Challenge Problem: Agent-Mediated Decentralized Information Mechanisms

David C. Parkes

Division of Engineering and Applied Sciences, Harvard University,
33 Oxford Street, MA 02138, USA
parkes@eecs.harvard.edu
http://www.eecs.harvard.edu/~parkes

Abstract. Pervasive computing, driven by faster, cheaper and smaller devices, and wireless networking technology, promises to make people perpetual users of a massive and decentralized computational system. Pervasive computing blurs the boundary between the physical and the virtual. Information, always cheap to duplicate, now becomes cheap to generate everywhere. At once both the opportunities, such as for data mining and collaborative filtering, and the dangers, such as for privacy abuse, are clear. As a challenge to the agent community this paper proposes the development of agent-mediated, decentralized information-processing mechanisms, that enshrine the principles of information property rights and provides economic incentives to support efficient information sharing. We discuss the complementary roles of markets, information-degradation and aggregation, and reputation, within such a mechanism, and propose a straw-man model.

1 Introduction

We are at the frontier of a world of pervasive computing, with fast, cheap, and small computational devices everywhere. Devices are embedded in buildings and physical infrastructure [17, 13], mobile devices continue to get smaller and more powerful, and ad hoc wireless networks continue to emerge. As pervasive computing blurs the boundary between the physical and the virtual, we become perpetual users of a massive and decentralized computational system.

It will soon be possible to collect information, for example about the location and activities of individuals everywhere, and infer contextual information about actions and current goals. As an example, maybe I walk into a store and the store can dynamically negotiate a price on a particular good or service based on information that I have recently spent a long time comparison shopping online.[1]

[1] We already hear of cases where cell phone locations have been used to solve crime, and there is a hot debate about the privacy implications of the practice of tracking cell phone locations. The Electronic Privacy Information Privacy Center (EPIC) maintains an archive of news articles debating privacy and wiretaps on digital communication technology. http://www.epic.org/privacy/wiretap/

Electronic commerce web sites already collect information about our on-line purchases and browsing habits, to use for personalization and service differentiation, and also to aggregate information with other sites and perhaps sell to third parties. Consider now the opportunities for the personalization of activities and services in the physical world that can be enabled by pervasive information gathering and pervasive information processing. Of course, at the same time the dangers, for example of privacy abuse, are clear [3].

As a challenge to the agent community, we propose the development of a well-functioning agent-based infrastructure for *decentralized information mechanisms*. We define a well-functioning information mechanism as a system in which autonomous computational agents, representing self-interested users, make decisions about information revelation and information sharing that are beneficial from the perspective of the system as a whole. For example, if my agent knows a good place for the user of another agent to go and buy the book that she is looking for, then we would like that information to be exchanged between agents (perhaps for a price).

It is proposed that well-functioning information mechanisms will require: (1) *information property rights* and *markets* to reward users for the fair value of information; (2) *control* of information release, via pseudonymity, data degradation, and information aggregation; and (3) reputation mechanisms to address asymmetries that exist during negotiation for information. Computationally, the challenge is to build decentralized information processing systems that can handle the ability for massive and distributed data acquisition and processing. We will describe a straw-man model for the components of a decentralized agent-mediated solution.

2 Some Design Principles of Information Mechanisms

The intention of this section is to lay out the three components, or principles, for the design of decentralized information mechanisms.

2.1 Information Property Rights and Markets

The first principle is to provide users with *property rights* over information, to allow users to control when and how information is shared and used, and to provide markets to allow them to extract surplus from the value of their information. With property rights, there is in fact no fundamental conflict between the right to privacy and the ability to leverage the value of shared information. Rather, an economic view of privacy holds that users should receive a fair price for the use of personal data (this price can be set arbitrarily high by a user that requires absolute privacy) [18, 11, 3].

Earlier work has suggested *markets* for information-sharing, for example "markers for evaluations" [4] and "markets for secrets" [3]. From a mechanism design perspective [9], we can also imagine the formulation of a simple *information sharing game*. Consider a system of agents, each with private information, and another agent with a query to execute. Mechanism design supposes

that agents will follow game-theoretically rational strategies, and choose to reveal information and formulate queries to maximize their own expected payoff in equilibrium, given beliefs about the strategies of other agents. It will be an interesting exercise in mechanism design to formulate an incentive-compatible mechanism for information sharing, in which truth-revelation is an equilibrium strategy. One direction is to provide an expressive bidding language, in which agents can express valuations for the accuracy of query responses, and costs for the accuracy of information revelation.

A well-functioning mechanism would select a social-welfare maximizing level of information revelation, for example selecting the accuracy of the response to a query to maximize the difference between the value of the query response and the cost in terms of information revelation.

2.2 Control of Information Revelation

As a second principle, a well-functioning information mechanism should provide a number of tools to give users control over the sharing and release of information.

We assume that once information is revealed to any agent (even implicitly via the response to queries), then the ability to extract additional surplus from that information is lost, because the marginal cost of duplication of information to the receiving agent is assumed to be zero. Information is very easy to duplicate and disseminate, via powerful content distribution networks can be constructed on top of the existing Internet infrastructure [2]. The challenge, then, is to provide responses to queries without revealing too much information. For example, it is better to respond "yes, you'll like that book", than "yes, you'll like that book because of all the information that I have, and here it is."

One control tool is the ability for users to adopt pseudonyms, and multiple identities. This prevents the aggregation of information across multiple queries ("tell me the first letter of your street address, then the next letter, then the next letter, etc.") Another control tool is *aggregation*, in which queries are evaluated on the basis of summary information from multiple (probably similar) users. Data mining methods, such as collaborative filtering and reputation mechanisms, allow useful knowledge about user populations to be mined from sparse data about the preferences and actions of a population of users [16, 14].

Even aggregated and anonymized query responses can leak information if an adversary knows some information about the users and can reverse-engineer the identity of the user associated with a response [3]. As such, another control tool is provided by the ability to perform statistical query-filtering, for example to check the sensitivity of the response to a query to the information content in any one user's data. Automatic *information degradation*, via coarse-level data clustering (e.g. the "binning" of data into crude data bins) and the injection of noise (e.g. the addition of random perturbations to data) are intriguing methods to address this problem. Similarly, boot-strapping methods, in which queries are queued up until enough there is enough data in aggregate across a number of agents to allow individual users to hide behind other users and allow query responses without undesirable information leakage.

2.3 Reputation Mechanisms

Finally, an information mechanism will need to provide methods to address the information asymmetries that exist between buyers and sellers in negotiation about information. The seller knows the quality of the information, but a seller with high quality information needs a way to demonstrate the quality a prospective buyer, and avoid spiraling towards Akerlof's "Market for lemons" [1], where only low-quality information survives.

This problem is often addressed with *branding*; for example, the *New York Times* would not survive if any newspaper could pretend to be the *Times*, or if the *Times* had no recognizable identity [19]. Free samples are another way to handle this problem, for example as delivered by listening stands in music stores. Branding is not useful in ad hoc and large scale decentralized mechanisms, and in particular with pseudonymous and anonymous identities.

Reputation mechanisms [15, 6] can provide a similar effect to branding in highly dynamic and highly distributed multi-agent systems. A reputation mechanism provides a trusted method to aggregate and track feedback from participants in transactions, creating a *shadow of the future* [5]. The ability to adopt pseudonymous and anonymous identities can itself reduce the efficiency reputation mechanisms; for example, one effect is that the default reputation for a newcomer to a system must be low reputation [8]. There can be other problems, arising from the ability to *trade reputations* (for example via identity swapping), and the ability to form collusive rings to artificially inflate reputations [7].

A reputation mechanism for an information market would also need to rely on *accurate* reports of the quality of a transaction because it is unlikely that there will be an *objective* measure of quality. Methods to promote truthful reporting include checking for outliers [4], and also using proper scoring rules and payments to implement truthful reporting in equilibrium [12].

3 Information Clusters and a Market for Queries

We imagine a dynamic "agent soup", with autonomous and self-interested agents associated with user information able to form *ad hoc* coalitions, that compete with other information clusters to provide responses to queries. A similar vision, of *information crystals*, was recently outlined in Adar & Huberman [3]. A well-functioning information mechanism would provide incentives for the emergence of information clusters of the right size and content: making a tradeoff between the higher computational efficiency (from less coordination costs) of smaller clusters and the higher informational value and better privacy properties (via aggregation) of larger clusters.

As a straw-man model, we propose the following basic components for an information mechanism:

property rights/market Users can control the release of information through multiple information agents, each of which chooses to join an *information cluster*. The user can decide what subset of information to provide to each

information agent (and can provide the same information to multiple agents). Information agents can participate within clusters, to respond to queries as appropriate to maximize the payoff to the user, making appropriate tradeoffs between information revelation and payments. Information clusters compete in a *market for queries*, in response to queries that specify values for different accuracy-levels in responses. Clusters share collected payments across the information agents within a cluster.

control Each information agent controls the level with which data is aggregated and shared with other agents in a cluster in order to respond to queries, and is free to adopt a pseudonymous identity. In responding to queries, clusters aggregate input information from each agent in the cluster, and select the appropriate level of accuracy to quote in response to query request. Agent clusters can also support boot-strapping, in which queries are queued at an information cluster until the cluster gets large enough to respond to queries without compromising privacy.

reputation A reputation-web maintains reputation, with agents able to rate the information provided by a cluster of agents and evaluations weighted by the individual reputation of reporting agents. The reputations of individual agents within a cluster are adjusted based on the reported evaluations, and the reputation of a cluster depends in turn on an aggregation of agent reputations within the cluster. Incentive-based schemes can be incorporated within the reputation mechanism to promote truthful rating about the quality of responses to queries.

Looking forward, it will be interesting to understand the extent to which the roles of reputation mechanisms and incentive mechanisms orthogonal. Traditional price-based mechanism design focuses on one-shot interactions, while tending to ignore long-term strategic interactions. On the other hand, reputation mechanisms focus on repeated interactions, but perhaps lack the fidelity to implement optimal equilibrium behavior at any single point in time. Kalai & Ledyard [10] previously demonstrated that *repeated implementation* is more powerful than one-shot implementation, but no theory that combines reputation mechanism methodologies with one-shot mechanism design methodologies has yet been developed.

4 Conclusion

We propose the design of decentralized information-processing mechanisms as an important challenge problem for the agent research community. The problem is proposed as a response to the rapid emergence of pervasive computing and pervasive information acquisition. We imagine that these information mechanisms will provide incentives for the emergence of *ad hoc* information clusters, in which agents collaborate to combine user information within clusters, and clusters compete to respond to queries.

References

1. G. A. Ackerlof. The Market for Lemons: Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, 84:488–500, 1970.
2. E. Adar and B. Huberman. Free riding on Gnutella. *First Monday*, October 2000.
3. E. Adar and B. A. Huberman. A market for secrets. *FirstMonday*, August 2001.
4. C. Avery, P. Resnick, and R. Zeckhauser. The market for evaluations. *The American Economic Review*, 89:564–584, 1998.
5. R. Axelrod. *The evolution of cooperation*. New York:Basic books, 1984.
6. C. Dellarocas. Analyzing the economic efficiency of eBay-like online reputation reporting mechanisms. In *Proc. 3rd. ACM Conf. on Electronic Commerce (EC'01)*, pages 171–179, 2001.
7. R. Dingledine, M. J. Freedman, and D. Molnar. *Peer-to-Peer*, chapter 12: Free Haven. O'Reilly, 2000.
8. E. Friedman and P. Resnick. The social cost of cheap pseudonyms. *Journal of Economics and Management Strategy*, 10:173–199, 2000.
9. M. O. Jackson. Mechanism theory. In *The Encyclopedia of Life Support Systems*. EOLSS Publishers, 2000.
10. E. Kalai and J. O. Ledyard. Repeated implementation. *Journal of Economic Theory*, 83(2):308–317, 1998.
11. J. Kleinberg, C. H. Papadimitriou, and P. Raghavan. On the value of private information. In *Conf. on Theoretical Aspects of Rationality and Knowledge (TARK'01)*, 2001.
12. N. Miller, P. Resnick, and R. Zeckhauser. Eliciting honest feedback in electronic markets. Technical report, University of Michigan, 2002.
13. K. S. J. Pister, J. M. Kahn, and B. E. Boser. Smart Dust: Wireless networks of millimeter-scale sensor networks. Technical report, U C Berkeley, 1999. Highlight Article in 1999 Electronics Research Laboratory Research Summary.
14. A. Popescul, L. H. Ungar, D. M. Pennock, and S. Lawrence. Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. In *Proc. UAI'01*, August 2001.
15. P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara. Reputation mechanisms. *Comm. of the ACM*, 43:45–48, 2000.
16. J. B. Schafer, J. Konstan, and J. Riedl. Recommender systems in E-Commerce. In *Proc. 1st ACM. Conf. on Electronic Commerce (EC'99)*, 1999.
17. D. Tennenhouse. Embedding the Internet: Proactive computing. *Comm. of the ACM*, 43:43–43, 2000.
18. H. Varian. Economic aspects of personal privacy. Technical report, NTIA, 1996. Privacy and Self-Regulation in the Information Age.
19. H. Varian and C. Shapiro. *Information Rules: A Strategic Guide to the Network Economy*. HBS Press, 1998.