A Model-Based Approach to Vacant Vehicle Routing of a Ride-Sourcing Fleet in Transportation Networks

Guocheng Jiang and Song Gao, University of Massachusetts Amherst

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

Keywords: Ride-Sourcing, Vehicle Routing, Infinite-Horizon Markov Decision Process, Endogenous State Transition Probability, Large Transportation Network

1 RESEARCH GOAL

One major operational question faced by ride-sourcing drivers is where to go for the next passenger. Efficient relocation balances supply and demand, so that passengers can be served promptly and drivers have higher proportions of revenue-earning miles. The current business model of Transportation Network Companies (TNCs) does not permit direct fleet control since drivers are independent contractors, therefore a fleet-level relocation study in the literature generally assumes a fleet of automated vehicles or does not discuss the application context specifically. The current study is motivated by another possible application context, where drivers organize as a cooperative (Conger, 2021) or unionize (Treffeisen, 2024), whose goal does not fully align with that of the TNC's. The TNC generates revenue from commissions charged to all drivers yet do not bear drivers' vehicle ownership or operating costs. It could be argued that drivers are more likely to comply with routing guidance from an entity that represents drivers. Furthermore, the inchoate organization might not possess enough data and be better off with a model-based approach.

We thus take a model-based approach and formulate the optimal vacant ride-sourcing vehicle routing problem for a fleet as an average-reward maximizing, infinite-horizon semi-Markov Decision Process (SMDP) problem in a physical transportation network such that the routing policy is directly operational at a turn-by-turn level. Major inputs are the demand rate and background supply (vacant vehicles outside of the fleet in question) rate at each node, which can come from direct observations or other models. There is not a time stamp in the state and the optimization is performed on a steady-state basis for each separate time period, yet the evaluation of the routing policies is done in a time-dependent setting. The method is intended as an offline planning tool to provide initial guidance to drivers based on long-term demand and supply patterns.

2 CONTRIBUTION

Table 1 provides taxonomy of the literature along three dimensions with selected studies. The first dimension concerns whether system dynamics are modeled (model-based) or not (model-free) in the optimization, and a third hybrid category that combines the previous two (not shown due to space limit). The second dimension concerns whether limited or more elaborate spatial stochastic processes in system dynamics are accounted for. In a model-based setting, the former is typically associated with a lack of spatial friction in passenger-vehicle matching, e.g., the point queue at a station (Zhang & Pavone, 2016), and the omission of matching possibilities along the way of moving to the assigned relocation destination (Braverman *et al.*, 2019), while the latter group considers these spatial uncertainties from the viewpoint of vehicle agents. In the

	Spatial Stochastic Processes			
	Limited		More Elaborate	
Dynamics	Zone-Based	Network-Based	Zone-Based	Network-Based
Model-Based	Zhang & Pavone (2016)	Iglesias *et al.* (2019)	Zhang *et al.* (2023)	***This study***
Model-Free	Mao *et al.* (2020)		Lin *et al.* (2018)	Yu & Hu (2022)

Table 1 – *Taxonomy of Related Work with Selected Studies*

model-free setting, the former is associated with a single fleet manager agent, which greatly simplifies the learning, while the latter group needs to deal with multiple agents via multi-agent reinforcement learning (RL), mean-field RL or ad hoc adjustments to single vehicle agent based learning. The third dimension concerns the elementary spatial units: zone-based vs. network-based. A network-based method fully considers the spatial connectivity constraints and directly generates turn-by-turn operational policies, in contrast to the ad hoc adjustments needed in zone-based methods.

This study contributes by developing a model-based approach to vacant vehicle routing of a ride-sourcing fleet in a physical transportation network that accounts for spatial friction and sequential matching. A proof is provided for the existence and uniqueness of a stationary distribution for the infinite-horizon SMDP with endogenous state transition probabilities. Computational tests in a large network with around 30,000 links demonstrate its real-world applicability.

3 METHODOLOGY

A fleet of M vehicles travel in a network $\mathcal{G} = (\mathcal{N}, \mathcal{A})$. $\mathcal{N}$ is the set of nodes and $\mathcal{A}$ the set of links. Vacant vehicles' routing decisions are modeled as an infinite-horizon SMDP. An SMDP behaves like an MDP in terms of the state to state transition given an action, and the difference is that the process stays at a state for a certain amount of time, the so-called holding time, with a probabilistic distribution depending on the action and next state, before a transition. The corresponding MDP where the holding time is ignored is called the embedded MDP of the SMDP. Hired vehicles' routing decisions are not explicitly modeled and assumed to follow shortest paths. In the remainder, a vehicle is by default vacant.

The state of a vehicle, $s \in \mathcal{A}$, is the link it is located at, and thus $\mathcal{S} = \mathcal{A}$. All vehicles follow a uniform, probabilistic policy. A probabilistic policy avoids crowding and allows for optimization over the spreading of vehicles at each node.

The action set for state s is the forward star $\mathcal{A}^+(s)$. A parameterized probabilistic policy, π_ϑ, defines $\pi(a|s, \vartheta)$, the probability that action a is chosen given parameters $\vartheta = \{\theta^1, \theta^2, \ldots\}$ with an exogenous feature vector. Example features include passenger arrival rate and background vacant taxi density, and should be customized case by case.

Let m_s be the probability that a vehicle is matched with a passenger at state s. We adopt a reduced form expression for the matching probability as a monotonically increasing function of the ratio of the number of passenger arrivals during the link traversal, $\lambda_s \tau_s$, to the number of eligible vehicles for matching on link s, x_s, that is, $m_s = 1 - \exp\left\{-\alpha \frac{\lambda_s \tau_s}{x_s}\right\}$. The expression for x_s is developed later.

Let $p(u|s, a)$ be the transition probability to state u from s with action a. Two types of transition could happen. 1) The taxi is not matched with any passenger while traversing link s, and thus the next state is the chosen link, that is, $u = a$, with probability $1 - m_s$. 2) The taxi is matched with a passenger when traversing s, and $u \notin \mathcal{A}^+(s)$ is the passenger's destination link. Denote the exogenous destination probability as $q(u|s)$ and the transition probability is $m_s q(u|s)$, with $\sum_{u \notin \mathcal{A}^+(s)} q(u|s) = 1$.

The holding time at state s depends on the action and next state, $t(s, a, u) = \tau_s$, if $u = a$, and $t(s, a, u) = \tau_s + c(s, u)$, if $u \notin \mathcal{A}^+(s)$, where τ_s is the travel time of link s and $c(s, u)$ the shortest path travel time from the end of link s to the end of link u.

We further simplify the notation by defining the expected additional holding time if matched, $\omega_s = \sum_{u \neq a} c(s, u) q(u|s)$. The expected holding time at state s thus can be written as $\mathbb{E}[t_s] = (1 - m_s)\tau_s + m_s(\tau_s + \omega_s) = \tau_s + m_s\omega_s$, during which the vehicle is eligible for matching for the duration τ_s traversing link s.

Let $\mu_s(\vartheta)$ be the stationary distribution at state s of the embedded MDP if all vehicles follow policy π_ϑ, and $\sum_{s \in \mathcal{S}} \mu_s(\vartheta) = 1$. The existence of such a stationary distribution will be discussed later. Under the assumption that such a distribution exists, the long-term fraction of time that the SMDP spends at state s, $\phi_s(\vartheta) = \frac{\mu_s(\vartheta)\mathbb{E}[t_s(\vartheta)]}{\sum_u \mu_u(\vartheta)\mathbb{E}[t_u(\vartheta)]} = \frac{\mu_s(\vartheta)(\tau_s + m_s(\vartheta)\omega_s(\vartheta))}{\sum_u \mu_u(\vartheta)(\tau_u + m_u(\vartheta)\omega_u(\vartheta))}$. ϑ will be dropped in the remainder of the paper when needed to simplify notation.

The long-term fraction of time that the SMDP spends at state s being eligible for matching, ϕ_s^E is the product of ϕ_s and the fraction of the holding time when the vehicle is eligible for matching, $\phi_s^E = \phi_s \frac{\tau_s}{\tau_s + m_s\omega_s} = \frac{\mu_s\tau_s}{\sum_u \mu_u(\tau_u + m_u\omega_u)}$. Therefore, the long-term average number of vehicles eligible for matching at state s is $x_s = M\phi_s^E = \frac{M\mu_s\tau_s}{\sum_u \mu_u(\tau_u + m_u\omega_u)}$. Combining it with the matching probability equation, and we have a fixed point problem, $m_s = -\exp\left\{-\alpha \frac{\lambda_s \sum_u \mu_u(\tau_u + m_u\omega_u)}{M\mu_s}\right\}$. We prove the existence and provide sufficient conditions for the uniqueness of the fixed point, which makes the mapping from μ to m_s a continuous function.

The state does not include a time stamp, and vehicles continue searching after dropping off a passenger. The process has no terminal states either temporally or spatially. Two approaches are typically applied to resolve the infinite-return issue. One is to apply a discount factor to future rewards, so that the return becomes finite. The discount however can be hard to defend considering the within-day context. The other approach is adopted in this study where the average reward per vehicle per time unit, $r_\pi(\vartheta)$ is maximized.

The calculation of the average reward $r_\pi(\vartheta)$ for a given parameter set requires the stationary distribution of vacant vehicles. Although the stationary distribution of an irreducible and aperiodic Markov chain with a constant transition probability matrix is well established, the dependence of transition probabilities on the policy parameter ϑ in this study requires additional theoretical work. We prove the existence of a stationary distribution, $\mu(\vartheta)$ and the corresponding stationary transition probability matrix, $p(u|s; \vartheta)$ by utilizing the Brouwer's fixed-point theorem. We also provide the sufficient condition for the fixed point mapping to be a contraction mapping, that is, when the network structure is regular and the effect of vehicle mass distribution on matching probabilities are moderate. The contraction mapping ensures the uniqueness of a stationary distribution for a given parameter set and thus makes the average reward $r_\pi(\vartheta)$ a proper function. The contraction mapping also leads to the natural choice of a fixed point iteration algorithm to solve for the stationary distribution and the resulting average reward.

The BFGS algorithm is then used to find the optimal parameter values that maximize the average reward. The parameter set is small and thus the memory is not an issue and the memory-efficient L-BFGS is not needed.

4 COMPUTATIONAL RESULTS

A proof-of-concept two-node, four-link (transition to oneself allowed) network case study shows that the optimal policy makes trade-off between moving to a high-demand area and the cost of relocation, while heuristics relying purely on demand distributions do not perform as well. It also shows the divergence of TNC's revenue maximization from an average driver's net income maximization: as the fleet size increases, the total revenue increases, while the average net income decreases.

A case study in a large network with around 10,000 nodes and 30,000 links is conducted. A number of measures are tested to improve the running time, including using parallelized matrix operations, sparse matrices, MPI (message passing interface) to utilize multiple computers of a cluster, and GPU to harness its better capability for parallelism. MPI is not effective as the overhead is large, while the other three measures are effective. The running time is only mildly increasing in the fleet size, thanks to the model-based system dynamics.

The policy is parameterized with two features: passenger arrival rate and background vacant vehicle density. Figure 1 shows that under the optimal policy vacant vehicles tend to stay in and move towards areas with high demand such as the city center and transportation hubs, consistent with intuition (the heat map is grid-based using the sum of stationary distribution over all nodes within each cell). As the fleet size increases, the optimal policy dictates more widely spread vacant taxis and the optimal unit profit decreases, suggesting that competition leads to certain loss of profit. A tabular, deterministic policy obtained from a single taxi routing problem serves as a benchmark. As shown in Table 2, when fleet size is 1, the unit profit of the probabilistic optimal policy is less than that of the benchmark policy due to the parameterization. The optimal policy outperforms the benchmark policy with a large enough fleet size. The improvement grows with the fleet size, reflecting the increasing benefit of taking competition into account as the competition becomes more consequential.

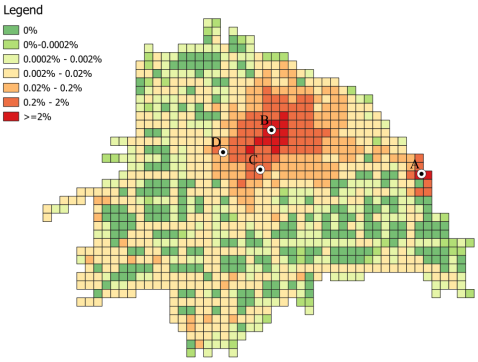

Figure 1 – *Stationary Distribution of Vacant Vehicles with Fleet Size 10000 (7:30am-9:30am)*

Strategy	Fleet Size	Unit profit (CNY/hour) (Occupancy rate)			
		5:30 am to 7:30 am	7:30 am to 9:30 am	9:30 am to 11:30 am	Overall
Optimal probabilistic policy	1	74.2 (0.45)	84.9 (0.52)	88.1 (0.55)	82.4 (0.50)
	5,000	71.3 (0.44)	82.0 (0.50)	84.1 (0.51)	79.1 (0.49)
	10,000	68.8 (0.43)	78.8 (0.48)	81.2 (0.50)	76.2 (0.46)
Deterministic policy	1	75.4 (0.46)	87.9 (0.54)	91.5 (0.56)	84.9 (0.52)
	5,000	66.8 (0.43)	78.6 (0.48)	79.4 (0.48)	74.9 (0.46)
	10,000	60.1 (0.41)	73.1 (0.45)	75.0 (0.45)	69.4 (0.43)

Note: the number in parenthesis is the occupancy rate.

Table 2 – *Average reward of different strategies based on trajectory simulation*

References

Braverman, Anton, Dai, J. G., Liu, Xin, & Ying, Lei. 2019. Empty-Car Routing in Ridesharing Systems. *Operations Research*, **67**(5), 1437–1452.

Conger, Kate. 2021. A Worker-Owned Cooperative Tries to Compete With Uber and Lyft. *The New York Times*, May.

Iglesias, Ramon, Rossi, Federico, Zhang, Rick, & Pavone, Marco. 2019. A BCMP network approach to modeling and controlling autonomous mobility-on-demand systems. *The International Journal of Robotics Research*, **38**(2-3), 357–374.

Lin, Kaixiang, Zhao, Renyu, Xu, Zhe, & Zhou, Jiayu. 2018 (August 19-23). Efficient Large-Scale Fleet Management via Multi-Agent Deep Reinforcement Learning. *In: KDD'18: The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.*

Mao, Chao, Liu, Yulin, & (Max), Shen Zuo-Jun. 2020. Dispatch of autonomous vehicles for taxi services: A deep reinforcement learning approach. *Transportation Research Part C*, **115**, 102626.

Treffeisen, Beth. 2024 (September). *Question 3: What to know about the ballot measure that would allow rideshare drivers to unionize.*

Yu, Zishun, & Hu, Mengqi. 2022. Deep reinforcement learning with graph representation for vehicle repositioning. *IEEE Transactions on Intelligent Transportation Systems*, **23**(8), 13094–13107.

Zhang, Kenan, Mittal, Archak, Djavadian, Shadi, Twumasi-Boakye, Richard, & Nie, Yu. 2023. Ride-hail vehicle routing (RIVER) as a congestion game. *Transportation Research Part B*, **177**, 102819.

Zhang, Rick, & Pavone, Marco. 2016. Control of robotic mobility-on-demand systems: a queueing-theoretical perspective. *The International Journal of Robotics Research*, **35**(1-3), 186–203.