



Supporting Organizational Accountability inside Multiagent Systems

M. Baldoni¹, C. Baroglio¹, K. M. May², R. Micalizio¹, and S. Tedeschi²

¹Università degli Studi di Torino, Dipartimento di Informatica

²Università degli Studi di Torino

Objective

We present and analyze the problem of realizing an accountability-supporting platform in multiagent systems technology. With the aim of achieving accountability as a design property, we provide a few principles that characterize an accountability-supporting multiagent system. We provide, then, an accountability protocol to regulate the interaction between an agent, willing to play a role in an organization, and the organization itself.

Accountability as a Design Property

The concept of accountability requires the presence of a number of attributes. We distill them in the following few founding principles at the basis of the way to achieve organizational accountability **as a design property**.

Principle 1

All the collaborations and communications subject to considerations of accountability among the agents occur within a single scope that we call **organization**.

Principle 2

An agent can enroll in an organization only by playing a **role** that is defined inside the organization.

Principle 3

An agent willing to play a role in an organization must be aware of all the **powers** associated with such a role before adopting it.

Principle 4

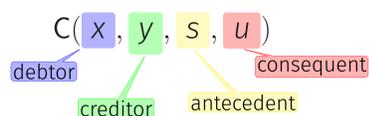
An agent is only accountable, towards the organization or another agent, for those **goals** it has explicitly accepted to bring about.

Principle 5

An agent must have the leeway for putting before the organization the **provisions** it needs for achieving the goal to which it is committing. The organization has the capability of reasoning on the requested provisions and can accept or reject them.

Social Commitments

A **social commitment** models a directed relation between two individuals [1], [2]:



Agent x commits to agent y that when condition s holds, it will bring about u

- Only the debtor can take on a commitment
- The creditor is not committed to bring about the antecedent
- Commitments are directly manipulable by agents via standard operations

An Accountability Protocol

Ag_i will commit towards the organization to exercise the powers, given to it by the role, when this will be requested by the legal relationships it will create towards other agents:

$$cpwr_{i,1} :: C(Ag_i, Org, C(Ag_i, Z_1, pwr_{i,1}), pwr_{i,1})$$

...

$$cpwr_{i,m} :: C(Ag_i, Org, C(Ag_i, Z_m, pwr_{i,m}), pwr_{i,m})$$

An agent willing to play a role is expected to create a commitment that takes the form:

$$cpwr_{R_i} :: C(Ag_i, Org, accept_player_{Org}(Ag_i, R_i), cpwr_{i,1} \wedge \dots \wedge cpwr_{i,m})$$

Org , then, promises to assign some goal to some agent should the agent accept to commit to pursue the goal:

$$cass_{i,1} :: C(Org, Ag_i, cg_{i,1}, prov_{i,1} \wedge assign_{Org}(Ag_i, g_{i,1}))$$

...

$$cass_{i,n} :: C(Org, Ag_i, cg_{i,n}, prov_{i,n} \wedge assign_{Org}(Ag_i, g_{i,n}))$$

$cg_{i,k=1,\dots,n}$ denote the commitments by whose creation the agent explicitly accepts the goals and possibly asks for provisions:

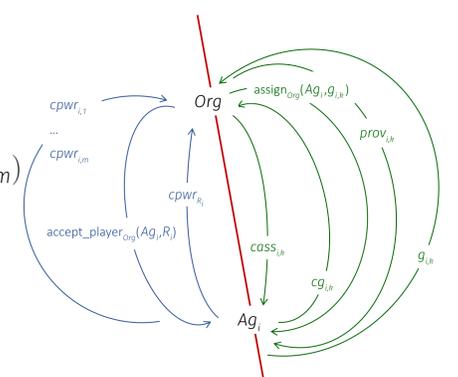
$$cg_{i,1} :: C(Ag_i, Org, prov_{i,1} \wedge assign_{Org}(Ag_i, g_{i,1}), achieve_{Ag_i}(g_{i,1}))$$

...

$$cg_{i,n} :: C(Ag_i, Org, prov_{i,n} \wedge assign_{Org}(Ag_i, g_{i,n}), achieve_{Ag_i}(g_{i,n}))$$

The resulting protocol is as follows:

- (1) $create(cpwr_{R_i})$
- (2) $accept_player_{Org}(Ag_i, R_i)$
- (3) $create(cpwr_{i,1}), \dots, create(cpwr_{i,m})$
- (4) $create(cass_{j,k}), k = 1, \dots, n$
- (5) $create(cg_{i,k}), k = 1, \dots, n$
- (6) $assign_{Org}(Ag_i, g_{i,k}), k = 1, \dots, n$
- (7) $prov_{i,k}, k = 1, \dots, n$
- (8) $achieve_{Ag_i}(g_{i,k}), k = 1, \dots, n$



Acknowledgements

This work was partially supported by the **Accountable Trustworthy Organizations and Systems (ATHOS)** project, funded by Università degli Studi di Torino and Compagnia di San Paolo (CSP 2014).

Contacts

- ① {matteo.baldoni,cristina.baroglio,roberto.micalizio}@unito.it
- ② {katherine.may,stefano.tedeschi}@edu.unito.it

References

- [1] R. Conte, C. Castelfranchi, and F. Dignum. Autonomous norm acceptance. In *Intelligent Agents V, Agent Theories, Architectures, and Languages, 5th International Workshop, ATAL '98, Paris, France, July 4-7, 1998, Proceedings*, pages 99-112, 1998. doi: 10.1007/3-540-49057-4_7.
- [2] M. P. Singh. An ontology for commitments in multiagent systems. *Journal of Artificial Intelligence in Law*, 7(1):97-113, 1999.