo Martin-Iradi[a], Francesco Corman[a], Nikolas Geroliminis[b]

[a] Institute for Transport Planning and Systems, ETH Zurich, Switzerland
bernardo.martin-iradi@ivt.baug.ethz.ch, francesco.corman@ivt.baug.ethz.ch
[b] Urban Transport Systems Laboratory, EPFL, Laussane, Switzerland
nikolas.geroliminis@epfl.ch

Extended abstract submitted for presentation at the 12th Triennial Symposium on Transportation Analysis conference (TRISTAN XII)
June 22-27, 2025, Okinawa, Japan

February 13, 2025

Keywords: Multimodal transit, Public transport, On-demand mobility, Stochastic optimization

1 INTRODUCTION

The rise of digitalization capabilities, the need to address new and changing commuting patterns, and the higher customer expectations underscore the need for new, more efficient and accessible transit systems. One of the emerging transport systems that aim to combine the benefits of conventional public transport (i.e., transit) and on-demand mobility is the On-Demand Multimodal Transit System (ODMTS). An ODMTS is a mixed-transport system where on-demand services complement the existing transit system and all are optimized in a synchronized and integrated manner. ODMTSs have been actively studied in recent years (Steiner & Irnich, 2020, Calabrò *et al.*, 2023) and pilot projects have proven their viability in practice (Van Hentenryck *et al.*, 2023), but modeling and planning these systems at scale remains a challenge. Most of the studies consider different assumptions such as deterministic passenger demands or homogeneous fleets (Banerjee *et al.*, 2021, Sumalee *et al.*, 2011) which can lead to under- and over-utilization of the system capacity and hinder the offered service level to passengers. Adjusting the frequency of the transit lines is one way of adapting the system capacity to the demand. However, this may not be enough, and it is also susceptible to unexpected variations in demand. We believe that including stochastic demand and heterogeneous fleet in the planning of an ODMTS can lead to system-optimum solutions by allocating the system capacity more efficiently, and provide a better level of service while maintaining or even reducing operational costs.

We present a novel two-stage problem formulation for the ODMTS in which we make strategic decisions in the first stage (i.e., system fleet sizing and transit scheduling) and operational decisions in the second stage (i.e., planning of on-demand vehicles and passenger routing) to respond to demand uncertainty. We propose an exact algorithm that leverages both Benders decomposition and column generation to solve large-scale instances. Results from a case study in the city of Zurich show that our method can solve instances with 10 lines and hundreds of requests outperforming commercial solvers. Comparison with unimodal, fleet-homogeneous and deterministic benchmarks, underscore the practical benefits of considering demand stochasticity and fleet heterogeneity. In overall, efficiently planning the system capacity of on-demand multimodal transit systems can provide win-win outcomes toward efficient, equitable and sustainable mobility.

2 TWO-STAGE STOCHASTIC OPTIMIZATION MODEL

We present a two-stage stochastic problem formulation that plans the transit schedule and its required fleet in the first stage, and defines the operations of on-demand vehicles and passenger routes in the second stage.

$$\begin{aligned} \min \quad & \text{Transit fleet and schedule costs} + \text{expected operational and travel costs} && (1)\\ \text{s.t.} \quad & \text{Passenger mode assignment constraints} && (2)\\ & \text{Mode service operating constraints} && (3)\\ & \text{Vehicle capacity constraints} && (4)\\ & \text{Passenger flow constraints} && (5) \end{aligned}$$

2.1 First stage

The first stage problem covers strategical decisions including, (i) which schedule and (ii) with which fleet should we operate the transit system, and (iii) which transit lines and on-demand services serve each origin-destination.

The set of public transport lines (e.g., bus or tram) are formed by a set of stations and operate in both directions. Each line can operate different schedules at different frequencies, and with different vehicles (differing in capacity). We divide the passenger demand in different origin-destination pairs and aggregate the demand flow during the operating time period.

Based on this setup, this first stage minimizes, on one hand, the operational costs of operating the transit schedule, which depend primarily on the frequency (i.e., number of vehicles needed) and vehicle type used, and on the other hand, the a-priori costs of assigning the aggregated origin-destination flow to be served by a specific line, on-demand service, or both. We model all these decisions using binary variables and linear constraints ensuring that each line operates one schedule, and that its capacity can cover the assigned demand.

2.2 Second stage

The second stage tackles the operations of on-demand services and the routing of the demand at the passenger level given the transit schedule defined in the first stage. This stage is divided into different scenarios, each corresponding to a different realization of the passenger demand. Each passenger request in this stage is characterized by an origin, destination, and request time. To capture each passenger's door-to-door trip, we use a time-expanded network where each node represents a time instant, spatial location, and mode of transport. We define multiple sets of arcs defining, walking and waiting time, (off)-boarding of vehicles, and running and dwelling times of the transport modes. A simplified example of such network is shown in Figure 1.

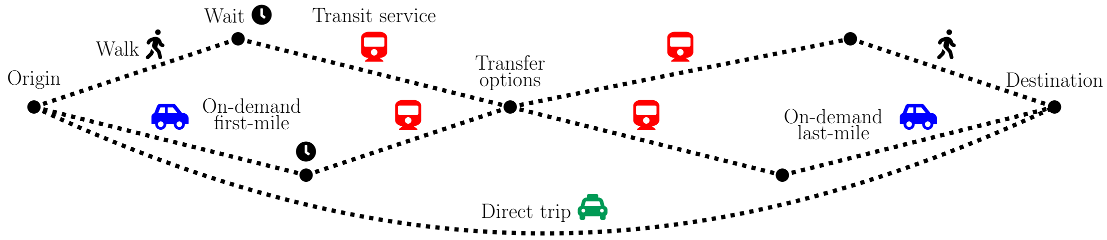

Figure 1 – *Example of a passenger graph where transit legs are shown in red, first and last-mile services in blue, and direct trips in green.*

The model has one set of binary variables depicting the arcs used by each passenger trip. The second stage minimizes the generalized cost of travel for the passengers (Desaulniers & Hickman, 2007) and the operational costs of the on-demand services. We operate on-demand vehicles as a first- and last-mile service or a fully door-to-door trip, and ensure capacity at the vehicle level for all system vehicles. Finally, we link the first and second stage with activating constraints that link the different transit schedules and demand mode choices to the corresponding arc sets in the time-space graph.

3 A DOUBLE DECOMPOSITION ALGORITHM

To solve large-scale instances of the problem, we exploit the decomposability of the model, and in particular, of the second-stage graph. First, we can apply Benders decomposition to the entire two-stage problem. The first-stage problem becomes the Benders Master Problem (BMP) and the second-stage problem becomes the Benders Sub-problem (BSP). Leveraging this decomposition, we can reformulate the second-stage problem as a set partitioning formulation where variables refer to paths instead of arcs in the time-space graph. Each path corresponds to a passenger trip from origin to destination.

Given the complexity of the second stage, enumerating all path-based variables is not tractable. Therefore, we opt to generate them dynamically using column generation. Due to the second-stage reformulation, we only convexify the flow-conservation constraints for passenger trips, meaning that the pricing problem (PP) is a shortest path problem in a directed and acyclic graph (DAG). Therefore, we can use efficient label-setting algorithms to solve them. We acknowledge that, when the PP is a pure shortest path problem, column generation does not provide a benefit in terms of root node relaxation (i.e., solving the root node with column generation and solving the relaxed version of the original formulation provides equal bounds), but we foresee to solve the root node significantly faster.

Once the column generation and Benders decomposition procedures converge, if the solution is still fractional, we solve the integer version of the BSP to guarantee an integer feasible solution. The BSP formulation is tight in itself, and therefore, we expect the integrality gap to also be near or fully optimal.

Finally, the model's solution indicates the number of first- and last-mile services to be operated. Based on the vehicle shareability network (VSN) from Santi *et al.* (2014), we employ a post-process to determine services that can be served simultaneously using the same vehicle, and we further combine them into vehicle itineraries to compute a valid and efficient final fleet of on-demand vehicles.

4 COMPUTATIONAL RESULTS

To assess the value of the model and method, we define a case study based on Zurich's tram and bus network. The instances comprise 3-10 lines, and 5-20 demand scenarios with hundreds of hourly requests.

We compare our double-decomposition algorithm with the direct model solved by Gurobi to assess the computational performance. While Gurobi can only find solutions to the smallest instances (i.e., 3 lines and 5 scenarios) within an 10 hour time limit, our double-decomposition method, achieves high-quality solutions for the largest instances (i.e., 10 lines and 20 scenarios) underscoring the value of the dynamic row and column generation processes.

Table 1 summarizes the modelling and practical benefits of our problem measured in different values, and we break down each value for each stakeholder (i.e., transit and on-demand operators, and passengers). The value of the stochastic solution (VSS) is measured by comparing our model to a deterministic equivalent, and we observe that considering more demand scenarios can lead to up 10% objective improvement and Pareto-optimal solutions to all stakeholders. We compute

Table 1 – *Value of Stochastic Solution (VSS), Value of Multi-modality (VMM), Value of Fleet Heterogeneity (VFH), and vehicle shareability analysis.*

		Transit costs (%)	On-demand costs (%)	Passenger costs (%)
Average # scenarios	$\lvert\frac{\text{VSS}}{Sol.}\rvert$	VSS breakdown		
5	3.9	9.9	-5.0	-4.0
10	7.7	18.6	-9.1	-7.6
20	10.3	-3.8	-11.4	-9.5
Average # lines	$\lvert\frac{\text{VMM}}{Sol.}\rvert$	VMM breakdown		
3	6.4	38.5	-7.7	-6.8
5	5.7	30.0	-8.5	-6.4
10	5.7	20.6	-7.4	-6.0
	$\lvert\frac{\text{VFH}}{Sol.}\rvert$	VFH breakdown		
Vs. homogeneous fleet	0.5	-10.4	-0.5	0.0
On-demand vehicle shareability analysis	Sareable F/L mile services (%)	Shared trips (%)	Service to fleet reduction (%)	Trip to fleet reduction (%)
	52	39	-54	-42

the value of multi-modality (VMM) by comparing it to a uni-modal approach (i.e., either on-demand or transit trips) and observe that using a mixed approach can improve solutions up to 6%, especially for on-demand operators and passengers. We also observe that the VMM is not affected by the network size, but rather by the additional flexibility that on-demand services provide for each line. We also compare the model using both a heterogeneous and a homogeneous transit fleet, and we observe that the value of fleet heterogeneity (VFH) does not have a significant impact on the overall objective but can significantly reduce transit costs. Finally, we measure the value of using shareability networks to cluster and optimize the on-demand fleet and observe that we can significantly reduce the fleet size without sacrificing passenger level of service.

Possible future research directions include scaling the algorithm to cover large-city networks, and investigating cost allocation mechanisms to incentivize the participation of all stakeholders.

In overall, the results suggest that ODMTS can become a relevant solution in the current mobility ecosystem and provide benefits to users, operators and society as a whole.

References

Banerjee, Siddhartha, Hssaine, Chamsi, Luo, Qi, & Samaranayake, Samitha. 2021. Plan your system and price for free: Fast algorithms for multimodal transit operations. *Available at SSRN 3972423.*

Calabrò, Giovanni, Araldo, Andrea, Oh, Simon, Seshadri, Ravi, Inturri, Giuseppe, & Ben-Akiva, Moshe. 2023. Adaptive transit design: Optimizing fixed and demand responsive multi-modal transportation via continuous approximation. *Transportation Research Part A: Policy and Practice*, **171**, 103643.

Desaulniers, Guy, & Hickman, Mark D. 2007. Public transit. *Handbooks in operations research and management science*, **14**, 69–127.

Santi, Paolo, Resta, Giovanni, Szell, Michael, Sobolevsky, Stanislav, Strogatz, Steven H, & Ratti, Carlo. 2014. Quantifying the benefits of vehicle pooling with shareability networks. *Proceedings of the National Academy of Sciences of the United States of America*, **111**(37), 13290–13294.

Steiner, Konrad, & Irnich, Stefan. 2020. Strategic Planning for Integrated Mobility-on-Demand and Urban Public Bus Networks. *Transportation Science*, **54**(6), 1616–1639.

Sumalee, Agachai, Uchida, Kenetsu, & Lam, William H.K. 2011. Stochastic multi-modal transport network under demand uncertainties and adverse weather condition. *Transportation Research Part C: Emerging Technologies*, **19**(2), 338–350.

Van Hentenryck, Pascal, Riley, Connor, Trasatti, Anthony, Guan, Hongzhao, Santanam, Tejas, Huertas, Jorge A., Dalmeijer, Kevin, Watkins, Kari, Drake, Juwon, & Baskin, Samson. 2023. MARTA Reach: Piloting an On-Demand Multimodal Transit System in Atlanta.