A Design Foundation for a Trust-Modeling Experimental Testbed

Karen K. Fullam[1], Jordi Sabater-Mir[2], and K. Suzanne Barber[1]

[1] Laboratory for Intelligent Processes and Systems
The University of Texas at Austin, Austin, TX 78712, USA
{kfullam,barber}@lips.utexas.edu
[2] Laboratory of Agent Based Social Simulation
ISTC-CNR, Viale Marx 15, 00137 Roma, Italy
jsabater@iiia.csic.es

Abstract. Mechanisms for modeling trust and reputation to improve robustness and performance in multi-agent societies make up a growing field of research that has yet to establish unified direction or benchmarks. The trust research community will benefit significantly from the development of a competition testbed; such development is currently in progress under the direction of the Agent Reputation and Trust (ART) Testbed initiative. A testbed can 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 easily-repeatable experiments. As a versatile, universal experimentation site, a competition testbed challenges researchers to solve the most prominent problems in the field, fosters a cohesive scoping of trust research problems, identifies successful technologies, and provides researchers with a tool for comparing and validating their approaches. In addition, a competition testbed places trust research in the public spotlight, improving confidence in the technology and highlighting relevant applications. This paper lays the foundation for testbed development by enumerating the important problems in trust and reputation research, describing important requirements for a competition testbed, and addressing necessary parameters for testbed modularity and flexibility. Finally, the ART Testbed initiative is highlighted, and future progress toward testbed development is described.

1 Introduction

Mechanisms for modeling trust and reputation to improve robustness and performance in multi-agent societies make up a growing field of research. A diverse collection of models and algorithms has been developed in recent years, resulting in significant breadth-wise growth. However, a unified research direction has yet to be established. In the pursuit of innovative trust theory, many experimental domains and metrics have been utilized. Yet, unified performance benchmarks which serve as the standards for comparing new technologies across representations have been neglected. In recent years, researchers [1–3] have recognized that the need for objective standards are necessary to justify successful trust modeling systems, rejecting inferior strategies and providing a baseline of certifiable strategies upon which to expand research and apply research results. As trust research matures, and trust modeling becomes an important tool in real-world use, some performance analysis must occur to assess relative worth among a multitude of emerging trust technologies. In order for trust algorithms and representations to crossover into application, the public must be provided with system evaluations based on transparent, recognizable standards for measuring success.

The trust research community would benefit significantly from the development of a competition testbed; such development is currently in progress under the direction of the Agent Reputation and Trust (ART) Testbed initiative [4]. A testbed can 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 easily-repeatable experiments. As a versatile, universal experimentation site, a competition testbed challenges researchers to solve the most prominent problems in the field. The development of a competition testbed can foster a cohesive scoping of trust research problems; researchers can be united toward a common challenge, out of which can come solutions to these goals via unified experimentation methods. Through the definition of objective, well-defined metrics, successful technologies can be identified and pursued; thus a testbed provides researchers with a tool 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.

This paper justifies the need for competition testbed development, explaining why current experimentation testbeds are insufficient. Further, this research initiates a movement toward testbed development by enumerating applicable research problems and desirable testbed characteristics. As a result, the paper is organized as follows. Section Two describes some experimental domains popular in trust research, explaining why each experimental setting falls short of achieving a unified testbed. In Section Three, the first task in designing a competition testbed is accomplished by enumerating the research objectives that must be addressed by the testbed's functionality. Desirable characteristics of a successful testbed are specified in Section Four, which also details important parameters that should be included to optimize the testbed's versatility. Finally, the Trust Competition Testbed Initiative is highlighted in Section Five, delineating future progress toward testbed development.

2 Existing Experimental Domains

Two approaches used by researchers to evaluate trust and reputation models are presented here: experiments based on the prisoner's dilemma game and common experiments used to compare SPORAS, ReGreT, AFRAS, and other systems. Both approaches fall short of the desired testbed capabilities for several reasons. First, neither has received universal acceptance within the trust research community. Second, each experimental domain is limited in flexibility and modularity, covering only a narrow range of scenarios. Finally, these experiment settings have failed to provide a competition environment in which researchers can compare their trust and reputation modeling strategies. Nonetheless, since these experimental domains are the most well-known within the research community, it is useful to discuss the characteristics of each in an effort to gain an understanding about useful testbed properties.

2.1 Prisoner's Dilemma Experiments

The prisoner's dilemma is a classic problem of game theory described in the following situation: two people have been arrested for a crime and placed in separate isolation cells. Each has two options, remain silent or confess. If both remain silent, each is subjected only to a reduced sentence (called the "payoff"). If one cooperates while the other remains silent, the confessor is set free while the other receives a harsh punishment. Finally, if both confess, each receives a moderate punishment. Each prisoner faces a "dilemma", since it is preferable to confess, yet the payoff when both confess is worse for each than the payoff when both remain silent. The iterated version of this game is the basis for several scenarios designed to evaluate trust and reputation models.

Schillo et al. [5] propose a disclosed iterated prisoner's dilemma with partner selection with a standard payoff matrix. It can be described as a five-step process:

1. Each player pays a stake.
2. Pairs of players are determined by negotiation and declaration of intentions. Agents are permitted to deceive others about their intentions. For this step, a contract net-like protocol is introduced that is executed until each player has had the chance to find a partner.
3. The prisoner's dilemma game is played, bearing in mind the previously declared intentions.
4. The results are published. Due to limited perception, each agent receives only the results of a subset of all players.
5. The payoffs are distributed.

Agents have a limited number of points, from which stakes are paid and to which payoffs are added. If an agent loses all its points, it must retire from the game.

Mui et al. [6] propose an iterated prisoner's dilemma game in which successful strategies yield greater descendant populations in the following generation. The game proceeds as follows: first, participants for a single game are chosen randomly from the population. After a generation, composed of a certain number of dyadic encounters between agents, an agent produces descendants in the next generation proportional to its success during the generation. The total population size is maintained from one generation to the next. Therefore, an increase in one agent's descendant population is balanced by a decrease in other agents' descendant populations.

The *Playground* experiment, designed by Marsh [7], consists of a cell grid, in which each cell may be occupied by at most one agent at a time. Agents have total freedom of movement. When an agent attempts to move into an occupied cell, the prisoner's dilemma game is played between the occupant and the visitor. Since an agent's range of vision is limited, its ability to move away from untrusted agents and toward trusted ones is limited. The *Playground* experiment makes use of concrete payoff matrices called situations. Participating agents know the payoff structures for all possible situations. For a given interaction, a random situation is chosen and the participants are informed. After both agents have chosen their actions, payoffs are made according to the given situation and each agent is permitted to adjust its trust values.

Prisoner's dilemma is a well-established game useful for trust experimentation. However, the game only allows players a Boolean action choice: whether to cooperate or not. In many trust-modeling cases, it is valuable to have more expressive choices representing degrees of trustworthiness. Recent papers [8] have investigated "continuous prisoner's dilemma", the possibility of extending this classical game by allowing a variable degree of cooperation, with payoffs scaled accordingly. This model, however, has not been used to compare trust and reputation models; a unified experimentation setting, agreed upon by the trust research community as a whole, is needed.

2.2 SPORAS, ReGreT and AFRAS

The set of experiments presented in this section was first used by Zacharia et al. [9] to test the SPORAS model. The same set of experiments was used by Sabater and Sierra [10] to compare the ReGreT model against SPORAS and the reputation mechanism used in Amazon auctions. Finally, an extended version of these experiments was proposed by Carbo et al. [11, 12] to compare their AFRAS model with SPORAS [9], ReGreT [10, 13], Yu and Singh's model [14] and two online reputation mechanisms (eBay [15] and Bizrate [16]).

The original experiment set focuses on convergence speed and abuse of prior performance. The convergence speed experiment proposes a marketplace scenario of a fixed number of traders with uniformly distributed, real-number reputations near a minimum reputation value. In each time period, traders are matched randomly, then they interact and rate each other according fulfillment of the transaction. The experiment measures the time for reputation models to reach true reputation values. In the abuse of prior performance scenario, a trader joins the marketplace, behaves reliably until reaching a high reputation value, then starts abusing the reputation to commit fraud. The experiment measures reputation models' ability to adapt to the new behavior.

Carbo et al. [11, 12] extend the set of experiments, studying the use of cooperation between agents to improve reputation value convergence and the impact of coalitions between sellers and buyers. However, even considering the extensions proposed by Carbo et al., the set of experiments remains too narrow in the range of problems addressed. In addition, these experiments are only oriented to evaluate reputation models based on single-agent metrics and do not consider the impact of that model on the society as a whole. A problem domain which more broadly encompasses the most prominent trust research objectives must be identified.

3 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 objectives. The following subsections summarize the most important research problems in the field and detail metrics previously used by researchers. Though a potential competition testbed domain problem may not explicitly express its challenge as a solution to these research goals, the domain problem should be designed such that solutions to these problems emerge as researchers attempt to compete within the problem domain. Once objectives of the research community have been unified, a domain problem, relevant to real-world applications, can be proposed that is suited to provide an arena for solving these research objectives. Care must be taken in designing domain parameters which test technologies against the identified research objectives. In addition, domain-specific metrics must be defined to provide a basis for experiment-based competition among researchers.

Social interactions in multi-agent systems generate the overarching research problem of modeling of inter-agent trust. To accomplish its goals, an agent often requires resources (tangible goods, information, or services) that only other agents can provide. It is to the agent's benefit to ensure that interactions are as successful as possible: that promised resources are delivered on time and are of high quality. Choosing to interact puts the agent at risk; agreements to exchange resources or avoid harmful activity may not be fulfilled. Resources the agent expects to receive may not be delivered, or resources the agent delivers may be used by the recipient to harm the agent. 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 both predict the outcome of interactions (will agreements be fulfilled?) and predict and avoid risky, or unreliable, agents. Modeling the trustworthiness of potential interaction partners enables the agent to make these predictions. However, an agent must be able to both 1) model trustworthiness of potential interaction partners, and 2) make decisions based on those models. The following subsections detail the implications of these two requirements and discuss some related, currently implemented metrics.

3.1 Modeling Trust

First, models of agent trustworthiness must be accurate predictors of interaction success. Trust models must be able to maintain accuracy even under dynamic conditions, adapting to changes introduced by other agents. For example, an adaptive trust model must adjust when the modeled agent's trustworthiness characteristics change suddenly (perhaps the agent suddenly loses competence or maliciously employs a strategy of varying its trustworthiness) [17]. Trust models must also be able to handle open multi-agent systems, in which agents can enter or leave the system, potentially attempting to change identities. When a new agent is introduced to the system, adaptive trust models should be able to build quickly an accurate picture of the agent's trustworthiness characteristics. 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 [18–20].

Several other characteristics make a trust modeling technique desirable. First, trust models should be efficient, both in terms of computational cost and time [21, 22]. Computational efficiency and accuracy can be gauged by assessing the time to converge to sufficiently accurate models [23]. For example, Barber and Kim [24] compare interaction-based and recommendation-based reputation strategies according to response time, steady-state error, and maximum overshoot (i.e. stability) metrics. Similarly, models should also be scalable, capable of functioning effectively in systems with large numbers of agents whose trustworthiness must be modeled. Trust modeling techniques should employ generic, domain-independent models, applicable to a variety of environments and conditions [25]. Finally, trust models should provide flexibility in the types of entities whose trust is modeled, predicting the trustworthiness not only of other agents, but other types of entities as well [26], such as centralized repositories or simple distributed information databases.

3.2 Acting on Trust

In addition to modeling trust accurately and efficiently, an agent must also be able to make effective decisions using its trust models; quality trust modeling can be measured by the resulting usability of the agent's models. The agent must be able to translate trust models into the best decisions about interacting with other agents. Given a potential interaction agreement, an agent must be able to correctly decide whether to participate in the agreement, predicting whether the agreement will be fulfilled by the other agent. For example, successful trusting can be defined in terms of the number of positive interactions as compared to total interactions [27, 5] or the agent's utility derived from an interaction [28]. If the agreement involves the receipt of a resource, the agent must estimate whether the resource will be delivered, the quality of the expected resources, and whether the resource will be delivered on time. If the agreement requires the agent to deliver a resource to another, the agent must assess what harm or benefit the other agent might cause upon obtaining the resource, as well as the resulting benefit or harm to its own reputation by participating in the agreement. If an agreement requires negotiation of some variable, such as the price or delivery date of a resource, the agent should be able to utilize its trust models to negotiate appropriately. For example, an agent should negotiate a lower price from a resource-providing agent who is predicted to deliver low-quality resources.

Methods for encouraging or enforcing good behavior from potential interaction partners are desirable, as well [28]. Since an agent may use its trust models to determine interactions and negotiate agreements, the agent can identify and isolate untrustworthy agents by refusing to interact with them [24, 29]. The agent may also have the ability to take disciplinary action against agents it deems untrustworthy due to malicious conduct; thus the agent benefits if it is able to distinguish between intentional and inadvertent behavior and act accordingly. In [29], success is measured by the ability to prevent manipulation of probabilistic reciprocity strategy by deceptive agents.

An agent can use its trust models to develop strategies for maximizing its benefit. The agent can explore methods for deceiving other agents to receive an unfair benefit, or restrict communication with harmful agents to avoid enabling malicious behavior. An agent can learn (possibly malicious) strategies, such as deception or collusion, or exploit flaws in trusting methods used by other agents. More altruistically, an agent can act defensively by learning to detect those strategies when used by others. Improving others' perceptions of the agent's trustworthiness is a way to encourage other agents to voluntarily participate in interactions. The agent can attempt to maximize the advantages of trusting other agents (ensuring the benefit of successful interactions) and of being trusted (improving likelihood of future interactions and increasing monetary benefit from being trusted). In addition to basing trust objectives on single-agent achievement, researchers can examine the social impact of trust-based actions to identify effective agent strategies. Researchers can better understand system-level behavior by examining contrasting cases of social versus isolationistic tendencies, or benevolent versus strategically malicious policies.

4 Toward Testbed Specification

Once 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. An effective domain-specific problem of the competition need not directly mirror a specific research objective; effort by researchers toward winning the competition can allow aspects of the domain-independent problem set to be solved along the way. This research does not yet seek to identify a suitable domain problem, but merely justify the need for such a testbed and describe its appropriate characteristics in terms of requirements and parameters.

4.1 Testbed Requirements

Several desirable properties, essential for an effective competition testbed, are enumerated below. This wide collection of experimentation, problem-design, and logistical requirements makes identification of the ideal problem domain difficult.

However, these characteristics must serve as guidelines for the entire testbed development process.

Modularity The testbed should allow simulation parameters to be adjusted easily. This modularity not only permits the testing of a wide range of capabilities, but also increases the competition challenge for subsequent competition editions by allowing rule changes. Multiple competition versions can be used to examine a variety of problem scenarios.

Versatile Experimentation Researchers should be permitted to both participate in competitions and use the testbed for independent experimentation. In competition settings, the testbed should allow researchers to participate as single agents competing against simulation agents and/or agents representing other researchers, attempting to maximize benefit to the single agent. During experimentation, researchers should have the freedom to generate all agents in the system, whether competitive or cooperative. This flexibility allows researchers to additionally control parameters for better observing benefit to the agent system, in addition to individual agent benefit.

Permitting Versatile Approaches A wide range of strategies for modeling trust have been explored in recent years, including direct interaction trust models, reputation mechanisms, group association, and hybrid methods [24, 6]. Additionally, various trust representations have been employed, such as Boolean trust values, rankings, single-value scales, and probabilistic models [30, 24, 31]. The competition testbed should not restrict the wide range of approaches used by researchers for modeling trust and making trust-based decisions.

Uniform Accessibility Testbed development should provide easy, standardized "hook-up" capability for varying numbers of agent participants, regardless of the modeling or decision-making algorithms and representations used by the agent.

Exciting, Relevant Domain The competition should address a currently popular, relevant domain problem which unites researchers under a common challenge. The popularity and applicability of the domain improves the competition testbed's likelihood of being accepted by the research community and of attracting public respect. The chosen domain should showcase trust as a required element of solving the problem, without emphasizing peripheral research areas (such as planning). However, the domain must have broad application which does not too narrowly stifle research on identified priority research problems.

Objective Metrics Metrics defined for the competition testbed should be objective success measures tied directly to the domain problem. Separate metrics should measure success from the single-agent perspective, as well as success for the multi-agent system as a whole, with the flexibility to gauge success achieved by designated cooperating agent subgroups. Metrics assessing the accuracy of an agent's trust models run the risk of restricting the wide range of possible modeling representations since comparisons must be made between some defined "true trustworthiness" and an agent's trust models. Measuring the success of an agent's decisions, based on its models of trust, is a more accommodating approach. However, action-based metrics must be careful to not value too heavily an agent's other capabilities (i.e. planning), instead focusing on decisions which require input from trust models.

4.2 Testbed Parameters

To enable modularity, as described in Section 4.1, the testbed must employ a set of adjustable parameters by which the agent environment changes according to experimenter or competition goals. Not only does parameterization allow the researcher flexibility while in experimentation mode, but this modularity also permits competition organizers to change the "rules of the game", or the environmental dynamics under which the competition is held, requiring players to adapt. Parameterization ensures a comprehensive set of relevant experimental scenarios and allows the testbed to adapt to future experimentation needs. The competition testbed must be designed such that both changing existing parameters and adding additional parameters are straightforward processes.

The UCI machine learning repository [32] is a prominent computer science example of facilitating testbed experimentation modularity. The UCI repository provides a set of databases to evaluate machine learning algorithms. As experimentation needs change, researchers can propose new datasets and methods for dataset analysis. The competition testbed for trust research poses a more complex need, requiring not only experimental data, but also a common framework flexible enough to allow different types of experimentation and metrics. The testbed environment should accommodate dynamics in the following areas.

Network Topology Network topology encompasses all factors affecting the number of agents in the system and the communication links between them. First, the testbed must be able to vary the number of agents in the system. In competition mode, the testbed includes all competing agents, but can adjust the total number of agents by including other agents standardized by the testbed simulation. In experimentation mode, the researcher should have the authority to set the number of agents as desired. The testbed should also be capable of changing the number of agents in the system dynamically by allowing agents to enter or leave the system, either by choice or as forced by the testbed simulation. Competition mode most likely would require competing agents to remain in the system for the duration of the game. However, allowing agents to leave and reenter, thus changing their identities, might encourage some novel strategies among competitors. Nonetheless, allowing agents to enter or leave the testbed simulation would be particularly valuable for researchers in experimentation mode.

Several parameters related to interagent communication can further increase the flexibility of the competition testbed. First, by controlling the creation and destruction of communication links, the testbed temporarily can affect agents ability to communicate. The testbed can employ a parameter called an *encounter factor*, which describes the likelihood of an agent to interact with the same partner repeatedly. For example, a system with few agents and high interaction frequency is associated with a high encounter factor. By specifying the number of interaction opportunities per competition session, the testbed can vary whether agents are able to interact frequently or rarely. In addition, by allowing the testbed to vary which communication protocols are permitted, the types of communications–whether reputation information, trade negotiations, or exchange of resources–can be controlled. Finally, the goals, resources, and utilities for accomplishing goals, as allocated to each agent, should be variable since they determine who agents choose to interact with and the importance of conducting those interactions.

Information Availability The testbed should be able to adjust the types of information available to the agents in the system. We distinguish among three types of information that can be used by agents to build trust and reputation:

– *Direct information* - information an agent gathers from a direct interaction with another agent, related to the agreement made between the agents and the resulting level of agreement fulfillment.
– *Witness information* - information an agent gets from a third party. The information can be related to the direct experiences of the witness but also to observations and information the witness has received from others.
– *Environment information* - information an agent obtains by observing the behavior or interactions of other agents or analyzing publicly available data. It is the information an agent gets directly from the environment without explicitly interacting with other agents.

How the different types of information are used depends on the environment of the agents. In environments where the cost of direct interactions is high, witness and environment information become very relevant. Conversely, if the cost of direct experiences is low, direct information is the best option for building trust models.

Quality observability is another parameter that the testbed should control; it relates to an agents ability to assess the quality of an interaction. For example, in an exchange of resources for currency, the buying agent is not able to observe the quality of the interaction until it receives the promised goods. In another case, in which an agent purchases information, the buyer can never observe quality (information accuracy) unless it has sufficient additional information sources against which to compare the purchased information. Quality observability is affected by an interactions feedback time, the time that passes between the completion of an interaction and the time at which the quality of that interaction is observable.

The testbed can have some control over the cost of trust-related information. For example, obtaining trust information by conducting a direct interaction exposes an agent to risk; the cost of the direct information is high if the agents interaction partner cheats. Similarly, if agents only sell witness information for outrageously high fees, an agent may exclude witness information from its trust models.

Agreement Fulfillment Degree of agreement fulfillment refers to the frequency and quality to which agreements between agents are fulfilled. Agreements to exchange accurate reputation information are included in this definition. In competition mode, the testbed cannot control the strategies of researchers competing agents. However, the testbed simulation can influence the degree of agreement fulfillment through additional agents inserted by the simulation itself.

There are three reasons that may cause an agreement to either not be fulfilled or be fulfilled to an unsatisfactory degree [1, 33]: 1) an agents malicious intent to disregard the agreement, 2) an agents honest inability to fulfill the agreement, or 3)an agent's honest attempt to "overhelp", or protect the requesting agent. In the context of agreements to exchange reputation information, malicious agents may deliberately communicate false information, whereas incompetent agents may truthfully convey information that is of poor quality. In cooperative systems, in which agents are altruistic, agents may be unable to fulfill agreements even when they do not adopt deceptive strategies. In competitive environments, both inability and malice may be factors in leaving agreements unfulfilled, but agents are unlikely to extend themselves by overhelping. In some cases, an incompetent agent may not be aware of its limitations. It is extremely difficult for an agent whose partner leaves an agreement unfulfilled to know the cause behind the partners shortcoming; often the reason is irrelevant to the resulting impact on the agents trust model. However, the testbed should consider incorporating both malicious and incompetent simulation agents since advanced modeling techniques may take advantage of that distinction.

The testbed can parameterize the amount of agreement fulfillment (as influenced by testbed-controlled agents) in terms of number of agreements left unfulfilled relative to total agreements and whether agreements were unfulfilled intentionally or due to incompetence. Additionally, the testbed can vary the degree to which agreements are partially fulfilled. Finally, the testbed can alter the distribution of agreement fulfillment; in one case, a few malicious agents may cheat frequently, while in another case, several agents may cheat occasionally.

Analysis Perspective Trust-modeling techniques can be analyzed from both the agent perspective and the system perspective. The agent perspective examines the utility of a strategy to a single agent without regard for the benefit to the overall agent system. The agents utility is measured as the benefit achieved from using the agents trust models to improve the agents decision-making mechanism. The possibility that the strategy can decrease system utility is not considered. Agent-based metrics are important in competition mode, in which each researchers agent holds its personal goals as highest priority.

The system perspective employs metrics that emphasize social welfare, or benefit to the agent system as a whole. In this case, the improvement in performance of a single individual is not as important as the sum of benefits among all agents. In experimentation mode, system-based metrics are valuable for observing the robustness of an agent society to cheaters or tendencies among agents to form coalitions, for example. A well-designed trust model should excel in both types of analysis, either improving the performance of the individual agent or the agent society, depending on the goals of the designer.

5 The Trust Competition Testbed Initiative

This paper has shown the shortcomings of existing experimentation domains, while demonstrating the need for a more comprehensive trust competition testbed. The foundation has been laid for testbed development by 1) enumerating the important problems in trust and reputation research, 2) describing important requirements for a competition testbed, and 3) addressing necessary parameters for testbed modularity and flexibility.

The Workshop on Trust in Agent Societies at The Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2004) showcased a panel discussion addressing the feasibility of a competition testbed for agent trust technologies. As a result, the Trust Competition Testbed Initiative was launched with the purpose of establishing a testbed to achieve these goals. A competition testbed surpasses the benefits of previous experimentation settings by providing researchers with easy access to the same experimentation setup. In addition, researchers are allowed to compete against each other to determine the most viable technology solutions. A competition testbed unites researchers through a domain problem applicable to many researchers objectives, while introducing rich, new problem spaces.

An international research team has been formed to coordinate domain specification, game design, testbed development, and competition administration. The teams first task is to identify a domain which fits the desirable characteristics described previously: modularity, versatility, dynamics, and relevance. As the domain is selected, the team begins testbed specification, structuring the behavior of the testbed simulation and detailing rules for agent participants. Once the domain specification is completed, development of the testbed proceeds, producing code for the simulation, a basic participating agent, and graphical user interfaces for monitoring experiments and competitions. Upon completion of the prototype competition testbed, experimental review, and revisions based on feedback from the research community, arrangements will be made to conduct the first competition using the experimental testbed. Plans are underway to complete the competition specification and prototype infrastructure. Next, researchers can prepare for participation by developing their agents for the first competition. Development progress can be monitored through the ART Testbed discussion board [4], where updates to competition development progress are posted periodically.

6 Acknowledgment

Jordi Sabater enjoys a Sixth Framework Programme Marie Curie Intra-European fellowship, contract No MEIF-CT-2003-500573. This research is sponsored in part by the Defense Advanced Research Project Agency (DARPA) Taskable Agent Software Kit (TASK) program, F30602-00-2-0588. In addition, the research reported in this document/presentation was supported in part by the Countermeasures to Biological and Chemical Threats Program at the Institute for Advanced Technology of The University of Texas at Austin, under contract number DAAD17-01-D-0001 from the U.S. Army Research Laboratory. The views and conclusions contained in this document/presentation are those of the authors and should not be interpreted as presenting the official policies of the U.S. Government unless so designated by other authorized documents. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

1. Barber, K.S., Fullam, K., Kim, J.: Challenges for trust, fraud, and deception research in multi-agent systems. Trust, Reputation, and Security: Theories and Practice (2003) 8—14
2. Fullam, K., Barber, K.S.: Evaluating approaches for trust and reputation research: Exploring a competition testbed. In: Proceedings of The Workshop on Reputation in Agent Societies at Intelligent Agent Technology (IAT2004), Beijing. (2004) 20—23
3. Sabater, J.: Toward a test-bed for trust and reputation models. In: Proceedings of The Workshop on Deception, Fraud and Trust in Agent Societies at The Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2004), New York, USA. (2004) 101—105
4. ART Testbed Team: Agent Reputation and Trust Testbed, http://www.lips.utexas.edu/~kfullam/competition/. (2005)
5. Schillo, M., Funk, P., Rovatsos, M.: Using trust for detecting deceitful agents in artificial societites. Applied Artificial Intelligence (2000) 825—848
6. Mui, L., Mohtashemi, M., Halberstadt, A.: Notions of reputation in multi-agent systems: A review. In: Proceedings of the first international joint conference on autonomous agents and multiagent systems (AAMAS-02), Bologna, Italy. (2002) 280—287
7. Marsh, S.: Formalising Trust as a Computational Concept. PhD thesis, Department of Mathematics and Computer Science, University of Stirling (1994)
8. Verhoeff, T.: The trader's dilemma: A continuous version of the prisoner's dilemma. Technical report, Computing Science Notes 93/02, Faculty of Mathematics and Computing Science, Technische Universiteit Eindhoven, The Netherlands (1998)
9. Zacharia, G.: Collaborative reputation mechanisms for online communities. Master's thesis, Massachusetts Institute of Technology (1999)

10. Sabater, J., Sierra, C.: Regret: A reputation model for gregarious societies. In: Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies, Montreal, Canada. (2001) 61—69
11. Carbo, J., Molina, J., Davila, J.: Trust management through fuzzy reputation. Int. Journal in Cooperative Information Systems **12** (2002) 135—155
12. Carbo, J., Molina, J., Davila, J.: Comparing predictions of sporas vs. a fuzzy reputation agent system. In: 3rd International Conference on Fuzzy Sets and Fuzzy Systems, Interlaken. (2002) 147—153
13. Sabater, J., Sierra, C.: Reputation and social network analysis in multi-agent systems. In: Proceedings of the first international joint conference on autonomous agents and multiagent systems (AAMAS-02), Bologna, Italy. (2002) 475—482
14. Yu, B., Singh, M.P.: Towards a probabilistic model of distributed reputation management. In: Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies, Montreal, Canada. (2001) 125—137
15. eBay: eBay, http://www.eBay.com. (2002)
16. BizRate: BizRate, http://www.bizrate.com. (2002)
17. Fullam, K., Barber, K.S.: A temporal policy for trusting information. In: Proceedings of The Workshop on Deception, Fraud and Trust in Agent Societies at The Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2004), New York, USA. (2004) 47—57
18. Fullam, K.: An expressive belief revision framework based on information valuation. Master's thesis, Department of Electrical and Computer Engineering, The University of Texas at Austin (2003)
19. Klos, T., la Poutre, H.: Using reputation-based trust for assessing agent reliability. In: Proceedings of The Workshop on Deception, Fraud and Trust in Agent Societies at The Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2004), New York, USA. (2004) 75—82
20. Whitby, A., Josang, A., Indulska, J.: Filtering out unfair ratings in bayesian reputation systems. In: Proceedings of The Workshop on Deception, Fraud and Trust in Agent Societies at The Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2004), New York, USA. (2004) 106—117
21. Ghanea-Hercock, R.: The cost of trust. In: Proceedings of The Workshop on Deception, Fraud and Trust in Agent Societies at The Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2004), New York, USA. (2004) 58—64
22. Yamamoto, H., Ishida, K., Ohta, T.: Trust formation in a c2c market: Effect of reputation management system. In: Proceedings of The Workshop on Deception, Fraud and Trust in Agent Societies at The Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2004), New York, USA. (2004) 126—136
23. Ding, L., Kolari, P., Ganjugunte, S., Finin, T., Joshi, A.: On modeling and evaluating trust networks inference. In: Proceedings of The Workshop on Deception, Fraud and Trust in Agent Societies at The Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2004), New York, USA. (2004) 21—32
24. Barber, K.S., Kim, J.: Belief revision process based on trust: Agent evaluating reputation of information sources. Trust in Cyber-societies: Integrating the Human and Artificial Perspectives, Lecture Notes in Computer Science (2002) 73—82

25. Huynh, D., Jennings, N., Shadbolt, N.: Developing an integrated trust and reputation model for open multi-agent systems. In: Proceedings of The Workshop on Deception, Fraud and Trust in Agent Societies at The Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2004), New York, USA. (2004) 65—74
26. Fujimura, K., Tanimoto, N., Iguchi, M.: Calculating contribution in cyberspace community using reputation system 'rumor'. In: Proceedings of The Workshop on Deception, Fraud and Trust in Agent Societies at The Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2004), New York, USA. (2004) 40—46
27. Falcone, R., Pezzulo, G., Castelfranchi, C., Calvi, G.: Trusting the agents and the environment leads to successful delegation: A contract net simulation. In: Proceedings of The Workshop on Deception, Fraud and Trust in Agent Societies at The Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2004), New York, USA. (2004) 33—39
28. Neville, B., Pitt, J.: A simulation study of social agents in agent mediated e-commerce. In: Proceedings of The Workshop on Deception, Fraud and Trust in Agent Societies at The Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2004), New York, USA. (2004) 83—91
29. Biswas, A.S., Sen, S., Debnath, S.: Limiting deception in groups of social agents. In: Proceedings of The Workshop on Deception, Fraud and Trust in Agent Societies at Autonomous Agents, Seattle, Washington. (1999) 21—28
30. Barber, K.S., Fullam, K.: Applying reputation models to continuous belief revision. In: Proceedings of The Workshop on Deception, Fraud and Trust in Agent Societies at The Second International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2003), Melbourne, Australia. (2003) 6—15
31. Sen, S., Sajja, N.: Robustness of reputation-based trust: Boolean case. In: Proceedings of the first international joint conference on autonomous agents and multiagent systems (AAMAS-02), Bologna, Italy. (2002) 288—293
32. UCI: UCI, http://www.ics.uci.edu/˜mlearn/MLRepository.html. (2004)
33. Castelfranchi, C., Falcone, R.: Towards a theory of delegation for agent-based systems. Robotics and Autonomous Systems **24** (1998) 141—157