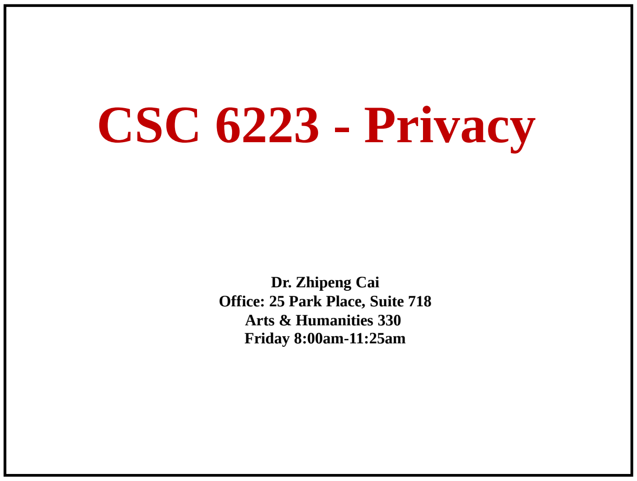
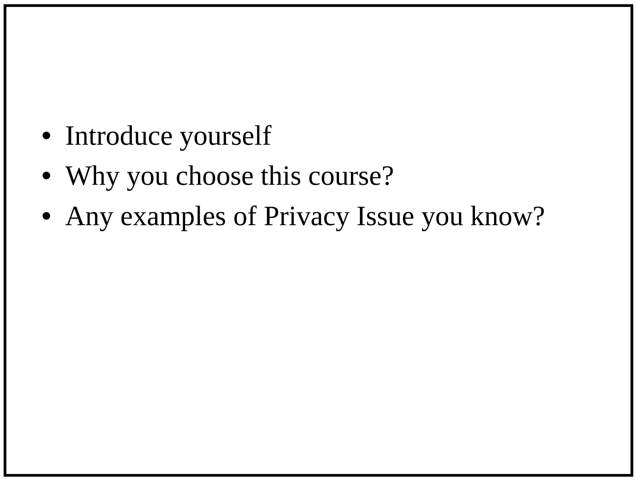
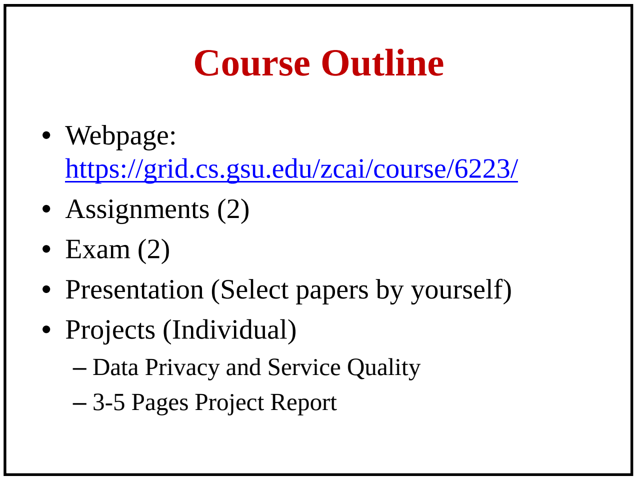


CSC 6223 - Privacy

Dr. Zhipeng Cai
Office: 25 Park Place, Suite 718
Arts & Humanities 330
Friday 8:00am-11:25am

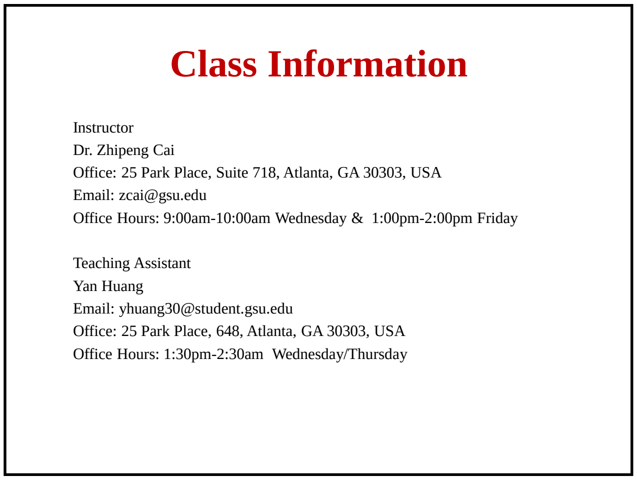


- Introduce yourself
- Why you choose this course?
- Any examples of Privacy Issue you know?



Course Outline

- Webpage: https://grid.cs.gsu.edu/zcai/course/6223/
- Assignments (2)
- Exam (2)
- Presentation (Select papers by yourself)
- Projects (Individual)
  - Data Privacy and Service Quality
  - 3-5 Pages Project Report



Class Information

Instructor
Dr. Zhipeng Cai
Office: 25 Park Place, Suite 718, Atlanta, GA 30303, USA
Email: zcai@gsu.edu
Office Hours: 9:00am-10:00am Wednesday & 1:00pm-2:00pm Friday

Teaching Assistant
Yan Huang
Email: yhuang30@student.gsu.edu
Office: 25 Park Place, 648, Atlanta, GA 30303, USA
Office Hours: 1:30pm-2:30am Wednesday/Thursday

Introduction

- Definition of Privacy

  [Alan Westin, Columbia University, 1967]

Privacy is the claim of individuals, groups or institutions to determine for themselves when, how, and to what extent information about them is communicated to others

Dimensions of Privacy

- Personal privacy
  - ➢ Protecting a person against undue interference (such as physical searches) and information that violates his/her moral sense
- Territorial privacy
  - ➢ Protecting a physical area surrounding a person that may not be violated without the acquiescence of the person
- Informational privacy
  - ➢ Deals with the gathering and selective dissemination of information

http://www.cs.purdue.edu/people/bb

Need for Privacy Guarantees

- Individuals [Cran *et al.* '99]
  - 99% unwilling to reveal their SSN
  - 18% unwilling to reveal their… favorite TV show
  - According to the 2015 Altimeter survey, Americans' privacy concerns are getting stronger, as "42% of Americans are more concerned about their digital privacy than they were just 12 months ago
- By businesses
  - Online consumers worrying about revealing personal data held back $15 billion in online revenue in 2001

Motivation for the Attacker

- Mobile payment transactions is projected to reach almost $630 billion by 2011.
- Put up a fake financial website, collect users' logins and passwords, empty out their accounts
- Insert a hidden program into unsuspecting users' computers, use it to spread spam
- Stage denial of service attacks on websites, extort money

Marketplace for Vulnerabilities

- Bug bounty programs for 2017
  - Apple: up to $200K
  - Google: up to $3133.7 in 2010, now up to $20K per bug
  - Facebook: up to $20K per bug
  - Microsoft: up to $150K per bug
  - Pwn2Own competition: $10-15K
  - GitHub: up to $10K
  - Intel: up to $30K
  - Uber: $10K

http://www.cs.purdue.edu/people/bb

Privacy Business

- Several companies specialize in finding and selling exploits
  - ReVuln, Vupen, Netragard, Exodus Intelligence
  - The average flaw sells for $35-160K
  - $100K+ annual subscription fees
- Nation-state buyers
  - “Israel, Britain, Russia, India and Brazil are some of the biggest spenders. North Korea is in the market, as are some Middle Eastern intelligence services. Countries in the Asian Pacific, including Malaysia and Singapore, are buying, too” -- NY Times (Jul 2013)

http://www.cs.purdue.edu/people/bb

Marketplace for Stolen Data

- Single credit card number: $4-15
- Single card with magnetic track data: $12-30
- “Fullz”: $25-40
  - Full name, address, phone, email addresses (with passwords), date of birth, SSN, bank account and routing numbers, online banking credentials, credit cards with magnetic track data and PINs
- Online credentials for a bank account with $70-150K balance: under $300

Prices dropped since 2011, indicating supply glut

http://www.cs.purdue.edu/people/bb

Marketplace

- Pay-per-install on compromised machines
  - US: $100-150 / 1000 downloads, “global mix”: $12-15
  - Can be used to send spam, stage denial of service attacks, perform click fraud, host scam websites
- Botnets for rent
  - DDoS: $10/hour or $150/week
  - Spam: from $10/1,000,000 emails
- Tools and services
  - Basic Trojans ($3-10), Windows rootkits ($300), email, SMS, ICQ spamming tools ($30-50), botnet setup and support ($200/month, etc.)

Facebook Privacy Leak

- Facebook has lost $35 billion in market value following reports that Cambridge Analytica, a data firm that worked with President Donald Trump in the 2016 elections, had unauthorized access to 50 million Facebook user accounts in one of its largest breaches yet.

http://fortune.com/2018/03/19/facebook-stock-share-price-cambridge-analytica-donald-trump/

Bad News

- Security often not a primary consideration
  - Performance and usability take precedence
- Feature-rich systems may be poorly understood
- Networks are more open and accessible than ever
  - Increased exposure, easier to cover tracks
- Many attacks are not even technical in nature
  - Phishing, social engineering, etc.

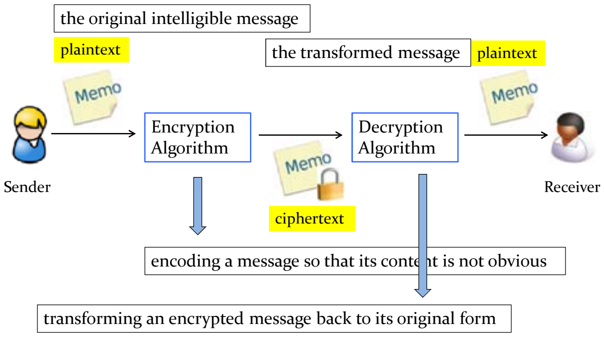
http://www.cs.purdue.edu/people/bb 14

Better News

- There are a lot of defense mechanisms
  - We'll study some, but by no means all, in this course
- It's important to understand their limitations
  - "If you think cryptography will solve your problem, then you don't understand cryptography… and you don't understand your problem"
  - Many security holes are based on misunderstanding
- Security awareness and user "buy-in" help
- Other important factors: usability and economics

http://www.cs.purdue.edu/people/bb 15

Cryptography - Basic Concepts



the original intelligible message

plaintext

the transformed message

plaintext

Sender

Encryption Algorithm

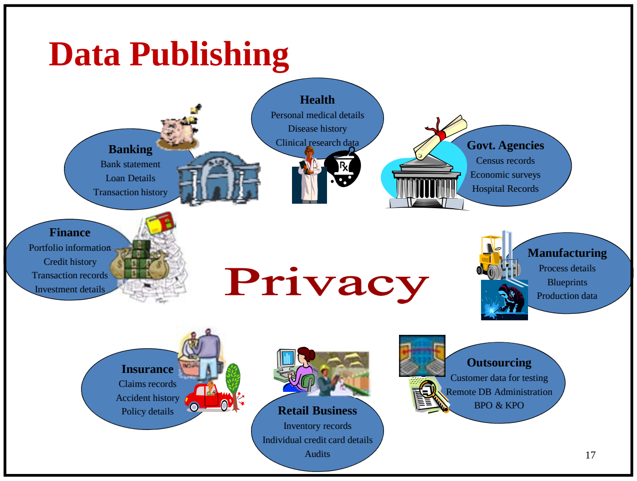
Decryption Algorithm

Receiver

ciphertext

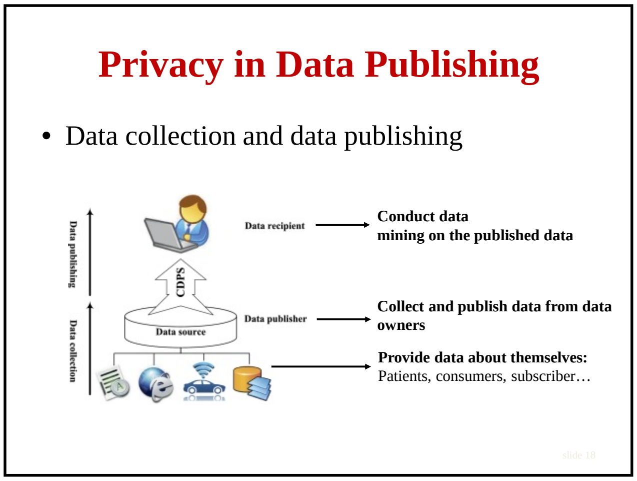
encoding a message so that its content is not obvious

transforming an encrypted message back to its original form



Data Publishing

17



Privacy in Data Publishing

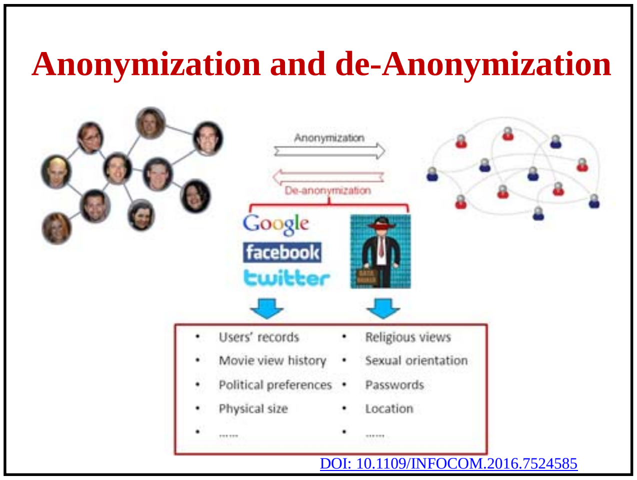
- Data collection and data publishing

Conduct data mining on the published data

Collect and publish data from data owners

Provide data about themselves:
Patients, consumers, subscriber…

slide 18

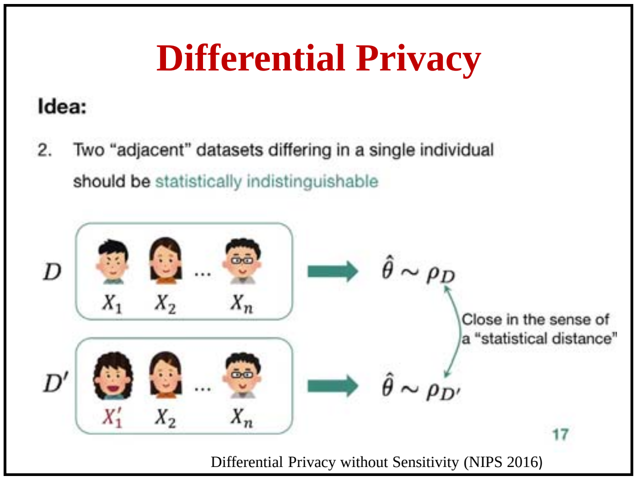


Anonymization and de-Anonymization

- Users' records
- Movie view history
- Political preferences
- Physical size
- ……

- Religious views
- Sexual orientation
- Passwords
- Location
- ……

DOI: 10.1109/INFOCOM.2016.7524585



Differential Privacy

Idea:

2. Two “adjacent” datasets differing in a single individual should be statistically indistinguishable

D
X_1
X_2
…
X_n
$\hat{\theta} \sim \rho_D$
Close in the sense of a “statistical distance”
D'
X'_1
X_2
…
X_n
$\hat{\theta} \sim \rho_{D'}$

17

Differential Privacy without Sensitivity (NIPS 2016)

CNN based Password Inference

Inferred / Action	1	2	3	4	5
1	42%	-	58%	-	-
2	32%	68%	-	-	-
3	-	-	60%	40%	-
4	26%	-	-	74%	-
5	-	-	-	19.2%	80.8%

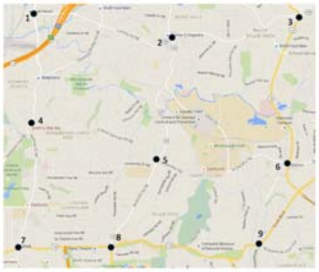
Threshold=35%

12345
12445
32345
32445

	4-Digit	6-Digit	8-Digit
Accuracy	94.3 %	92.0 %	89.9%

Deep Learning Based Inference of Private Information Using Embedded Sensors in Smart Devices, Accepted by IEEE Network Magazine (Impact Factor=7.230), 2018.

Experiments

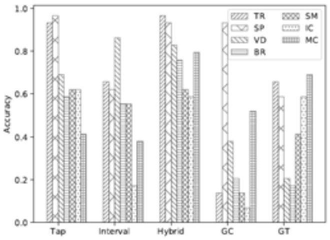


# of road segments	# of profiles	Ave success rate
10	1	84.2 %
5	1	90.3%
10	5	100 %

- 16 volunteers
- 48 unique routes
- 9 intersections
- Route lengths vary from 1 kilometer to 3 kilometers

Experiments --- App Usage Inference

Category	Apps
Gaming (TR)	Temple Run, Paris Metro
Shopping (SP)	TaoBao, Amazon
Video (VD)	Youku, Iqiyi
Browser (BR)	QQ browser, UC browser
Social Media (SM)	Weibo, Facebook
Instant Chat (IC)	WeChat, Messenger
Music Player (MC)	QQ Music, WangYi Yun Music



App usage inference

Pre-requisites

- **Discrete Mathematics**
- **Linear Algebra**
- **Statistics**
- **Calculus**
- **Optimization**