

Truthful Reputation Mechanisms for Online Systems

IFAAMAS-07 Victor Lesser Distinguished Dissertation Award

Radu Jurca

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Artificial Intelligence Laboratory (LIA)
CH-1015 Lausanne, Switzerland
radu.jurca@gmail.com

The internet is moving rapidly towards an interactive milieu where online communities and economies gain importance over their traditional counterparts. While this shift creates opportunities and benefits that have already improved our day-to-day life, it also brings a whole new set of problems. For example, the lack of physical interaction that characterizes most electronic transactions, leaves the systems much more susceptible to fraud and deception.

Reputation mechanisms offer a novel and effective way of ensuring the necessary level of trust which is essential to the functioning of any market. They collect information about the history (i.e., past transactions) of market participants and make public their *reputation*. Prospective partners guide their decisions by considering reputation information, and thus make more informative choices. Online reputation mechanisms enjoy huge success. They are present in most e-commerce sites available today, and are seriously taken into consideration by human users.

The economical value of online reputation raises questions regarding the trustworthiness of mechanisms themselves. Existing systems were conceived with the assumption that users will share feedback honestly. However, we have recently seen increasing evidence that some users strategically manipulate their reports.

This thesis describes ways of making online reputation mechanisms more trustworthy by providing incentives to rational agents for reporting honest feedback. Chapter 2 starts with a brief review of the state of the art related to online reputation mechanisms. Trust will be understood as an agent's subjective decision to rely on another agent in a risky situation. The *reputation* is one piece of information considered when taking the decision. The role of reputation information is two-fold: first, to provide information about the hidden characteristics of the trustee that are relevant for the given situation (i.e., a signaling role) and second, to make future agents aware about any cheating that occurred in the past (i.e., a sanctioning role).

Chapter 3 addresses signaling reputation mechanisms and is mainly focused on the reporting incentives of the participants. Carefully designed payment schemes can explicitly reward honest feedback by a sufficient amount to offset both the cost of reporting and the gains that could be obtained

Cite as: Truthful Reputation Mechanisms for Online Systems (Award Paper), Radu Jurca, *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril, Portugal, page 5.

Copyright © 2008, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

through lying. The main contribution of this chapter is to show that the principle of *automated mechanism design* [1] can significantly decrease the operation cost of reputation mechanisms, and can lead to collusion proof mechanisms.

Chapter 4 develops a practical application of truthful signaling reputation mechanisms for the domain of quality of service monitoring. As compared to traditional methods, QoS monitoring based on client reports and incentive compatible rewards can be significantly cheaper and more robust.

Chapter 5 addresses sanctioning reputation mechanism. The first contribution of this chapter is to extend existing results due to Dellarocas [2] for general settings, where the seller can choose between several effort levels, and buyers can observe several quality levels. The second part of the chapter discusses a mechanism for encouraging the submission of honest feedback.

The last chapter of the thesis takes a step towards better understanding existing online reputation mechanisms. The main objective is to investigate the factors that (i) drive a user to submit feedback, and (ii) bias a user in the rating she provides to the reputation mechanism. Based on an empirical study conducted on TripAdvisor hotel reviews, several conclusions were drawn. First, the groups of users who amply discuss a certain feature seem more likely to agree on a common rating for that feature. Second, we find a correlation between the effort spent in writing a review, and the risk posed by the trusting decision (e.g., choosing a bad hotel). Third, the rating expressed by a reviewer appears to be biased by the reviews submitted by previous users. The information already available in the forum creates a prior expectation of quality, and changes the user's subjective perception of quality. Forth, it appears that that human users are more likely to voice their opinion when they can bring something different to the discussion, and can contribute with new information.

1. REFERENCES

- [1] V. Conitzer and T. Sandholm. Complexity of mechanism design. In *Proceedings of the Uncertainty in Artificial Intelligence Conference (UAI)*, 2002.
- [2] C. Dellarocas. Reputation Mechanism Design in Online Trading Environments with Pure Moral Hazard. *Information Systems Research*, 16(2):209–230, 2005.