Privacy Issues in Big Data

Xiang-Yang Li， Lan Zhang
School of Computer Science and Technology, USTC
Professor, CS Department
2021

Today's Class

- Course Logistics
- A little bit about us
- A little bit about this course

Course Logistics

Course Information

- Instructor:
  - XiangYang Li 李向阳
  - Lan Zhang 张兰
- Classroom GT-B110
- Class-time 每周三晚上 7:30PM to 9:55PM

Textbook and Reading List

- **No specific textbook**
  - Big Data/economics/privacy issues are relatively new topics (so no fixed syllabus)
- **Reading List**
  - We will cover the state-of-art technology from research papers in big conferences
  - Many related papers are available on the course website
  - Check related conferences or journals

    ACM CCS, USENIX Security, IEEE Security and Privacy and so on
- **Related books:**
  - Bee-Chung Chen, Daniel Kifer, Kristen LeFevre and Ashwin Machanavajjhala.“Privacy-Preserving data publishing.” Foundations and Trends® in Databases, Vol. 2, Nos. 1–2 (2009) 1–167.
  - Dwork, Cynthia, and Aaron Roth. "The algorithmic foundations of differential privacy." Foundations and Trends® in Theoretical Computer Science 9, no. 3–4 (2014): 211-407.
  - https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf
  - Goldreich, Oded. Foundations of cryptography: volume 2, basic applications. Cambridge university press, 2009.
  - Rosulek, Mike. "The joy of cryptography." Oregon State University EOR (2020): 1.

Requirements & Grading

- **Seminar-Type Course**
  - Students will read research papers and survey them
  - Defense and Offense in discussions
- **Hands-on Course**
  - Coding project covering the entire semester
  - Survey of special topics

Grading

- **Class attendance (15%)**

  attendance 10%

  active and asking questions 5%

- **One term survey paper (35%)** (**Each** term is expected to read papers from a chosen topic and write a survey/summary on this topic.)

- **One term project (50%)** (the term project is formed by a team of three students). This project is about programming and implementation related to big data privacy--using real data, real application

  1) project proposal (10%)

  2) project preparation and presentation (10%) (10 mins per group)

  3) final project report (15%) (the project report need to report your methods, novelty, and results. Report need to follow IEEE/ACM conference format, and has at least 10 pages.)

  4) project code and demo (15%) (at end of semester, demo to TA)

Term Survey Paper

b) One term survey paper (35%) (Each term is expected to read SEVERAL papers from a chosen topic and be able to write a survey about the topic)

- The selection of the topic/paper from the list is first-come-first-service. No TWO students are allowed to select the SAME topics. You cannot COPY any material from any segment of results written by others (online material or published books, papers, reports), unless you need to cite some results or statements and clearly indicate in your report.
- The survey paper should be due at the end of the semester. The paper should be at least of 10 pages and in IEEE conference format, double column, font size (at most) 10,
- You can write it in Chinese or English. Need to have enough figures and formulas.

Term Project

c) One term project, presentation, and report (50%)

- We provide a list of possible projects. You will select one project. AT most two groups can select the same project. You can also propose project and ask for our approval.
- You have to really implement the project and show that it works.
- Each group needs to discuss the term project with the instructor or TA within 3 weeks of the first lecture. Each group needs to submit a 2 page project proposal by the end of the 4rd week. Submit your e-copy to TA, and hard-copy also to TA.
- Each student in the group will be graded equally unless it was reported to me and confirmed that some student did not do sufficient work for the project.
- Each group needs to do one 10 min presentations, at the end of the semester to demo the final results of your project.

Choosing Projects

- Pick a problem that is intellectually interesting And improves the practice.
- 更加重要的是：你们能够完成，别人感兴趣

Look for blind spots

- Question old school assumptions
- Open your heart and mind to people who question assumptions

Project Proposal

Components of your proposal

- Your project title
- Team members
- Challenges in your project
- Relation to the topic: privacy issues
- Current literature on this topic
- What are the data?
- Your evaluation plan and metrics (how to evaluate the success of your project)
- Management aspects such as your project plan, critical paths, means of team communication (e-mail, chat room, meetings, version control system).

Project Presentation

Covers the following material:

- Explain your design.
- Discuss design alternatives,
- Summarize privacy aspects of your project,
- Your system architecture.
- Your method or algorithms,
- Data to show the performance of your systems,
- The challenges faced by your group in implementing the project and how you address these challenges;
- Lessons learned from the project, and
- future plan for the project.

Project Report

Project report: Covers the following material: (at least 8 pages, IEEE Format, font size at most 10, two columns)

- Abstract, Introduction, ….
- Your system architecture: Explain your design, and compare with the literature; Discuss design alternatives,
- Your detailed method or algorithms and the technical challenges and how you address these challenges;
- Performance Data to show the performance of your systems,
- Discussion and Conclusion,
- References

Important Dates

Project:

- Project Proposal: in three weeks (3 students per group)
- Project Presentation: three weeks before the end of semester (10 mins per group)
- Project Code and Demo: end of semester (to TA)
- Project Report: end of semester (last week, to TA)
- Survey Paper: end of semester (To TA, need e-copy)

Classroom Policy

- Each of you is expected to contribute to each class session by arriving on time, being attentive, participating in the class discussion if needed, and being respectful to your instructor and fellow students.
  - Disruptive conversations, eating, sleeping and putting your feet on the furniture are not acceptable behavior in the class environment.
- In addition to arriving on time, students are expected to stay the whole class period.
  - Please avoid disrupting fellow students and the instructor by arriving late or leaving early. If a situation arises that consistently causes you to be late or absent, please contact me.
- Every electronic device (anything with an on/off button) should be off during the class (exception: disability-helping devices).

What you will learn

- Big Data and Privacy Issues:
  - Big Data implies big privacy leakages
  - Privacy issues in big data

    collection, storage, computing, visualization

About Instructor

Who I am and What I do

Prof. XiangYang Li

李向阳

http://staff.ustc.edu.cn/~xiangyangli

xiangyangli@ustc.edu.cn
xiangyang.li@gmail.com

ACM Fellow, IEEE Fellow, ACM Distinguished Scientist

School of Computer Science and Technology

USTC

Instructor

- **Info**
  - Professor, School of Computer Science and Technology, USTC
  - Professor, CS Department, IIT
  - IEEE Fellow, ACM Distinguished Scientist 2015
  - PhD/MS from UIUC 1997-2000
  - BS, BE Tsinghua University 1990-1995
- **Research Interest**
  - Wireless networks, mobile computing
  - Big data security and privacy
  - Internet of Things, Cyber Physical Systems
  - Algorithm design and analysis, Game theory
- Supported by MIST, NSF, NSF China, RGC HongKong
- Contact Information
  - Email: xiangyang.li@gmail.com
  - 计算机学院，电三楼6楼 627，中国科学技术大学

Who I am and What I do

Lan Zhang
张兰

zhanglan@ustc.edu.cn

School of Computer Science and Technology
USTC

Instructor

Info

- Researcher, School of Computer Science and Technology, USTC
- Poster Doctor, Tsinghua University 2014-2016
- PhD, Tsinghua University 2007-2014
- BS, Tsinghua University 2003-2007

Research Interest

- Big data understanding, protection and trading
- Privacy Protection
- Mobile computing

Award

- 2015 ACM China Doctoral Dissertation Award (1/2 nationally)
- CCF Outstanding Doctoral Dissertation Award (1/10 nationally)

Contact Information

- Email: zhanglan@ustc.edu.cn
- 科技实验楼105，中国科学技术大学

Learning Objectives

1) Learn the classic and state-of-the-art data privacy/security approaches

2) Study various applications where data privacy/security is needed and can be applied

3) Challenge existing solutions and identify new problems in data privacy and security

Some expectations

1) Participate in class, think critically, ask questions

2) Read and write reviews critically

3) Start on assignments and projects early

4) Enjoy the class!