Lecture Notes in Artificial Intelligence 4865

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Karl Tuyls Ann Nowe Zahia Guessoum
Daniel Kudenko (Eds.)

Adaptive Agents and Multi-Agent Systems III

Adaptation and Multi-Agent Learning

5th, 6th, and 7th European Symposium, ALAMAS 2005-2007
on Adaptive and Learning Agents and Multi-Agent Systems
Revised Selected Papers

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Karl Tuyls
Maastricht University
The Netherlands
E-mail: k.tuyls@micc.unimaas.nl

Ann Nowe
Vrije Universiteit Brussel
Belgium
E-mail: ann.nowe@vub.ac.be

Zahia Guessoum
University of Pierre and Marie Curie
France
E-mail: zahia.guessoum@lip6.fr

Daniel Kudenko
The University of York
United Kingdom
E-mail: kudenko@cs.york.ac.uk

Library of Congress Control Number: 2008920332

CR Subject Classification (1998): I.2.11, I.2, D.2, C.2.4, F.3.1,D.3.1, H.5.3, K.4.3

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-540-77947-7 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-77947-6 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12225323 06/3180 5 4 3 2 1 0

Preface

This book contains selected and revised papers of the European Symposium on Adaptive and Learning Agents and Multi-Agent Systems (ALAMAS), editions 2005, 2006 and 2007, held in Paris, Brussels and Maastricht.

The goal of the ALAMAS symposia, and this associated book, is to increase awareness and interest in adaptation and learning for single agents and multi-agent systems, and encourage collaboration between machine learning experts, software engineering experts, mathematicians, biologists and physicists, and give a representative overview of current state of affairs in this area. It is an inclusive forum where researchers can present recent work and discuss their newest ideas for a first time with their peers.

The symposia series focuses on all aspects of adaptive and learning agents and multi-agent systems, with a particular emphasis on how to modify established learning techniques and/or create new learning paradigms to address the many challenges presented by complex real-world problems.

These symposia were a great success and provided a forum for the presentation of new ideas and results bearing on the conception of adaptation and learning for single agents and multi-agent systems. Over these three editions we received 51 submissions, of which 17 were carefully selected, including one invited paper of this year's invited speaker Simon Parsons. This is a very competitive acceptance rate of approximately 31%, which, together with two review cycles, has led to a high-quality LNAI volume.

We hope that our readers will be inspired by the papers included in this volume.

Organizing a scientific event like ALAMAS, and editing an associated book, requires the help of many enthusiastic people. First of all, the organizers would like to thank the members of the Program Committee, who guaranteed a scientifically strong and interesting LNAI volume. Secondly, we would like to express our appreciation to the invited speakers of the the editions 2005, 2006 and 2007: Michael Rovatsos (2005), Tom Lenaerts (2006), Eric Postma (2007), and Simon Parsons (2007), for their distinguished contributions to the symposium program. Finally, we also would like to thank the authors of all contributions for submitting their scientific work to the ALAMAS symposium series.

November 2007

Karl Tuyls
Ann Nowé
Zahia Guessoum
Daniel Kudenko

Organization

Organizing Committee

Co-chairs	Karl Tuyls
	Ann Nowé
	Zahia Guessoum
	Daniel Kudenko

Program Committee (2005, 2006, 2007)

Bram Bakker
Ana Bazzan
Ozalp Babaoglu
Frances Brazier
Patrick de Causmaekcker
Philippe De Wilde
Marina Devos
Kurt Driessens
Saso Dzeroski
Marie-Pierre Gleizes
Zahia Guessoum
Pieter Jan't Hoen
Tom Holvoet
Steven de Jong
Dimitar Kazakov
Franziska Kluegl
Ville Kononen
Daniel Kudenko
Paul Marrow
Peter McBurney
Ann Nowé
Luis Nunes
Eugenio Oliveira
Liviu Panait
Simon Parsons
Paolo Petta
Enric Plaza
Marc Ponsen
Michael Rovatsos
Olivier Sigaud
Henry Soldano
Malcolm Strens
Kagan Tumer
Karl Tuyls
Katja Verbeeck
Danny Weyns
Marco Wiering
Niek Wijngaards

Table of Contents

To Adapt or Not to Adapt – Consequences of Adapting Driver and Traffic Light Agents

Ana L.C. Bazzan[1], Denise de Oliveira[1], Franziska Klügl[2], and Kai Nagel[3]

[1] Instituto de Informática, UFRGS
Caixa Postal 15064, 91.501-970 Porto Alegre, RS, Brazil
{bazzan,edenise}@inf.ufrgs.br
[2] Dep. of Artificial Intelligence, University of Würzburg
Am Hubland, 97074 Würzburg, Germany
kluegl@informatik.uni-wuerzburg.de
[3] Inst. for Land and Sea Transport Systems, TU Berlin
Salzufer 17–19, 10587 Berlin, Germany
nagel@vsp.tu-berlin.de

Abstract. One way to cope with the increasing traffic demand is to integrate standard solutions with more intelligent control measures. However, the result of possible interferences between intelligent control or information provision tools and other components of the overall traffic system is not easily predictable. This paper discusses the effects of integrating co-adaptive decision-making regarding route choices (by drivers) and control measures (by traffic lights). The motivation behind this is that optimization of traffic light control is starting to be integrated with navigation support for drivers. We use microscopic, agent-based modelling and simulation, in opposition to the classical network analysis, as this work focuses on the effect of local adaptation. In a scenario that exhibits features comparable to real-world networks, we evaluate different types of adaptation by drivers and by traffic lights, based on local perceptions. In order to compare the performance, we have also used a global level optimization method based on genetic algorithms.

1 Introduction

Urban mobility is one of the key topics in modern societies. Especially in medium to big cities, the urban space has to be adapted to cope with the increasing needs of transportation. In transportation engineering, the expression of the transport needs is called *demand*. This demand (in terms volume of vehicles, pedestrians, freight, etc.) is commonly used to evaluate transport *supply*. This is the expression of the capacity of transportation infrastructures and modes. Supply is expressed in terms of infrastructure (capacity), service (frequency), and other characteristics of the network. The increasing demand of transport needs 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 or economically attainable or feasible,

K. Tuyls et al. (Eds.): Adaptive Agents and MAS III, LNAI 4865, pp. 1–14, 2008.

transportation and traffic engineering seek to optimize the management of both supply and demand using concepts and techniques from intelligent transportation systems (ITS). These refer to the application of modern technologies in the operation and control of transportation systems [12].

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 centre (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 distribute the demand within the available infrastructure. In this context, it is now commonly 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, within the project "Large Scale Agent-based Traffic Simulation for Predicting Traffic Conditions", our long term goal is to tackle a complex problem like traffic from the point of view of information science. This project seeks to integrate microscopic modelling tools developed by the authors for traffic and transportation control and management. These range from traffic signal optimization [1], binary route choice, and effect of information on commuters [4], to microscopic modelling of physical movement [7].

An important milestone in the project is to propose a methodology to integrate complex behavioral models of human travellers reacting to traffic patterns, and control measures, focusing on distributed and decentralized methods. Classically, this is done via network analysis. Using this technique, it is assumed that individual road users seek to optimize their individual costs regarding the trips they make by selecting the "best" route among the ones they have experienced or have been informed about. This is the basis of the well known traffic network analysis based on Wardrop's equilibrium principle [17]. This method predicts a long term *average* state of the network. However, since it assumes steady state network supply and demand conditions, 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 both the variability in the demand and the available 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 equilibrium-based concepts overlook this variability, it seems obvious that they are not adequate for microscopic modelling and simulation. Therefore, the general aim of this paper is to investigate what happens when different actors adapt, each having its own goal. The objective of *local* traffic control is obviously to find a control scheme that minimizes queues in a spatially limited area (e.g. around a traffic light). The objective of drivers is normally to minimize their individual travel time – at least in commuting situations. Finally, from the point of view of the whole system, the goal is to ensure reasonable

travel times for *all* users, which can be highly conflicting with some individual utilities (a social dilemma). This is a well-known issue: for instance, Tumer and Wolpert [15] have shown that there is no general approach to deal with this complex question of collectives.

Specifically, this paper investigates which strategy is the best for drivers (e.g. adaptation or greedy actions). Similarly, traffic lights can act greedily or simply carry out a "well-designed" signal plan. At which volume of local traffic does decentralized control of Traffic Lights start to pay off? Does isolated, single-agent reinforcement learning make sense in dynamic traffic scenarios? What happens when many drivers adapt concurrently? These are hot topics not only in traffic research, but also in a more general multi-agent research as they refer to co-adaptation.

In this paper we depart from binary route choice scenarios and use a more realistic one, that shows features such as: heterogeneity of origin-destination pairs, heterogeneous capacity, and agents knowing about a set of routes between their origins and destinations. To the best of our knowledge, the question on what happens when drivers and traffic lights co-adapt in a complex route scenario has not been tackled so far.

In the next section we review these and related issues. In section 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 Traffic Engineering

Learning and adaptation is an important issue in multiagent systems. Here, we concentrate on pieces of related work which either deal with adaptation in traffic scenarios directly or report on 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, such as mathematics, physics, computer science, engineering, geography, and architecture. Therefore, several tools exist that target only a part of the overall problem. For example, 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 [11]. It consists of: trip generation, destination choice, mode choice, and route assignment. Route assignment includes route choice and a very basic traffic flow simulation that may lead to a Nash Equilibrium. Over the years, the 4-step-process has been improved in many ways, most mainly by (i) combining the first three steps into a single, traveller-oriented framework (*activity-based demand generation (ABDG)*) and by (ii) replacing traditional route assignment by so-called *dynamic traffic assignment (DTA)*. Still, in the actual implementations, all travellers' information gets lost in the connection between ABDG and DTA, making realistic agent-based modelling at the DTA-level difficult.

Another related problem is the estimation of the overall state of the complete traffic network from partial sensor data. Although many schemes exist for incident detection, there are only few applications 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 applies intelligent methods combining the current state with historical data in order to make short-term predictions. However, the travellers themselves 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 for simulating the real infrastructure. However, for evaluating the effects of travellers' flexible decision making, it is necessary to overcome this simplification for integrating additional information about dynamic decision-making context.

A true integration of these and other approaches is still missing. Agent technology offers the appropriate basis for this. However, until now agent-based simulations with a scale required for the simulation of real-world traffic networks have not been developed.

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. Alternatively, in a traffic-responsive system, at least one signal plan must be pre-defined in order to be changed on the fly.

In [1], a MAS based approach is described in which each traffic light is modelled as an agent, each having a set of pre-defined signal plans to coordinate with neighbours. Different signal plans can be selected in order to coordinate in a given traffic direction. This approach uses techniques of evolutionary game theory. However, payoff matrices (or at least the utilities and preferences of the agents) are required. These figures have to be explicitly formalized by the designer of the system.

In [10], groups of traffic lights 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. Also, the mediation process may take long in highly constrained scenarios, having a negative impact in the coordination mechanism.

Also a decentralized, swarm-based model of task allocation was developed in [9], in which the dynamic group formation without mediation combines the advantages of decentralization via swarm intelligence and dynamic group formation.

Regarding the use of reinforcement learning for traffic control, some applications are reported. Camponogara and Kraus [2] 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 [18]. In [8] a set of techniques were tried in order to improve the learning ability of the agents in a simple scenario. Performance of reinforcement learning approaches such as Q-learning and Prioritized Sweeping in non-stationary environments are compared in [13]. Co-learning is discussed in [19] (detailed here in Section 2.3).

Finally, a reservation-based system [3] is also reported but it is only slightly related to the topics here because it does not include conventional traffic lights.

2.3 The Need for Integration

Up to now, only few attempts exist to integrate supply and demand in a single model. We review three of them here.

Learning Based Approach. A paper by [19] 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. Additionally, 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: value functions are learned not only by the traffic lights, but also by the vehicles which thus can compute policies to select optimal routes to the respective destinations. The ideas and results presented in that paper are interesting. However, it makes strong assumptions that may hinder its use in the real world: the kind of communication and knowledge or, more appropriate, communication *for* knowledge formation has high costs. Traffic light controllers are supposed to know vehicles destination in order to compute expected waiting times for each. Given the current technology, this is a quite strong assumption. Secondly, 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 (few) past experiences? Finally, drivers being autonomous, it is not completely obvious that they will use the best policy computed by the traffic light and not by themselves. Therefore, in the present paper, we depart from these assumptions regarding communication and knowledge the actors must have about each other.

Game Theoretic Approach. In [16] a two-level, three-player game is discussed that integrates traffic control and traffic assignment, i.e. both, the control of Traffic Lights and the route choices by drivers are considered. 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 assumption that drivers always follow their shortest routes is 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 centres are in charge of the control of Traffic Lights. Rather, we follow a trend of decentralization, in which each 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 [16]. Moreover, 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.

Methodologies. Liu and colleagues [6] describe a modelling approach that integrates microsimulation of individual trip-makers' decisions and individual vehicle movements across the network. Moreover their focus is on the description of the methodology that integrates 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 of them can roughly be described as greedy, while the other is fixed signal plan based. In the present paper, we do not explore the methodological issues as in [6] but, rather, investigate in more details particular issues of the integration and interaction between actors from the supply and demand side.

3 Co-adaptation in an ITS Framework

Figure 1 shows a scheme of our approach based on the interaction between supply and demand. This framework was developed using the agent-based simulation environment SeSAm [5] for testing the effects of adaptation of different elements of the supply and demand. The testbed consists of sub-modules for specification and generation of the network and the agents – traffic lights and drivers. 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.

The basic scenario we use is a typical commuting scenario where drivers repeatedly select a route to go from an origin to a destination. As mentioned before, we want to go beyond simple two-route or binary choice scenario; we deal with route choice in a network with a variety of possible routes. Thus, it captures desirable properties of real-world scenarios.

We use a grid with 36 nodes connected using one-way links, as depicted in Figure 2. All links are one-way and drivers can turn to two directions in each crossing. Although it is apparently simple, this kind of scenario is realistic and, from the point of view of route choice and equilibrium computation, it is also

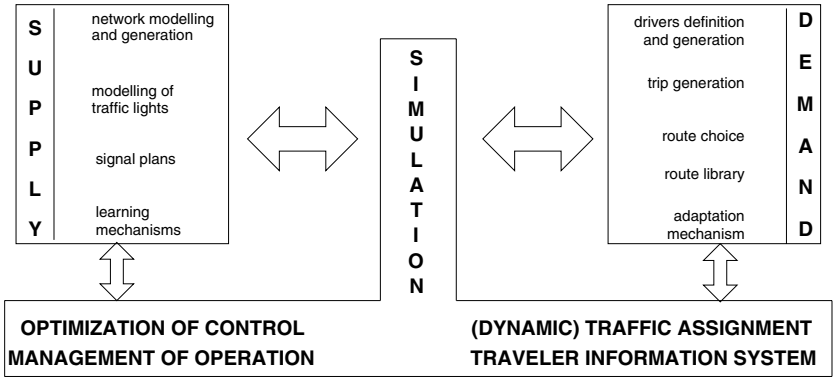

Fig. 1. Elements of Co-Adaptation in an ITS Framework

a very complex one as the number of possible routes between two locations is high.

In contrast to simple two-route scenarios, it is possible to set arbitrary origins (O) and destinations (D) in this grid. For every driver agent, its origin and destination are randomly selected according to probabilities given for the links: To render the scenario more realistic, neither the distribution of O-D combinations, nor the capacity of links is homogeneous. On average, 60% of the road users have the same destination, namely the link labelled as E4E5 which can be thought as something like a main business area. Other links have, each, 1.7% probability of being a destination. Origins are nearly equally distributed in the grid, with three exceptions (three "main residential areas"): 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%. Regarding capacity, all links can hold up to 15 vehicles, except those located in the so called "main street". These can hold up to 45 (one can think it has more lanes). This main street is formed by the links between nodes B3 to E3, E4, and E5.

The control is performed via decentralized Traffic Lights. These are located in each node. Each of the Traffic Lights has a signal plan which, by default, divides the overall cycle time – in the experiments 40 time steps – 50-50% between the two phases. One phase corresponds to assigning green to one direction, either north/south or east/west.

The actions of the Traffic Lights consist in running the default plan or to prioritize one phase. The particular strategies are:

i. fixed: always keep the default signal plan
ii. greedy: allow more green time for the direction with higher current occupancy
iii. use single agent Q-learning

Regarding the demand, the main actor is the simulated driver. The simulation can generate any number of them; in the experiments we used 400, 500, 600,

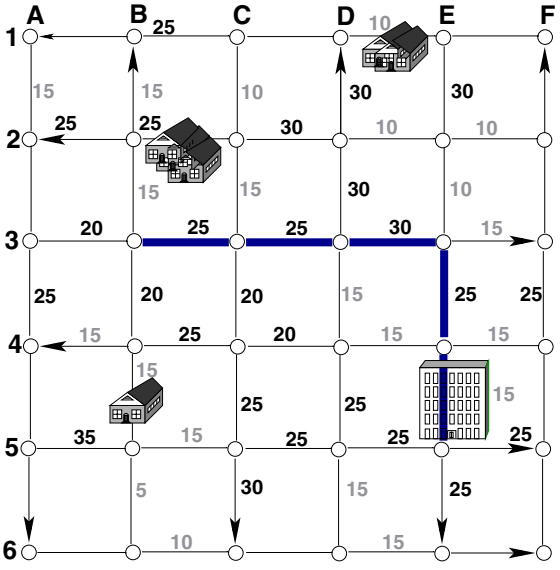

Fig. 2. 6x6 grid showing the main destination (E4E5), the three main origins (B5B4, E1D1, C2B2), and the "main street" (darker line). Numbers at the links represent the green times for the particular direction (determined by global optimization).

and 700 driver agents. Every driver is assigned to a randomly selected origin-destination pair. Initially it is informed about only a given number of routes. The experiments presented next were performed with each agent knowing five routes. These route options are different for each driver and were generated using an algorithm that computes the shortest path (one route) and the shortest path via arbitrary detours (the other four). We notice that, due to topological constraints, it was not always possible to generate five routes for each driver. One example is the following: origin and destination are too close. Thus, in a few cases they know less than this number, but at least one. Drivers can use three strategies to select a route (before departure):

i. random selection
ii. greedy: always select the route with best average travel time so far
iii. probabilistically: for each route, the average travel time perceived so far is use to compute a probability to select that route again.

The actual movement of the driver agents through the net is queue-based.

4 Results and Discussion

4.1 Metrics and Parameters

In order to evaluate the experiments, travel time (for drivers) and occupation (for links) were measured. We discuss here only the mean travel time over the

last 5 trips (henceforward $attl5t$) and travel time in a single trip. All experiments were repeated 20 times.

The following parameters were used: time out for the simulation of one trip (t_{out}) equal to 300 when the number of drivers is 400 or 500; 400 when there are 600 drivers; and 500 when there are 700 drivers.

The percentage of drivers who adapt is either 0 or 100 (in this case all act greedily) but any value can be used; percentage of Traffic Lights that act greedily is either 0 or 100; a link is considered jammed if its occupancy is over 50%; cycle length for signal plans is 40 seconds.

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 $\lambda = 0.9$.

4.2 Global Optimization

For the sake of comparison, we show the results of a centralized approach before we continue with the main focus of the paper on local (co-)adaptation approaches. We use a centralized and heuristic optimization method in order to compute the optimal split of the cycle time between two traffic directions at each intersection.

This centralized optimization was performed using the DAVINCI (Developing Agent-based simulations Via INtelligent CalIbration) Calibration Toolkit for SeSAm, that is a general purpose calibration and optimization tool for simulation. Although DAVINCI provides several global search strategies such as genetic algorithm (GA), simulated annealing or gradient based search, here we have used standard GA only, with a fitness proportional selection.

The input parameters for the GA are the default split values for each of the 36 traffic light agents (see next). The optimization objective is to minimize the average travel time over all drivers in a scenario with 400 drivers, where all drivers have only one route (the shortest path).

For a cycle length of 40 seconds, we have set seven possible values for the split at each intersection: 5/35, 10/30, 15/25, 20/20, ..., 35/5. Using four bits to codify each of these splits, for each of the 36 intersection, this leads to 144 bits for each GA string. We have allowed the GA to run for 100 generations.

The resulting optimized splits can be seen in Figure 2: numbers depicted close to the respective links indicate how much green time the link receives in the best solution found by the GA. Using these optimized splits, the average travel time of drivers is 105. This value can be used as a benchmark to assess the utility of adapting drivers and Traffic Lights in a decentralized way.

4.3 Drivers and Traffic-Lights Learning in a Decentralized Way

In this section we discuss the simulations and results collected when drivers and Traffic Lights co-adapt using different strategies, as given in Section 3. As a measure of performance, we use the $attl5t$ defined previously (Section 4.1). These are summarized in Table 1. For all scenarios described in this subsection, 400 drivers were used. As said, all experiments were repeated 20 times. Standard deviations are not higher than 4% of the mean value given here.

Table 1. Average Travel Time Last 5 Trips (*attl5t*) for 400 drivers, under different conditions

Type of Simulation	Average Travel Time Last 5 Trips
greedy drivers / fixed traffic lights	100
probabilistic drivers / fixed traffic lights	149
greedy drivers / greedy traffic lights	106
probabilistic drivers / greedy traffic lights	143
greedy drivers / Qlearning traffic lights	233
probabilistic drivers / Qlearning traffic lights	280

Greedy or Probabilistic Drivers; Fixed Traffic Lights. In the case of probabilistic drivers, the *attl5t* is 149 time units, while this is 100 if drivers act greedily. The higher travel time is the price paid for the experimentation that drivers continue doing, 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 of a simulation run, when experimentation does pay off, the probabilistic driver performs better.

Notice that this travel time is slightly better than the one found by the heuristic optimization tool described before, which was 105. 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 the case where drivers take longer routes.

Greedy or Probabilistic Drivers; Greedy Traffic Lights. When Traffic Lights also act greedily we can see that this does not automatically improve the outcome (in comparison with the case in which Traffic Lights are fixed): the *attl5t* is 106. This happens because the degree of freedom of Traffic Lights' actions is low, as actions are highly constrained. For example, acting greedily can be highly sub-optimal when, for instance, traffic light A serves direction D_1 (thus keeping D_2 with red light) but the downstream flow of D_1 is already jammed. In this case, the light might indeed provide green for vehicles on D_1 but these cannot move due to the downstream jam. Worse, jam may appear on the previously un-jammed 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. The travel time of 106, when compared to the travel time found by the centralized optimization tool (105), is of course similar. This is not surprising because the decentralized strategy does exactly the same as the centralized optimizer, namely drivers use their best route and Traffic Lights optimize greedily.

Q-Learning Traffic Lights. We have expected Q-learning to perform bad because it is already known that it does not have a good performance in noisy and non-stationary traffic scenarios [13]. In order to test this, we have implemented a

Q-learning mechanism in the traffic lights. Available actions are: to open the phase serving either one direction (e.g. D_1), or the other (D_2). The states are the combination of abstract states in both approaching links, i.e. $\{D_1\_jammed, D_1\_not\_jammed\} \times \{D_2\_jammed, D_2\_not\_jammed\}$.

The low performance of Q-learning in traffic scenarios is due basically to the fact that the environment is non-stationary, not due to the poor discretization of states. Convergence is not achieved before the environment changes again, and thus Traffic Lights remain in the experimentation phase.

4.4 Scenarios with More Drivers

For more than 400 drives, we only investigate the cases of greedy drivers / fixed Traffic Lights *versus* the scenario in which both drivers and Traffic Lights act greedily. This was done in order 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 increases, greedy actions by the drivers alone do not bring much gain; some kind of control in the Traffic Lights is expect to be helpful in case of high occupancy of the network. Notice that 400, 500, 600 and 700 drivers mean an average occupancy of $\approx$ 40%, 47%, 59%, and 72% per link respectively.

In Table 2 the *attl5t* for these numbers of drivers are shown. The case for 400 drivers was discussed above. With more than 600 drivers, the *attl5t* is lower when Traffic Lights also act greedily. In the case of 700 drivers, the improvement in travel time (411 *versus* 380) is about 8%. Thus, the greedy traffic lights are successful in keeping the occupancy of links lower, resulting in a reduction of travel times.

Table 2. Average Travel Time Last 5 Trips for Different Number of Drivers and Different Adaptation Schemes

Average Travel Time Last 5 Trips				
Type of Simulation	Nb. of Drivers			
	400	500	600	700
greedy drivers / fixed traffic lights	100	136	227	411
greedy drivers / greedy traffic lights	106	139	215	380

4.5 Overall Discussion

In the experiments presented, one can see that different strategies concerning the adaptivity of drivers, as well as of Traffic Lights have distinct results in different settings. We summarize here the main conclusions.

For the 6×6 network depicted, increasing the links capacity from 15 to 20 would lead to travel time levels that are the same we have achieved *without* this increase in capacity, i.e. substituting this increase by a better use of the available infrastructure. This is important because increasing network capacity is not always economically feasible, so that other measures must be taken. Diverting people by giving

information to them, has only limited performance. Thus the idea is to use the control infrastructure in a more intelligent way. Therefore, we have explored the capability of the Traffic Lights to cope with the increasing demand.

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}. This is not shown here but, to give a general idea of the figures, bad performance (around 75% arrived) was seen only when the drivers adapt probabilistically. The general trend is that when the traffic lights also adapt, the performance increases, for all metrics used.

Regarding the use of Q-learning, as said, single-agent learning, i.e. each agent learns isolated using Q-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 dependencies 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 multiagent Q-learning in its standard formulation.

5 Conclusion

Several studies and approaches exist for modelling travellers' decision-making. In commuting scenarios in particular, probabilistic adaptation in order to maximize private utilities is one of those approaches. However, there is hardly any 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*. In this case, co-adaptation happens in a decentralized fashion. 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 traffic network both affect and are affected by drivers' reactions to them. This leads to a feedback loop that 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 tool has modules to cope with the demand and the supply sides, as well as to implement the ITS modules and algorithms for the learning, adaptation etc.

Results show an improvement regarding travel time and occupancy (thus, both the demand and supply side) when all actors co-evolve, especially in large-scale situations e.g. involving hundreds of drivers. This was compared with situations in which either only drivers or only Traffic Lights evolve, in different scenarios, and with a centralized optimization method.