Privatizing Constraint Optimization, Thesis Abstract

Rachel Greenstadt
Harvard University
greenie@eecs.harvard.edu

My thesis proposes novel schemes for achieving privacy in constraint optimization and new ways of analyzing the privacy properties of constraint optimization algorithms. Thus, it aims to contribute to the fields of constraint processing and multiagent systems.

Motivation: Many problem domains within constraint processing require algorithms that better accomodate the need for privacy. Two salient examples are described in the following paragraphs.

Coordinating the schedules of muliple agents is a problem domain that naturally motivates the need for privacy preserving constraint optimization algorithms. In this domain, a number of agents wish to schedule events subject to individual constraints, which they wish to keep private. Their calendars may contain the intersection of multiple professional endeavors as well as events relevant to their personal lives. The agents involved want a solution that will optimize global social welfare, but are not willing for everyone to know their preferences. For example, scheduling constraints may expose attendance at support group meetings.

Resource allocation problems provide another motivating example. Imagine several government, corporate and nonprofit groups that come together to help with some disaster relief effort. They all wish to contribute resources that will lead to the optimal end, but they are mutually distrustful in most of their other endeavors and do not want the other entities to know the details of their individual holdings and constraints.

Background: Initially, all constraint processing algorithms were centralized; all constraints were communicated to and optimized by a central server. The central server thus knew everything about everyone. Concerns about privacy and efficiency led to the development of distributed constraint satisfaction and optimization algorithms (Yokoo *et al.* 1998; Modi *et al.* 2005). In these algorithms, each agent owns its own constraints and passes messages as necessary to other agents to collaboratively compute the solution. Since information is distributed among many agents, these algorithms were thought to provide more privacy. However, recent work has shown that in some cases, distributed algorithms lead to greater privacy loss than the traditional centralized solutions; everybody learning everything about everybody is not necessarily an improvement over a centralized agent learning everything about everybody (Maheswaran *et al.* 2005). The main thrust of this recent work, which I have made use of in my work to date, was to propose a framework for quantitative evaluation of privacy in distributed constraint optimization (DCOP) algorithms.

Other work has introduced a cryptographic approach to achieving privacy (Silaghi & Mitra 2004; Yokoo, Suzuki, & Hirayama 2002). These approaches use secure multiparty computation techniques to distribute and solve the constraint optimization or satisfaction problem, adding considerable overhead. Thus, they are severely resource intensive as constraint processing deals with NP-complete problems and adding exponentiations to *each* comparator operation is nontrivial.

Research Focus: My research aims to determine methods for analyzing privacy loss in existing algorithms and design new algorithms that lose less privacy without resorting to expensive cryptography.

We can investigate many parameters in traditional search strategies and create new algorithms that lead to increased privacy. Therefore my approach includes detailed experimental analysis of existing algorithms—to understand how these parameters interact to achieve privacy or privacy loss. I begin by analyzing the privacy properties of algorithms that are run once, then broaden my analysis to repeated use of the algorithm, for example, in the scheduling domain where similar sets of agents schedule new events after some existing events have already been scheduled. This analysis will aid in the development of new algorithms that are more suitable for use in settings where privacy matters.

Initial Results: The work by Maheswaran et al. provided a starting point (Maheswaran *et al.* 2005). This prior work produced a negative result, showing that the DCOP algorithms examined performed worse than centralized algorithms with regard to privacy. However, this work did not examine the newest and most commonly used DCOP algorithms, such as ADOPT (Modi *et al.* 2005) and DPOP (Petcu & Faltings 2005). My extensive experimentation shows that these new algorithms provide improved privacy over centralized algorithms, validating the approach of using the distributed paradigm to privatize constraint optimization. My work also introduces the MAX metric which measures the largest amount information accumulated by a

single agent. This metric helps to quantify intuitive ideas about privacy in these systems.

While previous work investigated the impact on efficiency of such distributed contraint reasoning design decisions as constraint-graph topology, asynchrony and message-contents, my work examines the privacy impact of such decisions, providing an improved understanding of privacy-efficiency tradeoffs. My work augments previous work on system-wide privacy loss, by investigating inequities in individual agents' privacy loss.

This experimental analysis concluded that: (1) Privacy results in Adopt and DPOP were better than for centralized algorithms. (2) Asynchrony in Adopt improves privacy by obscuring the identities of agents involved in a message and by making sophisticated inference difficult. This benefit is offset to a degree by its use of more messages. (3) Topology has significant impact on both system-wide and individual privacy loss.

This initial work will be presented as a short paper at AAMAS 2006 (Greenstadt *et al.* 2006). An extended version is under submission and further experimentation and analysis is ongoing.

Current Work: The privacy impact of repeated use of DCOP algorithms is an important area for further analysis. Repeated use is prevalent in the scheduling domain, where much of the same private information is used in multiple scheduling problems over time. Information not revealed by a single run of an algorithm may be inferred over time by repeated runs on slightly different sets of agents and constraints. In the anonymity literature, such inference is known as the intersection attack (Danezis 2003), and it is devastating to privacy. It is important to understand how repeated use affects the privacy properties of these algorithms so that agents do not believe that they are getting a higher level of privacy than is actually the case and also so that we can design more private algorithms with repeated use in mind. An initial treatment of this issue in the DisCSP setting is presented in (Modi & Veloso 2005).

My goal is to develop new algorithms or modifications to existing algorithms which provide less privacy loss than the distributed algorithms available now. My experimental analyses have identified key features of algorithms that affect privacy, and I will design algorithms around those features, drawing on analogous situations in the fields of anonymity and trust management.

My initial results have shown that key features of algorithm design where privacy is concerned are topology and asynchrony. I intend to produce algorithms that vary these features and provide increased privacy.

Topology and asynchrony are centrally important because privacy in these algorithms is derived from aggregating the results of multiple agents so that the agent receiving the message cannot infer the private information of individual agents. In the distributed algorithm topologies that rely on a tree or chain of agents, agents near the top send fewer messages, those that they send are aggregated with all the messages of agents below them. Asynchrony can cause an inferring agent to be uncertain about which agents' information is aggregated together in received messages, thus increasing privacy. On the other hand, asynchrony may cause unnecessary messages to be sent, reducing privacy.

Techniques for aggregating data to obscure its origin have been extensively studied in the anonymity literature, producing many strategies for batching and mixing data (Serjantov, Dingledine, & Syverson 2002). Delaying messages for better aggregation has often produced better privacy (Pfitzmann & Waidner 1985), suggesting that we too may be able to use these batching strategies to make effective privacy/efficiency tradeoffs.

We should also be able to leverage existing trust relationships between nodes to build topologies that increase privacy. Since nodes directly below other nodes tend to lose privacy to those nodes, we will explore algorithms that ensure that these pairs of nodes have a trust relationship. This use of trust relationships has proved useful in the reputation and peer-to-peer domains.

References

Danezis, G. 2003. Statistical disclosure attacks: Traffic confirmation in open environments. In *Proceedings of Security and Privacy in the Age of Uncertainty, (SEC2003)*.

Greenstadt, R.; Pearce, J. P.; Bowring, E.; and Tambe, M. 2006. An experimental analysis of privacy loss in dcop algorithms. In *AAMAS*.

Maheswaran, R. T.; Pearce, J. P.; Varakantham, P.; Bowring, E.; and Tambe, M. 2005. Valuations of possible states (VPS):a quantitative framework for analysis of privacy loss among collaborative personal assistant agents. In *AAMAS*.

Modi, P. J., and Veloso, M. 2005. Bumping strategies for the multiagent agreement problem. In *AAMAS*.

Modi, P. J.; Shen, W.; Tambe, M.; and Yokoo, M. 2005. ADOPT: Asynchronous distributed constraint optimization with quality guarantees. *Artificial Intelligence Journal* 161:149–180.

Petcu, A., and Faltings, B. 2005. A scalable method for multiagent constraint optimization. In *IJCAI*.

Pfitzmann, A., and Waidner, M. 1985. Networks without user observability – design options. In *Proceedings of EUROCRYPT 1985*. Springer-Verlag, LNCS 219.

Serjantov, A.; Dingledine, R.; and Syverson, P. 2002. From a trickle to a flood: Active attacks on several mix types. In *Information Hiding Workshop*.

Silaghi, M. C., and Mitra, D. 2004. Dist. constraint satisfaction and optimization w/ privacy enforcement. In *IAT*.

Yokoo, M.; Durfee, E. H.; Ishida, T.; and Kuwabara, K. 1998. The distributed constraint satisfaction problem: formalization and algorithms. *IEEE Transactions on Knowledge and Data Engineering* 10(5).

Yokoo, M.; Suzuki, K.; and Hirayama, K. 2002. Secure distributed constraint satisfaction: Reaching agreement without revealing private information. In *CP 2002*.