Privacy in Healthcare: Introduction

Alexander Schönhuth

UNIVERSITÄT
BIELEFELD
Faculty of Technology

Bielefeld University
April 11, 2023

WHO ARE WE?

- Research group "Genome Data Science"
  `https://gds.techfak.uni-bielefeld.de`
- Coordinates:
  Prof. Dr. Alexander Schönhuth
  email: aschoen@cebitec.uni-bielefeld.de
  office: UHG U10-128

Organization

MODULES

- Lecture part of modules
  - *39-Inf-BDS Biomedical Data Science for Modern Healthcare Technology* (graded, “benotete Prüfungsleistung”)
    - See here `https://ekvv.uni-bielefeld.de/sinfo/publ/modul/308594662`

PRESENTATION, REPORTS, PAPERS

- Presentations:
  - Individual presentations
  - To last for approx. 30 minutes, followed by discussion
  - Present contents of scientific paper
- Reports:
  - Reports summarize contents of paper
  - Reports 8-12 pages
- Papers:
  - Papers: some already available, list will be completed
  - Papers available via Wiki: `https://gds.techfak.uni-bielefeld.de/teaching/2023summer/privacy`

Schedule

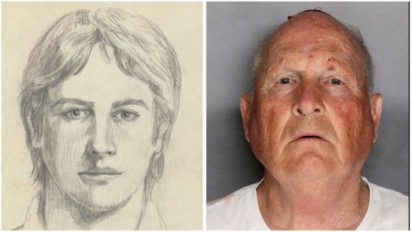
- Organization and introduction: *today*
- How to present (brief): *Apr 18* (online)
- How to write (brief): *Apr 25* (hybrid)

Schedule II

- **Presentations:** *from May 16* (earlier possible if desired)
  - Up to two presentations per week
  - Block seminar day possible as well (yet TBD)
- **Technical Report:** *after presentation:*
  - Each report 8-12 pages
  - Optimally, report profits from feedback provided after presentation
  - Drafts can be submitted for discussion
  - Improving drafts based on feedback
  - *Submission deadline: July 31*

Privacy in Healthcare: Overview

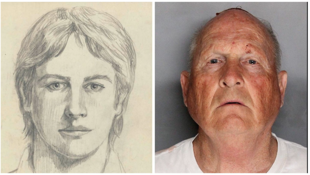
Example: Long Range Familial Searches



From www.stern.de

- ► Investigators uploaded crime scene sample to GEDmatch
  - ► GEDmatch contains 1 million DNA profiles
- ► GEDmatch search identified a third-degree cousin
- ► Genealogical search identified the perpetrator

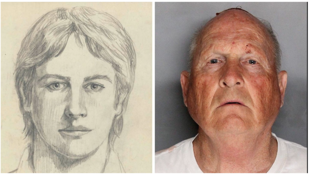
Exemplary Issues



From www.stern.de

- *Access control:*
  - Who has permission to run database searches?
  - How to organize access control?
- *Multiparty computation:*
  - Several parties share data to run computations
  - Each party's data should stay private
  - Everyone can use data to get anonymous summaries

Exemplary Issues

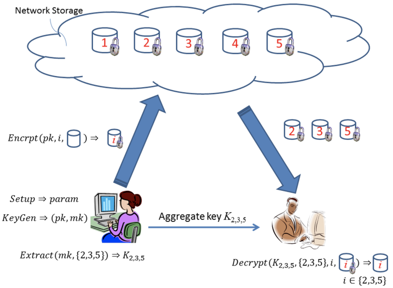


From www.stern.de

- *Homomorphic encryption:*
    - Encrypt data such that computations on encrypted data is possible
- *Differential privacy frameworks:*
    - Individual data should make no difference during analysis

Access Control

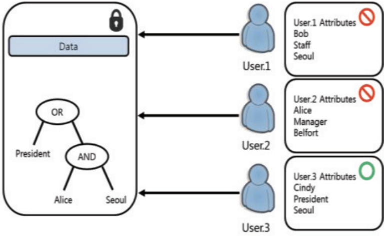
Access Control


Network Storage
1
2
3
4
5
$Encrpt(pk, i, \;) \Rightarrow$ i
2
3
5
$Setup \Rightarrow param$
$KeyGen \Rightarrow (pk, mk)$
Aggregate key $K_{2,3,5}$
$Extract(mk, \{2,3,5\}) \Rightarrow K_{2,3,5}$
$Decrypt(K_{2,3,5}, \{2,3,5\}, i, \; i \;) \Rightarrow$ i
$i \in \{2,3,5\}$

From [Chu et al., 2014]

- *Key aggregate cryptography:*
  - “Master” distributes key to potential users

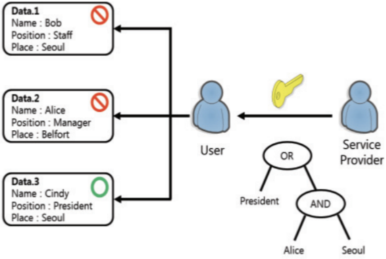
Access Control



From [Lee et al., 2015]

- *Attribute based access control:*
  - Keys depend on data characteristics

Access Control

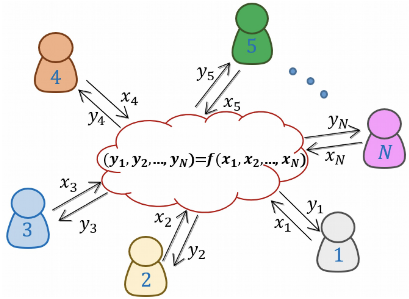


From [Lee et al., 2015]

- *Role based access control:*
  - Keys depend on user properties

Multiparty Computation

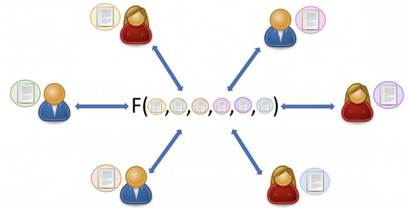
Multiparty Computation I



See www.mdpi.com

- *Multiparty computation principle:*
  - N parties provide data $x_1, \ldots, x_N$
  - Values $y_1, \ldots, y_N$ are computed
  - User providing x_i receives y_i (only)

Multiparty Computation II



See www.esat.kuleuven.be

- *Multiparty computation healthcare:*
  - Patients / doctors provide individual records
  - Individual analysis based on all records
  - Patients / doctors receive individual analysis results

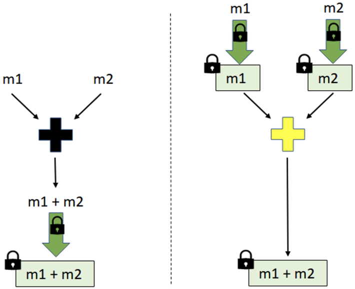
Homomorphic Encryption

HOMOMORPHIC ENCRYPTION I

See www.linksight.nl

- *Homomorphic encryption motivation:*
  - Important operations still possible after encryption
  - Decrypting data unnecessary
  - Allows users to carry out queries anonymously

Homomorphic Encryption II

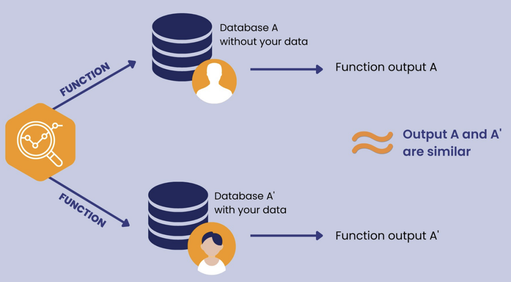


See akd13.github.io

- *Homomorphic encryption principle:*
  - Encryption and queries are mathematical operations
  - Exchanging these operations should lead to same results

Differential Privacy

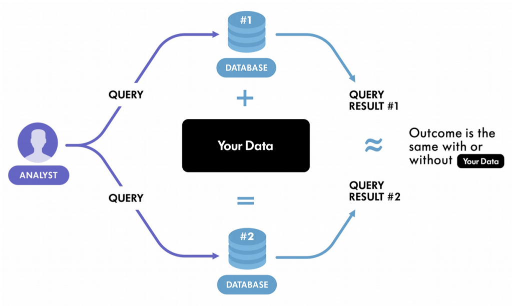
Differential Privacy I



See www.statice.ai

- *Differential privacy principle:*
  - Database A contains individual data, Database A′ does not
  - Running function returns same result on A and A′
  - *Individual data* makes no difference, so remains *unidentifiable*

Differential Privacy II



See www.winton.com

- *Differential privacy practice:*
  - Analyst runs (specially tailored) query on database with and without individual records
  - Outcomes do not differ: individual records remain anonymous

Thanks for your attention!