A Demonstration of the Agent Reputation and Trust (ART) Testbed for Experimentation and Competition

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

[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]École Nationale Supérieure des Mines, Saint-Étienne, France
[4]Institute of Cognitive Science and Technology (ISTC), National Research Council (CNR), Rome, Italy
[5]Multiagent Systems Research Group—Critical MAS, Hebrew University, Jerusalem, Israel

1. INTRODUCTION

The Agent Reputation and Trust (ART) Testbed [1] serves two roles: (1) as a competition forum for comparing technologies against objective metrics, and (2) as an environment for performing customizable, easily-repeatable experiments. In the testbed's art appraisal domain, agents valuate paintings for clients and gather opinions from other agents to produce accurate appraisals. In recent years, researchers [2–4] have recognized objective standards are necessary to justify successful trust modeling systems and provide a baseline for future work. For trust technologies to crossover into application [5, 6], the public must obtain system evaluations based on transparent, recognizable standards. As a versatile, universal experimentation site, the ART Testbed scopes relevant trust research problems and unites researchers toward solutions via unified experimentation methods. Through objective, well-defined metrics, the testbed provides researchers with tools for comparing and validating their approaches. The testbed also serves as an objective means of presenting technology features—both advantages and disadvantages—to the community. In addition, the ART Testbed places trust research in the public spotlight, improving confidence in the technology and highlighting relevant applications.

The ART Testbed unifies trust research, whereas other technologies do not: the Prisoner's Dilemma competition [7] lacks objective, system-based metrics and trust problem focus (agents can employ game theoretic strategies with minimal trust-modeling skills). The SPORAS experiments [8] are narrow in scope, evaluating reputation models based on only single-agent metrics and failing to compare multiple trust-modeling strategies in a competition setting. Schlosser et al. [9] propose a framework for evaluating reputation systems, but it does not easily extend to compare multiple trust-modeling algorithms in competition against each other.

2. TESTBED DOMAIN PROBLEM

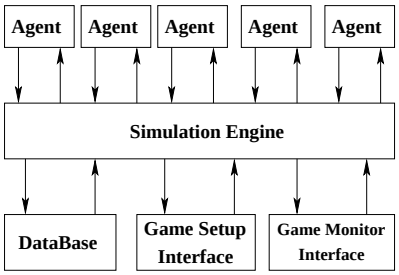
The ART testbed provides functionality for researchers of trust and reputation in multi-agent systems. As such, it operates in two modes: *competition* and *experimentation*. In competition mode, each participating researcher controls a single agent, which works in competition against every other agent in the system. Competition organizers can change parameters to permit the game structure to be adapted for subsequent competitions. To utilize the testbed's experimentation mode, the 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, the researcher has the flexibility of complete control over all experiment parameters.

The testbed operates in an art appraisal domain (see [10] for a detailed justification), where researchers' agents function as painting appraisers with varying levels of expertise in different artistic eras (e.g. classical, impressionist). Clients request appraisals for paintings from different eras; if an appraiser does not have the expertise to complete the appraisal, it may purchase opinions from other appraisers. Other appraisers estimate the accuracy of opinions they send by the cost they choose to invest in generating an opinion, and opinion providers may lie about the estimated accuracy of their opinions. Appraisers produce appraisals using their own opinion and opinions received from other appraisers. Appraisers receive more clients, and thus more profit, for producing more accurate appraisals. Appraisers may also purchase reputation information from each other about third-party agents. Appraisers attempt to accurately valuate their assigned paintings; their decisions about which opinion providers to trust directly impact the accuracy of their final appraisals. In competition mode, the winning agent is selected as the appraiser with the highest bank account balance.

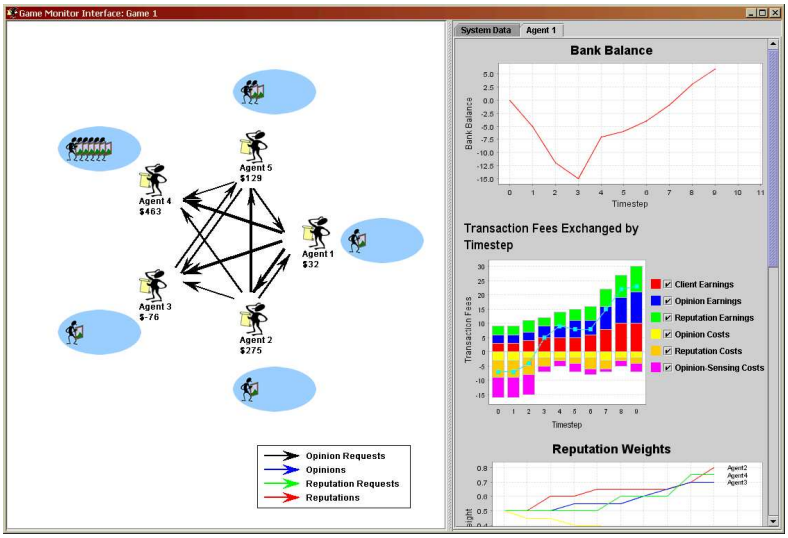
3. TESTBED ARCHITECTURE

As shown in Figure 1(a), the testbed architecture, implemented in Java, consists of four components: (1) Simulation Engine, (2) Database, (3) User Interfaces, and (4) Agent Skeleton (see [11] 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. In each timestep, the Simulation Engine assigns clients with paintings to each appraiser. Then appraisers conduct reputation transactions with each other about third-party appraisers, whereby reputation providers may lie about the reputations they sell. Appraisers then conduct opinion transactions with each other. Appraisers may lie about the opinions they sell. Finally, the Simulation Engine assesses each appraiser's accuracy based on the opinions the appraiser purchases and the 'weights' the appraiser places on those

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



(a) Testbed architecture.



(b) Game Monitor Interface.

Figure 1: Testbed architecture and Game Monitor Interface.

opinions. Weights are real values between zero and one that an appraiser assigns, based on its trust model, to another's opinion.

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. User Interfaces permit researchers to observe games in progress and access information collected in the Database by graphically displaying details. Figure 1(b) shows the Game Monitor Interface, by which observers can view opinion and reputation transactions between agents on the left and detailed statistics, such as bank balance, about each appraiser agent on the right. Finally, the Agent Skeleton is designed to allow researchers to easily implant customized internal trust representations and trust revision algorithms while permitting standardized communication protocols with entities external to the appraiser agent.

4. CONCLUSIONS

The ART Testbed 1) provides researchers with a common experimentation environment and 2) allows researchers to compete against each other to determine the most viable technology solutions. The necessity of selecting a single domain for competition and experimentation purposes is a limitation of the testbed, but a unified domain problem is necessary for comparing technologies. The art appraisal domain design addresses prominent trust research problems related to an agent's ability to model trust and make decisions based on trust. A second limitation is posed by the lack of experimentation previously conducted to validate domain rules. However, this experimentation cannot be conducted by the research community until the testbed software release takes place as planned in November of 2005. Based on prototype testbed experimental review and feedback from the research community, the first testbed competition will be conducted in July of 2006. The authors expect to encounter minor, unpredictable difficulties in coordinating the initial competition, but adjustments to domain rules and competition organization will be made for subsequent competitions. Development progress can be monitored through the testbed website [1], where updates to competition development are posted.

5. REFERENCES

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

[2] K. S. 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. Fullam and K. S. Barber. Evaluating Approaches for Trust and Reputation Research: Exploring a Competition Testbed. In Paolucci et al. [12], pages 20–23.

[4] J. Sabater. Toward a Test-Bed for Trust and Reputation Models. In *Proc. Trust Workshop at AAMAS*, 2005, pages 101–105.

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

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

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

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

[9] A. Schlosser, M. Voss, and L. Brückner. Comparing and Evaluating Metrics for Reputation Systems by Simulation. In Paolucci et al. [12].

[10] K. Fullam, T. Klos, G. Muller, J. Sabater, A. Schlosser, Z. Topol, K. S. Barber, J. S. Rosenschein, L. Vercouter, and M. Voss A Specification of the Agent Reputation and Trust (ART) Testbed In *Proc. AAMAS*, 2005.

[11] 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.

[12] M. Paolucci, J. Sabater, R. Conte, and C. Sierra, editors. *Proc. IAT2004 Workshop on Reputation in Agent Societies*, 2004.