Today

• Meet everyone in class
• Course overview
  – Why data privacy and security
  – What is data privacy and security
  – What we will learn
• Course logistics
Instructor

• Li Xiong
  – Web: http://www.mathcs.emory.edu/~lxiong
  – Email: lxiong@emory.edu
  – Office Hours: MW 11:15-12:15pm or by appt
  – Office: MSC E412
About Me

http://www.mathcs.emory.edu/~lxiong

• Undergraduate teaching
  – CS170 Intro to CS I
  – CS171 Intro to CS II
  – CS377 Database systems
  – CS378 Data mining

• Graduate teaching
  – CS550 Database systems
  – CS570 Data mining
  – CS573 Data privacy and security
  – CS730R/CS584 Topics in data management – big data analytics

• Research http://www.mathcs.emory.edu/aims
  – data privacy and security
  – Spatiotemporal data management
  – health informatics

• Industry experience (software engineer)
  – SRA International
  – IBM internet security systems
Meet everyone in class

• Group introduction (2-3 people)
• Introducing your group
  – Name
  – Goals for taking the course
  – Something interesting about your group
Today

• Meet everyone in class
• Course overview
  – Why data privacy and security
  – What is data privacy and security
  – What we will learn
• Course logistics
Quiz

• How many people know you are in this room now?
  (a) no one
  (b) 1-5 i.e. your immediate family and friends
  (c) 5-20 i.e. your department staff, your colleagues and classmates
Quiz

• How many people know you are in this room now?
  (a) no one
  (b) 1-5 i.e. your immediate family and friends
  (c) 5-20 i.e. your department staff, your colleagues and classmates
Quiz

• How many organizations have your medical records?
The Data Map
“THAT’S your Ark for the Big Data flood? Noah, you will need a lot more storage space!”
The 5 V’s of Big Data

1. Volume: The size of the data
2. Velocity: The speed at which the data is generated
3. Variety: The different types of data
4. Veracity: The trustworthiness of the data in terms of accuracy
5. Value: Just having Big Data is of no use unless we can turn it into value
Value of Big Data

- GPS traces, call records
- Syndromic surveillance, social relationships
Value of Big Data

- Electronic health records (EHR)
- Secondary use for medical research
Value of Big Data
“I’m Big Data, and this is my friend No Privacy.”
PRIVACY POLL
Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are studying those details to figure out what you like, what you need, and which coupons are most likely to make you happy. Target, for example, has figured out how to data-mine its way into your womb, to figure out whether you have a baby on the way long before you need to start buying diapers.
Location Privacy Risks

- Tracking
- Identification
- Profiling

The dark side of geo: PleaseRobMe.com

You know all those people who push their Foursquare and Gowalla locations out onto Twitter? Now there’s an aggregator for aspiring crooks who want to rob their houses.

by Caroline McCarthy @caro / February 17, 2010 9:55 AM PST

More than a social statement than an ad campaign, Please Rob Me has popped up as a way to expose the potential pitfalls of the geo-tagging trend. If you're pushing a "check-in" to your location-based services, you're basically advertising that you're not home. Which could be a big problem.

More than a social statement than an ad campaign, Please Rob Me has popped up as a way to expose the potential pitfalls of the geo-tagging trend. If you're pushing a "check-in" to your location-based services, you're basically advertising that you're not home. Which could be a big problem.
Privacy Risks

Fired Over Facebook Post

Daycare Worker Fired For Candid Facebook Post

Read More: Daycare Worker Fired Facebook, Fired Over Facebook Post, Facebook, Social Media, Fired Over Social Media, Facebook Privacy, Single Mom Fired Facebook Post, Kaitlyn Waits Facebook, Video, Business News

If you needed a reminder about watching what you post on social media, here’s yet another cautionary tale. According to reports this week, a woman i...

Read Whole Story

TSA Agent Fired Over Anti-Muslim, Racist Facebook Comments

Read More: Islamophobia, Fired Over Facebook Post, Video, Roy Egan Chicago, Roy Egan Tsa, Roy Egan, Racism, Roy Egan Facebook, Anti-Muslim Bias, Fired Over Facebook Profile, Chicago News

A former O’Hare International Airport baggage screener employed with the Transportation Security Administration has been fired as a result of habitual...

Read Whole Story
A Texas Teenager Got Fired for a Tweet Before Starting Her Job

Employers use social media too, kids

A Texas teenager got fired from her new job less than 24 hours before she started after she used a couple of choice expletives to describe it on Twitter.

“Ew I start this f*** a** job tomorrow,” tweeted the teen with username @Cella_. CBS reports that the job she was referring to was at a branch of Jet’s Pizza in Mansfield, Texas. Unfortunately for the luckless teen, her tweet was spotted by store owner Robert Waple.

“And no... you don’t start the FA job today! I just fired you!” he replied, “Good luck with your no money, no job life.”

Waple reportedly last tweeted in 2009, and logged on only to publicly terminate Cella’s employment.

His tweets have since been deleted, and the pizza chain’s corporate office told CBS: “We regret to see the manner in which this situation has been handled by both parties involved.”

Not that Cella seems too heartbroken about it.

“I got fired over Twitter actionBar

— Cella. (@cellla__) February 7, 2015
Benefits … and Risks

Fine line between benefit and risks

(Most people don’t even see it)
What is the course about

• Techniques for ensuring data privacy and security (while harnessing value of data)

• Not about
  – Network security
  – System security
  – Software security
Today

• Meet everyone in class
• Course overview
  – Why data privacy and security
  – **What is data privacy and security**
  – What we will learn
• Course logistics
What is Privacy

• Definitions vary according to context and environment
• right to be left alone (Right to privacy, Warren and Brandeis, 1890; Olmstead v. United States (1928) dissent, Brandeis)
• a: The quality or state of being apart from company or observation; b: freedom from unauthorized intrusion (Merriam-Webster)
Aspects of Privacy

• **Information privacy**
  – Collection and handling of personal data, e.g. medical records

• **Bodily privacy**
  – Protection of physical selves against invasive procedures, e.g. genetic test

• **Privacy of communications**
  – Mail, telephones, emails

• **Territorial privacy**
  – Limits on intrusion into domestic environments, e.g. video surveillance
Information Privacy

– Data about individuals should not be automatically available to other individuals and organizations
– The individual must be able to exercise a substantial degree of control over that data and its use
– The barring of some kinds of negative consequences from the use of an individual’s personal information
Models of privacy protection

• Laws and regulations
  – Comprehensive laws
    • Adopted by European Union, Canada, Australia
  – Sectoral laws
    • Adopted by US
    • Financial privacy, protected health information
    • Lack of legal protections for data privacy on the Internet
  – Self-regulation
    • Companies and industry bodies establish codes of practice

• Technologies
A race to the bottom: privacy ranking of Internet service companies

- A study done by Privacy International into the privacy practices of key Internet based companies in 2007
- Amazon, AOL, Apple, BBC, eBay, Facebook, Google, LinkedIn, LiveJournal, Microsoft, MySpace, Skype, Wikipedia, LiveSpace, Yahoo!, YouTube
A Race to the Bottom: Methodologies

• Corporate administrative details
• Data collection and processing
• Data retention
• Openness and transparency
• Customer and user control
• Privacy enhancing innovations and privacy invasive innovations
A race to the bottom: interim results revealed

Privacy-friendly and privacy enhancing
Generally privacy-aware but in need of improvement
Generally aware of privacy rights, but demonstrate some notable lapses
Serious lapses in privacy practices
Substantial and comprehensive privacy threats
Comprehensive consumer surveillance & entrenched hostility to privacy
A race to the bottom: interim results revealed

<table>
<thead>
<tr>
<th>Privacy-friendly and privacy enhancing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generally privacy-aware but in need of improvement</td>
</tr>
<tr>
<td>Generally aware of privacy rights, but demonstrate some notable lapses</td>
</tr>
<tr>
<td>Serious lapses in privacy practices</td>
</tr>
<tr>
<td>Substantial and comprehensive privacy threats</td>
</tr>
<tr>
<td>Comprehensive consumer surveillance &amp; entrenched hostility to privacy</td>
</tr>
</tbody>
</table>
Why Google

• Retains a large quantity of information about users, often for an unstated or indefinite length of time, without clear limitation on subsequent use or disclosure
• Maintains records of all search strings with associated IP and time stamps for at least 18-24 months
• Additional personal information from user profiles in Orkut
• Use advanced profiling system for ads
Are Google and Facebook and ... Evil?

- Targeted advertising
- Cross-selling of users’ data
- Personalized experience
They are always watching ... what can we do?

Who cares? I have nothing to hide.
If you do care …

• Use cash when you can.
• Do not give your phone number, social-security number or address, unless you absolutely have to.
• Do not fill in questionnaires or respond to telemarketers.
• Demand that credit and data-marketing firms produce all information they have on you, correct errors and remove you from marketing lists.
• Check your medical records often.
• Block caller ID on your phone, and keep your number unlisted.
• Never leave your mobile phone on, your movements can be traced.
• Do not use store credit or discount cards
• If you must use the Internet, encrypt your e-mail, reject all “cookies” and never give your real name when registering at websites
• Better still, use somebody else’s computer
Privacy Protection Techniques

• Finding balances between privacy and multiple competing interests:
  – Privacy vs. other interests (e.g. quality of health care; movie recommendation; social network)
  – Privacy vs. interests of other people, organization, or society as a whole (e.g. advertising, insurance companies, healthcare research; movie recommendation for others).
Industry awareness and trends

How Google tricks itself to protect Chrome user privacy

An open-source project called Rappor uses randomly muddled data to let Google gather information about people's software usage while keeping individuals' behavior private.

It's a sticky issue for software developers: how do you gather data about your product's users without invading their privacy?

One solution, as embodied in a new Google open-source project called Rappor, is to have the software send data that you know is wrong.
APPLE’S ‘DIFFERENTIAL PRIVACY’ IS ABOUT COLLECTING YOUR DATA—BUT NOT YOUR DATA
Security

• The quality or state of being secure: as a: freedom from danger; b: freedom from fear or anxiety (merriam-webster)

• National security

• Individual security

• Computer security (cyber security)
  – Protecting information systems including the hardware, software, data, network, and services
Security vs. Privacy

• Data surveillance
  – Surveillance cameras
  – Sensors
  – Online surveillance
Principles of Data Security – CIA Triad

• Confidentiality
  – Prevent the disclosure of information to unauthorized users

• Integrity
  – Prevent improper modification

• Availability
  – Make data available to legitimate users
Privacy vs. Confidentiality

• Confidentiality
  – Prevent disclosure of information to unauthorized users

• Privacy
  – Prevent disclosure of personal information to unauthorized users
  – Control of how personal information is collected and used
  – Prevent identification of individuals
Data Privacy and Security Measures

- **Access control**
  - Restrict access to the (subset or view of) data to authorized users

- **Cryptography**
  - Use encryption to encode information so it can be only read by authorized users (protected in transmit and storage)

- **Inference control**
  - Restrict inference from accessible data to sensitive (non-accessible) data
Access Control

• Access control
  – Selective restriction of access to the data to authorized users

• Access control policies and mechanisms

• Issues
  – Fine grained access control
  – Spatial and temporal context
  – Group access control in social network applications
Cryptography

- Encoding data in a way that only authorized users can read it
Applications of Cryptography

• Secure data outsourcing
  – Support computation and queries on encrypted data
Applications of Cryptography

• Multi-party secure computations (secure function evaluation)
  – Securely compute a function without revealing private inputs
Applications of Cryptography

• Private information retrieval (access privacy)
  – Retrieve data without revealing query (access pattern)
Inference Control

• **Inference control**: Prevent inference from accessible information to individual information (not accessible)

• Technologies
  – De-identification and Anonymization (input perturbation)
  – Differential Privacy (output perturbation)
Traditional De-identification and Anonymization

- Attribute suppression, encoding, perturbation, generalization
- Subject to re-identification and disclosure attacks

### Original Data

<table>
<thead>
<tr>
<th>Name</th>
<th>Thread pitch (mm)</th>
<th>Minor diameter tolerance (mm)</th>
<th>Nominal diameter (mm)</th>
<th>Head shape</th>
<th>Price for 50 screws</th>
<th>Available at factory outlet?</th>
<th>Number in stock</th>
<th>Part or Phillips head?</th>
</tr>
</thead>
<tbody>
<tr>
<td>M4</td>
<td>0.7</td>
<td>4g</td>
<td>4</td>
<td>Pan</td>
<td>$10.00</td>
<td>Yes</td>
<td>276</td>
<td>Flat</td>
</tr>
<tr>
<td>M5</td>
<td>0.8</td>
<td>4g</td>
<td>5</td>
<td>Round</td>
<td>$13.89</td>
<td>Yes</td>
<td>182</td>
<td>Both</td>
</tr>
<tr>
<td>M6</td>
<td>1</td>
<td>5g</td>
<td>8</td>
<td>Button</td>
<td>$10.42</td>
<td>Yes</td>
<td>134</td>
<td>Flat</td>
</tr>
<tr>
<td>M8</td>
<td>1.25</td>
<td>5g</td>
<td>8</td>
<td>Pan</td>
<td>$11.98</td>
<td>No</td>
<td>298</td>
<td>Phillips</td>
</tr>
<tr>
<td>M10</td>
<td>1.5</td>
<td>8g</td>
<td>10</td>
<td>Round</td>
<td>$10.74</td>
<td>Yes</td>
<td>438</td>
<td>Phillips</td>
</tr>
<tr>
<td>M12</td>
<td>1.75</td>
<td>7g</td>
<td>12</td>
<td>Pan</td>
<td>$18.29</td>
<td>No</td>
<td>998</td>
<td>Flat</td>
</tr>
<tr>
<td>M14</td>
<td>2</td>
<td>7g</td>
<td>14</td>
<td>Round</td>
<td>$21.19</td>
<td>No</td>
<td>235</td>
<td>Phillips</td>
</tr>
<tr>
<td>M16</td>
<td>2</td>
<td>8g</td>
<td>16</td>
<td>Button</td>
<td>$23.57</td>
<td>Yes</td>
<td>290</td>
<td>Both</td>
</tr>
<tr>
<td>M18</td>
<td>2.1</td>
<td>8g</td>
<td>18</td>
<td>Button</td>
<td>$25.37</td>
<td>No</td>
<td>864</td>
<td>Both</td>
</tr>
<tr>
<td>M20</td>
<td>2.4</td>
<td>8g</td>
<td>20</td>
<td>Pan</td>
<td>$20.59</td>
<td>Yes</td>
<td>406</td>
<td>Both</td>
</tr>
<tr>
<td>M24</td>
<td>2.55</td>
<td>9g</td>
<td>24</td>
<td>Round</td>
<td>$32.01</td>
<td>Yes</td>
<td>952</td>
<td>Phillips</td>
</tr>
<tr>
<td>M28</td>
<td>2.7</td>
<td>9g</td>
<td>32</td>
<td>Button</td>
<td>$30.66</td>
<td>No</td>
<td>1867</td>
<td>Phillips</td>
</tr>
<tr>
<td>M36</td>
<td>3.2</td>
<td>10g</td>
<td>36</td>
<td>Pan</td>
<td>$41.32</td>
<td>Yes</td>
<td>434</td>
<td>Both</td>
</tr>
<tr>
<td>M40</td>
<td>4.5</td>
<td>10g</td>
<td>50</td>
<td>Pan</td>
<td>$44.72</td>
<td>No</td>
<td>740</td>
<td>Flat</td>
</tr>
</tbody>
</table>

### Sanitized Records

<table>
<thead>
<tr>
<th>Name</th>
<th>Thread pitch (mm)</th>
<th>Minor diameter tolerance (mm)</th>
<th>Nominal diameter (mm)</th>
<th>Head shape</th>
<th>Price for 50 screws</th>
<th>Available at factory outlet?</th>
<th>Number in stock</th>
<th>Part or Phillips head?</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>Pan</td>
<td>4</td>
<td>4</td>
<td>$10.08</td>
<td>Yes</td>
<td>276</td>
<td>Flat</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>Button</td>
<td>5</td>
<td>$13.89</td>
<td>Yes</td>
<td>182</td>
<td>Both</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Button</td>
<td>8</td>
<td>$10.42</td>
<td>Yes</td>
<td>134</td>
<td>Flat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.25</td>
<td>Pan</td>
<td>10</td>
<td>$11.98</td>
<td>No</td>
<td>298</td>
<td>Philips</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>Round</td>
<td>$10.74</td>
<td>Yes</td>
<td>438</td>
<td>Philips</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.75</td>
<td>Pan</td>
<td>$18.29</td>
<td>No</td>
<td>998</td>
<td>Flat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Round</td>
<td>$21.19</td>
<td>No</td>
<td>235</td>
<td>Phillips</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Button</td>
<td>$23.57</td>
<td>Yes</td>
<td>290</td>
<td>Both</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>Button</td>
<td>$25.37</td>
<td>No</td>
<td>864</td>
<td>Both</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.4</td>
<td>Pan</td>
<td>$20.59</td>
<td>Yes</td>
<td>406</td>
<td>Both</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.55</td>
<td>Round</td>
<td>$32.01</td>
<td>Yes</td>
<td>952</td>
<td>Phillips</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.7</td>
<td>Button</td>
<td>$30.66</td>
<td>No</td>
<td>1867</td>
<td>Phillips</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td>Pan</td>
<td>$41.32</td>
<td>Yes</td>
<td>434</td>
<td>Both</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Button</td>
<td>$44.72</td>
<td>No</td>
<td>740</td>
<td>Flat</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Statistical Data Sharing with Differential Privacy

- Macro data (as versus micro data)
- Output perturbation (as versus input perturbation)
- More rigorous guarantee
Course Topics

• Inference control
  – De-identification and anonymization
  – Differential privacy foundations
  – Differential privacy applications
    • Histograms
    • Data mining
    • Location privacy

• Cryptography

• Access control

• Applications
Course Topics

• Inference control

• Cryptography
  – Foundations
  – Applications
    • Secure outsourcing
    • Secure multiparty computations
    • Private information retrieval

• Access control

• Applications
Course Topics

• Inference control
• Cryptography
• Access control
  – Foundations
  – Emerging issues and access control in new settings
    • Spatiotemporal context-driven access control
    • Access control to shared data
    • Access control to encrypted data

• Applications
Course Topics

• Inference control
• Cryptography
• Access control
• Applications
  – Healthcare data
  – Cloud computing
  – Location based applications
  – Online social networks and social media
  – Crowdsourcing
Learning Objectives

• Learn the classic and state-of-the-art data privacy and security approaches
• Study various applications where data privacy and security is needed and can be applied
• Challenge existing solutions and identify new problems in data privacy and security
Today

• Meet everyone in class
• Course overview
  – Why data privacy and security
  – What is data privacy and security
  – What we will learn
• Course logistics
Logistics

• Reading materials
  – Book chapters, papers, online articles

• Prerequisite
  – Some database and statistics background
  – Programming skills

• Class webpage
  – Lecture slides
  – Link to readings
  – Project/assignments

http://www.mathcs.emory.edu/~cs573000
Workload

- ~2 programming assignments (individual)
- weekly reading assignments and paper reviews
- ~1 paper presentation in class
- 1 course project (team of up to 2 students) with project presentation
  - Application and evaluation of existing algorithms
  - Design of new algorithms to solve new problems
  - Survey of a class of algorithms
- 1 midterm
- No final exam
Paper reviews

• 1 page
• NOT just a summary of the paper, but your critical opinion of the paper
• Summarize (at least 3) things you like or learned
• Point out (at least 3) limitations, extensions, or interesting applications of the ideas
• Connect and contrast the paper to what we have learned/read so far
Course Project

• Options
  – Application and evaluation of existing algorithms
  – Design of new algorithms to solve new problems
  – Survey of a class of algorithms

• Timeline
  – 10/17: proposal
  – 11/28, 11/30, 12/5: Project workshop/presentation
  – 12/17: project report/deliverables
Late Policy

• Late assignment will be accepted within 3 days of the due date and penalized 10% per day

• 2 late assignment allowances, each can be used to turn in a single late assignment within 3 days of the due date without penalty.
Learning Objectives (Non technical)

• Read papers and write paper critiques
• Present papers and lead discussions
• Learn/practice the life cycle of a research project
  – literature review
  – problem formulation
  – project proposal writing
  – algorithm design
  – experimental studies
  – paper/project report writing
• Time management
Grading

• Assignments/presentations 40%
• Final project 30%
• Midterm 30%
Some expectations

• Participate in class, think critically, ask questions
• Read and write reviews critically
• Start on assignments and projects early
• Enjoy the class!