Protecting Privacy While Sharing Information in Electronic Communities

Tad Hogg (hogg@parc.xerox.com)
Bernardo A. Huberman
Matt Franklin
Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

A great deal of economically useful information is distributed among many people. Collecting and utilizing this information could potentially provide many useful services not only for the people from whom the information is obtained but also others with similar interests. For instance, recommendation systems rely on feedback about the quality or suitability of various items from those who have used them in the past. As another example, the development of reputations through dealing with a number of people can help future transactions with new people by establishing well-regarded brand names. Reputations are extremely valuable in the context of electronic commerce, for they provide a mechanism to establish trust and reduce the costs and risks involved in transactions.

However, privacy concerns over inappropriate use of the information make it hard to successfully exploit this knowledge. We thus face the challenge of addressing these concerns so as to achieve the gains from sharing information. More generally, this issue is an example of a social dilemma, where the group as a whole gains from the willingness of people to share information, but individuals face a cost of possible adverse reactions to their revealed information.

Several electronic mechanisms can reduce privacy concerns. For example, a useful strategy for maintaining privacy consists in the anonymous posting of information. In recommender systems this can be useful when the recommendations are based on coarse characteristics such as the number of people voting for a particular choice. But anonymity has the drawback of preventing users from learning the usefulness of recommendations from particular people, track trends over time, and to use reputations which are built up over repeated interactions. For instance, in recommendations for various products, a user of anonymous reviews would not be able to determine whether a particular review came from the product's seller or a competitor.

The consistent use of pseudonyms can address some of these issues, but not all. One drawback of pseudonyms is that the very link which establishes reputation over time becomes a vulnerability if authorship can be established by other means for any pseudonymous message. Issues of privacy can also be tackled by the use of trusted third parties to mediate the exchange of information. However, it can be difficult to get everyone in a community to agree on a suitable third party, particularly when new users continually enter the system and live in different jurisdictions. Furthermore, the collection of all information by a single third party can lead to a system-wide failure if such a party is compromised.

Instead, we propose addressing the inherent tradeoff between individual privacy and the potential benefit to the community from using the information, with mechanisms allowing users to adjust the amount of information revealed. Such mechanisms could allow revealing those aspects of the transaction most useful to the group while protecting those aspects of most concern to the individual.

Specifically, existing cryptographic techniques for for secure function evaluation allow people to determine a desired function of their private information without revealing any additional details of that information. For example, in the context of a recommendation system, people could find recommendations from people with whom they have many common preferences without revealing the details of their preferences. Further examples include discovering communities with shared preferences and removing the disincentives posed by liabilities. These techniques can also allow an individual to negotiate on behalf of a group by proving membership in that group without revealing one's identity. This last example could be particularly useful for allowing automated "shopping agents" to negotiate group discounts while minimizing privacy concerns on the part of members of the group.

A specific implemented example of these mechanisms allows for identifying groups with common interests by comparing browser bookmarks. In this case, the number of matching bookmarks is computed without revealing either exactly which bookmarks match or the actual url's of the bookmarks.

These mechanisms involve trade-offs among computational efficiency, the leaking of information and ease of use. These trade-offs can be resolved differently depending on the specific application. For example, one may want to make it easier for new people to join a community by lowering the number of passwords and preferences that need to be listed, at the expense of reduced privacy. Another instance would be one in which everybody in a group shares the same key, which is a simple and secure procedure as long as no one leaves the group.

Additional trade-offs appear when on considers spoofing, whereby people can present false preferences in order to gain access to privileged information or to deter others from gaining an advantage from a weak adversary. One response might be anonymity, but at the cost of loosing the benefit of reputation building. Another one could be analogous to biological situations, where false signalling is used by many organisms to deter attack or to gain access to valuable resources. A strategy that has evolved to address the problem of spoofing in that context is for signals themselves to be costly to produce, and thus to imitate. Similar strategies could be applied to electronic communities by increasing the number of challenges needed to access a given group, or by imposing a waiting period. On the other hand, this could deter legitimate new people from joining the group. Moreover, even if the trade-offs could be negotiated successfully, there remains the problem of misusing these techniques, as in the case of fraudulent financial transactions, insider trading or the unauthorized collection of personal data.

In spite of the great potential for electronic commerce that the Web is enabling through its global reach, there are vast areas of knowledge and expertise that remain untapped for lack of mechanisms that ensure privacy and trust. Applications built on secure function evaluation make it easier to access vast repositories of information that are not readily known to producers and consumers, thus leading to improvements in economic efficiency through the more focused use of resources.

Further Reading

Discussions of reputation and trust, as well as some uses for on-line data collection are given in:

D. Klein, Reputation: Studies in the Voluntary Elicitation of Good Conduct, Univ. of Michigan Press, Ann Arbor, 1997

S. Singleton, “Data Collection as Free Speech”, Computer-Mediated Communication Magazine, Sept. 1997 (www.december.com/cmc/mag/1997/sep/single.html)

F. Fukuyama, Trust: The Social Virtues and the Creation of Prosperity”, Free Press (1996).

For introductions to secure function evaluation see R. Fagin, M. Naor and P. Winkler, “Comparing Information Without Leaking It”, Communications of the ACM, 39, 77-85 (1996) and the examples at http://www.wisdom.weizmann.ac.il/home/naor/public_html/PUZZLES/

Applications for community discovery, preference matching and deniable recommendations are described in B. Huberman, M. Franklin and T. Hogg, “Enhancing Privacy and Trust in Electronic Communities” in Proc. of the ACM Conf. on Electronic Commerce, pp. 78-86, 1999 (see also http://www.parc.xerox.com/iea/www/privacy.html).