A Novel Congestion Avoidance Algorithm for Autonomous Vehicles Assessed by Queue Modeling

Khashayar Kotobi and Mina Sartipi

March 2019

Abstract

Autonomous vehicle (AV) fleet management is one of the major aspects of AV development that needs to be standardized before AV deployment. There has been no consensus on whether AV deployment in general will be beneficial or detrimental in terms of road congestion. There are similarities between packet transmission in computer networks and AV transportation in road networks. In this work, we argue that congestion avoidance algorithms used in computer networks can be applied for AV fleet management. We modify and evaluate a novel adaptation of additive increase and multiplicative decrease (AMID) congestion avoidance algorithm. We propose assigning different priorities to transportation tasks in order to facilitate sharing the limited resources in such as usage of the road network. This will be modeled and assessed using a queueing model based on AVs arrival distribution. This will result in a load balancing paradigm that can be used to share and manage limited resources. Then, by using numerical study we merge congestion avoidance and load balancing to analyze our scheme in term of road network throughput (number of cars in network for a given time) for AV fleet management. Our evaluation demonstrates the improvement in terms of road network throughput.

1 Introduction

As the development of autonomous vehicles (AV) rapidly increases, there is a higher sense of urgency for engineers and regulator bodies to become more engaged in how these AVs will be integrated into the current road system. According to Federal Highway Association (FHWA), in 2016 there were more than 35,000 fatality crashes in the United States [1]. Implementation of AVs can reduce these numbers significantly, if their sensors maintained carefully, since AVs do not have human limitation [2]. The sensing technology used in AV provides a “360”- degree visualization of the dynamic surrounding environment that the regular driver cannot access. By 2030, 1 in every 5 citizens will be the share of older population according to U.S. Census [3]. As driver age increases,

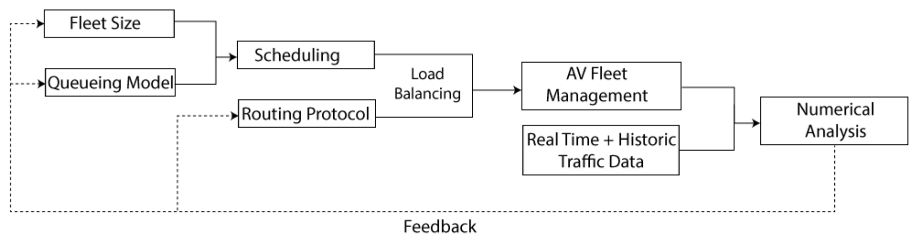

Figure 1: AV deployment using queueing theory and congestion avoidance algorithm by real time and historic data feedback block diagram.

decline in visual capability, reaction time, and memory will impact their ability to drive. AVs also increase productivity significantly; autonomous trucks do not require the resting time as it is necessary for the regular trucks. In this paper we aim to focus on AV fleet management deployment by introducing other benefits that can arise from studying the AVs as a group. Here, we study the AVs in a macro level to identify a global optimal AV fleet deployment. By quantitatively studying a fleet of AVs in terms of arrival rate and waiting time in road network, we aim to propose routing and scheduling algorithms that reduce and/or reshape the traffic. Next, we aim to develop optimization algorithms and models that consider necessary trade-offs (e.g., quality of service, traffic congestion, travel time and cost) to determine how to best deploy AVs spatially and temporally. With proper data gathering and analysis, these algorithms can be used to manage a feet of AVs.

AVs are being tested in few cities and they are expected to be deployed in more urban areas such as Pittsburgh and Seattle [4]. These tests have been mainly focused on the performance of a single AV and its interaction with a dynamic environment in a real time [5], [6]. We argue that there is an urgent need for a comprehensive protocol to manage the fleet of AVs (also called autonomous mobility-on-demand or AMoD) that is adaptable to a dynamic and fast paced road network. In order to achieve that protocol, we aim to build a scheduling protocol that will lead to a global optimal in terms of using the road capacity for the traffic containing both AV fleets and regular vehicles. By introducing a scheduling algorithm assessed by queueing theory based on real time and historical traffic data, we aim to avoid and/or reduce traffic congestion. The queue type is determined by the probability distribution of arrival time for each traffic type (regular or hybrid cars and AVs) in a queue. Then we allocate the number of AVs in each queue based on a corresponding routing protocol and the chosen queue type for each case study. Accordingly, fleet size will need to be tailored for each route. By combining routing, scheduling, and fleet size, we claim that one can provide a comprehensive algorithm for deployment of fleet of AVs in urban setting (Fig. 1).

In AV fleet management and routing schemes, multiple resources must be allocated among competing regular and hybrid cars and AVs. These resources include but not limited to electric charging, usage of road capacity, and parking. This multiple resource allocation is generally more difficult to solve than a single resource problem. These vehicles can achieve optimal allocation of these resources via regular communication with each other and traffic control units implemented in smart cities. Our resource allocation scheme is based on additive-increase and multiplicative-decrease (AMID) algorithm that is used to avoid congestion in transmission control protocol (TCP). Here by using numerical analysis we demonstrate that our new congestion avoidance algorithm for AVs can improve the road network throughout (number of cars in network for a given time) similar to algorithms used in computer networks.

The remainder of this paper is organized as follows: in Section 2, we investigate the current state of the art in AV routing and scheduling. In Section 3, we then define our scheduling protocol based on a congestion avoidance algorithm. We then describe the queueing model we use to evaluate our algorithm in section 4. In Section 5, we propose assigning different transportation priority to share the road network. In Section 6, we present our numerical analysis for the algorithms proposed. In Section 7, we discuss our results and present conclusions.

2 Literature Review

Recent studies on decision making for AVs focus mainly on:

1. Finding an optimal algorithm for AV allocation to select the next destination for the AVs performing multiple transportation tasks based on resource allocation [7],

2. Processing, gathering and updating data streams that are necessary for short and long term planing and decision making for AVs mentioned tasks,

3. Introducing and evaluation algorithms to improve the AVs' dynamic performance by means of improving high data rate communication among AVs,

4. Using Adaptive learnings to improve AV's driving pattern based on past recorded performance in case studies [8].

In this work, we aim to introduce a new scheduling (by using a congestion avoidance method) and routing algorithm (using queueing theory) for fleet of AVs with conditions that need to be satisfied to reach the ultimate goal of a congestion free road network. It has been claimed that AVs can increase the congestion since there will be more empty AV cruising to find a transportation task to perform [9]. One intuitive approach can be assigning transportation tasks to a fleet manager who is aware of the AVs' locations and their destinations and as result can avoid congestion. This approach is not practical since there are many competing entities that are going to act greedy to maximize their throughput [10]. Load balancing has been proposed for AV routing protocol [11] with the goal of reducing gas consumption or traveler waiting time. In [11], authors introduced an algorithm for maximizing the throughput of a mobility-on-demand transportation tasks. Here, We define a transportation task as a task to be performed by an AV or an autonomous truck. In this work, we aim to reduce traffic congestion by introducing different transportation task priorities. A queueing network framework has been proposed for driver to vehicle ratio balancing [12]; but not for AVs fleet deployment. They used a queueing approach to analysis and control of mobility on demand (MoD) systems for urban personal transportation. A MoD system consists of a fleet of vehicles providing transportation tasks and showed using a team of drivers; they can balance such vehicles. We use queueing theory to allocate transportation tasks based on their priorities to the road network. Almost all AV fleet management and routing protocols mainly consider point to point transportation [13], [14]. Our goal is to find an optimized routing and scheduling protocol to consider the road network entirely and find an optimal algorithm for AV deployment. Autonomous delivery can be expanded from autonomous trucks to autonomous delivery robots launched from trucks such as [15]. This shows the practicality of our approach to introduce different transportation tasks.

Scheduling for autonomous trucks and AVs has not been investigated extensively [16]. Similar to data packet scheduling and resource management in telecommunication networks, we claim that through efficient scheduling we can reach the sub-optimal usage of road network capacity. Inspired by Automatic Repeat reQuest (ARQ) and additive increase and multiplicative decrease (AMID), we argue that the AVs must be deployed as long as there is no traffic congestion on road network chosen to be the routing algorithm. In a computer network's ARQ, the sender transmits the data packets to the receiver and waits for acknowledgement (ACK) signals. If the ACK is not received, the data will be retransmitted by sender. Here, we use the real time data about the traffic congestion as our indicating signal regarding the road condition. If there is no congestion in the designated route, the AVs and autonomous trucks will be deployed based on transportation tasks' priorities. If a travel time for an AV for a given segment of the road network is more than certain threshold levels, the deployment will be decreased for the next time interval depending on severe or moderate congestion. Investigating the best interval time and threshold levels is subject to numerical analysis with real time and historic traffic data for each route.

3 Problem Definition

In this work, we aim to find an optimum routing protocol and scheduling using queueing theory for AVs to achieve the minimal traffic congestion using traffic data. By numerical analysis and simulation we will demonstrate that our AV fleet management (consisting of routing protocol and scheduling) leads to improve traffic throughput by a novel congestion control inspired by transportation control used in computer networks.

First, we present a road network as a weighted directed graphs to model the capacity of various roads in the network. Based on that, we develop congestion aware routing and rebalancing protocol. Second, we introduce a scheduling scheme using queueing theory and assigning priority levels for transportation tasks, we aim to efficiently manage the customer waiting time. Third, through numerical studies on traffic data (number of cars passed through a road network) obtained by our research group, we validate our models iteratively. As it can be seen from the Fig. 1, the fleet size to be deployed and the queueing model will determine our scheduling algorithm. This algorithm and routing protocol will define AV fleet management. We propose to use the real time and historic traffic data to analysis the road network performance and dynamically change the queue model and/or fleet size to improve the road network throughput.

We investigate a road network with an AV travel time as congestion indicator. The time used by an AV to pass a segment of the road will be used to define the state of the congestion in the system. This time indicator (t_i) will be compared to two threshold level (τ_1 and τ_2), representing moderate and severe congestion in the road network. Then, we introduce the number of AV that can be deployed in state of i as DW_i (deployment window for ith state). If the t_i is more than τ_2, the DW_i will be set to 1 and if the ti is more than τ_1, the DW_i will be half of DW_{i-1} inspired by Reno window update used in AMID for computer networks. As a result, the deployment window together with the threshold level represent a Markov chain.

In Fig. 2, two states based on two threshold levels defined for our road network is presented. For τ_1, we have a moderate congestions and for τ_2 we have a severe congestion. Fig. 2 represents the transmission between these two states based on the recorded amount for t_i. In case of moderate congestion the deployment increases until we reach saturation in road network. This ensures that the AV deployment will try to use the road network as much as it can to increase the network throughput. In the following the network throughput for AV is defined as the number of AV over time deployed in the road network. In the second state of Fig. 2, when the t_i is more than τ_2, we have a severe congestion and as result the AV deployment needs to be reduced to relieve the road network from congestion.

(a) Moderate Congestion

(b) Severe Congestion

Figure 2: Transition between different Markov state levels based on threshold levels for congestion

Here we want to find the throughput of road network by using Markov chain properties. We define EN_{ij} and ET_{ij} as expected number of AVs and time spent in the road network during the transition from state i to j respectively. We define the cost function for the number of AVs as

$$CN_i = \sum_j EN_{ij} P_{ij} \tag{1}$$

and for time spend in the road network

$$CT_i = \sum_j ET_{ij} P_{ij}, \tag{2}$$

where P_{ij} is the transition probability from state i to j. The steady state expected value for the number of AVs deployed and time in road network is can be found from the following expression [19]:

$$CN = \sum_i EN_{ij} \pi_i, \tag{3}$$

similarly we can use the following expression to find the total expected number of time spend in road network (CT):

$$CT = \sum_i ET_{ij} \pi_i, \tag{4}$$

where π_i is the steady state distribution of the Markov chain. $CN(t)$ will be CN for each epoch of Markov chain transmission.

Theorem 3.1. *Define the number of AV deployed in a road network over time as $CN(t)$ and $S(t)$ the number of Markov chain transmission. The road network throughput is $T_{AV} = CN/CT$.*

Proof. The road network throughput can be found using:

$$T_{AV} = \frac{CN(t)}{t}.$$

assuming t_0 is the epoch of the first transition, we can find $CN(t)$ and t upper limit from the following expression using the Markov Chain property:

$$CN(t) \leqslant CN(t_0) + \sum_{i=2}^{S(t)+1} CN_i$$

and

$$t \leqslant t_0 + \sum_{i=2}^{S(t)+1} T_i.$$

Therefore the upper limit for T_{AV} is

$$\lim_{S(t)\to\infty} \frac{\sum_{i=2}^{S(t)+1} CN_i / S(t)}{\sum_{i=2}^{S(t)+1} T_i / (S(t)-1)}$$

The lower limit for T_{AV} can be similarly found [20], therefore T_{AV} can be:

$$T_{AV} = CN/CT.$$

.

According to theorem 3.1, we can identify the throughput of the road network using the Markov chain transition probability. This transition probability demonstrate the expected amount of AVs in each state (moderate and severe congestion). The expected number of cars will be in severe or moderate congestion is going to be determined by choosing these threshold levels. By showing that a lower and upper limit for the T_{AV} merges to CN/CT, we proved 3.1. In the road network, a deployment scheme using the road traffic data and designed threshold levels, we can find the optimal T_{AV}. Using theorem 3.1 together with (3) and (4) can provide the designers with a trade off for AV deployment. In the numerical analysis section, we demonstrate that based on values used for τ_i, we can find the best T_{AV} for different road networks.

4 Scheduling via Queueing Theory

Queueing theory has been used extensively for various networks for resource management and scheduling [17]. Qeueing theory is a study of expected waiting times in queue. A queueing model is used to predict the expected waiting time and queue length. This helps to effectively mange limited resource that is going to be shared between various parties. The state of a queue is defined by a process in which the arrivals and departures from the queue, along with the number of users (in this work AVs) currently in the network. An arrival process is determined by the arrival probability distribution that best describes the users arrival.

Here we aim to use queuing theory to manage and implement the transportation tasks' priority (Fig. 4). We aim to implement a model so that queue lengths and waiting time for each transportation task can be predicted. We use a single queueing node for each deployment center using Kendall's notation. If the transportation task for each priority arrives with a Poisson process, and for the case of AVs they pass a route in deterministic time, then we have a M/M/1 queue (in the notation, the M stands for Markovian and M/M/1 means the queue in a road network having one road that is the server, where AV arrivals are determined by a Poisson process and passage times have an exponential distribution). Various scheduling policies can be used at queuing nodes such as first in first out or shortest job first. The waiting time for each AV can be predicted and an optimal scheduling scheme can be designed. Our goin in load balancing is that considering that road capacity is a limited resource, a pricing scheme can be introduced and numerically assessed by queuing theory in order to implement an optimal assignment for different transportation tasks.

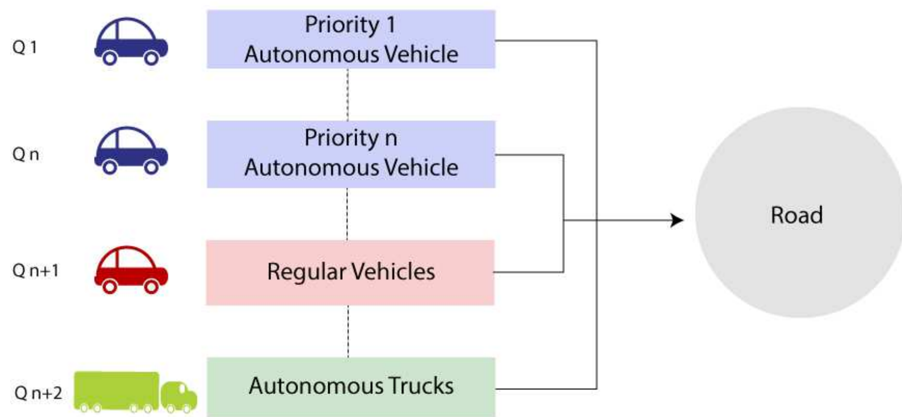

Figure 3: Assigning different transportation types for sharing a road network.

5 Traffic Load Balancing

We defined a transportation task as a task to be performed by an AV or an autonomous truck. We aim to find optimal scheduling for transportation tasks based on the constraints on resources, such as availability of AVs or autonomous trucks, time requirements for each task, and the traffic and road condition. Currently, popular delivery services are providing various options such as a range between one day and up to three week shipping options, each with different pricing. A well defined resource management algorithm with a pricing scheme can lead to an optimal scheduling for AVs and autonomous trucks. For example, an autonomous delivery can be scheduled to use the roads with less rush hour traffic for transportations tasks that have low sensitivity to delivery time.

Suppose we have n transportation tasks. By assigning each priority task described above (Fig. 4) to share a portion of the server in the queue (here the road network), we can find the waiting time for each task according to the following [17]:

$$W_s = \frac{1}{\mu - \lambda}, \tag{5}$$

where W_s is total waiting time including service time, μ is the average service rate, and λ is the average arrival rate. Here by assigning the portions of μ, the AV fleet management can assign each transportation task share of the road network (eg., high priority tasks get higher share). For different queue types based on different arrival distribution, different formulation for W_s can be found. In section 4, we introduced different transportation types that can have various arrival distribution. As one example, an emergency transportation has lower arrival rate, but should have assigned higher portion of μ.

6 Numerical Analysis

First, we investigate the service time according to (5) in the road network. In Fig. 3, we increase the number of AV in the road network from a share of 0.1 to share of 0.9 and see the waiting time will drop when most of the cars in the network are AVs. This is intuitive since the AVs are following the congestion avoidance algorithm introduced before and as a result they will reach a social optimal throughput. For regular cars, when the share of the AVs in road network increases, the portion of μ they are allocated according to (5) is decreased. As a result their waiting time increases, hence they do not change their scheduling algorithm.

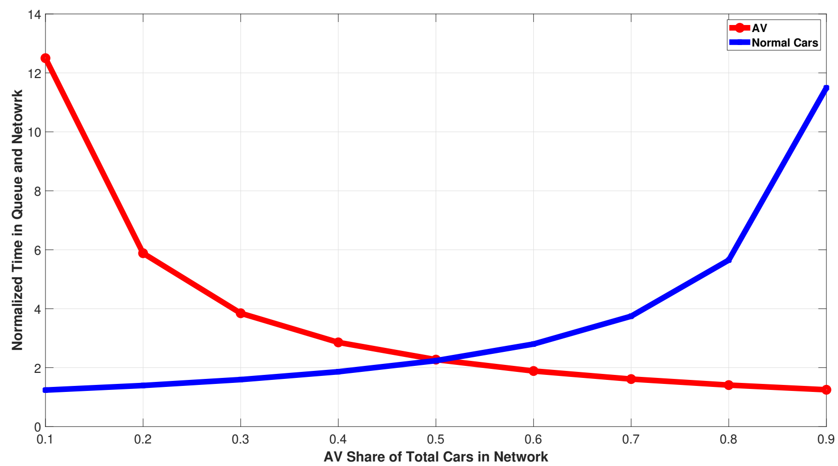

Figure 4: Normalized waiting time based on the share of AVs in a road network.

To investigate the threshold level used for Markov chain, τ_i, we used three methods for different pairs of τ_i. Method 1 has low τ_1 and τ_2. Method 2 has low τ_1 and high τ_2. Method 3 has high τ_1 and τ_2. The road network throughput with increasing number of AVs is presented in Fig. 5. This result demonstrates that for each road network based on the physical attributes of the road an optimal AV fleet size can be found.

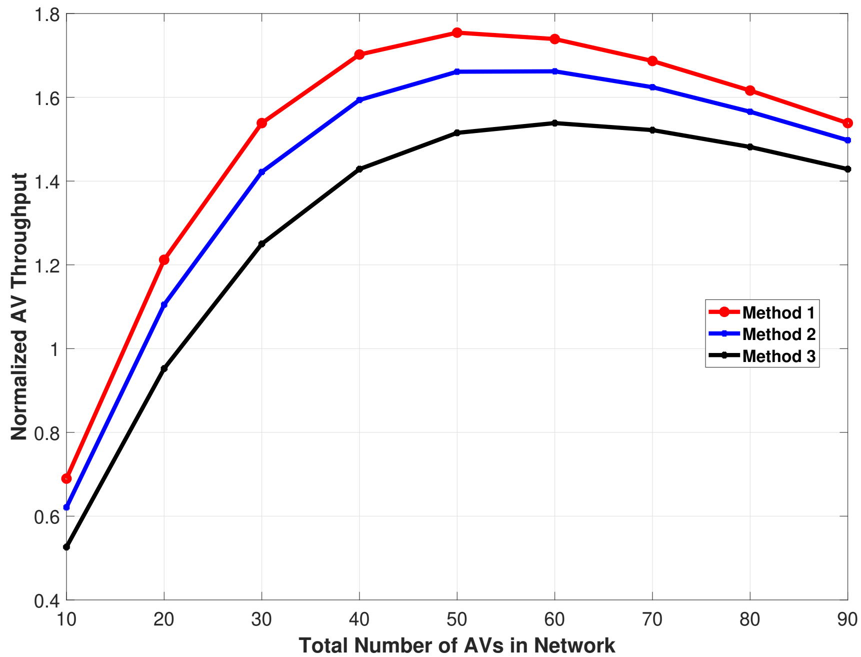

Figure 5: Normalized road throughput for various congestion threshold levels based on the share of AVs in a road network.

To validate our congestion avoidance algorithm, we modeled deployment of fixed number of regular cars using MATLAB. As it can be seen in Fig. 6, when AVs share increases in the road network, their throughput increases since they follow the congestion avoidance algorithm. This result closely correlates with our understanding that if the majority of cars follow a congestion avoidance algorithm, the overall throughput of the road network will improve.

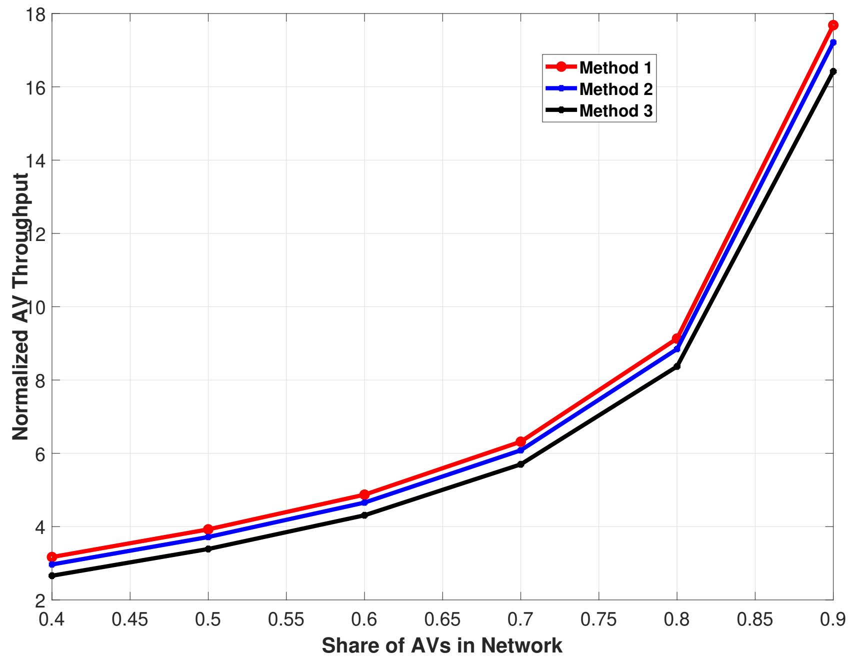

Figure 6: Normalized road throughput for a fixed number of regular car for various congestion threshold levels.

7 Conclusion

In this work, we proposed a novel adaption of a congestion avoidance algorithm based on additive increase and multiplicative decrease (AMID) in computer networks for AV fleet management. We define various transportation task to share a road network and by using queueing theory we implement and evaluate this assignment. Our numerical analysis suggests that our algorithms improve the road network throughput significantly. The results demonstrate that a congestion avoidance algorithm must be tailored to physical and traffic data of a road network in order to achieve an optimal network throughput.

Studying the impact of hybrid AV is the future research direction. During transmission to an AV transportation system, there will be cars that can operate in both autonomous mode and driver mode. Some drivers may want to optimize their route and as a result this may cause to a road network throughput that is less than social optimal. A game theoretic approach similar to ([17]) must be evaluated to force these greedy drivers to follow the congestion avoidance algorithm to achieve the social optimal road network throughput.

References

[1] FHWA Report *Online:* *https://safety.fhwa.dot.gov/facts_stats/*. (2018).

[2] Jerome M. Lutin, Alain L. Kornhauser, and Eva Lerner-Lam MASCE. *The revolutionary development of self-driving vehicles and implications for the transportation engineering profession*. Institute of Transportation Engineers. ITE Journal, vol. 83 no.7 (2013).

[3] United States Census Bureau *Online:* *https://www.census.gov/newsroom/press-releases/2018/cb18-41-population-projections.html*. 2018.

[4] AV Deployment *Online:* *https://www.inverse.com/article/42509-self-driving-uber-death-cities-still-testing-autonomous-cars*. 2018.

[5] Dmitri Dolgov, Sebastian Thrun, Michael Montemerlo, and James Diebel. *Path planning for autonomous vehicles in unknown semi-structured environments*. The International Journal of Robotics Research, vol. 29 no.5 (2010) pp. 485-501.

[6] Chad Goerzen, Zhaodan Kong, and Bernard Mettler. *A survey of motion planning algorithms from the perspective of autonomous UAV guidance*. Journal of Intelligent and Robotic Systems, vol. 57 no.1-4 (2010).

[7] Daniel J. Fagnant and Kara M. Kockelman. *The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios*. Transportation Research Part C: Emerging Technologies, vol. 40 (2014) pp. 1-13.

[8] Wilko Schwarting, Javier Alonso-Mora, and Daniela Rus. *Planning and decision-making for autonomous vehicles*. Annual Review of Control, Robotics, and Autonomous Systems, (2018).

[9] Autonomous Cars Likely To Increase Congestion *Online:* *https://cleantechnica.com/2016/01/17/autonomous-cars-likely-increase-congestion/*. 2018.

[10] Khashayar Kotobi, Philip B. Mainwaring, and Sven G. Bilén. *Puzzle-based auction mechanism for spectrum sharing in cognitive radio networks*. Wireless and Mobile Computing, Networking and Communications (WiMob), IEEE (2016) pp. 1-6.

[11] Macro Pavone, Stephen L. Smith, Emilio Frazzoli, and Daniela Rus. *Robotic load balancing for mobility-on-demand systems*. The International Journal of Robotics Research, vol. 31 no.7 (2012) pp. 839-854.

[12] Rick Zhang and Marco Pavone. *A queueing network approach to the analysis and control of mobility-on-demand systems*. American Control Conference (ACC), (2015) pp. 4702-4709.

[13] Kyle Treleaven, Marco Pavone, and Emilio Frazzoli. *Models and efficient algorithms for pickup and delivery*. Decision and Control (CDC), 2012 IEEE 51st Annual Conference on, (2012) pp. 5691-5698.

[14] Rick Zhang, Federico Rossi, and Marco Pavone. *Routing autonomous vehicles in congested transportation networks: Structural properties and coordination algorithms*. Autonomous Robots, vol. 42 no.7 (2018) pp. 1427-1442.

[15] Christian Ballarin and Martin Zeilinger.. *The Truck of the Future: Autonomous and Connected Driving at Daimler Trucks*. No. 2017-01-1931. SAE Technical Paper, (2017).

[16] Richard Viereckl, Dietmar Ahlemann, Alex Koster, and Sebastian Jursch. *Connected Car Study 2015: Racing ahead with autonomous cars and digital innovation.*. Strategy `http://www.strategyand.pwc.com/reports/connected-car-2015-study` (2015).

[17] Khashayar Kotobi and Sven G. Bilén. *Spectrum sharing via hybrid cognitive players evaluated by an M/D/1 queuing model*. EURASIP Journal on Wireless Communications and Networking vol. 1 (2017) pp. 85.

[18] Khashayar Kotobi, Philip B. Mainwaring, Conrad S. Tucker, and Sven G. Bilén. *Data-throughput enhancement using data mining-informed cognitive radio*. Electronics vol. 4 no.2 (2015) pp. 221-238.

[19] Robert G Gallager. *Discrete stochastic processes*. OpenCourseWare: Massachusetts Institute of Technology (2011).

[20] Chunmei Liu and Eytan Modiano. *On the performance of additive increase multiplicative decrease (AIMD) protocols in hybrid space-terrestrial networks.* Computer Networks vol. 47 no. 5 (2005) pp. 661-678.