TruMet: An approach towards computing trust in multi-agent environment.

TruMet: An Approach towards Computing Trust in Multi-Agent Environment

Sanat Kumar Bista, Keshav P. Dahal, Peter I. Cowling and Bhadra Man Tuladhar

***Abstract*— The growing popularity of multi-agent based approaches towards the formation and operation of virtual organizations (VO) present over the Internet, offer both opportunities and risks. One of the risks involved in such community is in the identification of trustworthy agent partners for transaction. In this paper we aim to describe our trust model which would contribute in measuring trust in the interacting agents. Named as *TruMet,* the trust metric model works on the basis of the parameters that we have identified as relevant to the features of the community. The model primarily analyses trust value on the basis of the agent's reputation, as provided by the agent itself, and the agent's aggregate rating as provided by the witness agents. The final computation of the trust value is given by a weighted average of these two components. While computing the aggregate rating, a weight based method has been adopted to discount the contribution of possibly un-fair ratings by the witness agents.**

***Index Terms*— Agent, Reputation, Trust, Trust Measurement, Unfair Rating**

I. Introduction

Trust is a crucial aspect for any form of interaction, be it between humans or that between agents interacting with each other to fulfill their needs as demanded by the cause of their interaction. With modern day systems becoming more and more distributed, dynamic and complex at the same time, solution approaches involving multi-agent systems, due to its ability to act autonomously and rationally have gained larger interest of academia and industries. The popularity of online trading businesses, virtual organizations over the Internet, the Grid etc, among researchers as well as business organizations and its acceptance in the form of widespread use by the users easily support the fact above. One of the challenges faced by an agent in a virtual community is in assessing the trustworthiness of the other agent with which it wishes to transact. Every agent interacts with others in the community to maximize its own gain, and it can be at the cost of the gain of other agents also [6]. For instance, in an e-commerce setting, where strangers are interacting, it might be more appealing to act deceptively for immediate gain rather than cooperation [2]. As an example, if a buyer pays first, the seller might be tempted to not provide agreed upon quality of goods or services [4]. Thus, it is important for an agent in a community to isolate malicious ones from reputed and trustworthy agents.

Our approach towards such a method of isolation is to develop a metrics for assessing the worthiness of trust of any transacting subject agent. Our approach of measuring the trust is similar to that of [1], in which the authors have first identified parameters for measuring trust and then based on those parameters a metrics model based on weighted average approach has been established to compute the trust value. However, in [1] the authors have primarily focused on the trust metrics in Peer-to-Peer systems. In our case, we consider the transacting bodies to be agents meaning that each component participating in the transaction is an autonomous software entity, rather than a simple peer in which we might not expect autonomy and intelligence. This extension has introduced number of complexities which has required us to design a new set of trust parameters with regards to computing it between agents in a multi-agent environment. Another contribution of our trust metrics model is its capability in handling contributions due to dishonest witness by agents. The metrics we have developed is balanced enough to take care of such contributions. We further define an approach in combining information from multiple witness agents. The concept of trust network as put forward by the authors in [7], [9] has been explored to provide solutions to the possible dishonest feed back problem.

In general, the major contribution of this paper is in formulating a useful trust metrics model, called *TruMet*, the outcome of which might be an easily readable figure for interpretation by agents. A system like this which considers a holistic approach towards Trust Measurement is the need of the hour and we have attempted to fulfill this need. Also, the enhanced information merging model is presented to bring out new methods in witness analysis and deception detection. In section V of this paper, we present the evaluation and result. Particularly we take a sample case in which we assume or infer some values for the parameters we outlined in the trust metrics model and then compute a trust value based on our

Sanat Kumar Bista is with the MOSAIC Research center at the University of Bradford, BD7 1DP, UK as a Research Student. (Corresponding author: phone: ++977-98510-86943, fax: ++977-1-4280612; e-mail: s.k.bista@bradford.ac.uk).

Keshav P. Dahal is with the MOSAIC Research Center at the University of Bradford, BD7 1DP, UK as a Senior Lecturer. (e-mail: k.p.dahal@bradford.ac.uk).

Peter I Cowling is with the MOSAIC Research Center at the University of Bradford, BD7 1DP, UK as a Professor. (e-mail: p.i.cowling@bradford.ac.uk).

Bhadra Man Tuladhar is with the Department of Natural Sciences (Mathematics) at Kathmandu University, Nepal, as a Professor. (e-mail: dean_sci@ku.edu.np).

model. We also present some meaningful graphical results in this section.

II. Related Works

The compelling factor for this research is mainly due to the absence of a general trust metric system for agent interaction in a multi agent system. There have been research and results in expressing trust as a probability distribution, which though realistic, is at the same time complex for a user (might be agent) to interpret [2].

In [1] the authors have proposed a reputation based trust metric system. The metric is mainly targeted towards evaluating trust between peers in a peer-to-peer based electronic community. We have adopted a similar approach to measuring trust, but with several enhancements to suit the needs of multi agent systems in particular.

One of the important characteristics of any trust measuring systems is its approach towards merging information from multiple sources. In [2], [7] the authors have presented approaches towards rating aggregations and information merging, particularly in relevance to trust measurement. The process of computing aggregate rating involves collecting opinions of all agents on one agent's rating. This clearly involves a series of queries through the network of the agents. In [7] and [9] the authors have defined such a network of agents through which referrals are propagated and computed as *TrustNet.* In computing the aggregate rating for our trust metrics also we will use a data structure like *TrustNet,* which we think of as a simple referral network. Such a network is constructed by a requesting agent through a series of referrals by incorporating each referral into the network with *depthLimit* as a bound to the length of chain [9].

Approaches towards handling dishonest feedback on an agent's capability has also been a focus of research for many. This is proven by the efforts of authors in [2],[3],[4],[6],[7]and [9]. The authors in these papers have explored multiple ways under Endogenous and Exogenous approaches to filtering of unfair ratings.

III. The TruMet System

We have identified the following parameters that would contribute towards the trust metric:

1. Reputation Value of the Subject Agent as provided by the agent itself (local rating).
2. Average Reputation: Ratio of total Reputation to the total number of transactions)
3. Aggregated rating, rating provided by entire agent community with which the agent has communicated.
4. Agent Community guarantee

A brief description and importance of these parameters is given below:

Agent Reputation: One of the parameters of trust is an agents own version of its reputation value. An agent however does not produce a random reputation value on its own, but it acquires such a value by accumulating the reputation offered to it by the agents it has worked with in past. Obviously, such a value might not be a true representation of an agent's actual reputation mainly because the raters for some reasons might intentionally offer a higher or lower reputation than the agent actually deserves. Hence, we are of the opinion that such a version of reputation should not be the only parameter of trust computation. We will explore other parameters to contribute in neutralizing the effects of deviated reputation on this part.

Average reputation: [1] discusses on why it is necessary to consider average while computing satisfaction an agent might have from its total number of transactions. The authors have provided examples where an agent might be interested in being honest for many smaller transactions, thus attaining reputation, and then turn deceitful when it comes to significantly important transaction. We consider average reputation of an agent to be a parameter contributing towards neutralizing the effects of acts as mentioned above. An average reputation of an agent is the ratio of an agent's reputation to the total number of transactions it has had before on the ground of which the reputation is claimed.

Aggregated Rating: In the society it is not necessary that an agent might have interacted with every other existing agent. In a case where an agent is considering to start transacting with unknown agents, it becomes necessary for the agent to asses the trustworthiness of the other agent. Hence the agent might enquire about the trust worthiness of the other agent with all other agents of the society with whom the subject agent has interacted before. Calculating such an aggregate rating is obviously a non-trivial task. Moreover, the problem is further aggravated by the witness agents providing dishonest feed-backs. Thus, it becomes necessary to compute the aggregate community rating with an inbuilt filtering mechanism for discounting unfair/dishonest ratings from the agent society. Our approach in TruMet explores the possibility of such a computation.

Agent's Community Guarantee: An agent might be a member of certain community which is reputed and has established itself as a trust worthy community. Further, it is much likely that an agent from such a community mostly interacts with other agents of similar reputation. Hence, it might sometime also be desirable that the agents own version of its reputation be given a greater weight. This can be adjusted by providing greater weight to the *Agent Community Guarantee* parameter in the model. The *Trust Metric* section below exemplifies a way in which this value can be assigned in the model. Our basic assumption here is that agents which are backed up by the norms of institutional guarantee are trust worthy and the agents wishing to transact might significantly reduce the overhead of gathering trust information on the subject agent from multiple witnesses.

A. *The Trust Metric*

Based on the parameters identified above, we endeavor to establish a general trust metric formula based on weighted average. Our approach towards measuring trust is similar to that of [1], but however, our overall model differs significantly from theirs' as stated already in the introduction section.

Establishing the trust metric: Let X be an agent whose trust value is being measured in any i^{th} instance of time *t*. Let R(X) denote the total reputation possessed by the agent X, which is a direct correlation to the number of successful transactions of X in past, and R(Xi) give the reputation of X at sometime t=i, due to the positive ratings of other agents. Similarly, $n(TR_x)$ gives the total number of transaction (successful as well as unsuccessful) that the agent X has had with all other agents for which it was reputed by them. Let, AGR(X) stand for the aggregate rating of the agent X by all member agents with whom it has interacted in the past. For simplicity reasons now, we will assume that the function AGR(X) will return a value that is compatible to the formula being established. The details of the operating procedure of this function are given in section *B* below. Also, let CGF(X) represent the community guarantee factor for the agent X. The function CGF (X) represents the contribution of all the communities of which the agent X is a member of. We assume here that an agent if it is not a member of at least one established community is not reliable enough for believing what it says of itself. In other words, in case where CGF(X) has a zero value, the total computation of trust relies on the aggregated rating obtained by the agent from other agents. For simplicity, here we will consider only two states of community guarantee, either an agent is backed up by established community (CGF(X) = 1), or it is not (CGF(X) = 0).

With these, we define the trust of agent X , T(X) as,

$$T(X) = WG_a * \left(CGF(X) * \left(\sum_{i=1}^{n(TR_x)} R(X_i) \right) / n(TR_x) \right) + WG_b * (AGR(X)) \quad (1)$$

Where, WG_a, and WG_b are the weight factors associated with each of the components of the agent X's trust measurement. Further, $0 \le WG_a \le 1$, $0 \le WG_b \le 1$, but $WG_a + WG_b = 1$. The full rating scale of trust is $0 \le T \le 1$.

B. *Computing the aggregate rating and Score*

In general, each agent is reputed by another agent after each transaction by providing a rating that can either be negative or positive. This rating can be expressed as a vector R of the number of positive (successful) and negative (unsuccessful) ratings. Following [2] we define the vector R and the aggregation process as below:

$$R = [s, u] \quad (2)$$

where, *s* stands for successful rating and *u* for unsuccessful rating such that $s \ge 0$ and $u \ge 0$.

Aggregation of ratings is obtained by simple vector addition. Let (X, Z) be a pair of agent. An aggregate rating R^t (X, Z) highlighting X's overall opinion of Z at time *t* is obtained as:

$$R^t(X,Z) = \sum_{Z}^{X,t} R \quad (3)$$

As established in [2], a simple point estimate of agent Z's probability distribution is given by the expected value E of the probability distribution. Thus, Z's reputation score at *t* is given by:

$$(PR^t(Z)) = E[\mathbf{beta}\ (\mathbf{R^t(Z)})] = (s+1) / (s+u+2) \quad (4)$$

Where $R^t(Z) = [s, u]$.

IV. Referral Network

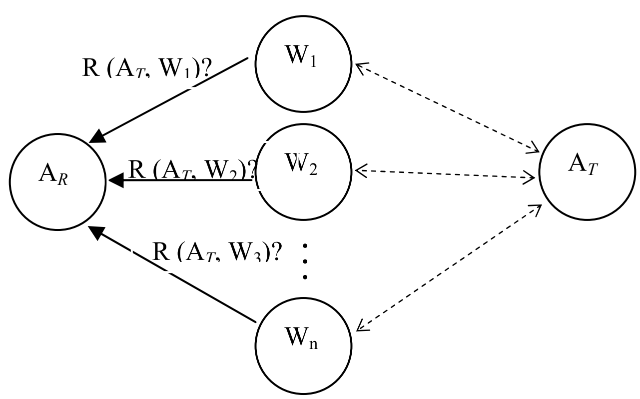

Fig. 1. Trust Referral Network

As specified in the figure above, a referral network is realized by four different entities, the *requesting agent*, the *target agent*, the *witness agent* and the *references*. Firstly, there is an agent, A_R, requesting for reference on some target agent, A_T. The agent A_R will query witness agents W_1, W_2 …Wn on the rating of the target agent A_T. The witness agents will provide the feedback based upon each of its experience of working with the target agent in the past. The dotted arrows in Fig.1 represent the fact that the interaction between the witness and target agent is an action in past. For conformance of the referral results to the aggregation procedure in section (III B*)* above, we pass the referral result as a rating vector consisting of the number of successful and unsuccessful transactions the witness agent has had with the target agent in the past. The structure of the vector is as specified in III B above.

A. Unfair Rating

While providing the ratings on the target agent by the witness agents, there is every chance that the agent provides an unfair and deceitful feedback. As already highlighted in the introductory section, agents work towards maximizing their own profit, and most of the time such profit is obtained at the cost of loss of other agents. Any trust rating system not considering this aspect is definitely incomplete and further, there also a risk of such system being misleading. Careful planning in the computation of ratings can minimize the negative effects in trust computation due to such contributions.

Ratings are primarily categorized into four different types [2],[4],[7],[9]:

- **Normal rating**: In which the witness agent provides the actual rating. In our case, this means that the witness agent returns the true value of successful and unsuccessful transactions.
- **Complementary rating:** In which the witness agent gives exactly opposite rating. In our case, this means that the value of successful and unsuccessful transactions is swapped.
- **Exaggerated Positive rating** ("ballot stuffing"): In which the witness agent gives a high value of positive rating. In our case, this means highly increased value of successful transaction.
- **Exaggerated Negative rating** ("bad mouthing"): In which the witness agent gives an overly negative rating. In our case, this means highly increased value of unsuccessful transaction.

B. Filtering Unfair Ratings

In [2],[3],[4],[7],[9] the authors have given their approaches towards filtering the unfair ratings. Broadly the approaches fall under two different categories [2]:

- **Endogenous discounting of unfair ratings**: In which the statistical properties of the ratings are analyzed to reveal any chances of unfair rating.

- **Exogenous discounting of unfair ratings**: In which the ratings are weighed according to the reputation of the rater.

For our trust metric system we follow the approach in [9], which concentrates on exogenous discounting of unfair ratings. The approach in [9] fulfills our requirement better as our trust model also relies on the feedback provided external sources. Relying on weighted majority continuous (WMC) (the authors' variation of Weighted Majority Algorithm (WMA)[11]), they compute weight of each witness agent providing the rating, in which the rating expressed as belief function is transformed into probabilities of telling or not telling truth. Thus, the weight W_i' of witness W_i, with R ($0 \leq R \leq 1$) as the rating from agent A_R is computed as:

$$W_i' = \theta\, W_i \tag{5}$$

Where, $\theta = 1 - ((|prob_i(\{T\}) - R|)/2)$ and, $0 < \theta \leq 1$

Here, $prob_i(\{T\})$ is the probability that the witness agent's feedback on the target agent's *i*th interaction is true and the $prob_i(\{T\})$ is computed by transforming the belief function by applying WMC [9]. The theory that we rely on here is that the witness provided by the agents with greater weight have greater value, and it is normally the agents with lower weight that provide dishonest or unfair rating.

This value of θ, thus computed can now be applied to strengthen the referral network conceptualized before. Further, it has direct influence in the reputation scoring defined in equation (4). We now redefine the equation as follows. Say, θ is the weight of witness X for agent Z whose rating is being scored. Thus, Z's reputation score at time *t* is given by:

$$\theta * (PR^t(Z)) = \theta * E[\mathbf{beta}\ (\mathbf{R^t(Z)})] = \theta * ((s+1)/(s+u+2)) \tag{6}$$

The overall computation of AGR(X) in equation (1) will now be obtained by taking an average of the scores weighted by θ.

V. Evaluation and Results

In the following section we provide experimentation with the metrics system established. We demonstrate with examples how trust is computed and what results can be inferred from the values. The context of a calculation like this can be thought of as an electronic commerce environment where a trading agent wishing to transact might consider assessing the trust worthiness of the other agent. Also, in the process of virtual organization formation over the Internet, such calculation can be meaningful. We take a case in which we provide equal weight to the trust computation parameters i.e. equal weight to first and the second part of the equation (1).The evaluation parameters and relevant assumptions are listed in the table below:

TABLE I Evaluation Parameters

Component	Parameter	Description	Value
Equation (1) first component parameters (agent's own version of data)	WG_a	Weight for first component in equation (1)	0.5, (*indicating a balanced weight*)
	CGF(X)	Community guarantee factor	1, (*indicating a community guarantee*)
	$N(TR_x)$	Total no. of agent X's transactions	Say, 45
	R(X)	Total reputation of agent X, *(total count of X's successful transactions)*	Say, 25 (*indicating a low success)*
Equation (1) second component parameters (witness version of data)	WG_b	Weight for second component in equation (1)	0.5, (*indicating a balanced weight*)
	RW_1, RW_2, RW_3, RW_4, RW_5	Vector representation of witness ratings	Say, [2, 6],[5, 5],[6, 2],[0, 8],[8, 0], Respectively
	θ_{W1},θ_{W2}, θ_{W3},θ_{W4},θ_{W5}	Weight assigned to each witness agent	Say, 0.5, 0.75, 0.8, 0, 1 respectively

The value for the trust of the agent X, T(X) is obtained after applying the given values to equation (1) and (6). Thus, we get the value for T(X) to be 0.43 for the given values in the table above. Towards this score for T(X), contribution of the first part of equation (1) is 0.28 and that of the second part of equation(1) is 0.15. In real world situations, where a mapping of score to the degree of trustworthiness is defined, a score like what we obtained above for agent X can be readily utilized to decide on whether to partner for business or not.

Many of the values assumed above are randomly chosen, however, it has been our attempt to pick values representing diverse conditions. The system that we have modeled is modular in nature and thus gives multiple reflections as we modify the parameters. Another interesting part in the calculation is to compute possible deviation of truth by the agent by comparing the value for the parameters of the first component of equation (1) to the second component of it. For instance, the reputation value and the total no. of transaction value in the first component (in which these values are the agent's own version) can be compared to that in the second component (in which these are obtained through the rating by all other applicable agents in the referral network). A graphical representation of this is given in figure 4 below. Appropriate measures can be taken if it is found that the corresponding values are significantly different to each other. In the example evaluation above, we have the taken the values to be non-conformant to each other, hence there lies a difference in the computed value of first and second part.

Some meaningful graphical representations based on the data above are given below:

TABLE II Some Specific data for analysis

Successful Rating	**Unsuccessful Rating**	**Reputation Score** (Equation (4))	**Theta**	**Theta * Reputation Score** (Equation (6))
2	6	0.3	0.5	0.15
5	5	0.5	0.75	0.375
6	2	0.7	0.8	0.56
0	8	0.1	0.01	0.001
8	0	0.9	1	0.9

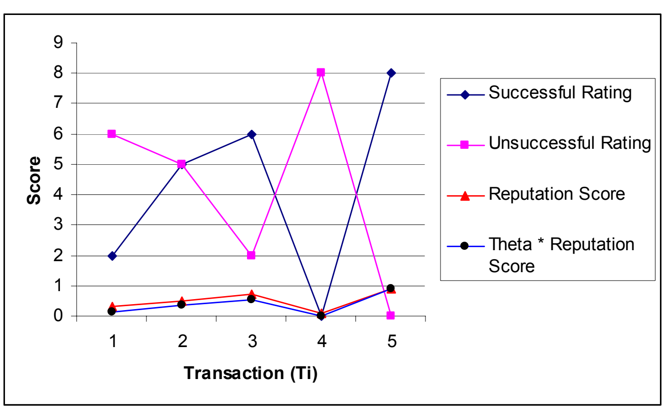

Fig. 2. Comparative view of successful unsuccessful ratings with the reputation score.

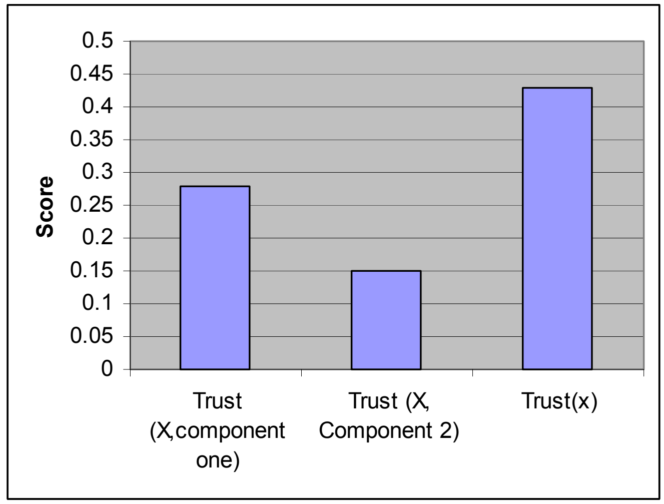

Fig. 3 Component wise contribution of overall trust computation

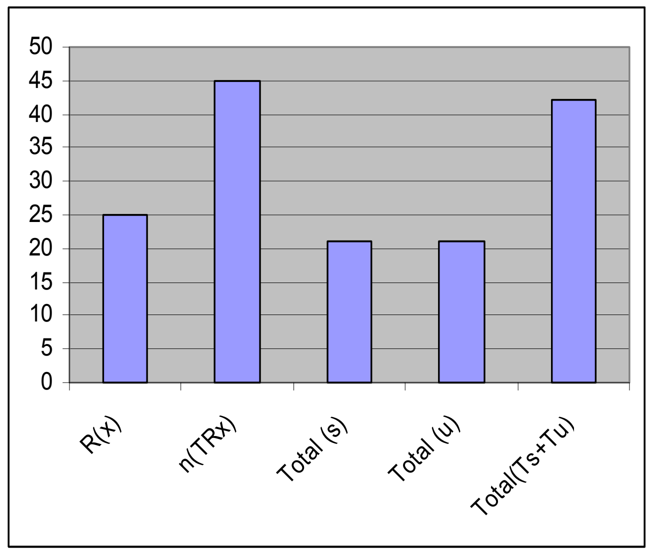

Fig. 4 Comparative view of ratings in component 1 and 2 of equation (1)

VI. Discussion and Conclusion

In this paper we have attempted to establish *TruMet* as a trust metric system for a multi agent environment. Our contribution in this has been particularly in deciding the trust parameters, applying them into the trust metrics model, gathering ways to compute aggregate ratings, define a referral network, and apply an approach towards filtering possibly unfair ratings. We are of the opinion that the establishment of the trust parameters and computation of trust value using those parameters have contributed in formulating a simple approach towards trust computation. The evaluation of the model in section V provides an example which results in a computation of a trust value. We hope that this simplicity can enable many users and systems to apply such trust metric system.

A limitation of our model is that it considers only the exogenous approach to filter the possibly unfair ratings. While the application of endogenous approaches could also have been meaningful in filtering the unfair ratings, our model does not currently incorporate this approach.

Future work in this model could be in the refinement of the model. Particularly, finding out ways to merge the Endogenous and Exogenous approaches to filtering of unfair ratings could be a really meaningful research. Such a combined model can be expected to filter the ratings and provide users a fair and reliable system. Contribution of such an enhanced filter can also be embedded in the current metrics model to provide better measurement capabilities.

VII. References

[1] L. Xiong and L. Liu, “PeerTrust: Supporting Reputation-Based Trust for Peer-to-Peer electronic communities”, *IEEE Transactions on Knowledge and Data Engineering,* vol. 16, no. 7, 2004.

[2] A. Whitby, A. J∅sang and J. Indulska, “Filtering Out Unfair Ratings in the Bayesian Reputation Systems ”, *Icfain Journal of Management Research,* vol. IV, no. 2, pp.48-64, 2005.

[3] C. Dellarocas, “Immunizing online Reputation Reporting Systems Against Unfair Ratings and Discriminatory Behavior”, *Proc. Second ACM conf. Electronic Commerce,* 2000.

[4] C. Dellarocas, “The Design of Reliable Trust Management Systems for Electronic Trading Communities”, *Working Paper Sloan School of Management MIT,* 2000.

[5] M.Chen and J.P. Singh, “Computing ad Using Reputations for Internet Ratings”, *Proc. Third ACM conf. Electronic Commerce,* 2001.

[6] W. T. Teacy, J. Patel, N.R. Jennings and M. Luck, “Coping with Inaccurate Reputation Sources: Experimental Analysis of a Probabilistic Trust Model”, *Proc. ACM AAMAS,* 2005.

[7] M. Schillo, P. Funk and M. Rovatsos, “Using Trust for Detecting Deceitful Agents in Artificail Societies”, *Applied Artificial Intelligence Journal, Special Issue on Trust, Deception and Fraud in Agent Societies,* 2000.

[8] E.M. Maximilien and M. P. Singh, “Agent-Based Trust Model Involving Multiple Qualities”, *Proc. ACM AAMAS,* 2005.

[9] B. Yu and M.P. Singh, “Detecting Deception in Reputation Management”, *Proc. ACM AAMAS,* 2003.

[10] A. J∅sang, C. Keser and T. Dimitrakos, “Can We Manage Trust?”, *Proc. Third International Conference on Trust Management (iTrust),* 2005.

[11] N. Littlestone and M.K.Warmuth, “The Weighted Majority Algorithm”, *Information and Computation,* vol. 2, no. 108, 1994.