university of groningen / faculty of science and engineering

Using Auctions for Traffic Moderation in Road Intersections

Bachelor's Project Thesis

Leonidas Zotos, l.zotos@student.rug.nl
Supervisor: Prof Dr D. Grossi

Abstract: A large portion of our everyday life is spent on the road traffic network during commutes. This has been linked to various poor health outcomes. To decrease the time spent on the road network, different measures have been taken (e.g. intelligent traffic lights). However, even state-of-the-art mechanisms do not take into consideration the individual urgency of drivers. To tackle this problem, auction-based approaches have been explored in the past. There, drivers waiting at a road intersection can place bids to receive priority for the use of the intersection. It was previously found that the utilisation of auctions for traffic moderation can lead to significantly reduced waiting times. However, an aspect that has not been sufficiently explored is the performance of different bidding strategies. In the current research, a multi-agent traffic network was set-up in which five different bidding strategies were tested. A number of simulations were run with varying distributions of present bidding strategies. At a single-agent level, the reinforcement-learning bidding strategy was found to perform best. However, using various measures, it was found that a traffic network only consisting of drivers that use adaptive bidding led to the best performing society.

1 Introduction

In today's world, we spend a large portion of our everyday life in commuting from one place to another, whether that is to have a nice dinner at a restaurant, be on time for work or catch a last minute flight. In fact, according to a study conducted in the United States, the mean time spent in commuting from home to work and vice versa is 62 minutes per day (Christian, 2012). The same study found that to accommodate this time cost, employees sacrificed sleep and physical activity time, which can consequently contribute to 'obesity and other poor health outcomes'.

A significant portion of this commuting time is spent at intersections waiting on traffic lights, or even worse, in traffic jams at 5.30 PM when everyone is making their way back home after a long day of work. As the population in cities increases, finding a solution to this problem has become increasingly important. To date, numerous solutions have been put in place with varying success. These solutions range from developing intelligent traffic lights to expanding the road network (in an attempt to uniformly distribute traffic).

Another solution that is becoming increasingly prominent is the use of autonomous vehicles (AVs). One of the main advantages that comes with the use of AVs is their ability to quickly react to environmental stimuli. This does not only decrease the time it takes for them to start at a green traffic light, but can also help in preventing traffic accidents. Another major advantage is the ability of AVs to quickly communicate with each other. For instance, to simply share internal properties (eg. speed, acceleration) or to make the intentions of the car known (eg. when overtaking or joining lanes).

The traffic moderation methods mentioned above attempt to tackle the problem from different sides. However, an aspect that has not been considered is the individual hurry, or rush, of the drivers. As an illustration, a traffic light will treat the driver rushing to catch a flight in exactly the same way as the driver going to the grocery store. That is to say, drivers value their time differently depending on whether they are in a rush or not.

As it has been previously established, a significant part of this commuting time is spent on road intersections, where the priorities are distributed using traffic lights. This can be seen as a supply/demand system, in which the supply or product is the priority (to cross the intersection) and the demand is the combined urgency of the drivers waiting to cross the intersection. Since the different drivers value the priority differently, it would make sense to have a system that takes this into account. To achieve this, auctions can be used where the drivers waiting at the front of the intersection can place bids and the highest bidder receives the priority. Using this system, the urgency of the individual bidders can be taken into account through the placement of high bids when the driver is in a rush and vice versa.

1.1 Previous Literature

The idea of traffic network moderation with the use of auctions has been explored in the past. Specifically, a research by Vasirani and Ossowski examined the use of *competitive traffic assignment* (or CTA for short) to moderate traffic (2012). In short, 'traffic assignment' refers to the problem of evenly distributing the traffic over the road network. In their approach, users of the road network can pay a certain *fee* to cross each intersection. This fee can vary between intersections and is set by the *intersection manager* of each intersection, who aims at increasing his revenue. As drivers can autonomously decide on what route to follow, intersection managers are incentivised to provide competitive prices. According to Vasirani and Ossowski, this market-based system will eventually lead to an efficient allocation of resources (i.e. access to intersections).

Another study conducted, at the University of Texas, found that using auctions for traffic moderation can lead to significantly reduced waiting times per car, compared to the status quo (Carlino, Boyles, and Stone, 2013). This research also used an agent-based simulator (named AORTA). In their study more focus was put in how individual agents can bid, compared to the study by Vasirani and Ossowski. They also introduced the idea of using a *benevolent system wallet* that 'intervenes to maintain various fairness properties'. As an illustration, this wallet can be used to prevent traffic jams by, for example, boosting the bid placed by a driver in a busy queue.

While the research by Vasirani and Ossowski focuses on the systemic level, Carlino et al. put more focus on individual road intersections. Research has also been done at a more granular level. Specifically, Schepperle and Böhm explored how the idea of traffic moderation using auctions can be combined with the concurrent utilisation of intersections, which comes closer to the current reality (2008). It was concluded that, the concurrent use of intersections by multiple drivers led to significantly higher user satisfaction compared to state-of-the-art mechanisms.

Moreover, a research by Rey and colleagues highlighted the issues of incentive compatibility (i.e. that users cannot achieve higher utility by bidding untruthfully) when using auctions in a dynamic traffic network (Rey, Levin, and Dixit, 2020). Here, by 'dynamic traffic network', a traffic network in which cars do not arrive at the same time is assumed. In their implementation, cars privately report their 'delay cost'/bid to the intersection manager/auctioneer. The auctioneer then can decide the order of priorities. This implementation also allows the concurrent use of intersections, as, for example, the car with the 2nd highest priority can already start if its path does not overlap with the preceding car's path. They also used Markov chain models to predict users' waiting time in both scenarios of static and dynamic traffic intersections. They also concluded that an online traffic intersection mechanism can be used to maximise social welfare by ensuring truthful placement of bids.

So far, the approaches that have been discussed use a centralised intersection moderator. In a novel approach, Censi and colleagues introduce a *karma* system. Here, karma functions as a credit system: cars can give karma to other cars waiting at the intersection to receive priority. From the opposite perspective, it is possible for cars to 'pass on' using the intersection now, so that they can receive karma, which they can use when they are in a hurry. It was concluded that a society of prudent (albeit individualistic) agents leads to a society with higher social welfare, compared to a society of agents that look for instant gratification (by quickly using up their karma) (Censi, Bolognani, Zilly, Sadat Mousavi, and Frazzoli, 2019).

A research by Lin and Jabari (2020) followed a

similar cooperative approach in which transferable utility games can be used for the pricing of the priority to the intersection. In game theory, a 'transferable utility game' is a game in which players are able to transfer a 'commodity' among themselves (often money), so that "any player's utility payoff increases one unit for every unit of money that he gets" (Myerson, 1991). In this case, agents with different valuations of time can engage in trading priority for direct monetary compensation (Lin and Jabari, 2020). In their research, empirical evidence is provided showing that adversarial behaviour is inhibited in such a system. However, as pointed out by Rey and colleagues, no formal proof of incentive-compatibility is presented (2020).

1.2 The Current Research

The aforementioned research has highlighted the possibilities offered by the use of auctions to moderate traffic in road networks from multiple aspects. At the same time, a side of this field that has not been explored sufficiently is the effect and performance of different bidding strategies. Thus, the current thesis is based on the research question: "How do different bidding strategies affect the bidders' waiting time?". Although there are numerous types of auctions, in this research only first-price sealed-bid auctions are used. Section 2.3 provides some more detail on the functioning of this auction and the rationale for choosing it.

The research question can be explored at 2 different levels. First, it is interesting to find the bidding strategy that performs best at an individual level (by reducing the bidder's overall waiting time). To give this some context, we can consider a society in which car manufacturers compete to develop superior bidding strategies to offer the best customer experience.

Moreover, the same topic can also be seen at a systemic level. Here, the interest lies in finding the bidding strategy that leads to the fairest distribution of waiting times. More in detail, societies with varying portions of bidding strategies can be examined. For example, there can be a society in which all cars are using bidding strategy A but there can also be a society in which 2/3 of cars use bidding strategy A the rest use bidding strategy B. Therefore, it is intriguing to find which society of car bidders leads to the highest equality, in terms of waiting time and driver urgency. Additionally, the efficiency of the most equal society is also worth evaluating, as it might be the case that an unequal society is overall more efficient. The exact way in which the systemic inequality and efficiency are measured will be described in section 3.2.

1.3 Operationalisation

Similarly to the literature (Carlino et al., 2013; Censi et al., 2019; Lin and Jabari, 2020; Rey et al., 2020; Schepperle and Böhm, 2008; Vasirani and Ossowski, 2012), a multi-agent system was designed to tackle the research questions stated above. In simple terms, a multi-agent system is a system that includes a number of 'autonomous entities' with potentially different knowledge and goals (Shoham and Leyton-Brown, 2008). In this instance, the system simulated road-traffic in a grid. Priorities for the use of the intersections are given through the use of first-price sealed-bid auctions (more detail on this decision is provided in section 2.3). Five different bidding strategies were implemented, of which two consider the individual hurry of the driver. The strategies will be described in detail in section 2.2.4.

For each research question, a different experiment was designed. The first experiment evaluated the individual performance of the bidding strategies by using an environment where all strategies are used by different agents (each strategy is used by 20% of the agents).

The second experiment aimed at finding the society with the highest equality. In this case, societies with various distributions of bidding strategies are evaluated. Naturally, homogeneous societies (e.g. a society where 100% of the agents use bidding strategy A) were expected to have the highest equality in terms of waiting time and driver urgency. However, it is still interesting to find which homogeneous society has the lowest variability, after a set number of iterations. The second experiment also evaluates, using two different measures, whether the fairest society is also the most efficient.

It is worth mentioning that for both experiments, performance was measured through a combination of the actual waiting time and the urgency of the driver at that point. That is to say, losing an auction when the individual driver urgency is high leads to a higher score penalty, compared to when the urgency of the driver is low. The exact way

in which the performance is calculated will be explained in section 2.3.

Last but not least, a description of the exact experimental setup is provided in section 3.

2 System Description

To answer the question raised in subsection 1.2, a multi-agent system approach was taken. The model used can be divided into 3 main parts. First, there is the road network, which simply consists of the set of intersections. The second part is the cars in the network. Finally, the auctions, or, in other words, the interaction between the road network and the cars, is another crucial part of the model. Naturally, these 3 parts are largely intertwined, but for clarity purposes, they have been separated in the following three subsections.

Last but not least, section 2.5 will describe some other, more general, aspects of the model. This includes, among others, the initialisation of the traffic network.

Figure 2.1 illustrates a grid of size 2x2. Its individual components are going to be explained in the following subsections.

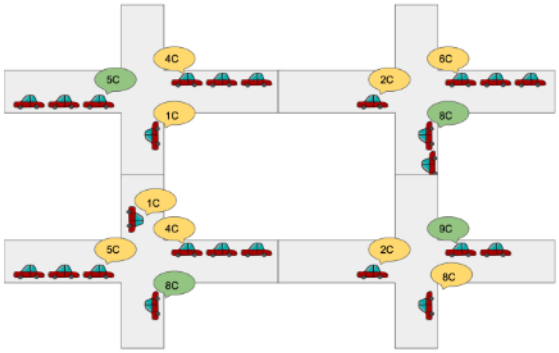

Figure 2.1: A 2x2 Grid with 4 intersections and some cars. The 1st car of each intersection bids an amount of credit (e.g. 4C). The highest bidder receives priority (green).

2.1 Environment

In this subsection, the general road network that was implemented will be described. The road network can be seen hierarchically as individual road lanes, that compose intersections, which in turn compose the grid.

2.1.1 Lanes

Starting from the granular level, each lane has an orientation/side and a maximum car capacity.

First, the orientation is simply one of the 4 basic orientations: North, East, South or West. This orientation does not signify the direction of the cars in the lane. Instead, it signifies the relative orientation of the lane with regards to the center of the 4-way intersection in which it belongs.

Moreover, for this implementation, it was chosen to use the same maximum car capacity for all lanes, as to make the model less stochastic, while also simplifying it. More in detail, if a lane has reached its car capacity, cars destined for that lane will not be able to enter and instead will have to wait.

2.1.2 Intersections

Lanes can be found in intersections. In this model, only 4-way intersections are used and a total of 4 lanes join at the intersection. In reality, intersections often have more than 1 lane joining the intersection from each side. However, since the focus of this research is on the performance of the bidding strategies, it was deemed appropriate to keep the model as simple as possible, by considering the simplest type of traffic network (i.e. with only 1 lane per side).

2.1.3 Grid

Intersections were placed next to each other in the form of a grid. This grid only consists of intersections and no empty lanes. That means that once a car exits an intersection, it immediately joins the queue of another intersection. In a way, this is realistic as road networks are effectively just a set of intersections.

It is worth mentioning that intersections at the edge of the grid also have 4 sides, as cars can also move into the grid from the edge. The behaviour of the cars will be described in more detailed in section 2.2. As such, since the grid only consists of intersections and each intersection has 4 sides, a 5x5 grid will have 25 intersections and 100 individual lanes. Furthermore, using a lane capacity of 4 cars,

the entire grid can contain up to 400 cars (although this causes complete traffic congestion).

2.2 Agents

Naturally, in this multi-agent system, the agents are the cars that transit through the road network. All cars have a destination, a bidding strategy, a driver-urgency and a credit balance. Each of these elements will be discussed in the following subsections.

2.2.1 Movement

The movement of the drivers is based on the direction they follow. Each time a car is placed on the grid, its direction is North-East, North-West, South-East or South-West. Consequently, all cars aim on moving diagonally. However, since a grid does not allow diagonal movement, in each intersection cars choose one of the two directions that interests them with equal probability. For example, a car moving North-West will always move North or West.

Another possibility would be to also include the main directions North, East, South and West. However, this was not done since realistically, the probability of a car only wanting to move in one of the cardinal directions (North, East, South or West) is quite low. This type of movement is still possible in the current implementation, as there is a small probability that a car will always pick the same direction (e.g. in a trip of 4 intersections, there is a 6.25% chance that the car will consistently move towards one direction).

So far, all cars have their own direction but they do not have a destination. For this, there are two alternative scenarios: a car might want to leave the grid or they might want to move to a different point in the grid. The former is implemented by simply extracting the car from the grid if it reaches its boundaries. For example, if a car with North-West as its direction chooses to go West while being at the left-end of the grid, the car is removed from the traffic network (i.e. no period boundaries condition was used).

To implement the scenario where the car's destination is within the grid, a probability approach was taken. More in detail, each car has a probability of reaching its destination after going through an intersection. This probability parameter is the same across all agents and has a default value of 20%. This mechanism entails that the probability of longer journeys is lower than the probability of shorter journeys. For example, there is approximately a 40% chance (0.8^4) that a car will go through 4 intersections without reaching its destination. An alternative to this mechanism is the implementation of route-planning. However, this was not done in order to increase the performance and speed of the overall simulation. Even so, the resulting behaviour is deemed to be sufficiently similar.

The two aforementioned ways in which cars can reach their destination result in the agents being temporarily removed from the traffic grid. At the end of each iteration, cars that left the grid are randomly placed across the grid. Section 2.5 describes this process in some more detail.

2.2.2 Credit System

Auctions are accommodated using a credit system. As previous literature has discussed, at least in the United States, transportation planning agencies are to avoid 'disproportionately high and adverse' effects on disadvantaged groups (Carlino et al., 2013). In other words, everyone should have the same rights while on the traffic network. Thus, a traditional currency should not be used for traffic auctions, as to not disadvantage people of a lower economical status.

In the current implementation, all agents receive a certain amount of credit after a number of iterations. This credit can only be used for traffic auctions and has, as its upper limit, the same credit amount as in every credit renewal. For example, after every 10 simulation iterations, the balance of all cars is set to 50 credit units. The difference between the agents is that they can decide how to spend their credits. For instance, they can spread it out over the iterations leading up to the next pay-day, or they can spend all the credit in the first few auctions. This credit system with a periodic renewal was also inspired by the work of Carlino and colleagues (2013).

2.2.3 Driver Urgency

A fundamental part of the performance of the different agents is their urgency, or hurry factor. This

hurry factor remains the same throughout each trip — from the moment the agent is placed on the grid up until the moment the agent reaches its destination and is removed from the grid. In this model, every time an agent is placed on the map, its hurry factor is picked from a uniform distribution ranging from 0 to 1, where lower hurry-factors entail lower urgency.

2.2.4 Bidding Strategies

Five bidding strategies are implemented. The bidding strategy determines the amount of credit that is placed as a bid in each auction. Agents are given one of the 5 strategies at the start of the simulation (the strategy remains the same, even after an agent completes a journey). It is important to mention that when assigning strategies to agents, they are not necessarily picked from a uniform distribution. Instead, custom distributions can be set. For example, it is possible to create a simulation in which only 2 bidding strategies are present.

For clarity, the bidding strategies will be described in detail in section 2.4, after introducing auctions in section 2.3.

2.3 Auctions

In this model, priorities for the use of the intersections are assigned through auctions. While there are numerous types of auctions, for this simulation, the first-price sealed-bid auction was used. In this type of auction, all participants place a single bid at the same time that cannot be changed/adapted later. Since bids are placed at the same time, bidders are not aware of the other bidders' bids. This auction type was mainly chosen due to its simplicity in the implementation. Specifically, knowledge does not have to be transferred between agents. Only the main administrator has to let the agent know if they won or lost the auction and then the winning agent will know how much they have to pay. This auction type was also preferred over auctions that allow successive bidding (e.g. English auction), for two main reasons. First, such auctions require more information to be transferred across agents, slowing down the overall network. Additionally, it is difficult to appropriately decide when the auction should close (e.g. closing at a fixed time is possible, but it can be difficult to decide on an appropriate closing time) (Shoham and Leyton-Brown, 2008). However, in contrast to second-price sealed-bid auctions (in which the winner pays the highest rejected bid), first-price sealed-bid auctions are not incentive compatible. That is to say, users can theoretically achieve higher utility by bidding untruthfully. For most bidding strategies, this is not a problem, as their bids are statically programmed as to not use untruthful bidding. However, the reinforcement learning bidding strategy (described in section 2.4.5), might have learnt to use untruthful bidding. To keep the structure of this section clear, this issue is further discussed in the discussion section (sections 6.2 and 6.3).

To further reduce traffic congestion, a handicap system was implemented in which the bids placed by cars in busy lanes are increased. By doing that, busy lanes are more likely to receive priority, thus reducing congestion. In this implementation, this handicap is a multiplier that changes depending on the number of cars waiting in the lane. This multiplier is defined by: $m = 1 + (l \cdot 0.15)$, where m is the multiplier and l is the length of the car queue. For example, if there are 2 cars waiting, the multiplier used for the bids of that lane is 1.3. Since the multiplier would not affect a bid of 0 credits, all bids are increased by 1 credit unit. It is worth noting that even if a single car is waiting, the multiplier is still 1.15. This entails that all bids are first increased by 1 and then multiplied by at least 1.15. The calculation of the final bid is described by equation (2.1).

$$b_f = (b_p + 1) \cdot (1 + (l \cdot 0.15)) \tag{2.1}$$

Here, b_f is the final bid, b_p is the bid placed by the bidder and l is the total number of cars waiting in the same queue of the bidder (including the bidder).

The first car of each lane participates in the auction and therefore up to 4 cars can participate. Naturally, no bids are placed from lanes that are empty. Bids are also not placed by cars that are destined towards a full queue, as, even if they win the auction, they will not be able to move. In general, the car with the highest final bid receives priority. In case of a tie, the priority is randomly given to one of the bidders of the tie. The car that received priority in the end pays the amount b_p and not the final amount b_f. In other words, they pay the amount

they bid and not the amount after the adjustment. This was mainly done to ensure that drivers have enough credits to pay, but also to increase the consistency of the bidding strategies.

Throughout the simulation, agents keep track of their score. After participating in an auction, the agent's score is adapted. The score reflects the time waited at the front of an intersection, weighted by the urgency of the driver. Therefore, a higher score is worse as it entails that the driver waited longer. Specifically, if a driver wins the auction, the score does not increase or decrease. However, if the auction is lost, the score increases by a hurry-factor. For example, if a car loses an auction when their hurry-factor is 0.75, its score is increased by 0.75. It is worth noting that if the hurry-factor is 0, then the score does not change (as the car does not care if it receives priority).

2.4 Bidding Strategies

As mentioned previously, five different bidding strategies were implemented. These are: Free-Riding, Random Bidding, Static Bidding, Adaptive Bidding and Reinforcement Learning (RL) Bidding. The first three are fairly simple and do not consider the individual urgency of the drivers, while Adaptive bidding and RL Bidding do consider it. It is important to note that the bidders are only aware of themselves. In other words, they do not know how many bidders participate in the auction or what their balances and strategies are. Therefore, they have to make a decision given (1) their balance, (2) their hurry factor and (3) the duration between pay-days. Furthermore, the bids placed by the agents are adapted using the handicap described in section 2.3. As mentioned previously, the highest bidder only pays the amount of credit that they bid, before the system adjusted it according to the handicap.

2.4.1 Free-Riding

The first and simplest strategy is inspired by the Free-Riding strategy as described by Carlino and colleagues (2013). Essentially, Free-Riders always bid 0 credits. This, of course, is a witless (for lack of a better word) strategy, as credit cannot exceed the renewal amount and does not influence the final score (i.e. not using credits does not lead to a better score). However, it was considered to be a simple base-line and exhibits the functioning of the traffic-congestion handicap, as described in section 2.3.

2.4.2 Random Bidding

This strategy is also fairly simple. Every time the agent has to bid, they place a bid ranging from 0 to the full balance (all credits). The amount that they place as a bid is picked from a uniform distribution. An alternative would be to pick the amount from a normal distribution with a mean of half the balance (as an example), but a uniform distribution was chosen for simplicity.

2.4.3 Static Bidding

Static Bidders also do not consider the urgency of the driver. However, in contrast to Free-Riders and random bidders, they do consider the period between pay-days. Effectively, at the start of the simulation, when the period between pay-days and the credit-renewal amount is announced to all agents, they decide how much to bid in every auction they partake. As the name implies, they bid the same amount in every auction. For example, if the period between pay-days is 10 iterations and the credit-renewal amount is 20 credits, they will bid 2 credits in every auction. It is worth noting that this strategy is based on the assumption that the agent will partake in an auction in every iteration, which is not always the case. In a way, it is fairly conservative and makes sure the agent has sufficient funds for every possible auction.

2.4.4 Adaptive Bidding

Adaptive bidding is based on Static Bidding, but also takes in consideration the hurry factor. In short, it starts off with the static bid (calculated as explained in section 2.4.3) and proportionally increases/decreases it depending on whether the driver is in a hurry. This is done according to equation (2.2).

$$b_p = b_s + (b_s \cdot (H - 0.5)) \tag{2.2}$$

Here b_p is the placed bid, b_s is the static bid as described in section 2.4.3 and H is the hurry factor of the driver. For an H value of 0.5, the placed bid is the same as the static bid. However, in extreme

cases, where the urgency 1, the placed bid is 50% higher than the static bid. The value of 0.5 was considered appropriate as increasing it would lead to very radical adjustments.

Once the desired bid is calculated according to equation (2.2), the agent checks if their balance is actually sufficient to place this bid. If it is not, the agent places a bid equal to their balance.

In short, this strategy tries to plan ahead by saving up when the urgency is low.

2.4.5 RL Bidding

To implement a Reinforcement Learning bidding strategy, the Monte Carlo (MC) algorithm was chosen. A main property of this algorithm is that it improves based on learning episodes (a series of decisions that lead to a result) (Barto and Sutton, 1998). It is important to mention that in this implementation the agent uses as a starting point the static bid (just like the static and adaptive strategies) and learns how to increase/decrease it depending on the situation.

The MC algorithm uses a lookup table to estimate the utility of different actions given the current model state. Formally, this utility is $Q_t(s, a)$ where Q is the expected utility at time point t of performing action a given the current state s. As such, each (s, a) pair has a different utility.

The state-space is defined as the number of total possible states the agent can encounter in the model. The larger it is, the harder for the agent to learn how to adapt. Therefore, when designing the RL agent, it is important to distinguish the aspects of the model that are essential for the agent to know and the aspects of the model that are not useful for the agent to make a good decision.

The Learning Episode In this implementation, a learning episode consists of the auctions between pay-days. If a shorter period was considered as the learning episode (e.g. a single auction), the strategy would adapt as to maximise the performance of the agent in a single auction, which would simply mean that the agent would learn to bid its entire balance. Thus, by defining a learning episode as the set of auctions between pay-days, the agent should learn to spread out the funds appropriately. It is important to note that in this case, not all episodes are of equal length. For example, if the period between pay-days is 10 iterations, the agent can participate in 10 auctions. However, if the network is busy, the agent might spend most of the time waiting in queues, which would result in them participating in only 4 auctions (as an example).

After each pay-day (except the first), the previous learning episode is evaluated and given a score, R. In this implementation, the score is the average score of all auctions in that episode. The score, s, of an individual auction is defined according to equation (2.3). If the auction is won, $+H$ is used, but if the auction is lost, $-H$ is used. This results in lost auctions having scores between 0 and 0.5 depending on the hurry factor. Similarly, auctions that are won have a score between 0.5 and 1. A score close to 0.5 means that the hurry-factor, H, is close to 0 (driver is indifferent).

$$s = \frac{(\pm H + 1)}{2} \tag{2.3}$$

States As mentioned previously, agents have to make their decisions only knowing the following: (1) their balance, (2) their hurry factor and (3) the duration between pay-days. Here, the RL states simply consist of the hurry factor and the time until the next pay-day. Although counter-intuitive, the current balance was not taken into consideration for two reasons. First, in this way, the state-space is significantly reduced. Secondly, since the RL agent learns how to adapt the static bid, adjustments that are likely to exceed the balance (eg. +400%) are indirectly inhibited. In the case where an agent makes such an extreme adjustment, the balance is immediately depleted. This results in the next few auctions to most likely be lost. Therefore, the overall episode receives a low score, which results in the extreme adjustment to (most likely) not be chosen in the future. The exact way in which the agent learns will be described later.

In short, the state is a pair $\langle H, P \rangle$, where H is the hurry factor and P is the number of iterations left until the next pay-day. Since H can technically take any value between 0 and 1, to reduce the state-space, binning was used. Specifically, the hurry factor was rounded, so that it can take up to 100 different values.

Actions The lookup table consists of state-action pairs where each of these pairs receives an expected

utility value. The actions here are the possible adjustments that can be made to the static bid. This adjustment ranges from -1 up to 4, with intervals of 0.1. An adjustment of -1 entails that the bid is reduced by 100% and similarly an adjustment of 4 entails that the bid is increased by 400%. Since 0.1 intervals are used, for each state there are 50 different possible actions/adjustments.

Learning Initially, all (s, a) pairs have the same utility, which was set to 0.5. After each episode, the utilities of the state-actions performed in the episode are adjusted. The adjustment is described by equation (2.4) (Barto and Sutton, 1998). Here, R is the score of the episode (ranging between 0 and 1, higher is better), $Q_{t+1}(s, a)$ is the utility of state-action pair in time point $t + 1$ and α is the learning rate (i.e. how radical the adjustments are).

$$Q_{t+1}(s, a) = Q_t(s, a) + \alpha(R - Q_t(s, a)) \quad (2.4)$$

An important aspect of this implementation of the Monte Carlo algorithm is that all RL-agents use the same lookup table. This was done for 2 reasons. First, since all RL-agents can contribute with their experiences towards the lookup-table, learning occurs much faster. From a more practical point of view, lookup tables are often very large and take-up a lot of memory. Therefore, having a different lookup table for each agent is impractical.

Action Selection In Reinforcement Learning, given a lookup table, agents face the exploration/exploitation dilemma (Barto and Sutton, 1998). That is to say, agents can either pick the action with the highest utility according to the lookup table (exploitation) or they can perform another (often random) action in hopes of finding a better alternative (exploration). In this implementation, this dilemma was solved with the commonly used ϵ-greedy approach (Barto and Sutton, 1998). Here, the hyperparameter ϵ is introduced, which is set between 0 to 1. Every time an action has to be chosen, there is an ϵ probability that a random action is chosen instead of the best action. For instance, using $\epsilon = 0.1$, there is a 10% chance to pick a random action.

2.5 General Remarks

In this section, some aspects of the model that do not fully fit in the other sections will be described.

2.5.1 Iterations

The model is based on iterations. In each iteration, 3 steps are taken. First, if it is a pay-day, the balance of all cars is renewed and RL-agents learn from the previous episode. Then, all auctions take place and the winners move. Lastly, cars that have reached their destination (and therefore are out of the network) are placed back in, with a new route and hurry factor. This placement of cars is done randomly: each car that has to be placed in the grid is placed at the end of a random queue.

It is worth mentioning that auctions take place asynchronously. Specifically, auctions take place starting from the top-left of the grid and finishing to the bottom right (row by row). Although unrealistic, this should not truly influence the results as cars that actually move are most likely to end up at the end of a new queue (and therefore they will not participate in 2 auctions during the same iteration).

2.5.2 Grid Initialisation

At the start of the simulation, the grid is populated with cars. This is done by going through each lane and populating it randomly with x cars, where x is in the range of 0 to N. Here, N is a parameter that has the same value for all lanes and is manually set. Naturally, if x is chosen to be higher than the capacity of the lanes (as described in section 2.1.1), the lane is populated up to its capacity.

3 Experimental Setup

As mentioned in section 1.2, the research question "How do different bidding strategies affect the bidders' waiting time?" can be tackled at the individual and systemic level. For each, a different experiment was designed. In terms of parameters, the only difference between the experiments was the distribution of bidding strategies present in the network. Both experiments were repeated 10 times in order to obtain accurate and representative results. Running more simulations was not possible

due to memory limitations. The set of all hyperparameters can be found in Appendix A, table A.1.

It is important to mention that the lookup table of the RL agents was only cleared after the completion of all 10 experiments. That is to say, RL agents did not have to go through the learning process every time a new experiment started. This choice was made in order to increase the realism of the experiments as cars would not reset their knowledge base in reality. Additionally, this enables RL agents to reach their full potential, making them more comparable to the other strategies that work at full capacity from the beginning.

For all experiments, an Ubuntu (version 20.04 LTS) virtual machine was used with 14GB of RAM and an AMD Ryzen 5 2600 six-core processor (although the program is single-threaded) running at 3.4GHz.

Since the data are created through simulations, conducting significance tests is a dubious method as significance (e.g. p-value below 0.01) can be achieved by running more simulations. However, since the current experimental setup is limited by memory resources (and therefore running an infinite number of experiments is impossible), significance testing is still reasonable and is done for completeness.

3.1 Experiment I

The first experiment tries to find the bidding strategy that performs best at the individual level. To do that, an environment where all bidding strategies are equally present was designed. That is to say, since there are 5 bidding strategies, 20% of agents used Free-Riding, 20% used Static Bidding etc. The performance of the agents was the average total waiting time of the drivers of each bidding strategy, weighted by their individual hurry-factors (as described in section 2.3).

Therefore, the independent variable is the bidding strategy and the dependent variable is the total score of each car. A pairwise t-test with Bonferroni correction (Galambos, 1977) is used to evaluate the significance of the results (i.e. the score difference between each possible pair of bidder types).

3.2 Experiment II

The second experiment attempts to find the society of bidders with the highest equality in terms of score. To evaluate this, all possible homogeneous societies were tested (i.e. 100% Free-Riders, 100% Static Bidders, 100% random bidders etc.). Therefore, the independent variable is the type of homogeneous society and the dependent variable is the total score of each car.

At this level, 3 measures will be used. First, the Gini coefficient will be calculated (using the total scores of each car) to evaluate the equality within the different societies. Equation (3.1) describes how the Gini coefficient can be calculated (Gini, 1912). In equation (3.1), G is the Gini coefficient, B is the area under the *Lorenz Curve* and A is the area between the *line of perfect equality* and the Lorenz Curve. Briefly explained, the *line of perfect equality* is a 45° line and the Lorenz Curve is the line created by cumulatively ordering the scores in increasing order. In short, a Gini coefficient of 0 implies that all agents have the exact same score (perfect equality). In contrast, a Gini coefficient of 1 implies that a single agent has the total score, while all other agents have a score of 0 (perfect inequality).

$$G = \frac{A}{A+B} \tag{3.1}$$

While the Gini coefficient measures the equality among agents, it is also important to measure the efficiency at a systemic level. To do that, the cardinal utilitarian social welfare function will be used, averaged by the total number of agents in each society (equation (3.2)) (Adler, 2019). The results will have to be averaged, as not all societies have exactly the same number of agents (i.e. societies with larger populations are more likely to have a higher overall score). Here, W is the social welfare, n is the number of agents and Y_i is the total score of agent i.

$$W = \frac{1}{n} \cdot \sum_{i=1}^{n} Y_i \tag{3.2}$$

Lastly, the 'least well-off' measure (also known as 'Rawlsian utility function') was used to evaluate the *worse case scenario* in each society. In this case, since higher scores are worse, the 'least well-off' measure is the highest individual total score of

each society (i.e. the score of the agent with the worst performance).

4 Results

The following two sections discuss the results obtained from experiment 1 and 2 respectively.

4.1 Individual Level

As mentioned previously, to evaluate which bidding strategy performs best, an environment with an equal proportion of agents was used. After each of the 10 simulations, the total score of each car was extracted. This resulted in a list of all cars from all simulations and their scores. To assess the performance, a box-plot was created (figure 4.1). It is important to emphasize that a higher score corresponds to a worse performance.

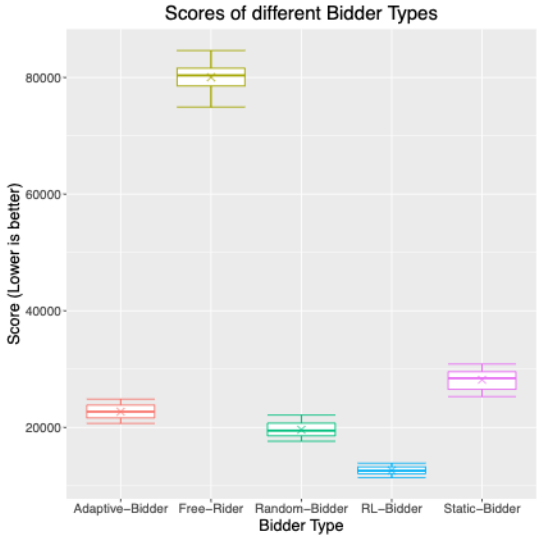

Figure 4.1: Performance of the different bidding strategies

As it can be seen in figure 4.1, RL Bidders perform significantly better compared to the rest.

As expected, Free-Riders have the worst performance and highest interquartile range. This is likely due to their dependence on the traffic network as a whole. Specifically, they mainly move when traffic jams are present, which are fairly stochastic.

Furthermore, it is also interesting to notice that random bidding performs slightly better than adaptive bidding and significantly better compared to Static Bidding. This is likely due to the boldness of the random bidding strategy, which will be discussed in more detail in section 6.1.

Table 4.1: Mean total score of each bidding strategy

Strategy	Mean Score	Standard Deviation
Free-Riders	80040	2051
Random Bidders	19604	1200
Static Bidders	28173	1550
Adaptive Bidders	22697	1140
RL Bidders	12629	641.7

To statistically evaluate the significance of the aforementioned differences, a pairwise t-test with Bonferroni correction was conducted. The score difference between each pair of bidder types was found to be significantly different ($p < 0.001$). The mean total score and standard deviation of all bidding strategies can be found in table 4.1.

4.2 Systemic Level

As mentioned in section 3.2, 5 different homogeneous societies were tested. Only homogeneous societies were tested, as they have naturally higher equality. Specifically, since all agents bid in exactly the same way over a long period of time, the stochastic differences in performance will slowly be eliminated, especially since there is no carry-over effect from previous pay-periods. Even so, equality is reached at different rates depending on which homogeneous society is considered. For each of the 5 homogeneous societies, the total score of each car was extracted. There were approximately 200 cars in each society.

To analyse the differences at a systemic level, 3 measures (described in section 3.2) are used. The following three subsections will briefly re-iterate the aim of each measure and present the results.

4.2.1 Gini Coefficient

To assess which society of bidders has the highest equality on a systemic level, the Gini coefficient is used (table 4.2). As explained in section 3.2, the Gini coefficient ranges from 0 to 1 and a lower value entails higher equality.

As expected, all societies have very high equality. However, some differences can still be observed.

Table 4.2: Gini coefficient of the different societies

Society	Gini Coefficient	Standard Deviation
Free-Riders	0.01357	6.29×10^{-4}
Random Bidders	0.01191	2.860670×10^{-4}
Static Bidders	0.02145	1.0715868×10^{-3}
Adaptive Bidders	0.00946	5.891306×10^{-4}
RL Bidders	0.01701	6.648237×10^{-4}

Specifically, a society that only consists of adaptive bidders has a significantly lower Gini coefficient. This means that after 250000 iterations, a homogeneous society of adaptive bidders is most likely to have the highest equality in terms of score.

To statistically evaluate the significance of the aforementioned differences, a pairwise t-test with Bonferroni correction was conducted. The Gini coefficient between each pair of homogeneous societies was found to be significantly different ($p < 0.001$). The mean Gini coefficient and standard deviation of each homogeneous society can be found in table 4.2.

4.2.2 Utilitarian Social Welfare

The utilitarian social welfare function (equation (3.2)) is used to evaluate the efficiency of each homogeneous society. The main purpose of this analysis is to assess whether the society with the highest equality is also the most efficient society (society with the lowest score). Table 4.3 shows the utilitarian social welfare scores (averaged over the number of agents in each society) and the respective standard deviations.

Similarly to the previous analyses, a pairwise t-test with Bonferroni correction was conducted and the utilitarian social welfare was found to be significantly different between each pair of societies ($p < 0.001$).

Table 4.3: Utilitarian social welfare (averaged over all agents) of different societies

Society	Util. Social Welfare	Standard Deviation
Free-Riders	37870	52.840
Random Bidders	36171	59.261
Static Bidders	38158	54.441
Adaptive Bidders	16853	18.816
RL Bidders	25868	69.132

As it can be seen from table 4.3, the homogeneous society that only consists of Adaptive Bidders has the highest efficiency. The society with RL Bidders performs significantly worse but still considerably better than the societies with Free-Riders, Random Bidders and Static Bidders.

The fact that the society with Adaptive Bidders outperforms the RL Bidders society is surprising as at the individual level, RL Bidders performed significantly better (as discussed in section 4.1). From the gathered results, it is not clear why the adaptive society overall performs better compared to the RL society. A possible explanation lies in the stochasticity of the two strategies. While the adaptive bidding strategy has a very deterministic way of generating bids, the RL bidding strategy constantly changes (in an attempt to improve). More importantly, the RL strategy is affected by the ϵ-greedy exploration/exploitation method which forces RL agents to place bids that have a lower predicted success (in this case, with a 10% chance). Section 6.2 discusses this problem (and a way to rectify it).

The other interesting result that can be observed in table 4.3 is that the 3 homogeneous societies of Free-Riders, Random Bidders and Static Bidders have very similar welfare scores. The similarity between Free-Riders and Static Bidders is not surprising as, in both cases, all agents always bid exactly the same amount (0, or a static bid). Random Bidders also, on average, all bid the same amount (50% of the current balance) but with significantly greater stochasticity. Lastly, the overall poor efficiency of these societies can potentially be explained by their lack of consideration of the individual hurry factors.

4.2.3 Least Well-Off Measure

The last measure at the systemic level is the 'least well-off' measure which essentially only considers the score of the worst-performing agent in each society. The idea behind this measure is that a society is successful if the 'least well-off' individual is not performing too badly. Therefore, to measure this, we can simply compare the highest (and therefore worst-performing) scores of each society (table 4.4).

Similarly to the previous analyses, a pairwise t-test with Bonferroni correction was conducted and the least well-off scores were found to be significantly different between each pair of societies ($p < 0.001$).

These results are on the same line with the utilitarian social welfare results (section 4.2.2). Specif-

Table 4.4: Worst individual scores of each society

Society	Least Well-Off Score	Standard Deviation
Free-Riders	39655	160.17
Random Bidders	38303	176.71
Static Bidders	40683	255.56
Adaptive Bidders	17712	101.65
RL Bidders	27578	145.00

ically, the Adaptive Bidders society has the best score, followed by the RL Bidding society. Then, the societies with Free-Riders, Random Bidders and Static Bidders perform considerably worse (but also very similarly compared to each other).

5 Conclusion

Linking back to the research question, we can see that at an individual level, the reinforcement learning agents performed considerably better than all other agents.

In a larger context, this conclusion is not surprising, as RL Bidders learn to constantly adapt to the model and are not bounded by many 'hard-coded' restrictions. For example, in contrast, the adaptive Bidding strategy can only adjust the static bid by $\pm 50\%$.

At a systemic level, all homogeneous societies had very high equality. However, the adaptive Bidding society stood out with a significantly higher equality in terms of the individual score. A homogeneous society of Adaptive Bidders was also overall significantly more efficient than a homogeneous society of RL Bidders. These findings are very interesting as, even though RL Bidders perform very well at the individual level, when put in a competitive environment (with all bidding strategies present) their performance is relatively poor at the systemic level in a homogeneous society (compared to Adaptive Bidders).

These results are also reflected when looking at the 'least well-off' score of each society, which showed that the worst performing agent of the homogeneous Adaptive Bidding society performed better than the worst agent of the RL Bidding society. The other societies (Static Bidding, Random Bidding and Free-Riding) performed similarly bad (worse than the RL society).

The reason why a homogeneous society of Adaptive Bidders has a much higher equality and efficiency compared to the RL society is unclear. A possible explanation has to do with the chance of these bidders to choose a sub-optimal action, due to their tendency to explore. This is further discussed in section 6.2.

6 Discussion

6.1 Emergent Behaviours

Due to how the different bidding strategies are implemented, some interesting emergent behaviours occur. In this section, some aspects of the observed behaviour of the different strategies will be discussed.

First, one of the most surprising strategies is the Free-Riding strategy, in which the car always bids 0 credits. This strategy is interesting, as it illustrates how the system helps agents that have run out of credits (so they do not have to wait indefinitely). Once a free-rider reaches the front of an intersection, they effectively just wait until all other lanes are empty (in which case the priority is given to them by default) or until a traffic-jam is formed behind them. In the latter case, the system gives them a large advantage in an attempt to resolve the congestion.

Random bidders also have a peculiar way of spreading out their funds. On average, random agents bid half of their current balance. In other words, on average, their balance is halved after winning an auction. Because of this, over the time period between pay-days, they spend a lot in the first few auctions but spend less and less as the pay-day comes closer.

At first, this behaviour might seem short-sighted. However, by being bold, the strategy ensures that most credits are used. For example, it is possible that between pay-days, the agent does not partake in many auctions, due to traffic congestion (less likely that they are the first in a queue). Therefore, by bidding boldly in the first few auctions, they use a large amount of their credit.

As mentioned in the system description, static and adaptive bidders use as a basis the static bid, which is an amount of credit that they can safely spend without risking running out of credits by the next pay-day. In hindsight, this is a fairly conservative strategy, as it is very unlikely that agents

participate in an auction in every iteration. Consequently, the busier the road network is, the worse these strategies perform. This is because they are less likely to participate in auctions (they instead wait in queues).

Last but not least, the fact that multiple RL agents are used in the same simulation results in an interesting problem, when tracking their learning. Since they are in the same road network, there is a high chance for them to compete with each other, instead of competing with agents that use others strategies. Consequently, as the RL agents improve, the auctions also get more difficult, as they bid against equally good RL-agents. So, even if the agents are improving, their results do not necessarily improve.

6.2 Limitations

In this research, a number of simplifications and assumptions were made that could have potentially influenced the results. This section discusses some of those, which should also be considered in similar future studies.

First, the traffic network that is used in this case is extremely simplistic. It is a square grid, only with intersections that consist of lanes of exactly the same size. Furthermore, concurrent use of the intersection is not possible, which significantly slows down traffic as shown by Schepperle and Böhm (2008). Additionally, all points in the grid are equally popular, which is also not realistic. In reality, any city or town has some places that are inherently busier than others (e.g. central square).

Another improvement that can be made is increasing the boldness of the adaptive and Static Bidders. To do that, the static bid amount could have been calculated at every iteration, splitting the current balance for all upcoming iterations until the next pay-day, instead of deciding on the amount at the start of each pay-period and not making adjustments throughout the period.

Moreover, as mentioned in section 2.3 sealed-bid auctions are not incentive compatible. Previous literature used an adaptation of the second-price sealed-bid auction (Carlino et al., 2013) which, in hindsight, should have also been used in this research, albeit its slightly increased complexity. This limitation only affected the RL bidding strategy, as it is the only strategy that bids completely dynamically and therefore is able to learn untruthful bidding. Unfortunately, it is not possible to systemically evaluate whether this problem truly occurred. Even so, untruthful bidding (learnt or otherwise), is arguably still a technique that agents can use to earn an individual and competitive advantage. In a broader and more realistic context, car manufacturers in the future could certainly use this (maybe unethical) advantage to increase the performance of their autonomous cars. Therefore, while this limitation is certainly one to look out for in future research, it can also be used to analyse the performance of unethical versus ethical bidding strategies in this context, while also potentially highlighting the importance of using an incentive compatible auction.

Furthermore, the RL bidding strategy could be optimised, for instance by increasing the state-space. For example, hurry factors and possible adjustments can be split using shorter intervals (e.g. 0.01 instead of 0.1). Given the resources, this was not possible for this project. Lastly, the pay-period was only set to be 10 iterations, which is fairly short. In a real context, that would mean that credits are replenished almost every second day, if we pass through 3-4 intersections per day. Such a short pay-period was chosen to increase the number of learning episodes that RL agents encounter. Additionally, the longer the learning episodes are, the harder it is to accurately adjust the utilities of the individual state-action pairs.

Last but not least, as discussed in section 4.2.2, the exploration/exploitation rate used by the RL-strategy (described in section 2.4.5), might have led to a lower systemic efficiency in a homogeneous society of RL bidders. In future research, to ensure that this does not occur, a *simulated annealing* approach can be used (Russell and Norvig, 2009). That is to say, the exploration rate ϵ can be slowly reduced over time (e.g. 0.05% reduction after each iteration). By doing this, the rate in which RL-agents learn would slowly decrease over time. If this is done, it is reasonable to also choose a significantly higher initial ϵ value (to boost learning at the start).

6.3 Further Research

Considering that using auctions for traffic moderation is a fairly new field of research, there are a few

ideas that are worth researching in the future.

First of all, as mentioned in section 6.2, it would be fruitful to consider traffic networks in which different destinations have varying popularity. In this way, the emergence of traffic jams can also be analysed and the bidding strategies might perform differently in such cases.

Moreover, while the hurry factor (urgency of the driver) remained constant for each trip, it is compelling (and more realistic) to consider a situation in which the urgency changes depending on the results of the auctions. That is to say, every time a bidder loses an auction, the individual hurry would increase and vice versa.

Furthermore, while in this scenario auctions were moderated independently in each intersection, it would also be interesting to consider a more systemic moderation in which a network administrator develops strategies to re-route drivers in the system, in an attempt to reduce the overall waiting time. Although, in this scenario, considering the individual urgency of drivers would be significantly harder.

Lastly, the scope of this research was contained to the use of a single auction type: first-price sealed-bid auctions. For future research, the bidding strategies can be assessed with alternative auction types, to evaluate the generalisability of the current findings. However, such a research would come with its own challenges, as it might not be possible to use the same bidding strategies in different auctions. For example, it is easy to use the same strategies in the Dutch auction (in which the auctioneer begins by announcing a high price and then proceeds to announce successively lower prices and the bidders can win by announcing that they are willing to pay the current price) (Shoham and Leyton-Brown, 2008). This is again a sealed-bid auction in practice and the same bids generated by the strategies can be used. In contrast, in the traditional English auctions, by which 'the auctioneer sets a starting price for the product and agents then have the option to announce successive bids, each of which must be higher than the previous bid' (Shoham and Leyton-Brown, 2008), it is impossible to use the exact same strategies, as they are not designed to respond to other bids.

References

Matthew D.. Adler. *Measuring social welfare: an introduction.* Oxford University Press., 2019.

Andrew G. Barto and Richard S. Sutton. *Reinforcement Learning: An Introduction (Adaptive computation and machine learning).* MIT Press., 1998.

D. Carlino, S. D. Boyles, and P. Stone. Auction-based autonomous intersection management. In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, pages 529–534, 2013.

Andrea Censi, Saverio Bolognani, Julian G. Zilly, Shima Sadat Mousavi, and Emilio Frazzoli. Today me, tomorrow thee: Efficient resource allocation in competitive settings using karma games. *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Oct 2019. doi: 10.1109/itsc.2019.8916911. URL `http://dx.doi.org/10.1109/ITSC.2019.8916911`.

Thomas J. Christian. Trade-offs between commuting time and health-related activities. *Journal of Urban Health*, 89(5):746–757, 2012. doi: 10.1007/s11524-012-9678-6.

Janos Galambos. Bonferroni inequalities. *The Annals of Probability*, 5(4): 577–581, 1977. ISSN 00911798. URL `http://www.jstor.org/stable/2243081`.

Corrado Gini. *Variabilita e mutabilita: contributo allo studio delle distribuzioni e delle relazioni statistiche.* Tipografia di Paolo Cuppin, 1912.

DianChao Lin and Saif Eddin Jabari. Pay for intersection priority: A free market mechanism for connected vehicles, 2020.

Roger B. Myerson. *Coalitions in Cooperative Games*, pages 417–482. Harvard University Press, 1991. ISBN 9780674341166. URL `http://www.jstor.org/stable/j.ctvjsf522.12`.

David Rey, Michael W Levin, and Vinayak V Dixit. Online incentive-compatible mechanisms for traffic intersection auctions, 2020.

Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall Press, USA, 3rd edition, 2009. ISBN 0136042597.

H. Schepperle and K. Böhm. Auction-based traffic management: Towards effective concurrent utilization of road intersections. In *2008 10th IEEE Conference on E-Commerce Technology and the Fifth IEEE Conference on Enterprise Computing, E-Commerce and E-Services*, pages 105–112, 2008.

Yoav Shoham and Kevin Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, USA, 2008. ISBN 0521899435.

M. Vasirani and S. Ossowski. A market-inspired approach for intersection management in urban road traffic networks. *Journal of Artificial Intelligence Research*, 43:621–659, Apr 2012. ISSN 1076-9757. doi: 10.1613/jair.3560. URL `http://dx.doi.org/10.1613/jair.3560`.

A Appendix

Table A.1: Set of hyper-parameters used during the experiments

Parameter	Value	Description
grid_size	5	Horizontal and vertical size of grid
init_cars	4	Maximum number of initial cars per lane
pay_period	10	Number of iterations between pay-days
payment	15	Credit renewal amount
num_sim	10	Number of simulations/experiments
queue_cap	4	Maximum capacity of each queue
RL_epsilon	0.1	RL ϵ-greedy epsilon value
RL_alpha	0.2	RL learning rate
prob_dest	0.2	Probability of a car to have reached its destination
handicap	0.15	Anti-traffic-congestion handicap
num_iter	250000	Number of iterations per experiment
bidder_dis	-	Distribution of bidding strategies that are present (varies between experiments)