Deviated Fixed-route Microtransit: Design and Operations

Bernardo Martin-Iradi[a], Alexandria Schmid[b,*], Kayla Cummings[d], Alexandre Jacquillat[b,c]

[a] Institute for Transport Planning and Systems, ETH Zürich, Zürich, Switzerland
[b] Operations Research Center, MIT, Cambridge, MA, USA
[c] Sloan School of Management, MIT, Cambridge, MA, USA
[d] Microsoft, Seattle, WA, USA
[*] Corresponding author

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

March 28, 2025

Keywords: Microtransit, stochastic optimization, Benders decomposition, column generation

1 INTRODUCTION

Major cities face critical challenges to meet mobility needs in the midst of rising congestion, greenhouse gas emissions and socioeconomic inequalities. Static transit infrastructure offers limited flexibility to respond to ever-changing mobility needs, resulting in a ridership decline and transit deserts. Simultaneously, ride-sharing provides flexible, on-demand mobility services, but low-occupancy vehicles still lead to high fares, congestion, and emissions. This context identifies opportunities to leverage emerging *microtransit* services toward efficient, equitable, and sustainable mobility. Broadly defined by the US DoT (2016) as "privately owned and operated shared transportation system(s) that can offer fixed routes and schedules, as well as flexible routes and on-demand scheduling," microtransit shepherds the digital capabilities and operating flexibility of ride-sharing into the realm of public transit. Yet, microtransit raises critical questions about how to combine transit and ride-sharing components into low-cost, high-quality services and how to develop corresponding routing capabilities, thus requiring dedicated analytics and optimization capabilities.

One possible microtransit model is to provide on-demand door-to-door transportation with high-capacity vehicles. However, Blanchard *et al.* (2023) showed that the optimal latency in a probabilistic traveling repairman problem grows with the dispersion of customers, and also grows at a supra-linear rate of $\Theta(n\sqrt{n})$ where n is the number of customers. The convexity of this function reflects negative externalities across customers in on-demand high-capacity operations, leading to detours, delays and long travel times.

This result outlines two approaches to ensure a high level of service with high-capacity vehicles: zone-based microtransit and deviated fixed-route microtransit. Zone-based microtransit restricts the geographic locale of on-demand operations (e.g., in Miami and Los Angeles). To serve longer trips, it often acts as a first- and last-mile feeder into fixed-route transit (e.g., in Dallas and Atlanta). Deviated fixed-route microtransit, in contrast, relies on transit lines to consolidate passenger demand into high-capacity vehicles—as in public transit—while allowing on-demand routing deviations to provide convenient mobility options in response to passenger requests—as in ride-sharing. It has been subject to limited experimentation and to limited academic research. In response, this paper aims to guide the design and operations of deviated fixed-route microtransit systems, in collaboration with transit operators toward pilot deployment and experimentation.

This paper proposes a scalable two-stage stochastic optimization methodology to support the design of deviated fixed-route microtransit systems at the strategic level and their operations at the tactical level. Specifically, the paper jointly optimizes network design (which reference lines to operate), service scheduling (frequency and timetable), and on-demand deviations (how to serve on-demand passenger requests). The main goals of the paper are to establish the scalability of the methodology to large instances arising in practice, and to assess the performance of deviated fixed-route microtransit in the urban mobility ecosystem.

2 MODEL FORMULATION

We formulate a *Microtransit Network Design (MiND)* model via two-stage stochastic optimization. The first-stage problem selects *reference trips*, each encapsulating a *reference line* and a service frequency. The second-stage problem reflects on-demand routing deviations to serve passenger requests. The model features a multi-objective structure to minimize planning costs, maximize ridership and maximize a passenger level of service metric encapsulating walking, waiting, in-vehicle travel, and arrival delays. We focus primarily on a MiND-VRP problem, corresponding to a vehicle routing setting in which all passengers have the same origin or the same destination, motivated by use cases such as airport shuttles. We extend the methodology and main results to a MiND-DAR problem, corresponding to a more complex dial-a-ride setting in which passengers request transportation from origin to destination.

The MiND features an adaptive optimization structure with two challenging discrete optimization problems: network design and capacitated vehicle routing with time windows. To retain a tight recourse formulation, we propose a subpath-based representation of second-stage microtransit operations in a load-expanded network, in which each node encodes a checkpoint on the reference line and a vehicle load, and each arc characterizes on-demand operations between checkpoints. Load expansion accommodates vehicle capacities without big-M constraints, leading to a continuous recourse function approximation. We show that our subpath-based variables enable a more effective formulation than a segment-based benchmark with variables connecting consecutive stops (by integrating time windows in the definition of subpaths without involving a time-load-expanded network) and than a path-based benchmark with variables connecting the start to the end of each transit line (by drastically quelling the rate of exponential growth in the number of variables).

3 DOUBLE-DECOMPOSITION ALGORITHM

We solve large-scale instance of MiND with a double-decomposition algorithm combining Benders decomposition and subpath-based column generation. The Benders decomposition scheme iterates between a first-stage network design problem and second-stage routing problems, exploiting the nested block-diagonal structure to decompose on-demand operations for each reference trip in each scenario. The column generation scheme adds subpath-based variables iteratively in the Benders subproblem. We develop exact and heuristic label-setting algorithms to generate subpaths of negative reduced cost while keeping track of vehicle load and level of service. As compared to typical combinations of Benders decomposition and (path-based) column generation, our modeling and algorithmic approach induces a double-decomposition structure: the column generation pricing problem adds subpaths between checkpoints, the Benders subproblem combines them to optimize the operating performance of each reference trip in each scenario, and the Benders master problem selects reference trips to optimize the overall network (see Figure 1).

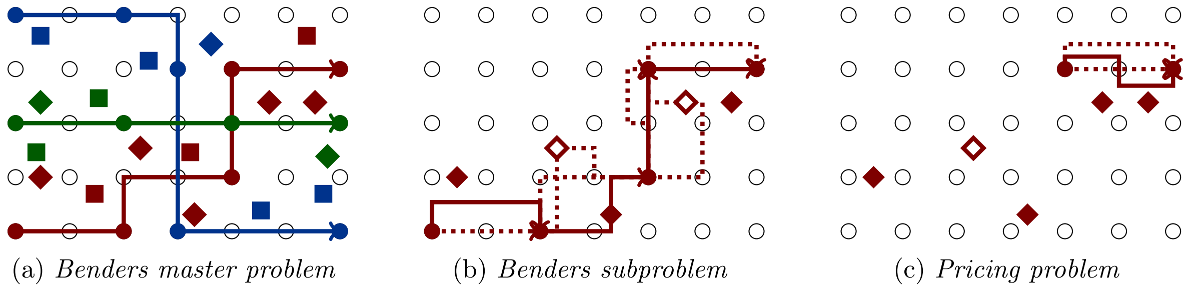

Figure 1 – *Double-decomposition algorithm.* **Left:** *Benders master problem with three reference lines (blue, red, green); passenger requests in two scenarios (squares, diamonds) with their first-stage assignments (colors).* **Middle:** *Benders subproblem for one reference trip and one scenario; full diamonds encode passengers served; solid lines characterize selected subpaths.* **Right:** *Pricing problem to generate new subpath between checkpoints (solid line).*

4 RESULTS

We develop a real-world setup in Manhattan using demand from the NYC Taxi & Limousine Commission (2021) during the morning rush (6–9 am). We consider a MiND-VRP setting corresponding to a shuttle service from Manhattan to LaGuardia Airport with vehicles of capacity 10 to 20 passengers. We vary the number of candidate reference lines (5 to 100), the planning horizon (1 to 3 hours), whether on-demand deviations can skip a checkpoint, and the number of scenarios (5 to 20).

A comparison to path-based and segment-based benchmarks shows that our subpath-based formulation requires orders of magnitude fewer variables, terminates much faster when solved with off-the-shelf methods, and returns a superior solution, scaling up to 10 candidate lines. Yet, no formulation scales to larger instances, which motivates our double-decomposition algorithm. We evaluate solution quality and computational time for our algorithm on larger-scale instances in Table 1, comparing Benders decomposition with full subpath enumeration ("Benders") to our double-decomposition algorithm with the exact and heuristic label-setting algorithms ("DD-Exact" and "DD-Heuristic").

Results show the combined benefits of Benders decomposition, subpath-based column generation, and our label-setting algorithm toward solving large-scale and otherwise-intractable instances. In particular, our double decomposition algorithm scales to instances of the size of the full Manhattan network with up to 100 candidate lines, hundreds of stations, thousands of passenger requests, and 5-20 demand scenarios, while standard Benders decomposition remains intractable in realistic instances because of the full subpath enumeration. Our methodology can handle much larger instances than previous approaches to microtransit operations, while also adding a first-stage design layer under demand uncertainty. Additional experiments show the practical benefits of our integrated stochastic optimization methodology, with a value of the stochastic solution of 5–7% against a deterministic benchmark.

Finally, we derive evidence that deviated fixed-route microtransit can provide win-win outcomes toward efficient, equitable and sustainable mobility. Table 2 compares microtransit to fixed-route transit and ride-sharing benchmarks for various system designs. As compared to ride-sharing, microtransit consolidates demand into high-capacity vehicles along reference lines, as evidenced by lower distance traveled per passenger. As compared to fixed-route transit, it achieves higher demand coverage and comparable levels of service by leveraging on-demand routing flexibility. In turn, the optimized microtransit network has a broader catchment area than its fixed-route counterpart, thus enhancing accessibility in otherwise-unserved regions. Finally, demand consolidation and high coverage result in a significant decrease in distance traveled per passenger, which yields environmental benefits and can enable more affordable on-demand mobility options. Altogether, deviated fixed-route microtransit can contribute to efficient (high

Table 1 – *Algorithm comparison for the subpath-based MiND-VRP model.*

			No skipped checkpoints								Skip up to 1 checkpoint				
			Benders			DD-Exact			DD-Heuristic		DD-Exact			DD-Heuristic	
Lines	Scenarios	Horizon	Sol.	Gap	CPU(s)	Sol.	Gap	CPU(s)	Sol.	CPU(s)	Sol.	Gap	CPU(s)	Sol.	CPU(s)
5	5	60	100	**0.0**	24	100	**0.0**	13	103.3	8	100	**0.0**	242	101.2	18
		120	100	**0.0**	325	100	**0.0**	56	101.8	26	100	**0.0**	5,753	101.5	45
		180	100	**0.0**	369	100	**0.0**	67	102	51	100	**0.0**	4,395	101	66
	20	60	│	│	│	100	**0.0**	94	102.6	39	100	**0.0**	3,536	100.9	85
		120	│	│	│	100	**0.0**	478	102.4	150	100	2.3	10,800	100.6	211
		180	│	│	│	100	**0.0**	529	102.3	230	100	3.3	10,800	100.5	306
10	5	60	100	**0.0**	48	100	**0.0**	82	102	57	100	**0.0**	6,222	100.3	75
		120	│	│	│	100	**0.0**	256	100.8	121	102.5	6.5	10,800	100	187
		180	│	│	│	100	**0.0**	407	101.1	200	104.4	10.7	10,800	100	280
	20	60	│	│	│	100	**0.1**	789	102	328	108.7	31.3	10,800	100	404
		120	│	│	│	101.2	4.4	10,800	100	1,228	│	│	│	100	912
		180	│	│	│	103	9.9	10,800	100	2,782	│	│	│	100	1,247
50	5	60	│	│	│	100	**0.2**	2,093	100.1	649	│	│	│	100	10,800
		120	│	│	│	100.6	3.2	10,800	100	10,800	│	│	│	100	10,800
		180	│	│	│	100	6.8	10,800	100.4	10,800	│	│	│	100	10,800
	20	60	│	│	│	104.6	7.3	10,800	100	10,800	│	│	│	100	10,800
100	5	60	│	│	│	100.9	1.1	10,800	100	2,802	│	│	│	100	10,800
		120	│	│	│	105.9	9.9	10,800	100	10,800	│	│	│	100	10,800
		180	│	│	│	│	│	│	100	10,800	│	│	│	100	10,800
	20	60	│	│	│	│	│	│	100	10,800	│	│	│	│	│

Optimality gap: integer MiND-VRP solution vs. lower bound from Benders decomposition, in percentage terms. [values in bold indicate that the algorithm has converged; for others, the algorithm reached the time limit.]

Table 2 – *Average level of service of fixed-route transit, microtransit, and ride-sharing.*

Mode	Design	Coverage	Walk	Wait	Detour	Delay	Distance
Fixed-route Transit	5 candidate lines	13.9%	2.06	7.06	158.56%	-1.17	356
	10 candidate lines	20.4%	2.21	6.91	146.22%	-0.79	384
	25 candidate lines	29.8%	2.03	6.80	136.98%	0.13	435
	50 candidate lines	33.6%	2.03	6.65	137.34%	-0.06	472
Microtransit	5 candidate lines	22.3%	1.68	6.22	159.99%	-0.01	419
	10 candidate lines	30.0%	1.68	6.22	146.11%	-0.15	462
	25 candidate lines	35.6%	1.53	5.82	138.52%	-0.16	471
	50 candidate lines	36.6%	1.36	5.55	141.00%	0.03	468
Rideshare	Vehicle capacity 4	36.3%	0	4.20	150.68%	13.4	1,883
	Vehicle capacity 2	44.7%	0	3.74	124.60%	8.17	3,359
	Vehicle capacity 1	50.5%	0	1.79	100.00%	1.79	5,671

Coverage: percentage of served requests; distance in kilometers; walk, wait, delay/earliness in minutes.

demand coverage, low operating costs per passenger, high service levels), equitable (broad geographic reach), and sustainable mobility (limited environmental footprint). These results have inspired ongoing collaborations with transit operators toward the pilot deployment of deviated fixed-route microtransit, based on the methodology from this paper.

References

Blanchard, Moïse, Jacquillat, Alexandre, & Jaillet, Patrick. 2023. Probabilistic bounds on the $k-$Traveling Salesman Problem and the Traveling Repairman Problem. *Mathematics of Operations Research.*

NYC Taxi & Limousine Commission. 2021. *TLC Trip Record Data.* Available at: https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page.

US DoT. 2016. *Shared Mobility Current Practices and Guiding Principles.* Tech. rept.