1.Case: Stolen data

Summary Target Data

- Data touchself can be extremely damaging (money & reputation)
- Full network access for 1st parties external vendors is risky
- Insufficient monitoring which logging led to late attack detection

Summary Target Data

- Lack of network segmentation allowed lateral movement of attackers
- Hacks like this can happen to overexpected only large organizations
- All companies working with Big Data should prioritize security

2.Types of threats

Summary Personal Data and Big Data Applications

- Typical threats include unauthorized access to data, data stealing, data theft, data manipulation
- Attacks often launched from inside an organization

Summary Personal Data and Big Data Applications

- Typical external attackers include:
  - Cleaner, cameras,
  - Hacktivists
  - Governments
  - Script kiddies

Individual by financial or personal gain

"Fat fun"

3.Security goals

Summary Security Goals

	Protection of...	
Security	Data	Unauthorized attacks
Safety	Environment	Negative impact of IT

Summary Security Goals

- Security goals
  - **Confidentiality**: Authorized use only
  - **Integrity**: No unauthorized change
  - **Availability**: Data and services available
  - Other goals can be added (e.g. privacy)

4.Implementing security for Big Data

Summary Implementing Security for Big Data

- 100% Security is not realistic
- Prioritize risks with an IT risk analysis
- Prioritize which assets shall be secured
- Unique characteristics of Big Data present unique security challenges

Summary Implementing Security for Big Data

- High level of Security for Big Data
  - Early risk analysis
  - Combination of preventive, detective and administrative security measures
  - Raise organizational awareness

5.Administrative measures for Big Data

Summary Administrative Measures

- Security needs
  - organizational-administrative measures
  - technical measures
- These measures define
  - required actions
  - and responsibilities
- Business and technical expertise required
- Standards (e.g., ISO 27001, IEC 62443) provide guidance

6.Preventive measures: Encryption

Summary Preventive Measures

- Encryption required to secure Big Data
  - Communication data
  - Stored data
- Cryptographic keys required
- Symmetric Encryption – shared secret key
- Asymmetric Encryption – combination of public key and private key

Summary Preventive Measures

- Asymmetric Cryptography applications
  - Securing communication data
  - Encrypting stored Big Data
- Certificates used to authenticate key owner
- Trust infrastructure – centralized or decentralized

7.Preventive measures: Access control

Summary Preventive Measures

- Authentication: Is the person who they claim to be?
- Authorization: Is the person allowed to access the resource, or perform operation on the resource?
  - Two forms of access control measures
  - Role based access control (RBAC)
  - Least Privilege and segregation
- Combine different authentication and authorization schemes for best results

8.Detective measures: Audits

Summary Detective Measures – Audits

- Detective security measures needed to increase security
- Audits: Check correct planning, configuration and set up
- Penetration tests: Find and remove weaknesses
- PIA or DPIA: Check compliance with requirements

9.Detective measures: Monitoring

Summary Detective Measures – Monitoring

- Outlier detection is used to identify anomalies
- One class classification is the most popular technique for outlier detection
- The monitoring of accessing of subjects can be done as part of an overarching SIEM system

10.Best practices for securing Big Data

Summary Best Practices

1. Apply Need to Know & Minimum User Rights
2. Update systems regularly
3. Track lineage of data provenance
4. Build systems with Security & Privacy in mind
5. Maintain growing systems with increasing complexity

11.Social costs of Big Data

Summary Social Costs of Big Data

- Big Data opens new opportunities but can lead to social costs
  - Prediction – preferences protected
  - Pre-emption – future options diminished
  - Growing quantity of data fosters Prediction + Pre-emption
- Linking information from sources makes inferences about people possible

12.Anonymizing and pseudonymizing data

Summary Anonymizing and Pseudonymizing Data

- Help mask relationship between data and a real person
- Anonymization is strongest masking method
  - All information linked to real person removed
- Pseudonymization replaces real data with pseudonyms
  - By linking attributes, people can be identified
- A pseudonyms addresses risk of re-identification of anonymized data through linkage to other datasets

13.Differential Privacy

Summary Differential Privacy

- Differential Privacy – guarantees that privacy of individuals is not noticed when running queries
- Use cases in health data
  - Protecting patient identity
  - Mining useful information
  - Sharing data with researchers
  - Makes inference and tracking attacks less likely

14.Differential Privacy methods

Summary Differential Privacy Methods

- Differential Privacy
  - Makes Big Data datasets available for analyses
  - Protects privacy of individuals
- Local Differential
  - Adds noise to computer function
  - Information about individual sub-groups protected

Summary Differential Privacy Methods

- Randomized Response
  - Ensure privacy at moment of data generation
- Exponential Mechanism
  - Publish anonymized dataset with same distribution patterns as original data

15.Homomorphic encryption

Summary Homomorphic Encryption

- Provides capability to perform operations on data while encrypted
- Possible to allow data processing, without need to provide access to data
- Limited practical implementations exist
  - High computational requirements
  - Organizing combination of encrypted parts complex

16.Secure Multiparty Computation

Summary Secure Multiparty Computation

- Problem solved:
  - Computation of function while keeping inputs secret
  - Use of trusted third party not needed
  - Used protocols need to operate
    - Input Privacy
    - Correctness
    - Independence
  - Complex to set up an online
  - High computational requirements

Secure Multiparty Computation Basic Properties

No more private information about input data can be inferred than what is revealed by function output

Input Privacy

Secure Multiparty Computation Basic Properties

Function is always computed correctly – dishonest parties cannot force honest parties to output incorrect result

Correctness

17.Data Protection Law for Big Data

Summary Data Protection Law for Big Data

Data Protection is needed to ensure that the service providers about how their data is handled by Big Data applications are obeyed

Summary Data Protection Law for Big Data

- The General Data Protection Regulation will come into force starting May, 2018.
- Its aim is:
  - Further harmonize Data protection across Europe
  - Strengthen rights of data subjects
- It replaces the Data Protection Directive 95/46/EC which did not account for emergence of data as part of Big Data and Cloud applications

Summary Data Protection Law for Big Data

- The GDPR applies directly as European law
- Vis-to-EU-citizens (citizens, Member States can introduce national provisions to organize or further specify particular rules
- The monetary expense of the GDPR will vary between Member States

18.General Data Protection Regulation (GDPR)

Summary General Data Protection Regulation

- The General Data Protection Regulation will come into force starting May, 2018.
- Its aim is:
  - Further harmonize Data protection across Europe
  - Strengthen rights of data subjects
- It replaces the Data Protection Directive 95/46/EC which did not account for emergence of data as part of Big Data and Cloud applications

Summary General Data Protection Regulation

- The GDPR applies directly as European law
- Vis-to-EU-citizens (citizens, Member States can introduce national provisions to organize or further specify particular rules
- The monetary expense of the GDPR will vary between Member States

19.Important changes resulting from GDPR

Summary GDPR

- Data Subject's able to easily access data and how it is processed
- Children's data needs specially protected legal guardian consent and authorization
- Privacy by design: Privacy is not an add-ons but more strongly ensured, data protection is integrated into the product

Summary GDPR

- Data protection impact assessment brought in, risk of processing personal data must be assessed and minimized before collected
- All international companies with services or EU market have to comply with standards
- Fines for data protection violations: data €20,000,000 or 4% of company's worldwide annual turnover, whichever is greater

20.Personal Data

Summary Personal Data

- Personal data includes
  - Name
  - Location
  - Personal images
  - Log data
- Aggregation of bits may allow re-identification of personal information

Summary Personal Data

- Pseudonymized data → personal data in which an identifiable person cannot be attributed to the data, without use of additional information
- Anonymized data → personal data where identifiable is not created under Data Protection Law

21.Personal data and Big Data applications

Summary Personal Data and Big Data Applications

- Big Data applications collect personal data
- They need to be aware of organizational organizational data
- For pseudonymized data, there is a chance that the natural person can be re-identified

Summary Personal Data and Big Data Applications

- Reasonable means should be taken into account when deciding whether or not the data subject can be identified.
- Establishing "without additional means" where the data subject is not willing to be re-identified required to be recognized, is not mentioned in the GDPR.