Enabling the 21st Century HEALTH CARE INFORMATION TECHNOLOGY REVOLUTION

The U.S. government's vision of the health care information infrastructure is possible using technologies that support the sharing of medical e-records while maintaining patient privacy.

The U.S. President's Information Technology Advisory Committee (PITAC) released a report in June 2004 entitled *Revolutionizing Health Care Through Information Technology* [10] that contained comprehensive findings regarding the potential of information technology (IT) to reduce medical errors, lower costs, and improve patient care. It also recommended a technological framework for transitioning from manual, paper-based health records to a modern, computerized electronic records infrastructure.

This article introduces several technologies, known collectively as the Hippocratic Database (HDB) [4], that address PITAC's findings and recommendations regarding electronic health records maintenance, computer-assisted decision support, and exchange of health information. We intend to show that the PITAC vision of

Illustration by PAUL WILEY

By RAKESH AGRAWAL, TYRONE GRANDISON, CHRISTOPHER JOHNSON, *and* JERRY KIERNAN

revolutionizing health care through IT is technically feasible by illustrating how some of its key findings can be realized using HDB technology.

The four core elements of PITAC's recommended framework for a 21st century health information infrastructure are to establish a nationwide system of electronic health records that provides caregivers with all relevant information about every patient; encourage the sharing of medical knowledge through computer-assisted clinical decision support; facilitate computerized order entry among providers for tests, medicine, and procedures; and ensure secure, private, interoperable exchange of health information.

The PITAC Report presents 12 specific findings and recommendations for research innovations necessary to promote the development of a modern electronic medical records system. These are intended to provide a roadmap for achieving PITAC's vision of a health records infrastructure that safeguards personal privacy; uses standard clinical terminology that can be read and interpreted by any health care provider and incorporated into a computerized system to assist decision making; eliminates errors associated with handwritten and paper-based records; and enables secure transfer of records to support patient care and electronic information sharing.

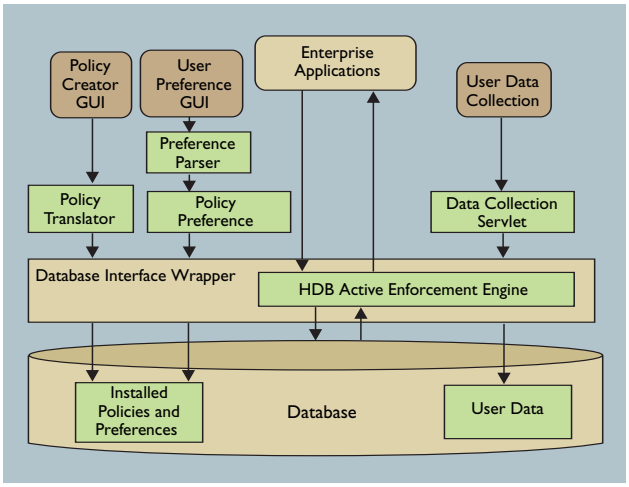

Figure 1. HDB Active Enforcement architecture.

This article provides an overview of HDB technologies that support PITAC's vision. We discuss how HDB Active Enforcement enables policy-based privacy management, describe methodologies for efficient data access and disclosure tracking, including HDB Compliance Auditing and Database Watermarking, and present Sovereign Information Integration, which enables the secure exchange of private health information. We also discuss techniques for de-identifying and analyzing sensitive information, including Privacy-Preserving Data Mining and Optimal *k*-anonymization.

Policy-Based Privacy Management

The PITAC Report strongly emphasizes the importance of safeguarding the personal privacy of patients in managing electronic health records. While e-records systems facilitate the sharing and transmission of health data, they also increase the potential for privacy abuses. PITAC stresses that "secure, private, interoperable, electronic health care information exchange" is critical to its vision of the 21st century health care IT infrastructure. HDB's Active Enforcement component advances this vision by enabling enforcement of fine-grained data disclosure policies.

Active Enforcement (AE) [4] is an agnostic middleware solution that responds to emerging privacy and security needs. It allows patients and health care institutions to negotiate policies governing the disclosure of their personal information. These policies are imposed between the enterprise applications and the database to ensure that any access and disclosure of health information is in accordance with patient preferences and applicable laws.

AE manages access to patient data in a secure and trusted manner that protects patient privacy. A decisive advantage of AE is its ability to manage disclosure at the cell level in the database, rather than just the row or the column level. AE includes three main phases—policy creation, preference negotiation, and application data retrieval (see Figure 1).

In the *policy creation* stage, the health care institution defines a data disclosure policy to specify who is allowed to access what information; for what purposes an item of information may be accessed, and to which recipients it may be disclosed. Policies are expressed in a privacy specification language such as P3P [8] and installed in the AE engine in a form amenable to symbolic manipulation.

In the *preference negotiation* stage, which occurs prior to providing any personal data, patients indicate their preferences regarding the use and disclosure of their personal information. These preferences are translated into a preference language [5] and communicated to the health care institution through a simple Web browser plug-in. HDB then matches these preferences with the institution's privacy policies to identify any conflicts. It advises the patient of these conflicts and provides him with an opportunity to resolve them or terminate the relationship. The patient may then be given a series of choices to allow

him to opt in or opt out of certain disclosures of his personal information. A successful negotiation confirms agreement between the patient and health care institution regarding access and disclosure of the patient's personal information.

In the *application data retrieval* phase, HDB programmatically modifies all queries to be executed on the data source so that the application only retrieves results that are compliant with the institution's disclosure policies (including legal requirements) and patient preferences (including opt-in and opt-out choices). This process is fully automated and is only dependent upon negotiated policies and preferences.

Depending on the application domain and the nature of the problem, the preference negotiation component may be excluded. For instance, certain organizations may be responsible for enforcing privacy policies that have already been negotiated by an affiliate company and therefore have no need to negotiate privacy preferences.

The key strengths of HDB AE include the following:

- It offers a general methodology for handling and codifying policy and preference information;
- Its policy enforcement is transparent to enterprise applications (integration assumes a database interface such as ODBC or JDBC);
- It is agnostic to underlying database technology; and
- It may improve query processing speed depending on application and choice selectivity.

Encryption provides a second layer of protection against direct intruder attacks on the database. AE can be used in combination with techniques that allow queries over encrypted data without significantly degrading performance [6]. Here are two scenarios illustrating AE's functionality.

AE Scenario One. Bob is a patient considering selecting NetHMO as his primary health care provider. He is concerned about the privacy of his personal information. NetHMO is a national organization with a large and active list of patients. It maintains a number of alliances with laboratories and drug research companies.

Policy Creation: After NetHMO installs HDB AE, its management reviews the company privacy policy to ensure it is consistent with the Health Insurance Portability and Accountability Act (HIPAA) and management's business objectives. Next, NetHMO's Chief Privacy Officer, Alex, articulates the policy in the chosen privacy language. He then installs the policy through the database administrative console, the HDB Control Center GUI.

Preference Negotiation: Bob logs onto NetHMO's Web site and submits his privacy preferences to NetHMO through his Web browser prior to entering any personal information. Among other things, Bob indicates that he does not want to share his medical information with his employer for insurance coverage purposes and that he does not want to share his personal information with third parties for marketing purposes.

AE compares Bob's privacy preferences with NetHMO's privacy policy and uncovers one potential conflict. NetHMO's policy is to release all medical data to employers for employees seeking workers compensation insurance. This conflicts with Bob's preference not to share any medical information with his employer, but there are no other conflicts. After being notified of this conflict, Bob decides to waive his preference regarding employer disclosure and completes an application to join NetHMO.

Prior to submitting his completed application, Bob is provided with two opt-in choices on the information collection screen. These choices are intended to allow the patient and health care provider to strike a balance concerning the provider's discretionary use of his personal information. For the first choice, Bob consents to share his medical information with third parties for research purposes. For the second choice, he consents to disclose his contact information, but not his medical information, for marketing purposes.

Application Data Retrieval: After joining NetHMO, Bob undergoes a series of routine medical tests in connection with his annual physical. Shortly thereafter, Wanda, a NetHMO marketing manager, requests the complete e-records of all patients that have undergone physicals in the past six months. She would like to generate leads for a joint marketing campaign with a manufacturer of a new high blood pressure medication. In the absence of AE controls, Wanda would be able to see the complete medical records of all patients meeting her search criteria, irrespective of the patients' privacy preferences. How-

Active Enforcement is an agnostic middleware solution that responds to emerging privacy and security needs. It allows patients and health care institutions to negotiate policies governing the disclosure of their personal information.

ever, with AE controls in place, the database returns only information the patients have consented to share for marketing purposes. In accordance with Bob's opt-in choices, the query results include his contact information, but not his medical information.

AE Scenario Two. Suppose that Jane is a drug researcher with Innovative Pharmaceutical Company (IPC), a NetHMO partner that accesses NetHMO's database periodically to conduct statistical research. Jane logs into IPC's Web portal to NetHMO's database and issues the following SQL query:

*Select * from patients where diagnosis = 'hypertension'*

Without any HDB enforcement controls, Jane would be given total access to the complete records of all patients with hypertension in NetHMO's database. This is a violation of NetHMO's privacy policy and the HIPAA Privacy Rule, because not all patients have consented to reveal their personally identifiable information to third parties for research purposes. However, with AE in place, Jane's query is rewritten to comply with NetHMO's privacy policy and each individual patient's opt-in and opt-out choices. Instead of merely accepting or rejecting Jane's query, HDB reveals data of those patients, including Bob, that agreed to share information with third parties for drug research purposes, and only data they consented to share.

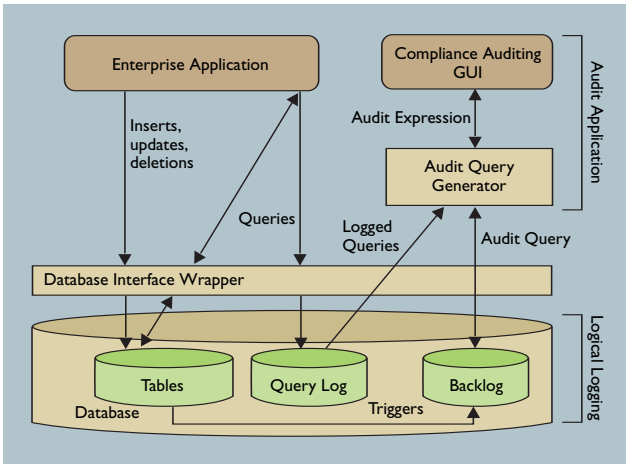

Figure 2. HDB compliance auditing.

Efficient Data Access Tracking

The PITAC report also encourages the development of data access tracking (or auditing) systems that enable patients, clinicians, and health care organizations to identify those who access patient information and the appropriateness of their access (Finding 12). The report specifically emphasizes the deterrent effect of such tracking systems on privacy breaches.

PITAC observes that current auditing systems are often turned off because of the computational and storage resources they consume. Thus, it emphasizes the critical need to develop cost-effective access and logging systems to protect the privacy of patient data. The report notes that "health information can only be accessed with adequate security and privacy if there are clear means for verifying the identities of those accessing and altering the data" (Finding 11).

HDB's Compliance Auditing component performs precisely the type of efficient data access tracking suggested by the PITAC Report. It allows health care institutions to track the identities of those who have accessed each cell in the database, and the exact version of the data they accessed, without consuming the computational and storage resources required by other audit systems. Another HDB component, Database Watermarking, tracks the origin of a leaked or misappropriated patient database by tracing a hidden bit pattern embedded in the data. Both technologies are discussed here.

HDB Compliance Auditing

HDB Compliance Auditing [1] enables health care institutions to determine whether they have complied with data protection laws, company policies, and patient preferences, either proactively or in response to a specific inquiry.

HDB auditing consists of a logical logging system and an audit application. The logical logging system records all queries and contextual information (identity, purpose, time, and so on) in query logs. It also stores all data updates, insertions, and deletions in backlog tables. The audit application uses the query logs and backlog tables to reconstruct the state of the database at any given time in the past. Upon receiving an audit query from an auditor interface, the application selects suspicious queries from the query logs. A suspicious query is defined as a query that shares an indispensable tuple (row) with the audit query. It then compiles all suspicious queries into a single SQL audit query, which it executes against the backlog tables. The audit application then returns an audit trail that identifies the user, recipient, purpose, and time of all suspicious queries, as well as the exact information disclosed by each query (see Figure 2).

In contrast, other auditing systems incur a substantial performance penalty to log the results of all database queries, including read queries. Moreover, the sensitive data disclosed by a query might not be reflected in its output, particularly if the query aggregates values from the records accessed. The alterna-

tive of logging the records accessed by the query also may not be feasible, as it is non-trivial to determine the precise records accessed by a complex query. Such a method would also dramatically increase the logging overhead. HDB avoids these problems by logging only minimal information (the query string, purpose, and recipient) during query processing and deferring all computation until the time of audit.

HDB can use several different methods to create and store the backlog tables. It can use triggers to record insert, update, and delete operations in local database tables. Alternatively, if the database supports replication or point-in-time queries, these features can provide the necessary backlog infrastructure for Compliance Auditing, with minimal additonal storage overhead. Methods such as those proposed in [12] can be used to detect any tampering to ensure the integrity of the query logs and backlog tables.

The following scenario illustrates the access tracking capabilities of Compliance Auditing in investigating whether a health care provider is responsible for the unlawful disclosure of private health records.

Compliance Auditing Scenario. Claire is a software engineer who has developed a business plan for a startup company. She presents her idea to a venture capital firm, with the goal of securing a first round of funding to launch her company. Immediate feedback is very positive, but soon thereafter, Claire receives a call informing her that her funding request has been denied. Upon investigation, Claire discovers that her funding was denied because the firm's (unauthorized) background check revealed her heart condition. The firm determined that Claire's health condition made an investment in her startup too risky.

Claire believes this information was disclosed by her health care provider, NetHMO. Therefore, she requests an accounting from NetHMO of all who have accessed her health information. Alex, NetHMO's privacy officer, would like to determine exactly who accessed Claire's private records and whether NetHMO was responsible for an unauthorized disclosure.

Audit Specification: Alex logs into the HDB Compliance Auditing interface and creates a new audit specification. The auditing interface is preset with a number of common tasks that a health care privacy officer might want to perform. To perform special tasks or advanced audits, an auditor can directly specify statements in a SQL-like syntax to execute the audits. He can also expand or restrict the time frame of the disclosures he would like to audit.

Disclosure Accounting: Alex's first audit task is to provide Claire with a written "disclosure accounting" that must contain certain access and disclosure information required by the HIPAA Privacy Rule. Alex selects the Disclosure Accounting task and requests an accounting of all persons who have accessed Claire's personal information. The application produces a written report of disclosures, which lists all searches (queries) that touched her records, sorted by time and purpose, and the exact results returned in response to each search.

Investigation of Suspicious Access: After obtaining the required disclosure accounting for Claire, Alex settles into the task of trying to pinpoint who may be at fault for the disclosed information.

Alex would first like to know which hospital workers accessed Claire's medical information. To accomplish this, Alex selects the Accounting of Access and Disclosure task and restricts his search to only Claire's medical records, rather than all of her personal information. The application returns a list of persons who have accessed Claire's medical records, the time and purpose of each access, and the exact data returned in response to each query.

The audit results indicate a large number of queries have accessed Claire's medical records, but not all of those queries revealed the diagnosis of Claire's heart condition. Thus, Alex adds a custom column based on diagnosis to further sort the information, so that he can isolate those queries that accessed a particular diagnosis. This view shows that many queries returned information about Claire's asthma condition, but nothing about her cardiac problem. These queries can be disregarded, as they could not have resulted in the wrongful disclosure of Claire's medical history.

Among the queries that accessed Claire's heart condition diagnosis, Alex sorts his results by user. Comparing the user identities with his record of Claire's treating physicians, he notes that her primary physicians and nurses frequently accessed her data. However, another physician, Dr. Richards, who is not listed as one of Claire's physicians, also accessed her data on several occasions over a short time period.

Alex is suspicious of Dr. Richards' access patterns, so he specifies another audit to determine the information that she has accessed. He notices that Dr. Richards has made only a few queries in the system, but has accessed a large number of patient's records, all with diagnoses related to heart conditions.

Alex wonders whether this is a common occurrence. Perhaps many doctors conduct these types of searches. Alex proceeds to specify another task, this time isolating physician queries that accessed over 50

cardiac patient records over a short period of time. Still, Dr. Richards is the only physician that has conducted such a query. Alex heads off to interview Dr. Richards to continue his investigation.

In this scenario, a manual audit would have required countless hours of searching through paper files and notes and interviewing various hospital employees, with little hope of locating the actual source of the leak, if any occurred. In contrast, HDB Compliance Auditing allows a privacy officer to conduct a series of audits, in a matter of minutes, to reliably isolate potential sources of the leak. In fact, Alex could have reduced the steps noted here by formulating a more precise initial audit expression. An audit may either reveal the actions of a malicious employee or show that the hospital is not responsible for the disclosure.

In the future, Alex can initiate proactive audits to investigate the effectiveness of NetHMO's disclosure controls. HDB Compliance Auditing provides the protective benefit of deterring unlawful access and disclosure among hospital employees.

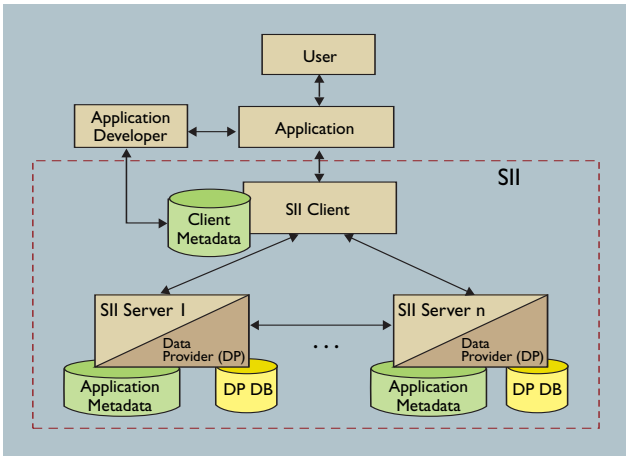

Figure 3. HDB Sovereign Information Integration (SII).

Database Watermarking. HDB Watermarking [3] is used to determine the origin of a leaked or misappropriated database. The watermarking algorithm introduces a pattern into the data that is difficult to find and is very unlikely to occur by chance. If it is difficult to find, the pattern is difficult to destroy, and therefore robust against malicious attacks. Existing watermarking techniques developed for multimedia are not effective for database tables because rows in a table are unordered; rows can be inserted, updated, deleted; and attributes can be added or dropped. In contrast, HDB's algorithm for watermarking database tables allows the watermark to be detected using only a subset of the rows and attributes of a table, is robust against updates, and is incrementally updatable.

Database watermarking deters the theft or misappropriation of sensitive data and allows a company to identify databases that have been leaked. For example, there have been many recent cases of health care workers unlawfully disclosing private patient data to drug companies, news outlets, or other third parties. In these cases, watermarking allows the health care institution to determine whether it was the source of the stolen database.

Secure Information Exchange

PITAC emphasizes the importance of secure information exchange in promoting medical research, assisting decision support, and increasing knowledge available for patient care. Finding 2 of the report underscores the need to develop technologies to help clinicians "integrate disparate data from multiple sources." Finding 9 asserts that "the ability to link patient data in an anonymous and secure fashion is critical to the national research enterprise, public health surveillance, and bio-preparedness." Sharing data between health care institutions in a private and secure manner is critical to a well-functioning e-records infrastructure. HDB's Sovereign Information Integration (SII) component meets this critical need by facilitating secure, private information sharing between autonomous entities.

SII [2] enables two or more autonomous entities to compute queries across their databases in such a way that only the results of the query are revealed. SII uses a Web Services infrastructure to apply a set of commutative encryption functions to uniquely identifiable data in different orders and at different locations. The resulting multiply-encrypted values can then be compared without compromising the privacy or security of either data set.

Unlike other data integration approaches, such as centralized data warehouses or mediator-based data federations, which reveal all data among the databases, SII does not reveal any information among the databases other than the results of the query. This allows institutions to perform a variety of joins and other operations across autonomous databases without revealing any confidential information. SII is a scalable middleware solution that can be integrated seamlessly into existing data environments without the need for a trusted third party or any anonymization of the original data. Figure 3 depicts the basic SII architecture.

As illustrated in the following medical research scenario, SII is an ideal solution to the problem of secure, privacy-preserving information sharing among health care institutions.

SII Medical Research Scenario. Wayne is a medical researcher with IPC who would like to research possible correlations between certain genetic traits and adverse reactions to IPC's cholesterol-lowering drug, Innostatin. To test these correlations, Wayne needs access to the medical records of patients who have taken Innostatin as well as the genetic information about these patients.

Wayne is aware that NetHMO and GeneBank have many common patients, many of whom have likely been prescribed Innostatin. However, privacy laws and company policies prohibit NetHMO and GeneBank from revealing personally identifiable information. Accordingly, Wayne would like to investigate the correlation between certain DNA sequences and adverse Innostatin reactions, without revealing any personally identifiable patient information among the three companies.

NetHMO, GeneBank, and IPC have installed SII to facilitate secure, privacy-preserving information sharing. To determine how many of those patients who have had adverse Innostatin reaction have a certain DNA sequence, Wayne sends an intersection query to the NetHMO's SII service via IPC's client application. NetHMO then encrypts the patient table with its own key and sends the table to GeneBank's SII service.

Next, GeneBank encrypts NetHMO's (encrypted) patient table and its own DNA table with its own key and sends both tables back to NetHMO's SII service. NetHMO then encrypts GeneBank's (encrypted) table so that both data sets are now doubly encrypted. Finally, SII joins both doubly encrypted tables and sends the number of matching results back to IPC's SII client application.

Using the information returned by the SII application, Wayne is able to isolate the particular genetic traits that correlated with an adverse reaction to Innostatin without communicating any personally identifiable information among the three companies. Thus, SII allows NetHMO and GeneBank to provide data necessary for IPC to perform valuable medical research, while maintaining the privacy of all patient records.

De-identification of Sensitive Information

The PITAC report stresses the importance of promoting anonymous and secure linking of health information to support national research, public health surveillance, and bio-preparedness (Finding 9). We present two HDB technologies that enable the de-identification of information for research and analysis. Its Privacy-Preserving Data Mining component randomizes personal data such that it can be accurately mined, but does not reveal personally identifiable information. On the other hand, its optimal *k*-anonymization component allows enterprises to de-identify personal data for publication and analysis purposes such that the patients cannot be re-identified through data linkage attacks.

Privacy-Preserving Data Mining. Analysis of large sets of patient data is necessary to perform epidemiological studies and other statistical medical research. However, HIPAA prevents health care institutions from sharing personally identifiable patient data without the consent of the patients or an Institutional Review Board (IRB). If analysis of such information is possible without disclosing any personally identifiable information, the institution may be able to derive valuable insights from patient data without obtaining patient or IRB consent.

HDB's Privacy-Preserving Data Mining (PPDM) technology adds random noise to individual values and reconstructs the distribution of the original data without revealing any personally identifiable information. Researchers can then mine the randomized aggregate data without compromising privacy. Algorithms for building classification models and discovering association rules on top of privacy-preserved data can be used on the randomized data with only small loss of accuracy [9].

Optimal k*-anonymization.* The HIPAA Privacy Rule allows health care institutions to disclose de-identified health information without restriction. HIPAA states that "health information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information." If an institution de-identifies patient data to a statistically and scientifically acceptable level, as determined by a qualified individual, it may use and disclose such data for research.

Naive approaches to data de-identification, such as removing specific personal identifiers, are prone to data linkage attacks that combine the subject data with other publicly available data to reidentify represented individuals [11]. For example, suppose a set of patient records has been "scrubbed" of any personal identifiers such as name or Social Security number. Although no record contains a single identifying value, many records are likely to contain unique value combinations. Thus, an individual who is the only male born in 1920 living in a sparsely populated area could have his age, gender, and zip code joined with another data set, such as a voter registry from the area, to obtain his name, thereby revealing his medical history.

The *k*-anonymity method in [11] was designed to avoid such linkage attacks, while preserving the integrity of the released data. In a *k*-anonymized data set, each record is indistinguishable from at least *k*-1 other records within the data set. The *k*-anonymization process involves applying operations to the input data set including data suppression and cell value generalization. Suppression involves deleting cell values or entire tuples, while generalization entails replacing specific values such as a phone number with a more general one, such as the area code. A larger value of *k* provides a higher level of privacy, since no individual can be identified with probability exceeding 1/*k* through linking attacks alone.

However, even simple restrictions of optimized *k*-anonymity are NP-hard. Therefore, we have developed a new approach [7], called Optimal *k*-anonymization, that explores the array of possible anonymizations to tame the combinatorics of the problem. The resulting algorithm finds optimizations under two representative cost measures and a wide range of *k*. It also produces useful anonymizations in circumstances where the input data or input parameters preclude finding an optimal solution in a reasonable amount of time. This process provides truthful de-identified data that is resistant to data linkage attacks.

Conclusion

Hippocratic Database technologies are well-suited to enable the transition to the 21st century electronic health records infrastructure. These technologies offer efficient methods of managing, auditing, sharing, and analyzing electronic health records that preserve the privacy of patients.

We have introduced a group of technologies that address problems raised in the PITAC report, providing the following capabilities:

1. Active enforcement of negotiated privacy policies at the database level;
2. Efficient data access tracking that identifies all who have accessed any information in the database;
3. A watermarking system that identifies leaked or misappropriated databases;
4. Information sharing across autonomous data sources that provides the results of the query without revealing any other data;
5. Accurate mining of aggregate data without compromising privacy of individual records; and
6. An optimal method of de-identification that enables useful data analysis without violating patient privacy.

We trust that these HDB concepts show that the PITAC vision is within reach. We hope they will serve as a model for future research and development of useful health care information management technologies that respect individual privacy. C

References

1. Agrawal, R., Bayardo, R., Faloutsos, C., Kiernan, J., Rantzau, R., and Srikant, R. Auditing compliance with a Hippocratic database. In *Proceedings of the 30th International Conference on Very Large Databases* (Toronto, Canada, Aug. 2004).
2. Agrawal, R., Evfimievski, A., and Srikant, R. Information sharing across private databases. In *Proceedings of the ACM SIGMOD Conference on Management of Data* (San Diego, CA, June 2003).
3. Agrawal, R., Haas, P., and Kiernan, J. Watermarking relational data: Framework, algorithms and analysis. *VLDB Journal,* 2003.
4. Agrawal, R., Kiernan, J., Srikant, R., and Xu, Y. Hippocratic databases. In *Proceedings of the 28th International Conference on Very Large Databases* (Hong Kong, Aug. 2002).
5. Agrawal, R., Kiernan, J., Srikant, R., and Xu, Y. An XPath-based preference language for P3P. In *Proceedings of the 12th International World Wide Web Conference* (Budapest, Hungary, May 2003).
6. Agrawal, R., Kiernan, J., Srikant, R., and Xu. Y. Order-preserving encryption for numeric data. In *Proceedings of the ACM SIGMOD Conference on Management of Data* (Paris, France, June 2004).
7. Bayardo, R. and Agrawal, R. Data privacy through Optimal k-Anonymization. In *Proceedings of the 21st International Conference on Data Engineering* (Tokyo, Japan, Apr. 2005).
8. Cranor, L., Langheinrich, M., Manchiori, M., Presler-Marshall, M., and Reagle, J. Platform for privacy preferences 1.0 (P3P1.0) specification. W3C Recommendation (Apr. 2002).
9. Evfimievski, A. Randomization in privacy-preserving data mining. *SIGKDD Explorations: Newsletter of the ACM Special Interest Group on Knowledge Discovery and Data Mining 4*, 2 (Dec. 2002), 43-48.
10. President's Information Technology Advisory Committee. *Revolutionizing Health Care Through Information Technology.* Report to the President of the United States. June 2004.
11. Samarati, P., and Sweeney, L. Generalizing data to provide anonymity when disclosing information. In *Proceedings of the 17th ACM SIGMOD-SIGACT-SIGART Symposium on the Principles of Database Systems 188* (1998).
12. Snodgrass, R., Yao, S., Collberg, and C. Tamper detection in audit logs. In *Proceedings of the 30th International Conference on Very Large Databases* (Toronto, Canada, Aug. 2004).

Rakesh Agrawal (ragrawal@us.ibm.com) is a Technical Fellow at Microsoft Search Labs in Mountain View, CA. All work described in this article was done while he was at the IBM Almaden Research Center, San Jose, CA.

Tyrone Grandison (tyroneg@us.ibm.com) is the leader of the Intelligent Information Systems Group at the Almaden Research Center, San Jose, CA.

Christopher Johnson (johnsocm@us.ibm.com) is an attorney and a researcher in the Intelligent Information Systems Group at the Almaden Research Center, San Jose, CA.

Jerry Kiernan (kiernan@us.ibm.com) is a senior software engineer in the Intelligent Information Systems Group at the Almaden Research Center, San Jose, CA.