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

Anupam Datta

Carnegie Mellon University

Personal Information Governance

Patient medical bills

Patient information

Hospital

Insurance Company

Drug Company

Advertising

Patient

Complex Process within a Hospital

Patient

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

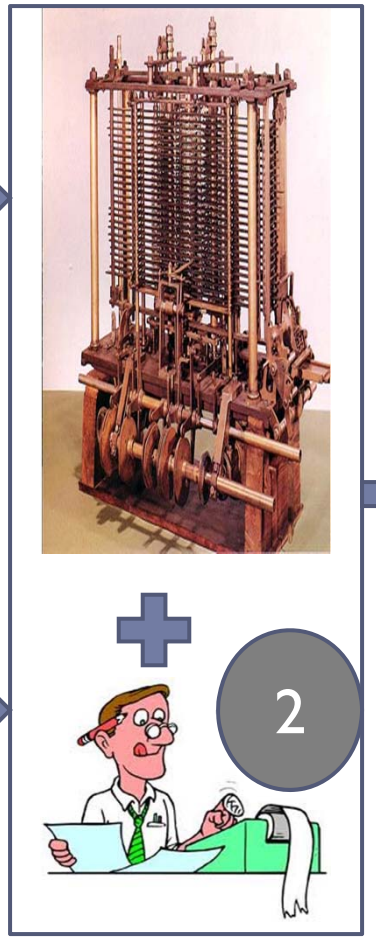
Privacy Law

Organizational audit log

1

$$\mathbf{G}\,\Big(\forall p_1, p_2, m.\ \mathsf{send}(p_1, p_2, m) \supset (\forall d, u, q, t.\ (m = \mathrm{info}(d, u)) \wedge \mathsf{contains}(m, q, t) \supset (\bigvee_i \varphi_i^+) \wedge (\bigwedge_j \varphi_j^-)) \wedge (\forall t.\ (m = \mathrm{req\_for\_access}(p_1, t)) \supset \varphi^-_{\mathtt{164.524b2i'}} \wedge \varphi^-_{\mathtt{164.524b2ii'}})\Big)$$

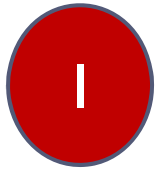
Computer-readable privacy policy



Monitoring and Audit

Detect policy violations

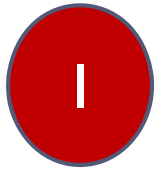
Representing Complex Privacy Laws



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

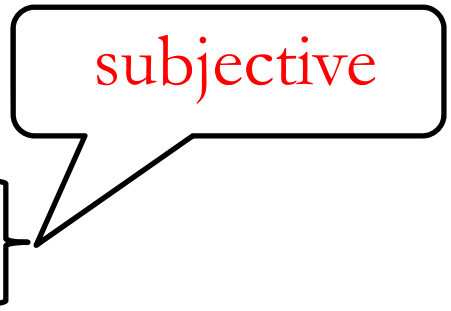


1. PrivacyLFP, a first-order logic (language) for representing privacy laws
2. First complete logical formalization of all disclosure-related clauses in the HIPAA Privacy Rule and the Gramm-Leach-Bliley Act

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

I

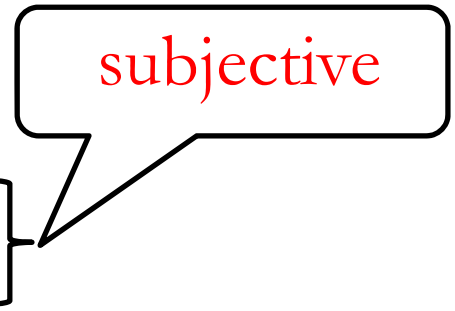
- **Basic concepts in privacy laws**
  - Actions: send(p1, p2, m)
  - Roles: inrole(p2, law-enforcement)
  - Data attributes: attr_in(prescription, phi)
  - Purposes: purp_in(u, id-criminal))
  - Beliefs: believes-crime-caused-serious-harm(p, q, m)



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

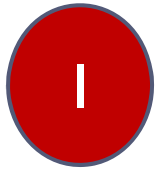
I

- **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)

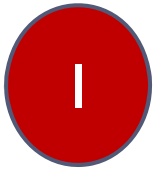
Example HIPAA Clause



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

$$
\begin{aligned}
&\forall p1, p2, m, u, q, t.\\
&\quad (\text{send}(p1, p2, m) \wedge\\
&\quad\ \text{inrole}(p2, \text{law-enforcement}) \wedge\\
&\quad\ \text{tagged}(m, q, t, u) \wedge\\
&\quad\ \text{attr\_in}(t, \text{phi}))\\
&\quad\ \supset (\text{purp\_in}(u, \text{id-criminal}))\\
&\qquad\ \wedge \exists\, m'.\, \Diamond\text{-state}(q,m') \wedge \text{is-admission-of-crime}(m')\\
&\qquad\ \wedge \text{believes-crime-caused-serious-harm}(p1, q, m')
\end{aligned}
$$

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

$$maysend(p_1, p_2, m) \triangleq \left(\bigvee_i \varphi_i^+\right) \wedge \left(\bigwedge_j \varphi_j^-\right)$$

Structure of HIPAA and GLBA

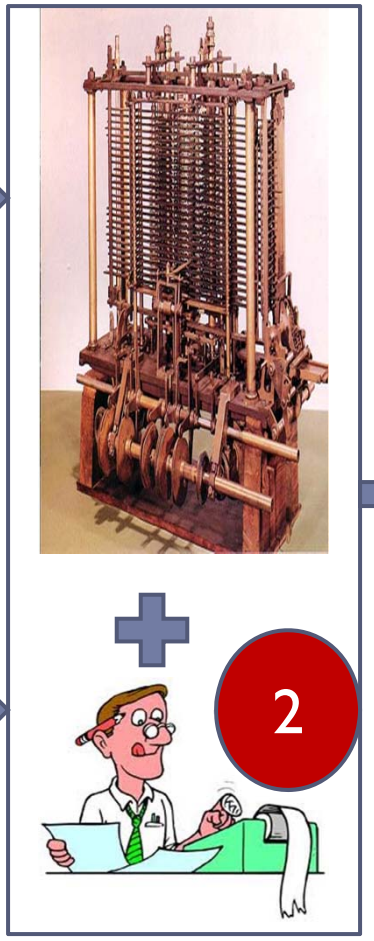
- HIPAA Privacy Rule
  - Deny all transmissions not explicitly allowed
  - 56 positive norms, 7 negative norms, 19 exceptions
  - Formalization in logic: 94 pages with explanation
- GLBA
  - Allow all transmissions not explicitly denied
  - 5 negative norms and 10 exceptions
  - Formalization in logic: 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

1

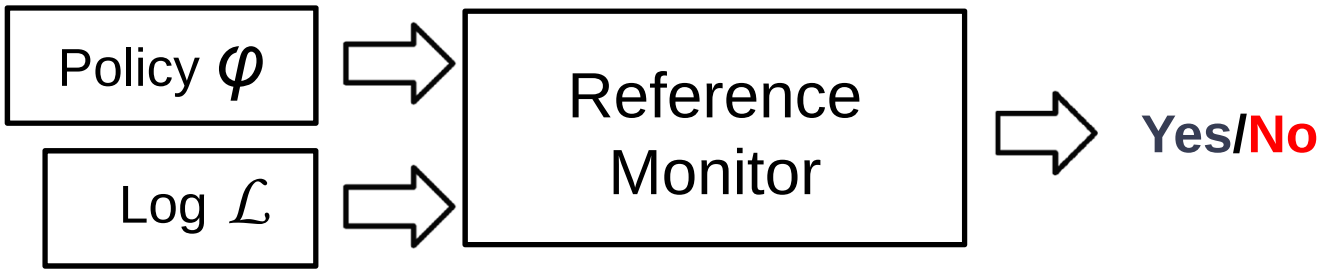
$$\mathbf{G}\,\Big(\forall p_1, p_2, m.\ \mathsf{send}(p_1, p_2, m) \supset (\forall d, u, q, t.\ (m = \mathrm{info}(d, u)) \wedge \mathsf{contains}(m, q, t) \supset (\bigvee_i \varphi_i^+) \wedge (\bigwedge_j \varphi_j^-)) \wedge (\forall t.\ (m = \mathrm{req\_for\_access}(p_1, t)) \supset \varphi^-_{164.524b2i'} \wedge \varphi^-_{164.524b2ii'})\Big)$$

Computer-readable privacy policy

Monitoring and Audit

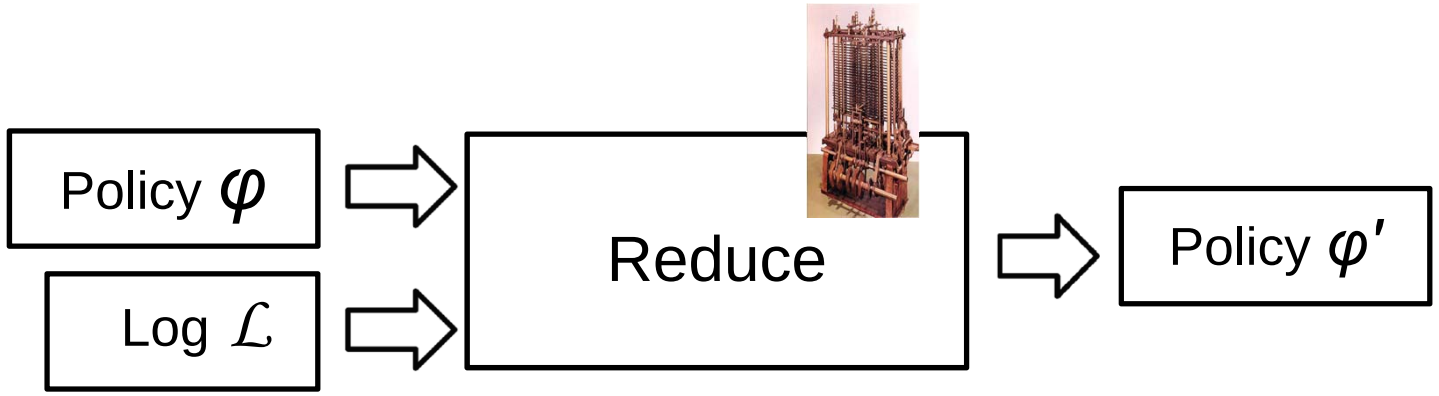
Main Challenge in Enforcing Privacy Laws

- 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 (reduce ($\mathcal{L}$, φ) = φ')
  - Output a policy that cannot be checked on the current log
  - Minimize human effort
    - Check as much of the policy as possible



Reduce Algorithm

2

$$\text{Reduce}(\mathcal{L}_1, \varphi_1) = \varphi_2$$

$$\mathcal{L}_2 > \mathcal{L}_1 \quad \text{Reduce}(\mathcal{L}_2, \varphi_2) = \varphi_3$$

$$\ldots$$

$$\mathcal{L}_{n+1} > \mathcal{L}_n \quad \text{Reduce}(\mathcal{L}_n, \varphi_n) = \varphi_{n+1}$$

If φ_1 only contains bounded future obligations, then eventually

- $\varphi_{n+1} \equiv \top$ (policy is satisfied); or
- $\varphi_{n+1} \equiv \bot$ (policy is violated); or
- φ_{n+1} contains only subjective predicates (needs human audit)

Example

2

φ =

∀p1, p2, m, u, q, t.

(send(p1, p2, m) ∧
tagged(m, q, t, u) ∧
attr_in(t, phi))

⊃ inrole(p2, law-enforcement) ∧

purp_in(u, id-criminal)

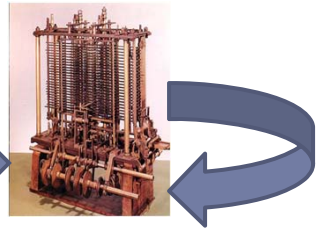
∧∃ m'.(◊state(q, m')

∧is-admission-of-crime(m')

∧believes-crime-caused-serious-harm(p1, m'))

{ p1→ UPMC,
p2→ allegeny-police,
m → M2,
q → Bob,
u → id-bank-robber,
t → date-of-treatment }

{ m' → 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)

φ' = T

∧purp_in(id-bank-robber, id-criminal)

∧is-admission-of-crime(M1)

∧believes-crime-caused-serious-harm(UPMC, M1)

Formal Properties

- Termination
- Correctness
  - If Reduce($\mathcal{L}_1$, φ_1) = φ_2 , then φ_1 and φ_2 enforce the same policies on extensions of $\mathcal{L}_1$
- Minimality
  - If Reduce($\mathcal{L}_1$, φ_1) = φ_2 , then $\mathcal{L}_1$ does not have sufficient information to determine truth values of atomic predicates in φ_2

Minimality

2

φ =

∀p1, p2 , m, u, q, t.
(send(p1, p2, m) ∧
tagged(m, q, t, u) ∧
attr_in(t, phi))
⊃ inrole(p2 , law-enforcement) ∧
purp_in(u, id-criminal)
∧∃ m'.(♢state(q,m')
∧is-admission-of-crime(m')
∧believes-crime-caused-serious-harm(p1, m'))

Log

Jan 1, 2011
state(Bob, M1)

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

φ' = T
∧purp_in(id-bank-robber, id-criminal)
∧is-admission-of-crime(M1)
∧believes-crime-caused-serious-harm(UPMC, M1)

HIPAA Case Study

- Reduce can automatically check 80% of all the atomic predicates

Degree of automation	# of clauses
100%	17
80% – 99%	24
50% – 79%	29
1% – 50%	8
0%	6

Remaining Challenge

2

φ' = purpose(u, treatment)

Was patient record accessed for treatment?

- 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)

Audit log records all accesses (100)

Accesses divided into types

Loss from each violation (internal, external detection)

Cost of each inspection

(5)

$ 500, 1000

$ 100

(95)

$ 250, 500

$ 100

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

(5)

(95)

Initial Budget Allocation

$500	$400	$300	$200	$100	$0
$1500	$1600	$1700	$1800	$1900	$2000
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

(5)

(95)

Allocated Budget	Observed Loss
$300	$2000
$1700	$1000

Higher loss from celebrity access violations

Updating Audit Budget

Total audit budget = $2000

Accesses divided into types

(5)

(95)

New Budget Allocation

$500	$400	$300	$200	$100	$0
$1500	$1600	$1700	$1800	$1900	$2000
2/6	2/6	1/6	1/12	1/24	1/24

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

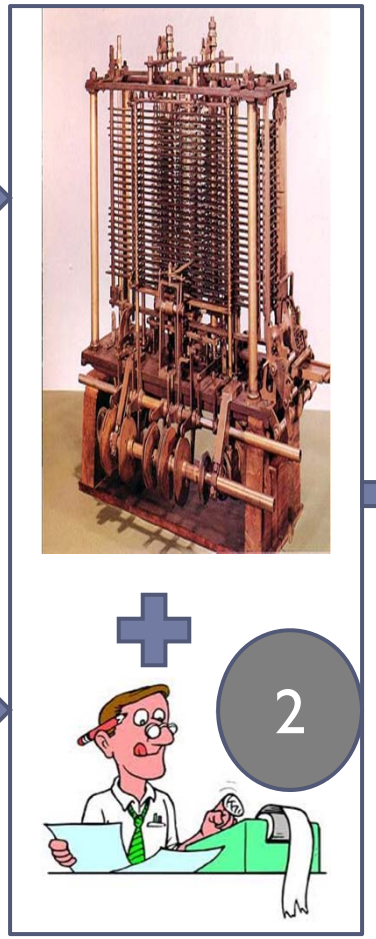
Privacy Law

Organizational audit log

1

$$\mathbf{G}\,\Big(\forall p_1, p_2, m.\ \mathsf{send}(p_1, p_2, m) \supset (\forall d, u, q, t.\ (m = \mathrm{info}(d, u)) \wedge \mathsf{contains}(m, q, t) \supset (\bigvee_i \varphi_i^+) \wedge (\bigwedge_j \varphi_j^-)) \wedge (\forall t.\ (m = \mathrm{req\_for\_access}(p_1, t)) \supset \varphi^-_{\mathtt{164.524b2i'}} \wedge \varphi^-_{\mathtt{164.524b2ii'}})\Big)$$

Computer-readable privacy policy



2

Monitoring and Audit

Detect policy violations

Bibliography

1. H. DeYoung, D. Garg, L. Jia, D. Kaynar, A. Datta, Experiences in the Logical Specification of the HIPAA and GLBA Privacy Laws, in *Proceedings of 9th ACM Workshop on Privacy in the Electronic Society*, October 2010.

2. D. Garg, L. Jia, A. Datta, A Logical Method for Policy Enforcement over Evolving Audit Logs, Technical Report arXiv:1102.2521, February 2011.

3. J. Blocki, N. Christin, A. Datta, A. Sinha, Regret Minimizing Audits: A Learning-Theoretic Basis for Privacy Protection, Technical Report CMU-CyLab-11-003, February 2011

Thanks!
Questions?