Managing Healthcare Data Hippocratically

Rakesh Agrawal[a] Ameet Kini[b] Kristen LeFevre[b] Amy Wang[c] Yirong Xu[a] Diana Zhou[b]

[a]IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120

[b]University of Wisconsin
1210 West Dayton Street
Madison, WI 53706

[c]UCLA Anderson School
110 Westwood Plaza
Los Angeles, CA 90095

1. INTRODUCTION

Data privacy is a growing concern among businesses and other organizations in a variety of sectors, such as healthcare, finance, e-commerce, and government. Every day, these organizations are entrusted with the responsibility of managing personal information. Unlike data security, which focuses primarily on preventing unauthorized individuals from inappropriately obtaining information, the privacy problem focuses on providing individuals the ability to control how their data is managed and used by a particular organization. We introduce a prototype implementation addressing several key issues in privacy management, and we demonstrate this prototype in the context of healthcare data management, a sector in which maintaining the privacy of individual information is of paramount importance.

The idea of a Hippocratic database was introduced by [1], and is founded on the premise that database systems should take responsibility for protecting the private data they manage. The authors describe ten principles governing the design of such a system, as well as a "strawman" architecture. In this demonstration, we describe the realization of several of the components and principles laid out by [1] in a relational database context:

- Privacy policy definition and installation
- Limited data collection that respects the privacy preferences of individual data subjects
- Privacy policy enforcement that limits the use and disclosure of data to that specified in the privacy policy and consented to by the data subject

2. PROTOTYPE OVERVIEW

Our prototype implements the three components just mentioned, and a high-level architectural diagram is provided in Figure 1.

The prototype uses an electronic privacy policy to define an organization's information management practices. In particular, this information can be encoded in one of two XML-based policy definition languages: the Platform for Privacy Preferences (P3P) [4], the W3C recommendation, or the Enterprise Privacy Authorization Languages (EPAL) [3], developed by IBM. The policy is shredded, and the necessary information stored inside the relational database as tables called the "privacy meta-data."

In practice, we found that it was possible to express privacy policies in the meta-data in a way that is largely language independent. Both P3P and EPAL encode rules for allowing or disallowing disclosure of data based on a combination of several factors, including some notion of *purpose*, *data recipient*, *data category* and *condition*. The primary type of condition is an opt-in/opt-out choice, specified by the user providing the private data, though EPAL also supports more complex conditions. The privacy meta-data stores a set of rules of the form <*purpose*, *recipient*, *data category*, *condition*>, indicating that the privacy policy allows for the disclosure of a particular category of data to a particular recipient for a particular purpose, provided that the indicated condition holds. For example, a rule might indicate that medical history is provided to external drug companies for research if an individual "opts in" to this choice. The meta-data also stores a mapping of data categories to relational attributes.

The next main component of the prototype controls private collection of personal information, matching an organization's privacy policy with individual user preferences. Here, the user specifies his or her preferences for data access and usage using the XPath-based preference language, XPref [2]. These preferences are submitted to the server, and matched against the meta-data for the policy stored in the database. This matching is accomplished using a standard XPath engine, and returns a result of either "allow," or "block." In the case of "block," the matching engine also returns the particular preference(s) that caused the conflict.

Finally, queries executed over the database are automatically modified to reflect the rules stored in the privacy meta-data. Specifically, the implementation captures incoming queries, and based on the context of the issuing application, the enforcement module infers the purpose and recipient, and augments the query as necessary to enforce the privacy policy, opt-in/opt-out choices, and conditions. The augmented query is then executed, and the results returned, transparent to the application. The policy enforcement can be quite granular, resolving conditions and opt-in/opt-out rules at the level of the individual data cell. This mechanism of query rewriting is generally much more efficient than the alternative, application-level enforcement.

3. DEMONSTRATION

Although healthcare is just one of the sectors that could benefit from Hippocratic database technology, the need has become particularly visible in the American healthcare industry due to new legislation mandating the protection of patient medical information. The Health Insurance Portability and

SIGMOD 2004, June 13–18, 2004, Paris, France.

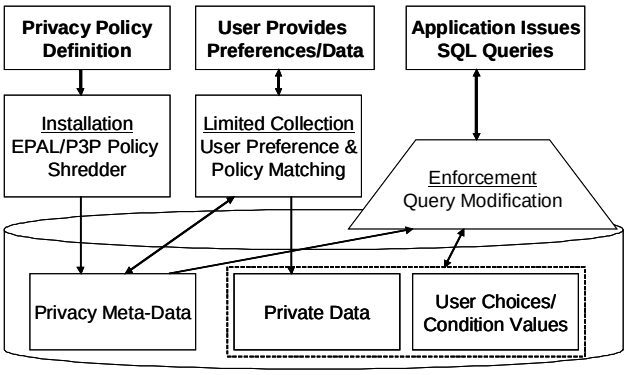

Figure 1: Hippocratic Database Prototype Architecture

Accountability Act (HIPAA) mandates that healthcare providers follow certain guidelines of privacy protection and that patients are given some specific control over how personal medical information is used and disclosed [5]. This has caused a flurry of activity in the medical privacy arena. For this reason, we demonstrate our prototype as applied to a medical information management system.

Our web-based demonstration is based on a hypothetical, yet realistic, healthcare provider, dubbed "NetCare." Through a familiar corporate-style web portal, we follow a cast of characters to depict the functionality of the system. We first demonstrate how an electronic privacy policy is created and installed into the Hippocratic database by our company's Chief Privacy Officer. The existing HIPAA-based privacy policy is first (manually) translated into electronic P3P or EPAL, and through NetCare's web portal, the Chief Privacy Officer installs the electronic policy into the database. This part of the demonstration highlights the ability of P3P and EPAL to express natural language privacy policies, as well as the ability of our architectural design to support disclosure rules expressed in either language.

To demonstrate limited data collection, we consider a new patient of NetCare. Using an Internet Explorer toolbar extension, the new patient defines her personal privacy preferences, choosing from a pre-defined set of rules typical of the healthcare domain. Each English statement corresponds to an XPref preference rule. After defining her preferences, the new patient attempts to create a new account through NetCare's web portal. The policy/preference matching process we described previously occurs behind the scenes, and displays a warning message if there is a conflict between the patient's personal preferences and NetCare's privacy policy. If a conflict occurs, the warning message points out the particular preference rule(s) that caused the mismatch. Here we demonstrate three main ideas: We showcase XPref as a preference specification language compatible with P3P and EPAL. We demonstrate server-side policy/preference matching, and we demonstrate the ease with which the new patient is able to choose between readable English preferences, and should a conflict occur, the ease with which she is able to understand the reason.

Finally, we demonstrate the enforcement of privacy policy limited disclosure, or access control, rules. More specifically, we illustrate how the data obtained by querying the database violates neither the organization's privacy policy nor the individual user's preferences. The demonstration includes two scenarios. The first scenario involves a query issued by a lab technician, a NetCare internal employee. For the purpose of lab work, we show how our form of privacy enforcement removes prohibited information from the result set of a query, and the results of the query conform to both the patient's privacy preferences, and the rules encoded in the corporate privacy policy. The second scenario involves a query from an external partner of NetCare that accesses patient data for research purposes, demonstrating the applicability of the system to both internal and external users. This portion of the demonstration highlights two important features: the ease with which Hippocratic database technology can be integrated into an existing infrastructure without application modification, and the cell-level granularity of enforcement provided by our enforcement mechanism.

4. ACKNOWLEDGMENTS

This work was done at IBM Almaden as part of an Extreme Blue project, sponsored by Arvind Krishna and Harriet Pearson. Our thanks to Jerry Kiernan, Ramakrishnan Srikant, Roberto Bayardo, Calvin Powers, Phil Fritz, and the Almaden Extreme Blue staff, particularly Dave Cheney, for their insights.

5. REFERENCES

[1] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. Hippocratic Databases. In *Proc. Of the 28th Int'l Conference on Very Large Databases*, Hong Kong, China, August 2002.

[2] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. An XPath-based Preference Language for P3P. In *WWW2003*, Budapest, Hungary, May 2003.

[3] P. Ashley, S. Hada, G. Karjoth, C. Powers, and M. Schunter. *Enterprise Privacy Authorization Language 1.2 (EPAL 1.2).* W3C Member Submission, November 2003.

[4] L. Cranor, M. Langheinrich, M. Machiori, M. Presler-Marshall, and J. Reagle. *The Platform for Privacy Preferences 1.0 (P3P 1.1) Specification*. W3C Recommendation, April 2002.

[5] US Department of Health and Human Services. *Summary of HIPAA Privacy Rule.* http://www.hhs.gov/ocr/hipaa/ 2004.