Unlinking Private Data

Alex Vaynberg
04/11/2006
Yale University

Sensitive Information in the Wired World

Privacy and Privacy Loss

- Ability to give information to certain individuals, while retaining the ability to keep that information secret from others
- Privacy loss occurs when information becomes known to those from whom it is kept secret

Aggregation = Privacy Loss

- Possible Cause:
  - One bit of data is considered private, but is public without being directly connected to an individual
  - One bit of data is not private, but gives out more information about an individual
    - allows connection with other record
  - Put together:
    - private information is known about a person

Story Time

A priest has been asked if people tell interesting stories during confessions.

He tells that his first confessor actually confessed to a murder

Later a new person comes in and greets the priest. People ask him, how does he know the priest?

He answers, “I was his first confessor”.

The Point

Seemingly nameless private data, can be combined with non-anonymous data, resulting in a privacy loss.

Linking Data to People

- Types of identification
  - Permanent
    - Uniquely identifies individual, follows him wherever
    - Examples: SSN, Passport ID, Name*
  - Semi-permanent
    - May id a real person, changing may involve a cost
    - Name*, address, telephone, credit card #
  - Transient
    - Almost no cost to change
    - Pseudonym, user id, e-mail

Linking Data to People

- Types of Data:
  - Public Data
    - Driver's license, property records
    - Kept open by government
    - Managed by applicable laws (HIPPA, ???)
  - Linked Private
    - Almost all business transactions
    - Data collected when dealing with business
    - Connected to person via (semi)permanent id
  - Unlinked Private
    - Website ids
    - No (semi)permanent id was recorded

Databases and Aggregation

- Semipermanent and permanent ids permit aggregation of data from private and public sources
- Results in digital dossiers, which many consider to be privacy concern
- Worse, these dossiers are scattered, unreliable, and frequently inaccessible by the person who they describe

Fixing The Problem

- Reduce public data to minimum
  - specifically remove associations between permanent and semipermanent Ids
- Force private data to be unlinked by creating a reliable system of certified pseudonyms
- Allow for undeletable, but commentable reports (with low privacy value) on pseudonyms that follow a real identity from pseudonym to pseudonym.

Certified Pseudonyms

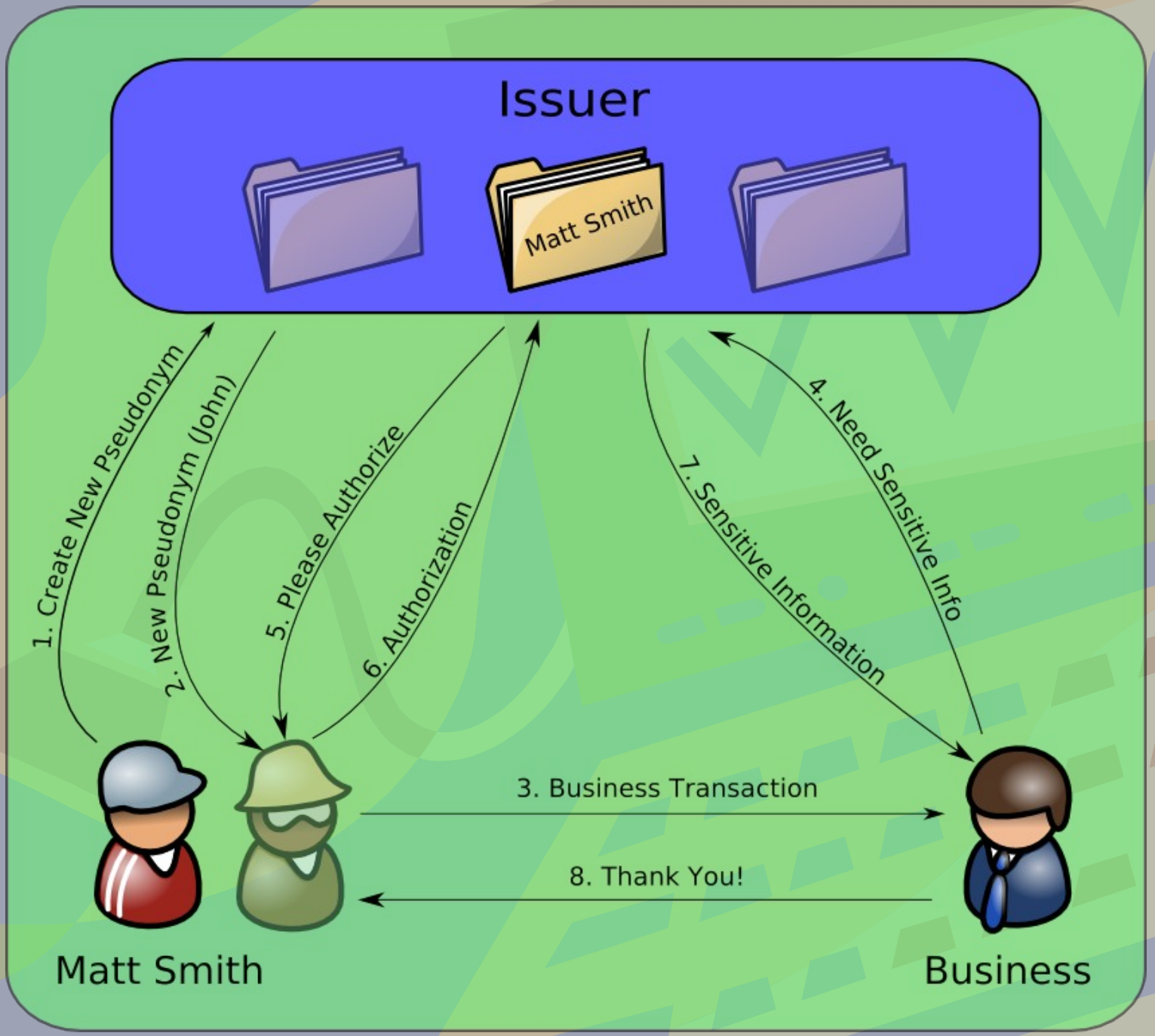
- A UID, but can be created at any time
- Comes attached with information that a person has authorized for a pseudonym
- Issued by a licensed pseudonym issuer
- No (semi)permanent Ids
  - Not linkable, except by issuer
- Issuers operate under strict legal guidelines
- Connection may be restored by courts upon necessity (lawsuit, etc.)

Ensuring Privacy

- A person creates as many identities as he wishes, selecting information that can be revealed by each one
- One of these identities will be used when dealing with another entity
- The other entity will be able to get authorized info from issuer
- Business dealing can proceed if enough information is attached to that identity
- Identity itself is completely throw-away

Pseudonym Issuers

- Private organizations
  - government will not get credit history without warrant, etc.
- Regulated by laws
  - minimum requirements / privacy guarantees
- Compete on ease of use, features, etc
  - Compare to credit card issuers
- Unify data from many pseudonyms
  - many ids, one credit history, no SSN involved
- One place to keep track / contest data


Issuer
Matt Smith
1. Create New Pseudonym
2. New Pseudonym (John)
5. Please Authorize
6. Authorization
4. Need Sensitive Info
7. Sensitive Information
3. Business Transaction
8. Thank You!
Matt Smith
Business

Advantages

- Businesses can not aggregate data
  - no (semi)permanent Ids
- Accountability preserved
- Free market / legal protections
- Anonymous guaranteed payment
  - similar to credit cards
- Ability to keep track of all personal data
- Can coexist with current system
- Allows for statistics for marketing use

Disadvantages

- Central point of failure
  - identity theft can be disastrous
- Complex management interface
- Standard protocol required for use
  - similar to credit cards
- Who will be charged, and how much?
- Inability for direct customer communication
- Semipermanent Id required for deals
  - house painting requires an address

Dealing with Difficulties

- Communication
  - Direct Communication requires semipermanent information about a person
  - Indirection needed; easy with e-mail, harder with phone and address
- Deals where semipermanent Id is required
  - Example: shipping, house painting, cable TV
  - Bad: can be aggregated with public data
  - Good: cannot be aggregated with private data
  - Similar to current method: trust

Portia Objectives

- Internet architecture
  - Protocols for sensitive information exchange
- Personal Information management
  - Gives users ability to monitor sensitive information about themselves
  - Enables placing of comments or contesting records about you

Portia Objectives (cont.)

- Enterprise Information Management
  - Ability to get reliable information about individuals
  - Unique ids that enable customer management
  - Decreased risk due to security leaks
- Cyber Rights + Responsibilities
  - Cheap pseudonimity without loss of accountability
- Use of Fair Information Principles
  - Mandated at pseudonym issuer level
  - No longer critical for every business