Learning-based Incentive Design for Eco-Driving Guidance

Jung-Hoon Cho[a,*], M. Umar B. Niazi[a], Siqi Du[b], Tianyue Zhou[c], Roy Dong[b], and Cathy Wu[a]

[a] MIT, Cambridge, MA, USA. {jhooncho, niazi, cathywu}@mit.edu
[b] UIUC, Champaign, IL, USA. {siqidu3, roydong}@illinois.edu
[c] ShanghaiTech University, Shanghai, China. zhouty1@shanghaitech.edu.cn
[*] 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: Sustainable Transportation, Intelligent Transportation System, Incentive Design, Reverse Stackelberg Game, Deep Reinforcement Learning

1 INTRODUCTION

Eco-driving is a practical strategy for mitigating greenhouse gas emissions in urban transportation systems. By integrating eco-driving principles, intelligent transportation systems (ITS) can nudge drivers towards greener and fuel-efficient behavior. This work introduces a novel learning-based incentive design promoting eco-driving within the framework of reverse Stackelberg games. The proposed game is played between a finite set of followers (human drivers) and a leader (system operator). While reverse Stackelberg games have been previously explored in traffic networks, particularly in the context of routing games (Groot *et al.*, 2015), it is typically challenging to model the complex dynamics of traffic, the interaction among drivers, and their responses to incentives. Therefore, we propose an approach where incentives are applied within the space of policy parameters rather than directly influencing specific actions.

Prior studies have highlighted the effectiveness of incentive design in influencing driving behavior (Littlestone & Warmuth, 1994, Niazi *et al.*, 2024). We use the framework of reverse Stackelberg games (Groot *et al.*, 2012) and incorporate learning-based strategies that optimize eco-driving incentives. We further draw inspiration from methodologies that enhance the performance of eco-driving vehicular control (Jayawardana & Wu, 2022). Deep reinforcement learning (RL) has shown promising results in various sequential decision-making problems. In this paper, we employ deep RL to design incentives that could influence human drivers in a transportation network to adopt eco-driving principles. These incentives are designed to ensure compliance by human drivers while adhering to the predetermined budget constraints of a system operator.

The contributions of this paper are twofold: firstly, we present a learning-based incentive design under the framework of reverse Stackelberg games that effectively capture and navigate the complexities of urban traffic dynamics; secondly, we integrate a regret minimization method to accurately model drivers' choices. Our results show the effectiveness of learning-based incentives for adopting eco-driving, significantly reducing overall emissions.

2 PROBLEM DEFINITION

Consider an urban transportation system where a system operator aims to minimize driver tailpipe emissions by incentivizing eco-driving. Let each driver in the system be indexed by i, for $i \in \{1, \ldots, N\}$. The emissions produced by driver i's vehicle is denoted as $x_i(\theta)$ and the corresponding travel time as $y_i(\theta)$, where $\theta = (\theta_1, \ldots, \theta_N)$ represents a set of driving policies chosen by the drivers. The cost to each driver i under policy θ, without any incentive, is given

by $c_i(\theta;\lambda_i) = \lambda_i x_i(\theta) + (1-\lambda_i) y_i(\theta)$, where λ_i is the preference parameter of driver i, reflecting the trade-off between emissions and travel time ($\lambda_i \in [0,1]$). When an incentive γ_i is introduced to driver i, the cost function modifies to $c_i(\theta;\lambda_i,\gamma_i) = (\lambda_i+\gamma_i)x_i(\theta) + (1-\lambda_i)y_i(\theta)$, incentivizing drivers to reduce emissions. Each incentive γ_i is bounded between 0 and $1-\lambda_i$. The planner's objective is to design an incentive scheme $\gamma = (\gamma_1,\ldots,\gamma_N)$ to minimize the total emissions under a budget constraint B, as the following optimization problem:

$$\min_{\gamma} \sum_{i=1}^{N} x_i(\theta) \quad \text{s.t.} \quad \sum_{i=1}^{N} \gamma_i x_i(\theta) \le B \quad \text{and} \quad \theta_i \in \arg\min_{\tilde{\theta}} c_i(\tilde{\theta}, \theta_{-i};\lambda_i,\gamma_i) \quad \text{for } i = 1,\ldots,N. \tag{1}$$

Here, θ_i is the chosen policy from driver i. The constraint ensures that the designed incentives induce an equilibrium where each driver is minimizing their own incentivized cost, subject to the actions of others and within the budgetary limit of B.

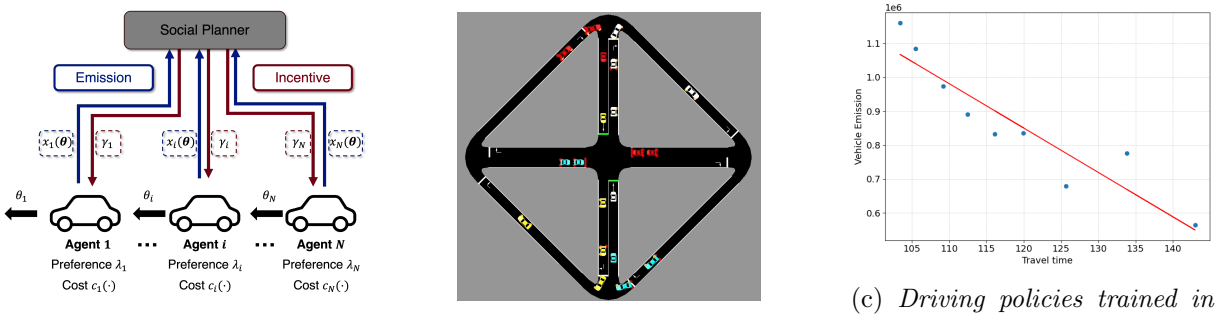
3 METHOD

We present our problem within the construct of a reverse Stackelberg game augmented by a learning-based approach, which unfolds in two sequential stages. In the initial stage, a system operator, the leader, determines an incentive rate γ offered to the drivers. Subsequently, each driver, the follower, must decide on their driving behavior, specifically their acceleration, based on a policy that integrates both their personal preferences and real-time observations.

Reverse Stackelberg Game The problem formulated in the previous section can be formulated as a reverse Stackelberg game between a system operator and N drivers. The system operator acts as a leader with decision variables $\gamma = (\gamma_1,\ldots,\gamma_N)$, where γ_i is an incentive given to driver i to influence their action θ_i. The cost $c_0(\theta_1,\ldots,\theta_N)$ of the system operator is the total emissions $\sum_{i=1}^N x_i(\theta)$ incurred by the drivers subject to the budget constraint $\sum_{i=1}^N \gamma_i x_i(\theta)$ as in (1). The cost of driver i is $c_i(\theta;\lambda_i,\gamma_i)$ as defined earlier. Under the full information case when λ_i's are known to the system operator, the game is played as follows. First, the system operator announces the incentive function $\theta \mapsto \gamma(\theta)$, which is presented to the drivers as a contract. Second, the drivers react to this incentive function and find policies that minimize their respective cost functions. Then, based on the outcome θ of the game, the drivers receive the incentive $\gamma_i(\theta)$ according to the commitment of the system operator. We consider the scenario of adverse selection where the system operator does not know λ_i's of drivers. In this case, the leader learns the incentive function by learning drivers' preferences through repeated interactions.

Deep Reinforcement Learning for Incentive Design For the decision-making of individual drivers, we employ deep RL to obtain a vector of driving policies, denoted by θ, which depends on respective driver preferences λ toward emissions. Each policy θ_j is trained using Proximal Policy Optimization (PPO) (Schulman *et al.*, 2017) to translate the observed state—the position and speed of both the ego vehicle and the surrounding vehicles—into an action that maximizes the expected reward. The reward function depends on preference parameter λ, which impacts drivers' proclivity towards eco-driving. On the other hand, the system operator's incentive design problem is non-trivial because of the complexity of traffic dynamics and vehicular interaction and the uncertainty in the transition matrix of sequential decision-making. To address this issue, we again leverage deep RL to learn the most effective incentive issuance strategy by adapting to both the traffic conditions and driver preferences.

Regret Minimization with Randomized Weighted Majority Algorithm After incentives are issued, drivers find policies that minimize their respective regret functions. To this end, we employ the randomized weighted majority (RWM) algorithm (Littlestone & Warmuth, 1994), which is an online learning method for iterative decision-making processes. The RWM algorithm maintains a set of weights, w_j, corresponding to each policy θ_j. The probability of selecting a policy is proportional to its weight, expressed as $p_j = \frac{w_j}{W}$, where the total weight W aggregates all weights. Following the outcome of each decision, RWM updates the weights to penalize the



(a) *Framework Overview* (b) *Road Network* (c) *Driving policies trained in different reward functions*

Figure 1 – *Overview of the incentive design framework, road network, and trained policies.*

Table 1 – *Results of emission and travel time on different scenarios*

Scenario	Emission	Travel Time	Incentive
Baseline	13.18 ± 0.15	149.01 ± 10.48	0.00
RL-based Incentive	12.87 ± 0.32	122.19 ± 21.37	25.30 ± 4.61
RL-based Incentive with RWM	12.18 ± 0.13	94.60 ± 5.21	25.67 ± 4.56

policies, with the update rule given by $w_j = w_j \cdot (1 - \beta)$ for policies leading to suboptimal outcomes. Through iterative updates, the RWM algorithm ensures that drivers' decisions are continually refined based on the dynamic interactions with other vehicles and traffic.

4 RESULTS

Road Network and Simulation Setup To validate our proposed incentive design, we designed a controlled experiment within a simulated environment: a simple closed-loop road network with a signalized intersection, where the diagonal of the network is fixed at 100 meters. The simulation uses the traffic microsimulation tool, Simulation of Urban MObility (SUMO) platform, which allows for detailed modeling of vehicular dynamics (Lopez *et al.*, 2018). In our experimental setup, illustrated in Figure 1b, we dispatched 15 vehicles within the network, each navigating the rhombus-shaped course and responding to the traffic signals.

Training Driving Policies We trained diverse driving policies on different preferences using the PPO algorithm (Schulman *et al.*, 2017). These policies, trained in 80% of high penetration scenarios, were tailored for a spectrum of driver preferences regarding emission reduction without any incentive, denoted by λ values ranging from 0 to 0.4 in increments of 0.05. To do this, the reward function was designed as $(\text{Reward}) = (1-\lambda)\frac{(\text{Speed})}{(\text{Speed limit})} - \lambda\frac{(\text{Emission})}{(\text{Max Emission})}$ to represent the trade-off between travel time and emission. Figure 1c exhibits a trade-off between travel time and emission. We denote these policies trained with different preference parameters as θ_i.

Learning-based Incentive Design Within the reverse Stackelberg game framework, the social planner's aim is to minimize total emissions, quantified as $\sum_i x_i$, while the followers, the drivers, adjust their driving strategies according to both their personal preferences λ_i and the incentive rates, γ_i, proposed by the social planner. These rates are drawn from a predefined set $\Gamma = \{0, 0.05, 0.1, 0.15, 0.2\}$. Drivers choose their actions based on the minimization of their individual cost functions, $\theta_i = \arg\min c_i(\theta; \lambda_i, \gamma_i)$.

Regret Minimization Adopting the RWM algorithm, we initialized the weights of all policies equally and updated them based on their performance. Each driver retains the individual weight and probability, which was adjusted in accordance with the RWM update rule, thus enabling drivers to make informed decisions that account for both other vehicles' dynamics and the incentives received.

Result Analysis Our findings are numerically summarized in Table 1, which presents the emission levels, travel times, and incentives calculated for the baseline scenario and those augmented

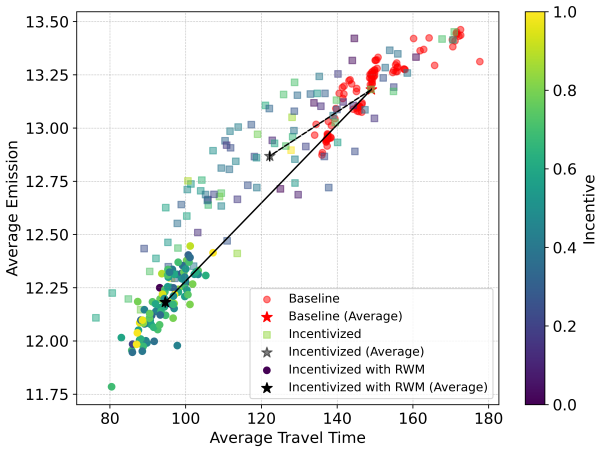


Figure 2 – *Scatter plot comparing average travel time and emissions under baseline and learning-based incentives with and without the Randomized Weighted Majority (RWM), with the average values indicated by stars. The color gradient represents the actual amount of incentive issued.*

with our RWM-based incentive scheme. Notably, the application of RWM yields a significant improvement over the baseline. Additionally, we illustrate these results graphically in Figure 2, where each data point encapsulates the outcome of an individual driving scenario. The color gradient in the scatter plot represents varying incentive levels, with the stark contrast between the baseline and RWM scenarios underscoring the efficacy of our approach. The data delineates a clear trend: implementing the RWM algorithm reduces emissions and effectively manages the incentivization budget. This trend affirms the potential of a learning-based incentive design to significantly elevate urban transportation systems' sustainability.

5 DISCUSSION

This research introduces a novel learning-based incentive design for eco-driving tailored for the reverse Stackelberg game framework. Our method is specifically designed to navigate complex traffic dynamics and uncertainty inherent in sequential decision-making under incomplete information. By integrating regret minimization techniques, the proposed approach demonstrably reduces emissions by influencing driver behavior toward eco-driving. Our results show the potential of this approach to achieve a significant reduction in traffic emissions. This work offers a novel approach for managing urban traffic, while simultaneously transforming how we understand driver behavior in the context of environmental sustainability.

References

Groot, Noortje, De Schutter, Bart, & Hellendoorn, Hans. 2012. Reverse Stackelberg games, Part I: Basic framework. *Pages 421–426 of: 2012 IEEE International Conference on Control Applications.*

Groot, Noortje, De Schutter, Bart, & Hellendoorn, Hans. 2015. Toward System-Optimal Routing in Traffic Networks: A Reverse Stackelberg Game Approach. *IEEE Transactions on Intelligent Transportation Systems*, **16**(1), 29–40.

Jayawardana, Vindula, & Wu, Cathy. 2022. Learning Eco-Driving Strategies at Signalized Intersections. *Pages 383–390 of: 2022 European Control Conference (ECC).*

Littlestone, N., & Warmuth, M.K. 1994. The Weighted Majority Algorithm. *Information and Computation*, **108**(2), 212–261.

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: The 21st IEEE International Conference on Intelligent Transportation Systems.*

Niazi, M. Umar B., Cho, Jung-Hoon, Dahleh, Munther A., Dong, Roy, & Wu, Cathy. 2024. Incentive Design for Eco-driving in Urban Transportation Networks. *In: 2024 European Control Conference (ECC).*

Schulman, John, Wolski, Filip, Dhariwal, Prafulla, Radford, Alec, & Klimov, Oleg. 2017. Proximal Policy Optimization Algorithms. *arXiv:1707.06347.*