### **CS-577** Data Privacy

### **Overall Introduction**

Erman Ayday Bilkent University

## Introduction

- Who you are, who I am
- Intended audience: Master and PhD students in CS
- Crucial and super-hot topic, dramatically under-represented in both research and education ("academic inertia")

# **Goals and Prerequisites**

#### Learning outcomes

- By the end of the course, the student must be able to:
- Assess / Evaluate the privacy risks of a given organization or system
- Propose a set of solutions
- Implement those solutions
- Estimate the appropriateness and effectiveness of the solutions

#### **Transversal skills**

- Set objectives and design an action plan to reach those objectives
- Assess progress against the plan, and adapt the plan as appropriate
- Communicate effectively with professionals from other disciplines

#### Prerequisites

- Information security (or security engineering) and some applied crypto
- Network security
- Probability theory

# Content

- Introduction
  - History of privacy protection; the legal framework
  - Anonymity, unlinkability, unobservability and related concepts
  - Privacy by Design; privacy-enhancing technologies (PETs)
  - The future: wearable computing, DNA sequencing, electroencephalogram interfaces,...
- Crypto-Based Solutions
  - Identity management and anonymous credentials (zeroknowledge proofs)
  - Secure multi-party computation, including garbled circuits
  - Secret sharing, homomorphic encryption

# Content

- Data Privacy Hiding Data from the Database User
  - k-anonymity, l-diversity, t-proximity
  - Differential privacy and Laplacian noise, composability
- Hiding Access Patterns from the Database Owner
  - Private information retrieval (PIR)
  - Oblivious RAM (ORAM)
- Privacy in the Internet
  - Anonymous routing and anonymous Web surfing; Tor
  - Privacy in online social networks

## Content

- Privacy in E-cash
  - Bitcoin
- Privacy in E-voting
- Privacy in Mobile Networks
  - Privacy in cellular and WiFi networks
  - Location privacy and its quantification
- Privacy of Healthcare and Genomic Data
- Economics and Incentives
  - The elusive value of private data
  - Economics of privacy; targeted advertisement and ad blocking; why privacy is often not

### **Tentative Schedule**

- Week 1 (today): Administrativia, Introduction
- Week 2: Introduction (cont.). Crypto-based solutions.
- Week 3: Hiding Data from the Database User
- Week 4: Hiding Data from the Database User. Final assignment of mini-projects and readings
- Week 5: Hiding Access Patterns from the Database Owner
- Week 6: Privacy in the Internet
- Week 7: Mid-term presentation of mini-projects
- Week 8: Privacy in E-cash and E-Voting
- Week 9: Privacy in Mobile Networks
- Week 10: Privacy of Healthcare and Genomic Data
- Week 11: Economics and Incentives
- Week 12: Reading Group Presentations
- Week 13: Mini Project Presentations
- Week 14: Oral Exam

# Mini-projects

- Carried out by maximum 2 students
- Tutoring by me and the grader
- All projects are different
- If you take this course for credit, be proactive on your choice of project (and possibly project partner identification)
- You can propose your own subject and we will discuss its appropriateness; It can be related to your ongoing research
- Ideally, a successful mini-project can *lead to* a publication

# Mini-projects

- Novelty and effort on the project are important
- Group projects require more effort than individual projects and will be graded accordingly
- Peer grading (for group projects) is very important
- Midterm report, final report, final presentation, (midterm presentation)
- No late submissions
- You should be physically present for the presentations and the oral exam

# Mini-project Format

- Research
  - Focus on a particular topic
  - Do a literature survey
    - NDSS, ACM CCS, IEEE S&P, Usenix Security, PETS
  - Analyze the existing work and criticize (determine weaknesses and potential improvements)
  - Make suggestions, propose your improvements
  - Examples:
    - Privacy in social networks and microblogging systems
    - Privacy-enhanced access control, authentication, and identity management
    - Traffic analysis
    - De-anonymization

# Mini-project Format

- Implementation
  - Focus on a particular application or dataset
  - Decide on the architecture and system model
  - Determine the privacy requirements
  - Implement PETs for your application or dataset
  - Examples:
    - CryptDB for genomic data
    - ORAM for large datasets
    - Applications of Garbled circuits
    - Web crawling and deanonymization

# **Examples from Last Semesters**

- Effect of Indirect Information Sharing on Privacy
- Kin Genomic Privacy: Inference Attacks
- De-anonymizing Unstructured Online Social Networks
- De-anonymizing Private Instagram Profiles via Twitter
- De-anonymization of Social Network Data
- Privacy Preserving Active Learning with Secure Multiparty Computation
- Microsoft Malware Classification Challenge
- Data Protection Legislation in Turkey, EU and USA
- Privacy-Preserving Community Detection
- Practical Differential Privacy via Grouping and Smoothing
- Side Channel Attacks: A Historical Survey
- Privacy Preserving Genome Wide Association Studies (GWAS) Using Hadoop
- De-anonymizing Call Records
- Data Sharing and Privacy in Genomics
- De-anonymizing medical databases
- Privacy Preserving Dynamic Time Warping
- Privacy preserving of RIMARC algorithm
- Detecting Fake Accounts on Social Networks
- Bioinformatic Data Sharing
- De-anonymizing Online Social Networks

### Last Semesters



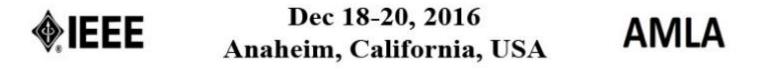
1

#### Privacy-aware computational genomics 2015 (PRIVAGEN 2015)

An official satellite workshop of GIW/InCoB 2015. September 8, 2015. Tokyo, Japan

Call for Talks/Posters

15th	IEEE	International	Conference	on	Machine	Learning	and Applications	
------	------	---------------	------------	----	---------	----------	------------------	--



### **IEEE/ACM Transactions on Computational Biology and Bioinformatics**

*IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* is a bimonthly journal that publishes archival research results related to the algorithmic, mathematical, statistical, and computational methods that are central in bioinformatics and computational biology. Read the full scope of TCBB

# Grading and Website

- Grading
  - Mini-project: 60%
  - Oral exam: 40%
  - Bonus: class participation: 10%
- Website
  - <u>http://www.cs.bilkent.edu.tr/~erman/Teaching.ht</u>
    <u>ml</u>
- That's it! Questions?