

Privacy and Anonymity in Data *Lecture 0*

School of Computer Science
Carnegie Mellon University
latanya@privacy.cs.cmu.edu

Question Addressed in this Course

The emergence of many new technologies becomes increasingly hampered by privacy concerns because these technologies leave society vulnerable to privacy abuses.

Current situation.

Let society choose between benefiting from the technology and maintaining privacy protections.

Our solution.

To proactively construct privacy technology (and integrated policy) with provable guarantees of privacy protection while allowing society to collect and share person-specific information for many worthy purposes .

An Objective – data detective

In this course, you will learn how to exploit existing data collections and how to reason about vulnerabilities in publicly available information.

An Objective – data protector

In this course, you will also learn how to render person-specific information available such that the ability to identify individuals contained in the released data is controlled, yet the data remain practically useful.

Privacy and Anonymity in Data

Course begins officially Tues 9/14 because of SCS Graduate IC

15-394 undergraduate /17-802, graduate version

Meets in Wean 5419ab

Highlights: importance of topic, research effort, programming requirement, privacy technology emphasis, term project

<http://privacy.cs.cmu.edu/courses/pad1/index.html>

<http://privacy.cs.cmu.edu/courses/pad1/index.html>

The screenshot shows a web browser window displaying the course page. The page title is "Privacy and Anonymity in Data" and the course number is "CS 15-394 / CALD 10-711". The professor is listed as Latanya Sweeney, Ph.D., with her email as latanya@privacy.cs.cmu.edu. The lecture schedule is Tuesdays and Thursdays, 3:00-4:20pm, Wean 5419ab. Office hours are 3:45pm on Tuesdays and Thursdays, Wean 1201, or contact Shantae Livingston at shantae@andrew.cmu.edu to make an appointment with Professor Sweeney. Links for "Schedule", "Handouts", and "Syllabus" are provided. The "Course Description" section begins with: "This course introduces students to concepts and methods for sharing person-specific data with provable guarantees of privacy protection using approaches that weave privacy technology into legal and social norms. Methods include those related to the identifiability of data, record linkage, data profiling, data fusion, data anonymity, de-identification, policy specification and enforcement and privacy-preserving data mining. This knowledge is timely because society is experiencing exponential growth in the number and variety of person-specific data collected, while unprecedented demand is also increasing for sharing collected data for many worthy purposes. Students learn a 3-prong approach that consists of (1) ways to learn sensitive information from disparate fragments of data, (2) methods that provably protect references that can be drawn

Privacy and Anonymity in Data Lecture

Overview of Course Requirements

1. Weekly lectures and discussions 10%
2. Weekly labs and homework assignments (grad students do more) 30%
3. Term Project, faculty showing at end of term. 60%
 - Powerpoint presentation
 - Poster
 - Project Report (undergrad)
 - Conference-style paper (grad)

No exams.

Term project has a steady flow of deliverables to move you through the process: proposal abstract, introduction, methods, experiments, discussion

http://privacy.cs.cmu.edu/courses/pad1/syllabus.html

Term Project

Each student must complete a term project, which constitutes the bulk of a student's grade. Special assignments will be provided throughout the semester to insure progress on projects. At the end of the semester is an SCS-wide exhibition of student projects. **EACH STUDENT MUST BE PRESENT, AND MUST EXHIBIT HIS HER PROJECT!** The overall grade for a project is broken down into:

- o Powerpoint presentation
- o Poster
- o Project report (undergraduate credit) or a conference-style paper (graduate credit)

Grading Policy

Your final grade in this course is based on:

1. 10% Class participation (in-class labs)
2. 30% Lab assignments (may drop 1 with no penalty)
3. 60% Project (Proposal, Presentation, Report)

Background for Labs and Projects

1. Programming (Java or equivalent)
2. Ability to express yourself in writing
3. Web searching beneficial

SQL beneficial, but there will be a primer provided.

Projects involve (data detective & data protective): given a stated problem in a setting,

provide a programming or technical solution
a synthesis of solution into the social/policy setting
analysis of result

http://privacy.cs.cmu.edu/courses/pad1/syllabus.html

Assignments

Assignments are provided weekly, though students may skip one assignment of their own selection without penalty provided the assignment is not an assignment directly related to the student's project. These weekly assignments are extensions to the in-class lab activities conducted each week. They engage students in a variety of diverse activities. Examples include: surveys, statistical analysis, database manipulations, image analysis, web searching, and more.

Students are expected to be uniquely responsible for their work. However, working and sharing ideas with other students on the course who are working on similar or complementary work is strongly encouraged.

Background

- o **Programming**
It is assumed students are comfortable writing programs in Java or equivalent. While there is not a large amount of programming involved, students should feel comfortable doing so.
- o **Communication skills**
Students should be able to express them verbally and in writing.
- o **Internet and SQL**
Students are assumed to be familiar with basic internet tools such as web searching, sending and receiving email messages. Familiarity with database-backed websites and SQL is beneficial, but a primer is provided for these activities.

Course Materials

There are no books to purchase. We will provide you with copies of the materials you will need.

The basic text this term will be Prof. Sweeney's manuscript:

The Science of Privacy

http://privacy.cs.cmu.edu/courses/pad1/schedule.html

Week No.	Date	Description	Handouts	Assignment Due
1	9-2, 9-4	Overview of field and course (PDF); Lab 1	see list	
2	9-9, 9-11	Human identification: SSN, identity theft; Lab 2	see list	Lab 1 assign
3	9-16, 9-18	Legal overview of Privacy Laws and Regulations, US and Europe. Guest lecture by Michael Shamos. Ethical overview; Lab 3	see list	Lab 2 assign
4	9-23, 9-25	Case study: face recognition	see list	Lab 3 assign
5	9-30, 10-2	Ubiquitous tracking: video surveillance, sensor networks	see list	Lab 4 assign
6	10-7, 10-9	Identifiability of data: data explosion, demographics; Lab 6	see list	Lab 5 assign, Project I
7	10-14, 10-16	Semantic learning: direct linkage, probabilistic linkage, iterative profiling, trails; Lab 7	see list	Lab 6 assign
8	10-21, 10-23	Distortion techniques: Formal protection models; Lab 8	see list	Lab 7 assign, Challenge
9	10-28, 10-30	Ubiquitous data sharing: bio-terrorism surveillance, counter-terrorism surveillance; Lab 9	see list	Lab 8 assign
10	11-4, 11-6	Statistical disclosure control: record linkage; privacy-preserving data mining; Lab 10	see list	Lab 9 assign, Project II
11	11-11, 11-13	Policy specification and enforcement: digital rights management, PIP, EPAL; Lab 11	see list	Lab 10 assign
12	11-18, 11-20	Internet privacy: SPAM; Personal information capturing tools: email, personal diaries, position location technologies; Lab 12	see list	Lab 11 assign
	11-23	Protecting textual documents	see list	Lab 12 assign
	12-2, 12-4	Project presentations: tentative data and time		Project presentation
	12-16	NO Final Exam, but final papers due		Project report

Emerging Technologies with Privacy Concerns

1. Face recognition, Biometrics (DNA, fingerprints, iris, gait)
2. Video Surveillance, Ubiquitous Networks (Sensors)
3. Semantic Web, Data Mining, Bio-Terrorism Surveillance
4. Professional Assistants (email and scheduling), Lifelog recording
5. E911 Cell Phones, IR Tags, GPS
6. Personal Robots, Intelligent Spaces, CareMedia
7. Peer to peer Sharing, Spam, Instant Messaging
8. Tutoring Systems, Classroom Recording, Cheating Detectors
9. DNA sequences, Genomic data, Pharmaco-genomics

Some Privacy Technology Solutions

- Face de-identification
- Self-controlling data
- Video abstraction
- CertBox (“privacy appliance”)
- Reasonable cause (“selective revelation”)
- Distributed surveillance
- Privacy and context awareness
- Data valuation by simulation
- Networks of people
- Video and sound opt-out
- Text anonymizer
- Privacy agent
- Blocking devices
- Query restriction

