

# Exploration and Exploitation in Adaptive Trust-Based Decision Making in Dynamic Environments

Mark Hoogendoorn, S. Waqar Jaffry, and Jan Treur  
Vrije Universiteit Amsterdam, Department of Artificial Intelligence  
De Boelelaan 1081, 1081 HV Amsterdam, the Netherlands  
<http://www.few.vu.nl/~{mhoogen, swjaffry, treur}>  
{mhoogen, swjaffry, treur}@few.vu.nl

**Abstract**— Trust is generally considered an important aspect in the decision making of agents. In the literature, a variety of computational models for trust have been proposed covering how an agent can make a decision by exploiting the trust levels it has for the different options. Within such a decision making mechanism the focus is usually on a single most trusted option, which as a side effect may lead to a lack of information on the other options over time. Therefore it may sometimes be worthwhile to be more explorative in the decision making, especially for dynamic environments. In this paper an adaptive trust-based decision making model is proposed that varies the extent of exploration and exploitation in the agent's decisions.

## I. INTRODUCTION

Within multi-agent systems research, the notion of trust has been under investigation for a substantial period (see, e.g. [9], [10] for an overview). The research is hereby focused on a variety of topics, ranging from cognitive models of trust (see e.g. [1]) to decision models based upon trust values of individuals (see e.g. [5], [11]). In most trust-based decision making models the selection is based upon the highest trust level, sometimes in combination with additional factors. In [8], for example expresses a decision making model based upon the trust value, the importance of the action, the risk of the situation, and the perceived competence of the options. In both [2] and [11] the decision to select an option is solely dependent on the trust level: the most trusted option is simply selected.

In [6] the idea has been proposed that agents should sometimes also be more explorative in their efforts to gain a good insight on the quality of other options. Hence, a decision to select a particular option (e.g., an agent to perform a certain task) is not solely based upon the current highest trust value, but should also depend on the desired amount of exploring to investigate the trustworthiness of other options. Typically it is assumed that agents should be explorative in the beginning, and thereafter exploit their gained knowledge. When participating in highly dynamic environments however, the behavior of the options and the experiences that these options give can fluctuate significantly over time, thereby possibly requiring new exploration phases once too many changes occur.

In this paper, an adaptive trust-based decision model is introduced which dynamically determines the amount of exploration and exploitation the agent performs. The choice

of the extent of exploration is based upon the agent's estimation of the extent of change of the environment. More specifically, this is modeled by taking the difference between a short term trust level, based on only the most recent set of experiences the agent has obtained, as opposed to the long term trust level, based on experiences over a longer time period. Hereby an existing trust model has been used (cf. [4]) and the newly designed model has been compared with existing decision making models (cf. [2]).

This paper is organized as follows. In Section 2 the existing trust model used is briefly introduced. The decision making model which incorporates both exploration and exploitation is presented in Section 3. The approach is evaluated by means of simulations in Section 4. Finally, Section 5 is a discussion.

## II. THE EXISTING TRUST MODEL USED

In the existing trust model used throughout this paper (cf. [4]) it is assumed that options  $S_1, S_2, \dots, S_n$  provide experiences  $E_i(t)$  to the human at each time step; these experiences have a value from the continuous interval  $[-I, I]$ . Here,  $-I$  indicates the most negative experience whereas  $I$  is the most positive. The trust values reside on the same interval. The human updates the trust value on an option by keeping trust values on other options under consideration along with the interdependency extent  $\eta$  of the options in the range  $[-I, I]$ . A negative value of  $\eta$  denotes cooperation (indicating that differences between trust levels are partially absorbed) while a positive value represents competition among the options (indicating amplification of differences between trust levels). Furthermore, human personality attributes like trust flexibility  $\beta$ , expressing how much an experience counts, and autonomous trust decay  $\gamma$ , indicating how fast trust goes back to a neutral value, also play a role in this process. The following differential equation then represents the trust update over time:

$$dT_i/dt = \beta \cdot (E_i(t) - T_i(t) + \tau_i(t)) - \gamma \cdot T_i(t) \quad (1)$$

This expresses that the change in the trust value  $T_i(t)$  for option  $S_i$  is based upon the height of the experience  $E_i(t)$  compared to the previous trust value plus the contribution of the relateness of the trust  $\tau_i(t)$ . This all is multiplied by the factor  $\beta$ , which expresses flexibility of the trust: how

much the new experience counts compared to the already existing trust. From this, the autonomous decay of the old trust value is subtracted. The relativeness of the trust is calculated by first normalizing all trust values to a value between 0 and 1:

$$T'_i(t) = (T_i(t) + 1)/2 \quad (2)$$

The value for  $\tau_i(t)$  is calculated in the following way:

$$\tau_i(t) = \eta \cdot (T'_i(t) - \sum_{j=1}^n T'_j(t)/n) \quad (3)$$

This equation determines the difference between the trust level of option  $S_i$  compared to the average. The value is then multiplied with the value of  $\eta$  resulting amplification in case of competition and absorption in case of cooperation.

### III. ADAPTIVE TRUST-BASED DECISION MAKING

As mentioned before, trust is generally considered an important criterion in the decision process to select a particular option among alternatives  $S_1, S_2, \dots, S_n$ . As a result, selection mechanisms have been expressed that take the trust level as input and generate a decision of the option to be selected. Mostly, the option with the highest trust level is selected (see e.g. [2], [4] and [11]). This is based on the idea that the trust levels give an adequate account of the world's state of affairs. In fact a silent assumption behind this is that the world is static or at least does not change very fast, so that adequate trust values for the different options can be built up over a longer time period. However, in case the world is changing over relatively short time periods, the trusting agent can be misguided when it always only chooses the option that was the best in the past, and thus prevents itself from acquiring new experiences with the other options.

In this section a trust-based decision model is introduced which varies between this exploitative behavior and more explorative behavior enabling it to obtain experiences with other options compared to the currently most trusted option. In order to create such a model a separation is made between two types of trust, namely *long term trust* and *short term trust*. Essentially, both types of trust are represented by the model described in Section 2. The only difference is the value for the decay  $\gamma$  chosen. For the short term trust a relatively high value for  $\gamma$  is selected (since this trust should only represent recent experiences and not consider a longer history of experiences) referred to as  $\gamma_S$ , whereas for the long term trust a relatively low value  $\gamma_L$ . The idea is that a difference between the short term and the long term trust levels indicates a recent change in the behavior of the options, and therefore is used to make an agent more explorative. In case the trend remains stable, indicated by no or only a small difference between long term and short term trust levels, the agent should be more exploitative by choosing the most trusted option. Assume that  $LT_i(t)$  represents the long term trust in option  $S_i$  at time  $t$ , whereas  $ST_i(t)$  represents the short term trust (both projected on the interval  $[0, I]$ ), then the estimated change in the environment  $C(t)$  is defined by:

$$C(t) = \frac{\sum_{i=1}^n |LT_i(t) - ST_i(t)|}{n} \quad (4)$$

Hence, the change is simply the average absolute difference between the short and the long term trust level for all options. Based upon this factor, the fraction of exploration behavior the agent should exhibit is updated in the following way:

$$\begin{aligned} dE(t)/dt &= Pos(\alpha \cdot C(t) - \rho \cdot E(t)) \cdot (1 - E(t)) - \\ &Pos(-\alpha \cdot C(t) + \rho \cdot E(t)) \cdot E(t) \end{aligned} \quad (5)$$

Here, the function  $Pos(V)$  is defined by:

$$\begin{aligned} Pos(V) &= V \text{ if } V \geq 0 \\ Pos(V) &= 0 \text{ if } V \leq 0 \end{aligned} \quad (6)$$

The function to determine the update of the extent of exploration to be performed considers two aspects. The first aspect expresses that the exploration extent has to be increased according to a factor  $\alpha$  times the estimated change  $C(t)$  in the environment. The second aspect specifies that there is an autonomous decay of  $E(t)$  by a factor  $\rho$ . When the first contribution is higher than the second one, the value of  $E(t)$  will increase; otherwise it will decrease. In particular, when the change in the world  $C(t)$  is 0, the value of  $E(t)$  will eventually approximate 0, which indicates a purely exploitative trust-based decision making approach.

Given the exploration extent  $E(t)$ , each option can be assigned a certain probability  $RP_i(t)$  of being selected. A difference is made between the option that currently has the highest trust value (equation 8), and the other options (equation 7).

$$RP_i(t) = (1 - E(t)) \cdot \left[ E(t) \cdot \frac{T(t)_i}{\sum_{j=1}^n T(t)_j} \right] + E(t) \cdot \left[ \frac{E(t)}{n} \right] \quad (7)$$

$$RP_i(t) = (1 - \sum_{j \neq i} RP_j(t)) \quad (8)$$

The first equation (7) expresses that the options that are not the most trusted option have a request probability of 0 in case the exploration factor is 0, whereas the probability is equal to all other options in case there is the maximum exploration extent 1. In case the exploration factor is between these extremes, a value proportional to the trust relative to all other options is taken in combination with a fraction of an equal request probability. In the second equation (8), the request probability of the option with the highest trust value is expressed, which is simply the remaining probability. This is 1 in case the exploration factor is 0, and an equal share in case the exploration factor is 1, and value with a combination of an equal share and a part dependent on the trust value in case of intermediate exploration values.

#### IV. SIMULATION RESULTS

The trust model described in Section 2 and the adaptive trust based decision model presented in Section 3 have been used to generate simulation results (using an implementation of the model in C++). Hereby, the models have been tested in various experiments with three options to judge their respective pros and cons. In the simulation results the parameter settings as shown in Table 1 have been used.

TABLE 1. CONFIGURATIONS USED FOR EXPERIMENTS

Parameter Name	Symbol	Values
Number of options	$n$	3
Options relativeness	$\eta$	1.00
Agent initial trust on options	$T_i(0)$	0.00
Agent initial exploration	$E(0)$	1.00
Rate of change of trust	$\beta$	0.50
Rate of change of exploration	$\alpha$	variable
Autonomous decay of trust	$\gamma_L$	0.10
Autonomous decay of short term trust	$\gamma_S$	variable
Autonomous decay of exploration	$\rho$	variable
Size of time step	$\Delta t$	0.10
Total time steps	TS	1000
Total time	$t$	100

In these experiments, a relativeness of 1 is taken which means that trustees are competitive with each other, so if an option gives a positive experience to an agent not only the trust value of this option will increase but it also has an effect on trust values of other options (in this case for those options with a more negative trust value their trust value will go down even more). The initial trust values  $T_i(0)$  for all options are set to neutral *zero* to give each a fair start in the beginning of the simulation. Furthermore, the rate of change  $\beta$  has been set to an average value 0.5. The initial exploration rate has been set to 1 to make the agent more explorative in the beginning. The parameters for short term trust decay  $\gamma_S$ , rate of change for exploration  $\alpha$ , and the autonomous decay for exploration  $\rho$  will all be varied in the scenarios to enable an investigation of the influence of these parameters.

In the following subsections simulation results of several experiments are presented. In these experiments, the experiences given by the options is varied. For those options that are not selected, an experience value equal to the current trust level of the option is taken (thereby thus only having a decay and an influence of the relativeness with the other options). In experiment 1 a one time change occurs in the world, resulting in different behavior of all options. In experiment 2 a highly dynamic environment with many changes is used and finally a non-changing world is represented in Experiment 3. As a measure for performance of a model the average value of the experiences until that time point is used.

##### A. Experiment 1

In this experiment a situation is modeled where one of the options initially gives very good experience while others give very bad experiences. After some time however (at time point 25) the first option starts giving neutral experiences while the other options start giving very good experiences. Here the options' behavior is assumed to be normally

distributed according to specific standard deviation and mean value (see Table 2). As a setting for the first set of graphs, values for  $\gamma_S$ ,  $\alpha$ , and  $\rho$  of 0.25, 0.50, and 0.25 respectively have been used.

TABLE 2. OPTIONS BEHAVIOR FOR EXPERIMENT 1

Option	Start Time	End Time	Standard deviation	Mean
T1	0	25	0.1	0.9
T2	0	25	0.1	0.1
T3	0	25	0.1	0.1
T1	25	100	0.1	0.5
T2	25	100	0.1	0.9
T3	25	100	0.1	0.9

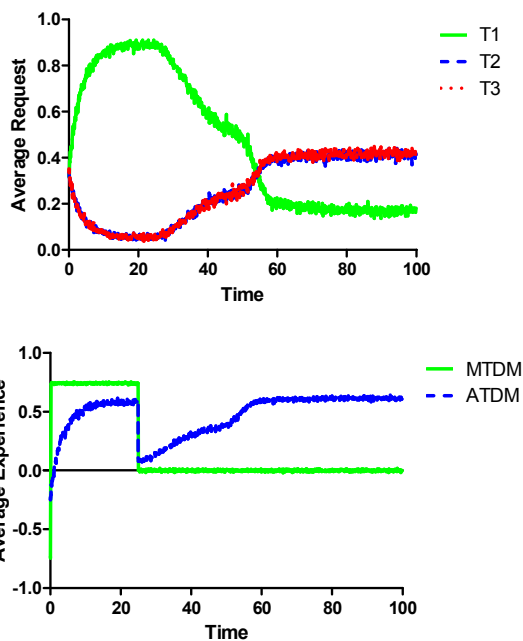


Figure 1. a) average request per option over time for ATDM, b) Performance of MTDM and ATDM

In Figure 1a the averages over the requests to each option is shown for ATDM (adaptive trust-based decision model) respectively, averaged over 1000 runs. Note the the value for MTDM (selection of the highest trust value) is not shown as it always requests T1. In the figures it can be observed that the selection of the highest option does not always lead to satisfactory results. MTDM sticks to T1 as he still gives neutral experiences which is far better than the trust level for the other two options (which is based upon the initial, very negative experiences). The adaptive ATDM model however, keeps exploring the environment to find better opportunities.

The average experience curve (i.e. the average experience for each time point taken over all 1000 runs, not using equation 9) for both models for this experiment is shown in Figure 1c. Here it can be seen that when the environment does not change (i.e. until time point 25) MTDM behaves better than ATDM. After the change however, ATDM starts to outperform MTDM.

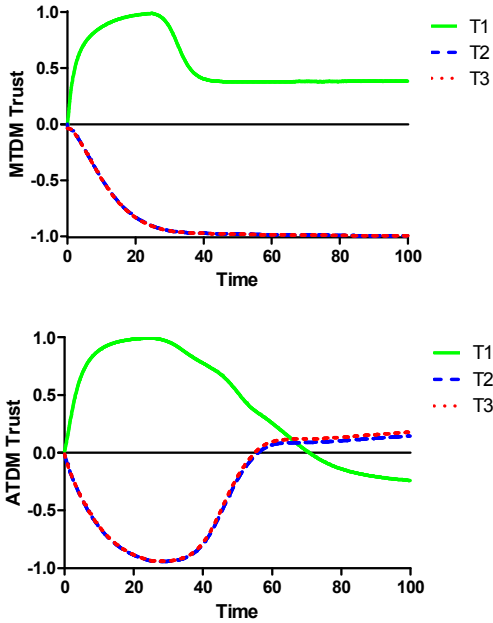


Figure 2. Trust dynamics, a) MTDM, b) ATDM (long term trust)

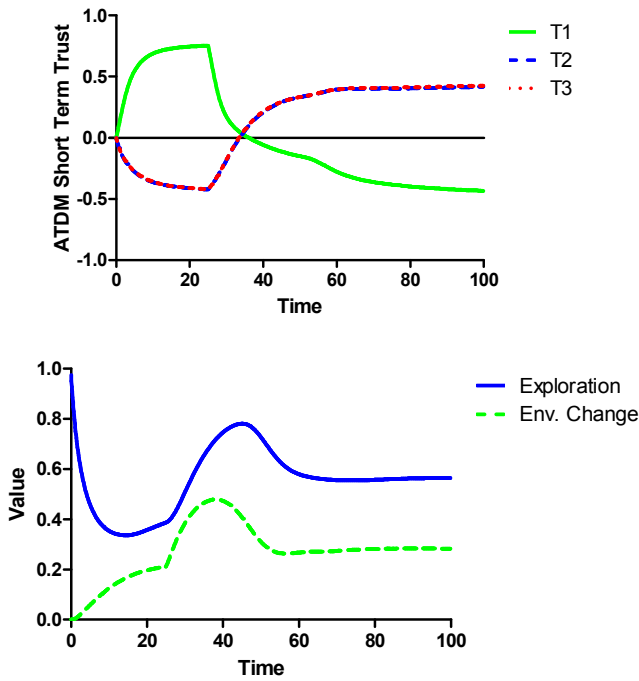


Figure 3. a) Short term Trust dynamics of ATDM, b) Change in Environment and Exploration dynamics for ATDM

As both models generate different request patterns and hence, different experience sequences, their respective trust dynamics curves are shaped very differently (see Figure 2a, 2b), whereas the same trust function has been used. In Figure 2a the trust curve of MTDM is shown. Here it can be seen that despite the fact that  $T2$  and  $T3$  give positive experiences after time 25, this model could not detect this change and

hence keeps the trust level of  $T2$  and  $T3$  very low while in Figure 2b the trust curves of ATDM detect the change very rapidly.

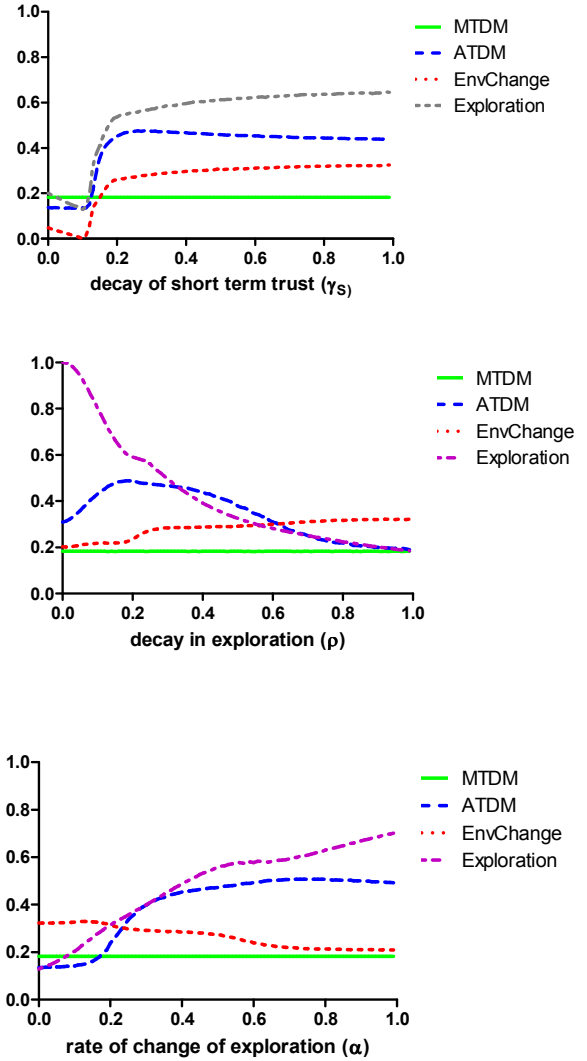


Figure 4. Performance of MTDM and ATDM, Environment Change, Exploration Dynamics for experiment 1 with change in, a)  $\gamma_s$ , b)  $\rho$  and c)  $\alpha$ .

In Figure 3a the short term trust values for ATDM are shown. Note that the difference in short term and long term trust (where the long term trust of ATDM is shown in Figure 2b) is the change in environment perceived by the model. This change in the environment leads to the exploration behavior of ATDM (see Figure 3b) that made the difference in performance of both models in this experiment.

To see the effect of the exploration parameters (i.e.  $\gamma_s$ ,  $\alpha$ , and  $\rho$ ) on the dynamics of exploration and eventually on the performance of the ATDM model, graphs are shown in Figure 4 with varying setting for each of these parameters. In these figures, the settings for the parameters are shown on the x-axis whereas the y-axis represents the value of different variables namely MTDM (representing the average performance of MTDM using equation 9), ATDM

(representing the average performance of ATDM, again using equation 9), *EnvChange* (representing the average change in environment perceived by the agent in ATDM) and *Exploration* (representing average value of exploration in EBDM).

### B. Experiments 2 and 3

Two other experiments have been conducted: (1) A highly dynamic environment. Hereby, one option starts giving good experiences while the other two give bad experiences, periodically these options change their behavior after a specified period of time (2.5 time points). Here, the performance of ATDM is better for almost every value of the parameter compared to MTDM; (2) No change in the experience behavior, due to the explorative behavior of ATDM this performs worse compared to MTDM as the latter approach immediately focuses on the appropriate trustee.

## V. DISCUSSION

In this paper, an adaptive trust-based decision making model has been presented that is able to handle dynamic environments in which the quality of the different decision options change over time. In most work on trust-based decision making, these decisions are made in a rather straightforward way: to select the most trusted option (see e.g. [2], [4], and [11]). In dynamic environments however, it might sometimes be worthwhile to be more explorative every now and then as agents might behave differently over time, and new agents might enter the system. Therefore, a decision making model has been presented that is able to do precisely this: start to explore when the world changes (based upon a difference between so-called long and short-term trust) and be more exploitative in case the world seems to be stable. The model has been evaluated and compared to other existing decision models by means of simulations for a series of cases of (nondeterministic) environmental dynamics. It has been shown that this decision model works better in these dynamic situations, whereas it performs approximately equivalent to the selection of the highest trustee in case of relatively stable environments.

Several other researchers within the domain of agent systems have proposed more complex decision making models than simply selecting the most trusted agent (or group of agents). In [6] it is also proposed to use a mechanism for switching between exploration and exploitation. In their setting however, they aim at discovering new agents that could potentially function well by giving these agents cases with known answers such that it can easily be determined whether the agent is good or not. In the approach, the precise mechanism when to explore and when to exploit is however not specified. In [1] the decision mechanism to select an agent is made more complex by considering the form of delegation that takes place: weak delegation and strong delegation, thereby differentiating between cases where there is a formal agreement between agents. The decision mechanism allows the reasoning about the consequences of these different delegation types.

Next to agent systems research, also in management sciences decision mechanisms have been studied. In [7] for example finding the balance between exploration and exploitation in alliance formation decisions is studied. This study is performed from a non-computational perspective, but does show the necessity of trying to find a good trade-off between the two in order for organizations to be successful.

For future work, it would be interesting to investigate how the model presented here matches with the human tendency towards exploration and exploitation. Hereby, it might even be possible to try and tailor the parameters in the function towards the human behavior such as for instance done for a trust function in [3]. Moreover, parameter tuning techniques can be used to select the best values of the parameters, given certain domain characteristics. Furthermore, additional factors to base a decision upon could also be taken into account.

## ACKNOWLEDGMENTS

This research has partly been conducted as part of the FP7 ICT Future Enabling Technologies program of the European Commission under grant agreement No 231288 (SOCIONICAL).

## REFERENCES

- [1] Falcone, R., Castelfranchi, C.: Trust dynamics: How trust is influenced by direct experiences and by trust itself. In: Proc. of AAMAS 2004, pp. 740–747, (2004).
- [2] Hoogendoorn, M., Jaffry, S.W., and Treur, J., Modeling Dynamics of Relative Trust of Competitive Information Agents. In: Klusch, M., Pechoucek, M., Polleres, A. (eds.), In: 12th Int. Workshop on Coop. Inf. Agents, CIA'08. LNAI 5180, pp. 55-70, Springer Verlag, (2008)
- [3] Hoogendoorn, M., Jaffry, S.W., and Treur, J., (2009). An Adaptive Agent Model Estimating Human Trust in Information Sources. In: Proc. of the 9th IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IAT'09. IEEE Computer Society Press, pp. 458-465, (2009)
- [4] Hoogendoorn, M., Jaffry, S.W., and Treur, J., Incorporating Interdependency of Trust Values in Existing Trust Models for Trust Dynamics. In: Proceedings of the Int. Conf. on Trust Management, TM'10. Advances in Information and Communication Technology, vol. 321. Springer Verlag (pp. 263-276), 2010.
- [5] Kim, D.J., Ferrin, D.L., and Rao, H.R., A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision Support Systems*, vol. 44, 2008, pp. 544-564.
- [6] Lashkari, Y., Metral, M. & Maes, P. (1994). Collaborative Interface Agents. In: Proceedings of the Twelfth National Conference on Artificial Intelligence, AAAIPress, pp. 444-449.
- [7] Lavie, D, and Rosenkopf, L., Balancing Exploration and Exploitation in Alliance Formation. *Ac. of Mngmt J.*, vol. 49, 2006, pp. 797-818.
- [8] Marsh, S. (1994). Formalising Trust as a Computational Concept. Ph.D. thesis, Department of Mathematics and Computer Science, University of Stirling.
- [9] Ramchurn, S.D., Huynh, D., Jennings, N.R., Trust in Multi-Agent Systems, *The Knowledge Engineering Review*, vol. 19, pp. 1-25, (2004)
- [10] Sabater, J., and Sierra, C., Review on Computational Trust and Reputation Models, *Artificial Intelligence Review*, vol. 24, pp. 33-60, (2005)
- [11] Vassileva, J., Breban, S., and Horsch, M., Agent Reasoning Mechanism for Long-term Coalitions Based on Decision Making an Trust, *Computational Intelligence*, vol. 18, 2002, pp. 583–595.