KOÇ UNIVERSITY

Data Privacy and Security

Introduction to Privacy and Security

M. Emre Gürsoy

Assistant Professor
Department of Computer Engineering
www.memregursoy.com

What is Privacy?

- Extremely overloaded term, hard to define

> "Privacy is a value so complex, so entangled in competing and contradictory dimensions, so engorged with various and distinct meanings, that I sometimes despair whether it can be usefully addressed at all."

Robert C. Post, *Three Concepts of Privacy*, 89 Geo. L.J. 2087 (2001).

- People have different opinions and preferences
- Multi-disciplinary: COMP, MATH, Law, Social Sciences
  - Within COMP: Databases, cybersecurity, AI/ML
- Has to be continuously updated due to changing technology, algorithms, ...

What is Security?

- **Wikipedia**: "... the protection of computer systems … from information disclosure, theft of or damage to their hardware, software, or electronic data, as well as from the disruption or misdirection of the[ir] services …".

- **Why should you care?**
  - Consumer devices (smartphones, laptops, …)
  - Large business corporations
  - Financial systems
  - Transportation systems: automobiles, aviation, …
  - … and many more rely on computer systems!

Right to Privacy (1890)

- "The Right to Privacy" is an article written by Warren & Brandeis, published in **Harvard Law Review**
  - Regarded as the first publication in the United States to advocate a right to privacy

- Some influential ideas:
  - Privacy protection needs to be a **fluid** principle that is reshaped according to political, social, and economic change
  - **"The right to be let alone"** - individuals should have the freedom to conduct their affairs without unwarranted publicity
  - **Privacy rights ≠ property or intellectual property rights**
  - Laws should recognize privacy violations have implications beyond physical harm, such as emotional or psychological harm
  - Limitations on freedom of press and journalism (e.g., newspapers)

Privacy Act of 1974

- Federal law in the US, developed in response to growing concerns about government surveillance and use of personal data

> "No agency shall disclose any record … to any person, or to another agency, except … with the **prior written consent** of the individual to whom the record pertains, unless disclosure of the record would be … used solely as a statistical research or reporting record, and the record is to be transferred in a form that is **not individually identifiable** …"

- Exceptions: for statistical purposes (Census Bureau, Labor Statistics), law enforcement purposes, etc.

PII

- What is “personally identifiable information”? (PII)
  - **Any information that can be used to distinguish or trace a person’s identity**
- Passport number, national ID number, biometrics (fingerprint, retina, x-rays, etc.), address, phone number
- Can be sector-specific: patient ID number (in a hospital), credit card number (in a bank)

- The challenging part: “information when combined with other information to collaboratively identify a specific individual should also count as PII”
  - Date of birth, place of birth, gender, race, …

Privacy Act of 1974

- Limits on record disclosure and sharing
  - Unless well-defined exceptions
- Right of individual access and amendment
  - **Access:** "Upon request by any individual … permitted to review the record in a form comprehensible to him …"
  - **Amendment:** request corrections if the record is inaccurate
- Public notice requirement
  - Agencies must publish the details of all their systems of records in a federal register
  - No "secret databases"
- Individuals can sue; federal agencies and officers can receive criminal penalties

Sector-Specific Regulations

- Individuals were particularly worried about specific sectors which were known to store sensitive personal data

- Education sector: Family Educational Rights and Privacy Act (FERPA) – 1974

- Health sector: Health Insurance Portability and Accountability Act (HIPAA) - 1996

- Financial sector: Gramm-Leach-Bliley Act (1999)

FERPA

- Parents have right to access their children's education records, the right to have the records amended, and the right to "control" the disclosure of information

- When the student turns 18, **FERPA rights transfer from the parents to the student**

- Promotes transparency and integrity for schools
  - Especially primary, middle, and high schools
- Consent required to disclose PII to third parties
- Again, some exceptions apply
  - E.g.: health or safety emergency

HIPAA

- "Personally identifiable health information" is a subset of all health information, including demographic information (gender, date of birth, etc.) and:
  - Relates to physical or mental health conditions of an individual
  - Healthcare services given to an individual
  - Payment for healthcare services by an individual
  - … with respect to which there is a reasonable basis to believe the information can be used to identify the individual

- Appropriate administrative, physical, and technical safeguards are **required** for HIPAA compliance
  - Administrative: internal or external audits, staff training, etc.
  - Physical: employment of security personnel, protection of servers
  - Technical: authentication, intrusion detection, encryption, …

HIPAA

- Right of access
  - What health data do you have about me?
- Disclosure to patients' relatives should be limited
  - Has caused some debate, as people couldn't find their relatives after car and plane accidents
- Notice of privacy practices
  - The need to inform patients about data usage
- Breach notification rule
  - If there is a data breach, affected individuals and authorities must be notified in timely manner

Gramm-Leach-Bliley Act

- Not just privacy, but has privacy-related aspects
  - Aka: “Financial Services Modernization Act”
- **Financial Privacy Rule**
  - Consumers must be given a privacy notice about what data is collected, used, shared, and protected
- **Safeguards Rule**
  - Financial institutions must develop and implement a written information security program, e.g., risk assessment, employee training, securing information systems
- **Pretexting Protection Rule**
  - Pretexting and social engineering must be prevented

Privacy by Design (PbD)

- Originally developed by **Ann Cavoukian**
  - Published in 2009, adopted by many authorities by 2010
  - Influenced the likes of GDPR

- PbD has seven principles:
  1. Proactive not reactive; preventive not remedial
     - Anticipate and prevent problems before they happen, not after
  2. Privacy as the default setting
     - Enabled by default, individual does not have to activate
  3. Privacy embedded into design
     - Available since the initial design phase, not added after on

Privacy by Design (PbD)

- Principles of PbD, continued:
  4. Full functionality – positive-sum, not zero-sum
     - Win-win for all parties
  5. End-to-end security – full lifecycle protection
     - Privacy and security are enforced from the birth of data until its death, throughout all phases
  6. Visibility and transparency – keep it open
     - Visible, verifiable, auditable by interested parties
  7. Respect for user privacy – keep it user-centric
     - Put users' demands and desires above everything else

Criticisms of PbD

- Common criticisms of PbD:
  - PbD is vague and not quantifiable
  - PbD doesn't say much about how to achieve or implement these principles
  - Some principles are impossible to implement
  - Some principles are missing incentives

What do YOU think about PbD?

GDPR

- The **General Data Protection Regulation (GDPR)** is an important component of the EU privacy law
  - Went into effect in May 2018
  - Applies to those who offer goods or services to persons in EU or EEA
    - EEA = European Economic Area (includes some nearby countries such as Iceland, Norway, ..)
    - Service provider's location doesn't matter, as long as it is processing data of individuals located in EU/EEA
  - Became a role model for privacy laws in other countries across the world
    - Turkey – KVKK
    - UK GDPR
    - California Consumer Privacy Act (CCPA)

GDPR Principles

- **Consent** must be requested from data subjects before their personal data can be processed
  - Must be unambiguous, explicit, and easy-to-understand
  - Must be specific: What data will be processed? Why?

- **Data minimization:** data collection should be limited to what is directly relevant and necessary to accomplish a specified purpose
  - **Purpose limitation** -> “data is collected for specified, explicit, and legitimate purposes only”
  - Also, retain the data only for as long as it is necessary for fulfilling that purpose

GDPR Principles

- **Security of processing**
  - Suggests the use of SOTA technical measures such as anonymization, encryption, …

"Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of [data] processing … as well as the risk … the [data] controller and processor shall implement appropriate technical … measures to ensure a level of security appropriate to the risk…"

- **Rights of the data subject**
  - **Right of access** – what data do you have about me?
  - **Right to object** – especially in the context of profiling & marketing
  - **Right to erasure** ("right to be forgotten")
  - …

Fines for Non-compliance

- Ranges from warning and periodic audits to 10-20 million Euros (or up to 4% annual worldwide turnover)
- Depends on several factors
  - Severity: How many people affected? What is the damage?
  - Intention: Negligence or bad intention?
  - Mitigation: Did the company try to address the damage?
  - Precautionary measures
  - History
  - Notification: Were the supervisory authorities notified?
  - Cooperation

- Some examples: https://en.wikipedia.org/wiki/GDPR_fines_and_notices

A Practical Perspective

- We worry about privacy and security when we have something of value and there is a threat to it (i.e., risk that leads to negative consequence).
- Example from the physical world:
  - Thing of value = gold
  - Risk/threat = thieves

Solution:
Lock your
gold in a safe!

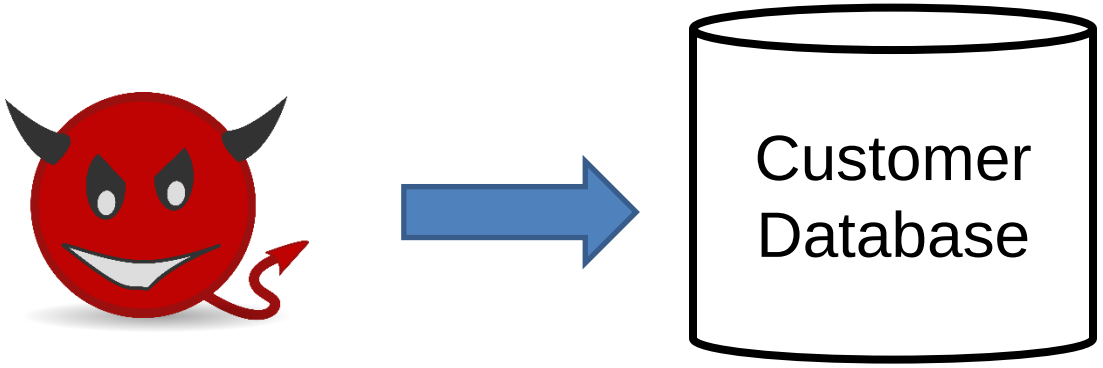
A Practical Perspective

- We worry about privacy and security when we have something of value and there is a threat to it (i.e., risk that leads to negative consequence).
- Another example:
  - Thing of value = building + people living in it
  - Risk/threat = terrorists, thieves, burglars, ...

Solution:
Hire a security guard!

The Digital World

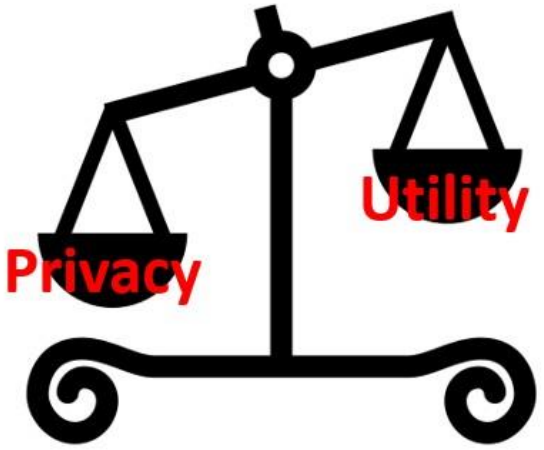
- We worry about **data** privacy and security:
  - Because data has value
  - There are actual threats to data privacy
  - And the threats have negative consequences
    - On us, on our friends/family, our company, our brand, …



Credit card numbers, purchase histories, ...

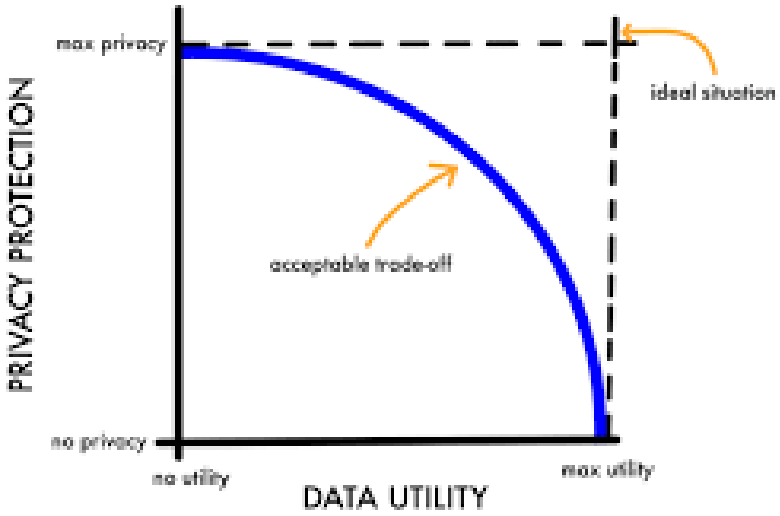
Trade-offs

- There are various trade-offs in enforcing privacy/security protection mechanisms.
- Aspects that are commonly traded-off:
  - **Efficiency** – e.g., encryption vs no encryption
  - **Accuracy** – e.g., noise addition
  - **Usability** – e.g., multi-factor authentication
  - **Financial cost** – e.g., you have to pay security experts



Managing Trade-offs

- "Good" solutions that get deployed in practice are often those that offer acceptable trade-offs.

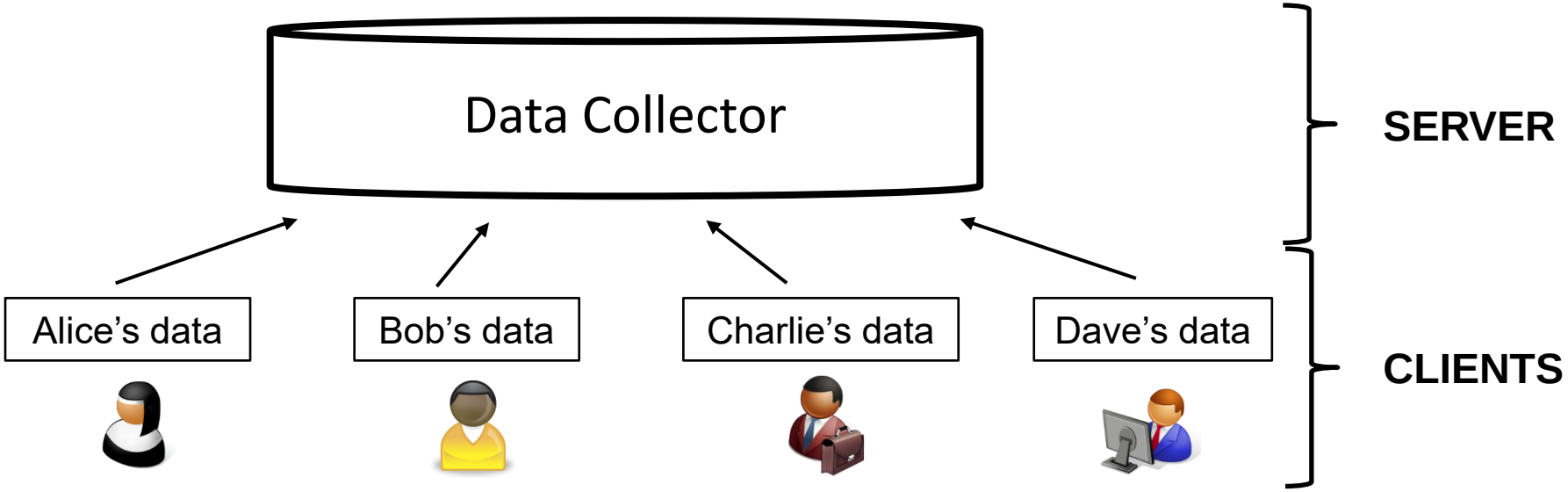


- An acceptable trade-off in one setting may be unacceptable in another:
  - Would you get a $100 safe to protect $10 gold? How about $100 gold? $1k? $10k? $10m?
- Trade-offs can be subjective

Modelling Adversaries

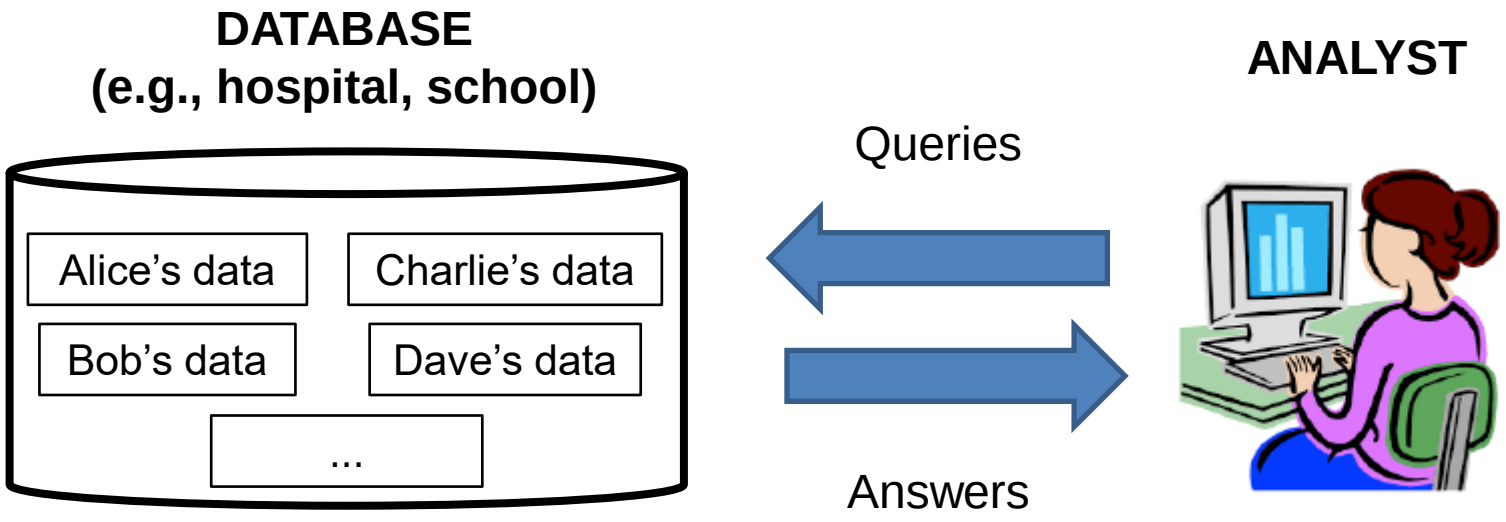
- **Real-world systems need multiple privacy and security protection mechanisms**
  - Different steps of the data management pipeline
    - Data collection, storage, analysis, sharing
  - Different threats and adversary models
    - Active vs passive adversaries
    - Adversaries may have varying resources (i.e., computational power, background knowledge, etc.)
- One mechanism does not solve all problems
  - You may encrypt client-server communication, but it doesn't protect against an attack on the server

Data Collection



- What are some potential adversaries/threats?
  - Honest-but-curious data collector
  - Man-in-the-middle
  - Malicious clients

Data Analysis



- What are some potential adversaries/threats?
  - Lack of authentication
  - Lack of authorization

A Practical Recipe

1) What setting are we operating in?
   - Data collection? Data analysis? Data sharing?
2) What is the valuable information we must protect?
   - Passwords? Locations? Health records?
3) Who is the potential adversary and what is the threat model?
   - Power and capability of the adversary
4) What solutions can be applied to this problem?
   - Can you make it similar to something you've seen before (or you'll see as part of this class)?
5) What are the trade-offs of these solutions? What is an acceptable trade-off?