Towards a Society of Robots: Behaviors, Misbehaviors, and Security

Antonio Bicchi, Adriano Fagiolini, Lucia Pallottino

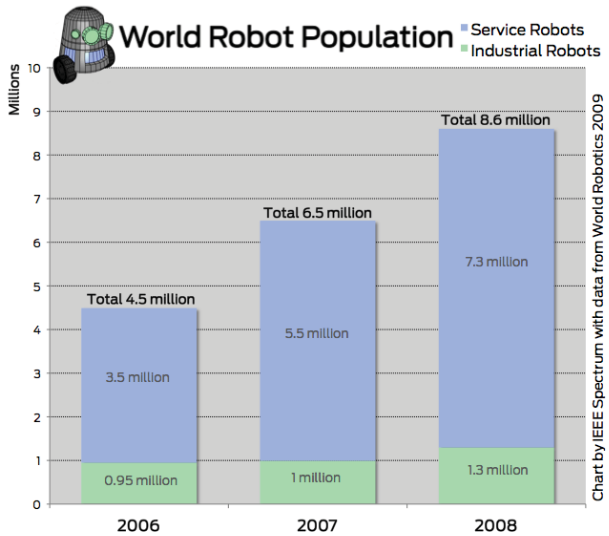


Figure 1. The world robot population. Source: IEEE Spectrum with data from World Robotics 2009.

Abstract—**In this paper we consider how very large numbers of robots, differing in their bodies, sensing and intelligence, may be made to coexist, communicate, and compete fairly towards achieving their individual goals, i.e. to build a "society of robots". We discuss some characteristics that the rules defining acceptable social behaviors should possess. We consider threats that may be posed to such a society by the misbehaviors of some of its members, due either to faults or malice, and the possibility to detect and isolate them through cooperation of peers.**

The paper presents examples of motion control protocols, for arbitrarily large groups of heterogeneous robots. We discuss intrusion detection algorithms, which allow detection of deviance from such rules, and algorithms to build a consensus view on the environment and on the integrity of peers, so as to improve the overall security of the society of robots.

I. Introduction

Since its birth 50 years ago, Robotics has witnessed a large growth and profound change in scope. Robots of the past were manipulators or vehicles designed to work in isolation. Robots of the present, or immediate future, are machines that are close to, and even in touch with humans: cognitive and physical Human-Robot Interaction (HRI) are nowadays among the most studied aspects of the discipline. However, the trend shown by available market data (fig. 1) seem to anticipate an even more surprising future, when personal robots will be so many and so ubiquitous that the core scientific and technical issues might become those of Robot–Robot Interaction.

Indeed, in recent years the increase in the use of service robots around the world has taken over the stagnating market of industrial robots. It has been estimated that $49,000$ professional service robots and 11.6 million personal service robots will be sold between 2009 and 2012 reaching a total world robot population of nearly 13 millions by around 2011 or 2012. Most of these service robots will be very different from the traditional stereotype of large heavy industrial manipulators, and probably no other stereotype will develop - as there will be possibly as many robot types as applications, and makers and models.

Large systems of many autonomous but networked units, capable of acting in and on the environment will soon be a reality. Robots will be many, autonomous, possibly fast, and very heterogeneous. The functions and structure of multirobot systems could be different [1]. One possibility is that robots could be organized in teams, flocks or swarms, to more effectively and robustly pursue a goal which is shared by all members. In this case, the paradigm of "emergent behaviors" is often used to describe the coordination of large numbers of robots with limited individual capabilities which achieve complex tasks (see e. g. [2]–[6]). When more complex robots put their specific capabilities at the disposal of a common goal, the paradigm of "intentional cooperation" is evoked [1].

In this paper we are interested in the case when the robots do not share the same goals, but have independent and possibly conflicting purposes, while they are supposed to coexist and behave so that the accomplishment of their mission does not jeopardize the chances of others. In other words, we will be concerned with the organization of a *society* of robots.

That societies of robots will soon happen is easy to predict - just consider present-day highways, where vehicles possess an ever increasing amount of sensors, actuators, communication and intelligence capabilities: an advanced car model is - as of today - already more of a "robot" than most industrial manipulators used in its own assembly. Most of the traffic rules which make coexistence of vehicles possible on the highway are currently implemented by drivers, but some automatic management systems are already commercially available (e.g., distance keeping in queues), and more are soon to come.

Looking a bit further ahead, one can easily imagine personal robots going to shop for the family (fig. 2). The user might take the robot to the local mall, where the robot would get authenticated and accepted. The robot could obtain information on goods and their locations, fill the cart (checking the list prepared by smart appliances in the house), wait in the queues etc. - while the owner makes time for more amusing activities.

Preprint submitted to IEEE Robotics and Automation Magazine
Received September 12, 2010 13:28:42 PST

Figure 2. A futuristic mall scenario. [Picture to be redrawn by the Magazine's editors]

It is interesting to speculate on what it is that makes this scenario impossible today, and what breakthroughs could enable it tomorrow. The mall could easily provide its customers with services such as localization beacons, navigation maps, and power recharging. In such a structured environment, the ability for robots to navigate the department store is already available, or soon will be, and so are those needed to fill and push the cart around. Electronic transactions are no problem, neither are wireless communications between the robots and the infrastructure or among the robots themselves. Perhaps the major obstacle would be that tens or hundreds of robots going about their individual programmed missions would compete for resources (e.g. room, power, goods), possibly creating conflicts and ending up in traffic deadlocks, collisions, or even safety hazards.

To negotiate the potential conflicts, communication and interaction among robots will have to be codified in a set of rules to which different robot producers and infrastructure authority adhere. Protocols and architectures for providing services such as communication or localization might be derived from current and developing standards for sensor networks and cyberphysical systems (see e.g. [7]). For example, a scalable component–based platform for decentralized traffic management of a multi–robot system was described in [8]. On the other hand, establishing rules for the physical interaction of robots, i.e. what behaviors are acceptable in the society, is a very challenging problem that has been explored only to a limited extent so far. Various available approaches differ in the extent to which robot team members are aware of, or recognize the actions of their teammates and discuss to which extent they should use this information to affect their own actions [9]–[11].

Pioneering work on methods for negotiating traffic and avoiding collisions based on rules includes those reported in [12], [13]. The idea of defining *group behaviors* was formalized and presented in [14], although in a cooperative framework. Behavior-based techniques are used for coordination in multi–agent RoboCup teams (see e.g. [15]). More recently, applications of protocol–based collision avoidance methods in a marine scenario has been presented in [16], where a multi–objective optimization is proposed to deal with situations where multiple rules are simultaneously activated.

Comparing the state of the art with the analysis of the mall scenario, several challenging problems are still outstanding. First, guarantees on safety (collision avoidance with robots and humans) and performance (ensuring that each robot eventually gets a chance to accomplish its mission) must be provided in the presence of many agents, whose number is not known nor can be bounded beforehand. The society should allow new robots to get in, or to leave, at any time, irrespective of their model, type, size or weight of the robots – provided only that they abide to the society's rules.

Scalability, *heterogeneity*, and *reconfigurability* are thus fundamental requirements for a system of behaviors for a society of robots. A very effective way of achieving these features is *decentralization*, i.e. decisions should be made by each agent autonomously and should be based on information limited to a local neighborhood of each robot, reducing the role of a central authority to the minimum necessary.

A system that relies on social behaviors to mitigate the excess of individualism is intrinsically very sensitive to the possibility that misbehaviors occur, due to either faults in some robots or malicious programming of agents. Thus, *security* requirements are crucial for a society of robots, which imply the capability to detect, isolate, and neutralize the threat posed by misbehaving robots (see e.g. [17], the papers on fault–tolerance in robot swarms [18] and on the ALLIANCE architecture [19], and references therein). In a society of autonomous robots, intrusion detection must also rely on information available locally and on limited knowledge of a model for the behavior of other robots.

A common problem with over–cautious security policies is that they can make the system too stiff and ineffective. In a heterogeneous robot society, a robot should not deem another robot to be a malevolent intruder just because it behaves differently, as far as that behavior does not pose a threat. Hence, a problem of detecting which type of behavior other robots in the neighborhood are following, or which "species" they belong to, is also in order.

In this paper we discuss the above challenges and present work toward solving some of them. The paper's first contribution is the formalization of a cooperation protocol by which societies of interacting robots can be described at a suitable abstraction level. We show examples of motion control protocols that guarantee collision avoidance for arbitrarily large groups of heterogeneous robots, and discuss intrusion detection algorithms, which allow detection of deviance from such rules. The description of a local misbehavior detector, representing the second contribution of the paper, is also presented. We also present algorithms to build a consensus view on the environment and on the integrity of peers, so as to improve the overall security of the society of robots. Furthermore, we show a Biologically–inspired example of social coordination protocol enabling a group of ant–like robots to cooperate during the foraging of the same group. This is based on the use of a local classifier by which individuals can distinguish neighboring robots obeying to a different set of rules and thus belonging to a different "species" or social

Preprint submitted to IEEE Robotics and Automation Magazine
Received September 12, 2010 13:28:42 PST

groups.

II. SOCIAL BEHAVIORS AS HYBRID AUTOMATA

Behavior–based societies of robots can be built by giving a set of rules that each agent should follow, which are only based on local information and communication between neighboring agents. Such rules can usually be described in the form of an automaton, with states corresponding to decisions or actions, and transitions triggered by locally evaluated conditions. A first example of a multi-robot system that has conflicting individual goals, but can negotiate crossroads by following a set of elementary rules is reported in fig. 3, while another example, where the mission goal is shared among all the members of the society, is the formation control protocol proposed by Arkin in [14] (fig. 4).

Although simple rule sets may well serve the purpose for a limited number of robots, a problem may arise when the rule same set is applied to larger and/or safety–critical systems as to whether it can be *guaranteed* that vehicles will not get into deadlocks, or even crash into each other. To provide such guarantees, a formal description of behaviors is in order. One should observe that, in dealing with physically embodied autonomous agents such as robots, traditional automata theory is limited, because of the lack of expressivity power to model continuous dynamics. The *hybrid automata* formalism and verification techniques can be effectively used to that purpose.

A *motion cooperation protocol* $\mathcal{P}$ that can describe the behavior of the individuals, $\mathcal{A}_1, \cdots, \mathcal{A}_n$, of a "robotic society" can be formalized as follows. The protocol must specify, for each robot $\mathcal{A}_i$:

- A configuration vector $q_i \in \mathcal{Q}$, where $\mathcal{Q}$ is a configuration space;
- A discrete state $\sigma_i \in \Sigma_i$, where Σ_i is the set of allowed actions or decisions;
- A dynamic map f_i describing how the agent's configuration is updated:
$$\dot{q}_i = f_i(q_i, u_i)\,,$$
where u_i is the input vector;
- A controller map g_i that, based on the agent's current configuration q_i and discrete state σ_i, returns the control value
$$u_i = g_i(q_i, \sigma_i)\,;$$
- A detector map $d_i = (d_{i,1}, \cdots, d_{i,\kappa_i})$ returning a binary vector e_i, whose j–th component activates if a *local* condition $d_{i,j}$ on the presence or absence of other agents in $\mathcal{A}_i$'s vicinity/neighborhood holds, i.e.,
$$e_i = d_i(q_i, v_i)\,,$$
where $v_i = \left\{q_{i_1}, \ldots, q_{i_p}\right\}$ is the set of configuration vectors of $\mathcal{A}_i$'s neighbors;
- An automaton δ_i describing how the agent's current action σ_i is updated based on the event vector e_i:
$$\sigma_i{}^{+} = \delta_i(\sigma_i, e_i)\,.$$

As such, the *behavior* of a robotic agent $\mathcal{A}_i$ adhering to $\mathcal{P}$ is described by the hybrid dynamics [20]

$$\begin{cases} (\dot{q}_i, \sigma_i^{+}) = \mathcal{H}_i(q_i, \sigma_i, q_{i_1}, \ldots, q_{i_p})\,, \\ q_i(0) = q_i^0\,,\ \sigma_i(0) = \sigma_i^0\,, \end{cases}$$

where q_i^0 and σ_i^0 are the initial configuration and discrete state, and $i_1, \ldots, i_p$ are the indices of $\mathcal{A}_i$'s neighbors. The architecture of an agent $\mathcal{A}_i$ adhering to the generic protocol is depicted in Fig. 5. We say that a cooperation protocol $\mathcal{P}$ is "fully distributed" if it involves only local interaction within a maximum number N of agents, i.e. $i_p \leq N$, where N is independent of the total number n of agents.

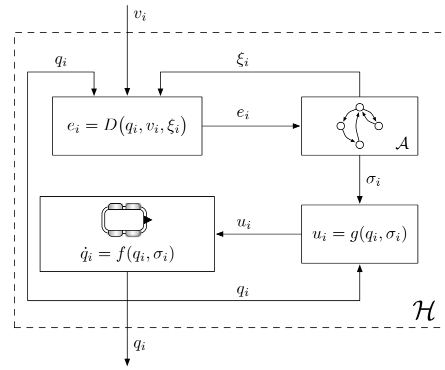


Figure 5. Architecture of a generic agent $\mathcal{A}_i$ following a motion cooperation protocol $\mathcal{P}$.

As an example of successful application of the hybrid automata theory, consider the problem of managing the traffic of an unbounded number of unmanned aerial vehicles described in [21]. Therein, Pallottino *et al.* proposed a motion coordination protocol, involving of a set of fully decentralized "behaviors", whose correctness in terms of safety and deadlock avoidance was proven through application of formal methods. The protocol, known as Generalized Round–About Policy (GRP), realizes a scalable ($N = 6$) and reconfigurable multi-robot system. Although the proof of correctness was given for aerial vehicles with the same size, it is straightforward to extend the result to vehicles with different sizes (see Fig. 6).

Another example of multi–robot system that can be formalized according to the above protocol, and that we will consider in more detail, is represented by n cars in a highway following a set of traffic rules to avoid collisions (Fig. 7). Each car has its own dynamics f_i and local controller g_i, and its pilot is supposed to decide decide a suitable maneuver, i.e., accelerate (FAST) or decelerate (SLOW), change to the next left lane (LEFT) or to the right one (RIGHT), based on the presence or absence of other cars in its neighborhood. E.g., the presence of a slower car in the front, and a free lane on the left requires the execution of an overtake that is a change from a FAST to a LEFT maneuver. These rules are those of a very large system with possibly hundreds of vehicles, but require that every car verify the existence and/or absence of N other cars in its vicinity only, where N is a small number depending only on the geometry of the lanes and of the vehicles.

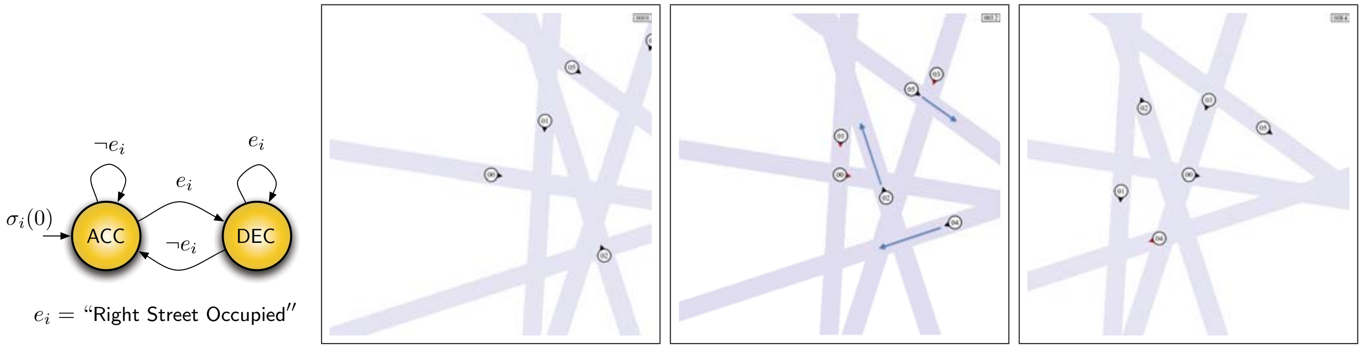


Figure 3. Autonomous LGVs in a warehouse can efficiently move products, from carrier tapes to storage piles, and avoid robot–robot and robot–human collisions, by following a very simple motion control protocol requiring that every robot give way on its own right.

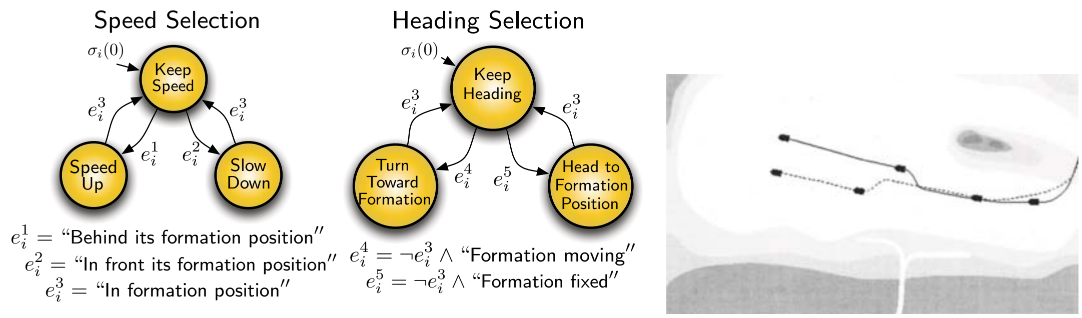


Figure 4. A behavior–based formation control protocol enabling a set of autonomous vehicles to achieve a common mission goal [14]. [[Reprint permissions for the image on the right to be asked to publisher/author]]

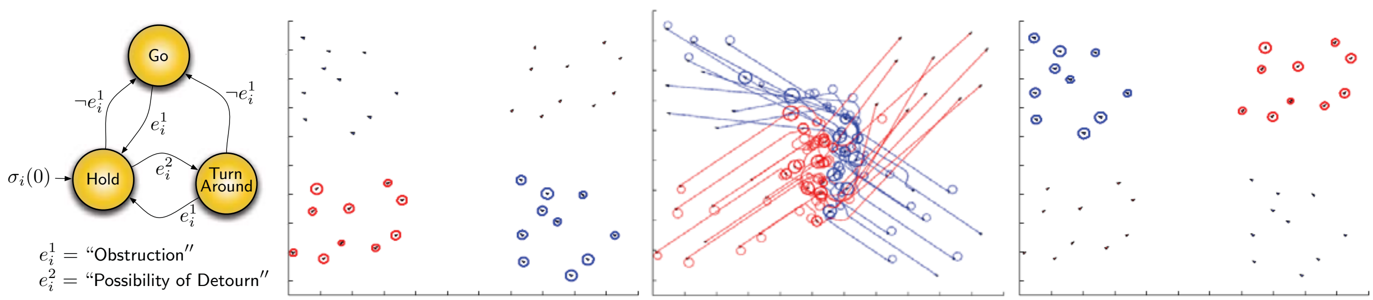


Figure 6. Automaton of the GRP motion coordination protocol and snapshots from a simulation run of the social behavior of twenty heterogeneous aerial vehicles following the protocol.

III. Misbehaviors and Local Detection

Heterogeneous robots may happen to have different protocols: for instance, in the automated highway example, a vehicle obeying the left–hand traffic rules may happen to access a right–hand automated highway. Some robots may experience sensor or actuator failure, and thus be unable to follow the protocol rules. Finally, since robot communication is based on a wireless network, an adversary could easily eavesdrop on communication as well as inject/modify packets. In such cases, safety requires that the society must be able to recognize such failures or intrusions in order to activate countermeasures to preserve the overall system and its individuals.

We define *misbehaving* an agent when the evolution of its state (q_i, σ_i) does not comply with the agent's hybrid model $\mathcal{H}_i$ described by the motion cooperation protocol $\mathcal{P}$. We use the term *intruder robot* as a synonym of "misbehaving agent". Misbehavior detection in fully distributed settings represents a tough challenge, mainly because the system's state is only partially known to a local observer. Consider e.g. the scenario depicted in Fig. 8, where a local observer on the car 00 is trying to learn whether the car 04 is cooperative or not. Not having full access to the information available for the car 04, it is difficult for the observer to decide whether the pilot is correctly driving or if it is *simulating* the presence of another car, that is hidden to the observer's view or outside of its range of visibility. In such cases, a decision on (mis)behavior classification should be postponed until enough evidence is collected.

A fully distributed misbehavior detector, or Intrusion Detection System (IDS), can be systematically generated once the motion cooperation protocol $\mathcal{P}$ is given [20]. The IDS endows an agent $\mathcal{A}_h$ with the ability to classify another neighboring

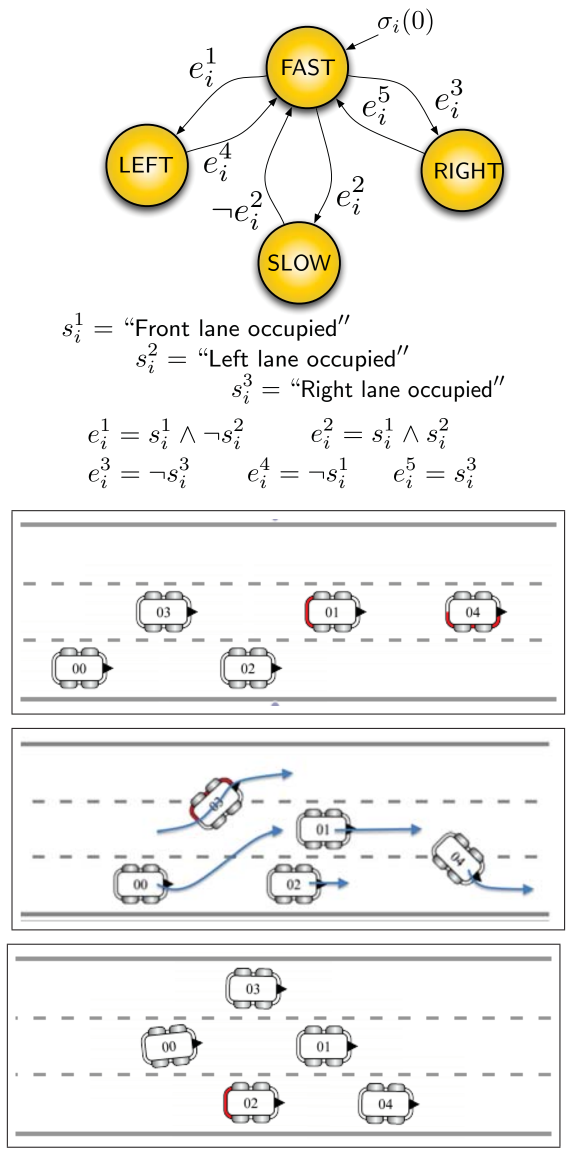


Figure 7. Automaton of the traffic rules of a society of cars that follow a collision avoidance and motion coordination protocol to guarantee their passengers' safety. Car 01 initially slows down due to the presence on its front lane of car 04, that in turn turn right as its next right lane is free. Car 03 later starts a left turn as car 01 occupies its immediate front lane and its next left lane is free. Finally, car 02 slows down as its front lane is occupied by car 04 and its next left lane is also occupied by the cars 00 and 01.

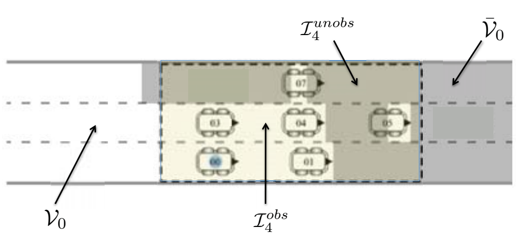


Figure 8. Example of partial visibility of a local observer. For the observer on the car 00 is difficult to discern whether the pilot of car 04 is correctly driving or it is faking the presence of other cars, that are outside the observer's view.

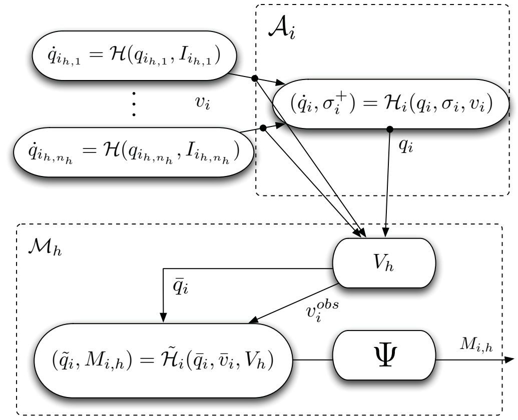


Figure 9. Architecture of a local monitor onboard of agent $\mathcal{A}_h$ that is able to estimate the cooperativeness of an agent $\mathcal{A}_i$ by using only local information.

agent $\mathcal{A}_i$ as certainly cooperative, certainly uncooperative, or still uncertain, based only on locally observed behaviors. The IDS consists of two components: a local *monitor process*, by which $\mathcal{A}_h$ can estimate a local map of occupancy $M_{i,h}$ of $\mathcal{A}_i$'s neighborhood, and a *consensus process*, described in the following section, that allows computation of a unique global view of the map.

More precisely, a local monitor is a set–valued observer $\tilde{\mathcal{H}}_i$ which computes all possible behaviors $\tilde{q}_i$ that an agent $\mathcal{A}_i$ can execute, based on the measure $\bar{q}_i$ of its current configuration and on the information of the neighbor set $\bar{v}_i$ that are visible from $\mathcal{A}_h$ (Fig. 9). The output of a local monitor is

$$(\tilde{q}_i, M_{i,h}) = \tilde{\mathcal{H}}_i(\bar{q}_i, \bar{v}_i, V_h)\,,$$

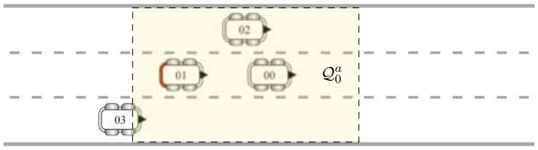
where V_h is the current visibility region of the observer on $\mathcal{A}_h$ [22].

Operation of each local monitor consists of a *prediction phase* and an *classification phase*. The predictor used during the first phase is composed of:

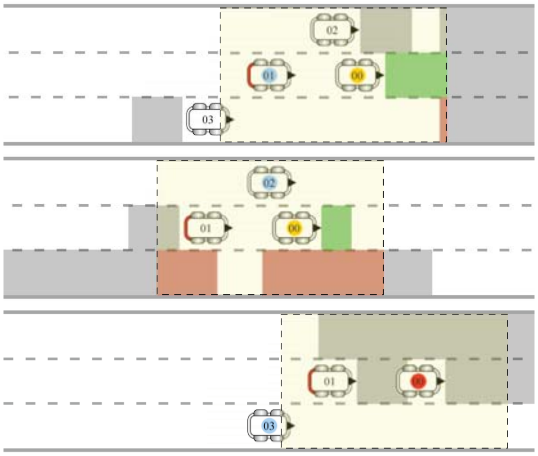
- A copy of the target agent's dynamics f_i, and a copy of its controller map g_i;
- An *uncertain encoder map* $\tilde{d}_i$ estimating the event vector $\hat{e}_i$, based on local measurement of $\mathcal{A}_i$'s configuration, $\bar{q}_i$, its known neighbors $\bar{v}_i$, and the observer's visibility region V_h:
$$\hat{e}_i = \tilde{d}_i(\bar{q}_i, \bar{v}_i, V_h)\,; \tag{1}$$
- An *uncertain automaton* defined through the nondeterministic map
$$\tilde{\delta}_i(\hat{\sigma}_i, \hat{e}_i) = \left\{\bar{\sigma} \in \Sigma_i \,|\, \exists \sigma \in \hat{\sigma}_i \,|\, \bar{\sigma} = \delta_i(\sigma, \hat{e}_i)\right\}\ ,$$
describing how the agent's estimated action $\hat{\sigma}_i$ is updated based on the estimated event vector $\hat{e}_i$:
$$\hat{\sigma}_i^+ = \tilde{\delta}_i(\hat{\sigma}_i, \hat{e}_i)\,. \tag{2}$$

During the second phase, all predicted behaviors $\tilde{q}_i$ are compared against the measured one $\bar{q}_i$. A behavior $\bar{q}_i(t)$ complies with model $\mathcal{H}_i$ if

$$||\bar{q}_i(t) - \tilde{q}_i(t)|| \leq \varepsilon\,, \text{ for all } t \in T_k\,,$$

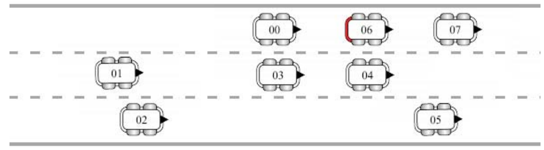


(a)

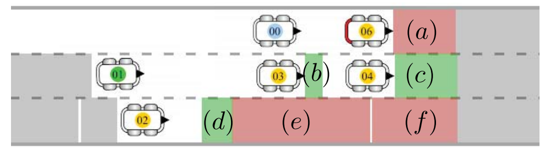


(b)

Figure 10. Misbehavior of car 00, running a FAST maneuver along the second lane, while its next right lane is free, has to be detected (a). Local maps of occupancy, $M_{00,01}$, $M_{00,02}$, and $M_{00,03}$, that local monitors on the cars 01, 02, and 03 has reconstructed (b). The yellowish area dashed box outlines the target agent neighborhood; a blue circle specifies the current monitor; red (green) areas are non–visible regions, where the presence (absence) of a car is required. A colored circle around the target robot (green, yellow, or red) specifies its estimated cooperativeness (cooperative, uncertain, or uncooperative, respectively).



(a)



(b)

Figure 11. Eight cooperative cars (a) and view of the monitor on car 00 (b). A local monitor's uncertainty in the classification of a neighbor can be reduced by cross–correlating maps of occupancies of different neighbors.

where $T_k = [t_k, t_{k+1})$ is the k–th observation period, ε is a suitable precision, and $||\cdot||$ is the Hausdorff norm. When no predicted behavior complies with the model, the agent is certainly uncooperative; if there is a unique behavior complying with the model and not depending on the region outside $\mathcal{A}_h$'s visibility range, the agent is certainly cooperative; otherwise it is possibly cooperative and thus the value uncertain is chosen.

Consider four cars in the highway example (Fig. 10–a). Misbehavior of car 00, running a FAST maneuver along the second lane, while its next right lane is free, has to be detected (the car should start a RIGHT maneuver to return to the first lane). A FAST maneuver of a car in the second lane implies that the region on its right is occupied by another car. Three local monitors on the other cars try to learn whether the car 00 is cooperative or not, but have no full view of its neighborhood of car. Fig. 10–b) reports the local maps of occupancy, $M_{00,01}$, $M_{00,02}$, and $M_{00,03}$, that each monitor has reconstructed. The figure shows that all local monitors remain uncertain on the cooperativeness of car 00, as possible cooperative behaviors could take place based on their partial visibility.

As a second example, consider eight cooperative cars (Fig. 11–a), and focus on the local view of car 00's monitor (Fig. 11–b). The presence of car 07 is detected (region a), based on the fact that car 06 is executing a SLOW maneuver. The presence of car 05 is detected (regions e, and f), based on the FAST maneuvers on the second lane executed by cars 03 and 04. This also allows the detection of car absence in front of car 03 (region b) and car 04 (region c). To the local monitor all these neighboring cars are uncertain, except car 01 that is certainly cooperative. The example is used to show the fact that — although this goes beyond the scope of the paper — a local monitor's uncertainty in the classification of a neighbor can be reduced by cross–correlating maps of occupancies of different neighbors: the occupancy map $M_{03,00}$ contains a free region (b in the figure) in front of car 03, and an occupied region (the union of d with e) on its right, while $M_{02,00}$ contains a free region (same d in the figure) in front of it. Therefore the region d in $M_{03,00}$ must be removed and the only possibly occupied region must be (e).

IV. Consensus for Misbehavior Detection

Reaching a global agreement on the presence of a misbehaving robot is essential to neutralize or reduce the threats that it may pose to the society. To this aim, for every robotic agent $\mathcal{A}_i$, a unique *consensus view* M_i of the occupancy map of its neighborhood explaining its actual behavior should be computed, through local information exchange in a communication network G. It is important to notice that we consider only misbehaviors at the motion execution level, while we assume that the exchange of information between agents is correct, and no collusion exists between a robot executing an incorrect motion and another robot trying to justify it. The problem of reaching consensus on information corrupted by intruders is a classical one in Computer Science [23], and is not investigated here. Obviously, the problem of detecting simultaneous motion and information misbehaviors is much more complex, and is left for further studies.

The approaches to consensus establishment traditionally developed within the Control Community involve algorithms that are modeled as linear systems, and that are able to combine data represented by real scalars or vectors (see e.g. [5], [24]). Theoretical results on the convergence toward a consensus to e.g. the average of the initial estimates of local agents are fully available. However, outputs from local monitors are continuous sets representing free and occupied

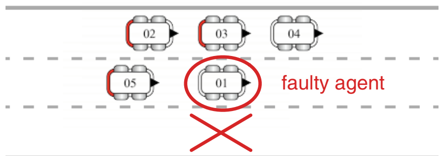


Figure 12. The misbehaving car 01 is executing a FAST maneuver on the second lane, while its next right lane is free.

regions in the neighborhood of a generic agent $\mathcal{A}_i$, and they cannot be trivially merged by using such algorithms. The need to overcome the limitations of the linear consensus framework is indeed emerging in various distributed robotic applications, and different forms of nonlinear iterative rules are under development (see e.g. [25]).

A solution to our problem is provided by the *set–valued consensus* approach, initially proposed in [20] and then furtherly investigated in [26]. The approach is based on the following theorem ensuring that every local agent consent on the decision of a hypothetical centralized monitor, that is able to collect and merge all initial estimates in one step:

Theorem 1: A network of n agents with fixed communication topology described by a graph G and evolving according to the distributed rule

$$\begin{cases} X_i(t+1) &= F(X_i(t), X_{i_1}(t), \cdots, X_{i_{n_i}}(t)), \\ X_i(0) &= U_i, \end{cases}$$

for all i, where U_i is agent $\mathcal{A}_i$'s initial measure, $i_1, \cdots, i_{n_i}$ are the indices of its communication neighbors in G, and F is a so–called *updated function*, *converges* to the consensus state

$$X = (\underbrace{X^*, \cdots, X^*}_{n\,\text{times}})^T, \text{with } X^* = F(U_1, \cdots, U_n),$$

in at most $\tilde{n} = \operatorname{diam}(G)$ steps, if G is connected and F is

- commutative ($F(X_1, X_2) = F(X_2, X_1)$),
- associative ($F(X_1, F(X_2, X_3)) = F(F(X_1, X_2), X_3)$), and
- idempotent ($F(X_1, X_1) = X_1$),

for all its input arguments X_1, X_2, X_3.

Application of this result to the misbehavior detection problem is straightforward. The state of each agent $\mathcal{A}_h$ trying to learn whether $\mathcal{A}_i$ is cooperative or not is initialized with the occupancy map $M_{i,h}$ that it has locally built, i.e. $U_h = M_{i,h}$, and it can be merged with data received from the monitors of its communication neighbors by using the set–theoretic intersection $\cap$ — the update function F above —, which satisfies Theorem 1's hypotheses. As an illustration of this approach, consider $n = 5$ agents in the highway example (Fig. 12). As above, suppose that a car (01 in the figure) misbehaves by remaining in the second lane. All other agents, 02, 03, 04, and 05, share local estimates by sending one–hop (immediate neighbor) messages through a communication network described by the connected graph $G = (V, E)$, with $V = \{2, 3, 4, 5\}$ and $E = \{e_{2,2}, e_{2,3}, e_{2,5}, e_{3,3}, e_{3,4}, e_{4,4}, e_{5,5}\}$, which gives the following instance of set–valued consensus system:

$$\begin{cases} X_2(t+1) = X_2(t) \cap X_3(t) \cap X_5(t), \\ X_3(t+1) = X_2(t) \cap X_3(t) \cap X_4(t), \\ X_4(t+1) = X_3(t) \cap X_4(t), \\ X_5(t+1) = X_2(t) \cap X_5(t), \end{cases}$$

The system's evolution is reported in Fig. 13, where the i–th row represents the evolution of $X_i(t)$ (from left to right). Although no single local monitor has initially detected the misbehavior, this is dynamically achieved first by cars 02 and 03 after two consensus rounds and then by the other cars. The simulation confirms that the consensus view of the occupancy map $M_h = U_2 \cap U_3 \cap U_4 \cap U_5$ (last column in the figure) is achieved after $\operatorname{diam}(G) = 3$ rounds.

V. Behavior Classification

This final section addresses the behavior classification problem for a set of autonomous agents. The objective is to classify heterogeneous agents that "behave" in a different way, due to their own physical dynamics or to the rules of interaction they are obeying, as belonging to a different species. The problem is not easy to solve in its generality, whereas a decentralized *classifier* (based on the IDS described above) can be constructed systematically, if the hybrid models describing all species' behaviors are available. Preliminary work on this topic is presented in [27] from which we recall the following examples.

Consider a society of robots composed of two colonies of polymorphic tree dwelling ants (*Daceton Armigerum*). Ant cooperation during the colony foraging arises whenever a prey cannot be moved by a single ant [28], [29]. This process, that can be described as an example of motion cooperation protocol $\mathcal{P}$, involves the recruitment of nestmates of the same colony, by issuing a *distinct*, colony–dependent visual or chemical marking. Suppose that two ant colonies exist, a *Green* and a *Red* one. Green ants starts moving around the prey to inform their neighbors of its impossibility to move it, whereas Red ants stops in front of it. The ants of both colonies are allowed to execute the following motions: EXPLORE $\stackrel{\text{def}}{=}$ "move straight along a random direction", STOP $\stackrel{\text{def}}{=}$ "remain fixed", ALERT $\stackrel{\text{def}}{=}$ "go toward a nestmate", RECRUITING $\stackrel{\text{def}}{=}$ "issue the visual signal to recruit neighboring nestmates", RECRUITED $\stackrel{\text{def}}{=}$ "come closer to a nestmate and check for the presence of a prey". An example where a Green ant is able to recognize and recruit its nestmates is shown in Fig. 14.

Finally, consider the example of a set of cars obeying the left–hand traffic rules sharing the same automated highway with other cars obeying the right–hand traffic rules, and emergency vehicles that are allowed to adopt both rules simultaneously. The different rules enable the presence of three existing "species" of drivers. Each car may require to understand what set of rules are followed by neighboring drivers, which can be achieved by a local classifier. Fig. 15 shows how an emergency vehicle (the white vehicle in the figure) is classified by neighboring cars (black, purple, and blue) running the local classifier. Three colored cells on top of the emergency vehicle are used to represent which species the observed behavior is compliant with (from top to down, right–hand, left–hand, and emergency species, respectively).

Preprint submitted to IEEE Robotics and Automation Magazine
Received September 12, 2010 13:28:42 PST

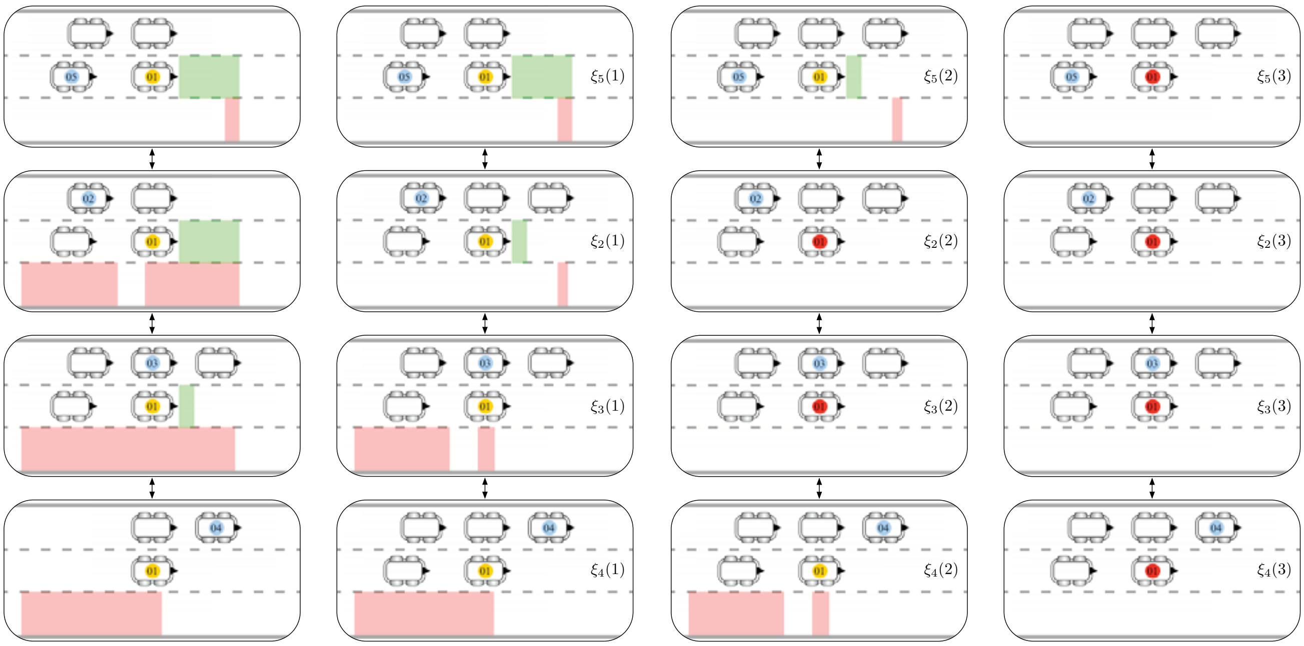
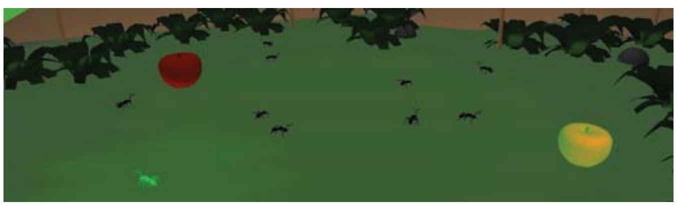
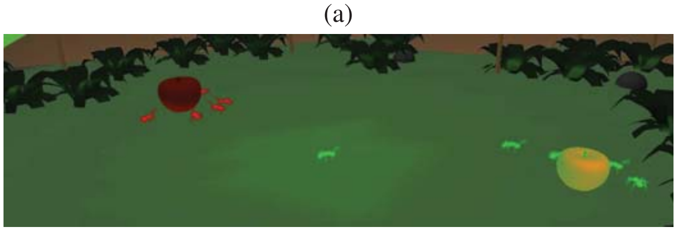


Figure 13. Misbehavior of car 01 is detected by the set–valued consensus algorithm, although no single local monitor was initially able to do it.



(a)



(b)

Figure 14. A Green ant is able to recognize and recruit nestmates of the same colony, by using only local observation. The ant is initially unaware of its nestmates (a), but finally classifies and detects all of them (b).

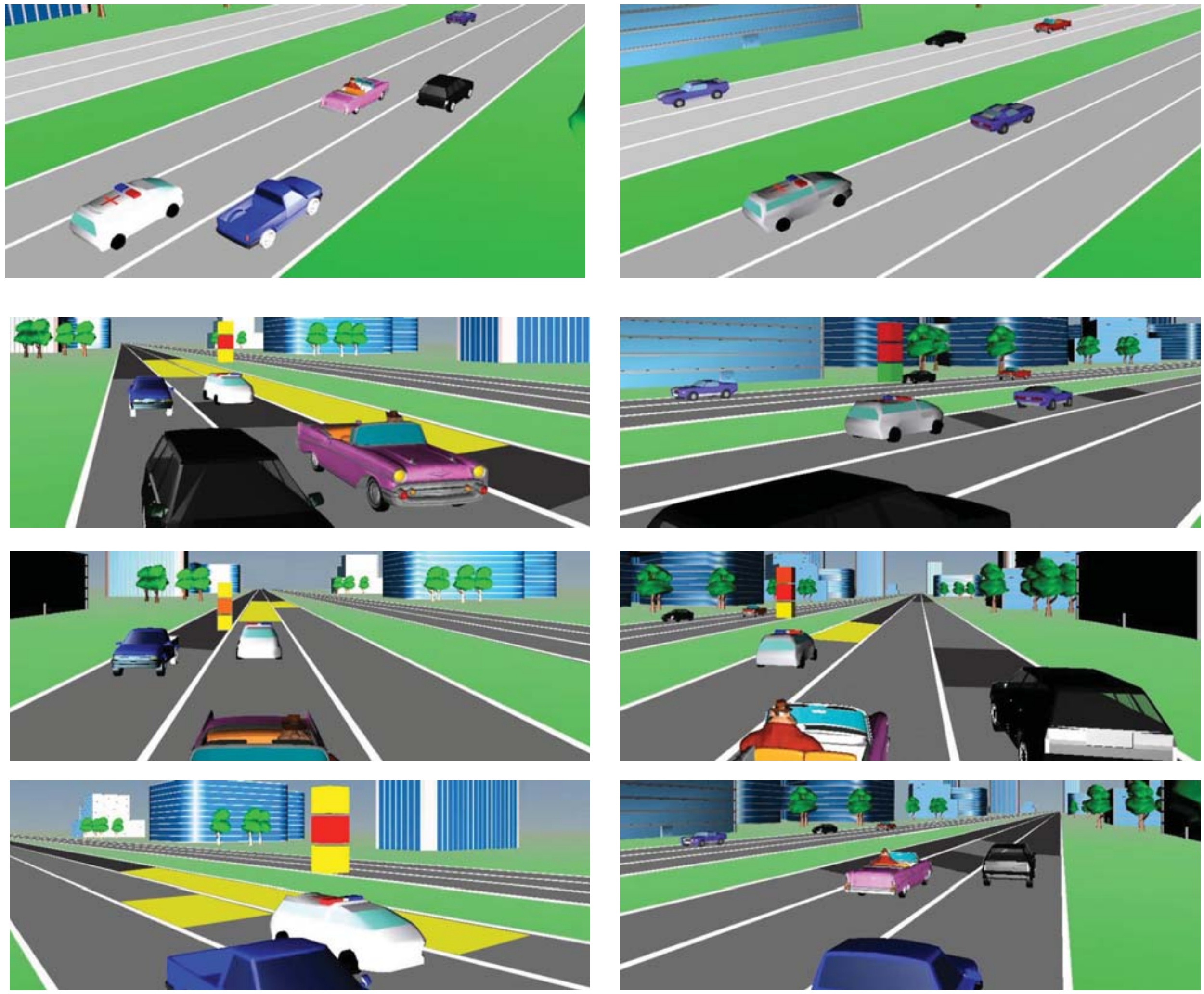
Fig. 15–a shows the instant at which the emergency vehicle changes from FAST to LEFT to overtake the purple vehicle. Note that a FAST to LEFT transition for an agent following the right–hand traffic rules implies that its frontal lane is occupied by another vehicle and its next left lane is free. To the contrary, a FAST to LEFT transition for a car following the left–hand traffic rules implies its frontal lane is occupied, but also that its next left lane is free, which is false in the example. Although their limited views, all classifiers are able to exclude the left–hand species (red color in the second cell), whereas the others all still possible (yellow color in the first and third cells). Fig. 15–b shows a successive time at which the emergency vehicle changes from FAST to RIGHT to overtake the violet vehicle in the figure. With similar reasoning this allows also the right–hand species to be excluded, which allows the vehicle to be recognized as one of the individual of the emergency species. The classifier on the black car is indeed able to distinguish the vehicle's species (green color in the third cell). The classifier on the purple car is able to distinguish the emergency species, but has still insufficient information to exclude possible unknown species (yellow color). The blue car has no more visibility of the emergency vehicle, and thus would need the execution of a consensus algorithm to correctly classify it.

VI. Conclusion

In this paper, we have considered some of the problems that will be encountered in the construction of large systems of autonomous and heterogeneous robots. The role of hybrid automata descriptions in providing verifiable safety properties, and in building general distributed intrusion detection systems for increasing the security of these systems has been shown. The method has also been applied to allow members of a society to classify other individuals based on their behaviors in case a model for such behavior is available. If this is not the case, a much harder and very interesting problem arises which requires construction of a model for a behavior that is observed in individuals of the society.

Acknowledgment

Authors would like to acknowledge Simone Martini for his help in working out the two examples for the behavior classification section. This work has been partially supported by the European Commission with contract FP7-IST-2008-224428 "CHAT - Control of Heterogeneous Automation Systems:



(a) (b)

Figure 15. Example of cars obeying the left–hand traffic rules and sharing the same automated highway with other cars obeying the right–hand traffic rules, and emergency vehicles that are allowed to adopt both rules simultaneously.

Technologies for scalability, reconfigurability and security”, and with contract number FP7-2007-2-224053 CONET, the “Cooperating Objects Network of Excellence”.

REFERENCES

[1] L. Parker, “Distributed intelligence: Overview of the field and its application in multi-robot systems,” *Journal of Physical Agents*, vol. 2, no. 1, p. 5, 2008.

[2] H. Tanner, A. Jadbabaie, and G. Pappas, “Flocking in fixed and switching networks,” *IEEE Transactions on Automatic Control*, vol. 52, no. 5, pp. 863–868, 2007.

[3] R. Olfati-Saber, “Flocking for multi-agent dynamic systems: Algorithms and theory,” *IEEE Transactions on Automatic Control*, vol. 51, no. 3, pp. 401–420, 2006.

[4] S. Oh, L. Schenato, P. Chen, and S. Sastry, “Tracking and coordination of multiple agents using sensor networks: system design, algorithms and experiments,” *Proceedings of the IEEE*, vol. 95, no. 1, pp. 234–254, 2007.

[5] J. Fax, R. Murray, N. Syst, and C. Woodland Hills, “Information flow and cooperative control of vehicle formations,” *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1465–1476, 2004.

[6] M. Egerstedt and X. Hu, “Formation constrained multi-agent control,” in *IEEE International Conference on Robotics and Automation*, vol. 4, 2001.

[7] R. Ganti, Y. Tsai, and T. Abdelzaher, “Senseworld: Towards cyber-physical social networks,” in *Proceedings of the 7th International Conference on Information processing in Sensor Networks*. IEEE Computer Society Washington, DC, USA, 2008, pp. 563–564.

[8] A. Bicchi, A. Danesi, G. Dini, S. La Porta, L. Pallottino, I. M. Savino, and R. Schiavi, “Heterogeneous wireless multirobot system,” *IEEE Robotics & Automation Magazine*, vol. 15, no. 1, pp. 62–70, 2008.

[9] C. Touzet, “Robot awareness in cooperative mobile robot learning,” *Autonomous Robots*, vol. 8, no. 1, pp. 87–97, 2000.

[10] L. Parker, “The effect of action recognition and robot awareness in cooperative robotic teams,” *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 1, pp. 212–219, 1995.

[11] S. Guirnaldo, K. Watanabe, and K. Izumi, “Enhancing the awareness of decentralized cooperative mobile robots through active perceptual anchoring,” *International Journal of Control, Automation and Systems*, vol. 2, pp. 450–462, 2004.

[12] H. Asama, K. Ozaki, H. Itakura, A. Matsumoto, Y. Ishida, and I. Endo, “Collision avoidance among multiple mobile robots based on rules and communication,” in *Proc. of the lEEE/RSJ International Conference onIntelligent Robots and Systems*, 1991, pp. 1215–1220.

[13] S. Kato, S. Nishiyama, and J. Takeno, “Coordinating mobile robots by applying traffic rules,” in *Proc. of the lEEE/RSJ International Conference onIntelligent Robots and Systems*, vol. 3, 1992.

[14] T. Balch and R. Arkin, “Behavior-based formation control for multirobot teams,” *IEEE Transactions on Robotics and Automation*, vol. 14, no. 6,

Preprint submitted to IEEE Robotics and Automation Magazine
Received September 12, 2010 13:28:42 PST

pp. 926–939, 1998.
[15] C. Candea, H. Hu, L. Iocchi, D. Nardi, and M. Piaggio, "Coordination in multi-agent robocup teams," *Robotics and Autonomous Systems*, vol. 36, no. 2-3, pp. 67–86, 2001.
[16] M. Benjamin, J. Leonard, J. Curcio, and P. Newman, "A method for protocol-based collision avoidance between autonomous marine surface craft," *Journal of Field Robotics*, vol. 23, no. 5, pp. 333–346, 2006.
[17] F. Pasqualetti, A. Bicchi, and F. Bullo, "Distributed intrusion detection for secure consensus computations," in *IEEE Conference on Decision and Control*, 2007, pp. 5594–5599.
[18] A. Winfield and J. Nembrini, "Safety in numbers: fault-tolerance in robot swarms," *International Journal of Modelling, Identification and Control*, vol. 1, no. 1, pp. 30–37, 2006.
[19] L. Parker, "Alliance: An architecture for fault tolerant multirobot cooperation," *IEEE Transactions on Robotics and Automation*, vol. 14, no. 2, pp. 220–240, 1998.
[20] A. Fagiolini, M. Pellinacci, G. Valenti, G. Dini, and A. Bicchi, "Consensus–based Distributed Intrusion Detection for Secure Multi-Robot Systems," *IEEE International Conference on Robotics and Automation*, 2008.
[21] L. Pallottino, V. Scordio, A. Bicchi, and E. Frazzoli, "Decentralized cooperative policy for conflict resolution in multivehicle systems," *IEEE Transactions on Robotics*, vol. 23, no. 6, pp. 1170–1183, 2007.
[22] A. Fagiolini, G. Valenti, L. Pallottino, G. Dini, and A. Bicchi, "Local monitor implementation for decentralized intrusion detection in secure multi-agent systems," in *IEEE International Conference on Automation Science and Engineering*, 2007, pp. 454–459.
[23] L. Lamport, R. Shostak, and M. Pease, "The byzantine generals problem," *ACM Trans. on Programming Languages and Systems (TOPLAS)*, vol. 4, no. 3, pp. 382–401, 1982.
[24] R. Olfati-Saber, J. A. Fax, and R. N. Murray, "Consensus and Cooperation in Networked Multi–Agent Systems," *Proceedings of the IEEE*, 2007.
[25] J. Cortès, "Distributed algorithms for reaching consensus on general functions," *Automatica*, vol. 44, no. 3, pp. 726–737, Mar. 2008.
[26] A. Fagiolini, S. Martini, N. Dubbini, and A. Bicchi, "Distributed consensus on boolean information," in *1st IFAC Workshop on Estimation and Control of Networked Systems (NecSys'09)*, Venice, Italy, September, 24 - 26 2009, pp. 72 – 77.
[27] S. Martini, A. Fagiolini, G. Zichittella, M. Egerstedt, and A. Bicchi, "Decentralized classification in societies of autonomous and heterogenous robots," in *Submitted to IEEE International Conference on Robotics and Automation*, Shanghai, May, 9 - 13 2011.
[28] B. Hölldobler and E. Wilson, "The multiple recruitment systems of the African weaver ant Oecophylla longinoda (Latreille)(Hymenoptera: Formicidae)," *Behavioral Ecology and Sociobiology*, vol. 3, no. 1, pp. 19–60, 1978.
[29] B. Hölldobler, R. Stanton, and H. Markl, "Recruitment and food-retrieving behavior in Novomessor (Formicidae, Hymenoptera)," *Behavioral Ecology and Sociobiology*, vol. 4, no. 2, pp. 163–181, 1978.

Antonio Bicchi is Professor of System Theory and Robotics at the University of Pisa. ÊHe graduated at the University of Bologna in 1988 and was a postdoc scholar at M.I.T. A.I. Lab in 1988–1990. His main research interests are in Dynamics, kinematics and control of complex mechanical systems, including robots, autonomous vehicles, and automotive systems; Haptics Êand dextrous manipulation; Theory and control of nonlinear systems, in particular hybrid (logic/dynamic, symbol/signal) systems. He has published more than 200 papers on international journals, books, and refereed conferences.
He currently serves as the Director of the Interdepartmental Research Center "E. Piaggio" of the University of Pisa, and as Editor in Chief of the Conference Editorial Board for the IEEE Robotics and Automation Society (RAS). Antonio Bicchi is an IEEE Fellow since 2005. He has served as Vice President of IEEE RAS, Distinguished Lecturer, and editor for several scientific journals including Transactions on Robotics and Automation and Int.l J. Robotics Research. He has organized and co-chaired the first WorldHaptics Conference (2005) and Hybrid Systems: Computation and Control (2007).

Adriano Fagiolini received his master degree in Computer Science Engineering, awarded *cum laude* in 2004, from the University of Pisa, with a thesis on Casting Manipulation. He was summer student at the European Center for Nuclear Research (C.E.R.N.), Geneva. He received his Ph.D. in Robotics and Automation from the University of Pisa in 2009. During his Ph.D., he enrolled in the ICO of Doctoral Studies in Hybrid Control for Complex, Distributed and Heterogeneous Embedded Systems, and he has actively participated in the EU Projects RUNES, CHAT, and HYCON. He led the UNIPI Team at the 1st ESA Lunar Robotics Challenge, where the team came in Second Place. His main research interests are in Boolean consensus and intrusion detection in distributed robotic systems. He is carrying on his activity at the University of Pisa, where he serves as a Post–Doc Researcher.

Lucia Pallottino received the "Laurea" degree in Mathematics from the University of Pisa in 1996, and the Ph.D. degree in Robotics and Industrial Automation degree from the University of Pisa in 2002. She has been Visiting Scholar in the Laboratory for Information and Decision Systems at M.I.T., Cambridge, MA and Visiting Researcher in the Mechanical and Aerospace Engineering Department at UCLA, Los Angeles, CA. She joined the Faculty of Engineering in the University of Pisa as an Assistant Professor in 2007. Her main research interests within Robotics are in optimal motion planning and control for nonholonomic vehicles, multi-agent systems and cooperating objects, optimal control, and Air Traffic Management Systems.