Effects of Co-Evolution in a Complex Traffic Network

Paper number 163

ABSTRACT

One way to cope with the increasing demand in transportation networks is to integrating standard solutions with more intelligent measures. This paper discusses the integration and co-evolution of decision-making (by drivers) and control measures (by traffic lights) We use microscopic modeling and simulation, in opposition to the classical network analysis. General questions here are whether co-evolution (drivers and traffic lights) pay off, and, if so, what kind of evolutionary approach shall be used. This is challenging for networks other than the two-route one due to the complexity of route-choice behavior, as well as control strategies by the traffic lights. Moreover, the more agents, the less effective learning strategies are, when the integration among them depicts complex interelationships. The approach was tested in different scenarios and results show an improvement regarding travel time and occupancy when all actors co-evolve.

1. MOTIVATION AND INTRODUCTION

Urban mobility is one of the key topics affecting both the policy-makers and the citizens/tax-payers. Especially in medium to big cities, the urban space has to be adapted to cope with the increasing needs of the commuters. In transportation engineering the expression of the transport needs is called *demand*. This demand (in terms of people, volume, etc.) is commonly used to quantify transport *supply*. This is the expression of the capacity of transportation infrastructures and modes. Supply is expressed in terms of infrastructures (capacity), services (frequency), and networks.

The increasing demand we observe nowadays has to be accommodated either with increasing supply (e.g. road capacity), or with a better use of the existing infrastructure. Since an expansion of the capacity is not always socially attainable or feasible, transportation and traffic engineering now seek to optimize the management of both the supply and the demand using concepts and techniques from intelligent transportation systems (ITS). These refer to the application of modern technologies to the operation and control of transportation systems [15].

From the side of supply, several measures have been adopted in the last years, such as congestion charging in urban areas (London), restriction of traffic in the historical center (Rome, Paris, Amsterdam), alternace of vehicles allowed to circulate in a given day (São Paulo, Mexico City).

From the point of view of the demand, several attempts exist not only to divert trips both spatially as well as temporally, but also to distributed the demand within the available infrastructure. Besides, it is now recognized that the human actor has to be brought into the loop. With the amount of information that we have nowadays, it is almost impossible to disregard the influence of real-time information systems over the decision-making process of the individuals.

Hence, our long term goal is to tackle a complex problem like traffic from the point of view of information science. This project is the result of an accumulated experience with microscopic models modeling tools for traffic and transportation management. These range from traffic signal optimization (refs. omitted due to blind review) and binary route choice and effect of information on commuters (idem), to microscopic modeling of physical movement (idem).

An important milestone in our project is to propose a methodology to integrate complex behavioral models of human travelers reacting to traffic patterns and control measures of these traffic patterns, focusing on distributed and decentralized methods. Classically, this is done via network analysis. To this aim, it is assumed that individual road users seek to optimize their individual costs regarding the trips they make by selecting the “best” route. This is the basis of the well known traffic network analysis based on Wardrop’s equilibrium principle [20]. This method predicts a long term *average* state of the network. By assuming steady state network supply and demand conditions from day-to-day, this equilibrium based method cannot, in most cases, cope with the dynamics of the modern transportation systems. Moreover, it is definitely not adequate for answering questions related to what happens in the network *within* a given day, as the variability in the demand and the capacity of the network tend to be high. Just think about changing weather conditions from day-to-day and within a single day! In summary, as equilibria based concepts overlook this variability, it seems obvious that it is not adequate to be used in microscopic modeling and simulation.

The reason why microscopic approaches are getting more and more popular within the community of transportation engineering is twofold. First, it is well recognized that individual decision making does affect the equilibrium and this cannot be disconsidered. The second factor is the improvements in hardware and in software paradigms (e.g. agent based simulation), which now allow the consideration of individual characteristics.

Based on this important assumption, the field of transportation engineering has seen a boom regarding methodologies for microscopic modeling as well as a trend towards development of real-time systems. As part of this effort, a new research area is studying how to integrate all pieces of work which have been produced in different fields such as traffic and transportation engineering itself, physics, psychology, computer science, geography, etc. Microscopic modeling can be accomplished in several ways. Agent-based modeling is one alternative.

Therefore, the general aim of this paper is to investigate what happens when different actors interact, each having its own goal. The objective of *local* traffic control is obviously to minimize queues in a spatially limited area (e.g. around a traffic light). The objective of road users is (normally) to minimize travel time. Finally, from the point of view of the whole system, the goal is to assure reasonable travel times for *all* user, which can be highly conflicting with some individual utilities as in a social dilemma like nature. This is a well-known issue. Tumer and Wolpert [18] for instance shown that there is no general approach to deal with this complex question of collectives.

Specifically, this paper investigates which strategy is better for drivers (e.g. adaptation or greedy actions). Also, what is better for traffic lights (acting greedily? just carry on a “well-designed” signal plan? Q-learning?)? After which volume of traffic does decentralized control of traffic lights starts to pay off? Does single-agent or isolated reinforcement learning make sense in traffic scenarios? What happens when drivers adapt concurrently? These are hot topics not only in traffic research, but also from a more general agent point of view as it refers to co-evolution.

The challenge of the present paper is to tackle more realistic scenarios, i.e. depart from binary route choice. To the best of our knowledge, the question on what happens when drivers and traffic lights adapt in a complex route scenario (e.g. a grid) has not been tackled so far.

In the next section we review these and other related issues. In Section 2.3.3 we describe the approach and the scenario. Section 4 discusses the results, while Section 5 presents the concluding remarks.

2. BACKGROUND: SUPPLY AND DEMAND IN TRANSPORTATION ENGINEERING

Learning and adaptation is an important issue in multiagent systems. It is not the purpose of this paper to review the existing literature on this issue; many subfields of multiagent systems report the advantages of learning or adapting agents, from which Robocup is just one example. Rather, we will concentrate on pieces of related work which either deal with traffic scenarios directly or report close scenarios.

2.1 Management of Traffic Demand

Given its complexity, the area of traffic simulation and control has been tackled by many branches of applied and pure sciences. Therefore, several tools exist which target the problem isolatedly. Simulation tools in particular are quite old (1970s) and stable.

On the side of demand forecasting, the arguably most used computational method is the so-called 4-step-process [14]. It consists of the four steps: trip generation, destination choice, mode choice, and route assignment. Route assignment includes route choice and a very basic traffic flow simulation, often, but not always, leading to a Nash Equilibrium.

Over the years, the 4-step-process has been improved in many ways, most notably by (i) combining the first three steps into a single, traveler-oriented framework (*activity-based demand generation (ABDG)*) and by (ii) replacing traditional route assignment by so-called *dynamic traffic assignment (DTA)*, where the traffic flow simulation is much more realistic. Still, in the typical implementations, all traveler information gets lost in the connection between ABDG and DTA, making realistic agent-based modeling at the DTA-level difficult. An important distinction exists between day-to-day replanning and within-day (on-the-fly) replanning. Only the latter allows simulated travelers to react to ITS measures, although some level of ITS functionality can be successfully emulated with day-to-day replanning only. For a discussion of these issues see, e.g., [2].

Another related problem is estimation of the state of the whole traffic network from partial sensor data. Although many schemes exist for incident detection, there are few deployments of large scale traffic state estimation. One exception is `www.autobahn.nrw.de`. It uses a traffic microsimulation to extrapolate between sensor locations, and it uses intelligent methods combining the current state with historical data in order to make short-term predictions. However, the particles (vehicles) are very simple: They do not know their destinations, let alone the remainder of their daily plan. This was a necessary simplification to make the approach work, but it is necessary to overcome this simplification since the effects of the travelers’ decisions are difficult if not impossible to estimate without these aspects.

What is missing it a true agent-based integration of these and other approaches. However, until now agent-based simulations with high-level agents on the scale required for traffic simulation of real-world networks have not been developed. The main reason why this has not happened is that the software tools for flexible and robust multi-agent simulations are currently just emerging.

Some steps towards that goal is to use concepts of microeconomics to approach decision-making and how drivers adapt to the previous experiences. Basically, simple binary scenarios have been used, based on the *El Farol* Bar Problem (EFBP) [1]. Adaptation can be achieved by computing a probability the driver puts in each alternative based on a past reward as in [8]. However, when the coordination emerges out of individual self-interest, sometimes a *user equilibrium* is achieved, but in general no *system optimum*. In consequence, in scenarios where agents act greedily, the performance of the overall system may be compromised. Information systems (advanced traveler information systems (ATIS), route guidance, etc.) may help in situations where travelers lack information, but are otherwise not expected to move the system away from the user equilibrium. In contrast, control systems such as dynamic tolls or dynamic traffic lights can move the system towards system optimum in spite of self-interested drivers.

2.2 Real-Time Optimization of Traffic Lights

Signalized intersections are controlled by signal-timing plans (we use signal plan for short) which are implemented at traffic lights. A signal plan is a unique set of timing parameters comprising the cycle length L (the length of time for the complete sequence of the phase changes), and the split (the division of the cycle length among the various movements or phases). The criterion for obtaining the optimum signal timing *at a single intersection* is that it should lead to the minimum overall delay at the intersection. Several plans are normally required for an intersection to deal with changes in traffic volume, or, in an traffic-responsive system, that at least one plan exist and can be changed on the fly.

For coordination of traffic lights, which is not the focus of this paper, other methods exist such as Transyt [17]. It runs off-line and aims at optimizing the bandwidth of an arterial via the design of phases and offsets from one intersection to the adjacent one. Transyt only computes a synchronized, optimal timing for a sequence of traffic lights in an arterial or corridor, in one traffic direction and in an *offline* fashion. SCOOT, SCATS, and TUC [7, 11, 5] respectively, work online but, when dealing with a whole network of traffic lights, they all seem to depend on a human expert to solve the conflict which arises regarding which direction of coordination to implement.

In [3], a MAS based approach is described in which each traffic light is modeled as an agent, each having a set of pre-defined signal plans to coordinate with neighbors. Different signal plans can be selected in order to coordinate in a given traffic direction. This approach uses techniques of evolutionary game theory. only information about their local traffic states. However, payoff matrices (or at least the utilities and preferences of the agents) are required, i.e. these figures have to be explicitly formalized.

In [13] groups were considered and a technique from distributed constraint optimization was used, namely cooperative mediation. However, this mediation was not decentralized: group mediators communicate their decisions to the mediated agents in their groups and these agents just carry out the tasks.

Camponogara and Kraus [4] have studied a simple scenario with only two intersections, using stochastic game-theory and reinforcement learning. Their results with this approach were better than a *best-effort* (greedy), a random policy, and also better than Q-learning.

Also, in [12] a set of techniques were tried in order to improve the learning ability of the agents in a simple scenario.

Finally, a reservation-based system [6] is also reported but it is only slightly related here because it does not include traffic lights.

2.3 The Need for Integration

It is obvious that the main actors, namely the driver and the engineering facilities interact. Broadly speaking these actors are responsible for the demand and the supply/capacity respectively. Of course things are not so simple as there is a well known loop between supply and demand: the improvement of the transportation systems or facilities makes nearby land more accessible and attractive, thus requiring furher increases in land-use development, which in turn result in even higher transportation demands.

The nature of the interaction between these actors, as well as its consequences, have only recently started to make its way towards main stream research. There are very few works dealing with these issues. Next, we focus on three of them which represent different views of the problem and how they differ from the present paper.

2.3.1 *Learning based approach*

In [21], the main focus is not exactly that interaction. Rather, the aim is to study reinforcement learning; traffic control is seen just as an application scenario. The paper describes the use of reinforcement learning by the traffic light controllers (agents) in order to minimize the overall waiting time of vehicles in a small grid. Agents learn a value function which estimates the expected waiting times of single vehicles given different settings of traffic lights. One interesting issue tackled in this research is that a kind of co-learning is considered: the value functions are learned not only by the traffic lights, but also by the vehicles which can thus compute policies to select optimal routes to the respective destinations.

The ideas and some of the results presented in that paper are important. However, strong assumptions turn difficult its use in the real world. First, the kind of communication and knowledge (or, more appropriate, communication *for* knowledge formation) has a high cost. Traffic light controllers are suppose to know vehicles destination in order to compute expected waiting times for each. Given the current technology, this is (still) a strong assumption.

Second, it seems that traffic lights can shift from red to green and opposite at each time step of the simulation.

Third, there is no account of experience made by the drivers based on their local experiences only. What about if they just react to the (few) past experiences after the route is completed and the driver takes it again the next day (typical commuting scenario)?

Finally, drivers being autonomous, it is not reasonable to expect that all will use the best policy computed, given the value function, which for this sake, was computed by the traffic light and not by the driver itself. Thus, in the present paper, we depart from the assumptions regarding communication and knowledge the actors have to have about each other.

2.3.2 *Game theoretic approach*

In [19], a two-level, three-player game is discussed which integrates traffic control and traffic assignment, i.e. both the control of traffic lights and the route choices by drivers are considered. The control part involves two players, namely two road authorities, while the population of drivers is seen as the third player. These are modeled as a game and the main aim of this work is to analyze the outcome when players are able to observe the previous move. Complete information is assumed, which means that all players (including the population of drivers) have to be aware of the movements of others. Although the paper reports interesting conclusions regarding e.g. the utility of cooperation among the players, this is probably valid only in that simple scenario. Besides, the assumptions that drivers always follow their shortest routes are difficult to justify in a real-world application.

In the present paper, we want to depart from both the two-route scenario and the assumption that traffic management centers are in charge of the control of traffic lights. Rather, we follow a trend of decentralization, in which each

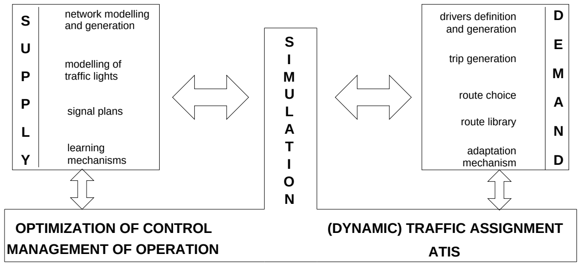

Figure 1: Schema of the Co-Evolution in an ITS Framework

traffic light is able to sense its environment and react accordingly and autonomously, without having its actions computed by a central manager as it is the case in [19].

Moreover, the two-route scenario is a very didactic one and serves the purpose of the main aim of [19]. However, it is questionable whether the same mechanism can be used in more complex scenarios, as claimed. The reason for this is the fact that when the network is composed of tens of links, the number of routes increases and so the complexity of the route choice, given that now it is not trivial to compute the network and user equilibria.

2.3.3 Methodologies

Liu and colleagues [10] describe a modeling approach which integrates microsimulation of individual trip-makers' decisions and individual vehicle movements across the network. At this stage, we do not focus on the movement of vehicles at the level of driving issues (car following, lane changing etc.) as the authors do, although this can be done using a microscopic simulator (ref. omitted due to blind review). Rather, we focus on how to manage and improve the capacity and operation of the network which will in turn affect the movement of vehicles. Moreover, in [10] the focus is on the description of the methodology which incorporate both demand and supply dynamics, so that the applications are only briefly described and not many options for the operation and control of traffic lights are reported. One scenario described deals with a simple network with four possible routes and two control policies. One can roughly be described as greedy, while the other is fixed signal plan based. In summary, in the present paper we do not explore the methodological issues as in [10] but, rather, investigates, in more details, particular issues of the integration and interaction between actors from the supply and demand sides.

3. CO-EVOLUTION IN AN ITS FRAMEWORK

Figure 1 shows an scheme of the our approach, based on the interaction among supply, demand, and an ITS module. The latter is related to strategic decisions and is composed of a simulation sub-module, as well as sub-modules to implement optimization of control (e.g. traffic lights), management of operation, traffic assignment (static or dynamic), and an ATIS.

Regarding the side of the supply, there are sub-modules for modeling and generation of the network, modeling of traffic lights and signal plans, learning mechanisms, etc. Regarding the demand side, the sub-modules generate the population of agents (drivers), the library of routes, the trips, the adaptation mechanisms, the route choice algorithm, etc. One sees that the interaction which is posed by the feedback loop mentioned above is tackled by the simulation submodule.

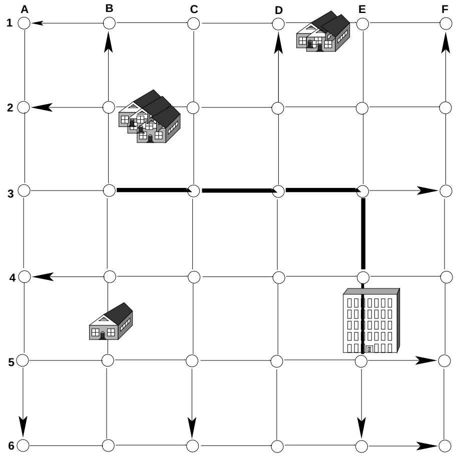

Figure 2: Grid 6x6 showing the main destination (E4E5), the three main origins (B5B4, E1D1, C2B2), and the "main street".

In the present paper we tackle some issues of the feedback loop between supply and demand by means of studying scenarios in which drivers adapt (as in [8]) and traffic lights use learning mechanisms (e.g. in [3, 13, 12]). This is just a part of the loop as we tackle here neither the ATIS issue nor the management of operation of the network.

Currently the approach generates the network (grid or any other topology), supports the creation of traffic light control algorithms as well as signal plans, the creation of routes (route library) and the algorithms for route choice. The movement of vehicles is queue-based. Moreover, it provides the simulation environment through the agent-based simulation environment SeSAm [9].

The scenario we use to exemplify the approach is a typical commuting scenario where drivers repeteadily select a route to go from an origin to a destination. It is not so simple as a two-route (binary decision) scenario. Rather, it is reasonably more complex and captures desirable properties real scenarios have regarding the aim of this study, namely the co-evolution among drivers and traffic lights. Next we detail the particular scenario used. Values such as size of grid, number of drivers, all probabilities, etc. can be changed.

To model the supply, we use a grid of size 6x6 where the 36 nodes are tagged from A1 to F6, as in Figure 2. All links are one-way and drivers can turn in each crossing. This kind of scenario is a realistic one and, in fact, from the point of view of route choice and equilibrium computation, it is also a very complex one as the number of possible routes from point O to point D is high.

Contrarily to simple two-route scenarios, in the grid one, it is possible to set arbitrary origins and destinations. Each driver has one particular origin and destination. To render the scenario more realistic, there is one main destination: on average, 60% of the road users have the link labelled as E4E5, associated with node E4, as destination. Other links have 1.7% probability of being a destination. Origins are nearly equally distributed in the grid, with three exceptions: links B5B4, E1D1, and C2B2 have, approximately, probabilities 3, 4, and 5% of being an origin respectively. The remaining links have each a probability of 1.5%. This was done to model residential neighborhoods and can be implemented via definition of weights for each link.

Regarding capacity, all links can hold up to 15 vehicles, except those located in the so called "main street" which can hold up to 45 (this is set via a parameter called increasing factor over the basic storage capacity). This main street is formed by the links B3 to E3, E4, and E5.

The control is done via decentralized traffic lights. These are located in each node. Each has a signal plan which, by default, divides the cycle time (in the experiments 40 time steps) 50-50% between the 2 phases. The actions of the traffic lights are to keep the plan as this default or to priorize one phase. The strategies are: i) always keep the default signal plan; ii) greedy (run green time for the phase (in this scenario, one link) with the higher occupancy); iii) use single agent Q-learning.

Regarding the demand, the main actor is the driver or road user. The simulation can generate any number of them, as well as any number of routes in the set of known routes. Normally the simulations were done with drivers knowing one to five routes. These were generated via an algorithm that computes the shortest path (one route) and the shortest path via arbitrary detours (the other four).

Drivers can use three strategies: i) select a route randomly (each time it departs); ii) select a route greedily (always pick the one with best average travel time so far); iii) select a route in an adaptive way meaning that the average travel times so far are used to compute a probability to select the route to use. The route choice is done before the trip. Notice that conventional single-agent Q-learning cannot be efficiently employed here as the space action policies is too big to be searched. In fact, as already reported by one of us in (ref. omitted) and by Tumer and Wolpert [18], in a system with a large number of agents where basically each competes with all others for the same resource, it is difficult for each to discern the effects of its own actions.

4. RESULTS AND DISCUSSION

4.1 Metrics and Parameters

In order to evaluate the experiments, four quantities were measured: the number of drivers who have arrived at their destinations up to the time out t_{out} for each particular trip; the mean travel time over all drivers for a given trip, as well as the mean of the average, over all possible routes, of the travel time over all drivers. Plots for these two are not shown here due to lack of space but we briefly discuss the patterns. Rather, we plot the average occupancy over all links in the network, as well as over the links of the node E4 (node closest to the main destination), and show, in tabular form, the mean travel time over only the last T trips to give a reference of the travel time at the end of the experiments.

All experiments were repeated 20 times. Plots in Section 4.4 show 15 trips.

The other parameters used were: t_{out} equal to 300 when the number of drivers is 400 or 500, 400 when it is 600, and 500 when there are 700 drivers; T is 5; percentage of drivers who adapt is either 100 or zero (in this case all act greedily) but any combination can be used; percentage of traffic lights which act greedily is either zero or 100; a link is considered jammed if its occupancy is over 50%; for the Q-learning there is an experimentation phase of $10 \times t_{out}$, the learning rate is $\alpha = 0.1$ and the discount rate is $\lambda = 0.9$.

Type of Simulation	Average Travel Time Last 5 Trips
homog. distr. O&D, greedy drvs	79
homog. distr. O&D greedy drvs, greedy TLs	82
random drvs / fix TLs	160
random drvrs / greedy TLs	150
greedy drvs / fix TLs	100
adapting drvs / fix TLs	149
greedy drvs / greedy TLs	106
adapting drvs / greedy TLs	143
greedy drvs / Qlearning TLs	233

Table 1: Average Travel Time Last 5 Trips for 400 Drivers, under Different Conditions

In Table 1 we summarize the average travel time over the last $T = 5$ trips (henceforward *attl*5*t*) under different conditions and for different number of drivers. These conditions are explained next.

4.2 Homogeneous Origins and Destinations

For the sake of calibration, we have run experiments in which the grid is the same as in all other experiments but the origins and destinations are evenly distributed, i.e. all links have the same probability of being an origin and/or a destination (when drivers are created, their origins and destinations can be any link with equal probability). This is of course an unrealistic situation because in reality there are locations of the urban area which are attractors (hence potential trip destinations) such as office centers, shopping centers, etc., while other locations originate trips with higher probability (residential areas etc.). This scenario was tested under two situations: i) traffic lights run the default signal plan and there are 400 greedy drivers; ii) with both drivers and traffic lights acting greedily. In the former case, the *attl*5*t* is only 79 time units, while this increases to 82 when the traffic lights also act greedily. The reason for this increase is explained in the next subsections.

4.3 Different Strategies by Drivers and Traffic-Lights

For all scenarios described in this subsection, 400 drivers were used.

Random drivers. Again for sake of comparison, here drivers select a route randomly from the set of known routes. The results in Table 1 show that, for this kind of drivers, it makes little difference whether or not the traffic lights act greedily. The *attl*5*t* is one of the highest from all experiments for 400 drivers and was not repeated for other number of drivers for obvious reasons.

Greedy or adaptive drivers; fix traffic lights. In the next scenarios, drivers either select greedily the route with the smallest average travel time experienced, or select a route in a probabilistic way, with this probability being computed based on the average travel time experienced. We call this adaptation (in opposition to greedy selection). In the case of adaptation, the *attl5t* is 149 units, while this is 100 in the former case. The higher travel time is the price payed by the experimentation which the drivers continue to do, even though the optimal policy was achieved long before (remember that the *attl5t* is computed only over the last 5 trips). The greedy action is of course much better after the optimal policy was learned. In the beginning, when experimentation does pay off, the adaptive driver performs better. Due to the characteristics of the scenario, the experimentation does not pay off for a long time since once the shortest path is learned it is of course better to remain on it. In summary, greedy actions by the drivers work because they tend to select the routes with the shortest path and this normally distributes drivers more evenly than longer routes (the longer the routes drivers have the more overlap they have leading to longer travel times).

Greedy or adaptive drivers; greedy traffic lights. When traffic lights also act greedily we can see that this is not automatically good: the *attl5t* is 106 (it is 100 if traffic lights do not act greedily). This happens because the degree of freedom of traffic lights' actions is low, as actions are highly constrained. For instance, acting greedily can be highly sub-optimal when for instance traffic light A serves direction D_1 (thus keeping D_2 with red light), and the downstream flow of D_1 is already jammed. In this case, the light might indeed be green for vehicles on D_1 but vehicles cannot move due to the downstream jam. Worse, jam may appear on D_2 too due to the small share of green time. This explains why acting greedily at traffic lights is not necessarily a good policy.

Q-learning traffic lights. We have expected Q-learning to perform at least as bad because Q-learning does not have a good performance in noisy traffic scenarios [16]. In order to test this, we have implemented a Q-learning mechanism in the traffic lights. Available actions are: to open the phase (i.e. not a time step!) serving either one direction (e.g. D_1), or the other (D_2). The states are the combination of states in both approaching links, i.e. $\{D_1\_jammed, D_1\_not\_jammed\} \times \{D_2\_jammed, D_2\_not\_jammed\}$. The rewards is the average occupancy in the incoming links of a given node. Please notice that the Q-learning was modified to deal with minimization of the Q-values. Increasing the level of discretization may improve the results but not to the point of being as good as the greedy strategy. In the best case, *individual* Q-learning would converge to the greedy policy.

The low performance of Q-learning in traffic scenarios is due basically to the fact that the environment is non-stationary, not to the poor discretization of states. Convergence is never achieved and so traffic lights keep experimenting. If we relax the value for convergence (e.g. increasing the ϵ value used to compare the current and the last Q-values), eventually traffic lights considered that they have learned an optimal policy. However, employing this policy in an environment which does not correspond to the time horizon used to "learn" degrades the performance badly. In fact it is even worse than random policy. For the same reason the combination of adaptive drivers and greedily traffic lights does not have a good performance.

4.4 Scenarios With More Drivers

Up to here we have kept the number of drivers on 400 in order to be able to compare different settings and/or learning strategies. Next we discuss some cases with more drivers. We only investigate the cases of greedy drivers / non-learning traffic lights versus the case in which both drivers and traffic lights act greedily. We do so because we want to test whether or not increasing volume of traffic (due to increasing number of drivers in the network) would cause greedy traffic lights to perform better. This is expected to be the case since once the number of drivers increase, greedy actions alone do not bring much gain; some kind of control in the traffic lights is expect to be helpful.

This is indeed the case when the network gets more crowded. 400, 500, 600 and 700 drivers mean an average occupancy of ≈ 40%, 47%, 59%, and 72% per link. In Table 2 the *attl5t* for these numbers of drivers are shown, both without greedy traffic lights and when traffic lights act greedily.

The case for 400 drivers was discussed above. With more than 500 drivers, the *attl5t* is lower when traffic lights also act greedily. In the case of 700 drivers, thus an initial occupancy over 70%, which is considered high, the improvement in travel time (411 vs. 380) is about 8%. The explanation

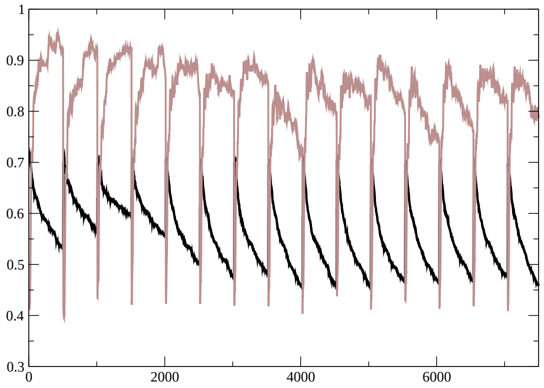

Figure 3: 700 greedy drivers: average occupancy over all links in the network (dark); over the links of the node E4 (gray)

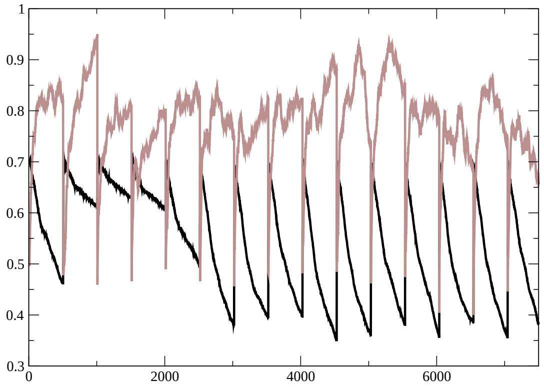

Figure 4: 700 greedy drivers and greedy traffic lights: average occupancy over all links in the network (dark); over the links of the node E4 (gray)

Average Travel Time Last 5 Trips				
Type of Simulation	Nb. of Drivers			
	400	500	600	700
greedy drvs / fix TLs	100	136	227	411
greedy drvs / greedy TLs	106	139	215	380

Table 2: Average Travel Time Last 5 Trips for Different Number of Drivers, under Different Conditions

for the performance of greedy traffic lights is that they tend to keep the occupancy of links lower.

Figures 3 and 4 show two curves regarding the occupancy of links for 700 drivers, without and with greedy traffic lights respectively. Occupancy is a measure related to the supply side, while the travel time which was presented in the previous sections is basically a measure related to the demand side (drivers).

Each graph has two curves. The darker depicts the average occupancy over *all* links of the network. We see an improvement regarding the case with greedy traffic lights: in Figure 3, where traffic lights are fixed, the occupancy is barely below 0.5, while it is reduced to slightly less than 0.4 when the traffic lights are greedy (Figure 4). Moreover, a similar improvement can be seen in *selected* links of the network, namely those which tend to be jammed for a longer time as they are near the main destination or lie in critical crossings. This is the case of the node $E4$ (see Figure 2) which lies exactly before the main destination so that the majority of the drivers has to drive through it. The occupancy for these links is depicted in Figures 3 and 4 in the lighter curves.

As a remark about Q-learning, it is not good here as well since the environment tend to be even more noisy with the increase of drives.

4.5 Drivers Have No Route Choice

Finally, in order to test whether greedy action of a traffic light has some effect on the travel time, we have run scenarios in which the drivers only know one route, the shortest one. Thus, their choices are irrelevant and what counts are the choices of the traffic lights. Again, this is an unrealistic assumption as drivers are likely to try more than one route. However it aims at testing traffic lights in isolation.

Table 3 shows the *attl5t* for three types of scenarios: homogeneous distribution of origin and destination, scenarios with a main destination and three main origins (as above), with 400 and 700 drivers, combined with the cases in which the traffic lights act greedily and remain fix.

We can see that, for instance, comparing the *attl5t* in the first and the second lines in that table, one sees that it decreases (126 to 100) when drivers know only one route (middle column). That means the greedy action of the traffic light does have an effect. When the drivers have more routes to select (last column is a repetition of the results in Table 1) and use their adaptation or greedy strategies, this trend is different depending on the scenario (homogeneous vs. heterogeneous) and the number of agents, as explained in the last two subsections.

4.6 Overall Discussion

In the experiments presented we could see that different strategies by the drivers, as well by the traffic lights have distinct results, in different settings. We give here the main conclusions. For the network depicted, increasing the links capacity from 15 to 20 would lead to much less jam (this was tested but is not shown here due lack of space). However, increasing network capacity is not always possible so that other measures must be taken. Diverting people and/or given them information both have limited performances. Thus the idea is to better use the control infrastructure. Therefore we have explored the capability of the traffic lights to cope with the increasing demand.

Average Travel Time Last 5 Trips		
Type of Simulation	Nb. of Known Routes	
	1	1–5
homog. / fix TLs	126	79
homog. / greedy TLs	100	82
400 drvs. / fix TLs	109	100
400 drvs. / greedy TLs	101	106
700 drvs. / fix TLs	400	411
700 drvs. / greedy TLs	345	380

Table 3: Average Travel Time Last 5 Trips When Drivers know One or More Routes

Regarding travel time, it was shown that the strategies implemented in the traffic lights pay off in several cases, especially when the demand increases. We have also measured the number of drivers who arrive before time t_{out}. Just to give a flavor of the figures, bad performance (around 75% arrived) was seen only when the drivers adapt probabilistically and when they select routes randomly. This of course is a consequence of the high travel times (see Table 1).

The general trend is that when the traffic lights also play a role, the performance increases, by all metrics used.

About the use of Q-learning, as said, single-agent learning is far from optimum here due to the non-stationarity nature of the scenario. This is true especially for those links located close to the main destination and the main street as they tend to be part of each driver's trip so that the pattern of volume of vehicles changes dramatically. A possible solution is to use collaborative traffic lights. In this case, traffic light A would at least ask/sense traffic light B downstream whether or not it shall act greedily. This however leads to a cascade of dependence among the traffic lights. In the worst case everybody has to consider everybody's state. Even if this is done in a centralized way (which is far from desirable), the number of state-action pairs prevents the use of MAS Q-learning in this format.

5. CONCLUSION

In Section 2 we have shown that, regarding the supply side, many tools exist for the optimization or management of traffic flow. One of the main tools is the control of traffic lights, for which many approaches exist. Regarding the demand, this is also true to some extent but the field has received more attention after the appearance of the so-called Advanced Traveler's Information Systems (ATIS). Several studies and approaches exist for modeling travelers' decision-making. In commuting scenarios in particular, the issue of how they adapt in order to maximize their utilities is one of those approaches.

However, there is no attempt to study what happens when *both* the driver and the traffic light use some evolutionary mechanism in the same scenario or environment, especially if *no central control exist*, i.e. the co-evolution happens in a decentralized fashion, in which case some form of auto-organization may arise. This is an important issue because, although ITS have reached a high technical standard, the reaction of drivers to these systems is fairly unknown. In general, the optimization measures carried out in the network both affect and are affected by drivers' reactions to them. This leads to a feedback loop which has received little attention to date.

In the present paper we have investigated this loop by means of a prototype tool constructed in an agent-based simulation environment. This has modules to cope with the demand and the supply side, as well to implement the ITS modules and algorithms for the learning, adaptation etc. The results show an improvement regarding travel time and occupancy (thus, both the demand and supply side) when all actors co-evolve. This was compared with situations in which either only drivers or only traffic lights evolve, in different scenarios.

This work can be extended in two main directions. First, we plan to integrate the tools developed by the authors independly for supply and demand (refs. omitted) which are simulators with far more user-friendly capabilities and permit the modeling of even more realistic scenarios as trips can be richer etc. The results are not expect to differ in the general trends, though. The second extension relates to the use of heuristics for a MAS reinforcement learning in order to improve its performance. This is not a trivial extension as it is known that reinforcement learning for non-stationary environments is a hard problem, especially when several agents are involved.

6. REFERENCES

[1] Arthur, B. Inductive reasoning, bounded rationality and the bar problem. Tech. Rep. 94–03–014, Santa Fe Institute, 1994.

[2] Balmer, M., Nagel, K., and Raney, B. Large scale multi-agent simulations for transportation applications. *Journal of Intelligent Transport Systems 8* (2004), 205–223.

[3] Bazzan, A. L. C. A distributed approach for coordination of traffic signal agents. *Autonomous Agents and Multiagent Systems 10*, 1 (March 2005), 131–164.

[4] Camponogara, E., and Jr., W. K. Distributed learning agents in urban traffic control. In *EPIA* (2003), F. Moura-Pires and S. Abreu, Eds., pp. 324–335.

[5] Diakaki, C., Dinopoulou, V., Aboudolas, K., Papageorgiou, M., Ben-Shabat, E., Seider, E., and Leibov, A. Extensions and new applications of the traffic signal control strategy tuc. In *Proc. of the 82nd Annual Meeting of the Transportation Research Board* (January 2003), pp. 12–16.

[6] Dresner, K., and Stone, P. Multiagent traffic management: A reservation-based intersection control mechanism. In *The Third International Joint Conference on Autonomous Agents and Multiagent Systems* (July 2004), pp. 530–537.

[7] Greenough, J. C., and Kelman, W. L. Metro toronto scoot: traffic adaptive control operation. *ITE Journal 68*, 5 (May 1998).

[8] Klügl, F., and Bazzan, A. L. C. Route decision behaviour in a commuting scenario. *Journal of Artificial Societies and Social Simulation 7*, 1 (2004).

[9] Klügl, F., Herrler, R., and Oechslein, C. From simulated to real environments: How to use SeSAm for software development. In *Proceedings of the 1st German Conferences MATES – Multiagent System Technologies* (2003), S. Berlin, Ed., no. 2831 in Lecture Notes in Artificial Intelligence, pp. 13–24.

[10] Liu, R., Van Vliet, D., and Watling, D. Microsimulation models incorporating both demand and supply dynamics. *Transportation Research Part A: Policy and Practice 40*, 2 (February 2006), 125–150.

[11] Martin, P. T. SCATS, an overview. http://signalsystems.tamu.edu, January 2001. Workshop on Adaptive Signal Control Systems.

[12] Nunes, L., and Oliveira, E. C. Learning from multiple sources. In *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi Agent Systems, AAMAS* (New York, USA, July 2004), vol. 3, New York, IEEE Computer Society, pp. 1106–1113.

[13] Oliveira, D., Bazzan, A. L. C., and Lesser, V. Using cooperative mediation to coordinate traffic lights: a case study. In *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS)* (July 2005), New York, IEEE Computer Society, pp. 463–470.

[14] Ortúzar, J., and Willumsen, L. G. *Modelling Transport*, 3rd ed. John Wiley & Sons, 2001.

[15] Roess, R. P., Prassas, E. S., and McShane, W. R. *Traffic Engineering*. Prentice Hall, 2004.

[16] Silva, B. C. d., Basso, E. W., Bazzan, A. L. C., and Engel, P. M. Dealing with non-stationary environments using context detection. In *Proceedings of the 23rd International Conference on Machine Learning (ICML 2006)* (June 2006), W. W. Cohen and A. Moore, Eds., ACM Press, pp. 217–224.

[17] TRANSYT-7F. *TRANSYT-7F User's Manual*. Transportation Research Center, University of Florida, 1988.

[18] Tumer, K., and Wolpert, D. A survey of collectives. In *Collectives and the Design of Complex Systems*, K. Tumer and D. Wolpert, Eds. Springer, 2004, pp. 1–42.

[19] van Zuylen, H. J., and Taale, H. Urban networks with ring roads: a two-level, three player game. In *Proc. of the 83rd Annual Meeting of the Transportation Research Board* (January 2004), TRB.

[20] Wardrop, J. G. Some theoretical aspects of road traffic research. In *Proceedings of the Institute of Civil Engineers* (1952), vol. 2, pp. 325–378.

[21] Wiering, M. Multi-agent reinforcement learning for traffic light control. In *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000)* (2000), pp. 1151–1158.