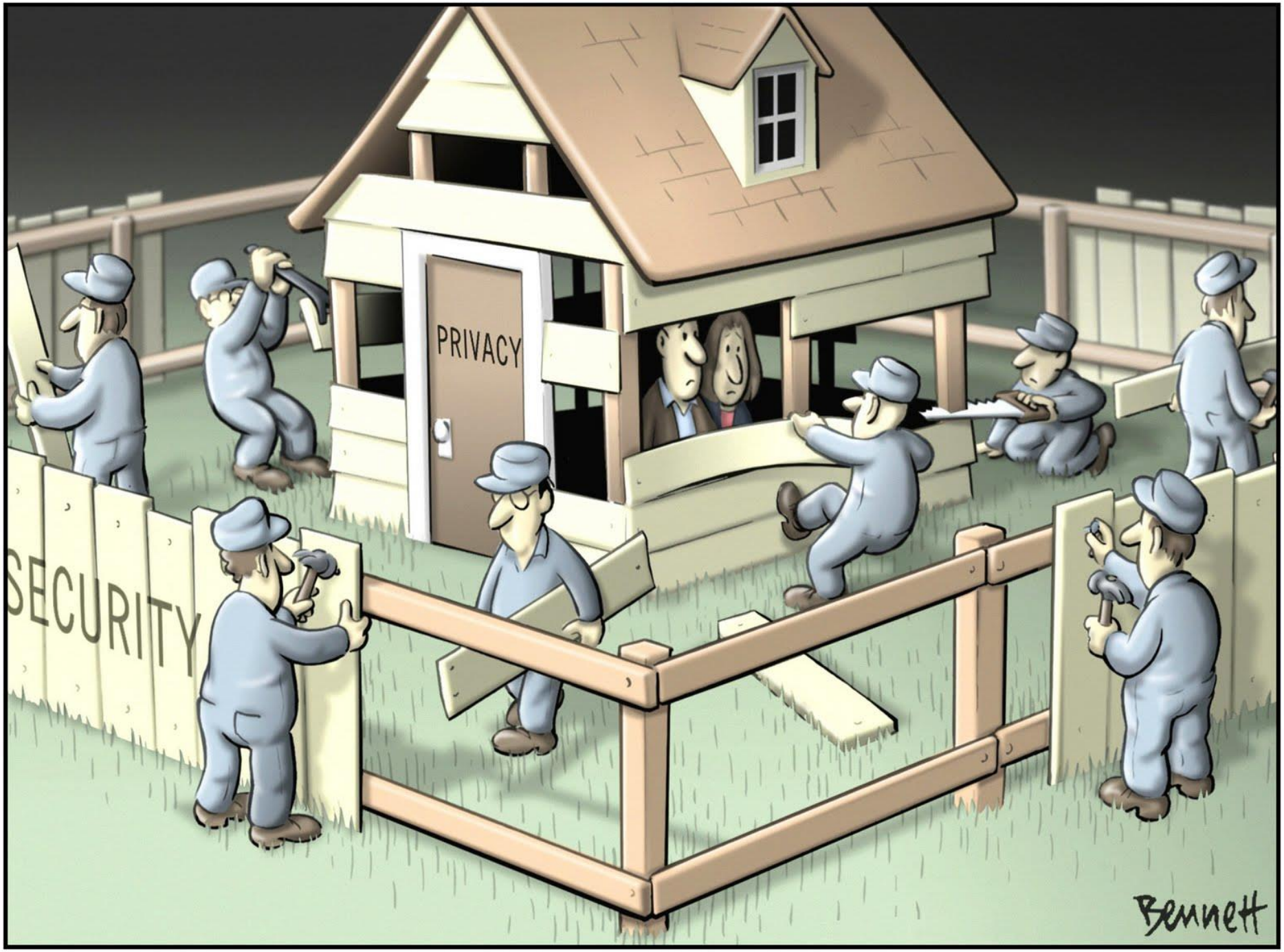
Privacy Issues in Big Data

CISC 6640 Privacy and Security in Big Data
Lecture 2a

Instructor:
Md Zakirul Alam Bhuiyan
Assistant Professor
Department of Computer and Information Sciences
Fordham University

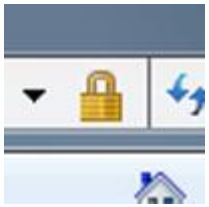
What We Are Going to Learn

- Privacy vs. Security
- Privacy Concerns
- Method to Protect Privacy Concerns
- Top Ten Big Data Security and Privacy Challenges


PRIVACY
SECURITY
Bennett

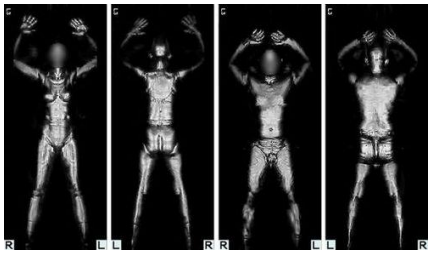
Privacy vs. Security

Privacy: what data goes where?



Security: protection against unauthorized access to data

- *Security helps* enforce privacy policies
- Can be *at odds with each other*
  - e.g., invasive screening to make us more "secure" against terrorism



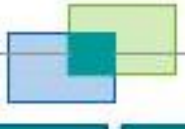
Privacy vs. Security

- **Privacy**
  - **Education Sector**
    - Student grade info is an asset whose confidentiality is considered to be highly important by students
  - **Medical Community**
    - Privacy is about a patient determining what information the doctor should release about him/her
  - **Financial community**
    - A bank customer determines what financial information the bank should release about him/her
  - **Government community**
    - FBI would collect information about US citizens. However FBI determines what information about a US citizen it can release to say the CIA

Privacy vs. Security

- **Security**
  - **Allowing access to individual's travel and spending data**
  - **Allowing access to web surfing behavior**
  - **Marketing, Sales, and Finance**
    - Allowing access to individual's purchases
  - **In case of Medical Community**
    - Security of patient information that should be available to the doctors
      - who can have access to a resource
      - under what conditions access can occur
      - what those accessing are allowing to do

Privacy vs. Security



	Information Security	Privacy
Accountability	• Focuses on tracking an individual's actions and manipulation of information	• Focuses on tracking the trail of PII disclosure
Integrity	• Protects against the corruption of data by authorized or unauthorized individuals	• Seeks to ensure that inaccurate PII is not used to make an inappropriate decision about a person
Aggregation	• Focuses on determining the sensitivity of derived and aggregated data so that appropriate access guidance can be defined	• Dictates that aggregation or derivation of new PII should not be allowed if the new information is neither authorized by law nor necessary to fulfill a stated purpose
Confidentiality	• Focuses on processes and mechanisms (e.g., authenticators) that prevent unauthorized access	• Focuses on ensuring that PII is only disclosed for a purpose consistent with the reason it was collected
Destruction	• Focuses on ensuring that information cannot be recovered once deleted	• Addresses the need for the complete elimination of collected information once it has served its purpose

Privacy-sensitive Data

- Identity
  - name, address, SSN
- Location
- Activity
  - web history, contact history, online purchases
- Health records
- Business secrete
- …and more

What We Are Going to Learn

- Privacy vs. Security
- **Privacy Concerns**
- Method to Protect Privacy Concerns
- Top Ten Big Data Security and Privacy Challenges

Privacy Concerns

Privacy Concerns

guardian.co.uk

News | World Cup | Comme

News › Technology › Rea

Series: Read me first

Facebook should compete on privacy, not hide it away

Bruce Schneier
guardian.co.uk, Wednesc
Article history

Facebook Wants You to Be Less Private - But Why?

Written by Marshall Kirkpatrick / July 1, 2009 1:56 PM / 35 Comments

« Prior Post Next Post »

Facebook he
the website's
A long list of
manageable

Websites 'keeping deleted photos'

User photographs can still be found on many social networking sites even after people have deleted them, Cambridge University researchers have said.

pular
addresses
ored - and
to find
including

Facebook says images are removed from its servers immediately.

Google Buzz Privacy Issues Have Real Life Implications

by **Robin Wauters** on Feb 12, 2010 **150 Comments** Like Buzz 134 retw

Who Knows Who Your Facebook Friends Are?

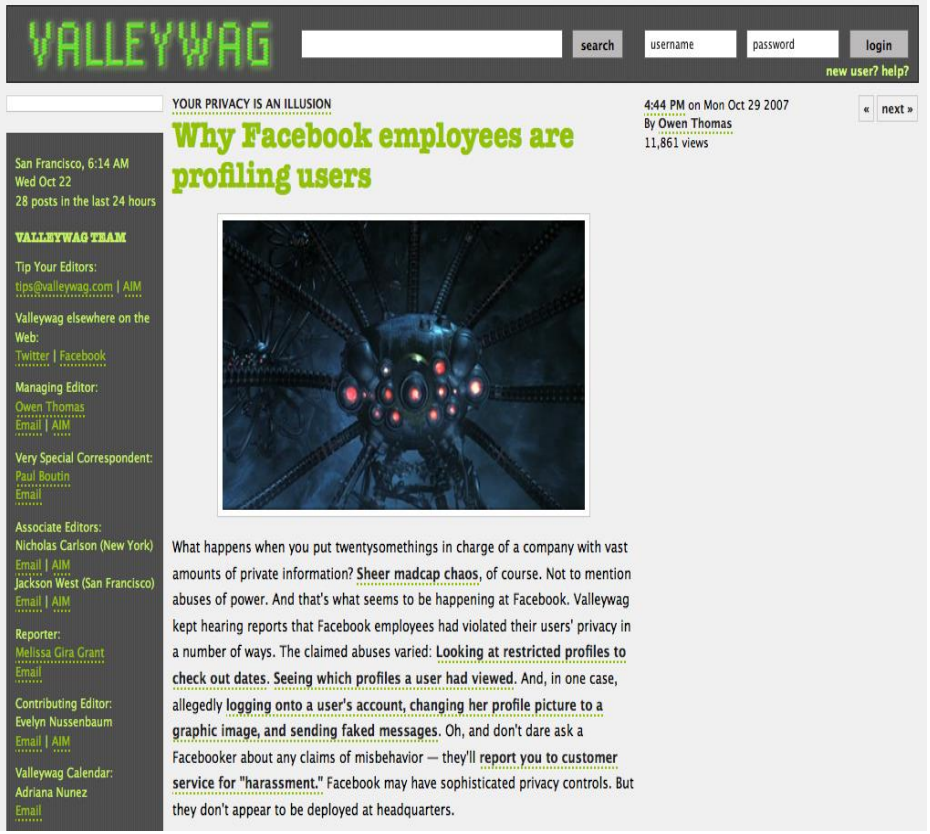
Deeplink by Tim Jones

As you may hav
overhaul was th
several change
Facebook user
a researcher wi

Facebook users unknowingly sharing personal data, warn researchers

Computer experts have warned that millions of Facebook users could inadverte
sharing personal information online because of the way the site's privacy settings work.

OSNs mishandle data



Facebook Beacon

Facebook employees abuse personal data

Big Data and The Insider Threat

Google fired engineer for privacy breach

David Barksdale, a Google engineer, was sacked earlier this year for improperly accessing the accounts of several Google users, Google confirms.

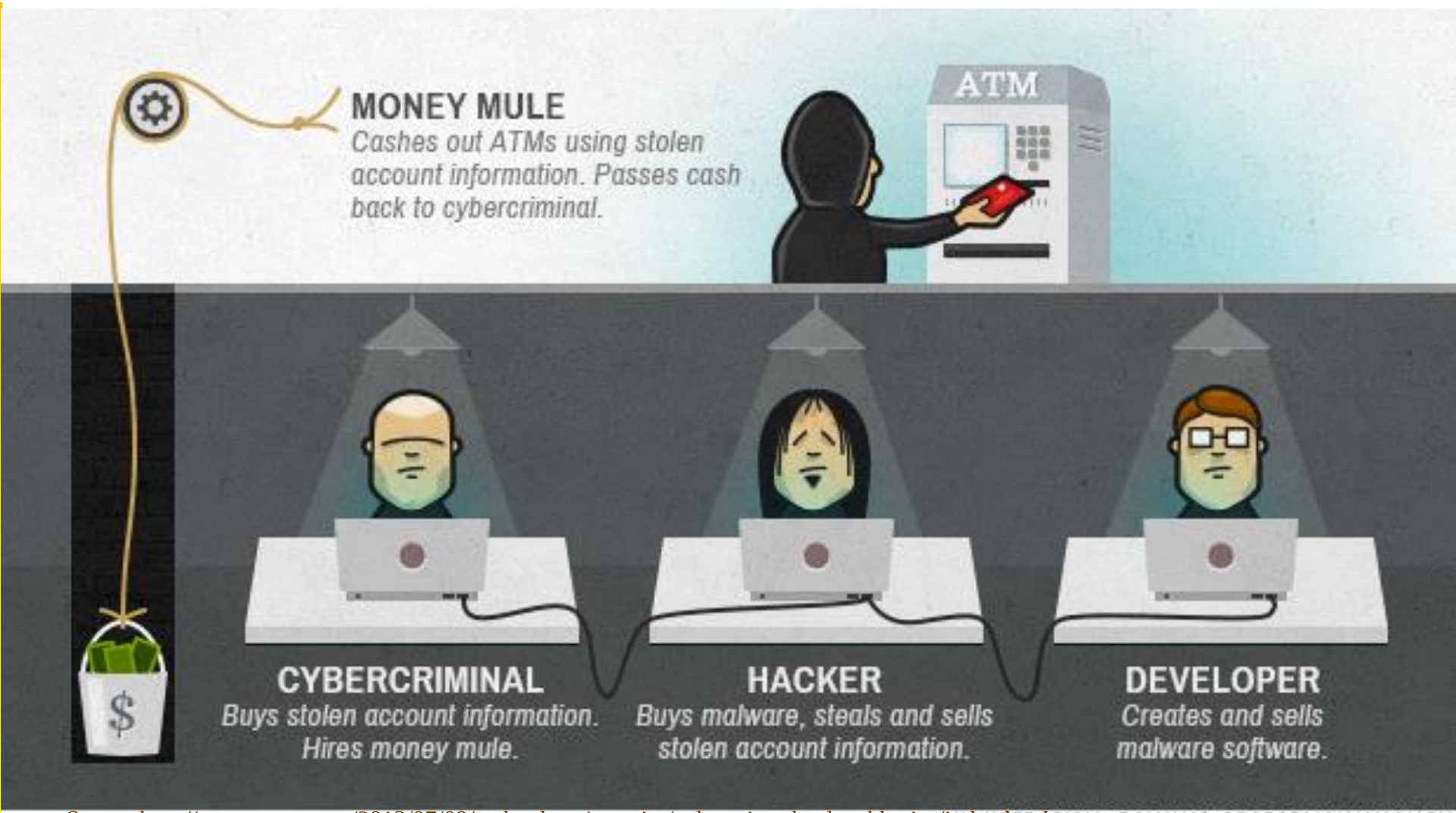
by Tom Krazit | September 14, 2010 5:27 PM PDT

Google confirmed on Tuesday that it fired an employee earlier this year for violating its policies on accessing the accounts of its users.

Earlier in the day, Gawker reported that David Barksdale, an engineer in Google's Seattle offices, used his position as a key engineer evaluating the health of Google's services to break into the Gmail and Google Voice accounts of several children. After parents of the children complained to Google, Gawker said Barksdale--who was not accused of anything with sexual overtones--was dismissed, and Google confirmed that move late Tuesday.

"We dismissed David Barksdale for breaking Google's strict internal privacy policies. We carefully control the number of employees who have access to our systems, and we regularly

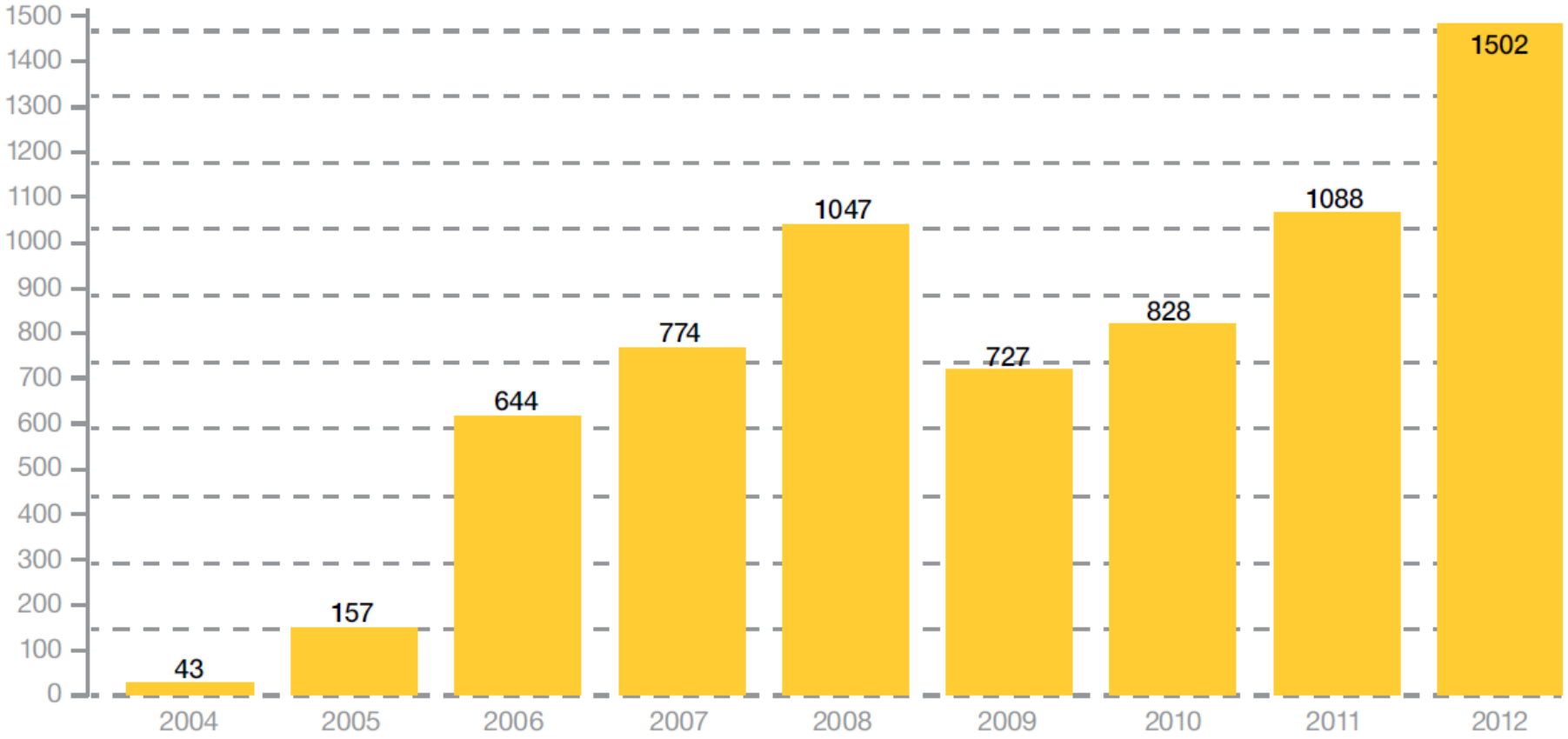
Fraud Detection & Prevention



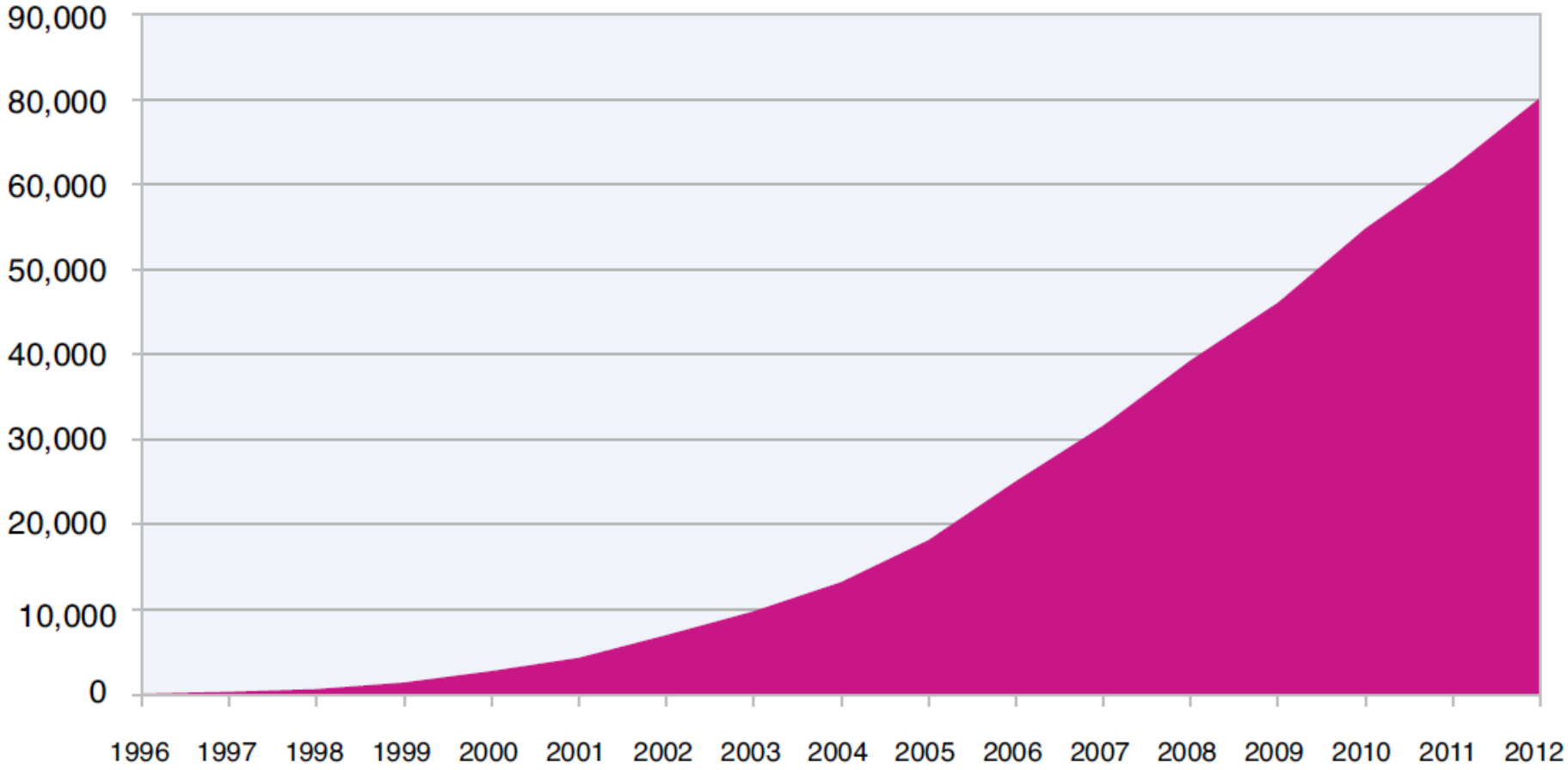
Source http://money.cnn.com/2013/07/09/technology/security/cybercrime-bank-robberies/index.html

Data Loss

- Reported Incidents of Data Loss



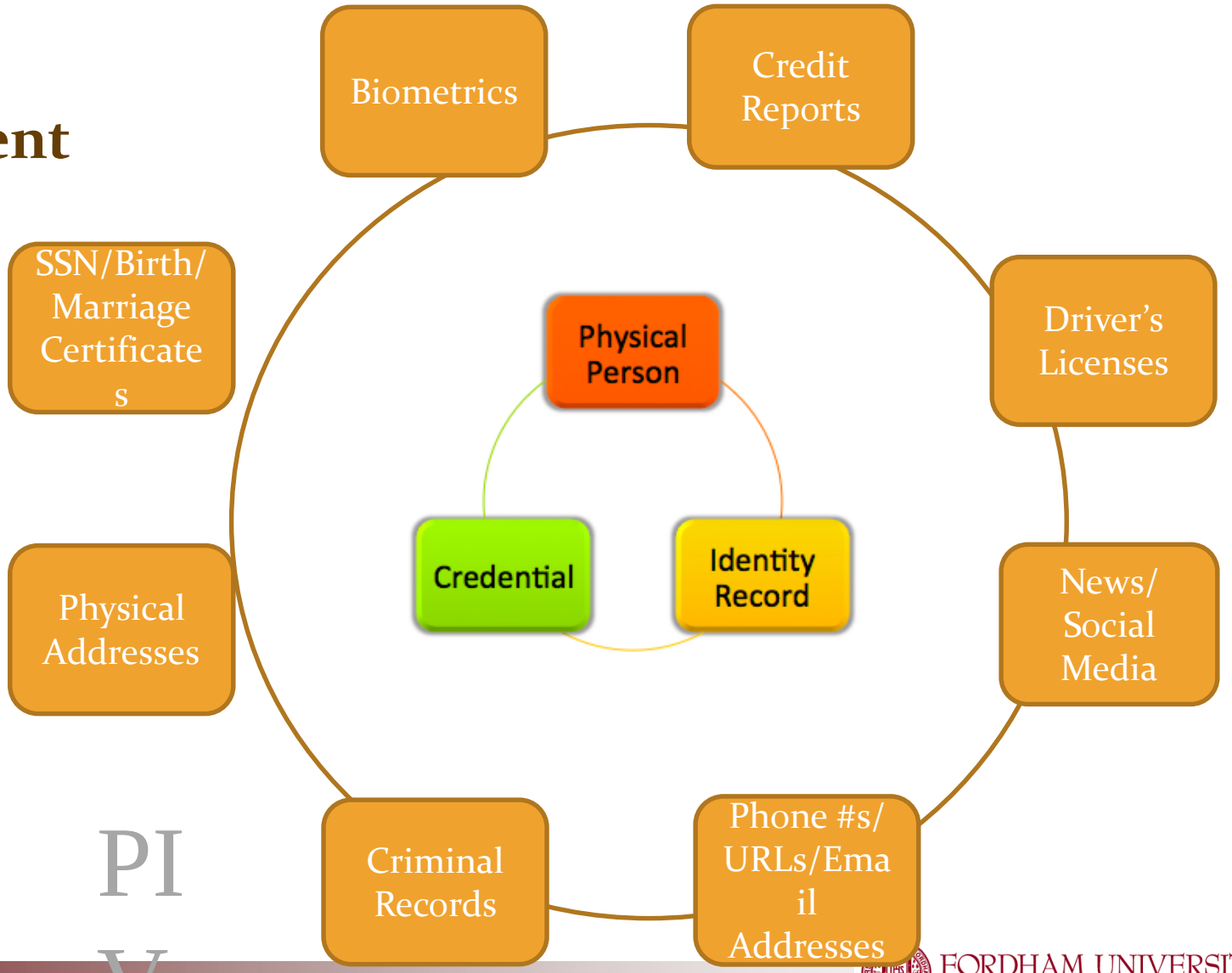
Reports of Vulnerabilities



IBM X-Force 2012 Trend and Risk Report *March 2013*

Big Data Privacy Concerns (1)

- **Identity Management**



PI

V

Big Data Privacy Concerns (2)

- "De-Identifed" Information Can Be "Re-Identified"
- Possible Deduction of Personally Identifiable Information
- Risk of Data Breach Is Increased
  - The higher concentration of data, the more appealing a target it makes for hackers, and the greater impact as a result of the breach

What We Are Going to Learn

- Privacy vs. Security
- Privacy Concerns
- **Method to Protect Privacy Concerns**
- Top Ten Big Data Security and Privacy Challenges

Method to Protect Privacy Concerns

- **Privacy Enhancing Technologies (PET):**
  - **PET is a term for a set of computer tools and applications**
    - This when integrated with online services allow online users to protect the privacy of their personally identifiable information.
  - **Numerous PETs have been proposed ranging from cryptographic techniques to data anonymization**
    - Such techniques either do not scale for large datasets and/or do not address the problem of reconciling security with privacy.

Method to Protect Privacy Concerns

- **Privacy Enhancing Technologies (PET):**
  - **Such techniques either do not scale for large datasets and/or do not address the problem of reconciling security with privacy.**
  - **Few approaches focus on efficiently reconciling security with privacy; these can be grouped as follows:**
    - Privacy-preserving data/record matching
    - Privacy-preserving collaborative data mining
    - Privacy-preserving biometric authentication

PET 1

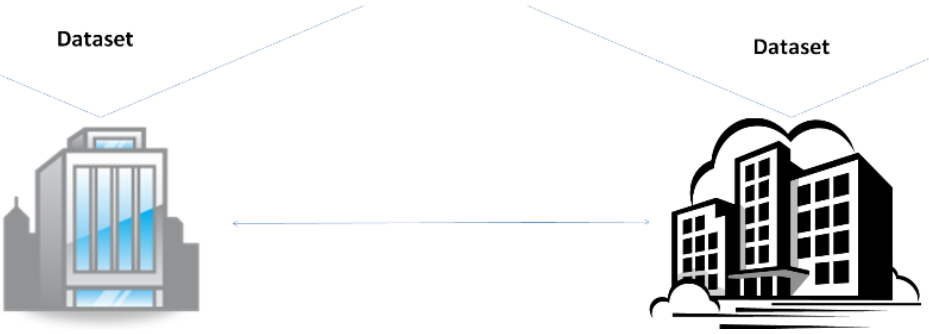
- **Privacy-preserving data/record matching**

Hospital A

Name	Sex	SSN	Age	Height (cm)
Angel Smith	Male	002-98-3445	20	180
Divine Scavo	Female	001-34-2356	24	162.5
Selene Paul	Female	000-22-6509	22	160
Sandrine Pal	Female	009-12-2222	23	167.5

Hospital B

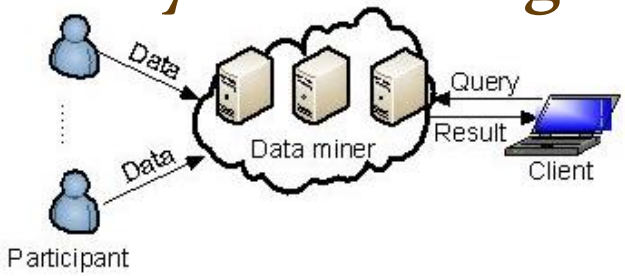
Name	Sex	SSN	Age	Height (cm)
Angel Smith	Male	002-98-3445	20	180
Divine Scavo	Female	001-34-2356	24	162.6
Ryan Solis	Male	033-24-0281	18	157.5
Katie Gomes	Female	243-30-2470	20	175



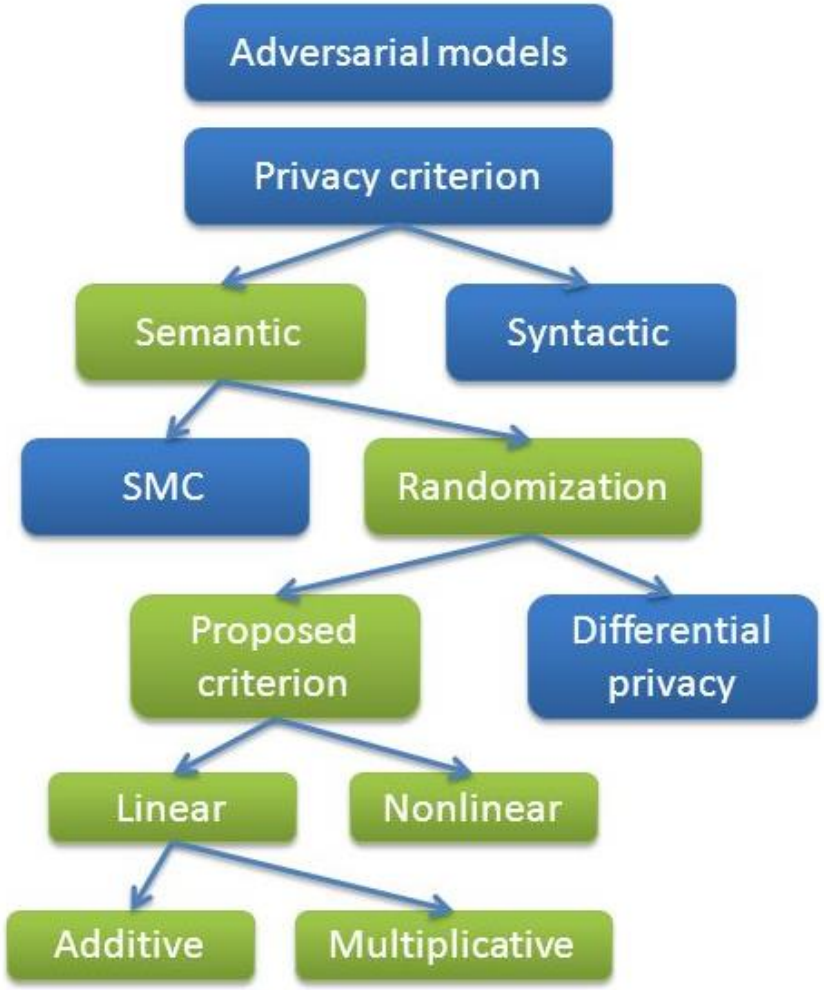
- Record matching is performed across different data sources with the aim of identifying shared common information
- Matching records from different data sources may conflict with privacy requirements of the individual data sources.

PET 2

Privacy Preserving Collaborative Data Mining

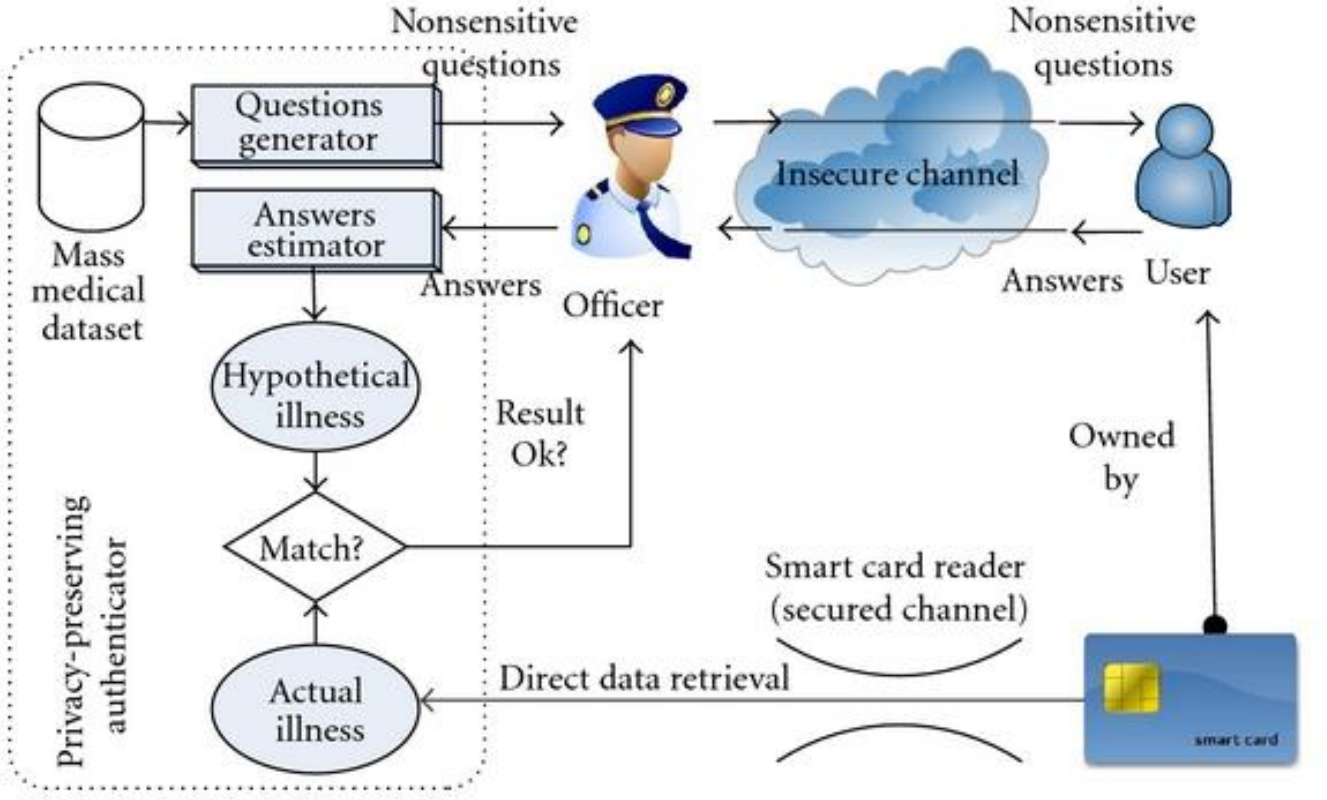


- **Requirement** imposed by participatory sensing:
  - online data submission, offline data processing
- **Design space**:
  - Data type:
    - continuous or categorical
    - voice, images, videos, etc.
  - Data structure:
    - relational or time series
    - for relational data: horizontal or vertical partitioned
  - Data mining operation



PET 3

- **Privacy Preserving Biometrics Authentication**



- Record biometrics templates of enrolled users and match with the templates provided by users during authentication

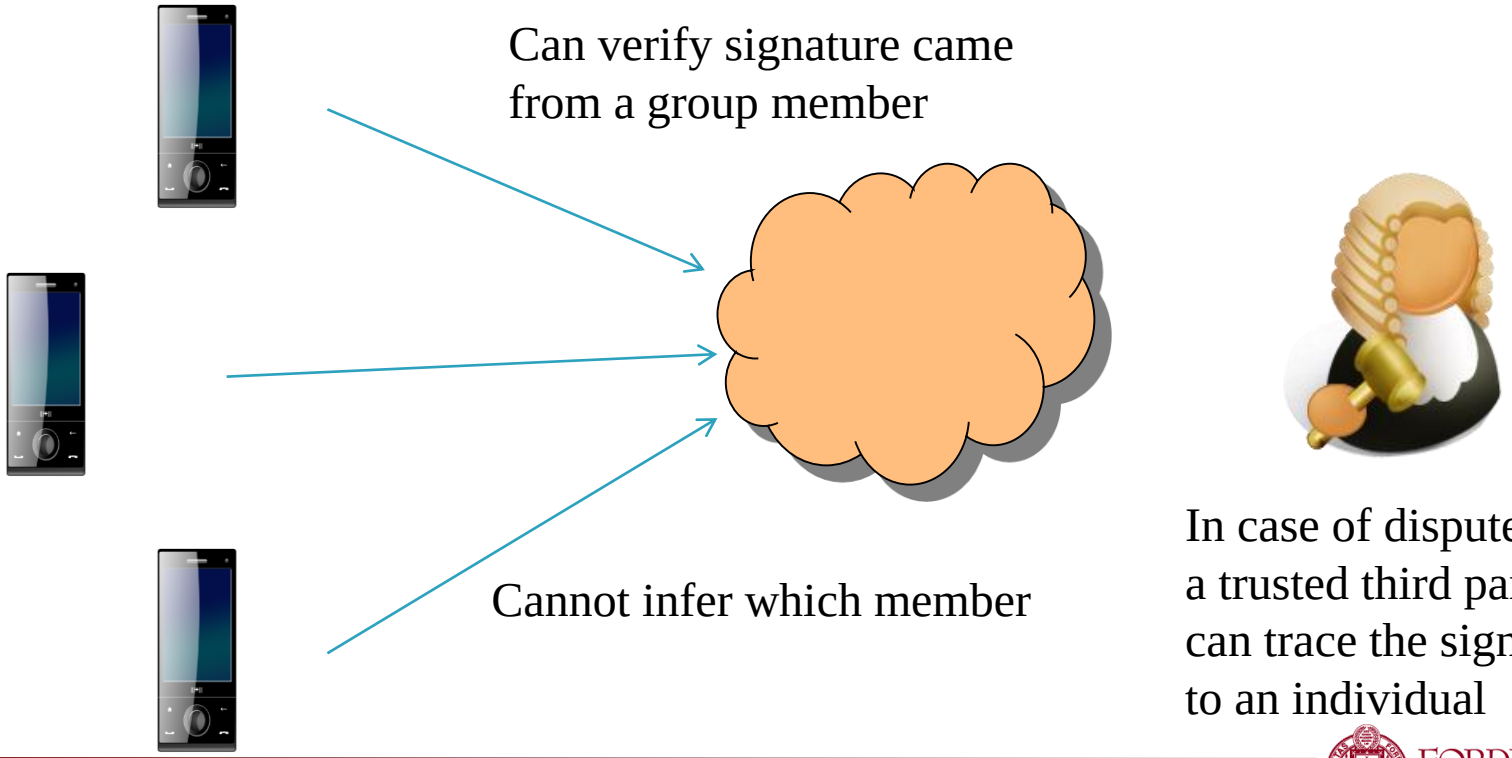
Crypto for Big Data Privacy

- Big data privacy
- Key management
- Data integrity and poisoning concerns
- Searching / filtering encrypted data
- Secure collaboration
- Secure outsourcing of computation

Crypto for Big Data Privacy

Secure and Privacy Preserving data collection

- How to make collection of data *private* as well as *authenticated*?



What We Are Going to Learn

- Privacy vs. Security
- Privacy Concerns
- Method to Protect Privacy Concerns
- **Top Ten Big Data Security and Privacy Challenges**

Top Ten Big Data Security and Privacy Challenges

Top Ten Big Data Security and Privacy Challenges

1. Secure computations in distributed programming frameworks
2. Security best practices for non-relational data stores
3. Secure data storage and transactions logs
4. End-point input validation/filtering
5. Real-time security/compliance monitoring

Top Ten Big Data Security and Privacy Challenges

6. Scalable privacy-preserving data mining and analytics
7. Cryptographically enforced access control and secure communication
8. Granular access control
9. Granular audits
10. Data provenance

Next Class

- Topics
  - Security in Big Data in details
  - Security algorithms/approaches in Big Data environments
  - Data Security
  - Secure data search
- Assignment 1
- Review Quiz