Hippocratic Databases

Presenter: Tyrone Grandison

Team: Rakesh Agrawal, Jerry Kiernan, Ameet Kini, Kristen LeFevre, Ramakrishnan Srikant, Amy Wang, Yirong Xu, Diana Zhou
Our Motivation

- New regulations requiring companies to protect personal information
- Privacy is a major concern for On-demand businesses
- Lack of technology for efficient privacy enforcement and data handling
- Loss of revenue
- Dilution of brand image
- Audit failure
Basics

- **Founding tenet**
  Database systems that take responsibility for the privacy of data they manage, while not impeding the flow of information

- **Principles**
  - **Collection Group:**
    - Purpose Specification, Consent, Limited Collection
  - **Use Group:**
    - Limited Use, Limited Disclosure, Limited Retention, Accuracy
  - **Security & Openness Group:**
    - Safety, Openness, Compliance

- **Driven by current privacy legislation**
Privacy Enablers: The Triad

Policy Creation

User Data Collection

Query Enforcement
Architecture: Policy

Privacy Policy

Converts privacy policy into privacy metadata tables.

Privacy Metadata Creator

For each purpose & piece of information (attribute):
- External recipients
- Retention period
- Authorized users

Limited Disclosure
Limited Retention

Privacy Metadata

Store
Architecture: Data Collection

Data Collection
- Privacy Constraint Validator
- Data Accuracy Analyzer
- Audit Info

Data cleansing, e.g., catch typos in address.

Purpose Specification

Accuracy

Record Access Control

Privacy Metadata
Audit Trail

Store
2. Query tagged “treatment” cannot see credit card info.

3. Query tagged “research” only sees records that include “research” in set of purposes.

1. Financial person cannot issue query tagged “treatment”.

Architecture: Queries

- Safety
- Limited Use
- Privacy Metadata
- Record Access Control
- Attribute Access Control
- Store
Architecture: Queries

- Payments query that asks for all credit card numbers.
- Compliance
  - Training data for query intrusion detector
- Audit Info
- Attribute Access Control
- Query Intrusion Detector
- Record Access Control
- Privacy Metadata
- Audit Trail
- Store
- Compliance
- Safety
Architecture: Other

- Limited Collection: Analyze queries to identify unnecessary collection, retention & authorizations.
- Limited Retention: Delete items in accordance with privacy policy.
- Safety: Additional security for sensitive data.

Other components include:
- Data Collection Analyzer
- Data Retention Manager
- Encryption Support
- Privacy Metadata

Store
Architecture: All Together

- Data Collection Analyzer
- Data Retention Manager
- Attribute Access Control
- Query Intrusion Detector
- Audit Info
- Privacy Constraint Validator
- Data Accuracy Analyzer
- Audit Info
- Privacy Metadata Creator
- Privacy Metadata
- Record Access Control
- Encryption Support
- Audit Trail
- Store
Hippocratic Databases

**NetCare Healthcare Business Scenario**

- John Cane, Chief Privacy Officer, NetCare Healthcare
- Jane Smith, New Patient, NetCare Healthcare
- Dr. Young, Physician, NetCare Healthcare
- Christine Jones, Lab Technician, NetCare Healthcare
- Phil Crew, Drug Researcher, Innovative Drug Research
Hippocratic Databases
NetCare Healthcare Business Scenario

John Cane, CPO installs corporate privacy policy

Jane, a new patient, defines her privacy preferences

Jane visits NetCare’s website to setup patient account

Jane submits her personal information

- Name, Address, SSN#, Email
- Opt-in to sharing data for research
- Opt-out of sharing full medical records to lab technicians

Installation

Negotiation

Corporate Policy

Jane’s Data (Personal/Medical Records)

DATABASE
Conclusion

- Minimal modification of existing applications
- More efficient than competing privacy solutions
- Increase customer trust and business opportunities
- Help mitigate legal risks
Thank You
Founding Tenets
Collection Group

1. Purpose Specification
   - For personal information stored in the database, the purposes for which the information has been collected shall be associated with that information.

2. Consent
   - The purposes associated with personal information shall have consent of the donor (person whose information is being stored).

3. Limited Collection
   - The information collected shall be limited to the minimum necessary for accomplishing the specified purposes.
4. **Limited Use**
   - The database shall run only those queries that are consistent with the purposes for which the information has been collected.

5. **Limited Disclosure**
   - Personal information shall not be communicated outside the database for purposes other than those for which there is consent from the donor of the information.
6. **Limited Retention**
   - Personal information shall be retained only as long as necessary for the fulfillment of the purposes for which it has been collected.

7. **Accuracy**
   - Personal information stored in the database shall be accurate and up-to-date.
Security & Openness Group

8. Safety
   - Personal information shall be protected by security safeguards against theft and other misappropriations.

9. Openness
   - A donor shall be able to access all information about the donor stored in the database.

10. Compliance
    - A donor shall be able to verify compliance with the above principles. Similarly, the database shall be able to address a challenge concerning compliance.
The Triad
Privacy Policy

- Who can access which types of data for what purposes, e.g.
  - Allows a physician to access patients' names and disease records for treatment purpose
  - Allows a public-affair person to disclose anonymous disease records for research purposes. However patients can opt-out this disclosure.
User Data Collection

- How does the user want personal information to be used? e.g.
  - Allow insurance companies to access my medical data
  - Restrict disclosure of personally identifiable information

- Does the corporate privacy policy conflict with my personal preferences?
Query Enforcement

- Automatically enforce the privacy policy
  - Shred the policy into metadata
  - Analyze queries with respect to the policy, and either
    - Allow the query to run as-is, or
    - Return a subset of the records/cells that reflects individual persons’ opt-in or opt-out preferences
    - Block the query if it is in violation of the policy
Query Enforcement through Rewriting
Query Rewrite for Privacy Enforcement

Consider a Simple Example…

<table>
<thead>
<tr>
<th>ID</th>
<th>NAME</th>
<th>PHONE</th>
<th>SALARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alice</td>
<td>111-1111</td>
<td>10,000</td>
</tr>
<tr>
<td>2</td>
<td>Bob</td>
<td>222-2222</td>
<td>20,000</td>
</tr>
<tr>
<td>3</td>
<td>Carl</td>
<td>333-3333</td>
<td>30,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>PhoneChoice</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

For a certain data accessor/purpose, Name is allowed under the privacy policy, Salary is prohibited, and Phone is allowed on an opt-in basis.
Query Rewrite for Privacy Enforcement

Original Query:
SELECT Name, Phone, Salary
FROM Patient

Rewritten Query:
SELECT Name, Phone, Salary
FROM ( SELECT Name,
        CASE WHEN EXISTS (SELECT 1 FROM Choices
                         WHERE Choices.PhoneChoice = 1
                         AND Choices.ID = Patient.ID)
              THEN Patient.Phone
              ELSE null
          END AS Phone,
        CASE WHEN (0 = 1)
              THEN Patient.Salary
              ELSE null
          END AS Salary
     )
FROM Patient
WHERE ((Name is not null OR Phone is not null) OR Salary is not null) )
Query Rewrite for Privacy Enforcement

Results of query…

<table>
<thead>
<tr>
<th>NAME</th>
<th>PHONE</th>
<th>SALARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Bob</td>
<td>222-2222</td>
<td>-</td>
</tr>
<tr>
<td>Carl</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- Forbidden values covered by null values in resulting tables
- Entirely null rows filtered from the result set
Performance Tests
Scenario 1

Table Size: 1 million, no index
Scenario 2

Table Size: 10 million, no index
Scenario 3

App Selectivity = .01

![Bar chart showing query execution time vs choice selectivity for original and rewritten queries. The chart indicates that rewritten queries perform better in terms of execution time across different levels of selectivity.]
Scenario 4

App Selectivity = .1
Scenario 5

App Selectivity = 1.0
Sources of Information
References

- R. Agrawal, J. Kiernan, R. Srikant, Y. Xu. *Hippocratic Databases*. 28th Int’l Conf. on Very Large Databases (VLDB), Hong Kong, August 2002.
- R. Agrawal, J. Kiernan. *Watermarking Relational Databases*. 28th Int’l Conf. on Very Large Databases (VLDB), Hong Kong, August 2002.