

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

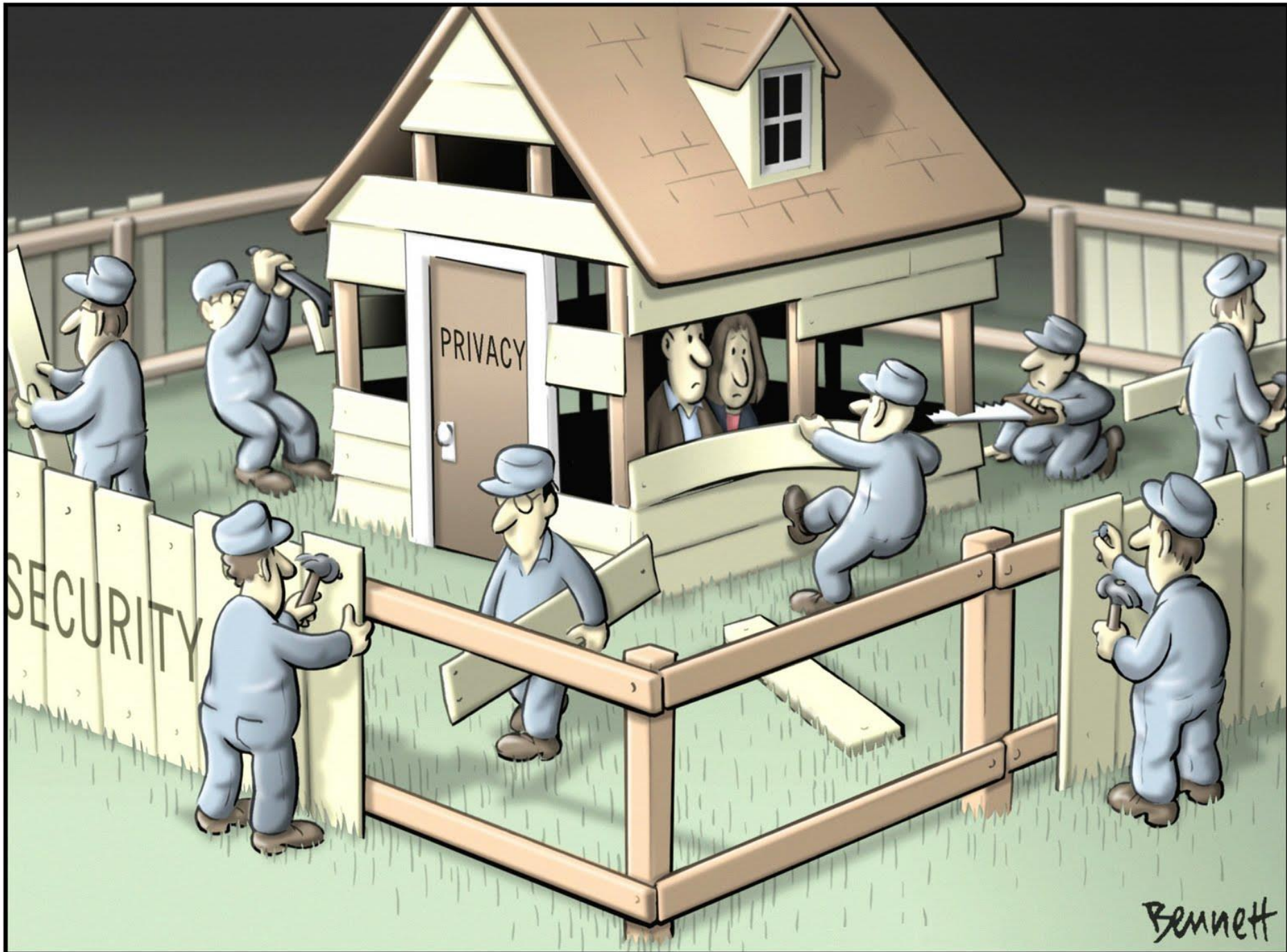


Review Quiz

- What makes Data “Big”?
- What are the 5Vs?
- Hadoop is an open-source software framework for structuring and storing data and running applications on clusters of commodity hardware. (agree/disagree)
- What are the two main functions of Hadoop?
- Securing a cluster through a gateway can be a way to provide security in Hadoop? (agree/disagree)
- What is MongoDB?
- Data security allows for the ciphertext to be protected (agree/disagree)
- How can Crypto be used for Big Data security?

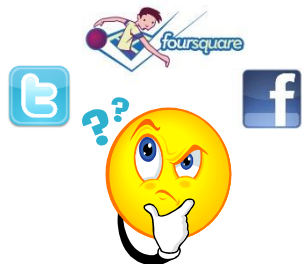
What We Are Going to Learn

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

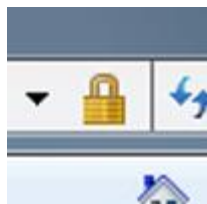


Bennett

Privacy vs. Security

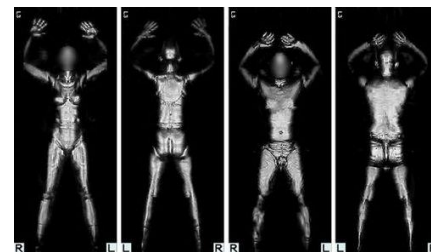


Privacy: what data goes where?



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

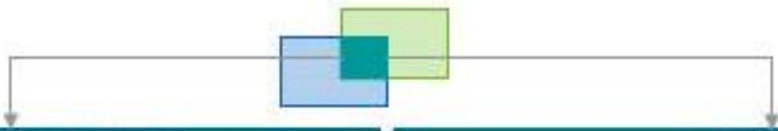
Privacy vs. Security

○ 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

○ Some related technical terms



	Information Security	Privacy
Accountability	<ul style="list-style-type: none"> Focuses on tracking an individual's actions and manipulation of information 	<ul style="list-style-type: none"> Focuses on tracking the trail of PII disclosure
Integrity	<ul style="list-style-type: none"> Protects against the corruption of data by authorized or unauthorized individuals 	<ul style="list-style-type: none"> Seeks to ensure that inaccurate PII is not used to make an inappropriate decision about a person
Aggregation	<ul style="list-style-type: none"> Focuses on determining the sensitivity of derived and aggregated data so that appropriate access guidance can be defined 	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> Focuses on processes and mechanisms (e.g., authenticators) that prevent unauthorized access 	<ul style="list-style-type: none"> Focuses on ensuring that PII is only disclosed for a purpose consistent with the reason it was collected
Destruction	<ul style="list-style-type: none"> Focuses on ensuring that information cannot be recovered once deleted 	<ul style="list-style-type: none"> 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 secrets
- ...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

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 »

News | World Cup | Comme

News | Technology | Rea

Series: Read me first

Facebook should compete on privacy, not hide it away



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.



Facebook says images are removed from its servers immediately.

Bruce Schneier
guardian.co.uk, Wednesday
Article history

Google Buzz Privacy Issues Have Real Life Implications

by Robin Wauters on Feb 12, 2010 150 Comments

Like

Buzz

134

retw

pular
addresses
ored - and

to find

them on seven sites - including

Who Knows Who Your Facebook Friends Are?

Deeplink by Tim Jones

As you may have heard, Facebook's privacy settings overhaul was the first of several changes to the site. Facebook users are now able to control who sees their profile, who can find them, and who can post on their wall. A researcher with

Facebook users unknowingly sharing personal data, warn researchers

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



OSNs mishandle data

5 of your friends became fans of Russell Peters

Worse than before, now Facebook doesn't even indicate that this is a sponsored ad.

Russell Peters
Come
28,791 fans

added A Dirty Shame to her queue on Blockbuster.

Updated: Justine Ezarik uploaded mobile photos.

Mobile Uploads - 66 photos
Check out Facebook Mobile

VALLEYWAG

YOUR PRIVACY IS AN ILLUSION

Why Facebook employees are profiling users

4:44 PM on Mon Oct 29 2007
By Owen Thomas
11,861 views

San Francisco, 6:14 AM
Wed Oct 22
28 posts in the last 24 hours

VALLEYWAG TEAM

Tip Your Editors:
tips@valleywag.com | AIM

Valleywag elsewhere on the Web:
Twitter | Facebook

Managing Editor:
Owen Thomas
Email | AIM

Very Special Correspondent:
Paul Routhin
Email

Associate Editors:
Nicholas Carlson (New York)
Email | AIM
Jackson West (San Francisco)
Email | AIM

Reporter:
Melissa Gira Grant
Email

Contributing Editor:
Evelyn Nussenbaum
Email | AIM

Valleywag Calendar:
Adriana Nunez
Email

What happens when you put twentysomethings in charge of a company with vast amounts of private information? **Sheer madcap chaos**, of course. Not to mention abuses of power. And that's what seems to be happening at Facebook. Valleywag kept hearing reports that Facebook employees had violated their users' privacy in a number of ways. The claimed abuses varied: **Looking at restricted profiles to check out dates. Seeing which profiles a user had viewed. And, in one case, allegedly logging onto a user's account, changing her profile picture to a graphic image, and sending faked messages.** Oh, and don't dare ask a Facebooker about any claims of misbehavior — they'll report you to customer service for "harassment." Facebook may have sophisticated privacy controls. But they don't appear to be deployed at headquarters.

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



Big Data and The Insider Threat

The "HoeflerText" font wasn't found.

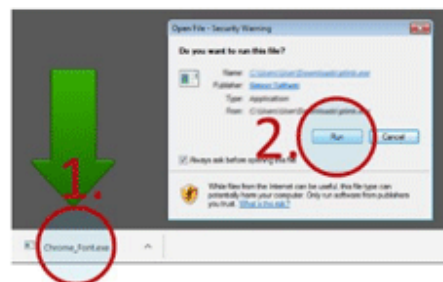
The web page you are trying to load is displayed incorrectly, as it uses the "HoeflerText" font. To fix the error and display the text, you have to update the "Chrome Font Pack".

Manufacturer: Google Inc. All Rights Reserved
Current version: Chrome Font Pack **53.0.2785.89**
Latest version: Chrome Font Pack **57.2.5284.21**

Update

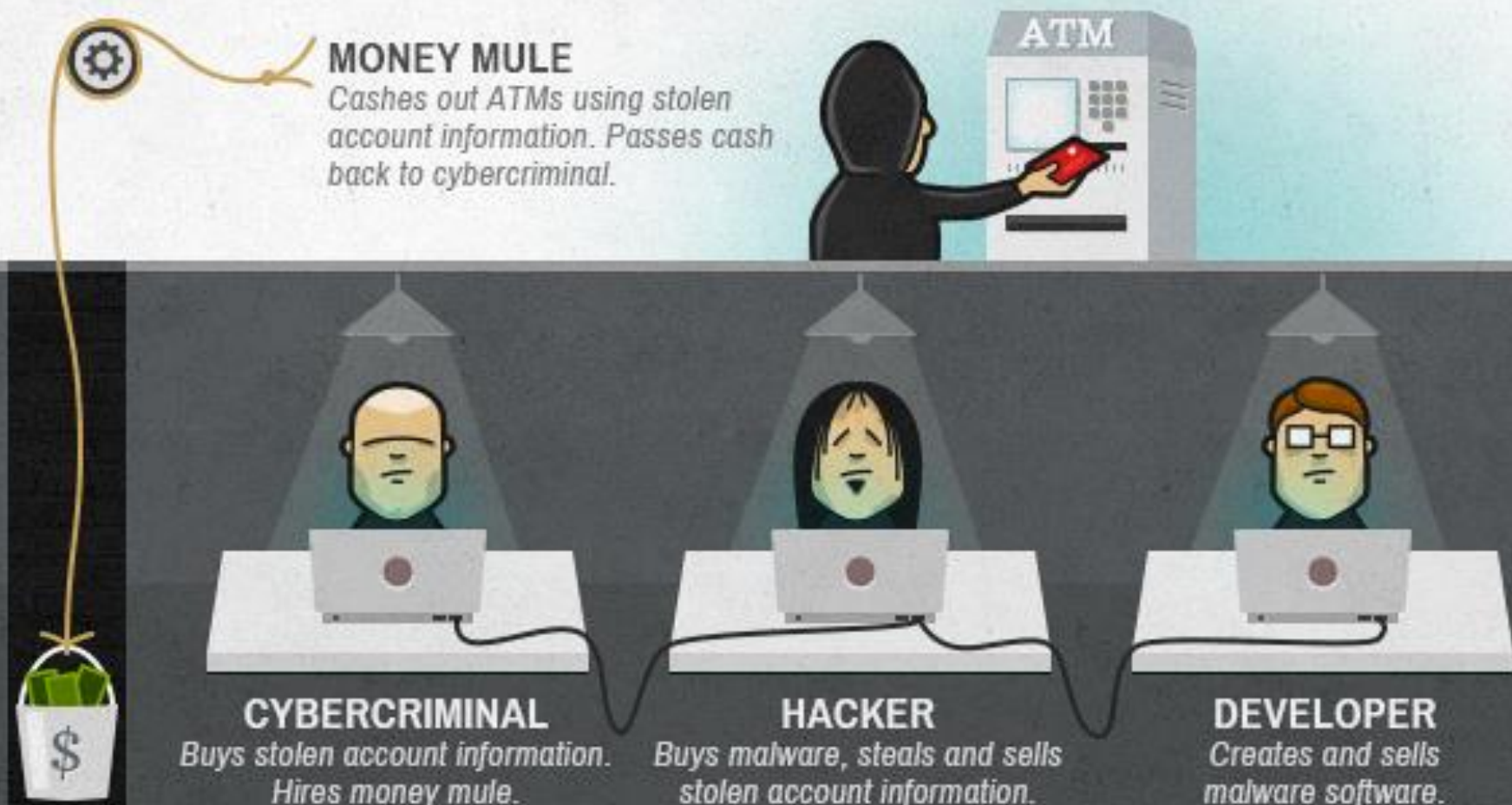
The "HoeflerText" font wasn't found.

Step 1: In the bottom left corner of the screen you'll see the download bar. **Click on the Chrome_Font.exe** item.
Step 2: Press **Yes(Run)** in order to see the correct content on the web page.



Update

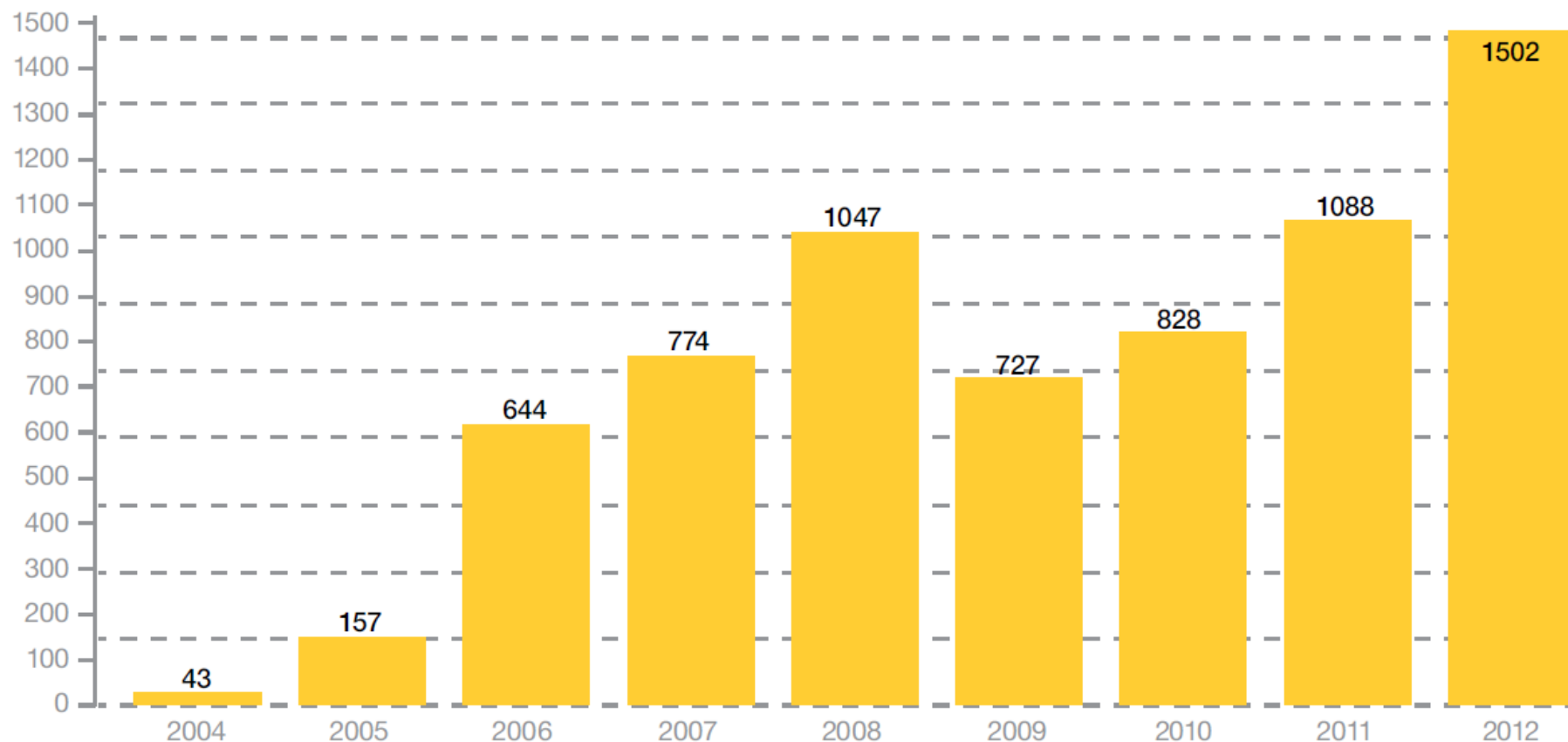
Fraud Detection & Prevention



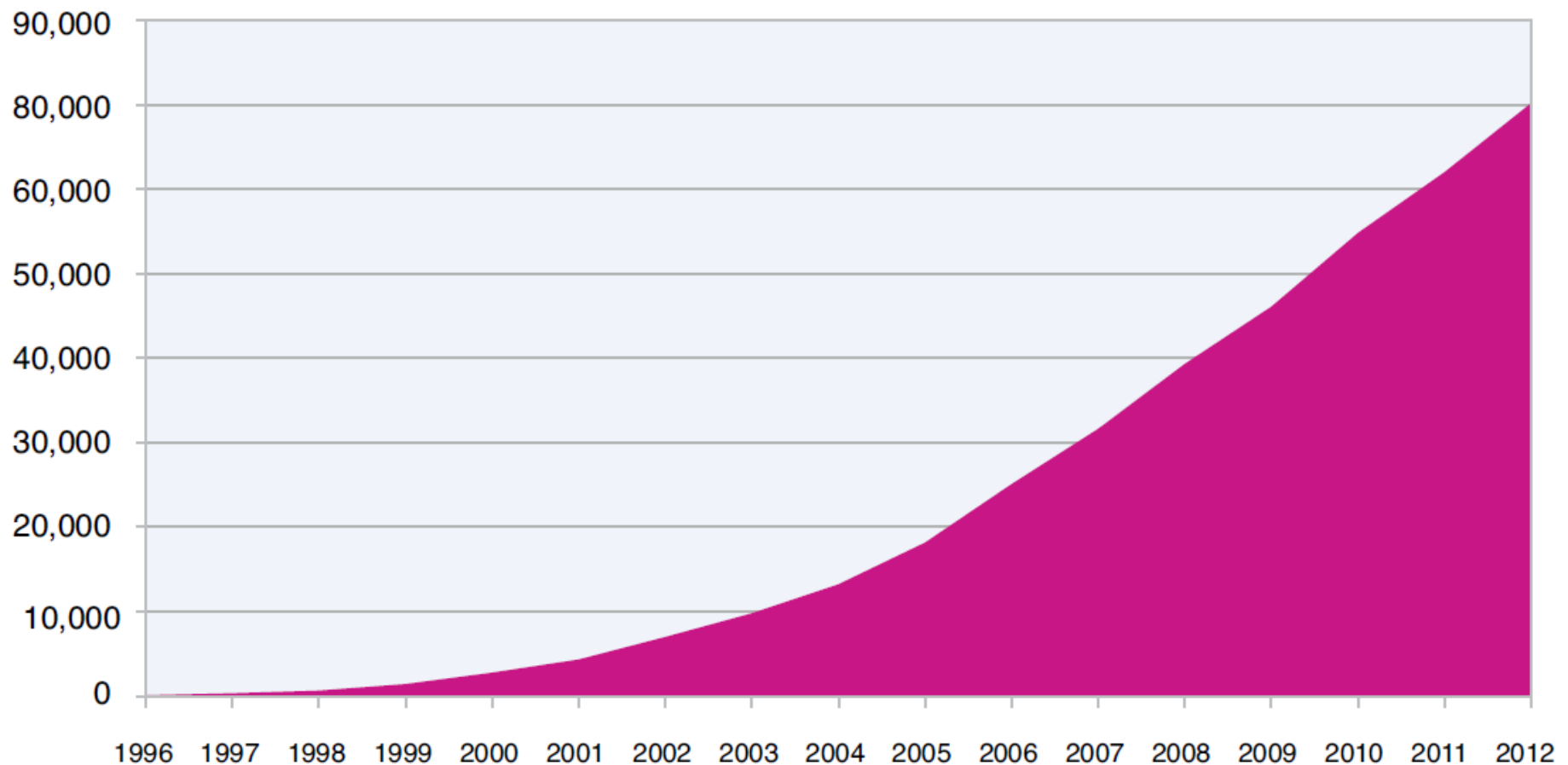
Source <http://money.cnn.com/2013/07/09/technology/security/cybercrime-bank-robberies/index.html>

Data Loss

○ Reported Incidents of Data Loss



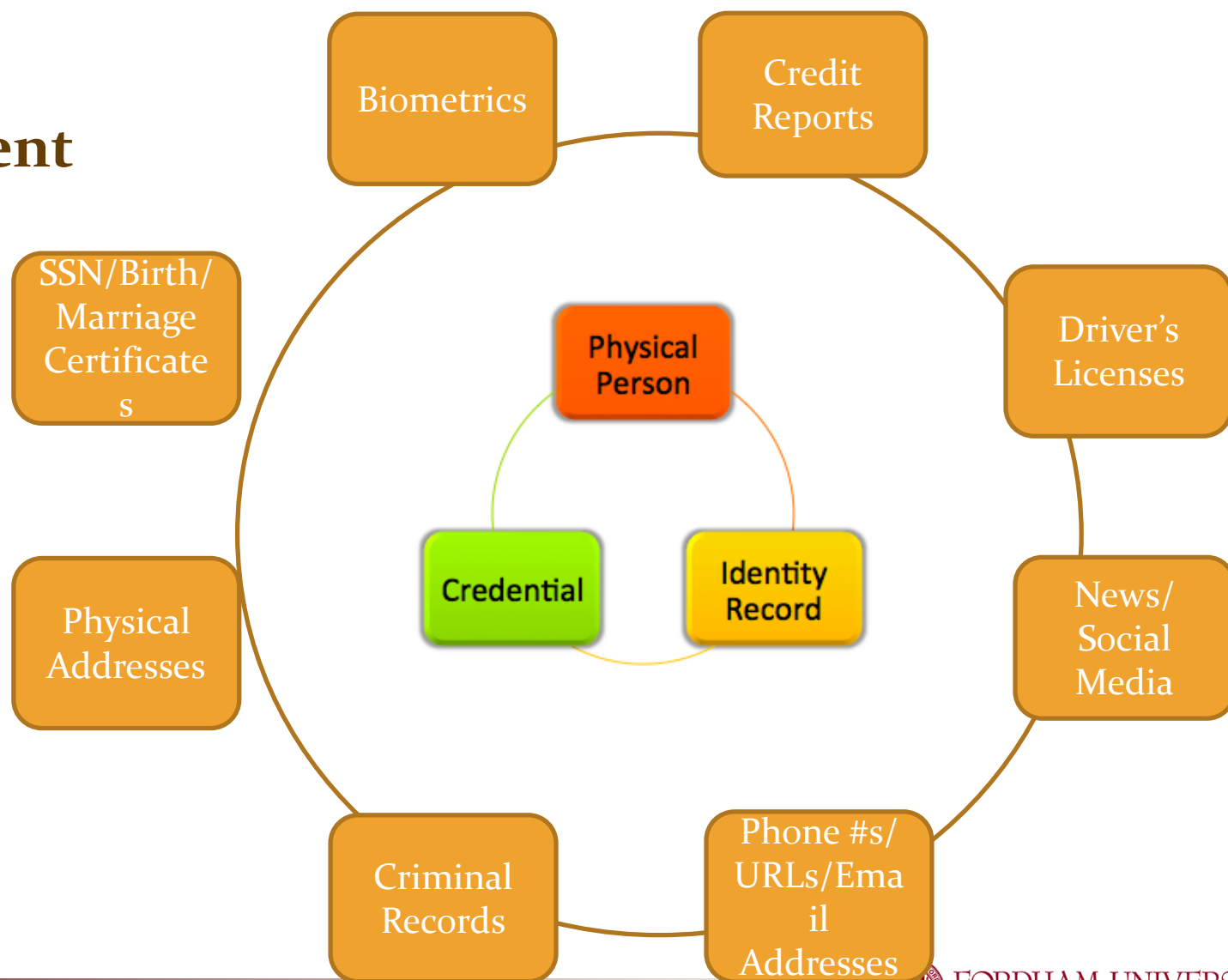
Reports of Vulnerabilities



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

Big Data Privacy Concerns (1)

○ Identity Management



Big Data Privacy Concerns (2)

- **“De-Identified” 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.

○

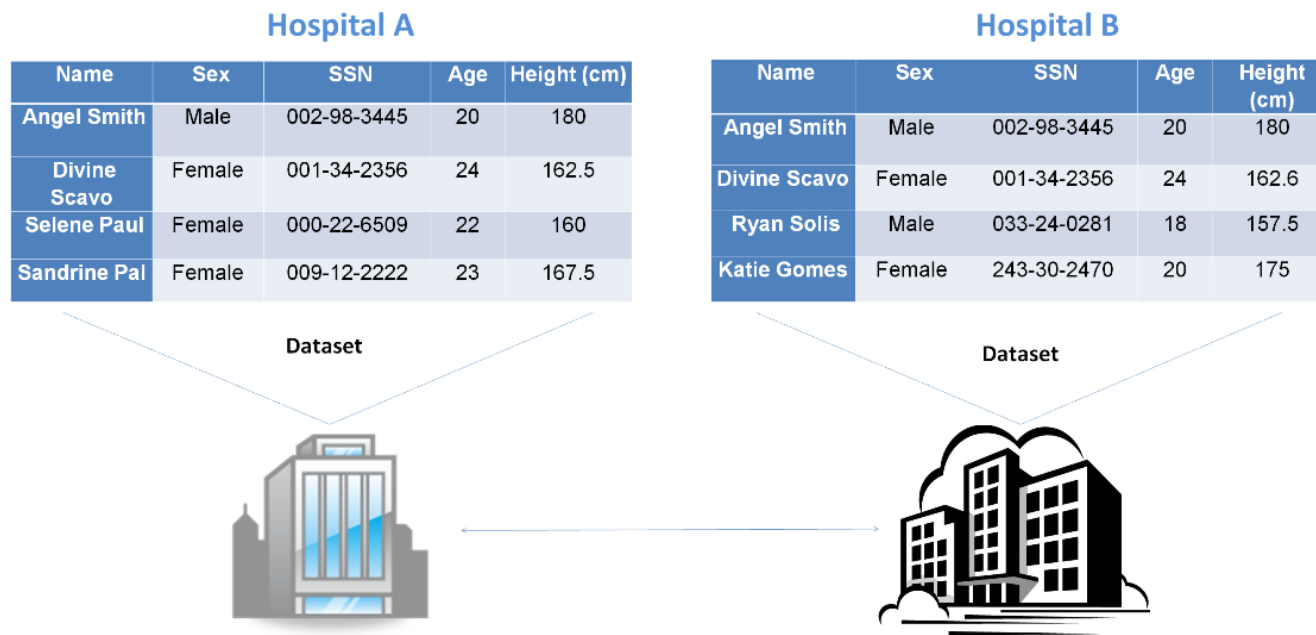
Method to Protect Privacy Concerns

○ 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

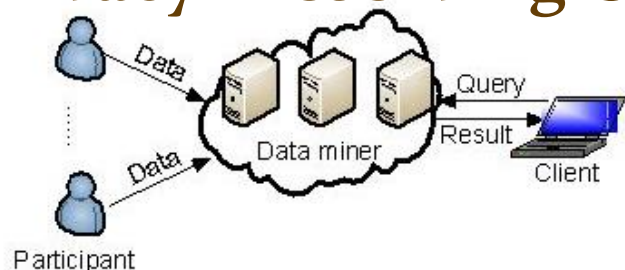
○ Privacy-preserving data/record matching



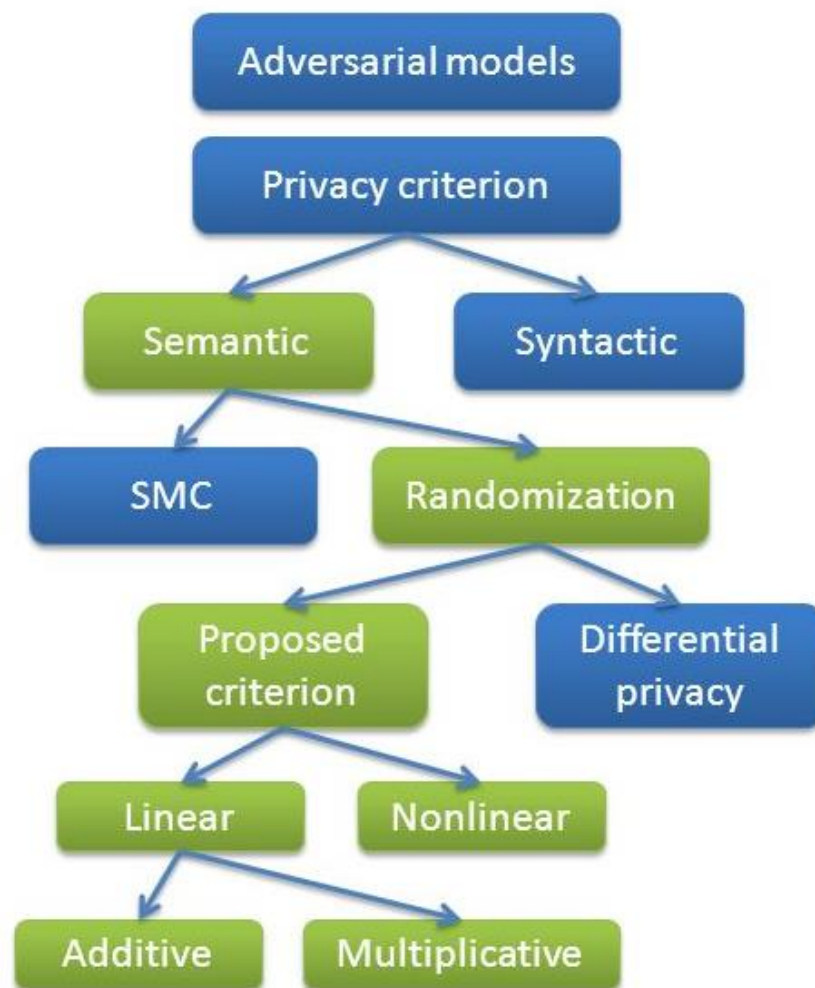
- Record matching is performed across different data sources with the aim of identifying shared common information
- Matching records from different data sources may conflict with privacy requirements of the individual data sources.

PET 2

○ Privacy Preserving Collaborative Data Mining

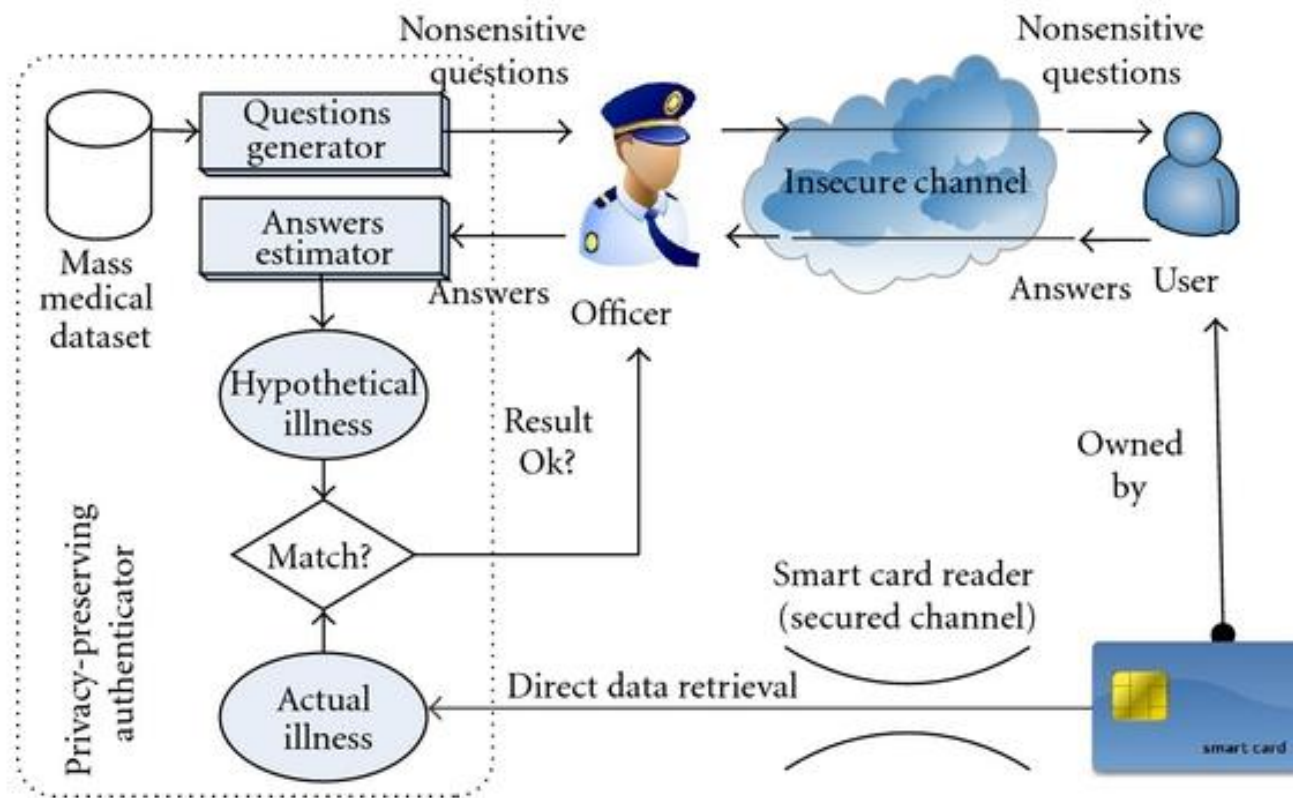


- **Requirement** imposed by participatory sensing:
 - online data submission, offline data processing
- **Design space:**
 - Data type:
 - continuous or categorical
 - voice, images, videos, etc.
 - Data structure:
 - relational or time series
 - for relational data: horizontal or vertical partitioned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



PET 3

○ Privacy Preserving Biometrics Authentication



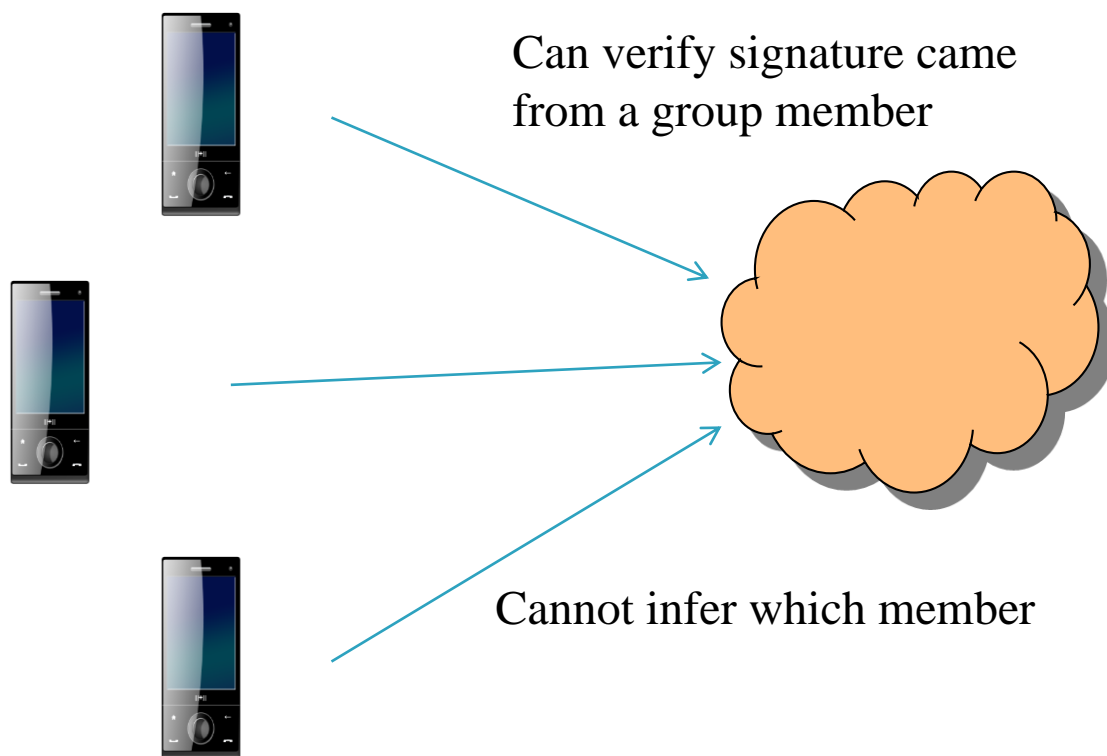
- Record biometrics templates of enrolled users and match with the templates provided by users during authentication

Crypto for Big Data Privacy

- 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*?



In case of dispute, a trusted third party can trace the signature to an individual

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