PORTIA: Privacy, Obligations, and Rights in Technologies of Information Assessment

Rebecca Wright
Computer Science Department
Stevens Institute of Technology
www.cs.stevens-tech.edu/~rwright
crypto.stanford.com/portia

17 March, 2004

Erosion of Privacy

"You have zero privacy. Get over it."

- Scott McNealy, 1999

- Changes in technology are making privacy harder.
  - increased use of computers and networks
  - reduced cost for data storage
  - increased ability to process large amounts of data
- Becoming more critical as public awareness, potential misuse, and conflicting goals increase.

Abuses of Sensitive Data

- Identity theft
- Loss of employment, health coverage, personal relationships
- Unfair business advantage
- Potential aid to terrorist plots

Historical Changes

- Small towns, little movement:
  - very little privacy, social mechanisms helped prevent abuse
- Large cities, increased movement:
  - lost social mechanisms, but gained privacy through anonymity
- Now:
  - advancing technology is reducing privacy, social mechanisms not replaced.

What Can We Do?

- Use technology, policy, and education to
  - maintain/increase privacy
  - provide new social mechanisms
  - create new models for better understanding

Problem: Using old models and old modes of thought in dealing with situations arising from new technology.

What is Privacy?

- Means different things to different people
  - seclusion: the desire to be left alone
  - property: the desire to be paid for one's data
  - autonomy: the ability to act freely

Product Design as Policy Decision

- product decisions by large companies or public organizations become de facto policy decisions
- often such decisions are made without conscious thought to privacy impacts, and without public discussion
- this has been particularly true in the United States, where there is not much relevant legislation

Example: Metro Cards

Washington, DC

- no record kept of per card transactions
- damaged card can be replaced if printed value still visible

New York City

- transactions recorded by card ID
- damaged card can be replaced if card ID still readable
- have helped find suspects, corroborate alibis

The PORTIA Project

Privacy, Obligations, and Rights in Technologies of Information Assessment

A five-year multidisciplinary project focusing on the technical challenges of handling sensitive data and the policy and legal issues facing data subjects, data owners, and data users.

Funded by the National Science Foundation as a Large ITR (Information Technology Research) grant, Oct 2003 - Sept 2008.

PORTIA Personnel

- Academic investigators:
  - **Dan Boneh**, Hector Garcia-Molina, John Mitchell, Rajeev Motwani, *Stanford*
  - **Joan Feigenbaum**, Ravi Kannan, Avi Silberschatz, *Yale*
  - **Stephanie Forrest**, *University of New Mexico*
  - **Helen Nissenbaum**, *NYU*
  - **Rebecca Wright**, *Stevens Institute of Technology*

PORTIA Personnel

- Research partners
  - Jack Balkin, *Yale Law School*
  - Greg Crabb, *Secret Service*
  - Cynthia Dwork, Brian LaMacchia, *Microsoft*
  - Sam Hawala, *US Census Bureau*
  - Kevin McCurley, *IBM Research*
  - Perry Miller, *Yale Center for Medical Informatics*
  - John Morris, *Center for Democracy and Technology*
  - Benny Pinkas, *HP Labs*
  - Marc Rotenberg, *Electronic Privacy Information Center*
  - Alejandro Schaffer, *DHHS/National Institutes of Health*
  - Dan Schutzer, *Citigroup*

PORTIA Personnel

- Supported PhD students
  - Stanford:
    - Gagan Aggarwal
    - Anupam Datta
    - Eu-Jin Goh
    - Krishnaram Kenthapadi
    - Robert Ledesma
    - Ben Lynn
    - Sergio Marti
    - Nagendra Modadugu
    - Aviv Nisgav
    - Justin Rosenstein
    - Dilys Thomas
    - Bevery Yang

PORTIA Personnel

- Yale:
  - Kevin Chang
  - John Corwin
  - Hong Jiang
  - Jian Zhang
- University of New Mexico:
  - Fernando Esponda
- NYU:
  - Bilge Yesil
  - Michael Zimmer
- Stevens:
  - Geetha Jagannathan
  - Zhiqiang Yang

PORTIA Goals

- Produce a next generation of technology for handling sensitive information that is qualitatively better than the current generation's.
- Enable end-to-end handling of sensitive information over the course of its lifetime.
- Formulate an effective conceptual framework for policy making and philosophical inquiry into the rights and responsibilities of data subjects, data owners, and data users.

Major Technical Themes

- privacy-preserving data mining
- identity theft and identity privacy
- database policy enforcement tools
- managing sensitive information in P2P systems
- using trusted platforms to provide trusted privacy-preserving services
- contextual integrity

Privacy-Preserving Data Mining

Allow multiple data holders to collaborate to compute important information while protecting the privacy of other information.

- Security-related information
- Public health information
- Marketing information
- etc.

Technological tools include cryptography, data perturbation and sanitization, access control, inference control, trusted platforms.

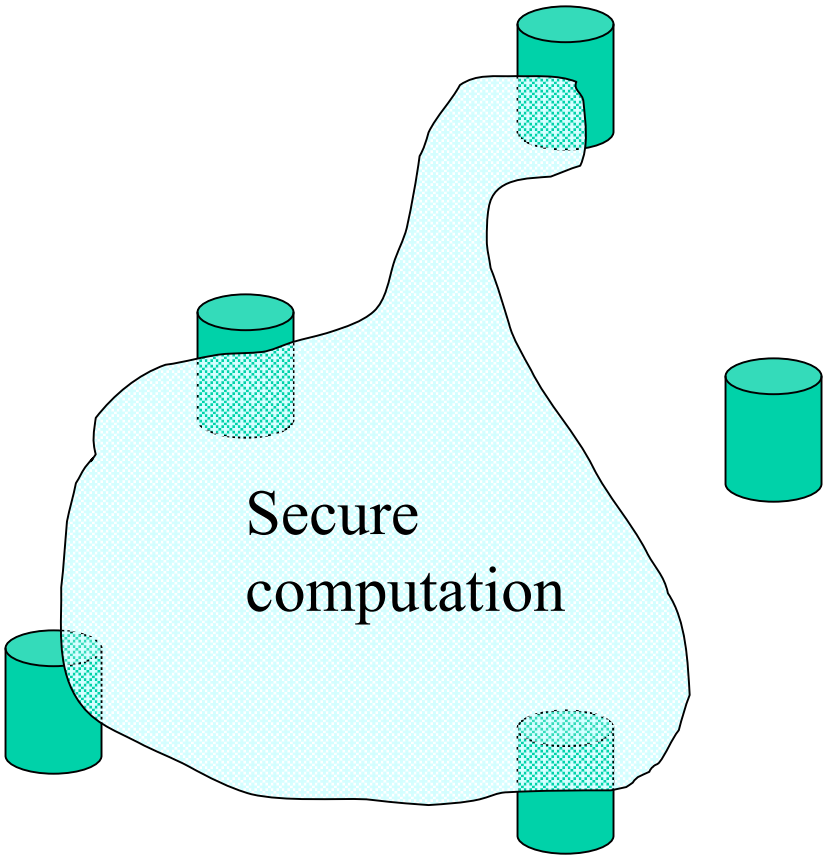
Advantages of privacy protection

- protection of personal information
- protection of proprietary or sensitive information
- enables collaboration between different data owners (since they may be more willing or able to collaborate if they need not reveal their information)
- compliance with legislative policies

Cryptography and Secure Computation

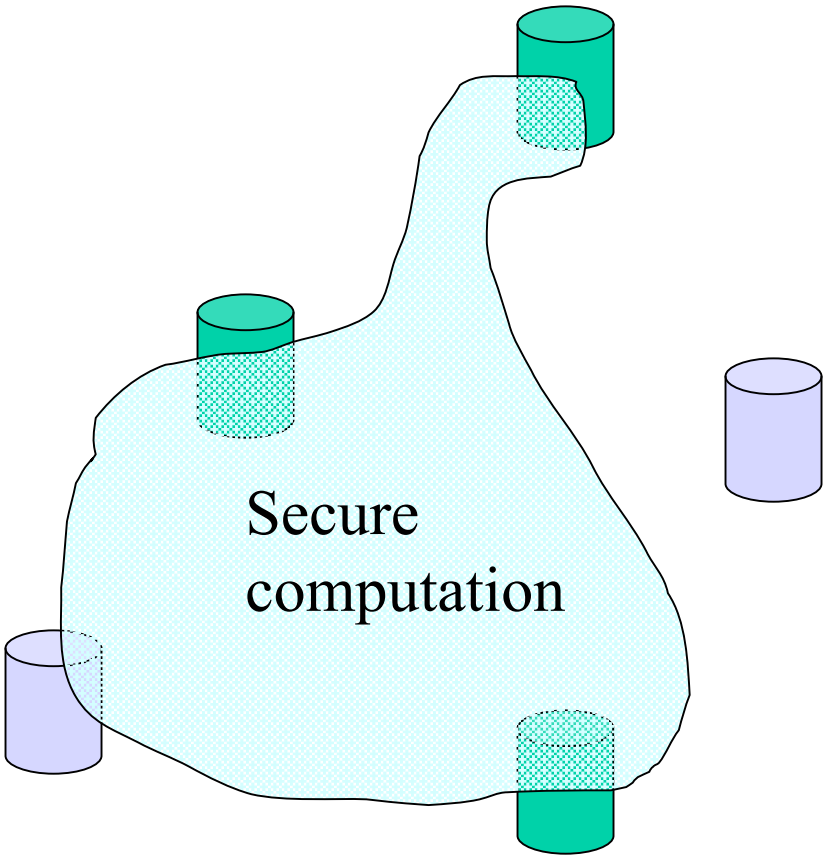
- Cryptography is a very useful tool.
- But, cryptographic secure multiparty computation definitions are both too strong and too weak for privacy-preserving data mining:
  - Too strong: do not allow leakage of innocuous information, and pay the price in efficiency.
  - Too weak: do not address leakage or misuse caused by the function itself (e.g., info implied by the outputs, misbehavior in choosing an input, poorly chosen ideal functionality).

Potential Integration



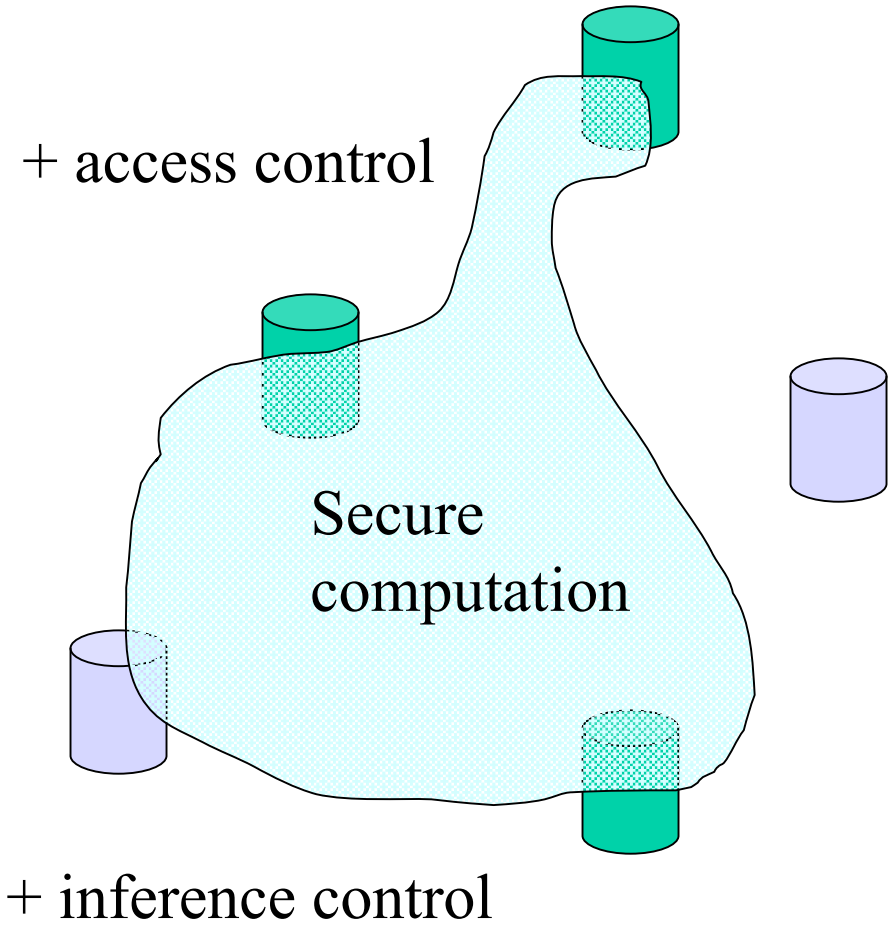
- Secure computation to protect critical data

Potential Integration



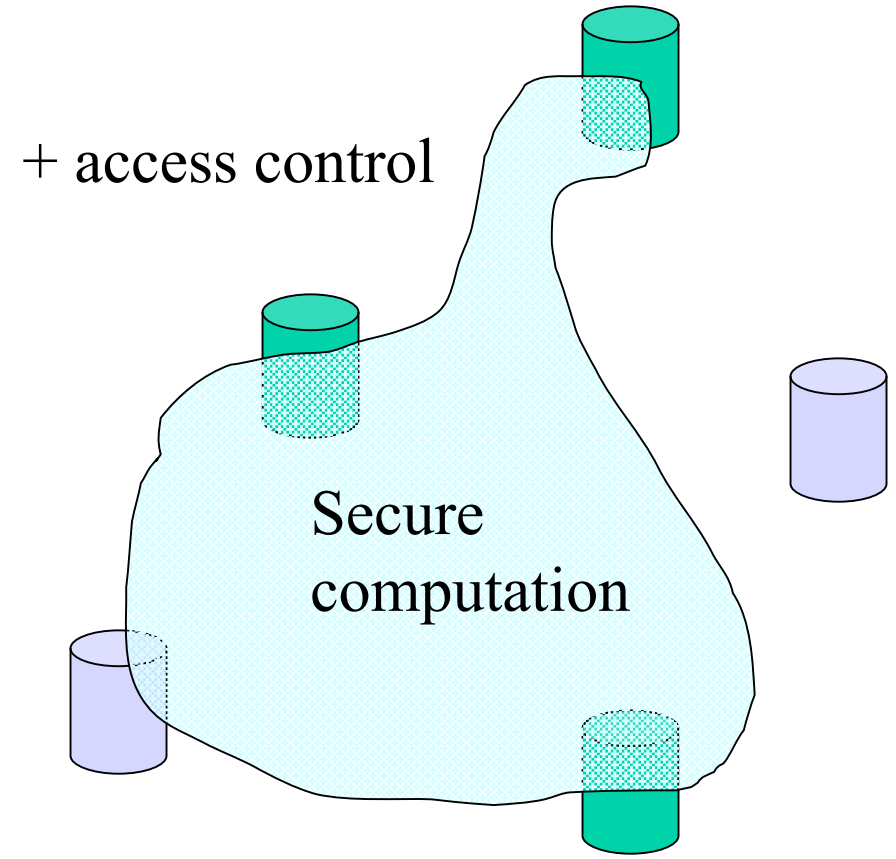
- Secure computation to protect critical data
- Perturbation or aggregation to protect possibly sensitive data
- No protection on completely innocuous data

Potential Integration



- Secure computation to protect critical data
- Perturbation or aggregation to protect possibly sensitive data
- No protection on completely innocuous data
- With policies, access control and inference control to prevent additional leakage

Potential Integration



Problems:

- How to determine which information is critical, possibly sensitive, innocuous?
- How to define appropriate policies?
- How to handle conflicting goals and desires?
- How to determine identities for access control?

Contextual Integrity

- Contextual integrity can help clarify privacy concepts: what are norms, expectations, and contractual obligations in various settings (and what *should* they be)?
- May be a helpful starting point for formalizing mathematical privacy definitions that allow for finer granularity than "output-only" crypto definitions.

Summary

- Increasing use of computers and networks has led to a proliferation of sensitive data.
- Without proper precautions, this data could be misused, misinterpreted, or mismanaged.
- The PORTIA project aims to develop a comprehensive, end-to-end technological infrastructure for handling sensitive data throughout its lifetime.