Learning to Dissipate Traffic Jams with Piecewise Constant Control

Mayuri Sridhar
MIT EECS
mayuri@mit.edu

Cathy Wu
MIT LIDS, CEE, IDSS
cathywu@mit.edu

Abstract

Greenhouse gases (GHGs), particularly carbon dioxide, are a key contributor to climate change. The transportation sector makes up 35% of CO_2 emissions in the US and more than 70% of it is due to land transport. Previous work shows that simple driving interventions can significantly improve traffic flow on the road. Recent work shows that 5% of vehicles using piecewise constant controllers, designed to be compatible to the reaction times of human drivers, can prevent the formation of stop-and-go traffic congestion on a single-lane circular track, thereby mitigating land transportation emissions. Our work extends these results to consider more extreme traffic settings, where traffic jams have already formed, and environments with limited cooperation. We show that even with the added realism of these challenges, piecewise constant controllers, trained using deep reinforcement learning, can essentially eliminate stop-and-go traffic when actions are held fixed for up to 5 seconds. Even up to 10-second action holds, such controllers show congestion benefits over a human driving baseline. These findings are a stepping-stone for near-term deployment of vehicle-based congestion mitigation.

1 Introduction

Greenhouse gases (GHGs) are gases that trap heat in the atmosphere and are a key contributor to climate change. Carbon dioxide (CO_2) makes up 80% of the United States GHG emissions in 2019. In particular, the transportation sector, primarily due to the combustion of fossil fuels like diesel and gasoline, causes significant CO_2 emissions, making up about 35% of the total CO_2 emissions and 29% of GHG emissions in 2019. The largest source of carbon dioxide emissions within transporation is land transport, representing emissions from passenger cars and light-, medium-, and heavy-duty trucks, which make up over 70% of the total CO_2 emissions within the sector [1].

We focus on reducing total emissions by improving traffic flow on the road and reducing congestion by designing vehicular controllers with deep reinforcement learning (RL). Previous work by Wu et al. shows that it is possible to nearly eliminate congestion and increase average speeds by up to 57% in idealized traffic settings using a small fraction of RL-controlled autonomous vehicles (AVs) on the road [2]. Further work by Sridhar et al. shows that these results can be reproduced with *piecewise constant policies*, a simplified class of reinforcement learning policies designed to be executable by humans. This suggests that we can use RL to empower human drivers to execute real-time congestion mitigation behavior in the near future, sidestepping the deployment obstacles for autonomous vehicles [3].

We expand upon prior work by investigating how well piecewise constant policies work in more complex settings by analyzing their potential to remove an existing traffic jam. In particular, our work offers two contributions:

- We show that piecewise constant policies, trained using deep reinforcement learning, can dissipate shock waves even after the formation of a severe traffic jam.

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2021.

- We show that piecewise constant policies are effective even with a local reward function, rather than a global one, which indicates potential for settings with limited cooperation.

2 Related work and preliminaries

2.1 Potential climate impact

There is a long history of work showing that simple changes to driving styles can have a large impact on emissions. For instance, the Environmental Protection Agency suggests avoiding unnecessary idling in cars and buses [1]. Previous work by Barth et al. shows that even simple traffic management strategies can significantly reduce emissions on the road. Based on typical traffic conditions in Southern California, they show that CO_2 emissions can be reduced by almost 20% by combining strategies for congestion mitigation, speed management, and shock wave suppression to reduce stop-and-go traffic [4]. Moreover, the Greek eco-driving program showed that bus drivers can reduce average fuel consumption (correlated with emissions) by 10.2% through simple strategies like maintaining steady speeds and anticipating traffic flows [5].

Advances in vehicle connectivity and control through RL allows us to take this a step further by directly optimizing for complex objectives. Our work focuses on leveraging RL to control vehicles to mitigate congestion and reduce overall emissions.

2.2 Policy class

We follow the model described in Sridhar et al. for formulating piecewise constant controllers. That is, previous human factors studies show that shared control on the road (for example, with semi-autonomous vehicles) can lead to increased safety risks [6, 7]. However, as discussed, studies suggest that even simple changes to driving styles can significantly impact congestion. Thus, we focus on a setting where human drivers are fully in control by designing policies that can be deployed as a minimally intrusive real-time app (similar to Google Maps) which provides personalized driving advice with minimal safety impacts [3].

In contrast with previous studies in human driver training, we focus on leveraging RL to provide real-time advice to directly optimize for congestion rather than general strategies. This is particularly compelling, since even simple driving strategies to reduce stop-and-go traffic waves show up to 7-12% emissions reductions [4]. We hypothesize that directly optimizing through RL techniques can increase the potential impact, while leveraging human drivers would permit a near-term deployment.

However, in order to ensure the “advice” from RL is executable by human drivers, we require that a new “suggestion” to the driver can only be provided every 5-8 seconds, due to human reaction times as studied by Mok et al [7]. We model this in simulation by requiring each action to be maintained for Δ timesteps where Δ is the action extension parameter. We first calculate performance using an “autonomous vehicle” baseline policy which chooses a new action per timestep (0.1 seconds), with $\Delta = 1$. This implies that policies with $\Delta = [50, 80]$ are structured to be executable by humans [3, 7].

2.3 Traffic jam formations

Previous work shows that piecewise constant policies with large action extensions can significantly mitigate congestion in simplified traffic settings [3]. Generally, traffic jams form due to small fluctuations in the movement of vehicles. If there is a large gap between vehicles, the fluctuation typically disappears and the vehicles remain in free flow, where cars are moving at a high velocity. However, when the traffic is dense enough, small fluctuations can compound, leading to cars clustering. These small fluctuations can grow until finally several vehicles are forced to stop completely, creating a stop-and-go wave [8].

Previous work by Sridhar et al. shows that a single controlled vehicle following piecewise constant policies with $\Delta \leq 140$ can successfully mitigate the buildup of traffic waves in the single-lane circular track setting with 21 simulated human drivers where small fluctuations are initially present [3]. We extend this work to a setting where the small fluctuations have already caused a stop-and-go traffic wave to develop. This is a much harder setting, akin to reducing congestion after a traffic jam has built up, compared to the original where the initial shock waves have just started.

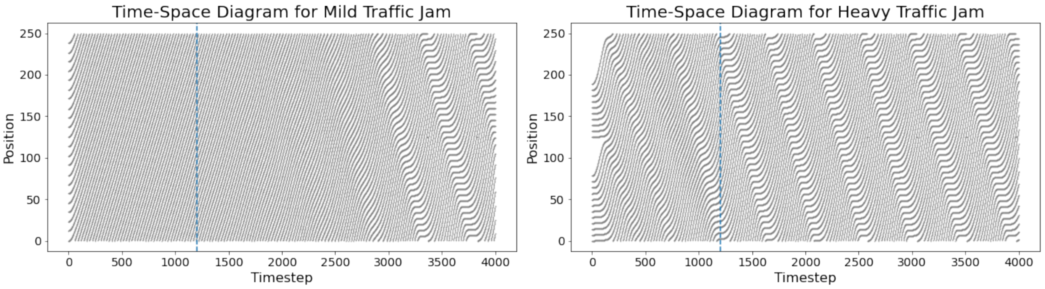

Figure 1: Time-space diagrams for varying levels of traffic jams. Each line represents a vehicle (with a human driver), where position is plotted on the y-axis and time in the x-axis. When a vehicle reaches position 250, it re-enters at position 0, since we are in a circular track environment. The end of the warmup period is marked by the blue line at 1200 time steps. **Left:** In the low-severity setting, we see mild disturbances build into stop-and-go waves after approximately 3000 timesteps. **Right:** In the high-severity setting, we observe significant stop-and-go traffic waves by 1200 time steps.

3 Problem

In this work, we study piecewise constant controllers to understand the potential of leveraging human drivers to mitigate congestion, particularly in severe settings. We consider the classic single-lane circular track. We integrate a single RL-controlled agent *after* the formation of stop-and-go traffic waves, and we train piecewise constant policies. Furthermore, we use local reward functions to explore congestion impacts with limited cooperation. The research questions we explore are: Under such severe traffic conditions, to what extent can piecewise constant controllers mitigate congestion? Is the regime of effectiveness within the limits of human reaction times?

4 Experimental results

4.1 Traffic severity

We can describe the severity of a traffic jam by analyzing the shock waves that build when only human drivers are present. We model human drivers using the Intelligent Driver Model (IDM) [9], a widely-used calibrated human driving model. In the original setting, described in the work of Sridhar et al., cars are initialized with random spacing around the track to create some initial small fluctuations [3]. With a warmup period (where the RL agent is also following IDM) of 1200 timesteps, we see small shock waves that can multiply in the human baseline setting into stop-and-go traffic, as seen in the left of Figure 1. However, by the end of the warmup period, the traffic waves are small.

In our setting, we start with a deterministic uneven spacing of the cars, where cars start out in two distinct clusters. This setup is designed for traffic jams to build faster. After a warmup period of 1200 steps, as seen in the right of Figure 1, we observe that strong traffic waves have developed. Thus, in our setting, the RL-controlled agent must solve the problem of smoothing a fully-formed traffic jam rather than much smaller shock waves, which are typically easier to equalize.

4.2 Experiments

We use the Flow framework for our vehicle control experiments [2]. In particular, we use the canonical single-lane circular track described by Sugiyama et al [8]. There are 22 vehicles on a 250m track and each driver in the track follows IDM. As described in Figure 1, we observe traffic waves build up without the presence of a controlled vehicle. We define the human baseline velocity as the average velocity on the track when no AVs are present. We note that the human driving baseline velocity is 3.45 m/s which is lower than the baseline of 3.6 m/s presented by Sridhar et al., most likely due to the increased severity of our traffic setting [3].

We then replace a single human driver with a learning agent (vehicle) on the track, constrained by a varying Δ. Every Δ timesteps, this agent is provided with a target acceleration. The agent translates this acceleration into a target speed and accelerates to that speed following IDM. Due to

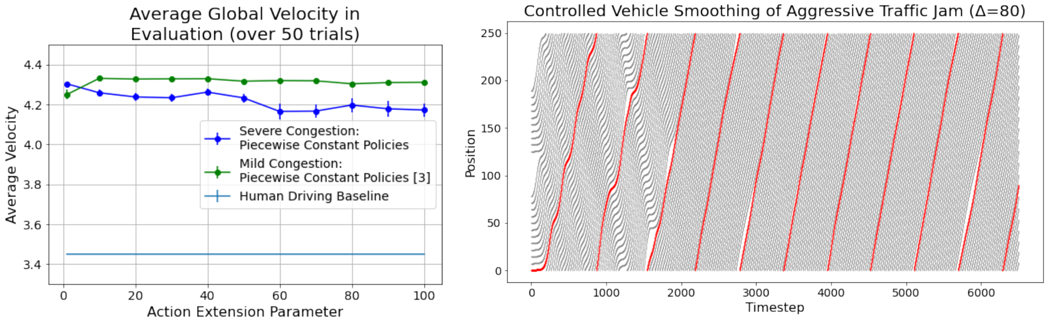

Figure 2: **Left:** Average velocity in evaluation for varying action extension parameters, evaluated over 50 trials for each Δ. We observe average speeds above 4.2 for $\Delta \leq 50$ and average speeds above 4.0 for $\Delta \leq 100$. We see comparable performance to past work for $\Delta \leq 50$ [3], even with the harder traffic setting. **Right:** Traffic smoothing with a singular agent (shown in red) with $\Delta = 80$. We observe traffic waves build in the first 1200 steps, then equalize due to the agent's behavior. We note that there are still minor waves near timestep 6000 since the agent cannot completely smooth the traffic.

safety considerations, the target speed is not always reached, thus modeling a relaxation of the exact piecewise constant requirements of the policy to be more representative of human behavior.

The learning agent observes its own speed, the speed of the preceding car, and the headway to the preceding car. We use a local reward function, which consists of the agent's own velocity [3]. The agent is trained for a horizon of 8000 steps, with 1200 warmup steps to allow the traffic waves to build. We train our agents using TRPO, with and without a critic [10]. We experiment with reward shaping where we penalize sharp changes in acceleration in order to discourage stop-and-go traffic waves. We train each setting for varying values of Δ and the results from the best model are described in the left of Figure 2.

First, we note that with $\Delta = 1$, we can smooth out traffic waves completely, with an average velocity of 4.3 m/s, which is close to the equilibrium velocity in the ring. We observe that up to $\Delta = 100$, we can consistently outperform the human baseline, even in the severe traffic jam settings. This is seen in the right of Figure 2 where an agent following a piecewise constant policy with $\Delta = 80$ is mostly able to smooth out heavy traffic waves over the horizon. Finally, up to $\Delta = 50$, we see that our average velocity is above 4.2, which is similar to the results shown by Sridhar et al [3].

5 Conclusions and discussion

In this work, we analyzed piecewise constant policies to study their performance in realistic traffic settings. We extend the work of Sridhar et al. to more complex settings by increasing the difficulty of the traffic flow problem and by reducing the sensing requirements [3]. That is, we focus on a single-lane circular track network where stop-and-go traffic waves have fully formed before the agent is deployed, representing situations where congestion is not predicted ahead of time. Moreover, we simplify the reward function to only use the RL vehicle's local observations.

We show that piecewise constant policies still work well under the increased constraints. In fact, for $\Delta \leq 100$ (10 seconds), the policies outperform the human baseline. Based on previous literature, we believe that $\Delta \in [50, 80]$ is large enough for the policies to be executable by humans [3, 7]. Studies in Southern California suggest that the elimination of such stop-and-go traffic could contribute to a 7-12% reduction in emissions, which could be a major climate change intervention [4]. Finally, we note that technological improvements in system efficiency, such as congestion mitigation, should be considered *jointly* with policy and regulation, in order to appropriately manage rebound effects.

In the future, we would like to experiment with increasingly complex traffic settings, including highway bottleneck and merging scenarios. We hypothesize that such robust simulation results, accompanied by user studies to evaluate the safety of piecewise constant policies would allow for a real-world deployment in the near future.

Acknowledgments and Disclosure of Funding

The authors thank Zhongxia Yan and Cameron Hickert for useful advice on the Flow framework and setup, Dr. Eric Horvitz for inspiring technical discussions and MIT Supercloud for computational resources. The authors also thank the MIT-IBM Watson AI Lab for funding this work and gratefully acknowledge that Mayuri Sridhar was supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate (NDSEG) Fellowship Program.

References

[1] US Environmental Protection Agency. Inventory of U.S. greenhouse gas emissions and sinks: 1990-2019. 2021.

[2] C. Wu, Aboudy Kreidieh, Kanaad Parvate, Eugene Vinitsky, and A. Bayen. Flow: A modular learning framework for mixed autonomy traffic. *IEEE Transactions on Robotics*, 2021.

[3] Mayuri Sridhar and Cathy Wu. Piecewise constant policies for human-compatible congestion mitigation. In *IEEE International Intelligent Transportation Systems Conference*, 2021.

[4] Matthew Barth and Kanok Boriboonsomsin. Real-world carbon dioxide impacts of traffic congestion. *Transportation Research Record*, 2058(1):163–171, 2008.

[5] Maria Zarkadoula, Grigoris Zoidis, and Efthymia Tritopoulou. Training urban bus drivers to promote smart driving: A note on a greek eco-driving pilot program. *Transportation Research Part D: Transport and Environment*, 12(6):449–451, 2007.

[6] M. Johns, B. Mok, D. Sirkin, N. Gowda, C. Smith, W. Talamonti, and W. Ju. Exploring shared control in automated driving. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 91–98, 2016.

[7] B. Mok, M. Johns, K. J. Lee, D. Miller, D. Sirkin, P. Ive, and W. Ju. Emergency, automation off: Unstructured transition timing for distracted drivers of automated vehicles. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pages 2458–2464, 2015.

[8] Yuki Sugiyama, Minoru Fukui, Macoto Kikuchi, Katsuya Hasebe, Akihiro Nakayama, Katsuhiro Nishinari, Shin-ichi Tadaki, and Satoshi Yukawa. Traffic jams without bottlenecks - experimental evidence for the physical mechanism of the formation of a jam. *New Journal of Physics*, 10:33001, 03 2008.

[9] Martin Treiber, Ansgar Hennecke, and Dirk Helbing. Congested traffic states in empirical observations and microscopic simulations. *Physical Review E*, 62:1805–1824, 02 2000.

[10] John Schulman, Sergey Levine, Pieter Abbeel, Michael Jordan, and Philipp Moritz. Trust region policy optimization. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 1889–1897, Lille, France, 07–09 Jul 2015. PMLR.