A Specification of the Agent Reputation and Trust (ART) Testbed: Experimentation and Competition for Trust in Agent Societies

Karen K. Fullam[1], Tomas B. Klos[2], Guillaume Muller[3], Jordi Sabater[4], Andreas Schlosser[5], Zvi Topol[6], K. Suzanne Barber[1], Jeffrey S. Rosenschein[6], Laurent Vercouter[3], and Marco Voss[5]

[1]Laboratory for Intelligent Processes and Systems, University of Texas at Austin, USA
[2]Center for Mathematics and Computer Science (CWI), Amsterdam, The Netherlands
[3]SMA/G2I—École Nationale Supérieure des Mines, Saint-Étienne, France
[4]Institute of Cognitive Science and Technology (ISTC), National Research Council (CNR), Rome, Italy
[5]IT Transfer Office, Darmstadt University of Technology, Darmstadt, Germany
[6]Multiagent Systems Research Group—Critical MAS, Hebrew University, Jerusalem, Israel

ABSTRACT

A diverse collection of trust-modeling algorithms for multi-agent systems has been developed in recent years, resulting in significant breadth-wise growth without unified direction or benchmarks. Based on enthusiastic response from the agent trust community, the Agent Reputation and Trust (ART) Testbed initiative has been launched, charged with the task of establishing a testbed for agent trust- and reputation-related technologies. This testbed serves in two roles: (1) as a competition forum in which researchers can compare their technologies against objective metrics, and (2) as a suite of tools with flexible parameters, allowing researchers to perform customizable, easily-repeatable experiments. This paper first enumerates trust research objectives to be addressed in the testbed and desirable testbed characteristics, then presents a competition testbed specification that is justified according to these requirements. In the testbed's artwork appraisal domain, agents, who valuate paintings for clients, may gather opinions from other agents to produce accurate appraisals. The testbed's implementation architecture is discussed briefly, as well.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*multiagent systems*

General Terms

Experimentation

Keywords

trust, reputation, competition testbed, multi-agent systems

AAMAS'05, July 25-29, 2005, Utrecht, Netherlands.

1. INTRODUCTION

Social interactions in multi-agent systems generate the overarching research problem of modeling inter-agent trust. To accomplish its goals, an agent often requires resources only other agents can provide. The agent benefits from ensuring interactions are as successful as possible: promised resources are delivered on time and are of high quality. Choosing to interact puts the agent at risk since agreements may not be fulfilled. An agent can attempt to minimize this risk by interacting with those agents it deems most likely to fulfill agreements. Toward this goal of minimizing risk, the agent must predict the outcome of interactions (will agreements be fulfilled?), avoiding risky, or unreliable, agents. Modeling the trustworthiness of potential interaction partners enables the agent to make these predictions.

A wide variety of trust-modeling algorithms has been developed in recent years, resulting in significant breadth-wise growth. However, a unified research direction has yet to be established. Many experimental domains and metrics have been utilized, but unified performance benchmarks for comparing technologies, in spite of representational differences, have been neglected. In recent years, researchers [2, 13, 22] have recognized objective standards are necessary to justify successful trust modeling systems and provide a baseline of certifiable strategies for future work. For trust algorithms and representations to cross over into application [5, 8], the public must be provided with system evaluations based on transparent, recognizable standards for measuring success.

As a versatile, universal experimentation site, a competition testbed fosters a cohesive scoping of trust research problems; researchers are united toward a common challenge, out of which come solutions to these problems via unified experimentation methods. Through objective, well-defined metrics, a testbed can provide researchers with tools for comparing and validating their approaches. A testbed also serves as an objective means of presenting technology features—both advantages and disadvantages—to the research community. In addition, a competition testbed places trust research in the public spotlight, improving confidence in the technology and highlighting relevant applications.

Based on enthusiastic response from the agent trust community, the Agent Reputation and Trust (ART) Testbed initiative [26] has been launched, charged with the task of establishing a testbed for agent trust- and reputation-related technologies. This international team of ten researchers from six countries has been formed to co-

ordinate domain design, game specification, testbed development, and competition administration. This paper proposes a competition testbed specification in an artwork appraisal domain; agents, who valuate paintings for clients, may gather opinions from other agents to produce accurate appraisals. The competition testbed is designed to serve in two roles: (1) as a competition forum in which researchers can compare their technologies against objective metrics, and (2) as a suite of tools with flexible parameters, allowing researchers to perform customizable, easily-repeatable experiments.

Section 2 describes testbed requirements by summarizing the most important trust research problems and valuable design characteristics, evaluating several related experimental environments which fall short of these research problem and design goals. In Section 3, the art appraisal domain specification is presented, explaining rules for both competition and experimentation modes of operation. Section 4 gives an overview of the implementation architecture, while Section 5 demonstrates how the testbed design facilitates solutions to the required research problems and incorporates each necessary design characteristic. Finally, Section 6 concludes, presenting plans for future work, including completion of testbed implementation and presentation to the research community for experimental evaluation.

2. TESTBED REQUIREMENTS

Before we can present a testbed specification, prominent trust research objectives and desirable testbed characteristics must be enumerated. The following subsections, briefly discussing requirements elaborated upon by Sabater [22] and Fullam and Barber [13], lay this foundation for testbed design and implementation.

2.1 Trust Research Objectives

To design the framework of an effective competition testbed, the research community must come to agreement regarding its primary research objectives, ensuring that the competition testbed facilitates solutions toward those problems. To minimize the risk of interacting with others, an agent must be able to both (1) model trustworthiness of potential interaction partners, and (2) make decisions based on those models. First, research objectives include building trust models that are:

Accurate: A trust model must be a correct predictor of another agent's future behavior. Accuracy of trust models can be measured in terms of the similarity between the agent's calculated trust model and the trusted entity's true trustworthiness [12, 18, 27].

Adaptive: Trust models must change to accommodate dynamic trustworthiness characteristics of other agents, who might suddenly lose competence or maliciously employ strategies to vary trustworthiness [14].

Quickly Converging: Trust modeling algorithms must be able to quickly create usable new models when unknown agents enter the system. Quick trust model bootstrapping is necessary to thwart agents attempting to change identities by repeatedly entering and leaving a system, and can be assessed by the time to converge to sufficiently accurate models [7].

Multidimensional: Trust models must be able to distinguish between another agent's varied trustworthiness characteristics across multiple categories [19].

Efficient: Trust algorithms must construct models with minimal computational cost and time [16, 28]. Computational efficiency can be gauged by time to complete a trust model update.

Second, research objectives also encompass an agent's ability to effectively translate its trust models to make decisions and take actions, such as:

Identifying and isolating untrustworthy agents: Agents have to be able to identify and isolate untrustworthy agents by refusing to interact with them [3, 4].

Evaluating an interaction's utility: An agent must estimate the utility of an interaction, or degree to which the agreement will be fulfilled, to better negotiate terms of the agreement, such as appropriate payment [20].

Deciding whether and with whom to interact: Given a group of trustworthy agents with which interaction potentially has a high utility, an agent must correctly choose its partner and predict whether the agreement they establish will be fulfilled. For example, successful trusting decisions can be defined by the number of positive interactions compared to total interactions [11, 24].

The artwork appraisal domain problem, detailed in Section 3, motivates participants to develop adaptive, efficient trust modeling algorithms for agent appraisers, capable of quickly forming accurate, multidimensional trust assessments of other appraisers. In addition, successful appraisers correctly determine when to use opinions from other appraisers and estimate the utility of those opinions, isolating inaccurate appraisers.

2.2 Testbed Characteristics

Now that the trust research community's research problems have been crystallized into a unified set of goals, a competition testbed can be designed to facilitate achievement of those objectives. Several desirable properties are essential for an effective competition testbed:

Modularity: Modularity permits the testing of a wide range of capabilities through adjustable parameters by which the agent environment changes according to experimenter or competition goals. Not only does parameterization allow the researcher flexibility while conducting experiments; in a competition setting, the environmental dynamics of the contest can be changed, requiring players to adapt their strategies between competitions.

Multipurpose Design: The testbed must allow researchers to both participate in competitions and use the testbed for independent experimentation.

Accessibility: The testbed must not restrict the wide range of trust-modeling algorithms used by researchers, instead providing easy, standardized "hook-up" capability regardless of individual agent's trust representations.

Objective Metrics: Testbed metrics should include objective success measures from both the single-agent and system-wide perspectives.

Problem Focus: The testbed scenario must be structured such that relevant trust problems are addressed, while out-of-scope research areas, such as belief revision, planning, or domain knowledge, are excluded.

The testbed specification explained in Section 3 is designed to incorporate these desirable characteristics.

2.3 Critique of Existing Experimentation Environments

Established experimentation environments have fallen short of the testbed requirements described in Sections 2.1 and 2.2. The Prisoner's Dilemma [1] is a well-established game in which two players can each choose to cooperate or defect, receiving utilities based on both player's choices. The recent Prisoner's Dilemma competition [17] has several drawbacks. First, multidimensional trust modeling is not promoted, since agents only evaluate one aspect of opponents' behavior. Second, agents have no opportunity to isolate untrustworthy opponents, since they are forced to interact with all competitors. The competition lacks objective, system-based metrics, instead focusing on a single, agent-based metric of total utility. Finally, the competition lacks trust problem focus, since agents can employ game theoretic strategies with minimal trust-modeling skills.

The SPORAS experiments [30], measuring time for electronic marketplace reputation models to converge to true reputations, have been widely-utilized for testing ReGreT [23], AFRAS [6], Yu and Singh's model [29], as well as two online reputation mechanisms (eBay [8] and Bizrate [5]). However, these experiments are too narrow in scope, evaluating reputation models based on only single-agent metrics, such as time to converge, while neglecting system-based success measures. The SPORAS experiments do not consider multidimensional trust, nor do they compare multiple trust-modeling strategies in a competition setting. While this experiment set emphasizes trust model accuracy, adaptivity, and efficiency, it does not measure an agent's ability to make trust-based decisions, such as determining whether to interact or isolating untrustworthy agents.

Schlosser et al. [25] propose a framework for evaluating reputation systems by simulation that is freely configurable by the user, provides several objective metrics, and is focused on trust problems, ignoring out-of-scope research areas. However, the experimentation environment is not easily extended to compare multiple trust-modeling algorithms in competition against each other.

3. TESTBED SPECIFICATION

The testbed operates in two modes: competition and experimentation. In competition mode, the testbed compares different researchers' strategies as they act in combination. Each participating researcher controls a single agent, which works in competition against every other agent in the system. The competition consists of several game sessions; the winner is selected by averaging results over all game sessions, to even out possibly unfair game settings. The duration of each session is randomly determined by the simulation and is unknown to each agent to prevent agents from exploiting end-game strategies. In each game session, 'dummy' agents, whose strategies are unknown to the other competitors, may be included in the competition to increase the number of players. In competition mode, dummy agents compete in the game throughout the duration of the competition. Adjustable parameters described here permit the game structure to be adapted for subsequent competitions.

To utilize the testbed's experimentation mode, the completed testbed will be downloadable for researcher use independent of the competition. Thus, results may be compared among researchers for benchmarking purposes, since the testbed provides a well-established environment for easily-repeatable experimentation. In experimentation mode, researchers may choose to allow agents (including dummy agents) to enter or leave the game as desired. The researcher also has the flexibility of complete control over all experiment parameters, further detailed below.

In the art appraisal domain, agents function as painting appraisers with varying levels of expertise in different artistic eras. Clients request appraisals for paintings from different eras; if an appraising agent does not have the expertise to complete the appraisal, it can request opinions from other appraiser agents. Appraisers receive more clients, and thus more profit, for producing more accurate appraisals. The following subsections outline the details of the domain problem and testbed rules, defining the appraising capabilities held by agents, appraisal information transactions conducted between agents, and exchange of reputation information among agents. In addition, system and individual agent metrics are discussed within both competition and experimentation modes. Later, Section 5 details how the requirements of Section 2 are satisfied by this specification. While this paper justifies important design decisions, further rationale can be found at the ART Testbed website's 'Frequently Asked Questions' page [26].

3.1 Client Appraisals

In each timestep, multiple clients present each appraiser (agent) with paintings to be appraised, paying a fixed fee f for each appraisal request. To increase business, appraisers attempt to valuate paintings as closely to market value as possible. A given painting may belong to any of a finite set of eras (a painting's era is known by all appraisers), and appraisers have varying levels of expertise in each era. An appraiser's expertise, defined as its ability to generate an 'opinion' about the value of a painting, is described by a normal distribution of the error between the appraiser's opinion and the true painting value. The simulation creates opinions according to this error distribution, which has a mean of zero and a standard deviation s given by

$$s = (s^* + \frac{\alpha}{c_g})t$$

where s^*, unique for each era, is assigned to an appraiser from a uniform distribution. t is the true value of the painting to be appraised and α is a parameter, chosen by the experimenter and fixed for all appraisers, relating opinion-generation cost to resulting opinion accuracy. c_g, the cost an appraiser is willing to pay to generate an opinion, is discussed in more detail below. An appraiser's expertise for each era does not change throughout the duration of a game, and appraisers know their levels of expertise for each era. However, the simulation does not inform appraisers of other appraisers' expertise levels. The true values of paintings presented by clients are chosen from a uniform distribution known only to the simulation; likewise, the eras to which paintings belong are also uniformly distributed among the set of eras. In experimentation mode, the researcher has flexibility to adjust appraiser expertise levels, true painting values, and number/types of painting eras.

An appraiser chooses a variable cost c_g, representing time taken to examine the painting, to pay in generating its own opinion about a painting's value. An appraiser is required to pay a minimum c_g of one monetary unit. By paying a higher cost c_g, analogous to spending more time studying the painting, an appraiser increases the accuracy of its opinion. However, an appraiser cannot perfectly judge a painting by spending an infinite amount of time studying it; the appraiser's accuracy is still limited by its expertise. Therefore, an appraiser cannot infinitely increase its accuracy by paying an extremely large value c_g; the minimum achievable error distribution standard deviation is $s^* \cdot t$. In addition to generating its own opinion, an appraiser may request opinions from other appraisers to improve its final appraisal. Request opportunities are especially important when an appraiser attempts to valuate paintings from eras for which it has low expertise.

Appraisers may request opinions from as many other appraisers

as desired for each painting, at a fixed cost c_p for each opinion transaction. In general, $c_p \ll f$ (where f is the price a client pays for an appraisal) to encourage opinion exchange. Appraisers may also provide opinions for as many paintings as desired in a single timestep. Appraisers are not required to truthfully reveal their opinions; they can communicate false opinions if desired, for example, in attempting to decrease the requester's client base and resulting profit. In addition to requesting opinions about a painting, an appraiser may similarly request reputation information about other appraisers, as discussed further in Section 3.3. The simulation oversees each portion of a timestep synchronously, including client requests, appraiser opinion generation, transactions between appraisers, and returning final appraisals to clients. Therefore, appraisers are required to perform required actions for each timestep portion within real-time limitations as monitored by the simulation.

3.2 Opinion Transactions

A central directory lists all appraisers in the system to assist appraisers in initiating opinion requests. However, no centralized information is maintained related to the trustworthiness of appraisers. In competition mode, transaction communication protocols are strictly regulated as described here and in Section 3.3; however, in experimentation mode, researchers may implement additional message formats.

The opinion transaction protocol, contained in a single timestep, is shown in Figure 1. To initiate an opinion transaction, a requester

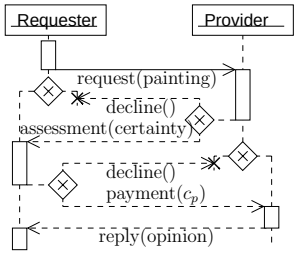

Figure 1: Opinion transaction protocol.

sends a request message to another appraiser (potential opinion provider), identifying the painting to be appraised. Upon receiving an opinion request, if the potential provider is willing to provide the requested opinion, it responds by sending a certainty assessment about the opinion it can provide, defined as a real number between zero and one (one represents complete *asserted* certainty). The potential provider is not required to provide a truthful certainty assessment. If the potential provider does not wish to participate in the requested transaction, it may choose to decline the request. By sending a certainty assessment, the provider promises to deliver the requested opinion should the certainty assessment be accepted by the requester.

After receiving the provider's certainty assessment, the requester either sends payment to the provider if it chooses to accept the promised opinion, or sends a 'decline' message if it chooses not to continue the transaction. The cost of each transaction is the non-negotiable amount c_p. Upon receipt of payment, the provider may choose to send an untruthful opinion or no opinion at all.

Upon paying providers, but before receiving opinions from providers, the requesting appraiser is required to submit to the simulation its roster of opinion providers and a set of corresponding weights. Weights are values between zero and one, loosely representing the appraiser's confidence or trust in each provider's opinion. Although researchers are permitted to internally represent an appraiser's trust in any desired form, a standardized format for communicating reputation information is required. In lieu of sending reputation information according to their internal trust models, appraisers are only permitted to communicate reputation information in the form of weights. Then the requester sends the set of opinions directly to the simulation upon receipt from providers.

The appraiser's final appraisal p^* is calculated *by the simulation* as a weighted average of received opinions:

$$p^* = \frac{\sum_i (w_i \cdot p_i)}{\sum_i (w_i)},$$

where w_i and p_i are the appraiser's weight for, and received opinion from, each provider i whose opinion it received (possibly including itself). The true painting value t, along with the calculated final appraisal p^*, is then revealed by the simulation to the appraiser. The simulation enforces this roster submission and final appraisal calculation protocol; requesting appraisers are not permitted to change their rosters or alter received opinions from providers. The simulation calculates final appraisals to prevent appraisers from developing non-trust-based appraisal calculation strategies and allow appraisers to focus on the more important task of assessing and selecting trustworthy opinion providers. Upon learning its final appraisal and the painting's true value, an appraiser may use this feedback to revise its trust models of other appraisers.

Each appraiser has a bank account, monitored by the simulator, from which it pays transaction costs and into which are deposited client appraisal fees. Bank accounts are instantiated with zero balances and may hold negative balances. Competing appraisers can not observe each other's clients or bank account balances.

3.3 Reputation Transactions

In addition to conducting opinion transactions, appraisers can exchange reputations, or information about the trustworthiness of other appraisers. The reputation transaction protocol is shown in Figure 2.

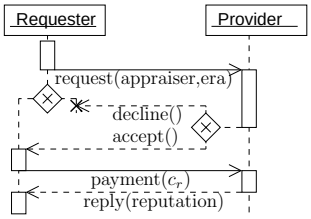

Figure 2: Reputation transaction protocol.

A requester sends a request message to a potential reputation provider, identifying the appraiser about whom (and era with respect to which) it is requesting reputation information. Upon receiving a reputation request, the potential reputation provider sends an 'accept' or 'decline' message depending on whether it is willing to provide the requested reputation. The requester then sends payment to the provider to receive the requested information. The cost of each reputation transaction is a non-negotiable amount c_r. Upon receipt of payment, the provider reports a reputation of the same form as weights submitted to the simulation for final appraisal calculation. The provider is not required to send its actual reputation value, neither is the provider forced to send any reputation value at all. In general, $c_r \ll c_p$ (where c_p is the cost of an opinion) to promote exchange of reputation information.

Appraisers are not required to report to requesting appraisers the same weights submitted to the simulation. Although these weights represent the providing appraiser's subjective trust measures, a requesting appraiser can learn how to interpret the provider's weights after observing the relationships among several weights sent by the same provider over time.

3.4 Assigning Client Shares

Although clients are initially evenly distributed among appraisers, those appraisers whose final appraisals are most accurate are rewarded with a larger share of the client base in subsequent timesteps. To calculate each appraiser's share of the client base, each appraiser a's average relative appraisal error, ϵ_a is first calculated:

$$\epsilon_a := \frac{\sum_{c \in C_a} \frac{|p_c^* - t_c|}{t_c}}{|C_a|},$$

where C_a is the set of appraiser a's clients, p_c^* is appraiser a's final appraisal for client c, and t_c is the true value of the painting client c submitted to a for appraisal.

Next, each appraiser a is assigned a preliminary client share $\tilde{r}_a$ according to its average relative appraisal error:

$$\tilde{r}_a = \left(\frac{\delta_a}{\sum_{b \in A} \delta_b} \right) \cdot |C|,$$

where A is the set of all appraisers, C is the set of all clients, and $\delta_a := 1 - \frac{\epsilon_a}{\sum_{b \in A} \epsilon_b}$.

Thus, the appraiser with the least average relative appraisal error achieves the highest preliminary client share. Finally, each appraiser a's actual client share r_a depends on the appraiser's client share from the previous timestep:

$$r_a = q \cdot r'_a + (1 - q) \cdot \tilde{r}_a,$$

where r'_a is appraiser a's client share in the previous timestep. The parameter q, a value between zero and one inclusive, reflects the influence of previous client share size on next client share size. Thus the volatility in client share magnitudes due to frequent accuracy oscillations is reduced for larger values of q, which is chosen by the experimenter and is the same for calculating all appraisers' client shares. Client shares are rounded to the nearest positive integer, while keeping the total number of clients constant and ensuring each appraiser has at least one client in each timestep.

3.5 Analysis Metrics

Appraiser strategies are analyzed in terms of both agent- and system-based metrics. The agent perspective examines the utility of a strategy to a single appraiser without regard for the benefit to the overall agent system. Because appraisers are permitted to employ any trust modeling techniques desired, assessing the accuracy of an appraiser's trust models, or comparing two appraisers' trust models, is difficult without restricting representation. Instead, the testbed assesses the quality of appraisers' trust-based decisions (cf. Section 2.1). In the testbed's appraisal scenario, appraisers attempt to accurately valuate their assigned paintings; their decisions about which opinion providers to trust directly impact the accuracy of their final appraisals. Therefore, in competition mode, the winning agent is selected as the appraiser with the highest bank account balance. In other words, the appraiser who is able to (1) estimate the value of its paintings most accurately and (2) purchase information most prudently, is deemed most successful. The testbed also provides functionality to compute the average accuracy of the appraiser's final appraisals and the consistency of that accuracy, represented as its final appraisal error mean and standard deviation, respectively. In addition, the quantities of each type of message passed between appraisers are recorded.

The system perspective employs metrics that emphasize social welfare, or benefit to the appraiser network as a whole. The testbed collects data such as system-wide bank account total, distribution of earnings among appraisers, number of messages passed (distinguished by type), number of transactions, and transaction distribution across appraisers. Finally, the testbed provides methods allowing researchers to define additional metrics and collect relevant data.

4. IMPLEMENTATION ARCHITECTURE

As shown in Figure 3, the testbed architecture, implemented

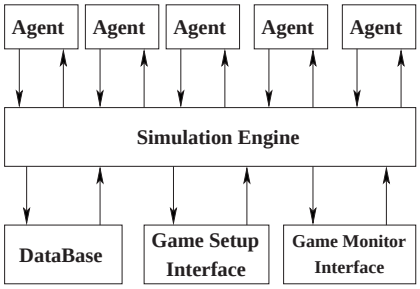

Figure 3: Competition testbed architecture.

in Java, consists of four components: (1) Simulation Engine, (2) Database, (3) User Interfaces, and (4) Agent Skeleton (see [15] for a detailed description of the ART Testbed architecture). The Simulation Engine is responsible for initiating the game and controlling the simulation environment by enforcing chosen parameters. Thus the Simulation Engine is also responsible for assigning clients to appraisers and coordinating communication among appraisers. In each timestep, the Simulation Engine manages appraiser actions, the generating of opinions by appraisers, opinion transactions, reputation transactions, calculation of final appraisals, and allocation of client shares.

Through the Simulation Engine, the Database collects environment and agent data, such as true painting values, opinions, transaction messages, calculated final appraisals, client share allocations, and bank balances. In addition, the testbed provides researchers with the functionality to log additional data types in the Database during experimentation mode. With access tools for navigating Database logs, data sets are made available to researchers after each game session for game re-creation and experimental analysis.

User Interfaces permit researchers to observe games in progress and access Database information by graphically displaying details such as transactions between appraisers, accuracy of appraisers' final appraisals, and money earned by each appraiser. Since some of these details are viewable by researchers in real time but are not programmatically available to appraiser agents, the competition forbids improving appraiser agents with 'on-the-fly' programming while the game is in progress. In experimentation mode, one User Interface allows researchers to initiate game sessions, setting parameters such as number of timesteps per game, number of appraisers, or appraiser expertise levels.

The Agent Skeleton is designed to allow researchers to implant customized internal trust representations and trust revision algorithms while permitting standardized communication with entities external to the appraiser agent. Java classes defining the Agent Skeleton implement all necessary interfaces to permit inter-agent

communication (via the Simulation Engine). The Agent Skeleton is also equipped to handle coordination tasks with the Simulation Engine, such as opinion formation and appraisal calculation. Developing agent skeletons using additional programming languages may be attempted as future work, to offer additional flexibility to the agent designer.

5. SPECIFICATION CRITIQUE

This section justifies the proposed testbed specification, demonstrating its superiority over other previously developed experimental environments. A testbed domain is required in which agents are less knowledgable than others about needed information to encourage information exchange. The art appraisal domain satisfies this requirement by limiting appraiser expertise so appraisers must request opinions from more skilled appraisers. In an appropriate domain, competitive agents must be tempted to cheat, yet also have an incentive to cooperate. This cheat vs. cooperate dilemma is evident in the art appraisal domain, since appraisers may save money by generating inaccurate opinions, yet sell more opinions by being accurate.

The art appraisal domain accommodates the research problems of Section 2.1, first by addressing trust-modeling accuracy; appraisers must know accurately which other appraisers are most trustworthy to provide correct opinions. Appraiser trust models must also be adaptive, since providers can change both opinion- and reputation-sharing strategies at any time. Fast trust model convergence is important, as well, because appraisers must build quickly their initial trust models of other appraisers, which allows a wide range of bootstrapping strategies. Multidimensional trust modeling is encouraged, since opinion providers have varying degrees of expertise, and possibly truthfulness, for each era; also, appraisers can choose (but are not required) to model the trustworthiness of other appraisers as both opinion providers and reputation providers. Further, appraisers must employ efficient trust modeling techniques to complete computation within the time allotted by the simulation.

Appraisers are judged by their decision-making abilities, as well. Appraisers must determine whether to interact by weighing the cost of purchasing an opinion against the value of receiving the information, depending on how trustworthy it perceives the opinion provider to be. Opinion-generating appraisers evaluate the utility of opinion-providing interactions by deciding how much to pay to generate an opinion. In addition, appraisers identify and isolate untrustworthy appraisers by refusing to buy opinions and reputations if the information is predicted to be too inaccurate.

Finally, necessary testbed characteristics from Section 2.2 are incorporated into the specification outlined here. To permit modularity, numerous parameters, such as transaction costs and appraiser expertise settings, are adjustable by the experimenter. The testbed is designed to be multipurpose, as well, serving in both competition and experimentation modes. Researchers find the testbed accessible for a wide range of trust modeling techniques, since all necessary communication interfaces are provided with the Agent Skeleton architecture. The metrics detailed in Section 3.5 give numerous objective measures for both agent- and system-based success. Lastly, the testbed demonstrates appropriate focus, avoiding out-of-scope research areas: 1) appraisers need not develop statistical appraisal calculation strategies since the simulation calculates final appraisals, 2) appraisers do not require painting appraisal domain knowledge since opinions are generated by the simulation, and 3) fixed transaction costs keep appraisers from needing price negotiation strategies.

6. CONCLUSIONS

This paper motivates the development of a comprehensive Agent Reputation and Trust competition testbed. A testbed provides researchers with easy access to a common experimentation environment and allows researchers to compete against one another to determine the most viable technology solutions. The art appraisal testbed design addresses prominent trust research problems related to an agent's ability to model trust (accurately, adaptively, quickly, multidimensionally, and efficiently) and make decisions based on trust (determining whether to interact, evaluating the utility of an interaction, and identifying and isolating untrustworthy agents). In addition, the testbed incorporates important characteristics such as modularity, multi-purpose design, accessibility, use of objective metrics, and appropriate problem focus. Through its coverage of these requirements, the testbed exceeds standards set by existing experimentation environments.

Plans for a competition testbed have received broad support from the agent trust research community. Upon completion of implementation, the testbed will be presented to the community for evaluation. Based on prototype testbed experimental review and feedback from the research community, the first testbed competition will be conducted in July of 2006. Development progress can be monitored through the ART Testbed website [26], where updates to competition development are posted periodically.

7. ACKNOWLEDGMENT

Jordi Sabater enjoys a Sixth Framework Programme Marie Curie Intra-European fellowship, contract No. MEIF-CT-2003-500573. Tomas Klos' research is sponsored by the Dutch government under project number TSIT2021 (SENTER). Karen Fullam's research is sponsored by DARPA TASK, F30602-00-2-0588.

8. REFERENCES

[1] R. Axelrod. *The Evolution of Cooperation*. New York: Basic Books, 1984.

[2] K. Barber, K. Fullam, and J. Kim. Challenges for Trust, Fraud, and Deception Research in Multi-agent Systems. In *Trust, Reputation, and Security: Theories and Practice*, volume 2631 of *LNCS*, pages 8–14. Springer, 2003.

[3] K. Barber and J. Kim. Belief Revision Process Based on Trust: Agents Evaluating Reputation of Information Sources. In R. Falcone, M. Singh, and Y.-H. Tan, editors, *Trust in Cyber-Societies*, volume 2246 of *LNAI*, pages 73–82. Springer, 2001.

[4] A. Biswas, S. Sen, and S. Debnath. Limiting Deception in Groups of Social Agents. In *Proc. of the Agents 1999 Workshop on Deception, Fraud and Trust in Agent Societies*, pages 21–28, 1999.

[5] BizRate. *BizRate*. http://www.bizrate.com, 2002.

[6] J. Carbo, J. Molina, and J. Davila. Trust Management Through Fuzzy Reputation. *Int. J. in Cooperative Information Systems*, 12(1):135–155, 2002.

[7] L. Ding, P. Kolari, S. Ganjugunte, T. Finin, and A. Joshi. On Modeling and Evaluating Trust Networks Inference. In Falcone et al. [9], pages 21–32.

[8] eBay. *eBay*. http://www.eBay.com, 2002.

[9] R. Falcone, K. Barber, J. Sabater, and M. Singh, editors. *Proc. of the AAMAS-2004 Workshop on Trust in Agent Societies*, 2004.

[10] R. Falcone, S. Barber, L. Korba, and M. Singh, editors. *Proc. of the Agents 2001 Workshop on Deception, Fraud and Trust*, 2001.

[11] R. Falcone, G. Pezzulo, C. Castelfranchi, and G. Calvi. Trusting the Agents and the Environment Leads to Successful Delegation: A Contract Net Simulation. In Falcone et al. [9], pages 33–39.

[12] K. Fullam. An Expressive Belief Revision Framework Based on Information Valuation. Master's thesis, Dept. of Electrical and Computer Engineering, U. Texas (Austin), 2003.

[13] K. Fullam and K. S. Barber. Evaluating Approaches for Trust and Reputation Research: Exploring a Competition Testbed. In Paolucci et al. [21], pages 20–23.

[14] K. Fullam and K. S. Barber. A Temporal Policy for Trusting Information. In Falcone et al. [9], pages 47–57.

[15] K. Fullam, T. Klos, G. Muller, J. Sabater, Z. Topol, K. S. Barber, J. S. Rosenschein, and L. Vercouter. The Agent Reputation and Trust (ART) Testbed Architecture. In *Proc. Trust Workshop at AAMAS*, 2005.

[16] R. Ghanea-Hercock. The Cost of Trust. In Falcone et al. [9], pages 58–64.

[17] G. Kendall, P. Darwen, and X. Yao. *The Iterated PD Competition*. http://www.prisoners-dilemma.com/, 2004.

[18] T. Klos and H. L. Poutré. Using Reputation-Based Trust for Assessing Agent Reliability. In Falcone et al. [9], pages 75–82.

[19] G. Muller, L. Vercouter, and O. Boissier. Towards a general definition of trust and its application to openness in MAS. In R. Falcone, K. Barber, L. Korba, and M. Singh, editors, *Proc. of the AAMAS-2003 Workshop on Deception, Fraud and Trust*, 2003.

[20] B. Neville and J. Pitt. A Simulation Study of Social Agents in Agent Mediated E-Commerce. In Falcone et al. [9], pages 83–91.

[21] M. Paolucci, J. Sabater, R. Conte, and C. Sierra, editors. *Proc. IAT Workshop on Reputation in Agent Societies*, 2004.

[22] J. Sabater. Toward a Test-Bed for Trust and Reputation Models. In Falcone et al. [9], pages 101–105.

[23] J. Sabater and C. Sierra. REGRET: A reputation model for gregarious societies. In Falcone et al. [10], pages 61–69.

[24] M. Schillo, P. Funk, and M. Rovatsos. Using Trust for Detecting Deceitful Agents in Artificial Societites. *Applied Artificial Intelligence*, Special Issue on Trust, Deception and Fraud in Agent Societies:825–848, 2000.

[25] A. Schlosser, M. Voss, and L. Brückner. Comparing and Evaluating Metrics for Reputation Systems by Simulation. In Paolucci et al. [21].

[26] ART Testbed Team. *Agent Reputation and Trust Testbed*. http://www.lips.utexas.edu/~kfullam/competition/, 2005.

[27] A. Whitby, A. J. sang, and J. Indulska. Filtering Out Unfair Ratings in Bayesian Reputation Systems. In Falcone et al. [9], pages 106–117.

[28] H. Yamamoto, K. Ishida, and T. Ohta. Trust Formation in a C2C Market: Effect of Reputation Management System. In Falcone et al. [9], pages 126–136.

[29] B. Yu and M. P. Singh. Towards a Probabilistic Model of Distributed Reputation Management. In Falcone et al. [10], pages 125–137.

[30] G. Zacharia. Collaborative Reputation Mechanisms for Online Communities. Master's thesis, Massachusetts Institute of Technology, September 1999.