University of Washington

Department of Electrical Engineering

EE 595 Advanced Topics in Communication Theory –

Algorithmic Introduction to Data Privacy

Spring 2017

Time: Wednesdays from 6:00-9:30pm in EEB 003

(**Voluntary** discussion about the assigned papers from 9-9:30pm)

Instructor: Tamara Bonaci (tbonaci@uw)

Office hours: By appointment

Teaching Assistant: TBD

Office hours: TBD

Course website: https://canvas.uw.edu/courses/1138434

Course assignments and dropbox: https://canvas.uw.edu/courses/1138434/assignments

Course discussion board: https://canvas.uw.edu/courses/1138434/discussion_topics

Course gradebook: https://canvas.uw.edu/courses/1138434/gradebook

Course mailing list: ee595b_sp17@uw.edu

Course Overview:

Consider the following scenario: You are an administrator of a large data set at a hospital, an internet service provider, a search engine or a social network. The data you hold is very valuable, and you would like to make it publicly available so that you and the rest of the world can make a better use of it. However, the data is also highly sensitive (e.g., consisting of patients' medical records, users' browsing records, etc.)! Moreover, the availability of fast and cheap computing resources, coupled with massive storage devices has enabled the collection and mining of data on a scale previously unimaginable. This opens the door to potential abuse regarding individuals' information. So, even though you are only planning to release aggregated data (and you were never planning to abuse that data), you must do so in a way that does not compromise the privacy of any individual in the data set.

In recent years, there has been considerable research exploring the tension between utility and privacy in the presented context, and this course provides an introduction to that work. Its goal is to explore techniques and issues related to data privacy. In particular, we will: define what is data privacy, introduce different techniques for achieving privacy, analyze modern anonymization techniques and possible attacks on data anonymization. We will then investigate privacy-preserving data mining method. We will also focus of differential privacy, and means for achieving it. Lastly, we will consider some limitations (i.e. lower bounds) on privacy in various settings.

Course Progression:

The following is the class progression covering the 10 weeks for the course. The class will meet once a week on Wednesdays from 6:00-9:30pm.

Week 1: Course overview. Introduction to data privacy and anonymization.

Week 2: Static data anonymization I – multidimensional data.

Week 3: Static data anonymization II – graphs and time series data.

Week 4: Privacy preserving data mining.

Week 5: Synthetic data generation.

Week 6: Threats to data anonymization. Introduction to differential privacy.

Week 7: Differential privacy I – oblivious transfer, secure function evaluation, differential privacy and impossibility proofs.

Week 8: Differential privacy II – Laplace and exponential mechanisms.

Week 9: Differential privacy III – composition theorems.

Week 10: Emerging topics – polymorphic encryption and pseudomization.

Finals week: Project presentations

About the Course:

The course will consist of ***readings and discussion***, ***homework assignments, and a project.***

Readings and Discussion:
Readings and discussions of the assigned papers are an important part of this course. The goal of this exercise is to work together on understanding broader implications of the seminal research. Additionally, we will also try to keep up with the state-of-the-art security and privacy research.

Prior to each class a research paper will be assigned, and you will be expected to post to the class discussion board about it. Your post should contain something original beyond what others have already posted. You may consider posting:

- A summary of the paper,
- Evaluation of the paper's strengths and weaknesses,
- Open research question on the topic, or
- Question you would like to discuss in class.

All post are due by **4pm on the day of each class,** and they will be graded on the scale of 0-2, where:

- 0 means missed or irrelevant post,
- 1 means a relevant post submitted, and
- 2 means a good and interesting post.

Post submitted after 4pm on the day of the class will receive no credit. There will be a total of nine reading assignments, and we will take eight best scores when determining your grade.

Homework:
There will be **four** homework assignments, related to the presented anonymization and differential privacy material. The assignments will be a mix of written questions and simulation problems, and you will have approximately two weeks for each assignment.

While coding/simulating parts of the homework may be designed with Matlab in mind, you are welcome to code them up using any software tool you prefer (e.g., Mathematica, Octave, Maple, Python). **You should, however, submit all your code and simulation models with your homework.**

Project:
The final component of this course is a project, and its goal is to give you a deeper understanding of how to think about, and how to solve a real-life problem from a security and privacy-oriented perspective.

For the project, you will be able to choose a topic related to any area of security and privacy (including those not directly covered in this course). You can work on the project either individually, or in groups of up to three persons. When working in a group, your end result should reflect the fact that it is a multi-person effort.

Your work on the project will consists of several milestones:

- Project proposal,
- Progress report,
- Final report, and
- Project presentation.

Grading:

Your grade in this course will be based on readings and discussion, homework assignments, and project. The expected grade breakdown is:

- Readings and discussion – 15%
- Homework – 40%
- Project – 45%

Course Material:

We will use two textbooks for this course:

- N. Venkataramanan, and A. Shriram, ***Data Privacy: Principles and Practice***. CRC Press, 2016.
- C. Dwork, A. Roth, ***The Algorithmic Foundations of Differential Privacy*****,** Foundations and Trends in Theoretical Computer Science: Vol. 9: No. 3–4**,** Now Publishers, 2014**.**

In addition, you may find the following books useful:

- D. Stinson*, Cryptography Theory and Practice, Third Edition,* CRC Press, 2006
- W. Stallings, *Cryptography and Network Security, Principles and Practice, 5th Edition,* Prentice Hall, 2006
- B. Schneier, *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*. WW Norton & Company, 2015.

Course Policies:

Collaboration: In this course, we want you to learn from each other. Therefore, you are allowed (and encouraged) to talk to your classmates and other students about all course assignments. You may also consult outside reference materials, or the instructor. However, all material that you decide to turn in should reflect your own understanding of the subject matter at the time of writing. This means that you should write your own posts about the assigned papers, your own security review, and your own homework. If you work with someone else on any assignment, please include their names on the material that you turn in.

Assignment Turn-in: Posts about the assigned papers should be submitted using the course discussion board. All other material (security reviews, homework, and project-related material)

should be submitted in a PDF form to the course dropbox. Please, ***do not use*** email for assignment submissions.

Late Assignment Turn-in: Discussion board posts are due **by 4pm on the day of the class**, and no late turn-ins will be accepted. All other assignments are due **by 11:59pm on the assigned date,** but we understand that you may have to sometimes turn them in late. The grading penalty is 20% of the grade that you would otherwise receive for each day, or part of the day, that you are late. No submissions will be accepted after 5 days.

Checking grades: Grades will be posted to the course gradebook.