Autonomous Agents and Multi-Agent Systems, 10, 131–164, 2005

A Distributed Approach for Coordination of Traffic Signal Agents

ANA L. C. BAZZAN bazzan@inf.ufrgs.br
Instituto de Informática, Universidade Federal do Rio Grande do Sul, CP 15064, 91.501-970 Porto Alegre, RS, Brazil

Abstract. Innovative control strategies are needed to cope with the increasing urban traffic chaos. In most cases, the currently used strategies are based on a central traffic-responsive control system which can be demanding to implement and maintain. Therefore, a functional and spatial decentralization is desired. For this purpose, distributed artificial intelligence and multi-agent systems have come out with a series of techniques which allow coordination and cooperation. However, in many cases these are reached by means of communication and centrally controlled coordination processes, giving little room for decentralized management. Consequently, there is a lack of decision-support tools at managerial level (traffic control centers) capable of dealing with decentralized policies of control and actually profiting from them. In the present work a coordination concept is used, which overcomes some disadvantages of the existing methods. This concept makes use of techniques of evolutionary game theory: intersections in an arterial are modeled as individually-motivated agents or players taking part in a dynamic process in which not only their own local goals but also a global one has to be taken into account. The role of the traffic manager is facilitated since s/he has to deal only with tactical ones, leaving the operational issues to the agents. Thus the system ultimately provides support for the traffic manager to decide on traffic control policies. Some application in traffic scenarios are discussed in order to evaluate the feasibility of transferring the responsibility of traffic signal coordination to agents. The results show different performances of the decentralized coordination process in different scenarios (e.g. the flow of vehicles is nearly equal in both opposing directions, one direction has a clearly higher flow, etc.). Therefore, the task of the manager is facilitate once s/he recognizes the scenario and acts accordingly.

Keywords: coordination of agents, evolutionary game-theory, learning in multi-agent systems, traffic signal control.

1. Introduction

The problems related to the increase of urban traffic jams call for innovative control strategies which allow vehicles to travel more freely as, for instance, coordination (also called synchronization) of traffic signals. In most cases these strategies are based on a central traffic-responsive control system, whose difficulty to implement and maintain increases with the number of traffic elements (traffic lights, detectors, specialized hardware, etc.). Another problem is the lack of interoperability among those elements, specially if purchased from different producers. Therefore, functional and spatial decentralization is desired. In fact, achieving spatial decentralization has been a major goal of traffic engineering [42].

Many authors have stressed the role of limited perception and communication capabilities and the need for implicit communication and learning mechanisms (e.g.

Author partially supported by CNPq.

[18,20,36]). This is a key issue in the traffic domain as it will be shown in Section 3, where other relevant contributions are also briefly discussed.

Decentralized systems in traffic may have several purposes: to attain synchronization of several neighboring intersections on an arterial [4], agent-based macroscopic [46] as well as microscopic simulation of traffic agents [12,25,27,34], and traffic control [33]. In some cases, cooperation is reached by means of a communication-based, centrally controlled coordination process. Nonetheless, this kind of coordination is not satisfactory for the domain of traffic signal control due to real-time and interoperability constraints. Real time constraints here means that agents cannot afford complex, time-consuming negotiation mechanisms due to the need to react immediately to the current traffic situation. Besides, the communication channel may be noisy or restrict. Interoperability constraints are posed by the producers of hardware. Despite effort to have a standard communication protocol, the reality is that the current available hardware does not provide such facilities.

In the present work, a coordination concept is used, which overcomes some of the disadvantages of the existing methods of centralized synchronization of traffic signals. This concept makes use of techniques of evolutionary game theory and of reinforcement learning. Intersections in an arterial are modeled as individually motivated agents or players taking part in a dynamic process, where not only their own local goals but also a global one has to be taken into account. Moreover, each agent possesses only information about their local traffic states.

In order to optimize the traffic flow, each agent may select a suitable action (a signal plan as it will be defined in Section 3). According to their local goals, agents are classified into different types. The information regarding this classification is not shared by neighbors. The joint selection of actions in the neighborhood of the agent defines the new traffic state, which yields a gain or loss for each agent. Decisions (action selections) are made at discrete time intervals. These action selection periods are interwoven with learning periods. In the latter, agents learn by reinforcement. According to the outcome of each action, the learning rule gives the probability with which every action should be played in the future. This way, the mechanism allows a sort of indirect agreement amidst agents toward specific action(s) aiming at reducing queues at intersections.

Ultimately, this paper focuses on distributed coordination strategies which provide the traffic manager with support to decide on traffic control policies both at an operational as well as at a tactical level. The strategical level could possibly be tackled by such a tool as well, but since this goes beyond the decision level of the traffic manager (as it potentially involves topological changes and/or infrastructural measures), it is not taken into account in the current work.

This text is organized as follows: the next section discusses the motivation for a paradigm shift in traffic control and coordination of traffic signals, as well as previous works in this direction. Section 3 gives a brief overview on some concepts related to traffic control and policies for coordination of traffic signals, and also on previous solutions related to decision-support for traffic control, most based on artificial intelligence (AI) techniques. In Section 4 some concepts of game-theoretic approaches already used in distributed artificial intelligence and multi-agent systems are highlighted. It is shown that for the traffic scenario tackled here, the evolutionary game-theoretic approach is more suitable. Sections 5 and 6 discuss the details of the approach proposed and the results respectively. Finally, Section 7 concludes the paper and outlines the future research directions.

2. Autonomous agents providing a shift in paradigm: from a centralized to a decentralized control policy

2.1. Motivation

Coordination in systems in which there is a shortage of resources has been a central issue of a number of works. In traffic domain, coordination has been usually tackled in a centralized fashion. This seems to be particularly true if communication channels have to be used efficiently, since it is important for the coordination mechanism not to consume much processing and communication resources.

One of our goals is to show that coordination may also emerge in the absence of a central authority, although it may require a longer time to achieve it since agents need time to observe other agents, build models of them, and learn their behaviors. In traffic domains this means that the traffic manager can be free to analyze high-level issues like changes in routine patterns; the low level issues would be solved by the agents in charge of analyzing usual tasks such as synchronizing traffic lights in neighboring intersections.

The motivation for using agents in this scenario is that there is a solid research line concerned with the study of how agents can cooperate in order to solve a global problem that is beyond their individual capabilities. However, achieving cooperation and coherent coordination is not trivial. Therefore, an initial motivation for this work is to adapt mechanisms for the emergence of cooperation among the participants of a system, especially when they have their own goals and are not necessarily cooperative. The second motivation is to validate distributed coordination mechanisms in traffic control.

In order to draw conclusions about the effectiveness of decentralized mechanisms, it is necessary to make an overview of how transportation and traffic engineering have coped with the question of traffic control. Initially, techniques of operations research, statistics, and computer simulation were tried [1,8,9,43]. More recently, the use of AI techniques (mainly expert systems and agent-based approaches) have also stirred increasing interest. Classical examples are the sub-areas of transportation logistics, where heuristic-based systems play an important role [15]; traffic management using intelligent systems [10,13,16,33,46,47]; and air-traffic control, where the inherent decentralized nature of the planning problem has motivated the use of agents techniques [7].

Recent technological developments concerning software and use of AI have enabled the control at intersections to be increasingly intelligent and autonomous. For instance, there is a trend to replace the centralized philosophy of traffic control by a hierarchical one composed of several sub-networks [4,6,11,13,19,24,29,33,42,46]. While this seems to reduce communication and to increase reliability by allowing data to be processed quickly, it demands more sophisticated software and data structures: when one speaks about *single* traffic light, they are normally embedded systems, whereas one has to design databases and interfaces for the end user (e.g., traffic engineer) when a centralized system is to be used.

An alternative is a totally decentralized policy: local units (for instance the signalized intersections) act independently while performing their tasks such as data acquisition, processing, reasoning, planning, and the carrying out of the signal plans. However, such units should be capable not only of processing local information, but also of interacting with neighbors when necessary in order to optimize traffic flow. Achieving such degree of decentralization while also reducing communication seem to be conflicting goals when one thinks that, the more decentralized a system, the more coordination it requires in order to improve efficiency. Because coordination is normally reached through communication, there is a clear need for a new mechanism of coordination which, if not capable of avoiding, at least minimizes communication.

Consequently, a third motivation of the present work is exactly the use of an approach which seeks coordination with reduced communication. A natural way to tackle the problem of developing a new coordination mechanism with low communication is to look at fields where a similar question has appeared and has been already studied, like microeconomics and, in particular, game-theory (see Section 4).

From the operative point of view, in the proposed approach, communication among participants can be minimized since they do not need to have full knowledge before making a decision. Since they somehow model neighbors, they are able to reason about others' states without need of an explicit communication process.

2.2. Related work

Rosenschein and Genesereth [36] have raised the question about the "benevolent agent assumption": in the real world benevolence cannot be taken for granted. Even in situations where all participants may have a single general goal, it is unlike that this remains true when talking about sub-goals. Conflict often arises after the decomposition of a task into subtasks and/or due to resource allocation. Their framework was later extended to an environment where less communication is allowed [20]. Apart from Rosenschein and colleagues' researches, many other comprehensive studies tackle the use of other aspects of the microeconomics in MAS. However we restrain ourselves here to evolutionary game-theoretic approaches. Agents which have some capacity to cope with complex situations should be able to react in particular situations. Hence, they should be designed with capabilities such as prediction and conflict resolution.

Finally, regarding distributed approaches in transportation, there are various research results worth noting. In urban traffic control, the Prodyn system [44] and the approach described in [13] are both based on communication of information from a single intersection within a determined neighborhood, which leads to complex protocols and high costs. Moreover there is no mechanism of conflict resolution. This point is addressed in [46] by means of a negotiation protocol (thus increasing communication). Other works on transportation are reviewed at the end of the next section.

3. A brief overview on policies for urban traffic control

3.1. Introduction

Transportation in general and logistics in particular are only slightly related to this paper. We refer the interested reader to [7] and [14] for early approaches of MAS to air-traffic control and to the logistics respectively. For agent-based traffic *simulation* please refer to [12,25,27,36,34].

The focus of this paper is on urban traffic control (UTC). The attractiveness of agent based models in UTC arises from the inherent distribution of the functionality of the components. Such distribution gives room for coordination and cooperation mechanisms among traffic-signal agents, in order to cope with their local goals, and also with overall goal of the traffic network. However, it is desirable that the coordination mechanisms do not require a high degree of communication during the negotiation process, as explained in Section 1.

3.2. Terminology and basic concepts in urban traffic control

The main goals of UTC systems are: to maximize the overall capacity of the network; to maximize the capacity of critical routes and intersections which represent the bottlenecks; to minimize the negative impacts of traffic on the environment and on energy consumption; to minimize travel times; and to increase traffic safety.

The control of traffic flow has the main functional objectives: provision for orderly movement of traffic; increase of traffic-handling capacity on the intersection; reduction of the frequency of accidents; and synchronization of traffic signals in order to provide for continuous movement of traffic at a defined speed along an arterial (in synchronized or progressive systems).

In order to achieve these goals, devices to control the flow of vehicles (e.g., traffic signals) have been used. Traffic signals can vary from hard-wired logic to computerized control, either centralized or not. It is also possible to acquire data from buried detectors (e.g., of loop-induced type) in order to perform a simple traffic-responsive local control. In a centralized and computerized system, the central computer sends instructions to several traffic signals either on a time base or according to detector information.

Signalized intersections are controlled by signal-timing plans which are implemented at traffic signals. A signal-timing plan (henceforth signal plan for short) is a unique set of timing parameters comprising basically the cycle length (the length of time for the complete sequence of the phase changes), the split (the division of the cycle length C among the various movements or phases), pedestrian requirements for timing, and the phase-change interval.

The design of traffic signal plans consists basically of two stages. The first is the division of the valid traffic movements at an intersection into different phases, so that the movements in each phase are free of conflict. This is normally made by a traffic expert. The second stage consists of finding the timing of the signal plan. This includes the determination of the cycle time, the split, and the phase-change interval.

The criteria for obtaining the optimum signal timing is that it should lead to the minimum overall delay at the intersection. This is usually achieved by using simulation or optimization programs. Several plans are normally required for an intersection (or set of intersections in the case of a synchronized system) to deal with changes in traffic flow.

3.3. Traffic signal control systems and structures

Concerning the level of aggregation, systems can be utilized to control: an individual intersection, an arterial, and a grid or network. According to the control strategy, two strategies can be distinguished: fixed time and traffic-responsive signal control. Thus, four variants can be described: fixed-time control, traffic-responsive signal plan selection, traffic-responsive signal plan modification, and traffic-responsive signal plan generation.

In the first strategy, a fixed-time (pretimed) controller works on a time-of-the-day basis. It is the cheapest and logical choice for network with stable or predictable traffic behavior. However, a fixed-time controller cannot cope with unexpected changes in traffic flow. Moreover, the key problem in developing fixed-time control systems is precomputing signal plans from historical traffic data. On the other side, actuated control, which comprises the three remaining strategies, makes use of buried detectors on all or some approaches to the intersections. These strategies are necessary where the traffic flow does not have stable pattern.

As for control structures, the principal ones are:

- **First generation** (non-computerized systems): the control functions are performed either by especially designed hard-wired logic in the form of an electromechanical device or by electronic logic;
- **Second generation** (centralized computer control): the individual control tasks can be carried out by a single computer if the number of intersections is relatively small. If several hundreds or thousands of intersections have to be coupled to the control center, then the installation of a computer hierarchy system may be necessary;
- **Third generation** (distributed computer control): microprocessors allows the individual intersections to be provided with their own processing unit. Measurements from detector can be locally evaluated and need no longer to be transmitted to the control center. Two versions of distributed traffic light control systems may be distinguished:
  1. hierarchically structured distributed control systems: in large networks, it is necessary to have sub-control centers;
  2. totally distributed traffic control systems: no control center exists; each local processor solves the control tasks occurring at its own intersection.

3.4. Synchronization in arterials: basics

The goal of coordinated systems (also called synchronized or progressive systems) is *to synchronize the traffic signals along an arterial* in order to allow vehicles, traveling at a given constant speed, to cross the arterial without stopping at red lights. This is known as a green wave. Normally, these are used for morning and afternoon rush hours (i.e., for fixed times of the day).

Apart from the control parameters discussed in Section 3.2, a coordinated system also needs the so-called *offset* (time between the beginning of the green phase of consecutive traffic signals), and the desired speed of vehicles V. The fixed parameters are the geometry of the intersection (number of lanes, etc.) as well as the geometry of the arterial (distance between intersections). Another important related concept is the bandwidth. There is a particular time period during which a vehicle is able to continue without stopping at any intersection if it maintains a fixed speed V.

The classical problem concerning synchronization systems is to find the optimal (larger) bandwidth for different cycle times and speeds. The most popular solution so far employed uses linear programming (e.g., [32]). Using these methods one can find the optimal cycle time and optimal speed. Two or more signalized intersections can be operated in synchronization when their traffic signals are commensurate in time. This synchronization can be imposed through a communication system or through a clock at each intersection.

Well designed signal plans can achieve acceptable results in *not completely congested streets in one flow direction*. However progression in two opposing directions of an arterial is difficult to achieve, if not impossible, in almost all practical situations. The difficulty is that the geometry of the arterial is fixed and with it the spacing between adjacent intersections. Only in very special cases (for instance when the spacing among intersections is almost the same, no left turns are allowed, and the traveling speed is constant and equal for both directions) the geometry allows progression in opposite directions. Synchronization in four directions is, for practical purposes, impossible.

As a measure of effectiveness of such systems, one generally seeks to optimize a weighted combination of stops and delays as in TRANSYT [35]. The commonly implemented approaches are synchronization with fixed time control and synchronization with traffic-responsive signal plan selection or modification. However for large networks or arterials where traffic patterns change often, there is a trend to use hierarchical or totally distributed approaches in order to reduce the complexity of control programs, as already discussed in Section 2.1.

3.5. Synchronization in arterials: approaches and tools

Some algorithms were proposed in the sixties and seventies to analyze traffic patterns and to set traffic signal cycle length and cycle splits (e.g., TRANSYT [35,43] and PASSER [8]).

A new generation of simulation tools and decision-support systems appeared in the 1980s. These were based either on a decentralized philosophy of control, or on the use of AI techniques. Some examples of the former approach are discussed below, followed by a brief description of the use of AI and multi-agent techniques.

Prodyn [24] as well as OPAC [19] and UTOPIA [11] are adaptive programs in which the control is not centralized. In Prodyn for instance, a decision is taken at each 5 seconds concerning whether to change phases or not. In a typical case, each intersection simulates all possible situations using detector information in adjacent areas. This information propagates from intersection to intersection with a decreasing weight. Both the relative complex computation and the communication system can increase the cost of implementation.

In OFSET [29], fully and semi-actuated signals located in several intersections aligned in a given direction communicate with each other via a simple databus, without a central computer. On the databus, a cycle is established where intersections periodically send information to one another about network volume, cycle length and offset. Each intersection computes a desired cycle length. Offsets are then calculated by a routine which searches for a value which minimizes total intersection delay. However this approach does not deal with the situation where the synchronization has to be shifted automatically to another direction since this procedure would need coordination.

In the 1980s, the status of traffic control problems has grown in importance amidst the AI and multi-agent community. However, the literature reports only a few proposals, ideas and tentative implementations. The reason for this may lie in the difficulty of modeling all constraints which appear in interactions among all traffic elements.

Initially, expert systems have motivated a series of studies in the field of UTC. The idea behind is to use expert knowledge from experienced traffic engineers, which is encoded in a knowledge base. An expert system approach to this problem was first tried by Zozaya-Goristiza and Hendrickson [47]. After this seminal work, expert systems such as the French System [16] were implemented. Although traffic waiting-time has been reduced, they suffer from several shortcomings: since they must deal with tremendous amounts of information, they are very slow. Also, they usually look after global problems and are then unable to deal with important local changes. Distributed and dynamic control philosophies have not been used although they offer a number of advantages.

The system described by Findler and Stapp [13] is based on the propagation of information in a network of traffic signals. The authors assume a street configuration in a grid form, with some simplifications. There is one processor at each intersection which communicates directly with the four processors at the adjacent intersections.

The operation of the whole system is based on a set of collaborating real-time expert systems working in conjunction with a simulation based planner. The decision to adjust cycle starting time is a function of the weighted recommendations transmitted from the adjacent and more distant processors. Although their work accounts for some degree of decentralization, it does not cope with conflict resolution, for instance.

Also multi-agent techniques have been tried in the UTC domain. The approaches can be classified into three levels: integration of heterogeneous traffic management systems, traffic flow control, and individualized traffic guidance. The platform called Multi-Agent Environment for Constructing Cooperative Applications - MECCA/UTS - [23] addresses the first and third levels of control just mentioned. More recently the third level has been the focus of many research studies [5,37,38,39,45]. For the second level – traffic flow control, which is our interest in this paper – previous works are the SAPPORO system [46] and extensions [3,4].

4. The motivation for evolutionary game theory

4.1. Game theory and decision-making

Microeconomics modeling usually considers no central authority and decentralized supply and demand. This modeling uses utility theory (to represent the preferences of each of the participants in a market), decision theory (to deal with uncertainty and to perform the operations regarding the utility of every preference), and game theory (to cope with conflicts of interest among the participants).

Game theory can be viewed as an extension of decision theory to the case where there is more than one decision-maker. The game-theoretic analysis of any interaction is fundamentally based on the concept of rationality: each participant acts in pursuit of maximizing its expected utility subject to its knowledge and capacities.

An additional, but not less interesting question is how participants form beliefs that are common knowledge. Halpern and Moses [21] have proved that arriving at common knowledge in practical distributed systems is impossible in finite time. Therefore it cannot be guaranteed that all participants are well informed. A more reasonable assumption is that, within time, participants learn a behavior by observing the interactions in which they are involved. Rationality is replaced by a process of natural selection with participants having no need to model their utilities, but measuring their fitness instead. In this way the forecast of the equilibrium point in a decision-making process is not reached by assuming participants behaving rationally, but by letting them reach a stable point out of the dynamics of the interaction.

Although game-theory has used many paradigms to explain certain types of interaction and conflicts of interest, the present work is primarily concerned with the so-called *coordination games*, a class of games in which there is more than one possible joint decision arising from players trying to maximize their payoffs.

4.2. Pure-coordination game

Here one solution may pareto-dominate the other. As an example of such a game, one can imagine a payoff matrix as shown in Table 1, with $a = 2$, $b = 1$ and $c = 0$ for instance.

By playing both $E_1 = (a_1, a_1)$ or $E_2 = (a_2, a_2)$ players have no reason to deviate. However the former is clearly better. The fact that not all games have a unique Nash equilibrium, is indeed one of the problems with the Nash concept. Uniqueness is crucial for the majority of real world problems modeled by game theory. In MAS, agents which rely on this modeling to negotiate with other agents may not be able to reason if they are not sure about which action other agents will choose. Therefore, when a game has more than one Nash equilibrium, a method is required which rules out some of them, eventually leading to a unique equilibrium point. In case some or all equilibria are equally plausible, it is impossible to predict how players are going to act. However in some cases, some equilibrium points are more plausible than others.

Table 1. Pure-coordination game: payoff matrix.

		agent 1	
		a_1	a_2
agent 2	a_1	a / a	c / c
	a_2	c / c	b / b

4.3. Evolutionary game theory

Classical game theory is based on rational behavior in interpersonal conflicting situations. If the rational assumption has been questioned when dealing with human beings, it is still less obvious that the theory can be applied when dealing with players without any intellectual capabilities.

Indeed, the paper by Maynard-Smith and Price [31] shows that a modified game theory can be applied in biology to model animal competing for limited resources such as territory or food. The idea behind their approach is that the rationality animals lack in order to carry out the process of maximizing their outcomes can be replaced by Darwinian fitness. Instead of consciously choosing strategies, animals are genetically programmed to present a determined behavior, beyond their control.

Although the classical solution concepts (e.g., Nash equilibrium) have been used extensively in many contexts, game theory has been unsuccessful in explaining, for instance, how players choose one Nash equilibrium if a game has multiple and equally plausible equilibrium points. Because of the cognitive limitations of individuals, the actual human rationality process and the global rationally model which is implied by game-theoretic solution concepts generally do not match each other. Introspective theories that attempt to explain equilibrium at the individual decision-making level by means of rationality, impose very strong informational assumptions and are widely recognized as having serious deficiencies. Assuming that each player knows all about the structure of the game, this knowledge may not be enough for him to decide how to play, for he must also predict the move of his opponents. Although the Nash equilibrium allows each player to correctly predict how his opponents will play, the understanding of this process requires an explanation of how players' predictions are formed.

More recent explanations on how players anticipate a solution are: to assume that they are able to extrapolate from what they have observed in past interactions, provide, they have played similar games [22]; that adaptive agents choose between alternatives in a ratio which matches the ratio of rewards [40]; and convergent prediction [30]. In these models, agents can pursue the goal of learning the equilibrium point, which seems a more plausible assumption. Players do not need to know explicitly how their actions influence those of their opponents. They will eventually learn that they do not play certain strategies, thus replicating the iterative dominance solution concept of the classical game theory. If players only know their own payoffs, they may asymptotically converge to a steady state represented by a set of evolutionary stable strategies (ESS) which is a Nash equilibrium.

4.4. Learning an evolutionary stable strategy

When considering a game having a unique iteratively undominated strategy s played dynamically, on the long run only s will be played. A plausible explanation is that players learn how to play their best replies, once the learning process requires strategies that are not doing well to be played less often.

While in biological applications of evolutionary game theory the ESS is genetically determined, in more general cases players can learn such strategies if they learn how to select the ESS. This can be done by analyzing the payoff obtained from each rule used to select strategies in the past. According to Harley [22], for a given player i and a set of actions $A_i = (a_1, \ldots, a_m)$, a learning rule is a rule which specifies the probabilities $P = (p_{i,1,t}, \ldots, p_{i,m,t})$ (for a given time step t) as a function of the payoffs obtained by playing those strategies in the past. He has also defined a rule for learning an ESS as the one which causes the members of a population with any initial $P = (p_{i,1,0}, \ldots, p_{i,m,0})$ to adopt the ESS of the game after a given time. He proved that such a rule must have the following property:

$$p(a_i) \geq \frac{\text{total payoff for playing } a_i}{\text{total payoff played so far}}.$$

Therefore, in the near future, each strategy is selected according to its probability. This leads the ESS to be selected asymptotically. It is assumed that the payoffs correspond to changes in fitness, and that the game is played enough times to ensure that the payoffs received after an ESS is reached exceed the payoffs received during the learning period. However, this kind of "passive learning" can lead to steady states which are not equilibria. Need of active learning, i.e., with experimentation, was emphasized in [17] where it is claimed that even if players play the same game many times, they may continue to hold incorrect beliefs about the opponents' preferences unless they perform enough experimentation. This can be done in several ways, such as considering mutations or players sometimes selecting strategies at random.

In the model proposed by Harley, on the other hand, learning rules lead to the ESS without completely fixing or deleting any possible strategy. In fact, in his model a poor strategy s_m is never completely discarded because it may become advantageous if the environment changes. This is done by weighing the recent payoffs more than the older ones, without however forgetting the whole past.

5. Decentralized coordination in networks of agents: a game-theoretic approach

This section describes the approach proposed to help in the decision-making level (traffic management). The idea is that the traffic manager transfers the operational issues to agents assigned to traffic objects such as lanes, intersections (the focus of the examples used in this work), etc. This way, the manager can concentrate on tactical issues such as what-if simulations by introducing perturbations in the traffic flow to see how the system would react.

This approach tackles the problem of complexity and cost associated with the communication in a many-agents environment with continually changing traffic patterns. The system uses a model which provides a qualitative description of the relations between traffic parameters and states. The development of this model starts with a macroscopic modeling, which reproduces the traffic state in the time–distance dimension by means of the two continuous functions, namely traffic volume (q) and traffic density (δ). This macroscopic, qualitative model is known as *fundamental diagram*.

To deduce the qualitative relation, it is assumed that vehicles in the traffic flow travel at an average speed, determined on the basis of the observed traffic density of the fundamental diagram. Qualitative density values are obtained by the relation between density and speed. The qualitative model thus contains a finite number of intervals which describe the assigned density, speed and volume values. An example can be seen in Table 2. To each qualitative description (e.g., "D-1" for the interval of the minimum density value and "STOP" for the interval of the maximum density value) there is a corresponding quantitative one (not shown here because it varies greatly from scenario to scenario) specifying the interval of density, speed, and volume. The fundamental diagram is used to generate the density calculus which combines all objects of the qualitative description and is used to simulate the traffic flow by the movement of density zones.

In short, a qualitative model is characterized by a traffic density interval, its symbolic name, and the intervals of the corresponding speeds and traffic volumes.

Since we use the concept of density intervals, the average density value on a lane or group of them is used as a measurement of performance. This is quite realistic since density represents the volume of vehicles per unit of distance. In order to compute the average density value for the lane during the simulation horizon T, the various time intervals determined by the events causing changes in the density pattern have to be considered. The average density value for lane j (in which several different densities over a length L are observed) during the time horizon T is thus computed by:

Table 2. Fundamental diagram in a qualitative description of the traffic model.

State of traffic	Qualitative density value
free flow	D-1
partial free flow	D-2 D-3
maximum flow	D-4
dense traffic	D-5 D-6 D-7
traffic jam	STOP

$$\bar{\delta}_j = \frac{\int_0^T \int_0^L \delta_j(l,t)\mathrm{d}l\,\mathrm{d}t}{T}. \tag{1}$$

Also an average density value for a set of lanes or for the whole network can be computed by simply weighing each $\bar{\delta}_j$ by each length L_j.

5.1. Coordination in evolutionary fashion

Other approaches (e.g., those cited at the end of section 2.2 and in section 3.5) have assumed traffic agents as being cooperative or benevolent. However, in the majority of the deals, real-world agents have conflicting local goals. After some attempts with protocols which consider benevolent agents, the conclusion was that a means for agents to model and represent other agents' knowledge and beliefs is necessary in order to reduce the communication required prior to reaching an agreement. Game theory was already used to provide agents a means to model and reason about other participants in a scenario in which each agent represents a lane in a single intersection [2]. The agent's goal is to negotiate the selection of a signal plan. In this case, conflicts arise because all lanes have to select the same signal plan (since only one can run in the traffic signals located at the intersection).

However, agents can be assigned to traffic elements according to other granularities. For instance, besides lanes, agents can be drivers, cars, intersections, and group of intersections. In the examples discussed in this section, agents are intersections located in a network. The central concern is how to select an equilibrium when more than one exists.

The network scenario differs significantly from that of a single intersection [2]. Moreover, it does not require the individually-motivated agents to know all about the structure of the interaction in which they are involved, when predicting the moves of their opponents. Instead, agents extrapolate from experience acquired by playing the game repeatedly in the past. If the interaction lasts long enough, then agents can asymptotically learn new equilibria. Even if the environment changes at a local level, thus demanding participants to select strategies to cope with the new situation, neighbors can learn the new behavior, provided the change rate is smaller than the time needed to learn. And finally, by a change at a global level, all neighbors have to learn the new global traffic pattern and adapt themselves to it.

The main motivation for this approach is that it has the benefit of coping with interactions in which agents are fully individually motivated and need only to know their own payoffs, but not those of the opponents, thus modeling coordination with no communication. Another positive characteristic is that this complex interactions can still be modeled as two-player games.

5.2. Description of the approach

The proposed approach departs from standard scenarios in game-theory. First, the geographical localization of each site of the network plays a role in defining which interactions happen between agents, i.e., interactions among all agents do not happen at random.

Second, learning is an important component of the approach. Therefore, the recent history of the game plays a significant role in deciding the selection of future strategies. This history is partially discarded only when a player detects a change in its environment, in which case it has to react to it in a new way. According to how it and the neighbors have reacted, a feedback is given and the process of learning restarts under a new condition. The process of updating strategies is done by computing, in each period, the probabilities of playing the strategies in next periods as a function of the previous payoffs, thus in a way similar to that proposed by Harley [22]. Further, in order to avoid the synchronization of behaviors which may arise from the use of deterministic updating of strategies, one may let players update them in a non-simultaneous way, which is more realistic.

Third, as discussed before, in the traffic scenario it is desired that communication be kept as low as possible. Therefore, as a way to reduce communication, players are not informed about the strategies selected in the neighborhood. Agents receive a reinforcement due to their actions performed in the near past, and this value is obtained from their local detectors only. However, if they are paid also according to a global goal, agents have an incentive to coordinate toward this goal, i.e., toward a joint action which allows the traffic over the whole road to flow better.

Thus the method of control is a traffic-responsive one. The proportion of time during which each stream has right of way is appropriately considered in order to achieve an equal use of all roads of the network or arterial. This aims at addressing the long run (hours), and is operationally implemented by calculating synchronized signal plans for neighboring intersections. In a short period, the system is able to respond to fluctuations in the arrival rates by changing to an alternative signal plan capable of dealing with this fluctuation.

Each agent i plays a two-person game G against each member of his neighborhood N_i in a network K. G is represented by the 3n-tuple $(1, \ldots, n, A_1, \ldots, A_n, \pi_1, \ldots, \pi_n)$, where:

- $I = (1, \ldots, n)$ is the set of players or agents, where player n is the Nature;
- $A_i = (a_{i,1}, \ldots, a_{i,k}, \ldots, a_{i,m})$ is the set of pure strategies of agent i;
- the mapping $\pi_i : X_{j \in I} A_i \rightarrow \Re$ is the payoff function of agent i;
- $\vec{P}_i = (p_{i,1}, \ldots, p_{i,k}, \ldots, p_{i,m})$ (the mixed strategy of agent i) is a probability distribution on A_i;
- $p_{i,k}$ is the probability assigned to the k-th pure strategy of agent i, with $p_{i,k} \geq 0$ and $\sum_k p_{i,k} = 1$ for each $a_{i,k} \in A_i$;
- S_i is the set of all mixed strategies of i;
- $S = X_{i \in I} S_i$ is the mixed strategy combination of G.

Each $i \in I$ updates its mixed strategy based solely on the payoff received by selecting an action. $\vec{P}_i$, the mixed strategy for agent i, is time dependent. After selecting an action $a_{i,t} \in A_i$ at time t, each i in K receives an individual payoff calculated as the sum of the payoffs obtained by playing G against each $j \in N_i$.

Before the actual beginning of the game, the types of the agents and their payoff functions are set by Nature. Her selection of payment is a metaphor to represent the stochastic dynamics of traffic flow. Nature determines the vector of probability distributions over the strategies ($\vec{P}_i$) for each site i and the payoff function for each possible combination of mixed strategy s in S. All agents are assumed to have detectors, therefore they acquire data about the local traffic conditions (i.e., about their vectors $\vec{P}_i$), and know their types. In each subsequent period, agents select a strategy with a probability which is determined by their beliefs about their environment.

A period in which $\vec{P}_i$ changes, irrespective of the site i in the network K, either by learning or by a local stochastic event, is called a learning period (for i) or an individual-state-change period (for i) respectively. Others are normal payoff-getting periods. In such periods agents only play the coordination game with their neighbors.

In learning periods a new distribution of players for the next generation is formed on the basis of their payoffs obtained in payoff-getting periods, according to some selection and reproduction rule and genetic operators. Even if agents possess only little local knowledge, which they get from sensoring their near environment, they are able to perform experimentation and, according also to the experimentation performed elsewhere in the neighborhood, they receive a reinforcement in their payment function. This feedback can be either communicated to each agent by the neighbors or by a controller of the network, or can be exclusively locally detected.

5.2.1 Individual-state-change periods. For practical purposes, changes in traffic volume are considered stochastic processes [28]. This volume may change both locally and at global level. An example of the latter is the arrival of a platoon of vehicles which shall travel the whole arterial and can be detected at the border of the network. This global alteration in traffic conditions also yields local perturbations.

An example of a local change is the arrival of vehicles coming from a crossing-street, either to cross the arterial at one intersection or to travel through a portion of it. These local stochastic events happen with probability σ_i at each site i (in an independent way), while global changes occur with probability γ.

Local events are detected at each intersection by pooling the hard-ware of the detector, which is generally installed only on the main lanes of the arterial. The flow of vehicles at each detector is compared to others. The agent at the intersection then classifies the detectors in decreasing order by the flow of vehicles and stores this information in a stack. Each time there is a change in the top position of the stack, an event at a local level is said to happen (i.e., the probability σ_i is met). Upon such an event, a change in the distribution of the mixed strategy is required. To each detector there is a unique correspondent pure strategy.

When a local event occurs at time $t = \rho$ at intersection i, agent i updates the vector $\vec{P}_i$ as a function of the flow of vehicles $q_{i,k}$ measured at each corresponding k-th detector:

$$
\begin{aligned}
&\vec{P}_{i,t} = (p_{i,1,t}, \ldots, p_{i,k,t}, \ldots, p_{i,m,t}) = [d(q_{i,1,t}), \ldots, d(q_{i,k,t}), \ldots, d(q_{i,m,t})] \\
&\text{with} \quad d(q_{i,k,t}) = q_{i,k,t} / \sum_k q_{i,k,t} \\
&\text{and, for each} \quad a_{i,k} \in A_i : \ p_{i,k,t} \geq 0 \ \text{and} \ \sum_k p_{i,k,t} = 1.
\end{aligned} \tag{2}
$$

At $t = 0$, these equations are used to set the initial vector $\vec{P}_i$. This distribution is then employed in the selection of strategies in the sub-sequent periods, until a learning period or an individual-state-change period occurs and that distribution is updated.

5.2.2 Payoff-getting periods. The global pattern of the traffic flow is known by Nature but remains unknown to the agents. It is assumed that Nature is informed about these global changes in the network, which can be detected at determined points of it. Therefore, Nature is able to set the correspondent payment functions for the agents according to the vector $\vec{P}_n$.

For the sake of example, Figure 1 schematically shows such a move for a 2×2 coordination game. For instance, if the traffic condition requires traffic signals to be synchronized through agents selecting an action s_1, agents which actually select this action are better paid. Nature changes the payment functions when a global change in traffic flow occurs.

The payment thus depends on the payoff matrix selected by Nature and on the actions selected by the agents. If they are being paid by Q_1 (where $a_1 > b_1$, $c < a_1$, $c < b_1$), and both agents select the action, say s_1 of $S_i = \{s_1, s_2\}$, then they both receive a payoff of a_1. If they select s_2 they receive a payoff of b_1. Otherwise they receive a payoff of c.

Both configurations $E_1 = (s_1, s_1)$ and $E_2 = (s_2, s_2)$ are Nash equilibria in pure strategies. A third equilibrium, in mixed strategies, is reached when strategy s_1 is played with probability $b_1/(a_1 + b_1)$ and s_2 is played with probability $a_1/(a_1 + b_1)$, in case the entire population is being paid by the payoff matrix Q_1. In the opposite case, the equilibrium is probability $b_2/(a_2 + b_2)$ on s_2 and probability $a_2/(a_2 + b_2)$ on s_1. Within time, it is expected that only the pareto-superior equilibrium be selected, i.e.

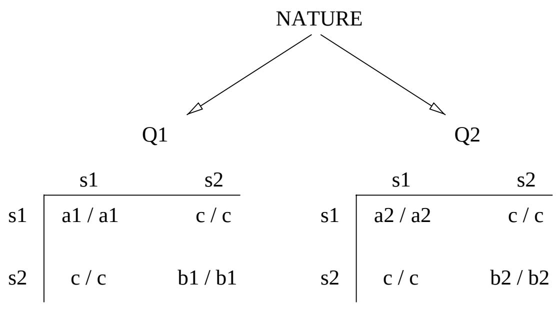

Figure 1. Payoff matrices for moves of Nature in a 2×2 coordination game.

(s_1, s_1) if Nature is paying agents according to Q_1 or (s_2, s_2) if Nature is paying them according to Q_2.

At each payoff-getting period, i plays a two-player coordination game with each of the elements $j \in N_i$, and selects a mixed strategy $\vec{P}_{i,t} = (p_{i,1}, ..., p_{i,k}, ..., p_{i,m})$ on A_i. The "raw" payoff for agent i when the set $s = (a_i, a_j)$ of actions is selected at time t is $\pi_{i,j,t}(s)$. Player i receives a summation of payoffs from these games. Hence, the payoff received by i at time t is given by:

$$\pi^{+}_{i,t}(s) = \sum_{j} \pi_{i,j,t}(s), \quad j \in N_i. \tag{3}$$

Let $\tau > 0$ be a time interval (generally τ represents the time between the last learning period and the current period), $a_{i,t} \in A_i$ be the action that agent i selects at time t, and $a_{i,k}$ each k-th pure strategy available to selection. The payoffs received in the last τ periods are represented by the vector $\vec{\pi}^{\star}_{i,\tau}$:

$$\vec{\pi}^{\star}_{i,k,\tau} = (\pi^{\star}_{i,1,\tau}, ..., \pi^{\star}_{i,k,\tau}, ..., \pi^{\star}_{i,m,\tau}), \tag{4}$$

where $1 \leq k \leq m$ is a pure strategy $a_{i,k} \in A_i$.

Finally, each element of the vector $\vec{\pi}^{\star}_{i,\tau}$ at $t \in \tau$, $t \geq 1$ is computed as follows:

$$\pi^{\star}_{i,k,\tau} = \begin{cases} \pi^{\star}_{i,k,t-1} + \pi^{+}_{i,k,t}, & \text{if } a_{i,k} = a_{i,t} \\ \pi^{\star}_{i,k,t-1}, & \text{otherwise.} \end{cases} \tag{5}$$

5.2.3 Learning periods. Depending on the frequency of the stochastic events, agents have time to learn rules about how to change strategies and are able to coordinate toward the global goal. In order for the pure strategy a_k with the highest expected value of environmental reaction to be selected, agents must learn using a selection rule. A performance function based on the payoffs obtained from the environment is used as fitness. Moreover, it is required that agents' selection rules do not assign zero probability to any pure strategy a_k. This is especially true in the scenario discussed here, since the stochastic changes in traffic flow may require an agent to respond with a given strategy, even if it has performed poor in the past and has, hence, low probability of being selected.

In the algorithm proposed, the learning rule assigns greater significance to recent than to past payoff information. To achieve this, a memory factor λ $(0 < \lambda \leq 1)$ is used in order to avoid the complete neglect of the payoffs obtained by one action in the past. At each period, the more recent payoff yielded by a given action is reduced by a factor of $(1 - \lambda)$ as shown in Equation (6).

The learning process does not occur at each period of time. Learning periods happen randomly and are determined by a learning frequency parameter f_l. At each period, there is a probability η for each agent to learn. If $t = \theta$ is a learning period for i, and $t = \rho$ $(\rho < \theta)$ is the last period in which an individual-state-change has happened before θ, then Δ is the learning interval, i.e., the time interval between the θ and ρ, for each θ. The "reduced" payoff, i.e., the payoff which also accounts for $\bar{\pi}_{i,k,\Delta}$ (the average payoff yield by the action $a_{i,k}$ during the interval Δ), then reads:

$$\pi_{i,k,t} = \lambda * \pi^{\star}_{i,k,t} + (1-\lambda) * \bar{\pi}_{i,k,\Delta}, \quad \text{for each } a_{i,k} \in A_i. \tag{6}$$

The cumulative and the average payoff after τ yield by action $a_{i,k}$ can be calculated by:

$$\pi_{i,k,\tau} = \sum_{\theta=t-\tau}^{t} \pi_{i,k,\theta}, \tag{7}$$

$$\bar{\pi}_{i,k,\tau} = \frac{\pi_{i,k,\tau}}{\tau}. \tag{8}$$

Equations (7) and (8) are also used to compute $\pi_{i,k,\tau}$ and $\bar{\pi}_{i,k,\tau}$ respectively, when $\tau = \Delta$ is the time interval as defined above.

The learning process consists of agents updating their $\vec{P}_i$ vectors according to the efficiency of every pure strategy in the past. This is a function of the fitness vector $\vec{F}_i$ defined over the elements of the set of strategies A_i and is computed locally as a function of $\pi_{i,k,\Delta}$:

$$\vec{P}_i = \vec{F}_i = (F_{i,1,\Delta}, \ldots, F_{i,k,\Delta}, \ldots, F_{i,m,\Delta}) = \left(\frac{\pi_{i,1,\Delta}}{\sum_k \pi_{i,k,\Delta}}, \ldots, \frac{\pi_{i,k,\Delta}}{\sum_k \pi_{i,k,\Delta}}, \ldots, \frac{\pi_{i,m,\Delta}}{\sum_k \pi_{i,k,\Delta}} \right),$$
$$1 \le k \le m, \quad a_{i,k} \in A_i. \tag{9}$$

This probability distribution on the pure strategies is then used in the subsequent periods until a new learning period or a change in the local environment happens.

5.2.4 Algorithm. Summarizing, the dynamics of the model is as the following algorithm:

```
begin
t := 0
set initial values agent-level parameters (σ_i, P⃗_i)
set initial values network-level parameters (γ, η, r, K, T, λ)
repeat for each i in K while not last period
    t := t + 1
    while not a global-state-change
      poll detector for a local change in traffic pattern
      while not an individual-state-change period
        while not a learning period
          perform action a_{i,t}
          collect payoff π_{i,t}
          accumulate payoff π^+_{i,t}
        compute new fitness F⃗_{i,t} and probability vector P⃗_{i,t}
        (due to learning)
    compute new probability vector P⃗_{i,t}(due to local
      change in traffic situation)
end
```

5.3. Simulation tool

In order to model the events which happen in a real traffic network, a simulation tool which shows the time evolution of the strategies used by the agents was developed [2–4, 46]. Selections made by them are independent and decentralized, that means, they happen at the individual level.

The initial parameters are set by the user. These include: the simulation horizon T, the memory factor λ, the range of interaction r, probabilities σ_i, η, γ, and the initial probability distribution Nature puts on paying agents according to Q_1 or Q_2, namely $\vec{P}_n = (p_{n,1}, p_{n,2})$. Each time a probability is met, the correspondent event happens. For instance, when σ_i at site i is met, agent i updates its $\vec{P}_i$ vector. If no probability is met, the period is a normal payoff-getting period (Section 5.2.2).

As events take place, the traffic manager sees a print-out with the main events. In case of a learning period at an intersection, the new probability distribution for this particular intersection is printed. After each repetition of the simulation, a file with the main information needed for further statistics and graphical presentation is written to a file. This file contains mean payoff, probability of selecting the strategies, and number of coordinated and miscoordinated interactions among neighbors.

5.4. Scenario

In the traffic signal coordination scenario, the goal is to bring as many neighbors in an arterial as possible to use the same signal plan since these are designed to allow vehicles to flow *in one of two opposite directions* (the reason behind this constraint is discussed in Section 3.4) through the intersections, without stopping at red lights.

Be $\mathcal{I}$ the number of agents and I the set of agents. In the scenario used as example here, the network K is an arterial composed of $\mathcal{I} = 10$ intersections (I11, I12, I13, I14, I15, I27, I31, I37, I40, and I44 in Figure 2), each being designed as an agent. However, simulations with further values of this parameter were evaluated for the sake of evaluation (see Section 6). The range of interaction among neighbors is $r = 1$. Therefore, the neighborhood $N_i = (i_{-r}, ..., i_{-k}, ..., i, ..., i_k, ..., i_r)$ of agent i is composed of the $2r$ neighbors.

Figure 3 shows one of these intersections in detail, namely the intersection I14. The main flow of vehicles are in direction west (lane-37 in section-19 towards lane-41

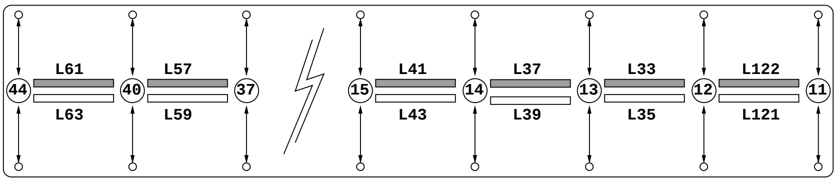

Figure 2. Arterial used in the simulations. Main intersections are I11, I12, I13, I14, I15, I27, I31, I37, I40, and I44. Only the main lanes (lane-122,..., lane-61 in direction W and lane-63,..., lane-121 in direction E) are shown; lanes in gray form the westward synchronization, lanes in white form the eastwards synchronization. Remaining lanes and nodes are secondary ones.

in section-21 of Figure 3), and in the direction east (lane-43 in section-22 towards lane-39 in section-20). Lanes in section-69, section-127, section-42, and section-38 play a minor role and interfere in the selection of signal plans only at a local level.

In order to reach a full synchronization of the signals, all agents have to select the same action from the set $A_i = (sp_W, sp_E)$ of signal plans, for each $i \in I$. The fact that agents synchronize their actions by selecting, say sp_W, only means that vehicles traveling in the direction W of the arterial are allotted more green time. However the general constraints posed by safety rules like minimum green time for each lane, minimum and maximum cycle time, etc. were respected when designing the signal plans.

Each agent at an intersection has local information acquired from detectors installed at the main lanes (lane-43 and lane-37 in the intersection shown in Figure 3). With this information, an agent i is able to detect a change in the local traffic situation. The agent then compares the detectors' data and decide the more appropriate signal plan.

Two assumptions are made: by selecting the appropriate signal plan, the local traffic condition in the intersection i improves, and by synchronizing with neighbors, the traffic condition in the neighborhood N_i also becomes better. These assumptions are quite realistic since the main lanes play the determining role in the kind of arterial considered here. By giving priority to the more congested of them, the density of vehicles is likely to decrease. If it does not, this means that the whole arterial is over-congested, in which case the synchronization of signals is not an appropriate method of control.

Further, if we consider a standard measurement used by traffic engineers, namely the total queue, a synchronization of signals in the more heavily loaded direction

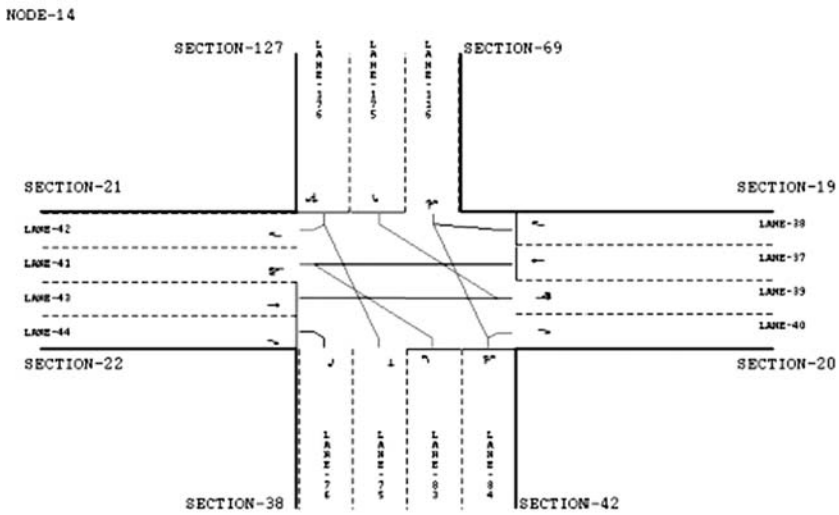

Figure 3. The intersection NODE-14 with sections and lanes.

would contribute to the decrease in the queues, once more vehicles would travel in this direction.

If two neighboring intersections have each to choose between running a signal plan which gives preference to the flow of vehicles traveling either westwards (sp_W) or eastwards (sp_E), then the payoff which both intersections get is presented in Table 1 where a_1 is sp_W and a_2 is sp_E. Besides: $a = 2$, $b = 1$, and $c = 0$ in case the global goal is to coordinate towards the direction west, or $a = 1$ and $b = 2$ in the opposite case. In this game, the intersections are better off when coordinating towards the same direction.

All stochastic events are addressed at discrete periods of time $t = 1, 2, ..., T$. Preferentially these periods are to be synchronized with the end of a cycle of the signal plan. As already discussed, at the beginning of the game the traffic flow pattern and other parameters are defined by a move of Nature. Then, agents interact according to their local information a number of times, i.e., during each period they have to decide which action to select. They do this myopically by assuming that strategies which proved to be effective in the recent past are likely to be effective in near future. Depending on the strategies played in the neighborhood, each agent obtains a payoff which is summed up.

Depending on the learning frequency, the probability distribution on the strategies for the near future ($\vec{P}_i$) is more adapted to the environment. By updating the mixed strategy according to the fitness of each pure strategy, agents are able to coordinate towards the global goal. Besides the probability of learning, at each period, agents have also a probability of experiencing a change in traffic condition at intersection level (σ_i). In this case, the vector $\vec{P}_i$ is modified according to Equation 2 and this distribution is employed in the selection of strategies, until a further learning period occurs and the distribution is updated.

By allowing $\vec{P}_i$ also to be updated according to the detectors values, the equilibrium may shift, once a local perturbation may cause a perturbation in the neighborhood and one equilibrium point may displace the other. The more often the local stochastic events, the more unstable the system. Usually, the frequency of a global change (γ) is low: one in a few hours. When it happens, agents must re-coordinate and reach the corresponding equilibrium in a short time.

5.5. *Measure of performance in the evolutionary approach*

As the algorithm for coordination presented here aims at leading as many agents as possible to coordinate toward a joint action, the performance of the algorithm is measured both by the number of agents reaching coordination and by the time needed to accomplish this.

Different forms of representing these quantities were tried. Initially, the average payoff received by the agents were recorded against time. When simulating an arterial K composed of $\mathcal{I}$ intersections, each choosing between k pure strategies, the average payoff received by each agent i at time interval t reads:

$$\bar{\pi} = \frac{\sum_t \sum_k \sum_i \pi_{i,k,t}}{\mathcal{I} * \tau}, \quad t \in 1, ..., \tau \;\; i \in I, \;\; a_{i,k} \in A. \tag{10}$$

Such average payoff gives a measure of the global performance of the network during the simulation interval τ.

All simulations discussed in this section assume payoffs as introduced in the last subsection. These values are commonly used in the literature of coordination games to represent the relative preferences of the agents regarding the possible actions. Theoretically, when all agents select sp_W with probability 1, the maximal average payoff possible is equal to 3.6 (remember that the two agents at the borders can receive at most 2 points while the other eight can receive 4 points each).

Although this gives an idea of how good the agents have performed during the simulation time, the only possible conclusion is whether or not a stable situation was reached on the long run, i.e., whether agents held on to one action, except when a local change happened. However, an average payoff of, say 2.8, says little about which agent miscoordinated, for how long, and where the miscoordination happened since this average value can be reached by several different configurations of the network.

Therefore, further forms of representation are needed, e.g., the accounting of the type of interactions occurring between all pairs of neighbors. These can be of three types: WW (both selecting sp_W,), EE (both selecting sp_E), and miscoordinated (WE + EW) (i.e., they select either sp_W,sp_E or sp_E,sp_W). Thus, a qualitative measure of the number of agents reaching coordination can be estimated. However, it says little about stability. When should one consider that coordination on WW or EE is reached?

The assumption that agents continually maximize their expected payoffs is quite strong for infinitely repeated games. While this assumption is common in game theory and economics, the solution of such a maximization problem in infinitely repeated games may be very demanding. For instance it is not realistic to expect that all pairs of agents coordinate all the time, since this would mean that the local conditions are not addressed when they are in conflict with the global goal.

Thus, regarding the time needed to coordinate, the number of interactions of type EW + WE seems to be only an approximate measure of the performance of agents. A much more interesting measure of the performance is the evolution of the average probability agents place on selecting the action which is better paid (determined by Nature). One expects that all agents place increasingly higher probability on the more profitable action, and that on the long run they select this action with probability close to one. How close the probability is to 1 depends on both the memory factor λ, and the frequency of individual changes in traffic state, since this requires that one or more agents respond to these changes in a random way.

While the plot of that probability as a function of time qualitatively characterizes the performance of agents, there is still a need to evaluate these curves as to what regards the time needed to reach a given pattern of convergence. This is done by reading, in the plot, the time needed to reach the probability Nature determines as the most profitable action (let us say sp_W).

For example, suppose the simulation was carried out setting $p_{n,W} = 0.9$; agents are expected to be able to put at least an average probability of $p_{i,W} = p_{n,W} = 0.9$ on their selection of sp_W, in order for the situation simulated to be considered a good one. To expect each agent selecting sp_W with probability $p_{i,W} = 0.9$ means, for 10 agents, that at each period, nine agents on average select the signal plan sp_W. Only one (on average) is expected to select the signal plan sp_E, thus miscoordinating.

6. Results

6.1. Parameterization

In this section the effect of the selection of a strategy in the simulator is verified using the measurement of performance discussed in section 5.5, and also according to the utility values as in Table 1.

Except when explicitly stated, the simulations are carried out under the following conditions: the probability distribution $\vec{P}_n$ on the payoff matrices (Q_W, Q_E), i.e., the probability distribution Nature puts on paying agents according to Q_W, and Q_E is $\vec{P}_n = (0.9, 0.1)$. In Q_W, $a = 2$, $b = 1$, and $c = 0$. In Q_E, $a = 1$, $b = 2$, and $c = 0$. Consequently, the pareto-efficient equilibrium (sp_W, sp_W) is expected to be selected.

Henceforward, the element $p_{n,W}$ of $\vec{P}_n$ will be referred simply as p_n, the element $p_{i,W}$ of $\vec{P}_i$ simply as p_W. We have performed statistical tests to determine the necessary number of repetitions of the simulations. Therefore, the experiments shown next were repeated at least 30 times, which accounts for 95% of significance.

In order to take a first look at the ideal range of certain parameters for further experiments, several simulations varying only one parameter at a time were carried out. Initially, several values for both the frequency of learning f_l and the memory factor λ have been separately tested. The parameter f_i (frequency of a local change in traffic condition) has been set to infinity, such that no local change in traffic flow happens: the more profitable strategy is assumed to be sp_W unless otherwise stated.

The evolution of the probability which the agents assign to select this strategy was studied, for $f_l =$ 3, 5, 8, 10, 15, 20, 25, and $\lambda = 0.95$. The probability of learning indicates that at each period, agents have one chance in f_l of being in a learning period, that means the frequency of learning is $\eta = 1/f_l$.

The main conclusion from these experiments is that, in the absence of individual changes in traffic conditions, agents have no incentive to deviate from the more profitable strategy. The system thus reaches an ESS as discussed in Section 4.4. However, it is also important to look at the *time needed for agents to reach the ESS*, i.e., the time needed for the probability of playing the more profitable strategy to reach the desired value of $p_w = p_n = 0.9$. This is shown in Table 3.

Another set of experiments was carried out setting $\lambda = 0.80$ and $f_l = 5, 10, 15, 25$. The comparative results concerning time are shown also in Table 3. In general, as expected, the lower the memory factor, the higher the time needed for agents to select sp_W with probability $p_W = 0.9$. This happens because the lower λ, the more the weight of past payoffs, and hence the higher the inertia.

Further, the lower the f_l (hence the higher the η), the faster the value $p_W = 0.9$ is reached. However, there is a limit for this rule. When f_l is too low, the wrong strategy (not evolutionary stable) may be learned. This happens because an agent may not have time to acquire information before it learns. Once the wrong pattern is learned,

Table 3. Time needed to reach $pw = 0.9$, varying f_l and λ.

Memory factor λ	Learning frequency f_l ($\eta = 1/f_l$)	Time needed to reach $p_w = 0.9$ (periods)
0.95	3 (0.33)	8
	5 (0.2)	12
	8 (0.125)	12
	10 (0.1)	20
	15 (0.067)	24
	20 (0.05)	30
	25 (0.04)	36
0.80	5 (0.2)	10
	10 (0.1)	20
	15 (0.067)	33
	20 (0.05)	36

without the interference of the traffic manager, it may take time to realize the mistake and converge to the right strategy.

Analyzing these simulations, one can conclude that a good range for further experiments lies between $f_l = 5$ and $f_l = 10$, since for higher values of f_l it takes time to reach the goal and for $f_l = 3$ the standard deviation is high, which means that in many of the runs, the population as a whole failed to reach a stable state.

Finally, further simulations were carried out varying the number of agents $\mathcal{I}$ (intersections). This parameter was also set to 35 in order to verify the scaling of the approach. The results do not differ significantly from those achieved by setting $\mathcal{I}$ to 10. This was a foreseeable result since the interactions happen in a defined neighborhood and in this case the total number of agents does not influence the performance of the algorithm. Table 4 depicts the comparative results.

6.2. *Experiment A: global payment*

In this experiment the focus is on the variation of the individual frequency f_i, once this verifies the robustness of the agent-based approach. The immediate interest here is to test several frequencies (f_i) and check whether agents learn how to coordinate and reach the global goal.

The probability σ_i means that at each period there is one chance in f_i for an individual change in traffic condition at each intersection. These probabilities are independent from site to site. Simulations are thus performed setting f_i to 10, 20, 50, 100, 200, and 300 and the learning frequency f_l to 5 and 10. Time (t_c) needed for the population of agents to reach $p_W = 0.9$ is measured for each condition and also depicted in Table 4 (third column).

As expected, a good pattern of coordination is reached faster in the most stable environment, i.e., with $f_i = 300$. However, the learning rate plays also a role, since, in general, it takes less time for a population with $f_l = 5$ and lower f_i (e.g., $f_i = 200$) to reach this pattern than for a population with higher f_i (e.g., $f_i = 300$) and $f_l = 10$.

Apart from the intuitive rule that coordination must be reached as soon as possible, knowledge about the superior limit of t_c is also interesting. When should it be considered not acceptable anymore? This is a function of the frequency with which

Table 4. Time needed to reach $p_W = 0.9$, varying *I*, f_l, f_i, and for $\lambda = 0.95$.

Learning frequency $f_l(\eta = 1/f_l)$	Frequency of a local change $f_i(\sigma_i = 1/f_i)$	Time needed to reach $p_W = 0.9$ (periods)	
		$\mathcal{I} = 10$	$\mathcal{I} = 35$
5 (0.2)	10 (0.1)	14	–
	20 (0.05)	19	13
	50 (0.02)	12	14
	100 (0.01)	12	15
	200 (0.005)	9	12
	300 (0.003)	8	10
10 (0.1)	20 (0.05)	22	–
	50 (0.02)	19	24
	100 (0.01)	18	17
	200 (0.005)	13	18
	300 (0.003)	10	15

the environment changes (not on an individual but on a global level). For the characteristics of the volume of vehicles in the arterial in question, in the time period considered, a global change, i.e., a change in the direction of the green wave occurs once every 90 cycles. As an arbitrary, but quite reasonable rule, one can think that a pattern of coordination reached in at most half this time can be considered satisfactory. This means 45 cycles in the case under consideration. As the results in the third column of Table 4 show this time was reached in all experiments.

In order to analyze the influence of the memory factor, similar experiments were done with $\lambda = 0.80$. The t_c's are then shown in Table 5.

Analyzing these results, one sees that the memory factor also has an effect on the time needed to reach a given pattern of coordination. In the case in which the population is expected to perform poorly (for $\lambda = 1.00$, $f_l = 10$, and $f_i = 20$), by setting $\lambda = 0.80$ the population even fails to reach $p_W = 0.9$ within the simulation time (50 periods). For an intermediate situation with $f_l = 10$ and $f_i = 100$, t_c differs only slightly under the three memory factors. And finally, in the situation expected to

Table 5. Time (t_c) needed to reach $p_W = 0.9$. Comparison of some cases for $\lambda = 0.80$, $\lambda = 0.95$, and $\lambda = 1.00$. ($\gg$ means that the pattern $p_W = 0.9$ was not reached during the simulation time).

f_l	f_i	λ	Time (periods)
10	20	0.80	$\gg$
		0.95	22
		1.00	28
10	100	0.80	21
		0.95	18
		1.00	18
5	200	0.80	10
		0.95	10
		1.00	10

be the best of those compared, namely with $f_l = 5$ and $f_i = 200$, results were the same. Here there is room for employing an even lower memory factor if necessary.

6.3. *Experiment B: Nature pays according to local states*

In the experiments discussed in the previous section, Nature pays all agents according to her knowledge of the global goal. Two neighbor agents selecting the same strategy, which fits the global goal, receive each the highest possible payoff. Selecting identical strategies, but not the one compatible with the global goal brings them a lower payoff, while not being able to select an equal strategy pays them nothing.

In this section the assumption that Nature pays an agent always according to the global goal is loosened. Agents are paid differently and according to their individual states, despite the global goal. For instance, even if the global goal is to coordinate using sp_W, if one of the agents detects a need to run sp_E, it receives from Nature the highest payoff, *if the neighbor also selects* sp_E. This gets also the highest payoff in case its state requires running sp_E, but only a lower payoff in case its state does not require sp_E. Further, both receive nothing if they select different strategies. The main interest is to find out how the population of agents select an equilibrium, if any. If they are able to select one, for which values of f_i, f_l, and λ does this happen? How long does it take?

A set of experiments were initially conducted, as usual, for $\lambda = 0.95$, varying f_l and f_i (Table 6, fourth column). In three cases the whole population reaches the desired pattern of coordination (selecting sp_W with probability above $p_W = 0.9$), namely for $f_l = 5$ and $f_i = 300$, for $f_l = 10$ and $f_i = 300$, and for $f_l = 5$ and $f_i = 200$. The t_c needed in each case are 12, 22, and 34, respectively. In extremely unstable scenarios, with $f_i = 20$, $f_i = 50$, and $f_i = 100$ the populations performed bad.

The experiment was performed for few other cases, for values of $\lambda = 0.80$ and $\lambda = 1.00$. The comparison is shown in Table 6. Under these conditions, the desired pattern of coordination is not reached for $f_i = 20$ and for $f_i = 50$, irrespective of λ and f_l. For $f_i = 100$, coordination is reached fast only when the memory factor is $\lambda = 1.00$ and $f_l = 5$. In general, low λ leads to longer time.

However, it is questionable whether it is still reasonable to expect agents reaching probability $p_W = 0.9$ since Nature actually does not pay all of them all the time according to the matrix Q_W with a probability $p_n = 0.9$.

6.4. *Experiment C: with communication between neighbors*

In this series of experiments, an agent (henceforward called changing-agent) is allowed to communicate with the immediate neighbors and inform them about its intention to change strategy. There is no further negotiation process. Neighbors can at this point decide on the gain by changing strategy if this is the case. After receiving such a message, an agent must decide whether or not to change the vector of probability distribution $\vec{P}_i$. If, after checking its traffic condition (by means of the detectors), the agent verifies that its traffic condition is compatible with the strategy proposed by the changing-agent, i.e., they both have similar traffic patterns, a change may be more profitable once agents are always better off

Table 6. Time (t_c) needed to reach $p_W = 0.9$. Experiment B. Comparison of some cases for $\lambda = 0.80$, $\lambda = 0.95$, and $\lambda = 1.00$ ($\gg$ means that the pattern $p_W = 0.9$ was not reached during the simulation time.

f_l	f_i	$\lambda = 0.80$	$\lambda = 0.95$	$\lambda = 1.00$
5	50	$\gg$	$\gg$	$\gg$
5	100	–	$\gg$	8
5	200	46	34	–
5	300	–	12	–
10	20	$\gg$	$\gg$	$\gg$
10	100	$\gg$	$\gg$	$\gg$
10	200	36	$\gg$	20
10	300	–	22	–

coordinating than selecting different strategies. Therefore, the agent which has received the message updates its vector $\vec{P}_i$ according to Equation (2), as in an individual-state-change period.

The goal of this series of experiments is to compare both situations, with and without communication (experiment A), as to what concerns performance, i.e., time needed to reach a given pattern of coordination.

In general, results are worse than those obtained for experiment A. Similar to experiment B, in the very unstable scenario, the performance was poor. And even in the cases where a probability $p_W = 0.9$ was reached, it takes more time than in experiment A. The expected pattern of coordination for $\lambda = 0.95$ was reached at time $t_c = 8$ for $f_l = 5$ and $f_i = 300$ (same as in experiment A); at $t_c = 12$ for $f_l = 5$ and $f_i = 200$ (9 in experiment A); at $t_c = 20$ for $f_l = 10$ and $f_i = 300$ (10); and at $t_c = 24$ for $f_l = 10$ and $f_i = 200$ (13).

These results are in accordance with the comparison made in [41], in which the authors find that a naive use of communication may even harm the efficiency of the system. In the scenario discussed in the present experiment, the poorer quality of the results can be explained by the fact that at each time a change in strategy happens, it introduces a perturbation in the stationary state. This occurs not only in the site where the change happens, but also in the neighborhood if neighbors are willing to change strategy. Consequently, the rate of individual change is not a function of the frequency f_i alone, but it is also a function of the rate with which agents accept changing strategies due to a mutation in the neighborhood, i.e., this rate is higher than that of experiment A.

6.5. *Comparison to a centralized method of control*

To compare the performance of different philosophies of traffic signal control is not an easy task, especially if one thinks that the measurements of the performance may vary enormously from simulator to simulator. Some abstractions and assumptions made require real data to be tackled in a simplified way. And finally, the literature reports almost no experiment with real data. The simulations described in [46] albeit aiming at the validation of the qualitative model, were done without such a comparison.

To validate the approach presented in this paper, simulations were carried out within a macroscopic simulator. The arterial represented in Figure 2 was mapped to test the synchronization of traffic signals.

Three scenarios were examined, comparing the agent-based approach with a traditional algorithm for synchronization of signals. In the first scenario, one of the direction of synchronization receives a clearly higher volume of vehicles than the other direction. In the second and third scenarios both directions receive a high and a medium-to-low volume of vehicles respectively.

In the traditional algorithm for synchronization of signals, the direction of synchronization is determined by a central computer, and changes only when a determined pattern of traffic is reached on the main detectors of the arterial. For instance, if the synchronization is in the direction west (W), vehicles traveling on the lanes indicated in gray in the Figure 2 are expected to cross the arterial without stopping. If the synchronization is in the direction east (E), the priority is given to those traveling on the lanes indicated in white in the same figure. For the sake of illustration, in the scenarios discussed next, volume is higher in direction W. The central computer thus initially determines the synchronization of *all signals* in this direction.

6.5.1 Scenario I. In this situation, the direction W receives a clearly higher volume of vehicles than the direction E. The flow measured by the detector at lane-122 (in direction W) is $q_W = 0.3$ veh/s (1080 veh/h), while that measured by detector at lane-63 is $q_E = 0.045$ (162 veh/h). These flows correspond to density intervals D-4 and D-2 in the fundamental diagram respectively (Table 2).
This is a typical situation where a central controlled progression performs good since the volume of vehicles in one direction is always higher than in the opposite one. An adaptive control does not bring a gain. This can be seen in the results of the simulation (Table 7). Comparing the average density intervals for the W-lanes obtained with the central progression (third column) and those with the agent-based approach (seventh column), one sees that the density intervals for the former method are the same or an interval lower. For instance, on intersection-12 the average density on lane-122, which is D-4 for the central progression, is higher with the agent-based approach (D-5).

As the volume of traffic in the E direction is low, it is not likely to interfere with the performance of the synchronization, neither in the central method, nor in the agent-based approach. This happens because during the simulation, the direction E never demands priority.

6.5.2 Scenario II. Next, a situation in which both directions, W and E, receive a medium to high volume of vehicles is simulated. The flow measured by the detector at lane-122 is $q_W = 0.3$ veh/s (1080 veh/h), while that measured by detector at lane-63 is $q_E = 0.09$ (324 veh/h). Both correspond to density intervals D-4.

Since both directions now present heavy traffic, the intersections have to cope with a competition for the synchronization in one or other direction. This reason alone is enough to justify a more flexible, adaptive method of control. However, the agent-based approach is superior not only because it can cope with local changes, but also because in this way the overall capacity of the arterial is increased.

Table 7. Comparison of the agent-based approach and the centralized method of control. Scenario I: direction W presents a clearly higher volume of vehicles than direction E.

	Central synchronization				Agent-based approach			
Inter-section	Lane (W)	Avg. dens.	Lane (E)	Avg. dens.	Lane (W)	Avg. dens.	Lane (E)	Avg. dens.
12	lane-122	**D-4**	lane-35	D-3	lane-122	D-5	lane-35	D-3
13	lane-33	**D-3**	lane-39	**D-2**	lane-33	D-5	lane-39	D-3
14	lane-37	**D-4**	lane-43	D-4	lane-37	D-6	lane-43	**D-3**
15	lane-41	**D-5**	lane-47	D-4	lane-41	D-6	lane-47	D-4
27	lane-45	D-7	lane-51	D-4	lane-45	D-7	lane-51	D-4
31	lane-49	**D-7**	lane-55	D-4	lane-49	D-8	lane-55	D-4
37	lane-53	D-7	lane-59	D-3	lane-53	D-7	lane-59	D-3
40	lane-57	D-7	lane-63	**D-2**	lane-57	D-7	lane-63	D-3

The individual average density intervals on intersections is shown in Table 8. Considering the W-lanes, the performance using the agent-based approach is equal or one interval of density better than the performance of the central-controlled progression. When comparing the lanes of the direction E, the results are significantly better. This happens because the central progression fails to give priority to this direction at the intersection level, since this means the dissolution of the synchronization. The agent-based approach, on the other hand, allows agents at intersections to break with the synchronization if necessary. However, this miscoordination takes place for short periods of time once selecting a different strategy generally is not worth, i.e., it is not evolutionary stable.

6.5.3 Scenario III. In this situation, both opposite directions have medium to low volumes of vehicles. The flow measured by the detector on lane-122 is D-1, while that measured by detector on lane-63 is D-2. In this case the agent-based approach also performs better than the central progression, although only slightly as it can be seen in Table 9. This is explained by the fact that, as flow of vehicles is relatively low in

Table 8. Comparison of the agent-based approach and the centralized method of control. Scenario II: volume of traffic is high in both opposing directions.

	Central synchronization				Agent-based algorithm			
Inter-section	Lane (W)	Avg. dens.	Lane (E)	Avg. dens.	Lane (W)	Avg. dens.	Lane (E)	Avg. dens.
12	lane-122	D-4	lane-35	D-5	lane-122	D-4	lane-35	**D-4**
13	lane-33	D-3	lane-39	D-7	lane-33	D-3	lane-39	**D-6**
14	lane-37	D-4	lane-43	D-6	lane-37	D-4	lane-43	D-6
15	lane-41	D-5	lane-47	D-5	lane-41	D-5	lane-47	**D-4**
27	lane-45	D-7	lane-51	D-4	lane-45	**D-6**	lane-51	D-4
31	lane-49	D-6	lane-55	D-5	lane-49	**D-5**	lane-55	**D-4**
37	lane-53	D-7	lane-59	D-4	lane-53	D-7	lane-59	**D-3**
40	lane-57	D-7	lane-63	D-2	lane-57	**D-5**	lane-63	D-2

both directions, traffic flows relatively free through the arterial and there are few local changes in traffic state.

6.6. *Summary of the results*

Several situations were discussed: Nature paying agents according to her knowledge of the global goal (experiment A – Section 6.2), this payment being a function of the local traffic state (experiment B – Section 6.3), and agents being able to communicate with neighbors (experiment C – Section 6.4). Further, in each experiment several values for the three parameters (f_l, f_i, λ) have been tested.

In experiment A, a stationary state was reached in the majority of the situations simulated. Also the time needed for an acceptable coordination pattern to be reached can be considered satisfactory. Results for experiment B proved that, when agents are paid according to their local states, the time needed for the same pattern of coordination to be reached is higher. This was expected since agents have more incentive to miscoordinate. Also the results for the experiment C have proved that the time needed to reach coordination is higher when communication among neighbors is allowed.

All three experiments have contributed to a deeper understanding of the emergence of an ESS in the presence of stochastic shocks. The high frequency of the stochastic shocks have immediate consequences here. Once an agent perceives a local change in its traffic state and changes strategy, it does not have reasons to believe that the neighbors will continue doing what they have done so far, because they all will get different payoffs from those they got in the past. This way, they behave like new agents which have solely the knowledge about their local states and about the efficiency of certain strategies in the past. Therefore, they must re-adapt to the new situation. Depending on the number of agents in this condition and of course if they interact (i.e., they are in the same neighborhood), the equilibrium point will eventually change.

In general, for one agent to coordinate towards the opposite strategy it is necessary that it be surrounded by neighbors already running this strategy most

Table 9. Comparison of the agent-based approach and the centralized method of control. Scenario III: volume of traffic is medium to low in both directions.

	Central synchronization				Agent-based algorithm			
Inter-section	Lane (W)	Avg. dens.	Lane (E)	Avg. dens.	Lane (W)	Avg. dens.	Lane (E)	Avg. dens.
12	lane-122	D-1	lane-35	D-4	lane-122	D-1	lane-35	D-4
13	lane-33	D-2	lane-39	D-6	lane-33	D-2	lane-39	**D-4**
14	lane-37	D-3	lane-43	D-5	lane-37	D-3	lane-43	**D-4**
15	lane-41	D-4	lane-47	D-4	lane-41	D-4	lane-47	D-4
27	lane-45	D-4	lane-51	D-4	lane-45	D-4	lane-51	D-4
31	lane-49	D-5	lane-55	D-4	lane-49	**D-4**	lane-55	D-4
37	lane-53	D-6	lane-59	D-3	lane-53	D-6	lane-59	D-3
40	lane-57	D-6	lane-63	D-2	lane-57	D-6	lane-63	D-2

of the time. And of course, once each agent has learned to play a strategy with probability near one, the only way for it to experiment new strategies is through a change in the traffic states in its neighborhood, since in this case it should react to that change. If such change does not happen, agents will have no further chance to coordinate with neighbors (which have already learned to play a different strategy).

Regarding the comparison to a central approach discussed in Section 6.5, the aim was to validate the proposed agent-based mechanism by means of the macroscopic simulator, where the control of signals using this algorithm can be compared with a centralized form of control (using a central synchronization), for the same conditions of traffic volume.

Results showed that a central synchronization performs better in stable scenarios *with the flow of vehicles being clearly higher in one direction than in the opposite* since few or no conflict occurs (scenario I). However, in scenarios where the volume of vehicles are nearly equal (scenarios II and III), the central progression does not perform well relative to the agent-based mechanism. This can be explained by the fact that the agent-based mechanism is adaptive and allows agents to break with the synchronization in order to cope with their local traffic conditions for a short time period, if necessary. By doing this, the agent-based mechanism has proved more efficient when comparing the values of the average densities at each lane of the arterial.

7. Conclusion and extensions of the work

The main motivation for the present work is the potential MAS approaches and agent-based, decentralized techniques offer as to what regards decentralized philosophies of control.

Centralized approaches to traffic signal control cannot cope with the increasing complexity of urban traffic networks. A trend towards decentralized policies of control was already pointed out by traffic experts in the 1980s [42]. Following this idea several implementations appeared [4, 6, 11, 13, 19, 24, 29, 33, 46]. However, only two [4, 33] succeed in providing mechanisms for conflict resolution (e.g., setting synchronized signal plans for the intersections in an arterial) other than those based on central coordination.

The present work addressed two problems. The first is the development of a framework to deal with the problem of maintaining the synchronization of traffic signals in a decentralized fashion and to permit the emergence of cooperation among individually-motivated agents in dynamic environment under communication constraints. The second is the use of the framework as a support for the traffic manager so that s/he can employ her/his time on more tactical issues like simulation of particular scenarios and decide which control policy to use: centralized or agent-based.

In the proposed framework, agents learn a behavior by observing interactions they are involved in. This is predominantly based on ideas and techniques of evolutionary game theory. It assumes that the expectations of agents concerning these intentions converge to a given pattern upon receiving a feedback from the environment.

Therefore, this approach explains, based on experimental learning, how a stationary stable equilibrium emerges.

Agents are involved in a long run interaction and, according to the actions of the neighbors, they get a feedback from the environment. Since this feedback usually regards the global traffic condition on the arterial (a state which is known only by the environment), every agent has an incentive to coordinate towards the strategy which permits the traffic to flow better over the whole arterial. Agents have knowledge only to a local extent. By gathering data from traffic detectors, they estimate the probability distribution over the set of signal plans available to their choice. This distribution is continuously modified according to the feedback received after selecting an action. This form of experimental learning permits agents to adapt to the environment (other agents included).

Coordination is reached in almost all conditions simulated, and within a time dependent upon the stability of the environment. The more frequent the changes in local traffic conditions, the slower the pattern of coordination is reached. The concept proposed in the present work may not perform well when the following combination of factors occurs: the environment is excessively unstable, participants of the system cannot learn with a periodicity compatible with the rate of these changes, and their excessive memory about past actions and payoffs leads them to respond with high inertia.

Concerning the operational level, where participants of the system cannot afford spending too much time on negotiations, a gain may be obtained by simply improving the quality of the communication. Instead of the naive form employed, participants should be able to distinguish between durable changes in their states (which should be communicated), and ephemeral changes (whose communication should be avoided).

References

1. R. Allsop, and J. Charlesworth “Traffic in a signal controller network: an example of different signal timings inducing different routing,”. *Traffic Eng. and Control*, vol. 18, pp. 262–264, 1977.
2. A. L. C. Bazzan, “A game-theoretic approach to distributed control of traffic signals,”. in *Proceedings of the First Int. Conf. on Multi–Agent Systems (ICMAS)* San Francisco. extended abstract, 1995.
3. A. L. C. Bazzan, “Evolution of coordination as a metaphor for learning in multi–agent systems,” in G. Weiss (ed.), *DAI Meets Machine Learning*, No. 1221 in LNAI. Springer–Verlag: Berlin Heidelberg, New York, pp. 117–136, 1997a.
4. A. L. C. Bazzan, “An evolutionary game–theoretic approach for coordination of traffic signal agents,” Ph.D. thesis, University of Karlsruhe, 1997b.
5. A. L. C. Bazzan, J. Wahle, and F. Klügl, “Agents in traffic modelling – from Reactive to Social Behaviour,” in *Advances in Artificial Intelligence, Berlin/Heidelberg. Also appeared as* Proc. of the U.K. Special Interest Group on Multi-Agent Systems (UKMAS), Bristol, UK, 1999.
6. F. Busch, and G. Kruse, “MOTION – Ein neues Verfahren für die städtische Lichtsignalsteurung und seine Erprobung in Rahmen des EG–Programms ATT,” in *Heureka 93*, pp. 79–93, 1993
7. S. Cammarata, D. McArthur, and R. Steeb, “Strategies of cooperation in distributed problem solving,” in *Int. Joint Conf. on Artificial Intelligence*. Karlsruhe, F. R. of Germany, 1983, pp. 767–770.
8. E. C. P., Chang, J. C. K. Lei, and C. J. Messer, “Arterial signal timing optimization using PASSER–II.” Technical Report 467, Texas Transportation Institute, 1988.

9. S. L. Cohen, and J. D. C. Little, "The MAXBAND program for arterial timing plans," *Public Roads*, vol. 46, no. 2, pp. 61–65, 1982.
10. J. Cuena, G. Ambrosino, and M. Boero, "A general knowledge–based architecture for traffic control: the KITS approach," in *Int. Conf. on Artificial Intelligence Applications in Transportation Engineering,*. California, 1992.
11. M. Di Taranto, "UTOPIA, in *Proceedings of the IFAC-IFIP-IFORS Conference on Control, Computers, Communication in Transportation*. Paris, 1989.
12. S. Elhadouaj, A. Drogoul, and S. Espié, "How to combine reactivity and anticipation: the case of conflicts resolution in a simulated road traffic," in *Proceedings of the Multiagent Based Simulation (MABS),* 2000.
13. N. Findler, and J. Stapp, "Distributed approach to optimized control of street traffic signals," *Journal of Transportation Engineering*, vol. 118, no. 1, pp. 99–110, 1992.
14. K. Fischer, N. Kuhn, H. Müller, J. Müller, M. Pischel, and A. Schroth, "Verteiltes Problemlösen in Transportwesen," *Information Management*, vol. 2, pp. 32–40, 1993.
15. K. Fischer, J. Müller, M. Pischeland, and D. Schier, "A model for cooperative transportation scheduling," in V. Lesser (ed.), *Proceedings of the First International Conference on Multi–Agent Systems (ICMAS)*,1995, pp. 109–116.
16. B. Foraste, and G. Scemana, "Surveillance and congested traffic control in Paris by expert system," in *IEEE Int. Conf. on Road Traffic Control,* London, pp. 91–94, 1986.
17. D. Fudenberg, and D. Levine, "Steady state learning and nash equilibrium," *Econometrica*, vol. 61, pp. 547–573, 1993.
18. A. Garland, and R. Alterman, 2004, "Autonomous agents that learn to better coordinate," *Autonomous Agents and Multi-Agent Systems*, vol. 8, no. 3, 267–301, 2004.
19. N. Gartner, "OPAC – A demand–responsive strategy for traffic signal control," *Transportation Research Record*, vol. 906, pp. 75–81, 1983.
20. M. R., Genesereth, M. L. Ginsberg, and J. S. Rosenschein, "Cooperation without communication," in *Proceedings of the Fifth Nat. Conf. on Art. Intelligence,* 1986.
21. J. Halpern, and Y. Moses, "Knowledge and common knowledge in a distributed environment," in *Third Annual ACM Symposium on Principles of Distributed Computing,* Vancouver, 1984.
22. C. B. Harley, "Learning the evolutionary stable strategy," *Journal of Theoretical Biology*, vol. 89, pp. 611–631, 1981.
23. H. Haugeneder and D. Steiner, "MECCA/UTS: A multi-agent scenario for cooperation in urban traffic'. in *Proceeedings of the Special Interest Group on Cooperating Knowledge Based Systems,* 1993.
24. J. Henry, J. Farges, and J. Tuffal, "The PRODYN real time traffic algorithm," in *Proceedings of the Int. Fed. of Aut. Control (IFAC) Conf,* Baden-Baden, 1983.
25. F. Klügl, and A. L. C. Bazzan, "Simulation of adaptive agents learning heuristics for route choice in a commuter scenario," in *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, vol. 1. Bologna, Italy, extended abstract, pp. 217–218, 2002.
26. F. Klügl, and A. L. C. Bazzan, 2004, "Route decision behaviour in a commuting scenario," *Journal of Artificial Societies and Social Simulation*, vol. 7, no. 1, http://jasss.soc.surrey.ac.uk/7/1/1.html.
27. F. Klügl, A. L. C. Bazzan, and J. Wahle, "Selection of information types based on personal utility - a testbed for traffic information markets," in *Proceedings of the Second International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, Melbourne, Australia, 2003, pp. 377–384.
28. W. Leutzbach, *Introduction to the Theory of Traffic Flow*. Springer, Berlin, 1988.
29. G. List, and J. Pond, "A strategy for real-time control demand-sensitive signal networks," in *Advanced Air Transportation Technologies (AATT) Conference,* 1989.
30. M. W. Macy, and A. Flache, "Learning dynamics in social dilemmas," in *Proceedings of the National Academy of Sciences,* Vol. 99, pp. 7229–7236, 2002.
31. J. Maynard-Smith, and G. R. Price, "The logic of animal conflict," *Nature*, vol. 246, pp. 15–18, 1973.
32. J. Morgan, and J. Little, "Synchronizing traffic signals for maximal bandwidth," *Operations Research*, vol. 12, pp. 897–912, 1964.

33. S. Ossowski, *Co-ordination in artifical agent societies, social structures and its implications for autonomous problem-solving agents*., No. 1535 in Lecture Notes in Artificial Intelligence, Berlin Heidelberg New York: Springer, 1999.
34. P. Paruchuri, A. R. Pullalarevu, and K. Karlapalem, "Multi agent simulation of unorganized traffic," in *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, Bologna, Italy, 2002, vol. 1. pp. 176–183.
35. Robertson, "TRANSYT: A traffic network study tool," Rep. LR 253, Road Res. Lab., London, 1969.
36. J. Rosenschein, and M. Genesereth, "Deals among rational agents," in *Proceedings Nineth Int. Joint Conf. on Art. Intelligence*,. 1985, pp. 91–99.
37. R. Rossetti, R. Liu, D. Van Vliet, S. Bampi, and H. Cybis, "Applying an agent–based approach to traffic modeling," in *Workshop on Agents in Traffic and Transportation (at Autonomous Agents)*, Barcelona, 2000a.
38. R. Rossetti, R. Liu, D. Van Vliet, and H. Cybis, "An agent–based framework for the assessment of drivers' decision–making," *IEEE Intelligent Transportation Systems*, Dearborn (MI), USA, 2000b.
39. R. J. F. Rossetti, R. H. Bordini, A. L. C. Bazzan, S. Bampi, R. Liu, and D. Van Vliet, "Using BDI agents to improve driver modelling in a commuter scenario," *Transportation Research Part C: Emerging Technologies*, vol. 10, 5–6, pp. 47–72, 2002.
40. A. E. Roth, and I. Erev, "Learning in extensive form games: experimental data and simple dynamic models in the intermediate term," *Games Econ. Behav*, vol. 8, pp. 164–212, 1995.
41. A. Schaerf, Y. Shoham, and M. Tennenholtz, "Adaptive balancing load: A study in multi-agent learning," *Journal of Artificial Intelligence Research*, vol. 2, pp. 475–500, 1995.
42. H. Strobel, *Computer Controlled Urban Transportation. A Survey of Concepts, Methods and International Experiences*, vol. 10 of *IIASA Int. Series*, John Wiley and Sons, 1982.
43. TRANSYT-7F, "TRANSYT-7F user's manual," Transportation Research Center, University of Florida, 1988.
44. F. Vanderborght, "PRODYN in Brussels," in *Traffic Technology International*, Surrey, UK: International Press, 1995.
45. J. Wahle, A. L. C. Bazzan, F. Klügl, and M. Schreckenberg, "The impact of real time information in a two route scenario using agent based simulation," *Transportation Research Part C: Emerging Technologies*, vol. 10, 5–6, pp. 73–91, 2002.
46. B. Wild, "SAPPORO – Toward an intelligent traffic management system," in *Int. Conf. on Artificial Intelligence Applications in Transportation Engineering*, California, 1992.
47. C. Zozaya-Goristiza, and C. Hendrickson, "Expert systems for traffic signal setting assistance," *Journal of Transport. Engineering*, vol. 2, pp. 113, 1987.