KOÇ UNIVERSITY

Data Privacy and Security

Course Introduction

M. Emre Gürsoy

Assistant Professor
Department of Computer Engineering
www.memregursoy.com

Today's Plan

- Course logistics
  - Going through the course syllabus
  - Description of course policies and grading
  - Q&A about course logistics
- Motivation and sneak peek
  - Why study data privacy and security?
  - What are you going to learn in this course?

Recording Disclaimer

- TL;DR: **The lectures are recorded**

The synchronous sessions are recorded (audiovisual recordings).

The audiovisual recordings, presentations, readings and any other works offered as the course materials aim to support remote and online learning. They are only for the personal use of the students. Further use of course materials other than the personal and educational purposes as defined in this disclaimer, such as making copies, reproductions, replications, submission and sharing on different platforms including the digital ones or commercial usages are strictly prohibited and illegal.

The persons violating the above-mentioned prohibitions can be subject to the administrative, civil, and criminal sanctions under the Law on Higher Education Nr. 2547, the By-Law on Disciplinary Matters of Higher Education Students, the Law on Intellectual Property Nr. 5846, the Criminal Law Nr. 5237, the Law on Obligations Nr. 6098, and any other relevant legislation.

The academic expressions, views, and discussions in the course materials including the audio-visual recordings fall within the scope of the freedom of science and art.

Course Description

- This course will cover:
  - Major developments in statistical data privacy and security in the past two decades
  - Some of the state-of-the-art privacy-enhancing technologies (PETs) used in the digital world
  - Recent research trends in privacy and security

- The course is intended to provide:
  - **Breadth** through lectures + HWs + midterms
  - **Specialization (depth)** through the course project

- Offered simultaneously to grads (**COMP530**) and undergrads (**COMP430**)

Instructor

M. Emre Gürsoy
Assistant Professor @ Koç University
www.memregursoy.com

PhD in Computer Science, 2020, Georgia Tech
MS in Computer Science, 2015, UCLA
BS in Computer Sci. and Engr., 2013, Sabanci Uni.
High school, 2009, Robert College

Research Areas: Data privacy, security, data analytics and mining, adversarial machine learning, security&privacy in AI

- > 45 journal, conference and workshop publications, 3 US patents
- Frequently invited reviewer for intl. IEEE/ACM journals and conferences
- Several awards and grants (EdgeSys 2020 Best Paper Award**,** TUBITAK CAREER Grant in 2021, etc.)

Logistics

- **Physical (on-campus)**
  - Tue + Thur @ 2:30-3:40 PM
  - SNA Building, room: A52
- **Lecture recordings will be released via Panopto**
- **I will open Zoom but not interact with Zoom participants**

- No labs, PS or DS sections
- TAs and office hours will be announced soon
  - My office: SNA Z27, e-mail: emregursoy@ku.edu.tr

Attendance Policy (Fall 2024)

- You should follow lectures by physically coming to class
- I will open Zoom but not interact with Zoom participants
  - Do **not** rely on Zoom as a regular lecture participation option
- Recordings will be made available via Panopto
  - Do **not** rely on Panopto recordings alone for learning
- Few bonus points for attendance
  - Especially in the second half of the semester, when attendance is low or has been decreasing
  - I may collect attendance regularly or irregularly (unannounced)

Student Background

- **Pre-requisites (will be enforced)**
  - Undergrads: COMP202 (Data Struct. & Algorithms)
  - Grads: None, but equivalent knowledge is expected
- **Strongly recommended**
  - Maturity in Computer Science-related concepts
    - COMP106: Discrete Math, ENGR200: Probability
    - Statistics, algorithms, databases, AI/ML … are beneficial
    - *(If you studied COMP in your undergrad, you should be OK)*
  - Programming proficiency in Python
    - Not heavy programming, but homework assignments are in Python
    - Writing few hundred lines of Python code should not be difficult for you
    - *(If you studied COMP in your undergrad or you are used to coding in Python, you should be OK)*

List of Topics

- Introduction and preliminaries
- Data anonymization
- Differential privacy
- Local differential privacy
- Machine learning for security
- Security for machine learning
- Authentication mechanisms
- Database security (access control, SQL injection)
- [tentative] Principles of modern privacy laws and regulations, e.g., GDPR, KVKK

List of Topics

- Although highly related to data privacy and security, I **aim not to overlap** with material that is already covered in other courses:
  - Modern Cryptography
    - COMP 443/543
  - Computer and Network Security
    - COMP 434/534

Grading (Tentative)

Midterm Exams	50%
Course Project	25%
Homework Assignments	22%
Peer Reflections	3%

- I typically offer some extra credit opportunities
  - Few extra points on midterms and/or HWs
  - Exceptionally good course projects

Midterm Exams

- We'll have two face-to-face midterms
  - 2 * 25% = 50%
- Scheduled by the Registrar's Office
  - Roughly by the end of week 6-7 and week 12 or 13
- Types of questions:
  - Mixture of T/F, multiple choice, short answer (conceptual), math problems, small proofs, …

Average of your two midterms must be ≥ 30 to pass the course.

Midterm Coverage (Tentative)

- Introduction and preliminaries
- Data anonymization
- Differential privacy
- Local differential privacy

MIDTERM 1

- Machine learning for security
- Security for machine learning
- Authentication mechanisms
- Database security

MIDTERM 2

- [tentative] Principles of modern privacy laws and regulations, e.g., GDPR, KVKK

Homework Assignments

- Tentatively 4-5 HW assignments (expected: 4)
  - Mostly programming in Python + some verbal and math questions, reporting experiment results, etc.
  - You get around 2 weeks for each HW
    - **Start early!**

- Weights of HWs will be different
  - HW1 is more difficult, hence higher weight (~8-9%)
  - HW4 is less difficult, hence lower weight (~4%)

HOMEWORK ASSIGNMENTS MUST BE COMPLETED INDIVIDUALLY.

USING CHATGPT OR SIMILAR AI TOOLS IS FORBIDDEN.

HW Coverage (Tentative)

- Introduction and preliminaries
- Data anonymization

HW 1

- Differential privacy
- Local differential privacy

HW 2

- Machine learning for security
- Security for machine learning

HW 3

- Authentication mechanisms
- Database security

HW 4

- [tentative] Principles of modern privacy laws and regulations, e.g., GDPR, KVKK

Late Homework Policy

- For late submission of HWs:
  - Up to 10 mins late: -5% penalty
  - 10 mins to 1 hour late: -20% penalty
  - > 1 hour late: not accepted

- Exceptions: emergencies (e.g., medical) with proper documentation (accepted by KU Health Center)
  - As much as possible, let me know ahead of time

- Do not ask for extensions close to a deadline (or after a deadline) – I want to be fair to all students

Course Project

- Goal is for you to specialize in a privacy or security-related topic of your choice (total worth: 25%)
  - Project proposal (mid-semester): 5%
  - In-class presentation (last 2 weeks): 5%
  - Final report and deliverables (finals week): 15%

- We are looking for research-style projects
  - Should go beyond what is taught in lectures
  - You should read, learn, experiment by yourself
    - E.g.: read research papers, articles, find and run code…
  - Novelty in terms of idea, design, development, experimentation, etc. is expected

Course Project

- Project topics:
  - You may choose freely, as long as it's related to the course
  - You may be influenced by lectures, HWs, your personal interests, ...
  - You're encouraged to discuss your ideas with me
  - I will provide a list of sample project topics
- Can be done in groups
  - Min and max group size TBD
  - Last year: ~80 students, max: 5 students/group, this year: ??
  - Larger group -> expectation is higher
- Course projects should be fun and enjoyable for you
- Promising projects may be extended to future semesters for Independent Study, publications, etc.
  - Talk to me if you are interested

Peer Reflections

- 3% of your course grade
- Relevant only during project presentation weeks

- Pick two presentations from that day, write a peer reflection for each
  - One paragraph summary of the presentation (5-6 sentences)
  - Strengths (2-3 bullet points)
  - Weaknesses (2-3 bullet points)
  - Each peer reflection is approx. half page long

- **Goal:** Learn about other groups' work, learn how to critically analyze others' work and give feedback

Exam and Make-Up Policy

- If you miss a midterm with a **valid excuse**, you can apply for a make-up
  - Valid excuse = accepted by KU and sent to me via official channels (e.g., Health Center, DoS)

- There will be a **single, joint make-up exam** at the end of the semester
  - Contains all topics
  - Your grade counts in place of the exam you missed
  - Friendly advice: **take the actual midterms** ☺

Academic Honesty

- You must work alone in everything other than the course project
  - Generative AI is also forbidden in HWs

- In the course project, you can use external code or libraries, but you must acknowledge them explicitly
- Generative AI can be used only in the final report of the project, and only **to improve your writing**

- I report cases of cheating and plagiarism directly to the University Disciplinary Committee

Consent Requests

- This is (used to be) a research-oriented course intended for 30-40 students who are enthusiastic about data privacy and security
  - **Is this you?**
  - **Will you actually do an interesting course project?**

- We have too many students + more on the waitlist

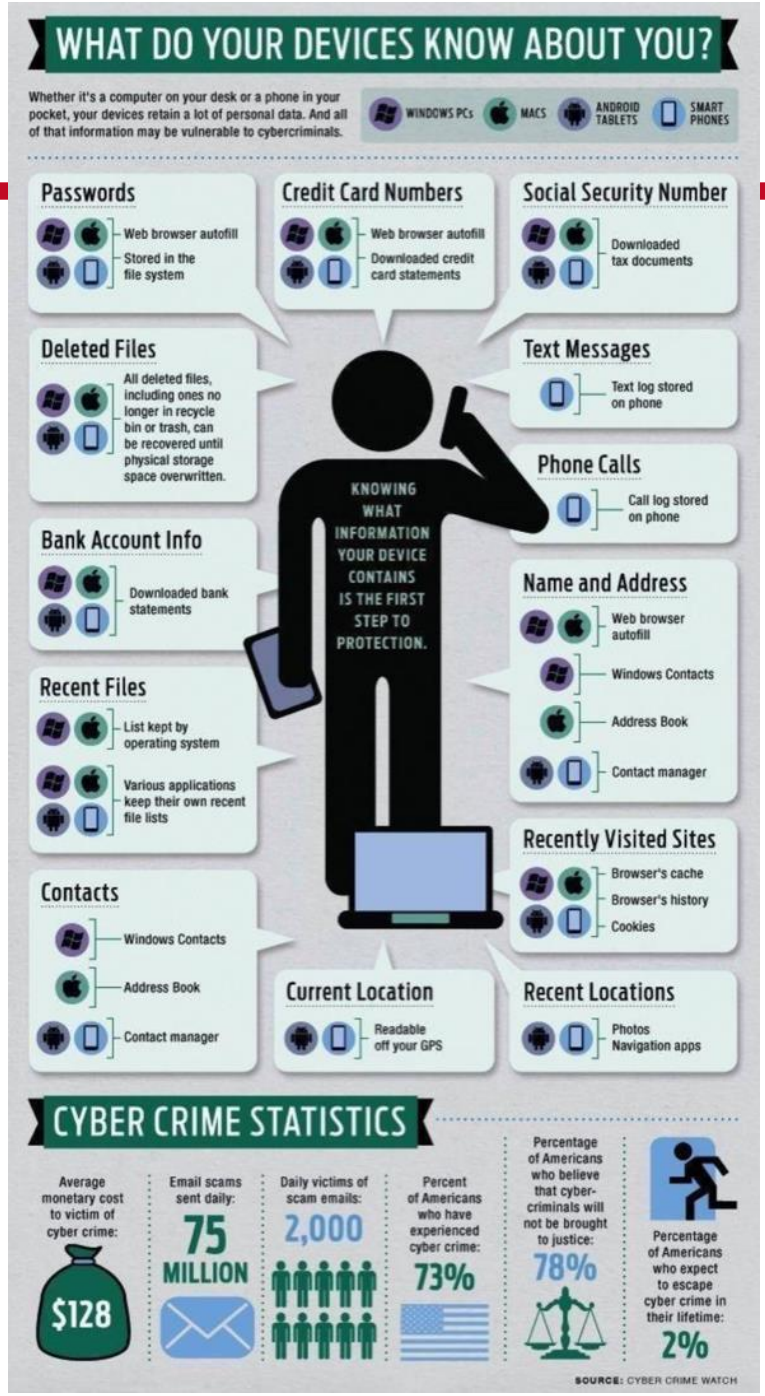
- We can't increase capacity much further
  - We are already way above what I'd like it to be
  - Resources are limited (classroom size, TAs, …)

Done with the logistics part...

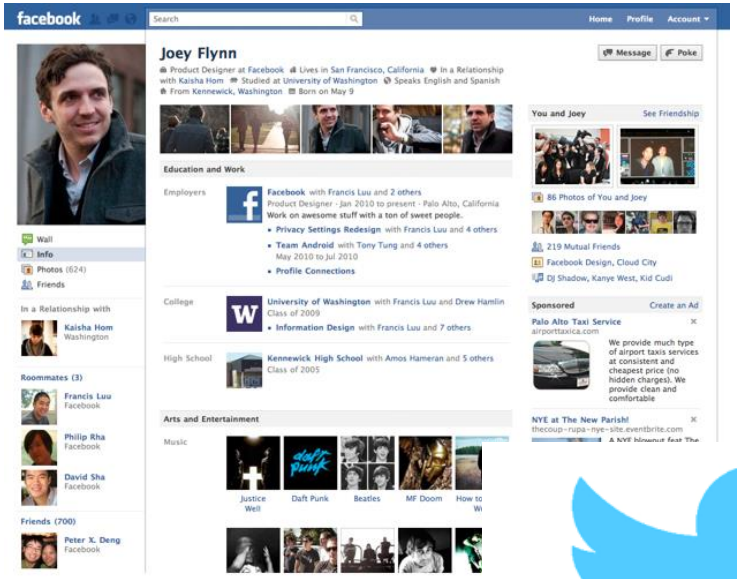
Questions?

The Digital Age

- What data do your devices have about you?
  - A lot ☺



The Digital Age



twitter

LinkedIn

- What data do social media sites have about you?
  - Friends, pictures, likes, interests, hobbies, personal life, work life, opinions and political views, ...

Value of Big Data

- Market transactions and purchase histories
- Tracked, e.g., via rewards cards

Your Data

- There's data about you everywhere!
  - We sometimes give our data voluntarily, to receive a certain service
  - Sometimes data collection is implicit
- There's great benefit in turning data into value

Is it worth studying data privacy and security, or should we just let it go?

We Need Privacy

- From an ethical perspective
  - Basic human need
- From a law perspective
  - GDPR, KVKK, ...
- From a business perspective
  - Customer trust, loyalty, company brand value

When Things Go Wrong...

- AOL search log release scandal
  - In 2006, AOL releases “anonymized” search engine logs
  - NY Times finds the identity of certain users from anonymized AOL logs
  - Class action law suit against AOL
  - AOL’s CTO + 2 other people fired

Thelma Arnold
Lilburn, GA
62-year-old widow

Search queries of user #4417749

“dog that urinates on everything”
“mature living”
“60 single men”
“landscapers in Lilburn, GA”
“… Arnold”

Sneak Peek

- Data anonymization
  - Also features in KVKK's Guidebook for "Erasure, Destruction or Anonymization of Personal Data"

Zip	Age	Nationality	Disease
13053	28	Russian	Heart
13068	29	American	Heart
13068	21	Japanese	Flu
13053	23	American	Flu
14853	50	Indian	Cancer
14853	55	Russian	Heart
14850	47	American	Flu
14850	59	American	Flu

Zip	Age	Nationality	Disease
130**	<30	*	Heart
130**	<30	*	Heart
130**	<30	*	Flu
130**	<30	*	Flu
1485*	>40	*	Cancer
1485*	>40	*	Heart
1485*	>40	*	Flu
1485*	>40	*	Flu

Sneak Peek

- Differential privacy (used by Uber, American Census)
- Local differential privacy (used in Apple iOS, Google Chrome, Windows 10)

Apple's 'Differential Privacy' Is About Collecting Your Data—But Not *Your* Data

Microsoft Research Blog

Collecting telemetry data privately

December 8, 2017 | By Bolin Ding, Researcher; Jana Kulkarni, Researcher; Sergey Yekhanin, Sr Principal Researcher

Windows 10

Sneak Peek

- Threats to modern AI/ML-based systems
  - Training-time and test-time attacks
  - Potential defenses

Skewing Gmail's spam filter using fake spam/non-spam reports

A data poisoning attack

Sneak Peek

Yemeksepeti

Değerli Kullanıcımız;

Siber saldırılar bugün devletlerden, en büyük şirketlere; ünlülerden, finans kurumlarına kadar pek çok veri kaynağını tehdit etmekte olup, çağımızın en büyük suçlarından ve güvenlik sorunlarından biri haline gelmiştir. Dünyada, siber saldırılara karşı %100 güvenli bir sistemden söz etmek mümkün olmasa bile, Yemeksepeti olarak veri güvenliğiniz adına bugüne kadar elimizden gelen tüm önlemleri aldığımızı vurgulamak isteriz.

Ancak, şu an sizlere bu konuda yaşanan bir gelişme üzerine, ilk ağızdan net bir bilgilendirme yapmak üzere ulaşıyoruz.

25.03.2021'de sabah saatlerinde tespit ettiğimiz üzere, Yemeksepeti kullanıcı veri tabanı, kimliği tespit edilemeyen siber korsan ya da korsanlar tarafından bir saldırıya uğradı ve bir güvenlik ihlali yaşandı. Sizin de içinde bulunduğunuz Yemeksepeti kullanıcılarının he sap bilgilerinin bir kısmı korsanlar tarafından ele geçirildi.