Special Topics in Security and Privacy of Medical Information

Sujata Garera

Hippocratic Databases

- Privacy preservation inherent in the database
  - “And about whatever I may see or hear in treatment or even without treatment, in the life of human beings - things that should not ever be blurted out outside- I will remain silent, holding such things to be unutterable” - Hippocratic Oath

Privacy guidelines

- US Privacy Act of 1974 set out comprehensive regime limiting the collection, use and dissemination of personal information held by Federal agencies.
  - Collection limitation
  - Data quality
  - Purpose specification
  - Use limitation
  - Security safeguards
  - Openness
  - Individual participation and accountability
Hippocratic Databases: Principles

- **Purpose Specification**
  - For personal information stored in the database, the purpose shall be associated
- **Consent**
  - The purposes associated will have consent of the donor of the personal information
- **Limited Collection**
  - The personal information collected should be limited to minimum necessary

---

Hippocratic Databases: Principles

- **Limited Use**
  - The database shall run only those queries that are consistent with the information being collected
- **Limited Disclosure**
  - The personal information stored in the database shall not be communicated outside the database unless consent received

---

Hippocratic Databases: Principles

- **Limited Retention**
  - Information will be retained only as long as necessary
- **Accuracy**
  - Personal information stored will be accurate and up-to-date
- **Safety**
  - Personal Information will be protected by security safeguards against theft
Hippocratic Databases: Principles

- Openness
  - A donor shall be able to access all information about donor
- Compliance
  - A donor shall be able to verify compliance with above principles

Strawman Design: Use case

- Online bookseller
- Needs minimum information to complete transaction
  - Name, shipping address and credit card
  - Email address
- Gives recommendations
- Publishes book popularity
**Design: Usecase**

- Alice is a privacy fundamentalist
  - Doesn’t want any information about purchase retained
- Bob is a privacy pragmatist
  - Convenience of providing information only once
  - Likes recommendations
  - Does not want his transactions used for purchase circles

**Strawman Design: Architecture**
Privacy Metadata

![Diagram of Privacy Metadata]

**Figure 2: Privacy Metadata Schema**

<table>
<thead>
<tr>
<th>Table</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>privacy-policies</td>
<td>purpose, table, attribute, { external-recipients }, retention</td>
</tr>
<tr>
<td>privacy-authorizations</td>
<td>purpose, table, attribute, { authorized-users }</td>
</tr>
</tbody>
</table>

**Figure 3: Database Schema**

<table>
<thead>
<tr>
<th>Table</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>customer</td>
<td>purpose, customer-id, name, shipping-address, email, credit-card-info</td>
</tr>
<tr>
<td>order</td>
<td>purpose, customer-id, transaction-id, book-info, status</td>
</tr>
</tbody>
</table>

**Figure 4: Privacy Policies Table**

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Table</th>
<th>Attribute</th>
<th>External Recipients</th>
<th>Retention</th>
</tr>
</thead>
<tbody>
<tr>
<td>purchase</td>
<td>customer</td>
<td>name</td>
<td>{ delivery-company, credit-card-company }</td>
<td>1 month</td>
</tr>
<tr>
<td>purchase</td>
<td>customer</td>
<td>shipping-address</td>
<td>{ delivery-company }</td>
<td>1 month</td>
</tr>
<tr>
<td>purchase</td>
<td>customer</td>
<td>email</td>
<td></td>
<td>1 month</td>
</tr>
<tr>
<td>purchase</td>
<td>customer</td>
<td>credit-card-info</td>
<td>{ credit-card-company }</td>
<td>1 month</td>
</tr>
<tr>
<td>registration</td>
<td>customer</td>
<td>name</td>
<td></td>
<td>3 years</td>
</tr>
<tr>
<td>registration</td>
<td>customer</td>
<td>shipping-address</td>
<td></td>
<td>3 years</td>
</tr>
<tr>
<td>registration</td>
<td>customer</td>
<td>email</td>
<td></td>
<td>3 years</td>
</tr>
<tr>
<td>payment</td>
<td>order</td>
<td>back-info</td>
<td></td>
<td>10 years</td>
</tr>
<tr>
<td>payment</td>
<td>order</td>
<td>shipping-address</td>
<td>{ aggregated-all }</td>
<td>1 year</td>
</tr>
<tr>
<td>payment</td>
<td>order</td>
<td>back-info</td>
<td>{ aggregated-all }</td>
<td>1 year</td>
</tr>
</tbody>
</table>
Privacy Metadata

- Assumes that purpose and set of attributes completely determines the set of recipients and retention period

- Assumption that set of attributes collected for a purpose is fixed

Data Collection
Data Collection

- Data Insertion
  - After privacy policy checks, data transmitted from user and stored in tables
  - Recall that each table has attribute purpose
    - Alice’s records would have a single purpose, purchase
    - Bob’s records would have three purposes, purchase, registration and recommendations

Data Collection

- Data Preprocessing
  - Data Accuracy Analyzer
    - Run checking and/or cleansing functions before or after data insertion
    - Alice’s address may be checked against database of street addresses to catch typos in the address

Queries
Queries

- During query execution
  - Record Access Control ensures that only records whose purpose attribute includes the query purpose will be visible to that query
  - E.g. Queries tagged recommendations will be able to see Bob’s set of books but not Alice’s since Alice’s purpose attribute only lists purchase

- After query execution
  - E.g. Mallory decides to steal email addresses instead of credit card information
  - Customer service regularly accesses email addresses
  - Neither Attribute Access Control nor the Record Access Control will be able to stop the query
  - Query intrusion detector is run before on the query results to spot queries with different access patterns
  - Mallory’s queries would be marked as highly suspicious

<table>
<thead>
<tr>
<th>purpose</th>
<th>table</th>
<th>attribute</th>
<th>authorized-users</th>
</tr>
</thead>
<tbody>
<tr>
<td>purchase</td>
<td>customer</td>
<td>customer-id</td>
<td>all</td>
</tr>
<tr>
<td>purchase</td>
<td>customer</td>
<td>name</td>
<td>shipping, charge, customer-service</td>
</tr>
<tr>
<td>purchase</td>
<td>customer</td>
<td>email</td>
<td>shipping</td>
</tr>
<tr>
<td>purchase</td>
<td>customer</td>
<td>shipping-address</td>
<td>shipping, customer-service</td>
</tr>
<tr>
<td>purchase</td>
<td>customer</td>
<td>credit-card-info</td>
<td>charge</td>
</tr>
<tr>
<td>purchase</td>
<td>order</td>
<td>customer-id</td>
<td>all</td>
</tr>
<tr>
<td>purchase</td>
<td>order</td>
<td>transaction-id</td>
<td>all</td>
</tr>
<tr>
<td>purchase</td>
<td>order</td>
<td>book-info</td>
<td>shipping</td>
</tr>
<tr>
<td>purchase</td>
<td>order</td>
<td>store</td>
<td>shipping, customer-service</td>
</tr>
<tr>
<td>registration</td>
<td>customer</td>
<td>customer-id</td>
<td>all</td>
</tr>
<tr>
<td>registration</td>
<td>customer</td>
<td>name</td>
<td>registration</td>
</tr>
<tr>
<td>registration</td>
<td>customer</td>
<td>shipping-address</td>
<td>registration</td>
</tr>
<tr>
<td>registration</td>
<td>customer</td>
<td>email</td>
<td>registration</td>
</tr>
<tr>
<td>recommendations</td>
<td>order</td>
<td>customer-id</td>
<td>mining</td>
</tr>
<tr>
<td>recommendations</td>
<td>order</td>
<td>transaction-id</td>
<td>mining</td>
</tr>
<tr>
<td>recommendations</td>
<td>order</td>
<td>book-info</td>
<td>mining</td>
</tr>
<tr>
<td>purchase-circles</td>
<td>customer</td>
<td>customer-id</td>
<td>stop</td>
</tr>
<tr>
<td>purchase-circles</td>
<td>customer</td>
<td>shipping-address</td>
<td>stop</td>
</tr>
<tr>
<td>purchase-circles</td>
<td>order</td>
<td>customer-id</td>
<td>stop</td>
</tr>
<tr>
<td>purchase-circles</td>
<td>order</td>
<td>book-info</td>
<td>stop</td>
</tr>
</tbody>
</table>

Figure 8: Purchase Authorizations Table
Data retention

- Delete items that have outlived their purpose
  - Alice’s information in order table will be deleted after a month while Bob’s will be kept for a longer period
  - Recall Bob’s purposes include both purchase and recommendations

Challenges

- Efficiency
  - Can current database systems afford the additional cost of privacy checking during a record fetch?

- Limited Collection
  - Query access only data values to fulfill its purpose and database store minimal information necessary to fulfill all the purposes

Challenges

- Limited Retention
  - How do you completely delete a record from logs and past checkpoints without affecting recovery?

- Safety
  - Encrypting database on disk protects it from adversarial access but hinders usability
  - How do you index encrypted data? How do you run queries against such data?
Suggested Reading

- Hippocratic Databases
  - By Agarwal et al.