"Control of Personal Information: A Dialogue Between a Technologist and a Lawyer"

Radcliffe Inst. and Harvard Div. of E&AS
Symposium on Security and Privacy
April 23, 2004

The Technologist:
Professor Joan Feigenbaum
Yale Computer Science Department
http://www.cs.yale.edu/homes/jf

The Lawyer:
Professor Peter Swire
Moritz College of Law
The Ohio State University
http://www.peterswire.net

Overview 1

- Defining “control of personal information” (Peter)
- The power and limitations of technology (Joan)
- The power and limitations of law (Peter)
- Combining the two approaches (Joan)

Overview 2

What can and should be achieved by

- Software systems
  - Encryption
  - (DRM-like) Permissions systems
- Hardware-supported “trusted systems”
- Sector-specific regimes such as HIPAA
- Broader legal regimes such as FIPs

I. Defining "Control of Personal Information"

- Some meanings of "privacy" not primarily addressed today:
  - Roe v. Wade and right to privacy in bodily autonomy
  - Intellectual property rights such as right of publicity (they can't use your face in ads)
  - Rules for search warrants and other compelled access to data that is held in private

Control of Personal Information

- Focus on data protection for personally identifiable information
- Today's task includes control over:
  - Information in transit
  - Information in storage
  - Often held by (partially trusted) "third parties"
- Less focus on spam and other intrusions

Some examples of data protection

- You send an e-mail to a friend.
  - Can the ISPs and others read it?
- You see a doctor.
  - Who else sees data? Nurse, insurer, employer
- You buy software on-line.
  - What do advertisers learn about you? Your credit-card company? Can the vendor track your usage of the software?

II. The Power and Limitations of Technology

Current state of the art:

+ We have the ability (if not always the will) to prevent *improper access* to information.

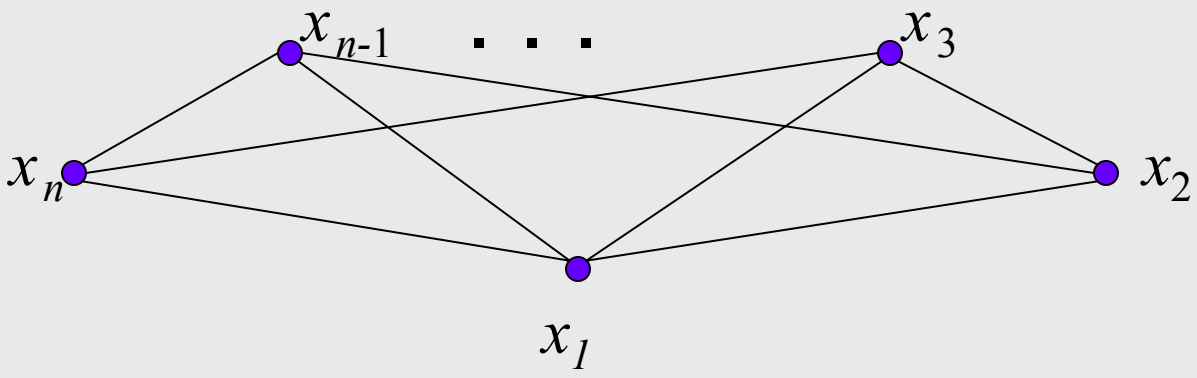
– We have little or no ability to prevent *improper use* of information by *parties authorized to access it*.

Performing Tasks on a "Need to Know" Information Basis

- Mundane tasks, *e.g.*, storage or transmission of information: Use encryption!
- More exotic tasks, *e.g.*
  - Privacy-preserving data mining
  - Privacy-preserving surveillance

  Computing exactly one fact about a distributed data set is precisely what cryptographic-protocol theory enables *in principle*!

Example 1: Secure, Multiparty Function Evaluation



$$y = F(x_1, \ldots, x_n)$$

- Each i learns y.
- No i can learn anything about x_j (except what he can infer from x_i and y).
- Very general positive results. Not very efficient.

Some Recent Progress on Special-Purpose SMFE Protocols

- Lindell and Pinkas: Efficient 2-party protocol for ID3 data mining on $x_1 \cup x_2$
- Aggarwal, Mishra, and Pinkas: Efficient n-party protocol for order statistics of $x_1 \cup \cdots \cup x_n$
- Freedman, Nissim, and Pinkas: Efficient 2-party protocol for $x_1 \cap x_2$

Example 2: Protection of Digital IDs

- General Problem: People use the same uid/pwd at many sites.
- Example: Same uid/pwd at eBay and at a high-school alumni site
- Threat: A break-in at a low-security site reveals many uid/pwd pairs that can be used at high-security sites.

Some Recent Progress on ID Protection

http://crypto.stanford.edu/WebSecPwd/

Blake Ross, Dan Boneh, John Mitchell

Browser plug-in that converts the user's pwd to a unique, *site-specific* pwd.

Basic Algorithm

- Locate all pwd HTML elements on page:
  `<INPUT TYPE=password NAME=pass>`
- When form is submitted, replace contents of pwd field with

  $$\mathbf{HMAC}_{\mathbf{pwd}}\mathbf{(domain\text{-}name).}$$

- Send pwd *hash* to site instead of pwd.

Features

- Conceptually *simple* solution!
- Implementation includes:
  - pwd-reset page
  - remote-hashing site (used in, *e.g.*, cafés)
  - list of domains for which domain of reset page is not domain of use page (*e.g.*, MS Passport)
- Dictionary attacks on hashes are much less effective than those on pwds and can be thwarted *globally* with a high-entropy plug-in pwd.

PORTIA

Privacy, Obligations, and Rights in Technologies of Information Assessment

Large-ITR, five-year, multi-institutional, multi-disciplinary, multi-modal research project on end-to-end handling of sensitive information in a wired world

http://crypto.stanford.edu/portia/

Core Technical Problem: The Unreasonable Effectiveness of Programmability

- Many machine-readable permissions systems
  - Rights-management languages
  - Privacy policies
  - Software licenses
- None is now technologically enforceable.
  - All software-only permissions systems can be circumvented.
  - Once data are transferred, control is lost.

Will "Trusted Systems" Help?

- Hardware-based, cryptographic support for proofs that a data recipient's machine is running a particular software stack
- Potential problems:
  - Technical: *Very* hard to build
  - Business: Adoption hurdles
  - Philosophy: Privacy, fair use, MS hatred, *etc*.
- Potential benefits:
  - Copyright enforcement? Maybe
  - Privacy enforcement? *Much* harder!

Dan Geer at YLS: “DRM ≡ Privacy”

I don’t think so. Circumvention is *not* the worst threat to privacy. Instead, information leaks because:

- Data objects are small.
- Privacy regimes (*e.g.,* HIPAA) are hard to automate. Complementary software systems are lacking, and relevant people are poorly trained.
- Lots of leakage is *permitted* by these regimes!

III. The Power and Limitations of Law

- Outline:
  - Technology as necessary but not sufficient
  - Data protection and Fair Information Practices
  - Toward “partially trusted systems” such as medical records under HIPAA

Technology as a Necessary Condition

- Control over data only possible, even on average, if have good tech protections
- Can have legal policy of "no sharing"
  - What if every script kiddie can get it?
  - Would you do $1 million transfers of funds in that environment?
  - Need technological protections against the malicious third party
  - Need technological *support* for well intentioned parties

Technology is Not a Sufficient Condition

- Joan's discussion:
  - Can prevent *improper access* to information
  - Can't prevent *improper use by parties authorized to access it*
- The role of law:
  - Laws today often prohibit transfers to 3rd parties, working with tech solutions.
  - Laws and institutions will be important to limiting improper uses by authorized users.

Data Protection and Fair Information Practices

- Standard legal approach of FIPs
  - Notice
  - Choice (limits on secondary uses)
  - Access and Correction
  - Security
  - Accountability
- If implemented, significant “control over personal information”

Fair Information Practices

- "Security" recognized as a necessary condition for privacy
- "Choice" and how to handle improper uses by those who can see the data
- "Accountability" and the law
  - Go to jail, pay fines, lose your job
  - Laws state and help establish norms.
  - Publicity and the press

HIPAA as a "Partially Trusted" System

- Reasons to share medical data
  - Treatment -- the nurse, second doctor, etc.
  - Payment -- insurance verifies that you received service
  - Research -- scan records for patterns to improve treatment; don't double-count
  - Public health -- new case of contagious disease
  - Many others: oversight, litigation on malpractice, anthrax and alert Homeland Security; etc.

HIPAA (medical records)

- Reasons not to share medical data
  - Democracy: people say they want medical privacy
  - Discrimination: employer or neighbor treats you differently
  - Encourage useful treatment: substance abuse, mental health, HIV need confidentiality
  - In short, we want confidence in the system.

Basic HIPAA approach

- Fair information practices
  - Notice, choice, access, security, accountability
  - Chief privacy officer and administrative oversight
  - Free sharing for treatment, payment, and other listed purposes
- For security
  - “Minimum necessary” use and sharing
  - Access controls often used to enforce that; let the nurse see data but not the janitor.

Technology & Law in HIPAA

- Now is time of big shift from paper to electronic medical records.
- HIPAA creates law, reg, norms, and institutions to protect privacy.
- Access controls and other technical speed bumps to sharing; not strict “trusted system”
- Diversity of settings, with 14% of GDP in health care; so level of protection varies
- Overall institutional and legal structure designed to provide better privacy than if no rule.

Summary on the Law

- "Defense in depth" as a strategy for providing control over personal data
  - Technology as a necessary condition
  - Legal rules: sanctions, norms
  - Institutions: administrative oversight in the organization, by CIO, CPO, and others
  - Publicity and the press lead to accountability.
- No formal proof that privacy is protected

IV. Combining the Two Approaches

- Technology cannot determine what is (or should be) legal.
- Laws cannot determine what is technologically feasible.
- “Systems” include people and procedures as well as hardware and software.

Use Technology for What It's Good At

- Storing large amounts of data
- Transmitting large amounts of data
- Retrieving or computing *one* piece of data that's needed while hiding the rest of the data
- Encoding and applying complex but *deterministic* rules for processing data

Don’t Rely Exclusively on Technology for Things It’s Not Good At

- Deciding what personal information “should” and “should not” be used for
- Pre-empting political disagreement about how to use this information
- Encoding and applying rules that inherently involve human judgment

Institutional Support for Data Privacy

- CIOs and CPOs
- Privacy-impact assessments
  - Now required for *new* federal computer systems
  - Should facilitate technical and legal input *early in design phase*

? Strategy for legacy systems

Components of (Really) Trustworthy Systems 1

- Ownership, rights, and responsibilities in the personal-information domain
  - Laws
  - Enterprise policies
  - Public awareness and understanding
- Technological support for compliance with the rules
  - Prevention of misuse
  - Detection and auditing
  - “Warning signs” for users

Components of (Really) Trustworthy Systems 2

- People and procedures to complement technological support for compliance
- Penalties for failures to comply
  - No one can violate law with impunity.
  - Enterprises cannot violate *their own policies* with impunity.

Example: Airline Passenger-Record Transfers

Would trustworthy systems of this sort have prevented the recent problems at Jet Blue, Northwest, and American ?

- Misleading user agreements ?
- Inadequate tech support for policies and inadequate warnings about potential violations ?
- Inadequate penalties for violations ?

Components of (Partially) Trustworthy Systems

- HIPAA as example of complex data uses, for diverse set of users and systems
- Won't get (really) trustworthy system
- The rationale for the regulation is that defense in depth by (laws + technology + institutions + norms) create an overall regime that is better than alternatives.
- Improved tech improves overall system.

~~V. Conclusion~~

- It is too early to draw conclusions!
- As a society, we are still at the beginning of our attempt to gain (at least some) control over personal information.
- Questions?