Trust Mechanisms for Online Systems (Extended Abstract)

Jens Witkowski
Institute of Computer Science
Albert-Ludwigs-Universität
Freiburg, Germany
witkowsk@informatik.uni-freiburg.de

1 Introduction

Almost every e-commerce site employs a so-called reputation mechanism that collects and publishes ratings from its users which then allow other market participants to make better-informed choices. It is instructive to distinguish between two kinds of reputation mechanisms in accordance with the problems they address: those that are employed by online opinion forums, such as Amazon Reviews, are built to eliminate asymmetric information while those at online auction sites are primarily intended to induce cooperation and trust between the market participants [e. g., Dellarocas, 2006]. Consider the online auction site eBay as an example for this latter kind: its procedure is such that the winning bidder (henceforth the *buyer*) first pays for the good and that the seller is required to send it only after receipt of payment. Without any trust-enabling mechanisms in place, the seller is best off keeping the good for himself[1], even if he received the buyer's payment. Since a rational, self-interested buyer can anticipate this, she will not pay for the good in the first place and no trade takes place. While sanctioning reputation mechanisms, such as the one employed by eBay, address this problem, some of their game-theoretic assumptions are too strong for real-world marketplaces. Three of these assumptions stand out in particular: the assumption of truthful buyer feedback, the assumption of long-lived sellers and the assumption of sellers being incapable of whitewashing. In my thesis, I develop incentive-compatible trust mechanisms that do not require any of these assumptions. Furthermore, I focus on designs which avoid the strong common knowledge assumptions that prevented the application of previous proposals.

2 Truthful Feedback

A common feature of trust mechanisms is the dependency on honest buyer feedback. Most mechanisms in the literature simply assume that every buyer reports truthfully which is problematic from a game-theoretic point of view for two reasons: first, the rating process is time-consuming and there is usually no reward for reporting feedback, so that the customers have no incentive to participate at all. Second, the involved players often have external interests, i. e. biases towards dishonest reporting. On eBay, for example, some sellers have an incentive to buy from their competitors and badmouth them since the majority of sellers has very high numbers of positive feedback, so that even a few negative reports by a competitor can have a large impact on market share. The truthful elicitation of the buyers' private experiences is thus crucial to incorporate into the design of trust mechanisms. A solution for online opinion forums is the so-called *peer prediction method* that was introduced by Miller, Resnick and Zeckhauser [2005] who propose to pay a buyer for her feedback report contingent on the report of another buyer. This comparison is meaningful if a product's underlying quality is essentially identical for all customers. Consider a digital camera bought from Amazon as an example: while different customers may have different experiences due to noise in the production process, every customer receives the identical model. This is different for online auction sites like eBay since the customers' experiences primarily depend on the sellers' actions, i. e. if the respective seller sent the good in the prescribed quality, which the seller potentially variates from one buyer to the other. We study the mechanism design space of peer-prediction-based feedback elicitation for reputation mechanisms that are situated in such online auction settings and show that it is impossible to design a truthful scheme for the basic setting with only a single-type seller. However, drawing on the literature on reputation building in game theory, we prove that it is possible to design a mechanism that elicits honest buyer feedback for settings with a small prior belief that the seller is of a cooperative commitment type [Witkowski, 2010].

Unfortunately, the peer prediction method relies on strong common knowledge assumptions that have so far prevented its application to real-world marketplaces. Consider again the digital camera bought from Amazon as an example: the basic idea of the peer prediction method is that every buyer has the same prior belief about the camera's true quality and that, once a buyer receives the camera, she experiences a noisy signal of this true quality. For example, the buyers might belief that the camera will be of high quality with a probability of 75% which would change to 90% after a positive experience. Following such a positive experience, it is then more likely that another buyer of the same camera model also had a good experience. It is assumed that all buyers and the mechanism share the same beliefs, e.g. 75% probability for high quality before experiencing the camera, 90% after a good experience and 30% after a bad experience.

[1] I refer to buyers and sellers as female and male, respectively.

In ongoing work, we design mechanisms that do not rely on these commonly-held beliefs. To understand the basic principle of our proposal, observe that if the mechanism has neither a prior belief of the quality nor of the signal probabilities, a buyer's belief that the camera if of good quality still rises following a positive experience. Moreover, Amazon knows that within some hours of ordering the camera, the buyer has not yet received it, so that the mechanism can ask for two probabilistic belief reports: one before the buyer receives the good and another one thereafter. The crucial part is that by truthfully eliciting these two beliefs, the mechanism can infer the binary signal that the buyer must have received. That is, if the second belief report is higher, the experience must have been good and vice versa. Note that the inferred signal is required to condition the other buyers' payments. In future research, we will also study how to aggregate the elicited beliefs into a joint belief using *proper scoring rules* for both the computation of the payments that ensure truthfulness and the weighting of reports that corresponds to the expertise of each individual buyer.

3 Incentive-Compatible Escrow Mechanisms

In a recent paper, we introduce a new class of trust mechanism [Witkowski *et al.*, 2011]. These *escrow mechanisms* are "history-free" in that they do not rely on the publication of reported feedback which improves on the state-of-the-art in that it avoids two assumptions of reputation mechanisms: first, we no longer need the assumption that sellers are long-lived, i.e. that every seller is in the market long enough to be sufficiently incentivized by future returns offsetting the immediate incentive to cheat, and, second, they remove the whitewashing problem, i.e. the seller's ability to create a new account with a fresh reputation profile once an old one is ran down. The main idea of an escrow mechanism is that a buyer does not pay the seller directly but through a trusted third party (henceforth the *center*). Once the seller has sent the good, the center asks the buyer for her feedback and forwards the payment to the seller only if the buyer acknowledges that the good arrived in the promised condition. The key question is how to proceed with the withheld payments following a negative report. Obviously, if the center reimbursed every buyer who reports negatively, a rational buyer would always do so. While we could use the reports solely to determine the seller's payment and leave the surplus that this generates with the center, this mechanism would not be efficient since buyers would take into account the probability that the good is lost and, consequently, would lower their bids. The idea of our mechanism is that the report of one buyer—instead of determining whether she herself receives a payback—determines whether another buyer receives a payback. In addition to being fully efficient, this mechanism is incentive compatible, interim individually rational and ex ante budget balanced. Moreover, and in contrast to previous work on trust and reputation [e. g., Dellarocas, 2003], our approach does not rely on knowing the sellers' cost functions or the distribution of buyer valuations. We also show how to make escrow mechanisms robust against colluding buyers by introducing *cross-seller matching*, where every buyer is matched with a buyer from a different seller to determine her payback. As a consequence, in large markets like eBay, the chances for two colluders to be matched with one another are very small and, thus, the expected utility for collusion is negative even under the assumption of minimal coordination costs.

The general escrow mechanism technique is applicable to a wide class of settings. A market that is particularly in need for a trust mechanism is Amazon Turk. One of the particularities of this market is that the verification of a task, i.e. whether the task was duly completed, is costly and it is therefore not uncommon that the verification of a task is itself made a task. In fact, for every original task, a requester usually creates two verification tasks and other workers are asked to vote if the original task was properly executed. Unfortunately, this scheme does not properly incentivize effort and fraudulent behavior is a major problem. In ongoing work, we develop an escrow mechanism for this market that provides proper incentives and increases efficiency by reducing the number of necessary verification tasks. Once we have designed the mechanism and proven its theoretical properties, we will run an experiment with different designs on Amazon Turk itself. We believe that it will be particularly interesting to study the optimal trade-off between the cognitive costs incurred by the respective level of complexity and the mechanism's theoretical properties.

Acknowledgements

I want to thank my coauthors for inspiring collaboration. My special thanks go to David Parkes for hosting me at Harvard University, where most of this research agenda developed. I am also grateful for financial support through PhD fellowships from the Landesgraduiertenförderung Baden-Württemberg and the German Academic Exchange Service.

References

[Dellarocas, 2003] Chrysanthos Dellarocas. Efficiency through Feedback-contingent Fees and Rewards in Auction Marketplaces with Adverse Selection and Moral Hazard. In *Proceedings of the 4th ACM Conference on Electronic Commerce (EC'03)*, 2003.

[Dellarocas, 2006] Chrysanthos Dellarocas. Reputation Mechanisms. In Terry Hendershott, editor, *Handbook on Information Systems and Economics*. Elsevier Publishing, 2006.

[Jurca, 2007] Radu Jurca. *Truthful Reputation Mechanisms for Online Systems*. PhD thesis, School of Computer and Communication Sciences, EPFL, 2007.

[Miller *et al.*, 2005] Nolan Miller, Paul Resnick, and Richard Zeckhauser. Eliciting Informative Feedback: The Peer-Prediction Method. *Management Science*, 51(9):1359–1373, 2005.

[Witkowski *et al.*, 2011] Jens Witkowski, Sven Seuken, and David Parkes. Incentive-Compatible Escrow Mechanisms. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence (AAAI'11)*, 2011.

[Witkowski, 2009] Jens Witkowski. Eliciting Honest Reputation Feedback in a Markov Setting. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI'09)*, 2009.

[Witkowski, 2010] Jens Witkowski. Truthful Feedback for Sanctioning Reputation Mechanisms. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI'10)*, 2010.