Alleviating Road Traffic Congestion with Artificial Intelligence

Guni Sharon

Texas A&M University, Computer Science and Engineering

guni@tamu.edu

Abstract

This paper reviews current AI solutions towards road traffic congestion alleviation. Three specific AI technologies are discussed, (1) intersection management protocols for coordinating vehicles through a roads intersection in a safe and efficient manner, (2) road pricing protocol that induce optimized traffic flow, and (3) partial or full autonomous driving that can stabilize traffic flow and mitigate adverse traffic shock waves. The paper briefly presents the challenges affiliated with each of these applications along with an overview of state-of-the-art solutions. Finally, real-world implementation gaps and challenges are discussed.

1 Introduction

Traffic congestion is one of the world's most costly and pressing problems. It leads to lost time and productivity, reduced air quality, and increased operating expenditures. While infrastructure expansions may help mitigate this problems, they are costly, disruptive, and require predicting demand and capabilities years, if not decades, into the future. On the positive side, recent advancements in AI pave a path towards improved utilization of existing infrastructure through autonomous and self-learning controllers.

A large body of work [Kockelman *et al.*, 2017; Zantalis *et al.*, 2019] was devoted in recent years to AI technology integration within intelligent transportation systems. In this paper, we present and discuss a subset of these works which relate to congestion alleviation. Specifically, we discuss three applications where AI controllers present significant potential towards congestion mitigation.

(1) Intersection Management. Road intersections introduce traffic bottlenecks which are a known and well studied source of traffic congestion [Tirachini, 2013]. We discuss two established AI approaches towards optimized intersection management. Namely, signal control optimization through deep reinforcement learning and reservation based trajectory coordination.

(2) Road Pricing for Optimized Traffic Flow. Self-interested drivers follow routes which optimize their own personal utility while potentially acting suboptimal with respect to the entire system (utility summed over all drivers). This phenomena, measured by *the price of anarchy* [Roughgarden and Tardos, 2002], is a known source of avoidable congestion. We discuss two established AI approaches towards road pricing protocols which attempt to align the drivers' personal utility with that of the system's. Namely, a game theoretical approach and a reinforcement learning approach.

(3) Fully or Partial Autonomous Driving. Erratic driving, of even a single vehicle, can lead to congestion that is induced by *shock waves* [Richards, 1956], which are also known as *stop-start traffic waves*, *ghost jams*, or *traffic shocks*. Recent autonomous driving controllers present a viable approach towards mitigating such adverse behaviors. We discuss AI approaches towards autonomous motion planing and control that stabilizes traffic flow progression by damping shock waves.

We conclude the paper by presenting open challenges relating to safety assurances in reinforcement learning, performance guarantees, interpretable and foreseeable autonomous control, warm start learning in reinforcement learning and knowledge transfer issues, public and governmental acceptance issues, and infrastructure compatibility issues.

2 Intersection Management

Travel time studies in urban areas show that 12–55% of commute travel time is due to delays induced by signalized intersections (stopped or approach delay) [Tirachini, 2013]. Hence, optimized intersection management has the potential of reducing commute time, traffic congestion, emissions, and fuel consumption, while requiring minimal infrastructure changes. A line of publications attempted to harness modern AI techniques towards this control optimization problem. This section surveys two such approaches. Namely, signal control optimization and reservation-based coordination.

2.1 Signal Control Optimization

A signalized intersection is composed of incoming and outgoing roads where each road is affiliated with one or more lanes. A signal controller must assign right of passage to phases where each phase is affiliated with a specific traffic movement through the intersection (incoming to outgoing roads/lanes). Two phases are defined to be in *conflict* if they cannot be enabled simultaneously (their affiliated traffic movement is

intersecting). During the signal operation, vehicles are assumed to continuously arrive on incoming roads where each vehicle is associated with a specific outgoing road. At every time step, the signal controller is required to assign right-of-passage (green light) to a set of non-conflicting phases such that some utility measurement is optimized. The utility to be optimized is commonly defined as the sum of vehicles' delay imposed by the intersection [Mousavi *et al.*, 2017; Ault *et al.*, 2020].

Given that signalized intersections vary in their layout and demand profile, optimized control policies are unique and instance dependent. Consequently, signal controllers should be trained and optimized according to observed conditions at the associated intersection. Such online optimization of signal controllers requires: (a) sensing the state of approaching traffic (e.g., number and position of approaching vehicles, approaching speeds, vehicle queue length, waiting times) aggregated by approaching lanes, and (b) defining a controller that takes the current state of traffic as input and outputs the next phases to be enabled (which is translated to a green, yellow, and red assignment for each light box).

Recent publications [Mousavi *et al.*, 2017; Ault *et al.*, 2020] proposed to utilize state-of-the-art reinforcement learning (RL) algorithms for online optimization of signal controllers. In this approach, the state of the intersection is usually defined by the set of incoming vehicles (incoming lane, speed, waiting time, queue length) and the current signal (enabled phase) assignment. An RL agent is tasked with optimizing a policy which maps states to signal (phase) assignment. Such learning is commonly done using a value-based approach (specifically a deep Q-learning approach [Mnih *et al.*, 2015]) which attempts to learn the expected future utility from every action at a given state. The controller is then directed towards actions that maximize the expected future utility. Such an approach showed a potential reduction of up to 73% in vehicle delays when compared to fixed-time actuation [Mousavi *et al.*, 2017].

Recent work showed further potential for accumulated delay reductions when utilizing Multiagent RL for coordinated control of multiple adjacent intersections [Chen *et al.*, 2020].

2.2 Reservation-based Coordination

Looking towards a future when most vehicles are autonomous and connected, researchers are developing reservation-based intersection management protocols. By relying on the fine and accurate control of *connected and autonomous vehicles* (CAVs) along with their communication capabilities, intersection management protocols coordinate multiple vehicles simultaneously across an intersection. Such protocols have been shown to lead to significant traffic delay reductions when compared to traditional traffic signals. One prominent example of such a protocol is the autonomous intersection management (AIM) protocol [Dresner and Stone, 2008]. The AIM protocol defines two types of autonomous agents: intersection managers (IMs), one per intersection, and driver agents, one per vehicle. An intersection manager is responsible for directing incoming vehicles through its assigned intersection, while the driver agents are responsible for controlling the CAV to which they are assigned. To improve the

throughput and efficiency of the system, the driver agents "call ahead" to an intersection manager and request path reservations (space-time sequence) within the intersection. The IM then determines whether or not this request can be met by checking whether it conflicts with any previously approved reservation or any potential *human operated vehicles* (HVs). If the IM approves a driver agent's request, the driver agent must follow the assigned path through the intersection. On the other hand, if the IM rejects a driver agent's request, the driver agent may not pass through the intersection but may attempt to request a new reservation.

A line of publications followed Dresner and Stone's 2008 seminal work by considering variants of the originally presented reservation based approach. These approaches include reservation coordination methods that are modeled as a quadratic program [Riegger *et al.*, 2016], a model predictive control [Kim and Kumar, 2014], and a joint optimization (path coordination and trajectory control) problem [Malikopoulos *et al.*, 2021]. Other approaches consider platooning and batch reservations [Bashiri *et al.*, 2018], game theoretic/incentive based [Sayin *et al.*, 2018], auction based [Levin and Boyles, 2015], and RL-based [Wu *et al.*, 2019] techniques for allocating reservations. Recent work also considered mixed traffic scenarios where human operated vehicles and CAVs coexist [Sharon and Stone, 2017].

3 Dynamic Road Pricing for Optimized Traffic Flow

Another significant sources of traffic congestion stems from the fact that drivers are self-interested when making travel choices: for instance, they (quite naturally) select the route that optimizes their own utility. Under such conditions, the sum total of individual travel times (i.e., the *social welfare*) can be arbitrarily worse than it would be were a centralized controller to assign routes for the purpose of achieving maximum social welfare [Roughgarden and Tardos, 2002]. In order to combat this inefficiency, researchers have devised various approaches for computing and assigning optimized travel routes. A line of publications assume the ability to assign dynamic road pricing per road segment and that by doing so, the route choice of self-interested drivers can be manipulated. In this sense, the problem of assigning optimized travel routes is reduced to assigning a set of road prices that induce optimized flow.

3.1 Game-theory Approach

Self interested agents that are routed in a congestible network, such as vehicles in a road network, impose a *user equilibrium* (UE) that is often far worse than the *system optimum* (SO) flow [Roughgarden and Tardos, 2002]. Charging *marginal cost tolls* (MCT), in which each agent is charged a toll equivalent to the damage it inflicts on all other agents, results in a UE that achieves SO performance [Pigou, 1920].

Calculating the MCT for a given passenger, on a given path, i.e., the damage that the passenger inflicts on other passengers by traversing the path in question, is challenging without making several simplifying assumptions (e.g., well-defined and known latency functions) that do not hold in most

traffic models and certainly not in real-life traffic. Recent work [Sharon *et al.*, 2017a; Sharon *et al.*, 2017b] suggested a model free technique, denoted Δ-tolling, for evaluating MCT. Δ-tolling was proven to result in exact MCT only when traffic accumulation follows the *Bureau of Public Roads* (BPR) latency function [Moses and Mtoi, 2017]. By contrast, when the BPR assumption does not hold, Δ-tolling provides only an evaluation of the true MCT. Nonetheless, experimental results using a cell transmission model [Daganzo, 1994] presented a reduction of up to 32% in total system travel time when using Δ-tolling and up to 52% when tuning the affiliated hyper parameters with standard black-box optimization techniques [Mirzaei *et al.*, 2018].

Looking towards real-world application of MCT-based systems, recent publications examined various implementation issues. Among them, the impact of partial participation in a MCT scheme [Sharon *et al.*, 2018; Hanna *et al.*, 2019] and worst-case performance guarantees for imposing inaccurate (evaluation-based) MCT [Sharon *et al.*, 2019].

3.2 Reinforcement-learning Approach

RL approaches for learning and setting pricing policies were considered in previous publications [Chen *et al.*, 2018]. In such cases, multiagent RL formalization is used to train a pricing policy. The state space is defined as the observed traffic flow per road segment. A set of priced road segments in a given road network defines the action space as a Cartesian product of the allowed price change over these road segments. The reward function is simply minus total system travel time per time step. The pricing policy is trained such that the expected sum of future rewards is maximized.

Chen et al. 2018 showed that applying a policy gradient approach within an actor-critic architecture can outperform Δ-tolling by up to 15% (reduced total system travel times) when only a subset of road segments can be priced in a given network. These results are to be expected as Δ-tolling—and any MCT based pricing scheme for this matter—assume that all road segments can be priced.

4 Autonomous Driving

Erratic or unpredictable driving of even a single vehicle can lead to substantial traffic congestion due to *shock waves* [Richards, 1956]. Luckily, autonomous vehicles (AVs) provides a viable way towards predictable and smooth vehicle control. Moreover, such AV technology showed potential for proactively stabilizing traffic flow and thus lead to reduced congestion. AV technology has evolved significantly in recent years to a point where many researchers agree that it is not a question of 'if' but a question of 'when' will the technology be viable.

4.1 Motion Planning

Researchers have proposed to address the AV motion control problem using classic Motion planning techniques [Frazzoli *et al.*, 2002]. However, it is challenging to apply such techniques in highly dynamic and stochastic environments such as the driving domain. The uncertainties introduced by other actuators (e.g., vehicles, pedestrians, traffic signals) and control objectives (e.g., Car-following, lane-keeping, trajectory

following, merging) necessitate a large variety of coordinated low-level control tasks as well as frequent replanning operations. Consequently, many researchers turn towards RL techniques that can learn a general and robust control policy.

Aradi 2020 provided a survey on deep RL hierarchical motion planning approach in which different layers of motion planning, namely, strategic decisions, trajectory planning, and control, are presented and combined to create the desired robust driving control policy. In a more specific example, Shalev-Shwartz et al. 2016 addressed the high-level task of long term driving strategies and trajectory planning. They addressed this planning problem by applying the REINFORCE algorithm [Williams, 1992] (Monte-Carlo policy gradient) along with a Q-estimator as a baseline. Both the policy and the Q-estimator were defined using deep neural-networks. This approach was shown to increase driving stability and predictability while satisfying safety constraints.

4.2 Stabilizing Traffic Flow

Leveraging previously proposed trainable AV motion controllers, researchers examined incorporation of flow stabilizing behaviors. This is noteworthy as traffic flow instabilities are known to contribute to adverse congestion, e.g., through traffic shock waves. Zheng et al. 2020 demonstrated how few AVs can potentially stabilize mixed (human operated vehicles and AVs) traffic flow progression. Such stabilizing is usually achieved through speed and headway distance control that optimize *string stability*. Results by Zheng et al. suggest that as few as 5% AV presence can lead to 6% increase in traffic flow speed. Vinitsky et al. 2018 suggested a deep RL approach for AV control with the objective of stabilizing traffic flow when entering a road bottleneck. Example of such a bottleneck is a bridge with a reduced number of lanes as compared to its incoming road. Vinitsky et al. showed that training AVs towards flow stabalizing behavior using trust region policy optimization (TRPO) [Schulman *et al.*, 2015] can lead to 25% throughput improvement at high inflows when as little as 10% AV penetration levels are considered. Addressing the scalability issues in these approaches, Cui et al. [Cui *et al.*, 2021] proposed to apply a transfer RL approach that can handle open-road settings where vehicles can dynamically join and leave a traffic flow composed of hundreds of vehicle.

Finally, Stern et al. 2018 provided a real-world demonstration of how flow-stabilizing AV controllers contribute to smooth and predictable driving which mitigates adverse shock waves. Moreover, Stern et al. demonstrated how a single AV can actively dampen shock waves initiated by other vehicles in a real-world setting.

5 Open Challenges

Several open challenges hinder the real-world implementation of the aforementioned technologies. Next, we briefly highlight some of these challenges.

Safety concerns. Common reinforcement learning (RL) algorithms rely on extensive exploration of the state and action space to discover an optimized control policy. In the discussed traffic domains, such exploration can lead to unacceptable safety violations. Notably, a large body of work

was devoted towards safe RL [García and Fernández, 2015] and stable performance during RL training [Schulman *et al.*, 2015]. Still, applicability and algorithmic adaptations studies towards the discussed traffic applications need to be established.

Predictability.   Due to the non-deterministic nature of real-world control problems, RL-based controllers are commonly limited to a single-step planning horizon. That is, given the current state of the environment, the RL agent computes an action to be taken at the current time step, but does not commit to any future actions. While allowing the RL agent flexibility to immediately adapt and respond to the environment, this fact reduces the ability to regulate the RL agent's behavior and predict its actions. Moreover, the aforementioned traffic applications involve environments where multiple actuators (in the form of vehicles, infrastructure, pedestrians) operate in conjunction. In such cases, actuator predictability can foster more efficient and safe coordination and should thus be aspired for.

Efficient Controller Training.   Training an RL controller is often slow, sample inefficient, and brittle. When it comes to traffic applications, such training can be prohibitively expensive. As a result, integration of warm start learning from demonstrations [Hester *et al.*, 2018] and knowledge transfer approaches [Parisotto *et al.*, 2015] should be examined and adapted towards RL-based traffic applications.

Public and Governmental Acceptance.   Following the aforementioned safety, predictability, and efficiency concerns, some people deem AI technology as not trustworthy when safety-critical tasks are considered. Such trust issues can lead to reduced public and governmental willingness to apply the discussed AI-based methods. AI algorithms and approaches that integrate legal, ethical, societal and economic impact considerations can address such trust issues and should thus be sought after. Relevant impact considerations can be found in legal documents such as the framework for trustworthy AI proposed by the European Commission [European Commission Press Release, 2021].

Infrastructure Compatibility.   Real-world implementation of traffic controllers is restricted by stringent standards and regulations. In the area of traffic controls, the National Transportation Communications for Intelligent Transportation Systems Protocol (NTCIP) [NEMA, 1996] is a family of standards that provides protocols and the vocabulary necessary to allow electronic traffic control equipment from different manufacturers to operate with each other as a system. In spite of such well defined protocols, AI-based traffic controllers are often developed with no regards to standardization. This inconsistency might lead to substantial infrastructure compatibility issues and should thus be avoided.

Acknowledgments

References

[Aradi, 2020] Szilárd Aradi. Survey of deep reinforcement learning for motion planning of autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–20, 2020.

[Ault *et al.*, 2020] James Ault, Josiah Hanna, and Guni Sharon. Learning an interpretable traffic signal control policy. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2020)*. International Foundation for Autonomous Agents and Multiagent Systems, May 2020.

[Bashiri *et al.*, 2018] Masoud Bashiri, Hassan Jafarzadeh, and Cody H Fleming. Paim: Platoon-based autonomous intersection management. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 374–380. IEEE, 2018.

[Chen *et al.*, 2018] Haipeng Chen, Bo An, Guni Sharon, Josiah Hanna, Peter Stone, Chunyan Miao, and Yeng Soh. Dyetc: Dynamic electronic toll collection for traffic congestion alleviation. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, 2018.

[Chen *et al.*, 2020] Chacha Chen, Hua Wei, Nan Xu, Guanjie Zheng, Ming Yang, Yuanhao Xiong, Kai Xu, and Zhenhui Li. Toward a thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, pages 3414–3421, 2020.

[Cui *et al.*, 2021] Jiaxun Cui, William Macke, Harel Yedidsion, Aastha Goyal, Daniel Urieli, and Peter Stone. Scalable multi-agent driving policies for reducing traffic congestion. In *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, May 2021.

[Daganzo, 1994] Carlos F. Daganzo. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological*, 28(4):269–287, 1994.

[Dresner and Stone, 2008] Kurt Dresner and Peter Stone. A multiagent approach to autonomous intersection management. *Journal of artificial intelligence research*, 31:591–656, 2008.

[European Commission Press Release, 2021] European Commission Press Release. Europe fit for the digital age: Commission proposes new rules and actions for excellence and trust in artificial intelligence, April 2021.

[Frazzoli *et al.*, 2002] Emilio Frazzoli, Munther A Dahleh, and Eric Feron. Real-time motion planning for agile autonomous vehicles. *Journal of guidance, control, and dynamics*, 25(1):116–129, 2002.

[García and Fernández, 2015] Javier   García   and   Fernando Fernández. A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research*, 16(1):1437–1480, 2015.

[Hanna *et al.*, 2019] Josiah P. Hanna, Guni Sharon, Stephen D. Boyles, and Peter Stone. Selecting compliant agents for opt-in micro-tolling. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI-19)*, January 2019.

[Hester *et al.*, 2018] Todd Hester, Matej Vecerik, Olivier Pietquin, Marc Lanctot, Tom Schaul, Bilal Piot, Dan Horgan, John Quan, Andrew Sendonaris, Ian Osband, et al. Deep q-learning from demonstrations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[Kim and Kumar, 2014] Kyoung-Dae Kim and Panganamala Ramana Kumar. An mpc-based approach to provable system-wide

safety and liveness of autonomous ground traffic. *IEEE Transactions on Automatic Control*, 59(12):3341–3356, 2014.

[Kockelman *et al.*, 2017] Kara Kockelman, Stephen Boyles, Peter Stone, Dan Fagnant, Rahul Patel, Michael W Levin, Guni Sharon, Michele Simoni, Michael Albert, Hagen Fritz, et al. An assessment of autonomous vehicles: traffic impacts and infrastructure needs. Technical report, University of Texas at Austin. Center for Transportation Research, 2017.

[Levin and Boyles, 2015] Michael Levin and Stephen Boyles. Intersection auctions and reservation-based control in dynamic traffic assignment. *Transportation Research Record*, 2497(1):35–44, 2015.

[Malikopoulos *et al.*, 2021] Andreas A. Malikopoulos, Logan Beaver, and Ioannis Vasileios Chremos. Optimal time trajectory and coordination for connected and automated vehicles. *Automatica*, 125:109469, 2021.

[Mirzaei *et al.*, 2018] Hamid Mirzaei, Guni Sharon, Stephen D. Boyles, Tony Givargis, and Peter Stone. Enhanced delta-tolling: Traffic optimization via policy gradient reinforcement learning. In *Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC-18)*, November 2018.

[Mnih *et al.*, 2015] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

[Moses and Mtoi, 2017] Ren Moses and Enock T. Mtoi. Calibration and evaluation of link congestion functions. *Journal of Transportation Technologies*, 4(2), 2017.

[Mousavi *et al.*, 2017] Seyed Sajad Mousavi, Michael Schukat, and Enda Howley. Traffic light control using deep policy-gradient and value-function-based reinforcement learning. *IET Intelligent Transport Systems*, 11(7):417–423, 2017.

[NEMA, 1996] NEMA. National transportation communications for intelligent transportation system protocol, 1996. National Electrical Manufacturers Association (NEMA) and the Institute of Transportation Engineers (ITE) and the American Association of State Highway and Transportation Officials (AASHTO).

[Parisotto *et al.*, 2015] Emilio Parisotto, Jimmy Lei Ba, and Ruslan Salakhutdinov. Actor-mimic: Deep multitask and transfer reinforcement learning. *arXiv preprint arXiv:1511.06342*, 2015.

[Pigou, 1920] Arthur Pigou. *The Economics of Welfare*. Palgrave Macmillan, 1920.

[Richards, 1956] Paul I Richards. Shock waves on the highway. *Operations research*, 4(1):42–51, 1956.

[Riegger *et al.*, 2016] Lea Riegger, Markus Carlander, Niklas Lidander, Nikolce Murgovski, and Jonas Sjöberg. Centralized mpc for autonomous intersection crossing. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1372–1377. IEEE, 2016.

[Roughgarden and Tardos, 2002] Tim Roughgarden and Éva Tardos. How bad is selfish routing? *Journal of the ACM (JACM)*, 49(2):236–259, 2002.

[Sayin *et al.*, 2018] Muhammed O Sayin, Chung-Wei Lin, Shinichi Shiraishi, Jiajun Shen, and Tamer Başar. Information-driven autonomous intersection control via incentive compatible mechanisms. *IEEE Transactions on Intelligent Transportation Systems*, 20(3):912–924, 2018.

[Schulman *et al.*, 2015] John Schulman, Sergey Levine, Pieter Abbeel, Michael Jordan, and Philipp Moritz. Trust region policy optimization. In *International conference on machine learning*, pages 1889–1897. PMLR, 2015.

[Shalev-Shwartz *et al.*, 2016] Shai Shalev-Shwartz, Shaked Shammah, and Amnon Shashua. Safe, multi-agent, reinforcement learning for autonomous driving. *arXiv preprint arXiv:1610.03295*, 2016.

[Sharon and Stone, 2017] Guni Sharon and Peter Stone. A protocol for mixed autonomous and human-operated vehicles at intersections. In *International Conference on Autonomous Agents and Multiagent Systems*, pages 151–167. Springer, 2017.

[Sharon *et al.*, 2017a] Guni Sharon, Josiah P. Hanna, Tarun Rambha, Michael W. Levin, Michael Albert, Stephen D. Boyles, and Peter Stone. Real-time adaptive tolling scheme for optimized social welfare in traffic networks. In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2017)*, May 2017.

[Sharon *et al.*, 2017b] Guni Sharon, Michael W Levin, Josiah P Hanna, Tarun Rambha, Stephen D Boyles, and Peter Stone. Network-wide adaptive tolling for connected and automated vehicles. *Transportation Research Part C: Emerging Technologies*, 84:142–157, 2017.

[Sharon *et al.*, 2018] Guni Sharon, Michael Albert, Tarun Rambha, Stephen Boyles, and Peter Stone. Traffic optimization for a mixture of self-interested and compliant agents. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, 2018.

[Sharon *et al.*, 2019] Guni Sharon, Stephen D. Boyles, Shani Alkoby, and Peter Stone. Marginal cost pricing with a fixed error factor in traffic networks. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-19)*, May 2019.

[Stern *et al.*, 2018] Raphael E Stern, Shumo Cui, Maria Laura Delle Monache, Rahul Bhadani, Matt Bunting, Miles Churchill, Nathaniel Hamilton, Hannah Pohlmann, Fangyu Wu, Benedetto Piccoli, et al. Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments. *Transportation Research Part C: Emerging Technologies*, 89:205–221, 2018.

[Tirachini, 2013] Alejandro Tirachini. Estimation of travel time and the benefits of upgrading the fare payment technology in urban bus services. *Transportation Research Part C: Emerging Technologies*, 30:239–256, 2013.

[Vinitsky *et al.*, 2018] Eugene Vinitsky, Kanaad Parvate, Aboudy Kreidieh, Cathy Wu, and Alexandre Bayen. Lagrangian control through deep-RL: Applications to bottleneck decongestion. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 759–765. IEEE, 2018.

[Williams, 1992] Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256, 1992.

[Wu *et al.*, 2019] Yuanyuan Wu, Haipeng Chen, and Feng Zhu. Dcl-aim: Decentralized coordination learning of autonomous intersection management for connected and automated vehicles. *Transportation Research Part C: Emerging Technologies*, 103:246–260, 2019.

[Zantalis *et al.*, 2019] Fotios Zantalis, Grigorios Koulouras, Sotiris Karabetsos, and Dionisis Kandris. A review of machine learning and iot in smart transportation. *Future Internet*, 11(4):94, 2019.

[Zheng *et al.*, 2020] Yang Zheng, Jiawei Wang, and Keqiang Li. Smoothing traffic flow via control of autonomous vehicles. *IEEE Internet of Things Journal*, 7(5):3882–3896, 2020.