Collaborating Without Compromising Privacy: Traffic Signal Control via Vertical Federated Reinforcement Learning

Qiqing Wang[a], Chaopeng Tan[b] and Kaidi Yang[a,*]

[a] National University of Singapore, Singapore, Singapore
qiqing.wang@u.nus.edu, kaidi.yang@nus.edu.sg
[b] Delft University of Technology, Delft, Netherlands
c.tan-2@tudelft.nl
* Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)
September 02-03, 2024, Crete, Greece

April 29, 2024

Keywords: Traffic signal control, data privacy, vertical federated learning, reinforcement learning

1 INTRODUCTION

The boom of real-time trajectory data, enabled by the rapid development of wireless communication technologies, has enormous potential to revolutionize traffic control. Unlike the traditional roadside sensors owned by municipal authorities (MAs) that can only monitor traffic at fixed locations, such trajectory data can provide a holistic picture of the entire transportation network.

To exploit the benefits of trajectory data in the near future, one promising solution is to enable collaboration between MA and mobility providers (MPs), such as mapping and ridesharing companies, whereby MPs leverage the real-time trajectories provided by their fleets to help MA make traffic control decisions. The reason for considering such a solution is two-fold. First, we envision MPs to be the main source of trajectory data before the massive adoption of privately owned connected vehicles (CVs). Due to high vehicle costs and the need for infrastructure upgrades, the penetration rate of privately owned CVs is expected to remain low in the near future, hence not sufficient to enhance traffic control. Second, such a collaboration has been demonstrated effective in both academia and practice (Feng *et al.*, 2015, Moradi *et al.*, 2022, Zheng *et al.*, 2018). For example, DiDi has collaborated with MA in Jinan, China, to improve traffic signal control with its trajectory data, reducing delay by 5-20% (Zheng *et al.*, 2018).

However, one obstacle that may hinder the collaboration is MPs' privacy concerns about sharing their data. On the one hand, MPs can be concerned that the shared trajectory data would disclose sensitive individual mobility patterns of customers, such as origin-destination pairs and routing preferences, which can be used to infer identities, personal profiles, and social relationships (De Montjoye *et al.*, 2013). On the other hand, sharing aggregated MPs' trajectories can leak sensitive information about MPs' operations, such as service coverage, fleet composition, and key algorithm parameters (He & Chow, 2020), which may compromise their competitive advantages and lead to economic losses. To the best of our knowledge, the research on privacy-preserving collaboration between MA and MPs for traffic control is rare. Some existing works attempt to tackle MPs' privacy concerns by outsourcing the decision-making processes to MPs, which do not allow the integration with MA's detector data and hence may not be effective. Other works devise synthetic data generation algorithms that may lead to the loss of data accuracy.

To effectively enable the collaboration between MA and MPs while addressing MPs' privacy concerns, we leverage a promising framework of federated learning (FL), specially designed to

train state-of-the-art learning-based models for distributed data owners without anyone having to explicitly exchange their private data. Despite FL's advantages, its application in traffic control is sparse. Existing frameworks are generally based on *horizontal FL* (Liu *et al.*, 2020), which assumes the datasets of all data owners to be homogeneous. However, the data provided by MA and MPs can be highly heterogeneous with different structures (e.g., traffic counts, fleet trajectories, etc.), so horizontal FL is not suitable for our case.

To fill in these research gaps, we propose a vertical federated reinforcement learning algorithm (herein named **VFedlight**) that combines *vertical* FL (Liu *et al.*, 2022) and reinforcement learning (RL) to enable the collaboration between multiple data owners with heterogeneous data for traffic signal control (TSC). Our contributions are two-fold. First, we initiate the research of designing privacy-preserving traffic control algorithms based on heterogeneous data provided by multiple data owners (i.e., MA and MPs). Second, we develop an effective algorithm for collaborative TSC without compromising the privacy of data owners. To the best of our knowledge, this is the first work that combines RL and vertical FL in transportation research.

2 METHODOLOGY

Consider an isolated intersection including a set of incoming lanes $\mathcal{L}$ and a set of signal phases $\mathcal{P}$. We choose isolated intersections as an initial building block that can be extended to large-scale urban transportation networks in future work. Let us denote the considered time horizon as a set of discrete intervals $\mathcal{T} = \{1, 2, \cdots, T\}$ of a given size Δt. We consider $K+1$ data owners, i.e., MA (indexed by $k = 0$) and K MPs (indexed by $k = 1, \cdots, K$), which are interested in collaborating to fuse their data and develop an RL-based TSC agent to make optimal signal control decisions without compromising privacy. The MA has access to loop detector measurements, and each MP k has access to the real-time trajectories of their fleets.

We formulate the TSC problem as a Markov Decision Process (MDP) $\mathcal{M} = (\mathcal{S}, \mathcal{A}, P, R, \gamma)$, where $\mathcal{S} = \mathcal{S}^0 \otimes \mathcal{S}^1 \otimes \cdots \otimes \mathcal{S}^K$ represents the state space with $\otimes$ indicating the Cartesian product, $\mathcal{S}^0$ the state space of MA, and $\mathcal{S}^1, \cdots, \mathcal{S}^K$ the states of MPs, $\mathcal{A}$ represents the action space, $\mathcal{P}$ represents the system dynamics (i.e., SUMO simulation in this paper), $\mathcal{R}$ represents the reward function, and γ represents a scalar discount factor. The details of the MDP are as follows:

State space. We define MA's state at time step t as $\boldsymbol{s}_t^0 = \left\{p_t, d_t, \{c_{tl}\}_{l \in \mathcal{L}}\right\}$, where $p_t \in \{0, 1\}^{|\mathcal{P}|}$ represents a one-hot vector with the element corresponding to current signal phase being 1, d_t denotes the elapsed green time in the current signal phase, and c_{lt} denotes the traffic count on lane $l \in \mathcal{L}$. We define the state of MP k (following Mo *et al.*, 2022) at time step t as $\boldsymbol{s}_t^k = \left\{n_{tl}^k, d_{tl}^k\right\}_{l \in \mathcal{L}}$, where n_{tl}^k and d_{tl}^k denote the number of vehicles and average delay of MP k's fleet on lane l at time step t, respectively.

Action space. The action at each time step is to decide whether to keep the current phase or to switch to the subsequent phase. If the agent decides to switch, an intergreen time t_y is allocated to ensure safety. We further set a minimum green time t_{min} and a maximum green time t_{max}.

Reward. We define the reward as the change of cumulative delay between two consecutive time steps $r_t = W_{t-1} - W_t$. Here, $W_t = \sum_{i=1}^{N_t} w_{it}$ denotes the cumulative delay at the end of time step t, where N_t represents the cumulative counts of all vehicles that have entered the intersection from time step 1 to time step t, and w_{it} represents the delay of the ith vehicle.

To solve the MDP, we propose a Soft Actor-Critic (SAC) (Christodoulou, 2019) algorithm with vertical FL-based privacy-preserving actor and critic networks (see Figure 1). Such a design allows us to train the RL agent without any party having to exchange raw private states, which preserves the privacy of both MA and MPs. In particular, the actor and critic networks are jointly held by all data owners, whereby each data owner (MA or MP) maintains a local model.

We next use the actor network as an example to describe the integration between vertical FL and RL, while a similar procedure applies to the critic networks. Specifically, MP k computes a local output $\boldsymbol{z}_{t,\mathrm{a}}^k = \boldsymbol{f}_{\mathrm{a}}^k(\boldsymbol{s}_t^k; \boldsymbol{\theta}^k)$ using its private state $\boldsymbol{s}_t^k$, where $\boldsymbol{f}_{\mathrm{a}}^k$ represents its local actor

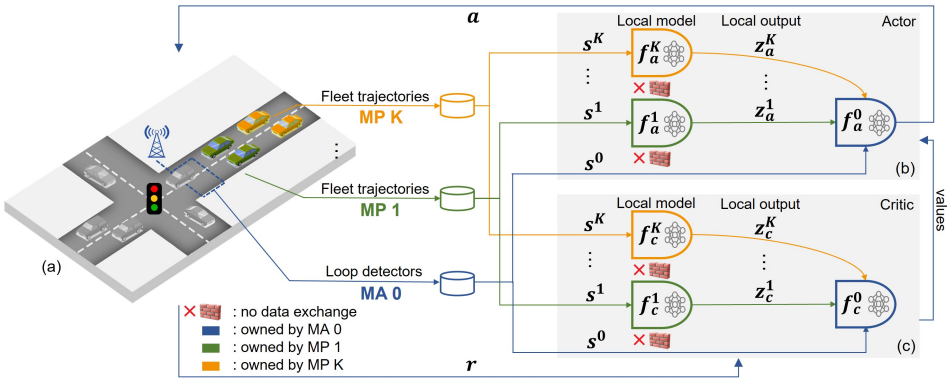

Figure 1 – *The framework of VFedlight: (a) environment consists of multiple private data owners (i.e., MA and MPs), (b) privacy-preserving actor, (c) privacy-preserving critic.*

model parametrized by $\boldsymbol{\theta}^k$. These local outputs from MPs $\left\{\boldsymbol{z}_{t,\mathrm{a}}^k\right\}_{k=1}^K$ are sent to MA, which, in turn, computes the probability of taking each action using its local model $\boldsymbol{f}_{\mathrm{a}}^0\left(\boldsymbol{s}_t^0; \boldsymbol{z}_{t,\mathrm{a}}^1 \ldots, \boldsymbol{z}_{t,\mathrm{a}}^K; \boldsymbol{\theta}^{\mathbf{0}}\right)$ parametrized by $\boldsymbol{\theta}^{\mathbf{0}}$. To facilitate the training of SAC, MA and MPs jointly update their local models using Federated Stochastic Gradient Descent (Fed-SGD) (Liu *et al.*, 2022) without sharing their raw data, which requires the calculation of the gradients of the SAC actor loss function $\ell(\cdot)$ with respect to the parameters of each data owner, i.e., $\frac{\partial \ell}{\partial \boldsymbol{\theta}^k}$. Specifically, MA calculates two types of gradients: (1) the gradients $\frac{\partial \ell}{\partial \boldsymbol{\theta}^0}$ for updating $\boldsymbol{\theta}^0$ and (2) $\frac{\partial \ell}{\partial \boldsymbol{z}_t^k}$ which will be sent to MP k for computing $\frac{\partial \ell}{\partial \boldsymbol{\theta}^k}$ and updating $\boldsymbol{\theta}^k$. Note that the local models are only accessible by data owners themselves for both deployment and training, and the local output dimension is typically much lower than the input dimension, meaning that the private state of each MP is transformed via an unknown transformation with reduced dimensions. The only information being exchanged is the local outputs and the corresponding gradients. It has been proved that adversaries cannot estimate the true states or the local model only from such limited information (Liu *et al.*, 2022).

3 RESULTS

We evaluate the performance of VFedlight via a 30-min SUMO simulation (Lopez *et al.*, 2018) of an isolated intersection during a typical morning peak (see Figure 2 (a)). We compare three algorithms: (1) **Actuated** that uses the default actuated traffic signal control algorithm in SUMO, (2) **SAC** that trains an RL agent using MA and MPs' states without privacy protection, and (3) our proposed **VFedLight** with privacy protection. We simulate the vehicle arrivals of each approach as a time-dependent Poisson process with the arrival rate sampled from a Gaussian distribution, where the mean is calculated as the product of a pre-defined base demand and a scaler shown in Figure 2 (b), and the standard deviation is 10% of the mean value. Left- and right-turn proportions are set to be 10%, respectively. The actor and critic architectures in SAC and the local models in VFedlight are chosen to be Multilayer Perceptron (MLP). The intergreen, minimum, and maximum green time are 3, 10, and 42 seconds, respectively. SAC and VFedlight are trained on 300-600 demand (i.e., the base demand for NS and EW are chosen as 300 and 600 vehicles/h, respectively). The episode rewards during training are shown in Figure 2 (c), which indicates that VFedlight converges to a similar performance as SAC, while still protecting the privacy of data owners.

We then test the trained policies on three demand scenarios (i.e., 300-600, 250-550, 350-650) shown in Table 1. By fusing MPs' trajectory data, the average delay (AD) decreases significantly by 13-37%. Moreover, we notice that the AD and the average queue length (AQL) of VFedlight are very close to that of SAC. This shows that the consideration of privacy comes at a minimal cost for our proposed method. Overall, VFedlight can reduce AD and AQL with satisfactory

control performance without compromising privacy in the collaboration.

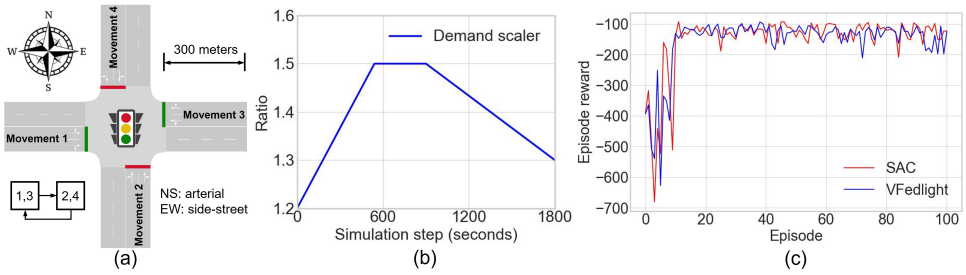

Figure 2 – *(a) An isolated intersection, (b) demand scaler, (c)episode reward during RL training.*

Table 1 – *Policy performance comparison on average delay (AD) and average queue length (AQL)*

	300-600		250-550		350-650	
	AD (second)	AQL (veh)	AD (second)	AQL (veh)	AD (second)	AQL (veh)
Actuated	21.5 (± 5.2)	10.4 (± 3.1)	17.3 (± 3.1)	7.7 (± 1.7)	33.9 (± 6.8)	18.9 (± 4.6)
SAC	15.3 (± 2.8)	6.6 (± 1.6)	13.4 (± 1.1)	5.3 (± 0.6)	21.2 (± 4.0)	10.6 (± 2.4)
VFedlight	17.6 (± 3.1)	8.1 (± 1.8)	15.1 (± 1.5)	6.3 (± 0.9)	24.9 (± 4.8)	13.2 (± 3.0)

4 CONCLUSION

In this paper, we propose VFedlight, a privacy-preserving vertical federated RL approach, for MA and MPs to collaborate on TSC without compromising privacy. Results show that VFedlight can yield satisfactory control performance while protecting MPs' privacy.

References

Christodoulou, Petros. 2019. Soft actor-critic for discrete action settings. *ArXiv Preprint ArXiv:1910.07207*.

De Montjoye, Yves-Alexandre, Hidalgo, César A, Verleysen, Michel, & Blondel, Vincent D. 2013. Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, **3**(1), 1–5.

Feng, Yiheng, Head, K Larry, Khoshmagham, Shayan, & Zamanipour, Mehdi. 2015. A real-time adaptive signal control in a connected vehicle environment. *Transportation Research Part C: Emerging Technologies*, **55**, 460–473.

He, Brian Yueshuai, & Chow, Joseph YJ. 2020. Optimal privacy control for transport network data sharing. *Transportation Research Part C: Emerging Technologies*, **113**, 370–387.

Liu, Yang, Zhang, Xinwei, Kang, Yan, Li, Liping, Chen, Tianjian, Hong, Mingyi, & Yang, Qiang. 2022. Fedbcd: A communication-efficient collaborative learning framework for distributed features. *IEEE Transactions on Signal Processing*, **70**, 4277–4290.

Liu, Yi, James, JQ, Kang, Jiawen, Niyato, Dusit, & Zhang, Shuyu. 2020. Privacy-preserving traffic flow prediction: A federated learning approach. *IEEE Internet of Things Journal*, **7**(8), 7751–7763.

Lopez, Pablo Alvarez, Behrisch, Michael, Bieker-Walz, Laura, Erdmann, Jakob, Flötteröd, Yun-Pang, Hilbrich, Robert, Lücken, Leonhard, Rummel, Johannes, Wagner, Peter, & Wießner, Evamarie. 2018. Microscopic traffic simulation using sumo. *In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE.

Mo, Zhaobin, Li, Wangzhi, Fu, Yongjie, Ruan, Kangrui, & Di, Xuan. 2022. CVLight: Decentralized learning for adaptive traffic signal control with connected vehicles. *Transportation Research Part C: Emerging Technologies*, **141**, 103728.

Moradi, Hossein, Sasaninejad, Sara, Wittevrongel, Sabine, & Walraevens, Joris. 2022. The contribution of connected vehicles to network traffic control: A hierarchical approach. *Transportation Research Part C: Emerging Technologies*, **139**, 103644.

Zheng, Jianfeng, Sun, Weili, Huang, Shihong, Shen, Shengyin, Yu, Chunhui, Zhu, Jinqing, Liu, Bingbing, & Liu, Henry X. 2018. Traffic signal optimization using crowdsourced vehicle trajectory data. *In: Presented at the 97th Annual Meeting of Transportation Research Board, Washington, D.C.*