

Industrial Engineering and Computer Sciences Division (G2I)

**DEFINING AGENT ORGANIZATIONS
USING SOCIAL COMMITMENTS**

Cosmin CARABELEA, Olivier BOISSIER

Décembre 2004

RESEARCH REPORT
2004-700-003



Les rapports de recherche
du Centre G2I de l'ENSM-SE
sont disponibles en format PDF
sur le site Web de l'Ecole

G2I research reports
are available in PDF format
on the site Web of ENSM-SE

www.emse.fr

Centre G2I
Génie Industriel et Informatique

Division for
Industrial Engineering and Computer Sciences
(G2I)

Par courrier :

By mail:

Ecole Nationale Supérieure des Mines de Saint-Etienne
Centre G2I
158, Cours Fauriel
42023 SAINT-ETIENNE CEDEX 2
France

Abstract

The coordination in open multi-agent systems is difficult to achieve, mainly because agents are autonomous. Social commitments have been proposed as a new coordination paradigm, suitable to describe agent communications or interactions. As some authors have pointed out, commitment enforcement is necessary, i.e., autonomous agents that violate commitments must be punished. In related work this is done by means of agent organizations or institutions, of which several different models have been proposed so far. In this paper we go one step further and we extend commitment theories to define roles and organizational structures using social commitments. We thus obtain a unified model of coordination in multi-agent systems: the expected behaviour of an agent can be defined using social commitments in both organizational and non-organizational contexts.

1. Introduction

Agent coordination in open and heterogeneous systems is one of the main challenges faced by the research in the area of multi-agent systems. In an open and heterogeneous system, autonomous agents, potentially different, can enter and exit at any time. It is thus very difficult for the system's designer to specify what is the system's desired behaviour without making any assumptions on the agents' internal model. Moreover, agents' obedience to the system's coordination mechanisms cannot be taken for granted. Due to their autonomy, the agents can disobey the coordination mechanisms imposed to them.

A coordination model is thus needed, one that can be easily verified whether it is obeyed or not and that makes no assumptions on the agents' internal model. In recent years one paradigm has emerged as an answer to the need of such a model: *social commitments*. One of the most known utilizations of social commitments is as a useful new approach on agent communication languages, one that fills the gaps of the classical, speech acts-based, approach. Dialog Games [12][13] represent an example of this utilization of social commitments. However, besides agent communication, social commitments have been used to provide a generic approach for agent interactions in general [16].

Agent organizations, norms and/or institutions are coordination mechanisms imposed to autonomous agents in order to ensure a coherent behaviour of the system. There are many models and definitions proposed for these terms and this is still an ongoing area of research in the multi-agent community. The utilization of such mechanisms modifies the nature of agent interactions that

now take place in an organizational context. Some authors have thus investigated how the social commitments paradigm is influenced by the organization in which the agent interactions take place [17].

As most of the authors point out, autonomous agents can violate the social commitments they made towards other agents. A commitment enforcement mechanism is thus necessary, a mechanism that makes the agents pay for the violation of social commitments. Sometimes this mechanism takes the form of an agent institution or organization. Our aim in this paper is to provide a commitments-based definition for agent organizations. We will then obtain a unified model of agent interactions in both organizational and non-organizational contexts. In other words, using the same model – social commitments – agents will be able to represent both their interactions with other agents and the organization they belong to and that enforces the commitment fulfillment.

This paper is loosely based on the work of Singh [17] who tries to capture several normative concepts as different types of social commitments or social policies (i.e., meta-commitments). Our aim is not to propose a new definition of social commitments or to provide a complete model of the life cycle of a commitment. Instead, in this paper, we show how social commitments can be used to represent the expected behaviour of an agent playing a role in an agent organization, i.e., how roles and organizational structures can be defined using social commitments.

This paper is structured as follows. In Section 2 we describe the foundations of our work: the social commitment definition we use, the task model used to describe the object of a commitment, operations with commitments and social policies. We then show how these notions are used to define roles and organizational structures (Section 3). In Section 4 we discuss how social commitment enforcement could be done in agent institutions. In the final section we draw some concluding remarks and we trace directions for future work.

2. Foundations

There are several attempts in related work to define social commitments, to identify their life cycle and to propose means to enforce their fulfillment, e.g., [3],[14] or [17]. The objective of this paper is not to propose a new definition of social commitments, but to present another utilization of them: defining agent organizations. Towards this aim, we build our work on existing work on social commitments. In this section, we describe the definition of and the operations with social commitments that we use as a starting point of our approach. One of the problems of existing definitions of social commitments is how to describe the object of the commitment (be it an action to execute, a goal to achieve, a condition to satisfy,

etc.). A candidate model that unifies these objects, together with their duration and deadline, is the TAEMS task model [11], described in the following.

2.1 A TAEMS-based representation of tasks

Most of the proposed definitions of social commitments consider that an agent commits itself towards another for an action to perform. We believe however that a richer description of the object of a commitment is needed and that agents can commit to execute an action or a plan, to achieve a goal, etc. This is the reason why in this paper we use a TAEMS-based [11] task representation.

In this representation, a goal to achieve is divided into subgoals that can be further divided until a level containing elementary actions is reached. The agent's planning process does this goal decomposition and forms the tree-like structures that represent the agent's plans. A *task* is a goal-decomposition tree (a partial plan to achieve a goal), a goal to achieve, or an action to execute. Besides this decomposition of goals into subgoals, the TAEMS model allows the representation of relationships between tasks (e.g., facilitates, enables, etc.) and of the resources needed for the execution of actions.

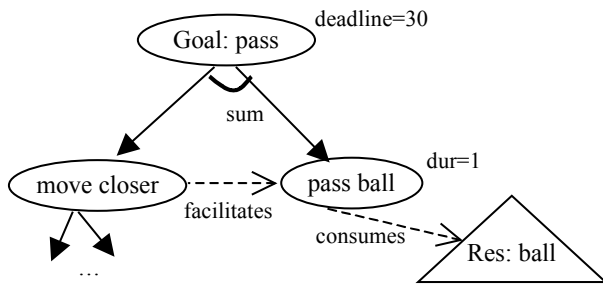


Figure 1. Example of a TAEMS task structure

An example of a task structure is given in Figure 1. This structure refers to the case of soccer-playing agents [15] and represents the plan used by an agent to satisfy the goal of passing the ball to a teammate. This goal has a deadline and can be achieved by executing an action (*pass*) that consumes the resource *ball*. However, to improve the chances of success of this action, the agent can try to execute a *move closer* task that *facilitates* the execution of that action. By reasoning on the tasks' durations and deadlines, an agent is able to form plans to achieve its goals. In the rest of this paper we will use this scenario of soccer playing agents to exemplify the concepts introduced.

We chose to use the TAEMS model because it is a generic one and it can be easily extended to cope with most task representation requirements. For reasons that will become clear later, we need to group tasks into sets depending on some of their characteristics. We thus use

the concept of *task type*, which is domain-dependent and for which we cannot provide a general definition. For example, in our scenario of soccer agents, examples of task types can be “all tasks done when having the ball” or “actions that modify the agent’s position”, etc.

2.2 Social commitments

In this paper we use the notion of *social commitment* to analyze agent organizations. Before going further, we would like to point out that there is a difference between different existing types of commitments: individual, collective or social [3]. Individual commitments are very similar with agent’s intentions in the BDI model. Collective commitments are the individual commitments of an agent representing a group of agents [6]. Social commitments are a relative notion: an agent is committed towards another to do something. Intuitively, social commitments are similar, although not identical, to the notion of contract [2].

There are several definitions and formalisms proposed to represent social commitments, for example the ones from [3], [14] or [17]. As most of these approaches, we define a social commitment (SC) as a predicate:

$$SC(\text{debtor}, \text{creditor}, \text{object}, \text{deadline}, \text{IA}, \text{status}) \quad (1)$$

When a social commitment is formed, the *debtor* agent commits itself towards the *creditor* to fulfill the *object* of the commitment before the *deadline*. The creditor can be another agent but also a collective one (e.g., a group), as we will see later. The object of a commitment can be an action to be performed, a goal to be achieved, a condition to be fulfilled, etc. We generalize this and we represent the object of a commitment as a task to be done. However, as we will see later, there are two special cases of social commitments that have different kinds of objects: roles to be played or expressions containing commitments. Although each task has associated a duration and a deadline, we decided to follow the example of [14] to specify the deadline of a commitment too. We believe that there is a conceptual difference between the deadline of a task and the deadline of a commitment to do that task.

The *IA* represents the *institutional agent* and it will be defined later in section 4. Intuitively, it represents what Castelfranchi calls the witness of the commitment [3]. It is however more close to the commitment’s context of Singh [17]: the organization in which the commitment is formed and which ensures its fulfilment. The IA represents the institution that, among other things, is responsible with the enforcement of commitments. It contains police agents that detect the commitment violation and agents that sanction this violation. Thus, in our model, the IA contains the two sanction parameters of a commitment proposed by Pasquier [14]. These sanctions can be applied to both the debtor and the creditor of the commitment.

We can thus talk about the beliefs or the desires of a role, i.e., the beliefs and the desires that an agent playing the role must have. Dastani *et al.* [4] use a similar model and define a role in terms of the goals it must achieve and the permissions it has to do so. They also consider the problem of an agent playing the role and having to conciliate its own goals with the ones of the role it plays. One of the drawbacks of these two approaches is that they make assumptions on the agent’s internal model. For example, it will be difficult for a non-BDI agent to understand what it means that the role it plays has an associated desire.

In fact, one of the reasons why defining agent organizations is still a challenge is the dynamic property of roles. For example, norms are context-dependent. One cannot simply specify what are the obligations of a role, but one needs also a mechanism to tell the agent playing the role when these obligations are active. Or, instead of just listing from the beginning all the goals a role must achieve, one should be able to specify authority relationships like “whenever your superior delegates a goal to you, you must adopt that goal”.

We believe that social commitments and policies are a suited tool to address these issues and that one can use them to define the constraints imposed to an agent playing a role in an agent organization. We are not the first to hold this belief: in this paper we extend the work of Singh [17] by introducing the permissions and prohibitions associated with a role and by identifying different types of social policies that target a role. An attempt to a common formalization of roles, commitments and obligations is presented in [8]. However, this attempt takes into considerations only the obligations an agent has because it is committed towards another, and not the organizational obligations imposed to roles.

Several authors (e.g., [2] or [5]) have used the concept of *contract* to define agent organizations. Their work is similar to ours because contracts and social commitments are similar concepts. The main difference between our commitment-based approach and the contract-based ones is due to the subtle difference between a contract and a social-commitment. Contracts are institutional, i.e., a contract must be signed in front of an institution (e.g., a notary) in order to be considered a contract. While some definitions of social commitments (like ours) use an institutional agent in one form or another, social commitments are not intrinsically institutional. They have a broader meaning and use, e.g., social commitments have been used to define agent communications not only interactions within and with an institution. We believe that the main advantage of using social commitments is that we will use the same *concept* to describe agent interaction in both institutional and non-institutional contexts.

3.2 Social commitments imposed to roles

We can classify the constraints imposed to an agent playing a role in an organization into several categories:

- *goals to achieve*: when it accepts to play a role, an agent accepts to try to achieve several goals, the role’s goals.
- *authority relations*: a role can have authority over another goal for something.
- *context-dependent obligations*: when playing a role, an agent might have to fulfill several obligations towards the organizations.
- *permissions and prohibitions*: when it accepts to play a role, an agent receives permissions to do some tasks and prohibitions to do others.

In the following we will show how the first three types of constraints can be expressed using social commitments or policies, while in the next section we will show how the fourth type completes the proposed role definition.

3.2.1 Role goals. The description of a role in an organization usually contains the goals that must be achieved by an agent playing that role. However, because we do not want make any assumptions on the agent’s internal model (e.g., goal-based), we will not express these characteristics of a role as goals, but as social commitments. In order to understand what it is expected to do when playing a role, an agent must simply understand for what it is socially committed. It can then translate these commitments into whatever notion it uses for its internal reasoning (e.g., goals, individual commitments, etc.). The *role goal commitment* (RGComm) has the form:

$$RGComm(x:R, OA, t, dl, IA, status) \quad (5)$$

We denote by $x:R$ the fact that an agent x plays a role R . The agent is committed towards the organization in which it plays the role. However, to be consistent with the existing definitions of social commitments, we say the agent is committed towards another agent that represents the organization and we note here by OA . Very often this agent is the same with the institutional agent in charge with the enforcement of the commitment ($OA \equiv IA$), but this is not always the case.

Let us now develop the soccer agents scenario by considering organizational concepts. We consider an organization *team* in which there are at least two roles: *player* and *coach*. The goal to win a match is a high-level goal for which the agents do not necessarily have a unique or explicit plan. However, it represents a goal that all agents playing the role *player* have. We can represent this by saying that an agent playing this role is committed towards the organizational agent *TeamAgent* (TA) to achieve this goal:

The *status* field represents one of the states in which a commitment can be ([14]): *inactive*, *active*, *violated*, *fulfilled* or *cancelled*. When a commitment is formed, it passes from the inactive state into the active one. If the debtor does the task-object of the commitment before the commitment's deadline has passed, then the commitment is fulfilled, otherwise it is violated. Finally, the creditor can cancel a commitment. Several commitment operations can be defined, operations that have as effect the status change of a commitment. Note that it is the responsibility of the IA to validate the status changes of a commitment and to take the appropriate measures.

In the soccer agents scenario we can represent the fact that an agent *striker* is committed towards another agent *defender* to execute the task *t* – to move to a given position and wait the ball there. If, for example, *striker* has done the task before the commitment's deadline – *10*, the institutional agent considers the commitment has been fulfilled:

$$SC(\text{striker}, \text{defender}, \text{move/wait}, 10, \text{IA}, \text{fulfilled}) \quad (1')$$

2.3 Operations with commitments and social policies

In order to enable agents to manipulate commitments, one must define operations that modify the commitments' status, such as the creation or cancellation of commitments. As this is the most relevant operation from the point of view of agent organizations (as we will see in the following section), in this paper we describe only the creation of commitments:

$$\forall x, y \in A \quad \forall t \in T \quad \forall IA \quad \text{create}(x, SC(x, y, t, dl, IA, \text{inactive})) \rightarrow SC(x, y, t, dl, IA, \text{active}) \quad (2)$$

In this paper, \rightarrow stands for material conditional, while \Rightarrow stands for strict conditional. We denote by *A* the set of all agents, by *T* the set of all tasks and by *TT_i* subsets of tasks that correspond to specific task types.

The agent interaction with the aim of creating commitments can be represented in different ways, e.g., using a request performative in a FIPA-like protocol or as result of commitment-based dialog games. In this paper we use a special predicate *request* that signifies that an agent requests the creation of a social commitment by another agent:

$$\exists x, y \in A \quad \exists t \in T \quad \exists IA \quad \text{request}(y, x, t, dl, IA) \Rightarrow \text{create}(x, SC(x, y, t, dl, IA, \text{inactive})) \quad (3)$$

For example, we can specify that the social commitment of a striker towards a defender (Formula 1') has been created as the result of an explicit request from the defender:

$$\text{request}(\text{defender}, \text{striker}, \text{move/wait}, 10, \text{IA}) \quad (3')$$

Note that commitments can be created proactively by agents, i.e., without receiving a request from other agents. It is important to point out that a request to form a commitment is not always followed by the creation of the commitment. Singh [17] introduces the notion of *social policy* that represents a commitment of a higher order, i.e., a meta-commitment. A social policy is a commitment that has as object an expression containing a commitment. In this paper we are interested on a special social policy, the *commitment creation policy* (CCPol) that specifies what is the possible answer to a request to create a commitment:

$$CCPol(x, y, (\forall z \in A \quad \forall t \in T: \text{request}(z, x, t, dl, IA) \Rightarrow \text{create}(x, SC(x, z, t, dl, IA, \text{inactive})), dl', IA' \text{ active})) \quad (4)$$

This social policy is nothing else but a social commitment. An agents *x* can commit to an agent *y* in front of an institutional agent *IA'* to do this policy. This means that *x* commits to obey all requests to create social commitments from other agents, which means the *CCPol* is a very restrictive social policy. In the next section we will give several examples of how different, less-restrictive, forms of this policy can be used to define agent organizations.

3. Defining agent organizations using social commitments

Agent organizations are social structures and patterns of agent interaction that are imposed to agents in order to ensure a coherent behaviour of the system. While the high number of organizational models proposed in related work (e.g., [9] or [10]) makes it difficult to generalize, we believe that the key notions used in defining agent organizations are the *role* an agent plays and the *norms* imposed to this role.

To enable agents to understand the organization they belong to, one must describe the constraints that playing a role imposes on the behaviour of an agent. For example, what an agent should do when playing a role (obligations), what it may and may not do (permissions), what are the relations with other roles belonging to a group (hierarchy), and how playing a role influences the agent's own goals. In this section we will show what we believe is missing from the existing work on role and organizations definition and how social commitments can be used to fill these gaps.

3.1 The context of our work

Due to space reasons, among the existing approaches on role definition we cite here only two. Boella and van der Torre [1] define roles using norms and the BDI model.

$RGComm(x:player, TA, win, ever, TA, active) \quad (5')$

The TeamAgent is an agent representing the whole team. It can be a specific agent, e.g., the team's manager, or an abstract agent. In this example we considered that this commitment does not have a precise deadline, i.e., as long as an agent plays in a team, it is committed to win the matches it plays. This commitment shows the difference between the task's deadline (win before the match is over) and the commitment's deadline (it does not have one). The TeamAgent is also responsible with the enforcement of this commitment. In Section 4 we give an example of punishment in case of the violation of this commitment.

3.2.2 Authority relations. It is difficult to express from the beginning all the goals that must be achieved by a role in an organization. This is one of the reasons for the utilization of hierarchies of roles or authority relations between these roles. We then say that a role has authority over another and by this we mean that whenever the "superior" role delegates a task to the "inferior", the latter adopts that task. However, this authority relation is not absolute, one does not have authority over another for everything possible, but only for some tasks. In real-world organizations it is difficult, if not impossible, to express all the possible objects of such a relationship. This is where our concept of *task type* comes in handle. We do not define authority relationships for all tasks or for each individual task, but for task types. This is not a novel idea: Castelfranchi [3] uses the notion of generic commitments to describe commitments formed for a specific type of objects and not for only one object.

The authority relations can be expressed as meta-commitments, more precisely as a special commitment creation policy: the *role authority policy* (RAPol). For example, the authority a role R_2 has over a role R_1 for the tasks of type TT in an organization:

$RAPol(x:R_1, OA, IA, (\forall y:R_2 \forall t \in TT: request(y, x, t, IA) \Rightarrow create(x, SC(x, y, t, IA, inactive))) , active) \quad (6)$

When an agent x plays a role, it adopts the social policies that define the authority relations other roles have over its role. These policies specify that an agent forms social commitments when it receives justified requests, i.e., requests from agents playing the roles with authority for the object of the request.

To continue with the previous example, we specify that the role *coach* has authority over the role *player* for all the tasks concerning positioning on the field. This can be expressed using a role-authority policy: an agent playing the role *player* is committed towards the TeamAgent to accept all requests from an agent playing the role *coach* regarding its position during a match. For example, if x is a player and y a coach:

$request(x, y, play\ on\ the\ left, dl, TA) \Rightarrow create(x, SC(x, y, play\ on\ the\ left, dl, TA, inactive)) \quad (6')$

Although the coach has authority over the player for its position on the field, it does not have authority for everything. For example, in most teams of soccer agents, the authority of the coach concerns the tactics of the team and not the technical skills of the players.

3.2.3 Context-dependent obligations. As mentioned before, it is likely that there are obligations for an agent playing a role in an organization to do certain tasks. However, these obligations can be context-dependent, i.e., they are not active all the time, but only in specific contexts. For example, it is often impossible to specify a priori when a task must be executed, but only that it should be executed when certain conditions are fulfilled. It is often the case of the law officers in human societies: they have the obligation to arrest another agent only when that agent has violated a law. In other words, we do not say that there is a social commitment of an agent towards the organization to do a task, but a social policy specifying the conditions of creation of such a commitment. We call this a *role obligations policy* (ROPol) and its definition is very similar to the one for role authority:

$ROPol(x:R, OA, IA, (\forall t \in TT: request(OA, x, t, IA) \Rightarrow create(x, SC(x, OA, t, IA, inactive))) , active) \quad (7)$

Using our formalism, an agent's obligations towards the organization can be reduced to the authority a role (the *organizational agent*) has over the agent, or vice versa. However, we believe it is important to keep a distinction between two different concepts: the hierarchical structures of roles and the norms imposed to a role.

To continue with the previous example, let us assume that there is an obligation for a player to cover the place of a teammate that has left its normal position. While one of the goals of the coach is to use its authority in such situations, it is not always possible. During matches, a coach cannot always communicate with its players. We can specify this context-dependent obligation using a role obligation policy. Thus, whenever the TeamAgent asks an agent x that plays the role *player* to reposition itself to cover the position of a teammate, x will form the commitment to do so:

$request(TA, x, reposition, dl, TA) \Rightarrow create(x, SC(x, TA, reposition, dl, TA, inactive)) \quad (7')$

Note that in this example the TeamAgent is an abstract agent that represents the team of agents. The request to form the commitment can be implicit, i.e., behaviour-based communication, or it can be done by one of the players. In both cases, the agent is committed towards the *team* to reposition itself and not towards a specific agent.

3.3 Defining roles and organizations using social commitments

In the previous section we showed how one can use different types of social commitments and policies to define the expected behaviour of a role in an organization. However, something is still missing: a role cannot be defined only by the social commitments it has, but also by its deontic permissions and prohibitions. When an agent plays a role, it must gain access to some resources in order to fulfill the social commitments that come with its role. This access comes in two forms: either physical (the agent receives resources) or deontic (the agent receives the permission to use the resources or to execute actions). In our model we consider that anything that is not explicitly permitted is prohibited and thus it is not necessary to specify the prohibitions imposed to an agent. Thus, a role can be defined as a 6-tuple:

$$Role = \langle name, Res, Perm, RGComm, RAPol, ROPol \rangle \quad (8)$$

where:

- *name* is the identifier of the role
- *Res* is the set of resources received by an agent that plays the role
- *Perm* is the set of resources and tasks for which an agent receives permissions when playing the role
- *RGComm* is the set of role goal commitments associated with the role – the social commitments that an agent has when playing the role
- *RAPol* is the set of role authority policies for which the role is either the debtor or the beneficiary – the social policies describing the hierarchy of roles
- *ROPol* is the set of role obligation policies associated with the role – the social policies describing the context-dependent obligations of the role.

In the example of soccer playing agents, we considered the role *player*. This role can be defined by specifying its resources (e.g., equipment), its permissions and prohibitions (e.g., to communicate only with teammates), its goals (e.g., to win matches – Formula 5'), the authority relationships that concern it (e.g., the coach has authority over it – Formula 6') and its obligations (e.g., to reposition itself – Formula 7').

We can now specify what it is generally called an organization as a collection of roles defined as above. The organization will contain both the organizational structure (the hierarchy of roles) and the norms regulating agents' behaviour, everything being defined using only one concept: the social commitment. The biggest advantage of using this approach to define organizations is that it makes no assumptions on the agents' internal model. In order to understand the organization it belongs to and the constraints imposed on its behaviour by the role it plays,

an agent must simply be able to understand the social commitments (and policies).

We defined a role as a collection of different types of social (meta-)commitments. These commitments apply to an agent that plays a role (noted $x:R$). We can specify what it means that an agent plays a role using the same notion of social commitment. An agent plays a role in an organization when it forms the commitment towards the organization to play that role. We call this commitment the *role commitment* (RC):

$$x:R \equiv RC(x, OA, R, deadline, IA, active) \quad (9)$$

where R is a role defined like in Formula 8. When this commitment becomes active, all the commitments and policies that define the role become also active.

4. Institutional agent

We would like to point out that we make a distinction between an organization (a collection of roles as defined in the previous section) and an institution. In our view, an institution is an agent organization to which several coordination mechanisms have been added. In order to specify an institution, one must specify how the institutional decisions are made and by whom, e.g., who chooses the goals of the institution and how are they delegated to the institution members. Another interesting aspect is how the institution enforces the fulfillment of social commitments, e.g., how the violation of a social commitment is detected and how the associated sanctions are applied.

The definition of agent institutions and their utilization is an active area of research in the multi-agent community. Among the most relevant works in this field we can cite the Harmonia [19] and ISLANDER [7] frameworks. Due to space reasons we will not enter into more details here. In defining an institution, one must define who chooses the institutional goals, how the achievement of these goals is delegated to the institution's members, etc. An example of possible answers to these problems is the work of Tambe [18] concerning teams of agents. Another approach is the one of [6], where the authors use collective commitments or intentions – not the same with social commitments – to describe the goals of a group of agents and their collective achievement by the group's members.

The role of the Institutional Agent in our social commitment definition is the role of a witness and enforcer of the commitment. In the remainder of this section we describe the characteristics of an IA and we point directions on how such an agent can be designed.

4.1 Commitment enforcement by an IA

In our social commitments definition we used an institutional agent to enforce the fulfillment of these commitments. This means that whenever a social commitment is fulfilled, the IA will reward the debtor and whenever the commitment is violated, the debtor is punished by the IA. The IA may also apply sanctions for the cancellation of the commitment. Thus, in our work we are interested in a IA that:

- detects the change of status of a commitment (violated, fulfilled, cancelled)
- chooses the appropriate sanction (either positive – reward or negative – penalty)
- imposes this sanction to the agents, i.e., makes sure that the appropriate agent receives its reward or pays its penalty.

The detection of a status change of social commitment is the simplest among the three tasks mentioned above. Without loss of generalization, we can consider that the IA has access to all the created social commitments to which it plays the role of witness. This hypothesis is necessary because the IA must be aware of the existing commitments in order to ensure their fulfillment. The IA must then monitor all the changes in the environment, to verify whether a committed task has been done or not and thus to establish whether the commitment has been fulfilled or not. While this is not generally easy to do, at least the IA can detect the social commitments violated because their deadline or the one of their tasks has passed.

Choosing the appropriate sanction for the fulfillment or violation of a social commitment is not straightforward. An ontology of sanctions used in human societies is proposed by the authors of [14]. They propose that in artificial institutions the hypothesis of *strict liability* should be used: an agent that violates a social commitment pays a sum equivalent with the value of the task it was committed for. However, in the general case it is difficult to establish this value. Moreover, to have justice, the sanction imposed to an agent should depend on whether the commitment violation has been intentional (i.e., the agent did not want to fulfill it) or accidental (i.e., the external conditions did not allow it to fulfill its commitment). As this situation raises a lot of problems (even in human societies), we believe that in artificial institutions it is better to use the strict liability hypothesis.

This hypothesis does not state only that the amount of the sanction should reflect the value of the task committed for, but also that the agent that violated the commitment *pays* the sanction. It is obvious that it is not easy to impose sanctions to autonomous agents: how to make sure they pay them? Both the authors of [7] and [14] propose a similar solution to this problem. In order to interact in an institution, an agent must include a special component, called *Dialog Manager* or *Governor*. This component

manages the social commitments of an agent and whenever a commitment is violated, the associated sanction is applied (internally).

It seems that up-to-date this is the best solution for the problem of imposing sanctions to autonomous agents, but other solutions can be envisaged too. For example, the sanctions can be in the form of decreasing the reputation of the commitment violating agents. As their reputation is a notion distributed among other agents, the violating agents cannot refuse the sanction. Another solution is to use a model existing in human societies. When a social commitment is violated, another commitment is automatically created to pay the associated sanction. If this second commitment is violated, another one is automatically created to pay a higher sanction. This goes on until a commitment that has an associated sanction of removing the violating agent from the system.

4.2 Example

In the previous sections we used an example of soccer playing agents and we showed how the role *player* can be defined using different types of social commitments and policies. These commitments are formed in front of an Institutional Agent that enforces their fulfillment. If a commitment or policy is violated, then its debtor should pay a penalty. One of the advantages of identifying different types of commitments is that this penalty may depend on the type of the violated commitment.

For example, a *player* has the role goal commitment to win a match. The sanction associated for the violation of this commitment could be a decrease of its reputation. A player is also subject of a role authority policy saying that it should obey its coach. If it violates this policy, then the sanction could be the exclusion from the team. Finally, there is a role commitment stating that when an agent plays a role, it forms several commitments and policies. If an agent fulfills this commitment, i.e., it forms the associated commitments/policies, then it can receive a positive sanction in the form of different resources (e.g., money) or permissions.

5. Conclusions

Social commitments have already been used to represent agent interactions and communications. One of their main advantages is that by using social commitments, one does not make any assumptions on the agents' internal model. One of the aspects that is less present in related work on commitment-based agent interactions is the organizational context of interactions. Agents act on the environment and interact with other agents while being part of an organization, i.e., while playing one or several roles in an organization. In an organizational context, the agents' behaviour is

constrained by the permissions and obligations they receive from the organization. Our objective is to obtain a unified model of reasoning on agent interactions in both organizational and non-organizational contexts. Towards this aim, in this paper we proposed a definition of roles and organizational structures based on social commitments.

In this paper we did not propose a new definition of social commitments, but we based our approach on several existing models proposed in related work. We enriched these models by using a generic representation of commitments' objects based on the TAEMS model. We also regrouped different concept present in related commitment definitions into one – the Institutional Agent. We analyzed its characteristics and we discussed several possibilities of designing such an agent.

An organization has been defined in this paper as a collection of roles, where a role is considered to be the subject of different types of social commitments and policies that were identified in this paper. By using the same concept – social commitments, we defined what it means that an agent plays a role in an organization. As future work, we intend to merge this work with existing works concerning commitment-based agent interactions to thus obtain a unified model of agent interaction in both organizational and non-organizational contexts.

References

1. Boella, G., van der Torre, L.: Attributing mental attitudes to normative systems. In *Proceedings of the 2nd Int. Conf. On Autonomous Agents and Multi-agent Systems (AAMAS'03)*, ACM Press (2003), 942-943.
2. Boella, G., van der Torre, L.: Contracts as legal institutions in organizations of autonomous agents. In *Proceedings of the 3rd Int. Conf. On Autonomous Agents and Multi-agent Systems (AAMAS'04)*, IEEE Press (2004), 948-955.
3. Castelfranchi, C.: Commitments: From Individual Intentions to Groups and Organizations. In *Proceedings of the Int. Conf. on Multi-Agent Systems ICMAS95*, (1995), 41-48.
4. Dastani, M.M., Dignum, V., Dignum, F.: Role-Assignment in Open Agent Societies. In *Proceedings of the 2nd Int. Conf. On Autonomous Agents and Multi-agent Systems (AAMAS'03)*, ACM Press (2003), 489-496.
5. Dignum, M.V., Meyer, J-J.Ch., Weigand, H.: Towards an Organisational Model for Agent Societies Using Contracts. In *Proceedings of the 1st Int. Conf. On Autonomous Agents and Multi-agent Systems (AAMAS'02)*, ACM Press (2002), 694-695.
6. Dunin-Keplicz, B., Verbrugge, R.: Collective Intentions. *Fundamenta Informicae*, vol. 51(3) (2002), 271-295.
7. Esteva, M.: *Electronic Institutions: from specification to development*. Ph.D. Thesis, Technical Univ. of Catalonia (2003).
8. Fasli, M.: On Commitments, Roles and Obligations. In *Proceedings of the 2nd Int. Workshop of Central and Eastern Europe on Multi-Agent Systems (CEEMAS01)*, LNAI 2296, Springer-Verlag (2001), 93-102.
9. Gasser, L.: Organizations in Multi-Agent Systems. In *Pre-Proceedings of the 10th European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, Annecy, (2001).
10. Hubner, J., Sichman, J.S., Boissier, O.: A model for the structural, functional, and deontic specification of organizations in multiagent systems. In *Proceedings of 16th Brazilian Symposium on Artificial Intelligence (SBIA'02)*, Porto de Galinhas, Brazil (2002), 118-128.
11. Lesser, V., et al.: Evolution of the GPGP/TAEMS Domain-Independent Coordination Framework. *Int. J. of Autonomous Agents and Multi-Agent Systems*, vol. 9(1) (2004), 87-143.
12. Maudet, N., Chaib-draa, B.: Commitment-based and dialog-game based protocols – new trends in agent communication languages. *Knowledge Engineering*, 17(2) (2002), p.157-179
13. McBurney, S., Parsons, P., Wooldridge, M.: Desiderata for agent argumentation protocols. In *Proceedings of the 1st Int. Conf. On Autonomous Agents and Multi-agent Systems (AAMAS'02)*, ACM Press (2002), 402-409.
14. Pasquier, Ph., Flores, R., Chaib-draa, B.: Modeling flexible social commitments and their enforcement. In *Gleizes, M-P. et al. (eds.): Proc. of the 5th Int. Workshop Engineering Societies in Agents World ESAW05*, (2004), 111-116.
15. Robocup Competition. www.robocup.org.
16. Singh, M.: Commitments among autonomous agents in information-rich environments. In *Proceedings of the 8th European Workshop on Modelling Autonomous Agents in a Multi-agent World (MAAMAW)*, (1997), p.141-155.
17. Singh, M.: An Ontology for Commitments in Multiagent Systems: Towards a Unification of Normative Concepts. *AI and Law*, vol. 7 (1999), 97-113.
18. Tambe, M.: Towards Flexible Teamwork. *Journal of AI Research*, vol. 7 (1997), 83-124.
19. Vázquez-Salceda, J.: *The role of Norms and Electronic Institutions in Multi-Agent Systems*. Whitestein Series in Software Agent Technology, Birkhäuser Verlag AG (2004).



Ecole Nationale Supérieure des Mines de Saint-Etienne
Centre G2I
158, Cours Fauriel
42023 SAINT-ETIENNE CEDEX 2

www.emse.fr
