Deep Reinforcement Learning for Passenger Transfer Synchronization in a Bus Network

Jaroonrak Omsap[1] Andri Pradana[3] Ong Yew Soon[2] Chew Lock Yue[3]

Keywords: Transfer coordination, Passenger transfers, Staggered Headway strategy, Reinforcement Learning, Bus network

Extended Abstract

Motivation. Bus transit systems are inherently unstable and highly susceptible to bus bunching, a phenomenon driven by a feedback loop between headway delays and passenger boarding times. This degrades the passenger travel experience, leading to longer average waiting times and missed transfer connections at critical multi-line transfer nodes. While traditional schedule-based and headway-based holding strategies can mitigate these issues on individual services, coordinating multiple services under passenger transfer demand introduces a complex multi-objective optimization problem. Recent work[1] has addressed this problem at the tactical level of bus operations using offline optimization with a min(i)max formulation. The primary objectives of this study are to develop a deep reinforcement learning methodology that learns an online control policy to improve transfer synchronization in a bus network, and to evaluate its performance against a staggered-headway holding baseline without explicit network-level transfer coordination. These objectives are important to transit agencies seeking adaptive control systems that improve transfer convenience while preserving service efficiency.

Approach and Methodology. This study employs a tick-based bus network simulation with event-driven decision epochs, modeling an idealized two-service (Loop A and Loop B) with one transfer node bus network. Two real-time operational control strategies are evaluated: (1) **Staggered-headway holding:** Buses are initialized in an evenly staggered configuration along each loop. After passenger boarding and alighting, a bus is held at stops until the backward spatial gap to the preceding vehicle satisfies a predefined threshold, unless the bus is already full. (2) **Reinforcement learning (RL) control:** A dueling Double DQN[2][3] agent is trained to make discrete hold/go decisions at the transfer node decision epochs (when a bus is ready to depart). At all other bus stops, buses follow the staggered-headway strategy as in the baseline. If the agent selects hold, the bus is held for a fixed duration of 5 simulation ticks, otherwise it departs immediately. The reward is the negative of a weighted congestion cost accumulated over the inter-decision interval, including transfer-related waiting, the age of transfer waiting, total queuing, and onboard passenger load, with a small explicit penalty for choosing hold. Training uses a stricter transfer-delay metric (TD), defined as the time from alighting at the transfer node until connecting bus departs, while evaluation reports both TD and boarding-based transfer wait (TW), defined as

[1]Interdisciplinary Graduate Program, Nanyang Technological University
[2]College of Computing & Data Science, Nanyang Technological University
[3]School of Physical and Mathematical Sciences, Nanyang Technological University

the time from alighting at the transfer node until boarding the connecting bus. Variable decision intervals are handled using a time-aware exponential discounting scheme $\exp(-\beta\Delta t)$, corresponding to a semi-Markov decision process setting.

Results. Over 30 runs with different stochastic passenger-demand realizations (Poisson process) and a fixed staggered initial fleet configuration, evaluated over a 5000-tick horizon, the staggered-headway baseline achieved an average transfer wait time (ATW) of 174.73 ± 30.01 ticks and an average transfer delay (ATD) of 190.82 ± 29.98 ticks. The RL controller substantially improved transfer synchronization relative to the baseline. By strategically applying holding actions at the transfer node, the trained RL agent reduced the average transfer wait time to 78.65 ± 40.75 ticks and the average transfer delay to 114.66 ± 40.47 ticks, while maintaining the average trip time and overall throughput. These results show that an online RL agent can outperform a strong rule-based approach in network-wide transfer coordination, while preserving overall service performance.

Table 1: **Comparison between the staggered-headway baseline without explicit network coordination and the reinforcement learning transfer-node controller.**

Metric	Staggered headway	RL at a transfer node
ATW (ticks, mean ± sd)	174.73 ± 30.01	78.65 ± 40.75
ATD (ticks, mean ± sd)	190.82 ± 29.98	114.66 ± 40.47
Average Trip Time (ticks, mean ± sd)	1030.41 ± 11.42	1031.43 ± 10.35
Completed Trips (trips, mean ± sd)	2548.00 ± 43.17	2566.07 ± 51.10
Throughput (trips per 1000 ticks, mean ± sd)	509.60 ± 8.63	513.21 ± 10.22

Conclusions and Outlook. The RL controller shows strong potential for transfer synchronization in multi-service bus networks. By the time of the conference, additional studies are expected to examine alternative reward formulations and a state-space ablation study in order to better understand the performance gains. Furthermore, we intend to evaluate the robustness of the approach under diverse service structures and passenger-demand patterns.

References

[1] Konstantinos Gkiotsalitis, Oskar A. L. Eikenbroek, and Oded Cats. Robust network-wide bus scheduling with transfer synchronizations. *IEEE Transactions on Intelligent Transportation Systems*, 21(11):4582–4592, 2020.

[2] Hado van Hasselt, Arthur Guez, and David Silver. Deep reinforcement learning with double q-learning. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, page 2094–2100. AAAI Press, 2016.

[3] Ziyu Wang, Tom Schaul, Matteo Hessel, Hado Hasselt, Marc Lanctot, and Nando Freitas. Dueling network architectures for deep reinforcement learning. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1995–2003, New York, New York, USA, 20–22 Jun 2016. PMLR.