Privacy Protection via Monitoring and Audit: Computer Science + Healthcare + Law

Anupam Datta
Carnegie Mellon University
Desiderata: Respect privacy expectations in the transfer and use of personal information within and across organizational boundaries.
A Problem of Growing Importance

- Increased privacy legislation in the US and Europe
  - FERPA (educational institutions), HIPAA and HITECH (health care providers), GLBA (financial institutions), data breach notification laws

- Increased digitization implies higher volumes of inappropriate disclosures and uses

- Increased lawsuits and fines
  - ChoicePoint 2005 ($26M), TJX 2005 ($256M), DVA 2009 ($20M), CVS 2009 ($2.25M), Rite Aid 2010 ($1M)

- Increased public awareness
  - CDT, EPIC, Markle Foundation, Patient Privacy Rights
Research Goal

Develop methods and tools to help organizations be compliant with privacy regulations and internal policies
Approach
Representing Complex Privacy Laws

Challenges

- Identifying core privacy concepts in long, dense legal text
  - HIPAA has 84 operational clauses about disclosures of protected health information (~30 pages)
- Understanding how individual clauses should be combined
  - permitting clauses, denying clauses, cross-references, exceptions
Main Result

1. PrivacyLFP, a first-order logic (language) for representing privacy laws
2. First complete logical formalization of all disclosure-related clauses in the HIPAA Privacy Rule and the Gramm-Leach-Bliley Act
A covered entity may disclose an individual’s protected health information (PHI) to law-enforcement officials for the purpose of identifying an individual if the individual made a statement admitting participating in a violent crime that the covered entity believes may have caused serious physical harm to the victim.

### Basic concepts in privacy laws

- **Actions:** \(\text{send}(p_1, p_2, m)\)
- **Roles:** \(\text{inrole}(p_2, \text{law-enforcement})\)
- **Data attributes:** \(\text{attr$_\text{in}$(prescription, phi)}\)
- **Purposes:** \(\text{purp$_\text{in}$(u, id-criminal)}\)
- **Beliefs:** \(\text{believes-crime-caused-serious-harm}(p, q, m)\)
A covered entity may disclose an individual’s protected health information (phi) to law-enforcement officials for the purpose of identifying an individual if the individual made a statement admitting participating in a violent crime that the covered entity believes may have caused serious physical harm to the victim.

**Basic concepts in privacy laws**

- **Actions:** `send(p1, p2, m)`
- **Roles:** `inrole(p2, law-enforcement)`
- **Data attributes:** `attr_in(prescription, phi)`
- **Purposes:** `purp_in(u, id-criminal)`
- **Beliefs:** `believes-crime-caused-serious-harm(p, q, m)`

**Temporal constraints**

- **Past provision:** `◊state(q, m)`
- **Future obligation:** `◊send(p1, p2, m)`
Example HIPAA Clause

A covered entity may disclose an individual’s protected health information (phi) to law-enforcement officials for the purpose of identifying an individual if the individual made a statement admitting participating in a violent crime that the covered entity believes may have caused serious physical harm to the victim.

\[
\forall p_1, p_2, m, u, q, t.
\ (send(p_1, p_2, m) \land \\
  \text{inrole}(p_2, \text{law-enforcement}) \land \\
  \text{tagged}(m, q, t, u) \land \\
  \text{attr}_\text{in}(t, \text{phi})) \\
\implies (\text{purp}_\text{in}(u, \text{id-criminal})) \\
\land \exists m'. \Diamond \text{state}(q, m') \land \text{is-admission-of-crime}(m') \\
\land \text{believes-crime-caused-serious-harm}(p_1, q, m')
\]
Combining Clauses

- **Two types of clauses**
  - Positive norm: disclosure permitted *if* requirement satisfied
    - “A covered entity may disclose protected health information for treatment activities […]” [HIPAA 164.506(c)(2)]
  - Negative norm: disclosure permitted *only if* requirement satisfied
    - “A covered entity must obtain authorization for any use or disclosure of psychotherapy notes.” [HIPAA 164.508(a)(2)]

- A disclosure is permitted if it satisfies at least one positive norm and *all* the negative norms

\[
m\text{aysend}(p_1, p_2, m) \triangleq \left( \bigvee_i \varphi_i^+ \right) \land \left( \bigwedge_j \varphi_j^- \right)
\]
Structure of HIPAA and GLBA

- **HIPAA Privacy Rule**
  - Deny all transmissions not explicitly allowed
  - 56 positive norms, 7 negative norms, 19 exceptions
  - Formalization in logic: 94 pages with explanation

- **GLBA**
  - Allow all transmissions not explicitly denied
  - 5 negative norms and 10 exceptions
  - Formalization in logic: 12 pages with explanation

- Important property of formalization
  - **Traceability**: Each clause in law corresponds to one norm or exception in formalization (roughly)
Approach

Privacy Law

Organizational audit log

Detect policy violations

Computer-readable privacy policy

Monitoring and Audit
Main Challenge in Enforcing Privacy Laws

- Incompleteness of logs makes fully automated enforcement impossible
  - Subjective (stores only objective events)
  - Future (stores only past and current events)
  - Spatial (logs may be distributed)
Reduce Algorithm

- Define an iterative algorithm \((\text{reduce} (\mathcal{L}, \varphi) = \varphi')\)
  - Output a policy that cannot be checked on the current log
  - Minimize human effort
    - Check as much of the policy as possible
Reduce Algorithm

\[ \text{Reduce}(L_1, \phi_1) = \phi_2 \]
\[ L_2 > L_1 \quad \text{Reduce}(L_2, \phi_2) = \phi_3 \]
\[ \ldots \]
\[ L_{n+1} > L_n \quad \text{Reduce}(L_n, \phi_n) = \phi_{n+1} \]

If \( \phi_1 \) only contains bounded future obligations, then eventually

- \( \phi_{n+1} \equiv T \) (policy is satisfied); or
- \( \phi_{n+1} \equiv \bot \) (policy is violated); or
- \( \phi_{n+1} \) contains only subjective predicates (needs human audit)
Example

\[ \varphi = \forall p_1, p_2, m, u, q, t. \]
\[ (\text{send}(p_1, p_2, m) \land \text{tagged}(m, q, t, u) \land \text{attr}\_\text{in}(t, \phi)) \]
\[ \supset \text{inrole}(p_2, \text{law-enforcement}) \land \text{purp}\_\text{in}(u, \text{id-criminal}) \land \exists m'. (\varphi \text{state}(q, m') \land \text{is-admission-of-crime}(m') \land \text{believes-crime-caused-serious-harm}(p_1, m')) \]

\[ \varphi' = T \]
\[ \land \text{purp}\_\text{in}(\text{id-bank-robber}, \text{id-criminal}) \land \text{is-admission-of-crime}(M1) \land \text{believes-crime-caused-serious-harm}(\text{UPMC, M1}) \]

Log

Jan 1, 2011
state(Bob, M1)

Jan 5, 2011
send(UPMC, allegeny-police, M2)
tagged(M2, Bob, date-of-treatment, id-bank-robber)
Formal Properties

- **Termination**
- **Correctness**
  - If \( \text{Reduce}(L_1, \varphi_1) = \varphi_2 \), then \( \varphi_1 \) and \( \varphi_2 \) enforce the same policies on extensions of \( L_1 \)
- **Minimality**
  - If \( \text{Reduce}(L_1, \varphi_1) = \varphi_2 \), then \( L_1 \) does not have sufficient information to determine truth values of atomic predicates in \( \varphi_2 \)
Minimality

\[ \varphi = \]
\[ \forall p_1, p_2, m, u, q, t. \]
\[ (\text{send}(p_1, p_2, m) \wedge \]
\[ \text{tagged}(m, q, t, u) \wedge \]
\[ \text{attr_in}(t, \phi)) \]
\[ \supset \text{inrole}(p_2, \text{law-enforcement}) \wedge \]
\[ \text{purp_in}(u, \text{id-criminal}) \]
\[ \wedge \exists m'. (\varphi \text{ state}(q, m') \]
\[ \wedge \text{is-admission-of-crime}(m') \]
\[ \wedge \text{believes-crime-caused-serious-harm}(p_1, m')))\]

Log

Jan 1, 2011
\text{state}(Bob, M1)

Jan 5, 2011
\text{send}(UPMC, allegeny-police, M2)
\text{tagged}(M2, Bob, date-of-treatment, \text{id-bank-robber})

\[ \varphi' = T \]
\[ \wedge \text{purp_in}(\text{id-bank-robber, id-criminal}) \]
\[ \wedge \text{is-admission-of-crime}(M1) \]
\[ \wedge \text{believes-crime-caused-serious-harm}(UPMC, M1) \]
Reduce can automatically check 80% of all the atomic predicates

<table>
<thead>
<tr>
<th>Degree of automation</th>
<th># of clauses</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>17</td>
</tr>
<tr>
<td>80% – 99%</td>
<td>24</td>
</tr>
<tr>
<td>50% – 79%</td>
<td>29</td>
</tr>
<tr>
<td>1% – 50%</td>
<td>8</td>
</tr>
<tr>
<td>0%</td>
<td>6</td>
</tr>
</tbody>
</table>
Remaining Challenge

- Human auditor can only check a subset of subjective predicates due to budgetary constraints

- Question: How should auditor allocate the audit budget?
## Risk Management Model (by example)

<table>
<thead>
<tr>
<th>Audit log records all accesses (100)</th>
<th>Accesses divided into types</th>
<th>Loss from each violation (internal, external detection)</th>
<th>Cost of each inspection</th>
</tr>
</thead>
<tbody>
<tr>
<td>(5)</td>
<td></td>
<td>$ 500, 1000</td>
<td>$ 100</td>
</tr>
<tr>
<td>(95)</td>
<td></td>
<td>$ 250, 500</td>
<td>$ 100</td>
</tr>
</tbody>
</table>

Total audit budget = $2000, i.e., can inspect at most 20 accesses

How many accesses of each type to inspect?
Allocating Audit Budget

Total audit budget = $2000

Accesses divided into types

<table>
<thead>
<tr>
<th>Type</th>
<th>Budget Allocation</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$500</td>
</tr>
<tr>
<td>95</td>
<td>$400</td>
</tr>
<tr>
<td></td>
<td>$300</td>
</tr>
<tr>
<td></td>
<td>$200</td>
</tr>
<tr>
<td></td>
<td>$100</td>
</tr>
<tr>
<td></td>
<td>$0</td>
</tr>
</tbody>
</table>

1/6 1/6 1/6 1/6 1/6 1/6

Example: All possible allocations are equally likely
Observed Outcome

Accesses divided into types

<table>
<thead>
<tr>
<th>Allocated Budget</th>
<th>Observed Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>$300</td>
<td>$2000</td>
</tr>
<tr>
<td>$1700</td>
<td>$1000</td>
</tr>
</tbody>
</table>

Higher loss from celebrity access violations
Updating Audit Budget

Accesses divided into types

<table>
<thead>
<tr>
<th></th>
<th>500</th>
<th>400</th>
<th>300</th>
<th>200</th>
<th>100</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1500</td>
<td>$1600</td>
<td>$1700</td>
<td>$1800</td>
<td>$1900</td>
<td>$2000</td>
<td></td>
</tr>
<tr>
<td>2/6</td>
<td>2/6</td>
<td>1/6</td>
<td>1/12</td>
<td>1/24</td>
<td>1/24</td>
<td></td>
</tr>
</tbody>
</table>

New Budget Allocation

Observed loss used to update probabilities of allocations
Regret Minimizing Audits

- Learns from experience to recommend budget allocation for audit in each audit cycle
- Budget allocation is provably close to optimal fixed budget allocation

- Technical approach: New regret minimization algorithm for repeated games of imperfect information
  (Online learning-theoretic technique)
Take-away messages

1. Privacy laws represented in computer-readable language (logic)
   - Complete formalization of HIPAA and GLBA

2. Automatic monitoring of audit logs
   - Applies to significant part of HIPAA, GLBA
   - Outputs residual policy involving subjective predicates

3. Learning algorithm guides human audit of subjective predicates in a manner that minimizes risk (regret)
Approach

1. Privacy Law
2. Organizational audit log
   - Computer-readable privacy policy

Monitoring and Audit

Detect policy violations
Bibliography


Thanks!
Questions?