Combined On-line and Off-line Trust Mechanism for Agent Computing

Babak Khosravifar, Jamal Bentahar, Maziar Gomrokchi
Concordia University, Canada
b_khosr@encs.concordia.ca, bentahar@ciise.concordia.ca
m_gomrok@encs.concordia.ca

Philippe Thiran
University of Namur, Belgium
pthiran@fundp.ac.be

Abstract

In this paper, we propose an efficient mechanism dealing with trust assessment for agent societies, aiming to accurately assess the trustworthiness of the collaborating agents. In the proposed model, we formulate on-line assessment process, which applies a number of measurements affecting the evaluator agent's estimated value regarding to the evaluated agent's trust. We then represent the off-line trust assessment, which compares and adjusts the involved features with the actual performance of the evaluated agent. The off-line process updates the belief set of the evaluator agents in the sense that it adapts the agents with respect to the changes in the network. In this paper, the proposed framework is described, a theoretical analysis of its assessment and its implementation along with simulations comparison with other models are provided. We also show how our model is more efficient than the existing models, particularly in very dynamic environments.

Keywords. *Trust, Multi-Agent Systems, Agent Communication.*

1. Introduction

Over the past couple of years, agent communication languages and protocols have been of much interest in multi-agent systems. In such systems, autonomous agents are distributed in large scale network and mutually interact to collaborate, coordinate and share resources with each other. Therefore, trust is essential in effective interactions within open multi-agent systems [2], [1], [16], [8], [12]. Generally, an agent's trust in another is defined as the measure of willingness that the agent will make what he agrees to do. Attempting to maintain a trust-based network, different frameworks have been proposed in literature regarding to assess the trust agents have in one another. The most recent research areas in trust models are as follows: a) trust based on the direct interaction done between two parties [15], [14]; b) trust based on prior interactions [4], [5]; c) witness reputation, which is based on the reports provided by the third parties [11], [7], [13]; and d) certified reputation which is based on the certified reports provided by the agent being evaluated [2], [4].

In the proposed frameworks, generally the objective is to perform accurate trust assessment. But since agents are self-interested, it is hard to analyze an agent's likely behavior based on previous direct interactions. Moreover, the collected information from other agents may be unreliable and could lead to a non-accurate trust assessment. The main problem of these frameworks is that they do not act properly when selfish agents tend to change their behaviors. Therefore, they do not recognize the recent improvement or degradation in particular agent's capabilities. Considering these limitations, the trust models are aimed to act more precisely in terms of assessment and be adaptive to the environment inconsistency.

In this paper, we present an efficient assessment process for trust in a twofold contribution: on-line and off-line. In the first contribution, agents mutually interact and rate each other based on the interaction done (either satisfactory or dissatisfactory). The obtained ratings are accumulated to assess the direct interaction rating of a particular agent. Inter-agent communication is regulated by protocols (shared amongst agents and thus public) and determined by strategies (internal to agents and thus private). Upon evaluating an unknown (or not very well known) agent (*so called the trustee*), the evaluator agent (*so called the trustor*) is able to ask others (*so called consulting agents*) about their direct interaction ratings with the trustee agent. The consulting agents are composed of *trustworthy agents* (known by the trustor agent) and *referee agents* (introduced by the trustee agent), illustrated in figure 1. In the proposed framework, the trustor agent evaluates the credibility of the trustee by combining his own direct trust rating by the ratings provided by the consulting agents. The suggestions provided by the consulting agents are partially considered with respect to their time recency, interaction strengthen and accuracy. Taking these features into account, we represent a more accurate trust assessment. In the second contribution of this paper, the trustor agent after a period of direct interaction with the trustee agent performs an off-line assessment in order to adjust the credibility of the consulting agents (trustworthy and referee agents) that provided information regarding to the trust level of the trustee agent.

Compared to existing trust models for agent-based computing, our model is the first one equipped with this off-line

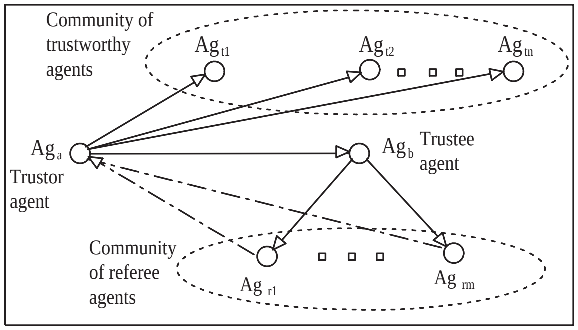

Figure 1. Overall trustworthy and referee agents protocol topology.

evaluation, which constitutes the maintenance process. In this process, the suggestions provided by consulting agents are compared with the observed behavior of the trustee agent. Exceeding some predefined thresholds, the trustor agent would either increase or decrease his belief about consulting agents. Doing so, gradually agents recognize more reliable consulting agents around in the network, which would cause a more efficient trust assessment process in future. Therefore, in our proposed model, agents would adapt themselves with new situations. In terms of game theoretical and mechanism design analysis [9], our model has the advantage of providing an incentive to the consultant agents, that are rational, to encourage them to reveal the truth when they send their ratings about other agents. Simply put, the incentive is computed in the off-line process in terms of increasing or decreasing the consultant's credibility according to the distance between the trustor's actual behavior and the revealed information by the consultants. In this paper, we analyze the effect of the maintenance process in different points of views and we compare the system efficiency with some other models.

The remainder of this paper is organized as follows. In Section 2, we define the on-line trust assessment process of our framework, which is composed of direct and indirect evaluation. In Section 3, we discuss the maintenance that a typical agent makes after a certain amount of time, since the interactions initiated. In Section 4, we outline the properties of our model in the experimental environment, represent the testbed, and compare our model results with two well-known trust models in terms of efficiency in trust assessment. Finally, Section 5 concludes the paper.

2. On-line Trust Estimation

In this section, we discuss the on-line evaluation process, in which the trustor agent collects some information and combines them to assess the credibility of the trustee agent. On-line evaluation is a twofold approach. In the former approach, the trustor only relies on what he has from his previous interactions with the trustee agent. In this case the number of interactions and the time recency of them would be considered (Section 2.1). In the later approach, the trustor prefers relying on the information provided by some other agents to assess a more accurate credibility for the trustee agent. Therefore, a number of consulting agents are to be selected, their credibility and the coherence of the information they provide is analyzed. We discuss these in depth in Section 2.2.

2.1. Direct Trust Evaluation

If the agents in the system know each other, this means they had a prior interaction history and consequently can compute the trust value of each other directly. Using their reasoning capabilities, agents evaluate the outcomes of their interactions. In the general case, they can evaluate these outcomes according to a scale of n types numbered from 1 (the most successful outcome) to n (the less successful outcome), such that the first m outcome types ($m < n$) are considered successful. Let $NI_{i\,Ag_a}^{\,Ag_b}$ be the number of interactions of type i that the trustor Ag_a had with the trustee Ag_b. Then, the trustor's estimated trust value for the trustee, $Tr_{Ag_a}^{Ag_b}$ can be computed by equation 1 as the ratio of the "*number of successful outcomes*" to the "*total number of possible outcomes*":

$$Tr_{Ag_a}^{Ag_b} = \frac{\sum_{i=1}^{m}(w_i \times \sum_{j=1}^{NI_{i\,Ag_a}^{\,Ag_b}} v_{ij})}{\sum_{i=1}^{n}(w_i \times \sum_{j=1}^{NI_{i\,Ag_a}^{\,Ag_b}} v_{ij})} \tag{1}$$

where w_i is the weight associated to the interaction type i and the v_{ij} is the measure taken into account in order to avoid two transactions with different values being treated equally (*i.e.* transactions with cost of $\$100,000$ and $\$100$). The weight w_i is application dependant and allows giving more importance to some interaction types than others. For example, considering the very successful interactions more important than the successful ones. In this particular case, agents will gain more trust by performing very good interactions (for example providing very good services). The factor v_{ij} is used to protect the model from attacks like reputation squeeze [3] in which one agent would obtain some positive ratings and make a bad interaction which actually makes a large damage.

In addition to the wight and measure of the interactions, another factor is used to reflect the *timely relevance* of these interactions. Since agents act in an environment, which is dynamic and may change quickly, it is more desirable to promote recent information and deal with out-of-date information with less emphasis. In our model, we assess this factor denoted by $TiR(\Delta t_{Ag_a}^{Ag_b})$ by using the function defined in equation 2. We call this function the timely relevance function.

$$TiR(\Delta t_{Ag_a}^{Ag_b}) = e^{-\lambda \; ln(\Delta t_{Ag_a}^{Ag_b})} \quad \lambda \geq 0 \tag{2}$$

$\Delta t_{Ag_a}^{Ag_b}$ is the time difference between the current time and the time at which Ag_a updates his information about Ag_b's trust (the interaction time). The intuition behind this formula is to use a function decreasing with the time difference. Consequently, the more recent the information is, the higher the timely relevance coefficient would be. Variable λ is an application-dependent coefficient. In some applications, recent interactions are more desirable to be considered. In that case, the trustor uses a higher value for λ to judge the credibility of the trustee. In contrast, in some other applications, even the old interactions are still valuable source of information. In that case, the trustor sets a relatively smaller value to λ. Considering the involved issues, we advance the equation 1 to the following to denote the direct trust estimation of a trustor agent Ag_a regarding to a trustee agent Ag_b based on interaction history between them:

$$Tr_{Ag_a}^{Ag_b} = \frac{\sum_{i=1}^{m}(w_i \times TiR(\Delta t_{Ag_a}^{Ag_b}) \times \sum_{j=1}^{NI_i{}_{Ag_a}^{Ag_b}} v_{ij})}{\sum_{i=1}^{n}(w_i \times TiR(\Delta t_{Ag_a}^{Ag_b}) \times \sum_{j=1}^{NI_i{}_{Ag_a}^{Ag_b}} v_{ij})} \tag{3}$$

2.2. Indirect Trust Evaluation

The second approach in on-line trust estimation of the trustee agent is to collect some information in terms of suggestions from some other agents. As indicated before, the consulting agents are divided into two groups: (1) the trustworthy agents, that the trustor agent Ag_a can rely on to request for information about the trustee Ag_b; and (2) the referee agents, that are introduced by the trustee agent as recommenders. In this section, we address the selection process of the consulting agents and how to deal with the information they provide in support of Ag_b.

Let $\mathcal{T}_{Ag_a}^{Ag_b}$ be the set of trustworthy agents that Ag_a knows from his belief set and that can report on Ag_b. Depending on the situation, how much Ag_a is aware of his surrounding environment and how restrictive Ag_a needs to be in consulting selection, a predefined trustworthy selection threshold (μ_T) is supposed to set in order to select a required number of trustworthy agents and fill the selected trustworthy set $\mathcal{T}s_{Ag_a}^{Ag_b}$. Basically, the elements of this set are the agents that are going to be asked about the credibility of the trustee Ag_b.

$$\mathcal{T}s_{Ag_a}^{Ag_b} = \{Ag_t \in \mathcal{T}_{Ag_a}^{Ag_b} | Tr_{Ag_a}^{Ag_t} > \mu_T\}$$

Another set to be involved in the evaluation process is the set of referee agents, which are introduced by Ag_b. Upon request from Ag_a, Ag_b replies by a list of the referee agents that he knows ($\mathcal{R}_{Ag_a}^{Ag_b}$). Following the same restriction policy by the predefined threshold μ_R, Ag_a consequently selects the appropriate referee agents ($\mathcal{R}s_{Ag_a}^{Ag_b}$). The elements of $\mathcal{R}s_{Ag_a}^{Ag_b}$ are the selected referee agents that Ag_a would consider their suggestions about Ag_b. However, there are some other referees that are introduced by Ag_b but somehow, because of not being reliable or unknown, they have not been selected to be asked about Ag_b (let us denote this set $\mathcal{R}s'^{Ag_b}_{Ag_a}$). In this case, Ag_a does not consider these agents' suggestions about Ag_b, but he saves them anyways in order to compare by the real behavior Ag_b performs after starting interaction with Ag_a. Thus, the referee is better known by Ag_a from now on and his trust level is calculated by the adjustment of what he provided and Ag_b's real behavior. The mentioned sets are defined as follows:

$$\mathcal{R}s_{Ag_a}^{Ag_b} = \{Ag_r \in \mathcal{R}_{Ag_a}^{Ag_b} | Tr_{Ag_a}^{Ag_r} > \mu_R\}$$

$$\mathcal{R}s'^{Ag_b}_{Ag_a} = \mathcal{R}_{Ag_a}^{Ag_b} - \mathcal{R}s_{Ag_a}^{Ag_b}$$

After selecting the proper consulting agents to ask about Ag_b, Ag_a asks each one of them about the rating he can provide. In the proposed framework, the obtained suggestions are partially considered in the total trust evaluation done by Ag_a. The first restriction factor is the time recency, which the affect is discussed in section 2.1 and derived from equation 2. The second restriction factor is denoted as the *relation strengthen*. We discuss its affect and how it is derived in the following. Let $NI_{Ag_x}^{Ag_y}$ be the total number of interactions between two agents Ag_x and Ag_y, which is computed by the equation 4 where n is the total number of outcome types.

$$NI_{Ag_x}^{Ag_y} = \sum_{i=1}^{n} NI_i{}_{Ag_x}^{Ag_y} \tag{4}$$

The total number of interactions between trustworthy agents Ag_t (resp. referee agents Ag_r) and the trustee Ag_b, $NI_{Ag_t}^{Ag_b}$ (resp. $NI_{Ag_r}^{Ag_b}$) is an important factor because it promotes information coming from agents knowing more about Ag_b. Generally, the agents that had high number of interactions with Ag_b are considered as good sources of information. The third factor considered by our trust model is the trustworthiness of the consulting agents. Considering his belief, Ag_a assigns a trust value for each consulting agent and thus restricts their contribution with respect to their relative trust value.

The trustor Ag_a derives the total trust estimation regarding to Ag_b using equation 5.

$$Tr_{Ag_a}^{Ag_b} = \frac{\Omega_T(\mathcal{T}s_{Ag_a}^{Ag_b}) + \Omega_R(\mathcal{R}s_{Ag_a}^{Ag_b})}{\Omega'_T(\mathcal{T}s_{Ag_a}^{Ag_b}) + \Omega'_R(\mathcal{R}s_{Ag_a}^{Ag_b})} \tag{5}$$

where

$$\Omega_T(\mathcal{T}s_{Ag_a}^{Ag_b}) = \sum_{Ag_t \in \mathcal{T}s_{Ag_a}^{Ag_b}} Tr_{Ag_a}^{Ag_t} \times Tr_{Ag_t}^{Ag_b} \times TiR(\Delta t_{Ag_t}^{Ag_b}) \times NI_{Ag_b}^{Ag_t}$$

$$\Omega'_T(\mathcal{T}s_{Ag_a}^{Ag_b}) = \sum_{Ag_t \in \mathcal{T}s_{Ag_a}^{Ag_b}} Tr_{Ag_a}^{Ag_t} \times TiR(\Delta t_{Ag_t}^{Ag_b}) \times NI_{Ag_b}^{Ag_t}$$

This equation takes into account the aforementioned factors, which are categorized as follows: (1) the trustworthiness of trustworthy/referee agents according to the point of view of the trustor Ag_a that is computed as agents' directed trust ($Tr_{Ag_a}^{Ag_t}$ and $Tr_{Ag_a}^{Ag_r}$ derived from equation 3); (2) Ag_b's trustworthiness according to the point of view of trustworthy/referee agents ($Tr_{Ag_t}^{Ag_b}$ and $Tr_{Ag_r}^{Ag_b}$); (3) the total number of interactions between these trustworthy/referee agents and Ag_b ($NI_{Ag_b}^{Ag_t}$ and $NI_{Ag_b}^{Ag_r}$); and (4) the timely relevance of interactions between trustworthy/referee agents and Ag_b ($TiR(\Delta t_{Ag_t}^{Ag_b})$ and $TiR(\Delta t_{Ag_r}^{Ag_b})$), as communicated by Ag_t/Ag_r to Ag_a. This equation is composed of two different terms representing the values got from two different consulting communities involved in trust evaluation. The functions Ω_T, Ω_R, Ω'_T and Ω'_R are used to compute the trustee's trust value by combining the trust values estimated by the trustworthy and referee agents. Let A be a set of agents, formally, these functions are defined as follows:

$$\Omega_T, \Omega'_T, \Omega_R, \Omega'_R : 2^A \rightarrow \mathbb{R}^+$$

Because of space limit, we have only defined the measure functions Ω_T and Ω'_T, which are related to the trustworthy agents, and Ω_R and Ω'_R are set likewise.

Following the ideology that Ag_a could to some extent rely on his own history interaction with Ag_b (direct trust evaluation approach) and partially use the second approach which is consulting other agents, Ag_a gives a 100% trustworthy rate ($Tr_{Ag_a}^{Ag_a} = 1$) to his history and considers himself as a member of his trustworthy community. This merging method takes into account the proportional relevance of each trust assessment, rather than treating them separately. Basically, the contribution percentage of each approach in the final evaluation of $Tr_{Ag_a}^{Ag_b}$ is defined regarding to how informative the history is in terms of the number of direct interactions between Ag_a and Ag_b and their time recency. Therefore, consultation with other agents is more considered if the history reflects lower uncertainty. Respectively, the higher uncertainty of the history makes the trustor uncertain and thus rely less on that history. Hence, consultation with other agents should be considered.

Generally, the second evaluation approach (consulting with others) is combined with the first evaluation approach (checking self history) to end up with an accurate trust estimation of the trustor agent Ag_a for the trustee agent Ag_b. To be more precise, we analyze the quality of the interactions of the trustee agent regarding to what expected (final trust evaluation $Tr_{Ag_a}^{Ag_b}$) and what is actually performed (so-called observed trust value $OTr_{Ag_a}^{Ag_b}$). To this end, we have a retrospect trust evaluation which is represented in Section 3.

3. Off-line Trust Estimation

3.1. Formulation

In the off-line trust evaluation we try to minimize the adverse affects the consulting agents may cause. For instance, two agents that have a strong relationship can support each other in trust evaluation and overestimate their trust level when they have been introduced as referee agents. That implicitly means Ag_a can expect more accurate suggestion from a consulting agent that had a large number of interactions with Ag_b comparing to the consulting agent who had less. Respectively, these agents should be affected more when the opposite of their suggestions turned out to be true. Generally, Ag_a is more confident about his trustworthy agents as they have shown an acceptable trustworthiness so far, but the referee agents are chosen by Ag_b, so we should always consider the possibilities like the cooperating partners may vote in favor of each other or competing agents may underrate their opponents. Therefore, Ag_a after a period of interacting with Ag_b performs maintenance in order to evaluate the accuracy of consulting agents' reports and therefore modify his belief about those agents' trust level. In addition, in evaluation process done by Ag_a, maybe there were some referee agents involved in, but their suggestions were not considered as they were not known by Ag_a. Consequently, these agents are not eligible to contribute in the trust evaluation process, but Ag_a does not discard their suggestions. This set of referee agents is denoted by $\mathcal{R}o_{Ag_a}^{Ag_b}$. After the maintenance, Ag_a would be able to estimate such referee agents' credibility as long as they are known (because of their referee history) by Ag_a from now on. To get into the maintenance procedure, we need to define some parameters in the following paragraphs.

Considering Ag_a as the trustor and Ag_b as the trustee, let $\mathcal{C}_{Ag_i}$ be the overall Ag_i's contribution in the trust evaluation of $Tr_{Ag_a}^{Ag_b}$. This contribution is defined as follows:

$$\mathcal{C}_{Ag_i} = \frac{Tr_{Ag_a}^{Ag_i} \times Tr_{Ag_i}^{Ag_b} \times TiR(\Delta t_{Ag_i}^{Ag_b}) \times NI_{Ag_b}^{Ag_i}}{\Omega'_T(\mathcal{T}s_{Ag_a}^{Ag_b}) + \Omega'_R(\mathcal{R}s_{Ag_a}^{Ag_b})} \quad (6)$$

Consequently, the percentage contribution of each consulting agent Ag_i would be computed as follows:

$$\mathcal{PC}_{Ag_i} = \frac{\mathcal{C}_{Ag_i}}{Tr_{Ag_a}^{Ag_b}} \quad (7)$$

Let $\overline{\mathcal{PC}}$ be the mean value of $\mathcal{PC}_{Ag_i}$ defined as follows:

$$\overline{\mathcal{PC}} = \frac{\sum_{Ag_i \in (\mathcal{T}s_{Ag_a}^{Ag_b} \bigcup \mathcal{R}s_{Ag_a}^{Ag_b})} \mathcal{PC}_{Ag_i}}{|\mathcal{T}s_{Ag_a}^{Ag_b} \bigcup \mathcal{R}s_{Ag_a}^{Ag_b}|} \quad (8)$$

Also, let $\mathcal{T}'s_{Ag_a}^{Ag_b}$ and $\mathcal{R}'s_{Ag_a}^{Ag_b}$ represent the set of trustworthy and referee agents, which perform high contribution. We define these sets as follows:

$$\mathcal{T}'s_{Ag_a}^{Ag_b} = \{Ag_t \in \mathcal{T}s_{Ag_a}^{Ag_b} \mid \mathcal{PC}_{Ag_t} > \overline{\mathcal{PC}}\}$$
$$\mathcal{R}'s_{Ag_a}^{Ag_b} = \{Ag_r \in \mathcal{R}s_{Ag_a}^{Ag_b} \mid \mathcal{PC}_{Ag_r} > \overline{\mathcal{PC}}\}$$

Rationally, Ag_a is more interested in the accuracy of information provided by the consulting agents who participated much in the evaluation process (agents belonging to $\mathcal{T}'s_{Ag_a}^{Ag_b}$ and $\mathcal{R}'s_{Ag_a}^{Ag_b}$) and those that could not contribute but their provided information is saved to assess in maintenance (agents belonging to $\mathcal{R}o_{Ag_a}^{Ag_b}$). Therefore, let $\mathcal{M}_{Ag_a}^{Ag_b}$ be the set containing all the agents that are to be involved in the maintenance process, which is defined as follows:

$$\mathcal{M}_{Ag_a}^{Ag_b} = \mathcal{T}'s_{Ag_a}^{Ag_b} \bigcup \mathcal{R}'s_{Ag_a}^{Ag_b} \bigcup \mathcal{R}o_{Ag_a}^{Ag_b}$$

Ag_a then adjusts the trust level for each agent Ag_i ($Ag_i \in \mathcal{M}_{Ag_a}^{Ag_b}$) by comparing the actual performance of Ag_b, as a result of a period of direct interaction experience ($OTr_{Ag_a}^{Ag_b}$), and what Ag_i provided as the suggested trust level for Ag_b ($Tr_{Ag_i}^{Ag_b}$). Therefore, thresholds ν_T and ν_R are associated as inaccuracy tolerance thresholds for the trustworthy and referee agents. We assign two different thresholds because the referee agents were supposed to deliver a more accurate information about Ag_b comparing to the trustworthy agents because they have been introduced by Ag_b. By doing so, if the difference is greater than the associated threshold for the agent, the consulting agent's trust level should be dropped to some extent, otherwise it will be enhanced regarding to the importance of suggestion provided. The incremental rate ϕ_i is defined by Ag_a. The trustworthiness level of the agent Ag_i is defined by the following equation:

$$Tr_{Ag_a}^{Ag_i} = \begin{cases} Tr_{Ag_a}^{Ag_i} \times \frac{NI_{Ag_b}^{Ag_i}}{NI_{Ag_b}^{Ag_i} + NI_{Ag_i}^{Ag_a}} \times (1 - D_i), & \text{if } D_i \geq \nu_R; \\ 1 - \phi_i, & \text{if } D_i < \nu_R. \end{cases}$$

$$D_i = |Tr_{Ag_i}^{Ag_b} - OTr_{Ag_a}^{Ag_b}|;$$
$$\phi_i = (1 - Tr_{Ag_a}^{Ag_i}) \times \frac{NI_{Ag_b}^{Ag_i}}{NI_{Ag_b}^{Ag_i} + NI_{Ag_i}^{Ag_a}} \times (1 - D_i) \tag{9}$$

The value D_i defines the inaccuracy of the trust level regarding to the specific consulting agent Ag_i. The inaccuracy is checked by the predefined threshold (here for referee agents ν_R and respectively for the trustworthy agents, ν_T would be used) to recognize whether the referee agent's suggestion was apart from the real value. If the inaccuracy is less than the predefined threshold, Ag_a increases the current trust level of Ag_i. Since the increment should not exceed 1, ϕ_i is used to be the decrease in the difference between the current trust value and 1. Indeed, decreasing the difference makes the trust value increase. The value ϕ_i, decreases regarding to the portion of the number of interactions done between the trustor Ag_a, the consulting agent Ag_i and the trustee Ag_b.

3.2. Off-line as an Optimization Problem

Strictly speaking, the objective of off-line process is to get the most accurate evaluation possible. This can be formalized as an optimization problem. In this problem we try to minimize the difference of the estimated trust and the observed value by changing the trust values provided by the consulting agents. These changes should respect the awards and penalties that the trustor agent would like to apply to some of the consulting agents. Therefore we define new constraints associated to each consulting agent that make such restrictions in the sense that after minimization, the consulting agents that are supposed to get higher trust value, cannot get decreased. Likewise the agents that are supposed to get less trust value, cannot get increased.

To define these constraints, let us first define the following vectors. Let $\underline{\mathcal{T}r}_{Ag_a}^{Ag_b}$ be a vector of size $|\mathcal{T}s_{Ag_a}^{Ag_b} \bigcup \mathcal{R}s_{Ag_a}^{Ag_b}|$, that contains the trust rates assigned by the trustor Ag_a to each consulting agent Ag_i involved in the trust assessment of Ag_b done by Ag_a ($Tr_{Ag_a}^{Ag_i}$). Also consider $\underline{\mathcal{CM}}_{Ag_a}^{Ag_b}$ as the vector of measures involved when computing the contribution of consulting agents in trust evaluation of Ag_b, except Ag_a's trust rate for the consulting agents. We note that each measure $\mathcal{CM}_{Ag_i}$ (each element of the vector) remains fixed in the maintenance process. Therefore, we have separated it from the partial rate considered by the trustee agent for the consulting agent Ag_i. The measure is derived as follows:

$$\mathcal{CM}_{Ag_i} = \frac{Tr_{Ag_i}^{Ag_b} \times TiR(\Delta t_{Ag_i}^{Ag_b}) \times NI_{Ag_b}^{Ag_i}}{\Omega'_T(\mathcal{T}s_{Ag_a}^{Ag_b}) + \Psi'_R(\mathcal{R}s_{Ag_a}^{Ag_b})} \tag{10}$$

The vector $\underline{\mathcal{T}r}_{Ag_a}^{Ag_b}$ is subject to change in the minimization function when resolving the optimization problem. For that reason, we save a copy of $\underline{\mathcal{T}r}_{Ag_a}^{Ag_b}$ in another vector called $\underline{\mathcal{CT}r}_{Ag_a}^{Ag_b}$ (with elements denoted by $\mathcal{CT}r_{Ag_i}^{Ag_b}$) to keep the current trust rates provided. For simplicity reasons, we only consider the case of referee agents $Ag_r \in \mathcal{R}s_{Ag_a}^{Ag_b}$. The same analysis could be applied for the trustworthy agents $Ag_t \in \mathcal{T}s_{Ag_a}^{Ag_b}$. The underlying constraints of the optimization problem would be obtained as follows:

$$\begin{cases} \textbf{add } \textit{constraint } Tr_{Ag_r}^{Ag_b} > \mathcal{CT}r_{Ag_r}^{Ag_b}, & \text{if } D_R < \nu_R; \\ \textbf{add } \textit{constraint } Tr_{Ag_r}^{Ag_b} < \mathcal{CT}r_{Ag_r}^{Ag_b}, & \text{if } D_R \geq \nu_R. \end{cases}$$

where: $\quad D_R = |CTr_{Ag_r}^{Ag_b} - OTr_{Ag_a}^{Ag_b}|$

Due to the fact that Ag_a only gives one chance to the consulting agents to release their suggestions, $\underline{\mathcal{CM}}_{Ag_a}^{Ag_b}$ and $OTr_{Ag_i}^{Ag_b}$ are assumed to be constant. Therefore, by the information provided, the vector $\underline{\mathcal{T}r}_{Ag_a}^{Ag_b}$ denotes the partial ratings of consulting agents. Respecting the direction of possible change for each agent, now we are able to perform a maintenance step in order to minimize the error of the combined partial ratings. To minimize the total error, we perform one iteration of the *steepest descent* minimization method to get the appropriate direction towards the minimum, together with the proper step size for updating the

Table 1. Simulation summarization over the obtained measurements.

	S.P. Agent Type	Density in S.P. Community	Provided Utility at Each RUN	
			Range	Standard Deviation
Service Provider Agents (S.P.)	Good	15.0%	]+5, +10]	1.0
	Ordinary	30.0%	]-5, +5]	2.0
	Bad	15.0%	]-10, -5]	2.0
	Fickle	40.0%	[-10, +10]	-
	S.C. Agent Type	**Density in S.P. Community**	**Number of Joining Agents at Each RUN**	
Service Consumer Agents (S.C.)	Proposed Model	33.3%	10	
	Travos	33.3%	10	
	BRS	33.3%	10	

initially obtained $\underline{\mathcal{T}r}^{Ag_b}_{Ag_a}$. Doing this, we obtain the best sequence of ratings that could have been assigned by Ag_a to the consulting agents in the sense that the difference of the estimated trust ($Tr^{Ag_b}_{Ag_a}$) and the observed value ($OTr^{Ag_b}_{Ag_a}$) is minimized. The corresponding optimization problem is shown in equation 11:

$$min_{\underline{\mathcal{T}r}^{Ag_b}_{Ag_a}} \quad \underline{\mathcal{CM}}^{Ag_b}_{Ag_a} \times {\underline{\mathcal{T}r}^{Ag_b}_{Ag_a}}^{T} - OTr^{Ag_b}_{Ag_a} \qquad (11)$$

subject to

List of obtained constraints;

The updated sequence of rates are assumed to be the best values the consulting agents could provide during the specific time period considered for the maintenance. Although these updates respect the derived constraints, there might be some anomalies in the sense that some particular consulting agents (most likely some with intermittent behavior) may get increased of their rates by some values while they actually do not deserve. These anomalies would be recovered by the consequent maintenance processes, which the trustor agent would make. This is because the intermittent behavior of those agent would affect the overcoming maintenance procedure, which may eventually cause the right direction of trust adjustment. Hence, by the consequent updates the trustor agent performs, he rates in such a way that the trust value of the consulting agents approach their actual behavior.

4. Experimental Results

In this section we describe the implementation of a proof of concept prototype. We also compare our proposed model with Travos [13] and BRS [7] trust models. In the implemented prototype, agents are implemented as $Jadex^{©TM}$ agents, i.e. they inherit from the basic class $Jadex - Simulator^{©TM}$ $Agent$. The agent reasoning capabilities are implemented as Java modules using logic programming techniques. The testbed environment (represented in table 1) is populated with two agent types: (1) service provider agents that are supposed to provide services (toward simplicity, we assume only one type of service is provided and therefore consumed); and (2) service consumer agents (equipped with the different trust models) that are seeking the service providers to interact with and consume the provided service.

The simulation consists of a number of consequent RUNs in which agents are activated and build their private knowledge and keep interacting with one another, gain utility and enhance their overall knowledge about the environment. Table 1 represents four types of the service providers we consider in our simulation: good, ordinary, bad and fickle. The first three provide the service regarding to the assigned mean value of quality with a small range of deviation. But fickle providers are more flexible as their range of quality covers the whole possible outcomes. To put the system in a more tight situation and make it open, we use high number of fickle agents and a certain number of agents to enter and leave to start interacting and thus, build their environment. Upon initiation of the simulation, some agents are initialized with some random parameters. However, some other agents and the ones that join the system within RUNs, are the agents that do not know any agent in the surrounding environment. These agents start interacting with their adjacent agents in order to build their connections to their surrounding network. After gaining enough connections, obviously they restrict their connectivity by evaluating the interacting agents. Due to space limit, we do not elaborate such scenarios. In this section, we compare the proposed model with Travos and BRS concerning how they survive in biased environment where agents constantly change their behaviors. Travos and BRS trust models differ from ours in the trust assessment mechanism and analysis they perform in order to choose the best possible provider. Therefore the utility gained by each model is considered as its efficiency in selecting reliable service providers. The experimental measurements of the comparison between these models are outlined in table 2. Travos and BRS are similar to the proposed model in the sense that they do consider other agents' suggestions while evaluating the trust in some specific agent and discard inaccurate suggestions aiming to adapt themselves to the environment inconsistency attitude. In BRS, the trustor agent evaluates the recommender agents' suggestions using beta distribution method and ignores the suggestions that deviate the most from the majority of the ratings. Concerning this, BRS is comparatively a static trust method, which causes a low-efficient performance in very dynamic and biased environments like open multi-agent systems. Cumulative gained utility vs. number of runs graph is shown in figure

Table 2. Results summarization over the obtained measurements.

Measurements and Characteristics	Proposed Model	Travos	BRS
No. of active agents in simulation	20	20	20
No. of RUNs in each simulation	300	300	300
Measured cumulative utility gained in five simulations	9,648	6,032	2,870
	9,721	6,736	1,678
	9,939	7,455	2,188
	9,652	5,909	1,573
	9,388	7,735	1,760
Average cumulative utility gained	*9,669.6*	*6,773.4*	*2,013.8*
Standard deviation of cumulative utility gained	176.23	733.14	476.42
Half value of confidence interval	218.52	909.09	590.76
Full interval with 95% confidence level	[9,451 - 9,888]	[5,864 - 7,682]	[1,423 - 2,604]

2. This model is not sensitive to an agile behavior change. This means that if a BRS agent decides to evaluate an agent that he is not acquainted with, he considers the majority of ratings, which are supposed to be truthfully revealed about the trustee agent. In such a case that the trustee agent is just changed his strategy, the trustor agent would loose in trust assessment and does not maintain any action to verify the accuracy of the gained information. It may take as much time that other agents perform a number of direct interactions to start rating about the spurious trustee agent. Therefore, as illustrated in figure 4, the BRS agents would have higher percentage of fickle providers selection and relatively less percentage of good providers selection (illustrated in figure 3). Generally, it would take more time for BRS agents to adapt themselves to the new environment conditions.

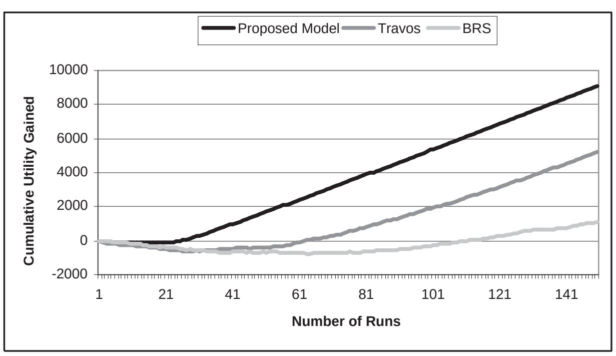

Figure 2. Comparison in terms of cumulative utility gained.

Travos [13] has a similar method like BRS. It also uses beta distribution to estimate the trustworthiness of an agent based on the previous interaction experience. Travos model also does not have partial rating. The trustor agent

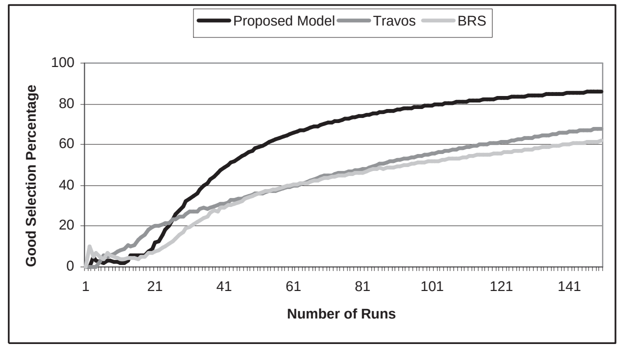

Figure 3. Comparison in terms of good provider selection percentage.

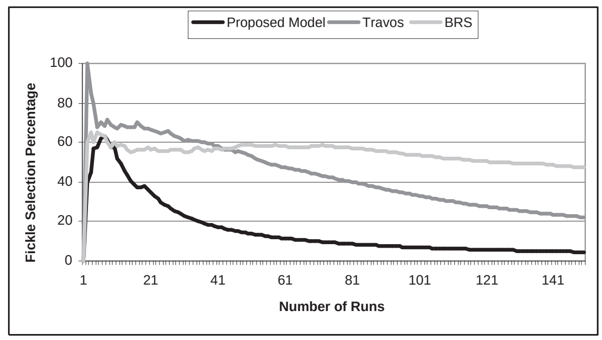

Figure 4. Comparison in terms of fickle provider selection percentage.

merges his own experience with recommendations from other agents. However, unlike BRS model, Travos filters the surrounding agents that are fluctuating in their reports about a specific trustee agent. To some extent, this feature would cause a partial suggestion consideration and thus, Travos agents would learn faster comparing to BRS agents. Rates concerning the good and fickle selection percentage shown in figures 3 and 4 reflect higher efficiency of Travos compared to BRS. Travos agents are capable of preventing the concept of fake reputation in which a group of agents artificially increase their reputation by their collusive behaviors. However, Travos model considers that agents do not change their behavior towards the elapsing time. These missing assumptions affect the accuracy of trust estimation in a very biased environment. On the other hand, lack of agile learning ability for agents will weaken the protection against collusion and fake behaviors. This is the case when a surrounding agent is being discarded because of providing diverse reports about a particular trustee agent. In this case, the deviation would be filtered by mistake if the reports are reflecting the fickle attitude of that particular provider.

In the proposed model we tried to improve the trust mechanism to recover these limitations that enable agents to adapt themselves while the environment is strictly intermittent. In our model agents update their beliefs about the quality of the service together with the accuracy of the ratings provided by the neighbor agents in support or against the specific provider. Considering all the involved parameters, the agent that is doing maintenance, balances his beliefs to be more accurate in terms of knowing the best provider and the best neighbors that can consult. Therefore, as shown in figure 2, the proposed model would gain more utility comparing to the other two models. Figures 3 and 4 reflect our model agile reaction to increase its good selection percentage very fast and thus decrease its fickle selection percentage the most possible.

5. Conclusion

The contribution of this paper is the proposition of a new probabilistic-based trust model for agent-based computing. The trust assessment procedure is composed of on-line and off-line evaluation processes. On-line approach is based on consulting agents presented as well as several other features in order to enhance the accuracy for agents to make use of the information communicated to them. Off-line process considers the communicated information to judge the accuracy of the consulting agents in the previous on-line trust assessment process. This process is incentive compatible in terms of game theory and mechanism design.

Our model has the advantage of being computationally efficient as it takes into account the important factors involved in the trust assessment process. Moreover, extra process of maintenance enables agents to dynamically adjust their beliefs, and consequently update their trustworthy community in a more efficient manner. The proposed mechanism efficiency is compared with other related models and discussed in details to prove the capabilities of the proposed model. In this paper, the size of the simulation environment is not taken into account as an issue. However, the scalably of the proposed model is subject to be analyzed in huge networks. Our objective for future work is to advance the assessment model to enhance the model efficiency and we would like to analyze such efficiency in large-scale networks. In the off-line process, we need to elaborate more on the optimization part, trying to formulate it in the sense to be adaptable to diverse situations. Finally, we plan to further the analysis in comparison with other models in order to capture more results reflecting the proposed model capabilities.

References

[1] J. Bentahar, B. Khosravifar. Using Trustworthy and Referee Agents to Secure Multi-Agent Systems, ITNG 2008, pp. 477-482, 2008.

[2] J. Bentahar, F. Toni, J-J. Ch. Meyer and J. Labban. A security framework for agent-based systems. In the International Journal of Web Information Systems, Emerald, 3(4):1102-1115, 2007.

[3] C. Dellarocas. The digitization of word-of-mouth: promise and challenges of online feedback mechanisms. Management science, 2003, 49(10):1407-1424, 2003.

[4] T. Dong-Huynh, N.R. Jennings and N.R. Shadbolt. Certified reputation: How an agent can trust a stranger. In Proceedings of The Fifth International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 1217-1224, Hakodate, Japan, 2006.

[5] T. Dong-Huynh, N.R. Jennings and N.R. Shadbolt. Fire: An integrated trust and reputation model for open multi-agent systems. Journal of Autonomous Agents and Multi-Agent Systems 13(2):119-154, 2006.

[6] P.M. Dung, P. Mancarella and F. Toni. Computing ideal sceptical argumentation. Artificial Intelligence, Special Issue on Argumentation in Artificial Intelligence, 2007.

[7] A. Jesang and R. Ismail. The beta reputation system. 15th Bled Electronic Commerce Conference e-Reality: Constructing the e-Economy, June 2002.

[8] B. Khosravifar, M. Gomrokchi, J. Bentahar, and Ph. Thiran. A Maintenance-based trust for Open MultiAgent Systems. Accepted in the 8'th International Conference on Autonomous Agents and Multiagent Systems, AAMAS, Budapest, Hungary 2009.

[9] N. Nisan, T. Roughgarden, E. Tardos, and V.V. Vazirani. Algorithmic Game Theory. Cambridge University Press, 2007.

[10] J. Sabatar. Trust and reputation for agent societies. Phd thesis, Universitat autonoma de Barcelona, 2003.

[11] J. Sabater, M. Paolucci and R. Conte. Repage: REPutation and ImAGE Among Limited Autonomous Partners. Journal of Artificial Societies and Social Simulation, 9(2), 2006.

[12] E. Shakshuki, L. Zhonghai, and G. Jing. An agent-based approach to security service. International Journal of Network and Computer Applications. Elsevier, 28(3):183-208, 2005.

[13] W. T. Teacy, J. Patel, N.R. Jennings, and M. Luck. Travos: Trust and reputation in the context of inaccurate information sources. Autonomous Agentsand Multi-Agent Systems, 12(2):183-198, 2006.

[14] Y. Wang, and M.P. Singh. Formal trust model for multiagent ststems. Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI), pp. 1551-1556, 2007.

[15] L. Xiong and L. Liu. PeerTrust: supporting reputation-based trust for peer-to-peer electronic communities. Journal of IEEE Transactions on Knowledge and Data Engineering, 16(7):843-857, 2004.

[16] P. Yolum and M.P. Singh. Engineering self-organizing referral networks for trustworthy service selection. IEEE Transaction on Systems, Man, and Cybernetics, 35(3):396-407, 2005.