An Operational Approach to Norms in Artificial Institutions *

Francesco Viganò[1], Nicoletta Fornara[1] and Marco Colombetti[1,2]
[1]Università della Svizzera italiana, via Buffi 13, 6900 Lugano, Switzerland
[2]Politecnico di Milano, Piazza L. da Vinci 32, I-20133 Milano, Italy
{francesco.vigano, nicoletta.fornara, marco.colombetti}@lu.unisi.ch

ABSTRACT

The notion of artificial institution is crucial for the specification of open interaction frameworks where heterogeneous and autonomous agents enter and leave dynamically and interact to face problems in various fields, like for instance electronic commerce, business-to-business applications, and personal assistant applications. In our view the specification of artificial institutions requires a clear standard definition of some basic concepts: ontology, authorizations, conventions, and norms. In this paper we propose an operational approach to the definition of norms that is mainly based on the generation of commitments. Norms can be employed to verify if the interacting agents are behaving in accordance with the normative specification of the system. In particular we regard norms as event-driven rules that are fired by events happening in the system and then modify commitments affecting the agents having a certain role. Furthermore we will discuss the crucial differences between the notion of authorization and permission and how permissions, obligations, and prohibitions can be expressed in our model.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent systems*

General Terms

Design, Languages, Theory

Keywords

Artificial Institutions, Norms, Commitments, ACLs

1. INTRODUCTION

Artificial institutions are a model, inspired by human institutions [6], used to specify open software systems as technological extensions of human society. They can be employed to perform, by means of artificial agents, certain computational intensive tasks such as: establishing appointments, signing contracts, and carrying out commercial transactions for example by means of an auction. In particular a concrete *open* interaction framework for heterogeneous agents is the result of the reification of the abstract specification of a set of artificial institutions. Given that the system is open, that is, heterogeneous agents made by different designers may enter and leave the system dynamically, it is crucial that the concepts used for its specification are clearly and unambiguously defined and commonly accepted as a standard by all agent designers.

In our view the specification of artificial institutions requires a clear standard definition of some basic components: (i) the *ontology* of the context where the interaction takes place, which is characterized by institutional entities and institutional actions; (ii) *authorizations* to perform institutional actions; (iii) *conventions* to bind the concrete performance of an action with its institutional counterpart; (iv) *norms* that impose obligations, prohibitions and permissions for the agents that interact within the institution. Our model of norms will be presented in this paper, whereas for a complete treatment of the other components see [4].

In open systems norms play a crucial role because they: regulate the behavior of agents, and create expectations on the behavior of other agents [2, 8]. We propose an operational approach to the specification of norms based on the notion of *social commitment*, that is, on a concept whose use in the specification of agent communication is becoming increasingly common [1]. Thus an important advantage of our approach is its coherence with an existing specification of an Agent Communication Language (ACL) [3]. Therefore an agent able to reason on commitments would, as consequence, be able to reason both on the effects of communicative acts and on norms defined within a system.

2. NORMS

We regard norms as event-driven rules that, when are fired by *events* happening in the system and if certain contextual conditions are met, create or cancel *commitments* affecting a set of *liable agents* described by a suitable selection expression. In general, the collection of liable agents corresponds to the set of agents that play a given role in the institution.

2.1 Commitments

We regard a commitment as an entity with the following attributes: a *debtor*, a *creditor*, a *content*, and a *state*, used to keep track of the temporal evolution of the commitment. In particular the content of a commitment can be *undefined*, *true*, or *false*. When the content is no longer *undefined*, as a consequence of the occurrence of a domain event, the state of that commitment is automatically set to *fulfilled* if the content has become *true*, otherwise it is set to *vio-*

*Supported by Swiss National Science Foundation project 200021-100260, "An Open Interaction Framework for Communicative Agents"

AAMAS'05, July 25-29, 2005, Utrecht, Netherlands.

lated. Moreover suitable institutional actions are defined to create, accept, and cancel commitments. For a complete description of our model of commitment see [3, 4].

2.2 Events

We represent events as entities of the system having attributes that provide information about the state transition that caused them. In our formalization we have singled out three main categories of interesting events that may activate norms: (i) a *TimeEvent*, that occurs when the system reaches a certain instant of time; (ii) a *ChangeEvent*, that happens when an institutional entity changes in some way, for example an attribute has changed its value; (iii) an *ActionEvent*, that happens when an agent perform an action, for instance the act of sending a message.

Event templates are event types with some restrictions on certain attributes that describe a set of possible event occurrences. They are used in the *on* section of a norm; when an event matches the given descriptor, the corresponding norm is fired and its variable *e* is filled with the event that activated it.

2.3 Definition of norms

When an agent fills a role in an institution, we assume that it accepts that norms create commitments binding the agent to a pseudo-agent representing the institution, which we call *institutional agent*. Such agent allows us to keep trace of commitments created by a certain instance of an institution, which also means that commitments created by norms of an institution can be cancelled only by norms defined by the same institution; this is because only the creditor of a pending commitment can set it to cancelled [4]. The general structure of a norm can be described as follows:

within $context\_name$: $ientity$
on e: $event\ template$
if $contextual\ conditions$ **then**
  foreach $agent$ **in** $liable\ agent$
  do $\{commitmentActionDescription(agent, institutional\_agent, parameters)\}^{+}$

Using norms, institutions can regulate in an uniform way both the communication protocol and protocol-independent normative aspects. Norms can be used to specify protocols, because they can dictate that in certain circumstances an agent ought to send a given type of message, or react to a message in a specific way. At the same time, norms can forbid the execution of institutional actions, in particular communicative acts, even if they are authorized. Furthermore, in correspondence of events that conclude the interaction process, norms can instantiate commitments to noncommunicative actions, like the payment of the purchased goods at the price negotiated during the interaction.

From our point of view, norms are not themselves commitments, but rules that manipulate commitments of the agents engaged in an interaction. In fact, norms are associated to roles rather than to individual agents, and strictly speaking they cannot be fulfilled or violated. Indeed, what can be fulfilled or violated is the commitment created by the application of a norm. Moreover by creating a new commitment whenever a norm is applied, we can compute how many times a norm has been violated or fulfilled.

3. NORMATIVE POSITIONS

A crucial aspect of our proposal is that commitments are used to represent fundamental normative positions [7] between agents, including as a special case the deontic relationships undertaken by the debtor through communicative acts [4]. In particular, commitments toward institutional agents are used to represent *obligations* (commitments to perform an action of a given type) and *prohibitions* (commitments to not to perform an action). Furthermore, we interpret the absence of positive or negative commitments to the execution of an action of a given type as *permissions*.

Usually in agent literature the concept of *authorization* is not distinguished from permission or the former encompasses the latter. Coherently with the concept of institutionalized power [5], we distinguish between the notions of authorization and permission. The main difference between authorization and permission resides in the effects of the action. Whereas the former represents a necessary condition for the execution of institutional action, permission represents the need to regulate the performance of authorized actions, but it cannot prevent the effects deriving from the performance of a forbidden act.

4. CONCLUSIONS

A crucial property of our approach is the possibility to verify if one or more agents have violated the norms of the system. That is, it is possible to keep track of the commitments generated by the activation of norms for all the agents having some role in the system. By identifying the violations of commitments it is possible to detect the violation of the associated norm.

Several research questions are still open, and will be tackled in our future work. In particular we are interested in verifying during the specification phase whether norms may create obligations to perform unauthorized actions, or under what conditions two norms may generate conflicting commitments. Finally, we plan to devise an explicit representation of the sanctions connected to the violation of a commitment.

5. REFERENCES

[1] F. Dignum, editor. *Advances in Agent Communication: International Workshop on Agent Communication Languages, ACL 2003, Melbourne, Australia, July 2003*, volume 2922 of *LNAI*. Springer, 2004.

[2] F. López y López and M. Luck. Modelling Norms for Autonomous Agents. In *Proceedings of Fourth Mexican International Conference on Computer Science*, pages 238–245, 2003.

[3] N. Fornara and M. Colombetti. A commitment-based approach to agent communication. *Applied Artificial Intelligence an International Journal*, 18(9-10):853–866, 2004.

[4] N. Fornara, F. Viganò, and M. Colombetti. Agent communication and institutional reality. In R. van Eijk, M. P. Huget, and F. Dignum, editors, *Agent Communication, Int. Workshop on Agent Communication AC 2004, New York, USA, July 2004*, volume 3396 of *LNAI*, pages 1–17. Springer, 2005.

[5] A. Jones and M. J. Sergot. A formal characterisation of institutionalised power. *Journal of the IGPL*, 4(3):429–445, 1996.

[6] J. R. Searle. *The construction of social reality*. Free Press, New York, 1995.

[7] M. Sergot. A computational theory of normative positions. *ACM Transactions on Computational Logic*, 2(4):522–581, 2001.

[8] W. W. Vasconcelos. Norm verification and analysis of electronic institutions. In *Declarative Agent Languages and Technologies (DALT) AAMAS'04 Workshop*, New York City, New York, USA, 2004.