

IT Security #4

Privacy, GDPR, and PrivacyByDesign

Niels Christian Juul

IT-security

Course book

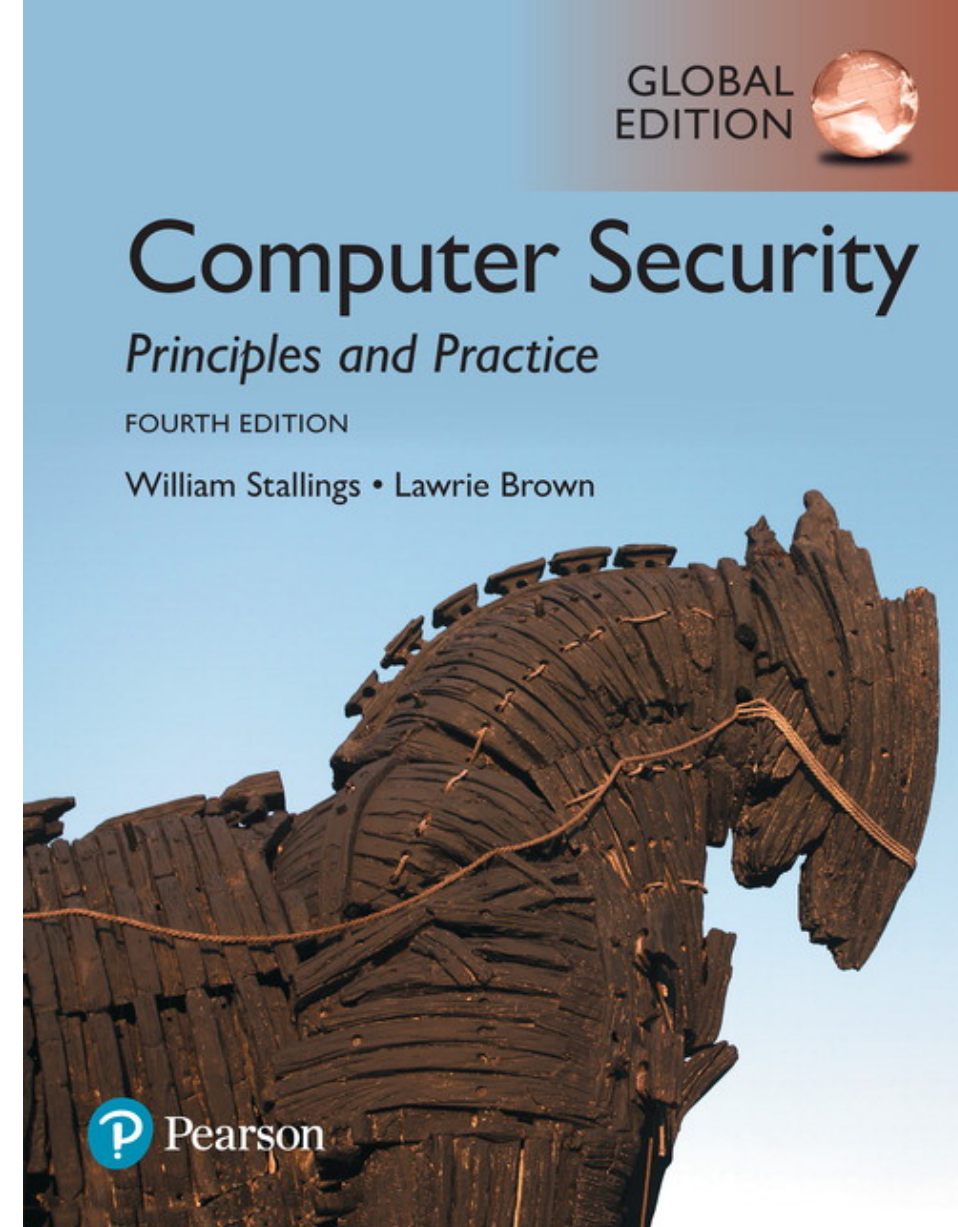
- Chapter 19 (19.1+19.3) today

Additional mandatory literature:

- [Presthus, W., Sørsum, H., & Andersen, L. R. \(2018\). GDPR compliance in Norwegian Companies. In: Proceedings from the annual NOKOBIT conference held at Svalbard the 18th-20th of September 2018, Vol 26 No 1.](#)
- [Colesky, M., Hoepman, J. H., & Hillen, C. \(2016\). A critical analysis of privacy design strategies. In 2016 IEEE Security and Privacy Workshops \(SPW\) \(pp. 33-40\). IEEE.](#)

Background:

- [Automatic Number Plate Recognition \(Wikipedia\)](#)



Learning outcome

- What is privacy?
- Why is it a requirement?
- How can we achieve it?

Exam themes/questions:

- What is GDPR and how should it be implemented in praxis?
- Show how Privacy can influence the design of an IT system

Additional knowledge not covered here, but part of the big picture:

- PrivacyByDesign (PbD), as defined by Ann Cavoukian, and Privacy Impact Assessment (PIA) is covered in BUITA class ITS1 (+ITS2)

Agenda

1. Intro (NC)
2. Cybercrime, Chapter 19.1 (Frederik).
3. Privacy, Chapter 19.3 (NC)
4. GDPR, article 1 (Natalia)
5. Privacy By Design, article 2 (Daniel)
6. ANPR (NC)
7. (Re-)design of ANPR for daily police work (NC)

Next time and next time again

- **Monday 11th March 13.15-17.00 (NJ)**

Buffer Overflows

Theme B: Software and system security.

- Student presentations:

-
-

- **Monday 18th March 13.15-17.00 (NC)**

Organizational IT Security Policy & Analysis & Implementation

Theme C: Management issues (ii).

- Student presentations:

- Security risk assessment (ch. 14.3)
- Security controls, (ch. 15.2)

Agenda

1. Intro (NC)

2. Cybercrime, Chapter 19.1 (Frederik).

3. Privacy, Chapter 19.3 (NC)

4. GDPR, article 1 (Natalia)

5. Privacy By Design, article 2 (Daniel)

6. ANPR (NC)

7. (Re-)design of ANPR for daily police work (NC)

Agenda

1. Intro (NC)
2. Cybercrime, Chapter 19.1 (Frederik).
- 3. Privacy, Chapter 19.3 (NC)**
4. GDPR, article 1 (Natalia)
5. Privacy By Design, article 2 (Daniel)
6. ANPR (NC)
7. (Re-)design of ANPR for daily police work (NC)

Privacy

- Overlaps with computer security
- Dramatic increase in scale of information collected and stored
 - Motivated by law enforcement, national security, economic incentives
- Individuals have become increasingly aware of access and use of personal information and private details about their lives
- Concerns about extent of privacy compromise have led to a variety of legal and technical approaches to reinforcing privacy rights

A Privacy Definition

- Privacy: *The right of people to choose freely under what circumstances and to what extent they will reveal themselves, their attitude, and their behavior to others*
- Threats to Privacy: Government and business
 - Regime spying on citizens
 - Employee surveillance
 - Customer surveillance (trading private information for business)
 - Use/abuse of transaction information (or citizen records)

Privacy and Data Surveillance

The demands of big business, government and law enforcement have created new threats to personal privacy

Ex.

- Scientific and medical research data collection for analysis
- Law enforcement data surveillance
- Private organizations profiling

This creates tension between enabling beneficial outcomes in areas including scientific research, public health, national security, law enforcement and efficient use of resources, while still respecting an individual's right to privacy

Privacy and Data Surveillance

Another areas of particular concern is the rapid rise in the use of public social media sites:

- These sites gather, analyze, and share large amounts of data on individuals and their interactions with other individuals and organizations
- Many people willingly upload large amounts of personal information, including photos and status updates
- This data could potentially be used by current and future employers, insurance companies, private investigators, and others, in their interactions with the individual

Privacy Protection

- Both policy and technical approaches are needed to protect privacy
- In terms of technical approaches, the requirements for privacy protection for data stored on information systems can be addresses in part using the technical mechanisms developed for database security
- In terms of policy approaches, overall regulations like OECD Guidelines, GDPR, etc must be implemented in each organization

How to protect privacy

System designers shall use:

- **PIA**
Privacy Impact Assessment
- **PET**
Privacy Enhancing Technology
- **PbD**
Privacy by Design

PET – Privacy Enhancing Technology

- Data minimization
- Unlinkability
- Informed consent
- Virtual identities
- Anonymous credentials
- Transaction logs

Data mining challenge: big data

- Privacy leakage problem

1. solution:

- Remove all identification data

2. Solution

- Anonymize (e.g. use pseudonyms)

But:

- Can de-identified data be re-identified?

PRIVACY

Credit card study blows holes in anonymity

Attack suggests need for new data safeguards

By John Bohannon

For social scientists, the age of big data carries big promises: a chance to mine demographic, financial, medical, and other vast data sets in fine detail to learn how we lead our lives. For privacy advocates, however, the prospect is alarming. They worry that the people represented in such data may not stay anonymous for long. A study of credit card data in this week's issue of *Science* (p. 536) bears out those fears, showing that it takes only a tiny amount of personal information to de-anonymize people.

The result, coming on top of earlier demonstrations that personal identities are easy to pry from anonymized data sets, indicates that such troves need new safeguards. "In light of the results, data custodians should carefully limit access

One correlation attack became famous last year when the New York City Taxi and Limousine Commission released a data set of the times, routes, and cab fares for 173 million rides. Passenger names were not included. But armed with time-stamped photos of celebrities getting in and out of taxis—there are websites devoted to celebrity spotting—bloggers, after deciphering taxi driver medallion numbers, easily figured out

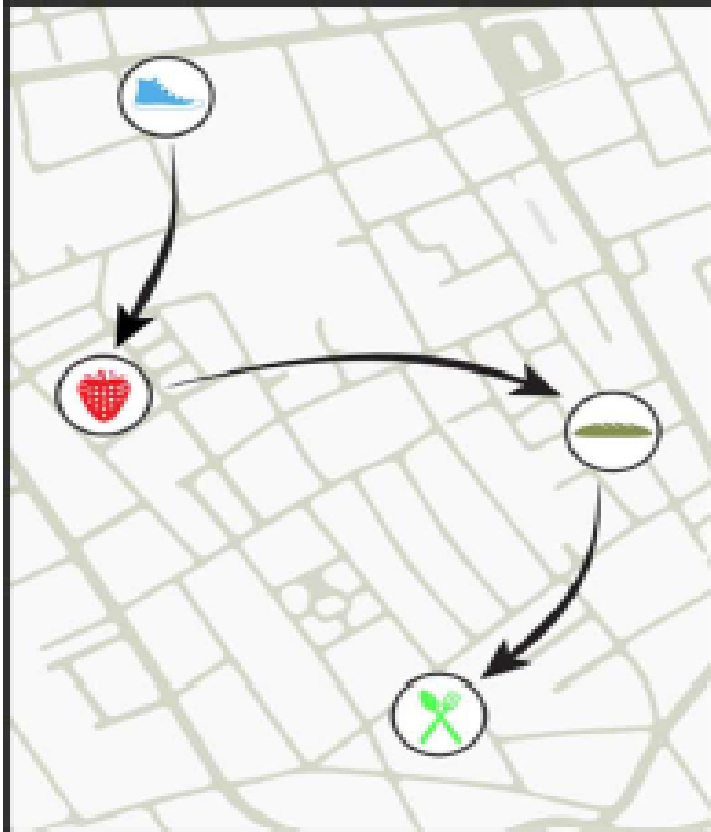
the amount spent on those occasions—the equivalent of a few receipts from someone's trash—made it possible to de-anonymize nearly everyone and trace their entire transaction history with just three pieces of information per person. The findings echo the results of a 2013 *Scientific Reports* study in which de Montjoye and colleagues started with a trove of mobile phone metadata on subscribers' movements and showed that knowing a person's location on four occasions was enough to identify them.






One way to protect against correlation attacks is to blur the data by binning certain variables. For example, rather than revealing the exact day or price of a transaction, the public version of the data set might reveal only the week in which it occurred or a price range within which it fell. Binning did not thwart de Montjoye's correlation attack; instead, it only increased the amount of information needed to de-anonymize each person to the equivalent of a dozen receipts.

These studies needn't be the death knell for social science research using big data. "We need to bring the computation to the data, not the other way around," de Montjoye says. Big data with sensitive information could live "in the cloud," protected by gatekeeper software, he says. The gatekeeper



Anonymous data about 1.1 mill people shopping in 10.000 shops during 3 month



shop	user_id	time
	7abc1a23	09/23
	7abc1a23	09/23
	3092fc10	09/23
	7abc1a23	09/23
	4c7af72a	09/23
	89c0829c	09/24
	7abc1a23	09/24

- Add the equivalent of a photo with timestamp, eg. name and data
- In 9 out of 10 cases we can link to the right user_id, if the person did four purchases
- Add prices to the purchases in our dataset, and we have 95% hits.

Unique in the shopping mall: On the reidentifiability of credit card metadata

Yves-Alexandre de Montjoye,^{1*} Laura Radaelli,² Vivek Kumar Singh,^{1,3} Alex “Sandy” Pentland¹

Large-scale data sets of human behavior have the potential to fundamentally transform the way we fight diseases, design cities, or perform research. Metadata, however, contain sensitive information. Understanding the privacy of these data sets is key to their broad use and, ultimately, their impact. We study 3 months of credit card records for 1.1 million people and show that four spatiotemporal points are enough to uniquely reidentify 90% of individuals. We show that knowing the price of a transaction increases the risk of reidentification by 22%, on average. Finally, we show that even data sets that provide coarse information at any or all of the dimensions provide little anonymity and that women are more reidentifiable than men in credit card metadata.

Exercises with your class mate

- Two by two
- 5 minutes

Find at least one additional case/study that prove privacy risks in big data?

Try to identify:

- What system
- Origin of data
- Risk including both positive and negative outcome of studies system

Privacy Protection

With regard to social media sites, technical controls include:

- The provision of suitable privacy settings to manage who can view data on individuals
- Notification when one individual is referenced or tagged in another's content
- Although social media sites include some form of these controls, they are constantly changing, causing frustration for users who are trying to keep up with these mechanisms
- Another approach for managing privacy concerns in big data analysis is to anonymize the data, removing any personally identifying information before release to researchers or other organizations for analysis

Data Privacy

- In terms of policy, guidelines are needed to manage the use and reuse of big data, ensuring suitable constraints are imposed in order to preserve privacy
 - **Consent**
Ensuring participants can make informed decisions about their participation in the research
 - **Privacy and confidentiality**
Privacy is the control that individuals have over who can access their personal information
Confidentiality is the principle that only authorized persons should have access to information
 - **Ownership and authorship**
Addresses who has responsibility for the data, and at what point does an individual give up their right to control their personal data
 - **Data sharing – assessing the social benefits of research**
The social benefits that result from data matching and reuse of data from one source or research project in another
 - **Governance and custodianship**
Oversight and implementation of the management, organization, access, and preservation of digital data

Fair Information Practices

- OECD (Organization of Economic Cooperation and Development) in 1980 developed the standard eight-point list of privacy principles:
- Limited Collection Principle
- Quality Principle
- Purpose Principle
- Use Limitation Principle
- Security Principle
- Openness Principle
- Participation Principle
- Accountability Principle

Same 8 principles in 2013 OECD Guidelines. Table 19.3, p.625

Following OECD Guidelines....

- European Union had European Data Protection Directive (OECD principles) since 1995
- EU Directive requires data on EU citizens to be protected at same standard even when it leaves their country (Directive 95/46/EC)
- EU General Data Protection Regulation (GDPR) in force May 2018
- China does not protect privacy
- U.S. has not adopted OECD principles

BUT

United States Privacy Initiatives: citizen vs government

Privacy Act of 1974

- Deals with personal information collected and used by federal agencies
- Permits individuals to determine records kept
- Permits individuals to forbid records being used for other purposes
- Permits individuals to obtain access to records and to correct and amend records as appropriate
- Ensures agencies properly collect, maintain, and use personal information
- Creates a private right of action for individuals

US Laws Protecting Privacy: single sector, single state

- Federal Trade Commission Act, 1914 (FTCA)
- Fair Credit Reporting Act, 1970 (FCRA)
- Electronic Communication Privacy Act, 1986 (ECPA)
- Video Privacy Protection Act, 1988 (VPPA)
- Telephone Consumer Protection Act, 1991 (TCPA)
- Driver's Privacy Protection Act, 1994 (DPPA)
- Health Insurance Privacy and Accountability Act, 1996 (HIPAA)
- Financial Services Modernization Act, 1999 (Gramm-Leach-Bliley Act (GLBA))
- Controlling the Assault of Non-Solicited Pornography and Marketing Act, 2003 (CAN-SPAM)
- Data Security and Breach Notification Act *Proposal 2017* (DSBN)
- California Consumer Privacy Act of 2018 (CCPA), eff. 2020

Agenda

1. Intro (NC)
2. Cybercrime, Chapter 19.1 (Frederik).
3. Privacy, Chapter 19.3 (NC)
- 4. GDPR, article 1 (Natalia)**
5. Privacy By Design, article 2 (Daniel)
6. ANPR (NC)
7. (Re-)design of ANPR for daily police work (NC)

Agenda

1. Intro (NC)
2. Cybercrime, Chapter 19.1 (Frederik).
3. Privacy, Chapter 19.3 (NC)
4. GDPR, article 1 (Natalia)
- 5. Privacy By Design, article 2 (Daniel)**
6. ANPR (NC)
7. (Re-)design of ANPR for daily police work (NC)

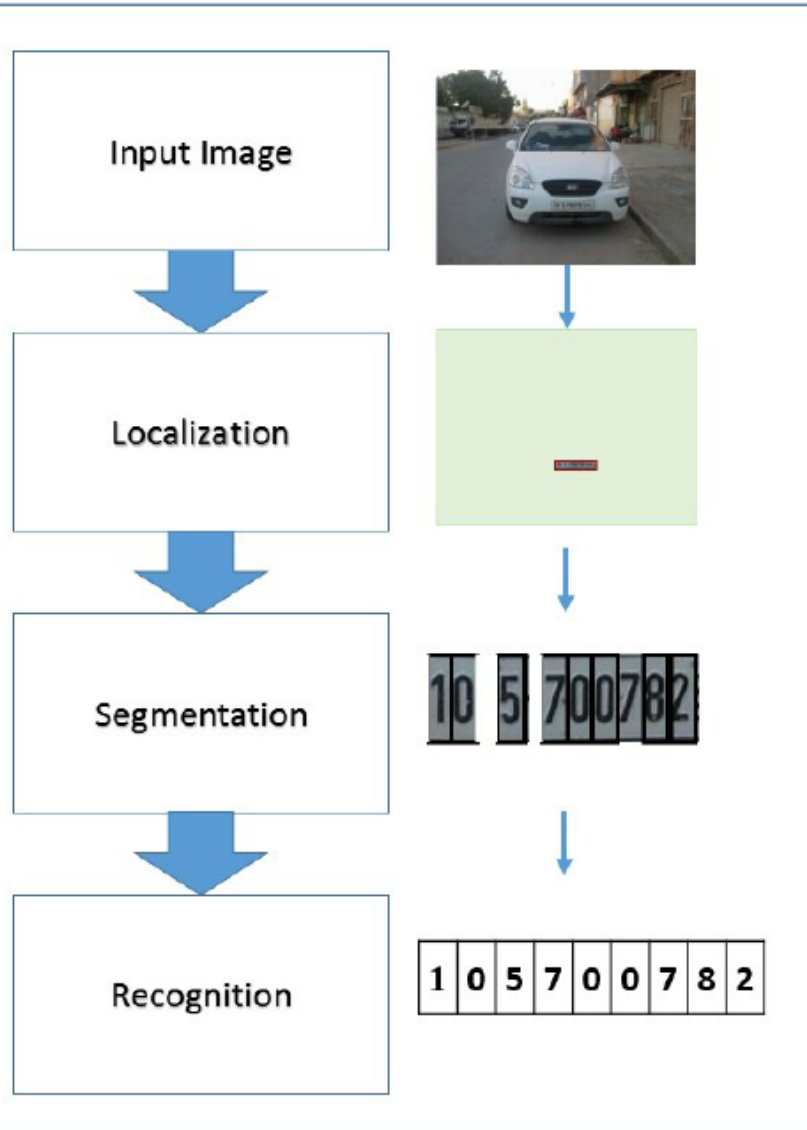
Agenda

1. Intro (NC)
2. Cybercrime, Chapter 19.1 (Frederik).
3. Privacy, Chapter 19.3 (NC)
4. GDPR, article 1 (Natalia)
5. Privacy By Design, article 2 (Daniel)
- 6. ANPR (NC)**
7. (Re-)design of ANPR for daily police work (NC)

ANPG stationary positions



Automated Number Plate Recognition



- System installed in Police cars and stationary on road sides.
- ANPR generates plate numbers and other data about cars seen by the cameras (infrared view, no blitz)
- Works with a dynamic and updated **hotlist** of "wanted" plate numbers managed by the police, i.e. cars worth looking for.
- Classifies each view as **hit** or **no-hit** according to hotlist.

Agenda

1. Intro (NC)
2. Cybercrime, Chapter 19.1 (Frederik).
3. Privacy, Chapter 19.3 (NC)
4. GDPR, article 1 (Natalia)
5. Privacy By Design, article 2 (Daniel)
6. ANPR (NC)
- 7. (Re-)design of ANPR for daily police work (NC)**

Exercise on ANPR system

With PbD principles in mind, design a system using these ANPR installations (stationary and/or on police cars) to achieve either both or one of the two official goals

1. Hits

Catch stolen and other wanted cars from the hotlist

2. No-Hits

Catch all plate numbers (not on the hotlist) during targeted police investigations limited in space and time for later search operations within 30 days maximum

Evaluate pros and cons for your proposed solutions wrt legal outcome and privacy protection.

Learning outcome

- What is privacy?
- Why is it a requirement?
- How can we achieve it?

Exam themes/questions:

- What is GDPR and how should it be implemented in praxis?
- Show how Privacy can influence the design of an IT system

Additional knowledge not covered here, but part of the big picture:

- PrivacyByDesign (PbD), as defined by Ann Cavoukian, and Privacy Impact Assessment (PIA) is covered in BUITA class ITS1 (+ITS2)