TAXI2: Joint Pairing and Vehicle Assignment for Two-Passenger Shared Taxis

Extended Abstract

Afzaal Hassan
Swinburne University of Technology
Melbourne, Australia
afzaalhassan@swin.edu.au

Irene Moser
Swinburne University of Technology
Melbourne, Australia
imoser@swin.edu.au

Mark Wallace
Monash University
Melbourne, Australia
mark.wallace@monash.edu

Daniel Harabor
Monash University
Melbourne, Australia
daniel.harabor@monash.edu

ABSTRACT

We study the two-passenger shared taxi problem where a dedicated fleet serves dynamic requests. Existing approaches typically decouple rider matching from fleet assignment by first identifying matches and then dispatching vehicles. This separation often leads to suboptimal fleet utilisation and excessive empty travel. In this work, we present TAXI2, an integer linear programme framework that jointly optimises rider pairing and vehicle assignment in a single decision step. We focus on a flexible service model where modest delays are permitted to maximise pooling efficiency. Using one hour of New York City taxi demand comprising approximately 24,000 requests, we demonstrate that our joint formulation achieves 100% service rates. The results demonstrate that our method efficiently solves city-scale problems within practical runtimes, confirming its suitability for real-time operator deployment.

KEYWORDS

Shared taxi problem, Dynamic ride matching, Joint optimisation, Integer linear programming, Urban mobility

ACM Reference Format:
Afzaal Hassan, Irene Moser, Mark Wallace, and Daniel Harabor. 2026. TAXI2: Joint Pairing and Vehicle Assignment for Two-Passenger Shared Taxis: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/RVZN9394

1 INTRODUCTION AND RELATED WORK

Urban ride-sourcing platforms increasingly promote pooling to mitigate congestion and enhance transport sustainability. While shared taxis utilising small-capacity vehicles for overlapping routes offer a promising operational model, their efficacy hinges on real-time rider grouping and fleet allocation. Conventional research frequently decouples passenger matching from vehicle assignment, which often results in systemic inefficiencies and underutilised fleets [5, 12]. Consequently, there is a critical requirement for methodologies that jointly optimise these processes at a city-wide scale under realistic temporal constraints.

Existing literature frequently prioritises scalability over optimality, relying on heuristics or decoupled processes to manage the computational complexity of the problem. Although capacity-two pooling is classified as NP-hard [2], empirical evidence suggests it offers significant potential for reducing total travel distance and enhancing urban mobility [11]. However, current exact models are largely confined to small-scale instances [8], leaving a distinct gap for city-scale solutions that maintain rigorous optimality.

This paper addresses this deficiency by introducing *TAXI2*, a framework designed for the two-passenger shared taxi problem using a dedicated fleet. We propose an integer linear program (ILP) that simultaneously selects feasible rider pairs and assigns vehicles to minimise vehicle hours travelled (VHT) while maximising service rates. By integrating these components into a joint formulation, the *TAXI2* framework achieves practical runtimes for large-scale applications without compromising solution quality.

2 METHODOLOGY

2.1 Overview

We address the two-passenger shared taxi problem using a two-phase framework. In Phase 1, all feasible rider pairs are generated subject to spatio-temporal constraints. A pair is feasible if both riders can be collected and delivered within their earliest departure and latest arrival times under either pick-up order. This yields a match graph whose vertices are riders and whose edges denote feasible pairs. The construction extends our reachability-graph principle for peer-to-peer ridesharing [7] to a shared-taxi fleet.

In Phase 2, an integer linear program (ILP) jointly assigns vehicles to either feasible pairs or solo riders. The objective minimises total vehicle hours travelled (VHT) with penalties for unserved requests, thereby inducing high service rates while preserving linearity. The model incorporates vehicle availability, empty repositioning costs, and disjoint assignment constraints. This unified formulation integrates matching and vehicle assignment, overcoming limitations of decoupled heuristics.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/RVZN9394

2.2 Variables and Parameters

The road network is a graph $G = (V, E)$ with positive edge costs. Shortest-path travel times are precomputed as $w(u, v)$. Riders $r \in R$ have origin o_r, destination d_r, earliest departure t_r^{ed}, and latest arrival t_r^{la}, with earliest arrival $t_r^{ea} = t_r^{ed} + w(o_r, d_r)$. Vehicles $i \in D$ have current location l_i and, when assigned, traverse the shortest route serving their allocated riders; unassigned vehicles remain parked [9].

The ILP minimises total cost over assignments of vehicles to solo riders $j \in U$ or feasible pairs $\langle k, l\rangle \in M$, with penalties c_n for unserved riders $n \in N$:

$$\begin{aligned} \min \quad & \sum_{i \in D} \left(\sum_{j \in U} (cost_{ij} + cost_j)\, x_{ij} \right. \\ & \left. + \sum_{\langle k,l\rangle \in M} (cost_{ik} + cost_{(k,l)})\, y_{i(k,l)} \right) + \sum_{n \in N} c_n \end{aligned} \tag{1}$$

$$\begin{aligned} \text{s.t.} \quad & (1)\ M \subseteq F,\ (N \cup U) \subseteq R \\ & (2)\ R = N \cup U \cup \bigcup_{\langle k,l\rangle \in M} \{k, l\} \\ & (3)\ (N \cup U) \cap \bigcup_{\langle k,l\rangle \in M} \{k, l\} = \emptyset \\ & (4)\ N \cap U = \emptyset \\ & (5)\ \sum_{j \in U} x_{ij} + \sum_{\langle k,l\rangle \in M} y_{i(k,l)} \le 1 \quad \forall i \in D \\ & (6)\ \sum_{i \in D} x_{ij} = 1 \quad \forall j \in U \\ & (7)\ \sum_{i \in D} x_{in} = 0 \quad \forall n \in N \\ & (8)\ \sum_{i \in D} y_{i(k,l)} = 1 \quad \forall \langle k,l\rangle \in M \\ & (9)\ x_{ij},\ y_{i(k,l)} \in \{0, 1\} \\ & (10)\ \langle k,l\rangle \in M \Rightarrow \langle l,k\rangle \notin M \end{aligned}$$

A pair $\langle j, k\rangle$ is feasible if at least one service order satisfies both riders' time windows, allowing waiting at pick-up if arrivals remain on time. The pair cost is the minimum feasible route cost; for an unmatched rider r, $cost(r) = w(o_r, d_r)$. The objective accounts for both in-vehicle and empty repositioning travel.

2.3 Dynamic Demand and Snapshot Optimality

TAXI2 operates in rolling time windows under snapshot optimality: each window is solved to optimality using current requests and vehicle states. Vehicles advance along assigned routes, and riders are picked up or dropped off as nodes are reached. Capacity constraints enforce continuity for riders already onboard. Routes are then issued to assigned vehicles for solo or paired service, with drop-off order determined by feasibility and cost.

Assignments follow an *Eager* policy, initiating trips at the earliest feasible time. Prior work shows comparable service quality to *Lazy* dispatch with substantially lower computational cost [6]. Our model permits minor delays, this is justified by empirical evidence suggesting that travellers typically accept modest detours in exchange for reduced fares [1, 4]. This flexibility allows operators to significantly improve service rates and fleet efficiency.

3 EXPERIMENTS

3.1 Setup

We evaluate *TAXI2* using the New York City taxicab dataset [3], extracting ∼24,000 peak-hour requests (08:00–09:00) within Manhattan. The network comprises ≈4,500 nodes, all serving as potential door-to-door trip endpoints. Edge costs are calculated as free-flow travel times, determined by dividing edge length by maximum speed, consistent with established methods [10]. A fleet of 2,000 dual-capacity vehicles is initialised with random starting positions. The simulation is implemented in Java, with the ILP formulated in MiniZinc and solved via Gurobi.

3.2 Results and Discussion

We evaluate the framework using varying levels of *slack*—defined as the additional travel time permitted beyond the shortest-path duration (20%–50%)—which facilitates pooling detours. Table 1 demonstrates that increasing slack enhances pairing opportunities, significantly reducing total VHT. Notably, as slack increases, the proportion of delayed riders and the average delay magnitude decrease, suggesting that greater temporal flexibility allows the optimiser to find more efficient global assignments. These results confirm that our method solves city-scale problems within practical runtimes, demonstrating its suitability for real-time operator deployment.

Table 1: Flexible assignment results (2,000 vehicles).

Slack	Delayed (%)	Delay (%)	Service (%)	Solve (ms)	VHT
20%	17.4	7.97	100	37,295	1853
30%	14.7	5.10	100	67,046	1661
40%	13.7	4.74	100	106,919	1578
50%	11.5	4.90	100	140,080	1536

4 CONCLUSION AND FUTURE WORK

This paper introduced *TAXI2*, an ILP-based framework for the two-passenger shared taxi problem. We demonstrated that jointly optimising rider pairing and vehicle assignment is computationally tractable at city scale when the service model permits modest delays. Our experiments on New York City data show that this flexibility allows for 100% service rates and efficient fleet utilisation within a rolling-horizon framework.

While the current approach balances efficiency with service coverage, real-world operations often require stricter reliability guarantees. Enforcing hard deadlines, where no delays are permitted, significantly increases the computational complexity of the assignment problem, potentially limiting scalability. Future work will extend this framework to address strict temporal feasibility. Specifically, we aim to investigate advanced match-pruning heuristics that can reduce the search space of the ILP, enabling the system to meet hard deadlines in real-time without compromising the optimality of the solution.

REFERENCES

[1] María J Alonso-González, Oded Cats, Niels van Oort, Sascha Hoogendoorn-Lanser, and Serge Hoogendoorn. 2021. What are the determinants of the willingness to share rides in pooled on-demand services? *Transportation* 48, 4 (2021), 1733–1765.

[2] Xiaohui Bei and Shengyu Zhang. 2018. Algorithms for trip-vehicle assignment in ride-sharing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

[3] Brian Donovan and DB Work. 2014. New york city taxi trip data (2010-2013). *Univ. Illinois Urbana-Champaign, Champaign, IL, USA, Tech. Rep* (2014).

[4] Panagiotis Fafoutellis, Eleni G Mantouka, Eleni I Vlahogianni, and Georgeta-Madalina Oprea. 2022. Acceptability modeling of autonomous mobility on-demand services with on-board ride sharing using interpretable Machine Learning. *International Journal of Transportation Science and Technology* 11, 4 (2022), 752–766.

[5] Daniel J Fagnant and Kara M Kockelman. 2018. Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in Austin, Texas. *Transportation* 45, 1 (2018), 143–158.

[6] Afzaal Hassan. 2024. *Snapshot Optimal Ridesharing*. Ph.D. Dissertation. Swinburne.

[7] Afzaal Hassan, Mark Wallace, Irene Moser, and Daniel D. Harabor. 2024. Snapshot-Optimal Real-Time Ride Sharing. *Information* 15, 4 (2024). https://doi.org/10.3390/info15040174

[8] Shijia Hua, Wenjia Zeng, Xinglu Liu, and Mingyao Qi. 2022. Optimality-guaranteed algorithms on the dynamic shared-taxi problem. *Transportation Research Part E: Logistics and Transportation Review* 164 (2022), 102809.

[9] Zhiguang Liu, Tomio Miwa, Weiliang Zeng, Michael GH Bell, and Takayuki Morikawa. 2018. Shared autonomous taxi system and utilization of collected travel-time information. *Journal of Advanced Transportation* 2018 (2018).

[10] Arthur Mahéo, Shizhe Zhao, Afzaal Hassan, Daniel D Harabor, Peter J Stuckey, and Mark Wallace. 2021. Customised Shortest Paths Using a Distributed Reverse Oracle. In *Proceedings of the International Symposium on Combinatorial Search*, Vol. 12. 79–87.

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

[12] Kien Hua Ting, Lai Soon Lee, Stefan Pickl, and Hsin-Vonn Seow. 2021. Shared mobility problems: a systematic review on types, variants, characteristics, and solution approaches. *Applied Sciences* 11, 17 (2021), 7996.