Privacy-Enhanced Data Management for Next-Generation e-Commerce

Chris Clifton
Department of Computer Sciences
Purdue University
clifton@cs.purdue.edu

Irini Fundulaki, Richard Hull, Bharat Kumar,
Daniel Lieuwen, and Arnaud Sahuguet
Bell Labs, Lucent Technologies
{hull,bharat,lieuwen,sahuguet,fundulaki}@lucent.com

Electronic commerce is becoming pervasive. Convergence of wireless, wireline, and telephony networks enables a new level of web services. Will these services be a benefit, or a new avenue for spam? Extensive profiling and information sharing is needed to ensure that people get the services they want, and only what they want. We must ensure this personal, private information is used properly – to deliver desired web services. We must ensure that *pervasive* doesn't become *invasive*. Meeting this goal requires advances in several technologies: profile data management, preference and policy management, personalized and privacy-conscious data sharing. Achieving these requires technology development in several areas:

Sharing profile data across devices and services. Sharing address books across different devices (e.g., PDA, cell phone, home/office phone) is typically cumbersome to impossible. Sharing data between web store fronts with today's technology is essentially impossible. How can we achieve an "enter once / use everywhere" state of being?

Combining real-time context data with web services. Cell phones will soon provide GPS/911 location information, which can be combined with web-resident restaurant locators, buddy finders, etc. More generally, how can the real-time context information about end-users be made accessible on a large scale, for use by huge numbers of web services?

Supporting intricate preferences. To properly respect a user's preferences, many kinds of information must be combined in intricate ways. An example service is "selective reach-me". This service uses presence information (e.g., from wireless network, from instant messaging, from onhook/ off-hook information for wireline phones), end-user calendar, and end-user preferences so that incoming calls will be routed to the appropriate device. Since peoples schedules and habits are varied, the logic to be used for routing may be complex and inter-related. Some example rules are:1) During working hours, if end-user's presence is "available" (e.g., verified with IM), then call office phone first, and then try cell phone. 2) During working hours, if end-user is in a scheduled meeting with her boss, or speaking to her boss on the phone, then convert all acceptable incoming calls into instant messages using speech-to-text (caller) and text-to-speech (callee).

Different classes of users (sales people, field reps, students, emergency workers, ...) may need highly different kinds of preference specifications. Also, it must be easy for end-users to provision their preferences, even if rather intricate, and easy to override them when special circumstances arise. For wide acceptance, much of this customization must be nearly automatic, leading to a need for data mining technology to help identify and establish such profiles.

The data required to generate and utilize these profiles raises significant privacy issues both for individuals and for businesses, who must keep their business intelligence secure. Therefore, advanced services must be provided with limited sharing of data in *real-time*.

This tutorial will survey reasons behind these privacy constraints and their impact on data management, emerging technologies that enable the necessary private data management, and research needed to achieve this vision. Specific topics include:

Problem space: What we would like to do but currently can't. This will be presented in the context of motivating examples for data mining, real time profile management, and a combined scenario.

Privacy and security constraints: Types of constraints and their sources, including legal regulations (Telecomm regulations, EU 95/46, etc.), contractual obligations, and others.

Survey: Web services issues and standards related to this problem (e.g., XACML, P3P/APPEL) systems and recent research results (e.g., Hippocratic databases, Houdini).

Privacy preserving data mining technologies: Data obfuscation and data partitioning/secure multiparty computation. Technical details of representative algorithms from each domain.

Research agenda: key problems to solve to allow personalized and privacy-conscious data sharing.

**Proceedings of the 29th VLDB Conference,
Berlin, Germany, 2003**