Operationalisation of Norms for Usage in Electronic Institutions

Huib Aldewereld*
Frank Dignum
Institute of Information and Computing Sciences
Utrecht University
The Netherlands
{huib, dignum}@cs.uu.nl

Andrés García-Camino†
Pablo Noriega
Juan Antonio Rodríguez-Aguilar
Carles Sierra
Artificial Intelligence Research Institute, IIIA
Spanish Council for Scientific Research, CSIC
Campus de la UAB, Barcelona, Spain
{andres, pablo, jar, sierra}@iiia.csic.es

ABSTRACT

Agent-mediated electronic institutions belong to a new and promising field where interactions between a group of agents are regulated by means of a set of explicit norms. Current implementations of such open-agent systems are, however, mostly using constraints on the behaviour of the agents, thereby severely limiting the autonomy of the agents. To increase the autonomy of agents and possibly boost the efficiency of the overall system, a more flexible norm enforcement is required. However, as norms make extensive use of vague and ambiguous concepts and lack operational meaning (not expressing how the norm should be enforced), translating norms for usage with such a flexible enforcement mechanism might be difficult. In this paper we propose an extension to electronic institutions to allow for a flexible enforcement of norms, and manners to help overcome the difficulties of translating abstract norms.

Categories and Subject Descriptors

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

General Terms

Theory, Legal Aspects, Design

Keywords

Norms, Electronic Institutions, Normative Systems

*The first author of this paper was supported by the Netherlands Organisation for Scientific Research (NWO) under project number 634.000.017

†A. García-Camino was supported by the Spanish Council for Scientific Research (CSIC) by means of an I3P grant.

AAMAS'06 May 8–12 2006, Hakodate, Hokkaido, Japan.

1. INTRODUCTION

Agent-mediated institutions, introduced in [12, 13], are open agent systems that allow agents to enter and perform tasks. Because of the heterogeneous nature of the agents joining an electronic institution (e-institution), measures have to be taken to control and regulate the behaviour of these agents. These measures are needed to improve and guarantee the safety and stability of the system, as agents joining the institution might, (un)intentionally, break the system by behaving in non-expected or non-accepted manners. It has been widely accepted that norms can be used for ensuring this safety, since norms, which are vague and abstract in order to express various, different circumstances without the need for change, can be used for defining the legality and illegality of actions (and states) in e-institutions [3].

For these norms to be used in e-institutions, thereby regulating the agents joining the institution, enforcement mechanisms must be devised to implement them.Although the norms should be available to the agents joining the institution to allow them to work more efficiently in the regulated domain, it is not the agents reasoning and trying to adhere to the norms that provides the safety that the institution needs; it is the institution itself that has to ensure that this safety exists. This enforcement of norms, as discussed in [14], comes down to either: 1) defining constraints on unwanted behaviour, or 2) detecting violations of norms and reacting to these violations. The first approach is simpler and used by current implementations of e-institutions such as [5]. However, it seriously reduces the autonomy of the agents. To allow the agents in e-institutions more freedom and flexibility, while still complying to the norms, we would like to extend the implementation of e-institutions with the second approach to enforcement.

Previous work on normative systems (mainly focussed on deontic frameworks [10]) is mostly *declarative in nature*, while the implementation of norms and norm enforcement in e-institutions, as mentioned above, requires norms to have an *operational semantics* as well. Where the declarative nature of norms is necessary for reasoning about norms (reasoning about what is and what is not accepted), the operational semantics define how norms are to be implemented (e.g. what to do when norms are violated). Recent approaches on normative systems have begun to research

and express this operational meaning of norms, as seen in [14, 1, 11, 7]. These approaches represent norms and their operational meaning, but are not conclusive on how the implementation in an agent system, such as an e-institution, should be obtained. In this paper we are trying to bridge this gap, by proposing a translation from the (operational) approach proposed in [14] to elements usable for norm enforcement in AMELI ([5]).

In this paper we assume institutions to be defined as a set of norms, which are to be enforced by a distributed set of (internal) agents. Secondly we assume that the norms can sometimes be violated by agents in order to keep their autonomy, which can also be functional for the system as a whole as argued in [2]. The violation of norms is handled from the organisational point of view by violation and sanction mechanisms. And finally, we assume that the internal state of agents is neither visible, nor controllable from an institution's point of view, which, basically, means that enforcement of norms can only use the observable behaviour of agents to detect the violations.

2. IMPLEMENTING NORMS

E-institutions, as viewed here [4, 12, 13], are agent environments that restrict the behaviour of agents to ensure that agents interact in safe conditions. E-institutions constrain agent behaviour by defining the valid sequences of (dialogical) interactions that agents can have to attain their goals.

In order to be able to handle violations of norms we need a manner to express the violations that can occur in an institution, for which we introduce *integrity constraints* (derived from [6]):

DEFINITION 1. *Integrity constraints are first-order formulas of the form*

$$\left(\bigwedge_{i=1}^{n} uttered(s_i, w_{k_i}, \boldsymbol{i}_{l_i}) \wedge \bigwedge_{j=0}^{m} e_j\right) \rightarrow \bot$$

where s_i are scene identifiers or variables, w_{k_i} is a state k_i of scene s_i or a variable, $\boldsymbol{i}_{l_i}$ is an illocution scheme l_i matching the schema labelling an outgoing arc from w_{k_i} and e_j are boolean expressions over variables from uttered predicates.

Integrity constrains define sets of states that *should not* occur within an e-institution, i.e. if grounded illocutions matching the illocution schemes $\boldsymbol{i}_{l_1}, \ldots, \boldsymbol{i}_{l_n}$ are uttered in the corresponding scene states, and expressions $e_1, \ldots, e_m$ are satisfied, then a violation occurs ($\bot$).

This, however, is not enough to implement norms, as we need to define the reactions to the violations of the norms as well. These reactions (normally seen as sanctions and repairs, [14]) are expressed by using *dialogical constraints*:

DEFINITION 2. *Dialogical constraints are first-order formulas of the form:*

$$\left(\bigwedge_{i=1}^{n} uttered(s_i, w_{k_i}, \boldsymbol{i}^*_{l_i}) \wedge \bigwedge_{j=0}^{m} e_j\right) \Rightarrow \left(\bigwedge_{i=1}^{n'} uttered(s'_i, w'_{k_i}, \boldsymbol{i}'^*_{l_i}) \wedge \bigwedge_{j=0}^{m'} e_j\right)$$

where s_i, s'_i are scene identifiers or variables, w_{k_i}, w'_{k_i} are variables or states of scenes s_i and s'_i respectively, $\boldsymbol{i}^*_{l_i}$, $\boldsymbol{i}'^*_{l_i}$ are illocution schemes l_i matching the schema labelling an outgoing arc from w_{k_i} of scenes s_i and s'_i respectively, and e_j, e'_j are boolean expressions over variables from uttered predicates. These boolean expressions can include functions to check the state of the institution.

The intuitive meaning of a dialogical constraint is that, if the violation of the norm has been detected (expressed in the left-hand side of the constraint), the enforcers in the e-institution become obliged to perform a specific pattern of illocutions to punish the violator (this sanction is expressed in the right-hand side of the constraint).

The integrity constraints are then implemented in the infrastructure of the e-institutions, thereby providing the means to detect violations of norms, where the dialogical constraints are implemented in the enforcing agents which use them to determine the illocutions that should be uttered when a norm has been violated (thereby punishing the agent that violated the norm).

Semantics of the mentioned constraints, and an implementation for the AMELI platform can be given, but is omitted here due to space limitations.

3. CONTEXTUALISING NORMS

In the previous section we described how norm enforcement could be implemented in AMELI. However, the description of the norms should be given completely in terms of the utterances used in the e-Institution. However, usually norms are described in more abstract terms and need first to be "contextualised". The following norm, expressed in the formalism developed in [14]:

OBLIGED($(buyer$ DO $pay(Price,seller))$ IF $done(buyer,won(Item,Price)))$

can be translated into the utterances that would be used in an e-institution:

$uttered(payment,W,inform(A,buyer,B,payee,pay(Item,Price),T))$
$uttered(auction,w2,inform(C,auctioneer,A,buyer,won(Item,Price),T'))$

However, if we regard institutional norms that are derived (or translated) from human laws and regulations, the contextualisation becomes much harder, as laws contain vague and ambiguous concepts and cannot always be related to a single integrity constraint. In order to implement such norms with a high level of abstraction two steps must be taken: 1) interpreting the abstract concepts and link them to concrete concepts used in the institution (by using a *counts as* operator as described in [8, 9]), and 2) adding procedural information and artifacts to the institution to simplify (or allow) the enforcement of the norm.

The operational approach to norms developed in [14] tries to implement norms from an institutional perspective. It views norms as a way to describe how someone should behave. The norm frame used in this approach contains the *norm condition*, which defines obligations, permissions and prohibitions also known as the *declarative meaning of norms* (cf. [10]), of which an example was given above. However, since a system needs responses to the violations that occur, the norm frame also includes a definition of the responses to violations of the norms, which are known as sanctions and repairs (also known as the *operational meaning of the norm*). Finally, the norm frame includes the *violation condition* (obtained from the norm condition to express when a violation occurs) and the *detection mechanisms* (to describe the available resources on the platform that can be used for detecting violations).

Norms expressed in this formalism can be translated to

the constraints mentioned earlier. The integrity constraint is directly derivable from the violation condition of the formalism of [14]. Dialogical constraints are then formed by combining the violation condition and responses of the system to violations. For instance, if a norm with sanction S and repairs R (which can be obtained from the norm framework), can be translated to the integrity constraint $IC \rightarrow \bot$, the following rule can be created (automatically) to oblige the enforcers to utter the illocution (DC) that is derived from contextualising S and R:

$$IC \Rightarrow DC$$

Since these dialogical constraints are considered obligations to the enforcers, we can specify another integrity constraint to detect that this obligation has been violated, which is when the original norm was violated, but the enforcer did not punish the violating agent.

$$(IC \wedge \sim DC) \rightarrow \bot$$

4. CONCLUSIONS

With the development of electronic institutions, with the aim of implementing normative open-agent systems, comes the problem of ensuring the safety and stability of the system. Previous implementations of electronic institutions enforced norms by ensuring that the agents joining the system followed a pre-defined protocol, thereby guaranteeing norm compliance of the agents. As this approach severely limits the autonomy of the agents, a more flexible enforcement was desired. This paper proposes the use of integrity constraints and dialogical constraints to implement such a flexible enforcement of norms. This norm enforcement is based on the detection of and reacting to violations of norms.

In order for any kind of norm enforcement to be implemented, norms need to be expanded with an operational meaning, as the declarative nature of norms only defines what is legal/illegal, but never expresses how this legality/illegality is obtained/averted. In [14] we introduced several mechanisms for operationalising norms, where we annotated norms (expressed in deontic logic) with operational aspects, like sanctions and repairs. In this paper we have used this normative frame to design an implementation scheme usable for implementing norm enforcement in electronic institutions. However, before norms can be implemented using this scheme, the norms need to be contextualised. This contextualisation amounts to 1) connecting the abstract concepts of the norm to the concrete concepts used in the institution, and 2) extending the norm with additional procedural information before attempting to implement it. The contextualisation of the norms is, in fact, a further operationalisation of the norms, where, in contrast to declarative norms (which never change the world), the second step of this operationalisation changes the world in order to enforce the norm.

Acknowledgements

This paper was partially supported by the Spanish Science and Technology Ministry as part of the Web-i-2 project (TIC-2003-08763-C02-00).

5. REFERENCES

[1] A. Artikis, L. Kamara, J. Pitt, and M. Sergot. A protocol for resource sharing in norm-governed ad hoc networks. In *Proc. of the Declarative Agent Languages and Technologies (DALT) workshop*. Springer, 2004.

[2] C. Castelfranchi. Formalizing the informal?: Dynamic social order, bottom-up social control, and spontaneous normative relations. *Journal of Applied Logic*, 1(1-2):47–92, February 2003.

[3] F. Dignum. Abstract norms and electronic institutions. In *Proc. of the Int. Workshop on Regulated Agent-Based Social Systems: Theories and Applications (RASTA '02), Bologna*, pages 93–104, 2002.

[4] M. Esteva. *Electronic Institutions: from specification to development*. Number 19 in IIIA Monograph Series. PhD Thesis, 2003.

[5] M. Esteva, J. Rodríguez-Aguilar, B. Rosell, and J. Arcos. AMELI: An Agent-based Middleware for Electronic Institutions. In *3rd Int. Joint Conf. on Autonomous Agents and Multi-agent Systems*, New York, US, July 2004.

[6] M. Esteva, W. Vasconcelos, C. Sierra, and J. Rodríguez-Aguilar. Verifying norm consistency in electronic institutions. In *Proceedings of The AAAI-04 Workshop on Agent Organizations: Theory and Practice (ATOP)*, San Jose, California, July 2004.

[7] A. García-Camino and J. Rodríguez-Aguillar. Implementing norms in electronic institutions. In *Proc. of the 4th Int. Joint Conf. on Autonomous Agents & Multi Agent Systems (AAMAS-05)*, Utrecht, The Netherlands, July 2005.

[8] D. Grossi, H. Aldewereld, J. Vázquez-Salceda, and F. Dignum. Ontological aspects of the implementation of norms in agent-based electronic institutions. *Computational and Mathematical Organization Theory*, to appear in 2006.

[9] D. Grossi, F. Dignum, and J.-J. Ch. Meyer. Contextual taxonomies. In J. Leite and P. Toroni, editors, *Proceedings of CLIMA V Workshop, Lisbon, September*, pages 2–17, 2004.

[10] A. Lomuscio and D. Nute, editors. *Proc. of the 7th Int. Workshop on Deontic Logic in Computer Science (DEON04)*, volume 3065 of *LNCS*. Springer Verlag, 2004.

[11] F. López y López and M. Luck. Towards a Model of the Dynamics of Normative Multi-Agent Systems. In G. Lindemann, D. Moldt, M. Paolucci, and B. Yu, editors, *Proc. of the Int. Workshop on Regulated Agent-Based Social Systems: Theories and Applications (RASTA '02)*, volume 318 of *Mitteilung*, pages 175–194. Universität Hamburg, 2002.

[12] P. Noriega. *Agent-Mediated Auctions: The Fishmarket Metaphor*. Number 8 in IIIA Monograph Series. PhD Thesis, 1997.

[13] J. A. Rodriguez-Aguilar. *On the Design and Construction of Agent-mediated Electronic Institutions*. Number 14 in IIIA Monograph Series. PhD Thesis, 2001.

[14] J. Vázquez-Salceda, H. Aldewereld, and F. Dignum. Norms in multiagent systems: From theory to practice. *International Journal of Computer Systems Science & Engineering*, 20(4):pages 225–236, 2005.