

A Hybrid Reputation Model Based on the Use of Organizations ^{*}

Viviane Torres da Silva¹, Ramón Hermoso², Roberto Centeno²
viviane@fdi.ucm.es, and {ramon.hermoso, roberto.centeno}@urjc.es

¹ Department of Computer Systems and Computation, University Complutense

² Artificial Intelligence Group - DATCCIA, University Rey Juan Carlos
Madrid, Spain

Abstract. In this paper we present a hybrid reputation model focused on organizational structures that attempts to solve problems associated with both centralized and decentralized reputation models. Agents in our approach are able not only to evaluate the behavior of others and store reputations values but also to send such information to a centralized mechanism and ask for reputations to this one and to other agents. The main objective of our approach is to allow agents to reason about the reputation values that they receive. Therefore, together with the reputation values, agents store and send information about norms violated and fulfilled and about the facts that contributed to such behavior. Furthermore, this model provides two different types of reputations, as *service provider* that is related to the behavior of an agent while providing a service to other agents and as *reputation source* that is related to the behavior of an agent while providing reputation of others.

1 Introduction

Several reputation models have been proposed with the aim to make available the reputations of agents interacting in multi-agent systems (MAS). The centralized approaches [2, 3] provide mechanisms that aggregate the feedback about the behavior of the agents and make available their reputations. Agents executing in those systems are able to *a)* evaluate the behavior of others with whom they have interacted and *b)* provide testimonies to the reputation model about such behavior. This kind of models presents some problems, such as: *i) scalability*, since it could be necessary to store too much information and, what is more important, to monitor every interaction in the system, and *ii)* most of those models does not allow the agents to know the reputation providers, since usually this type of social information is not stored.

^{*} The present work has been partially funded by the Spanish Ministry of Education and Science under projects TIC2003-08763-C02, TIN2006-14630-C03-02 (FPI grants program) and TIN2006-15660-C02-01; by the Juan de la Cierva program; by Comunidad de Madrid under project S-0505/TIC- 407; by the Spanish project “Agreement Technologies” (CONSOLIDER CSD2007-0022, INGENIO 2010) and by Comunidad de Madrid and University Rey Juan Carlos under project URJC-CM-2006-CET-0300

On the contrary, the decentralized reputation models such as [7, 9, 15] emerged in order to solve the problems of centralized models. They usually have no central authority and allow agents to assess reputation values by asking other agents about their past experiences, so solving *i*. Furthermore, agents can reason about those that behave as *reputation source*, i.e., those that are requested to provide reputation about third parties (solving *ii*). However those mechanisms still present some problems, such as: *iii*) how agents find out “good” *reputation sources* - what can become a problem as difficult to solve as finding a “good” counterpart to interact with. And *iv*) how to give incentives (and what kind of incentive) to agents in order to share their opinions with others.

As we have pointed out, there exist some pros and cons from using both centralized and decentralized reputation models. Nevertheless, they also present a problem shared by both of them: the interpretation of reputation values is never provided when agents exchange reputation values (decentralized models) or even when they ask for them to the system (centralized models). Only the reputation values themselves are presented and no information about fulfilled and violated norms or about the facts that have contributed to such fulfillments or violations are shared. Since two different agents can evaluate the same situation in different ways, it is really a hard task to interpret the reputation values and distinguish trustworthy and untrustworthy agents without additional information about their behavior.

In this paper we propose a hybrid reputation system that uses a centralized and a decentralized mechanism by taking the benefits provided by them while trying to solve some of their drawbacks. The main characteristics of our approach are: *a*) agents are able to evaluate the behavior of other agents, store such information, and provide to the organizations (they are not forced to do this) the reputations evaluated and the reasons for such evaluations (violated and fulfilled norms and the facts that have violated or fulfilled the norms); *b*) organizations implement centralized mechanisms and, therefore, are able to store and to provide the reputation values and related information, and also the identification of the agents that have provided the information; *c*) together with the reputation values, agents store and send information about norms violated and fulfilled and about the facts that contributed to such behavior and in addition *d*) the model allows considering two different dimensions for reputation: agents’ reputation as *service providers* and agents’ reputation as *reputation sources*. While the reputation as *service provider* represents the degree of satisfaction an agent has obtained after performing an interaction, the reputation as *reputation source* is related to the degree of satisfaction an agent obtains after requesting reputation about other agents to a third party, it evaluates the behavior of an agent while providing information about the reputation of others. Our hybrid model tries to solve the underlined problems mentioned above by giving more semantic meaning to traditional reputation techniques.

Our approach is supported by the scope of *organizations*, where agents enact some *roles* in different *interactions* in order to achieve some *goals*. Some benefits can be obtained from using this approach (i.e. in [6]), since organizational struc-

tures provide a semantically richer information when dealing with reputation values and their reasons.

The paper is organized as follows. In Section 2 we present an overview of the proposed hybrid reputation model. Section 3 details the decentralized mechanism and Section 4 presents the centralized one. Section 5 presents some discussions about related work and Section 6 summarizes and points out some future work.

2 Hybrid Model Overview

Organizational approaches are more and more used in order to build Multi-Agent Systems (MAS) since they allow facing complex problems using simple abstractions [8]. Those abstractions can be concepts that structure relationships among organization members, such as *roles* that agents can play and *interactions* that agents can use to communicate to each other, and also constraints, such as *norms* that establish undesirable agents' behavior. In this paper an organization is specified by the following definition:

Definition 1 *Let $\mathcal{O} \equiv (\mathcal{R}, \mathcal{I}, \mathcal{ON})$ be an organization formed by the following essential elements:*

1. *A set \mathcal{R} of roles that are involved in these interactions and can be played by agents in \mathcal{O} .*
2. *A set \mathcal{I} of interactions available for agents within \mathcal{O} .*
3. *A set \mathcal{ON} of organizational norms that regulate the behavior of the agents playing roles in \mathcal{O} .*

A typical organization establishes a set \mathcal{R} of *roles* as positions that agents have to put themselves in order to achieve some specific goals while interacting with other agents. Therefore, organizations have to be capable of providing a set \mathcal{I} of *interactions* that can be used by agents to interact with each other. In addition, organizations can define a set \mathcal{ON} of *organization norms* that regulate agents behavior by establishing how agents are expected to fulfill their roles in the organizations in terms of *rights* and *duties*.

Although our model assumes that norms can sometimes be violated. Those norms describe actions that agents are *prohibited*, *permitted* or *obligated* [13] to do and the sanctions to be applied in the case of violations and rewards to be provided in the case of fulfillment [14].

Reputation mechanisms are well-known techniques to fight against unexpected behavior (i.e. norm violations) since they provide agents with relevant information about the trustworthiness of others. The aim of this work is twofold: *i*) to present a hybrid reputation model that, based on the organizational structures (i.e. roles, norms, ...), tries to improve the performance of the organization; and *ii*) to improve the accuracy of reputation mechanisms by using these structures, taking into account the semantics they provide.

2.1 Model architecture

In order to endow agents with a hybrid behavior concerning reputation mechanisms within the organization, we propose the architecture described in Figure 1. There we can observe:

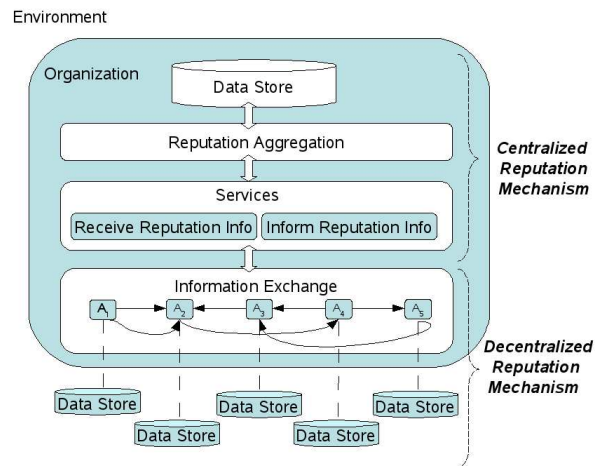


Fig. 1. Model architecture

- A *Centralized Reputation Mechanism* that contains the following entities:
 1. A service that allows agents to send information about others' reputation.
 2. A service that allows the organization to inform agents about others' behavior.
 3. A module that deals with the aggregation of information provided by agents.
 4. A central repository where the information is stored.
- A *Decentralized Reputation Mechanism* composed of the data agents gather by interacting with others.

2.2 Definition of model elements

Previous paragraphs pointed out our intention to present a reputation model based on the advantages provided by organizational structures. In this section we stress the relationship between the essential model elements defined in organizations and the agents' reputations.

Reputation of agents is usually associated with the behavior of an agent while playing a role in a given interaction. Therefore, we have defined the concept of

situation in order to relate agents, interactions and roles, three essential elements of organizations. A similar approach is described in [6, 5].

Definition 2 Let $S \equiv \langle \mathcal{A}_i, \mathcal{R}_j, \mathcal{I}_k, t \rangle$ be a situation where $\mathcal{A}_i \in \mathcal{A}$ is an agent among organization members, $\mathcal{R}_j \in \mathcal{R}$ represents a role enacted by \mathcal{A}_i and $\mathcal{I}_k \in \mathcal{I}$ represents an interaction performed by \mathcal{A}_i playing the role \mathcal{R}_j , at instant t .

A situation can be related to the violation or to the fulfillment of a norm. Both violations and fulfillments are associated with facts that were executed or with facts that should have been executed. In order to state such type of situations we have defined the concept of *(il)legal situation (I)LS*.

Definition 3 Let $(I)LS \equiv \langle S_l, \mathcal{N}_m, \mathcal{F}_n \rangle$ be a illegal situation *ILS* if the situation $S_l - \langle \mathcal{A}_i, \mathcal{R}_j, \mathcal{I}_k \rangle$ - entails a violation of a norm \mathcal{N}_m broken by \mathcal{A}_i due to some facts represented by \mathcal{F}_n or a legal situation *LS* if the situation S_l entails a fulfillment of a norm \mathcal{N}_m by \mathcal{A}_i due to the facts represented by \mathcal{F}_n .

Most of the reputation systems use quantitative values (opinions) to indicate the reputation of agents. However, such information is not sufficient to understand the behavior of the agents since such values are subjective, i.e. the same norm violation or fulfillment can be differently evaluated by two different agents. The subjective opinion of each agent about the same third party behavior could entail the problem of interpreting the meaning of the agent reputation. In order to tackle this, we propose a reputation model that not only takes into account a numerical value as the opinion an agent provides about a third party behavior, but also the set of norms that the latter has violated or fulfilled and the facts associated with them as a justification of the former's evaluation. This could be viewed as a single-step argumentation about how an agent evaluates the opinion about others.

Definition 4 Let $\mathcal{RI} \equiv \langle \mathcal{P}_q, (I)LS_r, Rep \rangle$ be a reputation information provided by the agent provider $\mathcal{P}_q \in \mathcal{A}$ about an *(il)legal situation (I)LS_r* $\in (I)LS$ associated with the reputation value *Rep*.

The reputation *Rep* of an agent can vary from $[-1, +1]$. Norm violations are described by illegal situations and are represented by negative reputation values $[-1, 0)$. Norm fulfillments are described by legal situations and are represented by positive reputation values $[0, +1]$.

In order to illustrate the need for stating the norm violated and the facts associated with such violation while informing the reputation of an agent, consider the following example. *Alice* is looking for a seller to purchase a new guitar. In order to choose the most trustworthy one, she decides to ask other agents for opinions about sellers that have sold guitars (or musical instruments) to them. Note that she is searching for opinions about similar situations such as $S_i = \langle Bob, Seller, SellInstruments, - \rangle$, where *Bob* is one of the available *sellers* with whom *Alice* is interested to interact using a *SellInstruments* interaction. In the case of seller *Bob* three agents have sent the following reputation values

$\langle -0.4, -0.1, -0.5 \rangle$. By analyzing such reputations she can only conclude that *Bob* has violated norms while interacting with those agents but she cannot understand why there exist different reputation values. She has received three different reputation values about the same agent in similar situations and she does not know what has happened when those agents have interacted with *Bob*.

If *Alice* had received not only the reputation values but also the information about the norms violated and the facts that have violated the norms, *Alice* would be able to understand the different agents' opinions. She would be able to understand that similar violations can be evaluated in different ways and diverse punishments can, thus, be applied by different reputation sources.

2.3 Types of reputation

In the previous section we have only presented the definition of *reputation* but we have not stressed the different types of reputation we consider in our approach. We distinguish between two different dimensions of reputation:

- *Reputation as a service provider*. This value represents the degree of satisfaction an agent has obtained after performing an interaction. This reputation reflects the fulfillment and violations of organizational norms (\mathcal{ON}) in an interaction playing a specific role.
- *Reputation as a reputation source*. This kind of reputation is related to the degree of satisfaction that an agent obtains after requesting opinions about others to a third party. The reputation of an agent as a reputation source evaluates the behavior of such agent while providing information about the reputation of others. If an agent has a bad reputation as *reputation source*, the opinions it has provided about other agents should not be trustworthy. The reputation of an agent as a reputation source is related to the role *reputation provider* and to a unique norm that prohibits the agent from lying when informing others about a counterpart's reputation.

It is fundamental to distinguish these two different dimensions of reputation since, on the one hand, the former deals with the quality, competence, availability, etc. of the agent which is requested for some kind of interaction to provide a service. On the other hand, the second reputation is calculated in order to measure how popular and accurate is another agent providing reputation information about third parties³. Both values are important depending on the information the agent needs in each moment.

2.4 Organizational and Individual Norms

At the beginning of this section we have defined an organization as an entity which is formed, among others, by a set of organizational norms (\mathcal{ON}) that regulate the behavior of agents in different situations.

³ Reputation as a reputation source could be seen as a particular case of the first type of reputation. It is the agent's reputation when providing the service of informing others about reputation of another one. We distinguish both types in order to make clear the different natures they have and the different ways of assess them.

Definition 5 Let $\mathcal{ON}(\mathcal{R}_j, \mathcal{I}_k)$ be an organizational norm applied to the agents playing the role $\mathcal{R}_j \in \mathcal{R}$ in the interaction $\mathcal{I}_k \in \mathcal{I}$. If \mathcal{R}_j is not specified, the norms are applied to all agents playing any role in the interaction \mathcal{I}_k . If \mathcal{I}_k is not defined, the norms are applied to all agents playing the role \mathcal{R}_j in any interaction. If neither \mathcal{R}_j nor \mathcal{I}_k are specified, the norms are applied to all members of the organization.

Note that when describing an (il)legal situation, the norm being mentioned in the situation must be a valid norm, i.e., must be a norm that is defined for the role being mentioned and for the interaction being described. From a social point of view - *macro level* - \mathcal{ON} are global norms commonly accepted by all organization members. They are imposed by the organization and are publicly advertised. On the other hand, from an individual point of view - *micro level* - agents may define their own norms by specializing the organization norms and making them more restrict. Such norms, called *individual norms*, reflect the relevance the global norms represent to the agents and cannot be used as a mechanism to modify or delete organization norms. In contrast to organizational norms, individual norms are not public but can be shared with anyone according to the agent decision. While evaluating the behavior of agents, an agent may consider both the organizational and individual norms. An individual norm that specializes an organizational norm \mathcal{ON}_x is also related to the same $(\mathcal{R}_j, \mathcal{I}_k)$ tuple defined by \mathcal{ON}_x , that is, the individual norm is applied to the same interaction and role as the organizational norm that it specializes.

Definition 6 Let $\mathcal{IN}_{\mathcal{A}_i}(\mathcal{R}_j, \mathcal{I}_k)$ be an individual norm applied by agent \mathcal{A}_i to the agents playing the role $\mathcal{R}_j \in \mathcal{R}$ in the interaction $\mathcal{I}_k \in \mathcal{I}$.

The reputation $Rep_{\mathcal{A}_i \rightarrow \mathcal{A}_j}$ resulting from applying agent \mathcal{A}_i 's assessment about other agent \mathcal{A}_j regarding to violations and fulfillment of norms is as follows:

- If \mathcal{A}_j has fulfilled the organizational norm⁴, then $Rep_{\mathcal{A}_i \rightarrow \mathcal{A}_j} \in [0, +1]$. The specific value will be generated by the agent according to the fulfillments of the individual norms. If no individual norm has been fulfilled then $Rep_{\mathcal{A}_i \rightarrow \mathcal{A}_j} = 0$.
- If the organizational norm has been fulfilled and so have the individual ones, then $Rep_{\mathcal{A}_i \rightarrow \mathcal{A}_j} = +1$.
- On the other hand, the violation of the organizational norm entails that $Rep_{\mathcal{A}_i \rightarrow \mathcal{A}_j} \in [-1, 0)$. The adjustment of the specific value calculated in this range will result from checking individual norms.
- If the organizational norm and the individual norms have been violated, then $Rep_{\mathcal{A}_i \rightarrow \mathcal{A}_j} = -1$.

By using this approach we allow agents, on the one hand, to specify at their will the reputation values to be associated with the agents behavior, and, on the

⁴ Here the organizational norm refers to that from which the individual norm is specialized.

other hand, to use a one-step argumentation when sending all the information related to a reputation value to others (included in the organization).

3 Decentralized Mechanism

In order to make use of our proposed reputation mechanism, the agents must be able to individually evaluate (as well as storing) the behavior of other agents with whom they have interacted and also of providing such information to the organization they belong to (centralized part) or to other agents if requested. Such evaluation must be made according to the set of organizational norms defined in the organization and also according to the set of individual norms defined by the agent itself.

3.1 Evaluating agents' behavior when providing a service

We are assuming that agents are able to interpret norms and also to identify that an action is violating a norm. In order to help agents in doing such tasks, approaches such as [10], can be used.

While evaluating the behavior of other agents, an agent must focus on the service being provided, i.e., on the interaction being checked and on the role being played by the other agent. A reputation value should be generated for each $\langle \mathcal{R}_j, \mathcal{I}_k \rangle$ tuple by separately considering each norm that regulates such interaction and/or role. Note that there may be norms applied to all agents regardless of the role being played and the interaction where the agent is participating. Those norms must also be considered.

An agent should start its evaluation by checking if the organizational norms have been fulfilled. The agent must check if the other agent has fulfilled the obligations and has not violated the prohibitions defined by the organization. The organizational norms are public to all agents executing in the organization. Every time an agent enters in the organization, it is informed about *i*) the norms to be fulfilled related to the role to be played, and *ii*) other norms applied to all agents in the organization regardless of the role they play. Agents must request for both types of norms to the organization before starting any interaction.

After verifying if the organizational norms have been fulfilled or violated⁵, the agent may verify if its individual norms were fulfilled, in the case they have been defined. The individual norms to be analyzed are those applied to the same $\langle \mathcal{R}_j, \mathcal{I}_k \rangle$ tuple and also those that do not depend on the role or interaction being analyzed.

To illustrate such idea, consider the following example. Let's suppose that *Alice* playing role *Flight Customer* has made a reservation of a flight ticket and that *Bob*, playing the role *Flight Provider*, has cancelled such reservation since the deadline for paying it has expired. The organization *Travel Agency* where

⁵ Note that verifying if a norm has been fulfilled is an objective action, while the interpretation of a violation of a norm - calculus of reputation - is clearly subjective.

both agents are involved has defined an organizational norm \mathcal{ON}_1 stating that *flight providers* can only cancel a reservation after the deadline for paying it.

\mathcal{ON}_1 : PROHIBITION FlightProvider EXECUTE ticket.cancelTicket IF
ticket.deadlineForPaying \leq TODAY

If *Bob* cancels the flight after the deadline for paying the reservation, he has not violated the organizational norm and its reputation will thus vary from $[0, +1]$. However, *Alice* can still be dissatisfied. *Alice* expects *Bob* to cancel the reservation only 5 days after the deadline, described by the individual norm \mathcal{IN}_{Alice_1} .

\mathcal{IN}_{Alice_1} PROHIBITION FlightProvider EXECUTE ticket.cancelTicket IF
ticket.deadlineForPaying+5 \leq TODAY

Such expectation characterizes an individual norm defined by *Alice*. If *Bob* cancels the reservation only 1 day after the deadline, he has violated such individual norm and, therefore, his reputation will be automatically lower than +1.

In order to be able to fulfil the individual norms defined by *Alice*, *Bob* can ask her about them. Note that we consider that all individual norms defined by *Alice* are specializations of organizational norms define by the *Travel Agency* organization, since they are more restricted.

3.2 Evaluating agents' behavior as reputation sources

The behavior of agents while providing information about the reputation of other agents is regulated by only one organizational norm. Such norm states that the agent playing role *ReputationSource* in the interaction *ReputationInformationExchange* cannot lie while providing that information. The norm that regulates such behavior is described by $\mathcal{ON}(\text{ReputationSource}, \text{ReputationInformationExchange})$ ⁶:

$\mathcal{ON}(\text{ReputationSource}, \text{ReputationInformationExchange})$: PROHIBITION
ReputationSource EXECUTE sendingWrongInformation

In contrast with other organizational norms, this is the unique norm that is domain-independent since it neither depends on the roles nor interactions defined in the organization.

The reputation of the agent as a *reputation source* can only be evaluated after the agent that is receiving the information has interacted with the desired agent (the third party it requested for). After the interaction, the agent that has received the information is able to evaluate if the reputation value it has received is consistent with the counterpart's behavior (from agent's individual

⁶ Although we are not describing sanctions to the norms used in this paper as examples, punishments and rewards can be specified.

view). This evaluation will form the reputation value for the agent that played the role *ReputationSource*.

Let's suppose *Alice* has never interacted with *Bob* before; then she asks *Carol* about how "good" is *Bob* playing role *Flight Customer*. After interacting with *Bob*, *Alice* evaluates his behavior and compares it with the information she has received from *Carol*. In the case the reputation values are similar, the reputation of *Carol* as a reputation source will be good, i.e., will be in the range $[0, +1]$. In the other case, *Carol's* reputation as reputation source will be bad from *Alice's* point of view, i.e., will be in the range $[-1, 0)$. The specific reputation value will depend on how different the reputation being informed is from the reputation being evaluated by the agent herself. To present the equation to be used by the agent is out of the scope of the paper, since we consider this is a function every agent should define by itself.

3.3 Sending and Receiving Agents' Reputations

Our proposed decentralized mechanism makes possible the sharing of reputation values between the agents. Agents can send and receive reputation values by using the *reputation information* (\mathcal{RI}) tuple defined in 2.2. Such information is composed by the following contents $\langle \mathcal{P}_q, ((\mathcal{A}_i, \mathcal{R}_j, \mathcal{I}_k, t), \mathcal{N}_m, \mathcal{F}_n), \text{Rep}_{\mathcal{P}_q \rightarrow \mathcal{A}_i} \rangle$:

- *Provider* (\mathcal{P}_q): the agent providing the evaluation about another agent's behavior;
- *(Il)legal Situation*: a legal or illegal situation evaluated according to a norm:
 - *Situation*: the situation being evaluated;
 - * *Client* (\mathcal{A}_i): the agent whose behavior is being evaluated;
 - * *Client role* (\mathcal{R}_j): the role that the client was playing and that is related to the behavior being evaluated;
 - * *Interaction* (\mathcal{I}_k): the relation between the provider and the client associated with the evaluation;
 - * *Time* (t): the time where the situation has occurred;
 - *Norm* (\mathcal{N}_m): the organizational or individual norm that is being considered. It can be a violated or fulfilled norm;
 - *Fact* (\mathcal{F}_n): the action(s) that was(were) executed (that is fulfilling the norm, in the case of an obligation, or violating the norm, in the case of a prohibition) or the action(s) that was(were) not executed (that is fulfilling the norm, in the case of an prohibition, or violating the norm, in the case of an obligation) during the interaction.
- *Reputation* ($\text{Rep}_{\mathcal{P}_q \rightarrow \mathcal{A}_i}$): a value that represents the evaluation about the agent's behavior;

In the case of reputations related to reputation sources, it is not necessary to state the norm since there is only one norm associated with such value, as stated in section 3.2. It is also not necessary to state the client role and interaction since it is always the same interaction type.

4 Centralized Mechanism

As we have introduced in section 2, our centralized mechanism should be implemented in organizations and, thus, be common to all participants in such organizations. The main objective of the mechanism is to receive the agents' past experiences as reputation values together with the information used to justify such value. The mechanism puts together such information and can provide it to any agent, including to new ones. The main advantage of our proposed centralized mechanism is its ability of pointing out the violated and fulfilled norms and the facts related to such violations and fulfillments while informing the agents' reputations. In addition, the mechanism is also able to inform who are the agents that have provided the information and also to inform about their reputations as reputation sources.

4.1 Sending and Receiving Agents' Reputations

Every time an agent evaluates the behavior of another one, it may send such evaluation as *reputation information* (\mathcal{RI}) to the organization. The organization is able to put together all reputations and provide them to any other agent.⁷ As a passive entity, the organization aggregates \mathcal{RI} tuples and makes rankings using such information. Therefore, agents cannot call the organization as an entity capable of lying, since it only acts as a means for publicizing information.

In order to illustrate that idea we show some examples of *reputation information* provided by the organization using the format of the tuple \mathcal{RI} : $\langle \mathcal{P}_q, ((\mathcal{A}_i, \mathcal{R}_j, \mathcal{I}_k, t), \mathcal{N}_m, \mathcal{F}_n), Rep_{\mathcal{P}_q \rightarrow \mathcal{A}_i} \rangle$

- Overall reputation of an agent \mathcal{A}_i as a service provider from the agent \mathcal{P}_q point of view. It gives a value representing aggregated opinions provided by \mathcal{P}_q about the agent \mathcal{A}_i in all situations in which it has participated regardless of the norms it has violated or fulfilled.

$$\langle \mathcal{P}_q, ((\mathcal{A}_i, -, -, -), -, \mathcal{F}_n), Rep_{\mathcal{P}_q \rightarrow \mathcal{A}_i} \rangle$$

- Reputation of an agent \mathcal{A}_i playing the role \mathcal{R}_j from the point of view of the agent provider \mathcal{P}_q . All reputation values provided by \mathcal{P}_q related to all situations where the agent \mathcal{A}_i is playing the role \mathcal{R}_j are grouped.

$$\langle \mathcal{P}_q, ((\mathcal{A}_i, \mathcal{R}_j, -, -), -, \mathcal{F}_n), Rep_{\mathcal{P}_q \rightarrow \mathcal{A}_i} \rangle$$

- Reputation of an agent \mathcal{A}_i while providing a specific service from the point of view of the agent provider \mathcal{P}_q . It gives the reputation of the agent in a given situation, i.e., while playing a role \mathcal{R}_j in an interaction \mathcal{I}_k . Such reputation value represents how trustworthy an agent is while providing a service from the point of view of the agent \mathcal{P}_q .

$$\langle \mathcal{P}_q, ((\mathcal{A}_i, \mathcal{R}_j, \mathcal{I}_k, -), -, \mathcal{F}_n), Rep_{\mathcal{P}_q \rightarrow \mathcal{A}_i} \rangle$$

⁷ Note that organizations must not modify the information that they receive.

- Overall reputation of an agent \mathcal{A}_i as reputation source from the point of view of the agent provider \mathcal{P}_q . It gives a value representing all opinions provided by \mathcal{P}_q about the agent \mathcal{A}_i while providing reputation values about other agents behavior, i. e. how the agent \mathcal{A}_i is regarding to the organizational norm \mathcal{ON}_l which describes that an agent cannot lie when providing reputation information.

$$\langle \mathcal{P}_q, ((\mathcal{A}_i, -, -, -), \mathcal{ON}_l, \mathcal{F}_n), Rep_{\mathcal{P}_q \rightarrow \mathcal{A}_i} \rangle$$

- Also, the organization could provide rankings⁸ such as the best agents providing any kind of service, the best agents being reputation sources, the agents with the highest *social prestige* (see section 4.2), the most collaborative agents based on the agents which send more opinions to the organization, etc.

Note that organizations are not limited to provide only the above mentioned services, other services can also be included. In addition, organizations can implement different algorithms to provide this information. For example, some can only consider the most recent reputations while others can consider all reputations they have received over time.

4.2 Agents Motivations

An agent can contribute to the organization by providing information about other agents as services providers and by providing information about other agents as reputation sources. The main goal of an agent on sending such information is to increase its *social prestige*, i.e., the image that the agent can offer about itself. The *social prestige* of an agent is increased when it advertises other agents about the ones they can trust as services providers. Such information helps then on distinguishing trustworthy agents from others. Its *social prestige* also increases when the agent provides information about the others that can be trusted as reputation sources. Such information helps the agents on distinguishing the ones that are sending trustful information about the services provided by other agents⁹.

In order to make public the *social prestige* of the agents, the organization provides a ranking of the best agents on sending trustful informations to the organizations (section 4.1). Since such ranking can be used by agents while selecting the ones to interact with, the definition of the concept of *social prestige* in our model motivates the agents to provide information to the organization.

5 Discussion

Our proposed approach is based on the benefits provided by the two most common mechanisms used for implementing reputation systems while trying to solve

⁸ All rankings are built based on the opinions received by the agents.

⁹ In order to avoid cheating agents increase their social prestige at the expense of others, agents should implement their aggregation functions - for received values - as an average value weighted by the reputation of the reputation source.

some of their drawbacks. The advantages of centralized mechanisms such as [3, 2] are *i)* the availability of the reputations to any agent in the system and *ii)* the use of global reputation values. Agents that have recently joined the organization or agents that have been playing roles can ask the centralized mechanism for reputations of other agents. It is not necessary to search for agents that could provide such information - since it is available in the organization -, or to ask for different testimonies in order to decrease the risk of misunderstanding - since the organization aggregates all provided reputation information about every agent. Some drawbacks of centralized mechanisms are (*C1*) they simply put the reputation values together - there is no reasoning about them, (*C2*) they may not be scalable, (*C3*) the agents that has provided the information might be unknown and (*C4*) there is not any information about what has influenced the agent reputation value.

Trying to solve the problems pointed out in *C2* and *C3*, some authors have proposed the use of decentralized mechanisms [1, 9, 15, 16]. However, those mechanisms had a serious problem: (*D0*) in order to know the reputation of a possible future partner, the agent should search for other agents that had previously interacted with the former, so providing their testimonies about such interactions. In large-scale MAS we can consider that the process of searching for those agents may be difficult and take a lot of time.

This problem was solved in decentralized mechanism that have adopted certified reputations [7, 4, 12]. Each agent has numeral certified reputations (or references) provided by the agents they have interacted with and, thus, can offer those references to the agents that intend to interact with it. The main advantages of those decentralized mechanism are *i)* the high availability of the reputations since any agent can easily learn the reputation of any other, *ii)* the fact that reputations are individually stored making the solution scalable and *iii)* the knowledge about the source agents that are providing the reputation values. The main disadvantages of such approaches are (*D1*) it may be difficult to find the agents that are trustworthy as reputation sources, (*D2*) agents must meet frequently in order to establish strong links among them and to pass on consistent reputations and (*D3*) there is not any information about what has influenced the agent reputation value.

In order to overcome some of those problems, the hybrid reputation system presented in [11] provides a centralized mechanism that may store the recently reputation values (solving problem *C2*) where agents can ask for reputations of other agents provided by several agents in different experiences (solving problem *D2* since one single agent does not need to meet frequently with other agents to be able to have a consistence evaluation of their behavior). In addition, the agents executing in this system are also able to evaluate the past behavior of other agents and store the reputation values. Different from the centralized approaches, the proposed centralized mechanism groups the reputations according to the violated norms and the roles being played (solving problem *C1*). The mechanism is able to provide the reputation of an agent considering a specific norm or one of the roles of the agent. But it is not able to provide information about the facts

that characterize the violations and about the agents that have provided the reputation values (it cannot solve problems *C3*, *C4*, *D1* and *D3*). Since agents make a personal evaluation about the behavior of other agents, it is fundamental to know the facts that violated the norms in order to interpret the reputation values.

The hybrid reputation model proposed in this paper solves the problems pointed out in this section. The agents using the decentralized mechanism are able to evaluate the behavior of other agents, to store such evaluation (contributing to solve problem *C2* since the centralized mechanism will probably store only the recent reputation information) and to provide to the organizations where they are involved information about their reputations, norms violated and fulfilled and facts violating and fulfilling the norms (solving problem *D3*). The organizations implement the centralized mechanisms that are able to store the (recent) reputation values (solving problem *C2*), the information related to such values (solving problem *C4*) and the identification of the agents that have provided them (solving problem *C3*). The centralized mechanism is able to provide such information (grouped in many different ways) to any agent in the system (solving problems *C1* and *D2*). The reputations are evaluated according to the norms that regulate a situation (role + interaction) where the agent being evaluated is involved. The model distinguishes between two different types of reputation: reputation as *service providers* and reputation as *reputation sources*. Therefore, it is easy to know the agents that are trustworthy as reputation sources (solving problem *D1*).

6 Conclusions

In this paper we have presented a hybrid reputation model for organizations which try to solve problems usually associated with centralized and decentralized reputation mechanisms.

The main advantages of our proposed model are: *i*) the reputation values are exchanged together with the violations and fulfillments of norms and the facts related which those violations or fulfillments (as a single-step argumentation process); *ii*) the organizations are able to provide non-manipulated information about the reputations of agents; *iii*) our proposed model distinguishes between two types of reputation: *reputation as a service provider* and *reputation as a reputation source*; and *iv*) the model motivates agents on sending information to the organization by making available a ranking stating the agents with highest *social prestige*.

As future work we plan to test our reputation model in some organizational domain such as travel agencies. We also intend to compare our model with other centralized or decentralized reputation models to observe the advantages of the hybrid proposals, as well as with other hybrid models to observe the benefits of our model opposite to others. Finally, we want to integrate our proposal in some proposed trust model in order to observe how it behaves working together with a confidence model.

References

1. A. Abdul-Rahman and S. Hailes. Supporting trust in virtual communities. In *Proceedings of the 33rd Hawaii International Conference on System Sciences*, volume 6, pages 294–301, 2000.
2. Amazon. <http://www.amazon.com>, 2008.
3. eBay. <http://www.ebay.com>, 2008.
4. T. Grandison and M. Sloman. A survey of trust in internet applications. *IEEE Communications Surveys & Tutorials*, 3(4), 2000.
5. R. Hermoso, H. Billhardt, R. Centeno, and S. Ossowski. Effective use of organisational abstractions for confidence models. In Oguz Dikenelli Gregory O’Hare, Michael O’Grady and Alessandro Ricci, editors, *Engineering Societies in the Agents World*, volume 4457 of *LNAI*, pages 368–383, Berlin, 2007. Springer.
6. R. Hermoso, H. Billhardt, and S. Ossowski. Integrating trust in virtual organisations. In Pablo Noriega and Javier Vázquez-Salceda, editors, *Coordination, Organization, Institutions and Norms in Agent Systems II*, volume 4386 of *LNAI*, pages 17–29, Berlin, In press. Springer.
7. T. Huynh, N. Jennings, and N. Shadbolt. Fire: An integrated trust and reputation model for open multi-agent systems. In *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI)*, pages 18–22, 2004.
8. A. Omicini and S. Ossowski. Objective versus subjective coordination in the engineering of agent systems. In Matthias Klusch, Sonia Bergamaschi, Peter Edwards, and Paolo Petta, editors, *Intelligent Information Agents: An AgentLink Perspective*, volume 2586 of *LNAI: State-of-the-Art Survey*, pages 179–202. Springer-Verlag, March 2003.
9. J. Sabater and C. Sierra. Reputation and social network analysis in multi-agent systems. In *Proceedings of First International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 475–482, 2002.
10. V. Silva. Implementing norms that govern non-dialogical actions. In J. Sichman, P. Noriega, J. Padget, and S. Ossowski, editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems III*, volume 4879 of *LNAI*, pages 232–244. Springer-Verlag, 2008.
11. V. Silva, F. Duran, J. Guedes, and C. Lucena. Governing multi-agent systems. *Journal of Brazilian Computer Science*, 2(13):19–34, 2007.
12. H. Skogsrud, B. Benatallah, and F. Casati. Model-driven trust negotiation for web services. *IEEE Internet Computing*, 7(6):45–52, 2003.
13. J. Vázquez-Salceda, H. Aldewereld, and F. Dignum. Implementing norms in multi-agent systems. *LNAI*, 3187:313–327, 2004.
14. F. López y López. Social power and norms: Impact on agent behavior. *PhD thesis, Univ. of Southampton, Faculty of Engineering and Applied Science, Department of Electronics and Computer Science*, 2003.
15. B. Yu and M. Singh. Distributed reputation management for electronic commerce. *Computational Intelligence*, 18(4):535–549, 2002.
16. B. Yu and M. Singh. An evidential model of distributed reputation management. In *Proceedings of First International Joint Conference on Autonomous Agents and Multi-Agent Systems*, volume 1, pages 294–301, 2002.