Privacy in Digital Healthcare

Alexander Schönhuth

Bielefeld University
October 11, 2023

Overview

- *Organization:*
  - How do lectures, tutorials etc work
  - What tools will be used
- What does *Differential Privacy* mean?
  - Modifying databases such that individual records cannot be tracked
  - Statistical framework
- What is *Federated Learning*? What is *Swarm Learning*?
  - Multiparty computation: use all individual records, without sharing them
  - Federated learning = multiparty (collaborative) learning
  - Swarm learning: blockchain orchestrated federated learning
- What are *Bitcoin and Blockchains?* What are
  - Electronic cash
  - Decentralization, immutability
  - Transaction blocks, Proof of Work

Prerequisites, Lectures, Exercises

- Lectures: Wednesdays, 12-14
- November 15 & 22: no lecture
- Hybrid or (exceptionally) online meetings
- Lectures will be recorded
- Edited videos and slides will be posted
- Exercises: 7 assignments + 1 exam preparation session

Assignments, Exam

- *Tutorials / Assignments:*
  - Exercise sheets distributed on Wednesdays Oct 18, Nov 1, Nov 15, Nov 29, Dec 13, Dec 20, Jan 3 after the lecture
  - Exercises to be submitted by Monday, **23:59** twelve days thereafter; Discussion on Wednesday, 10-12 same week
  - Submission of exercises in groups of 2-3 people possible
  - Everyone to present at least one exercise in the tutorials
  - Upload to corresponding folder in the "Lernraum Plus"
  - First exercise sheet uploaded on October 18 (next week)
  - Usage of ChatGPT (or similar) fine, if links to chats provided
- *Exam:*
  - Presence exam planned for **Wednesday, January 31, 2024 between 10:00 and 14:00** (may be subject to changes due to situation; we will communicate changes as timely as possible)
  - Admitted: everyone exceeding 50% of total exercise points

Tutorials

- Every **Wednesday, 10-12**
- Tutor: Johannes Schlüter
- Tutorials will be in English
- Presence meetings
- Presentation of solutions individually

Course material

- . . . available on course website: `https://gds.techfak.uni-bielefeld.de/teaching/2023winter/healthcare`
  - Slides and pointers to literature
  - Excercise sheets
- Moodle: `https://moodle.uni-bielefeld.de/course/view.php?id=3013`
  - Submission of exercise solutions
  - Self-managed forum

Literature and Links

- *Download:* `https://d28rh4a8wq0iu5.cloudfront.net/bitcointech/readings/princeton_bitcoin_book.pdf`
- *Further materials:* `https://bitcoinbook.cs.princeton.edu/`
- *Other literature:* See Lernraum Plus, course website and lecture slides
- *Literature for other topics yet TBD*

Course Curriculum

Part 1: Bitcoin / Blockchain

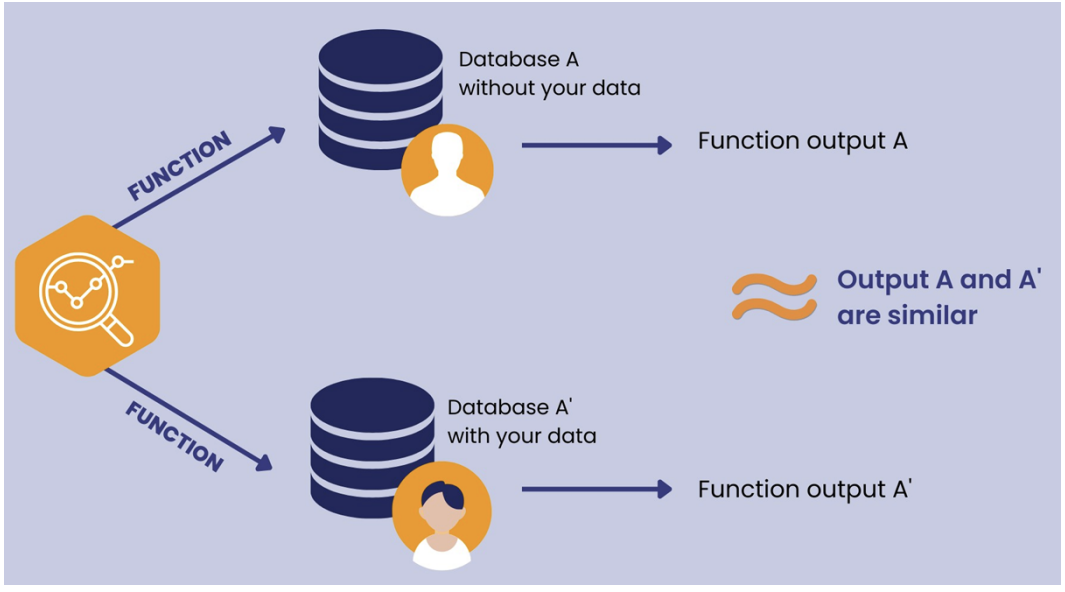
- Introduction / Motivation
- Cryptography / Cryptocurrencies
- Decentralization
- Cryptocurrency Mechanics
- Application I: Cloud Supported Medical Blockchain

Part 2: Ethereum / Privacy

- Smart Contracts, Ethereum, Solidity
- Applications: MedRec: Individual permission management, BlockTrial: Clinical trials data
- Differential Privacy
- Federated Learning
- Swarm Learning

Differential Privacy

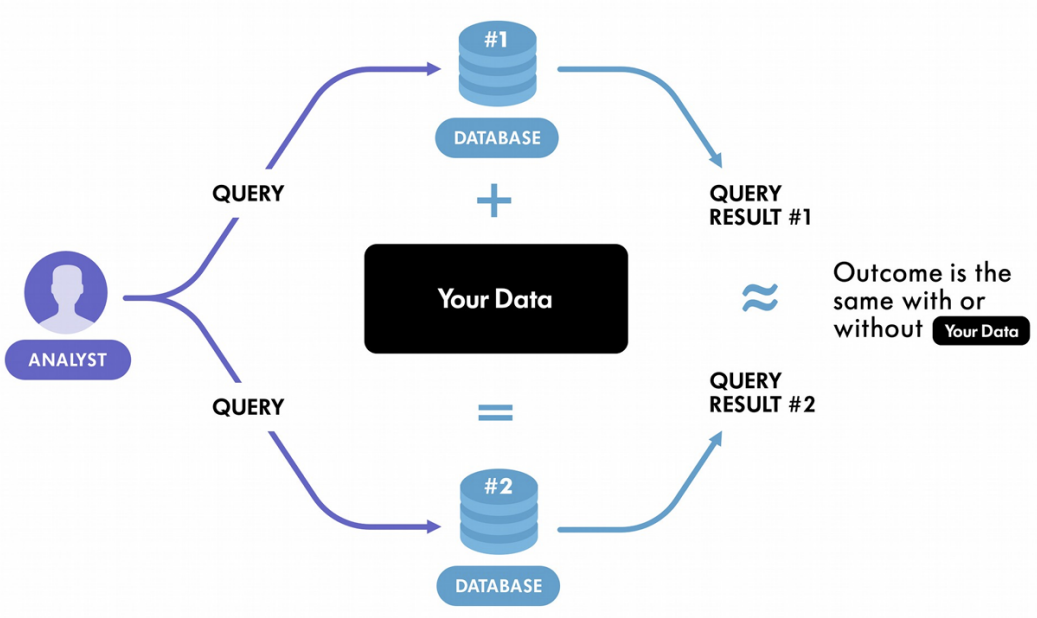
Differential Privacy I



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

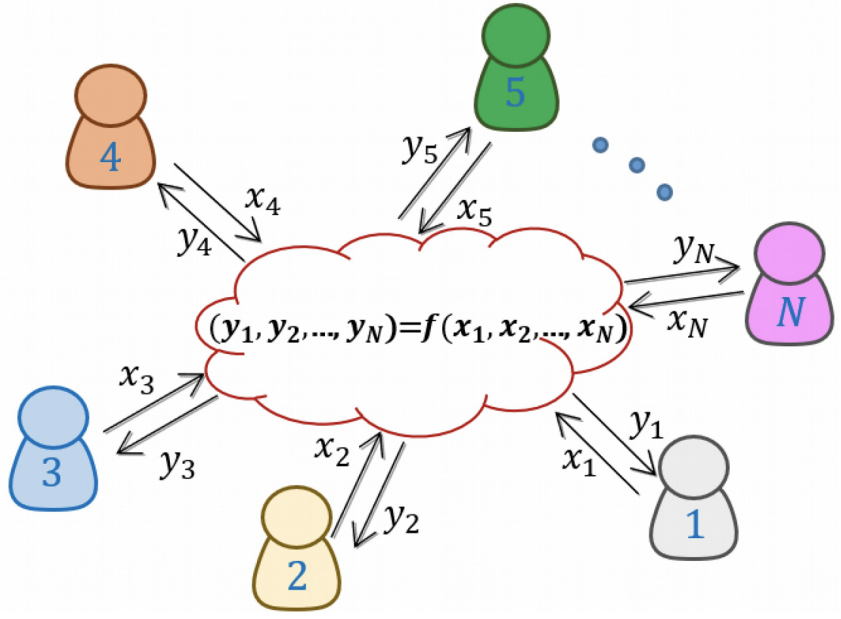


See www.winton.com

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

Federated Learning

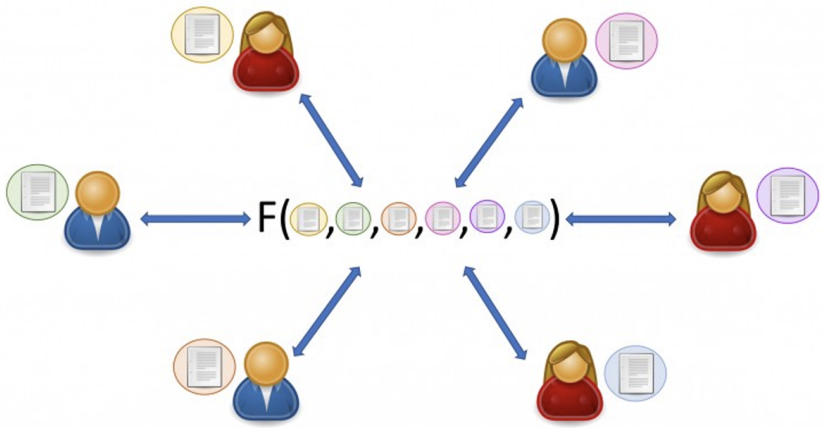
Multiparty Computation I



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)

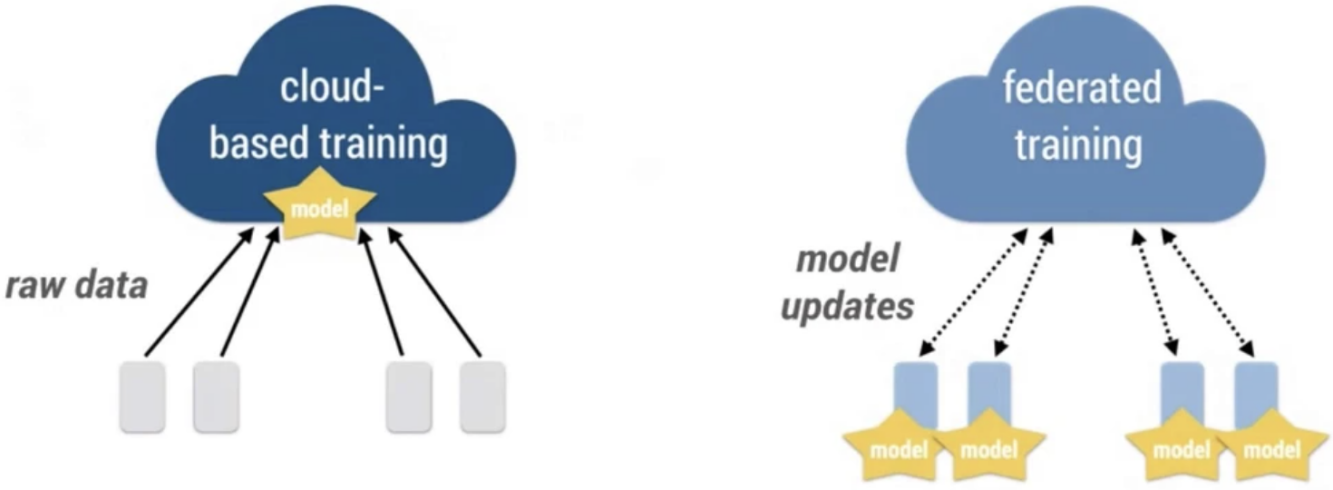
Multiparty Computation II



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

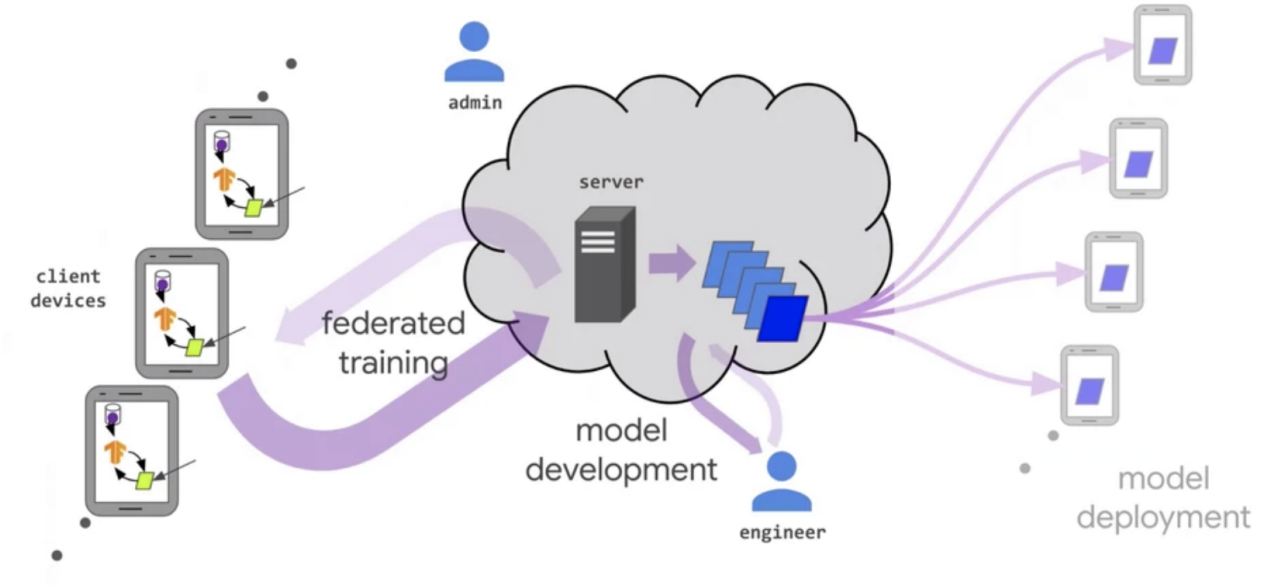
Federated Learning



See slideslive.com/38935813/federated-learning-tutorial

- *Cloud based learning:* Data transferred to cloud
- *Federated learning (FL):* Data remains stored locally
  - Reduced network strain
  - Enhanced privacy
  - Quick incorporation of new data

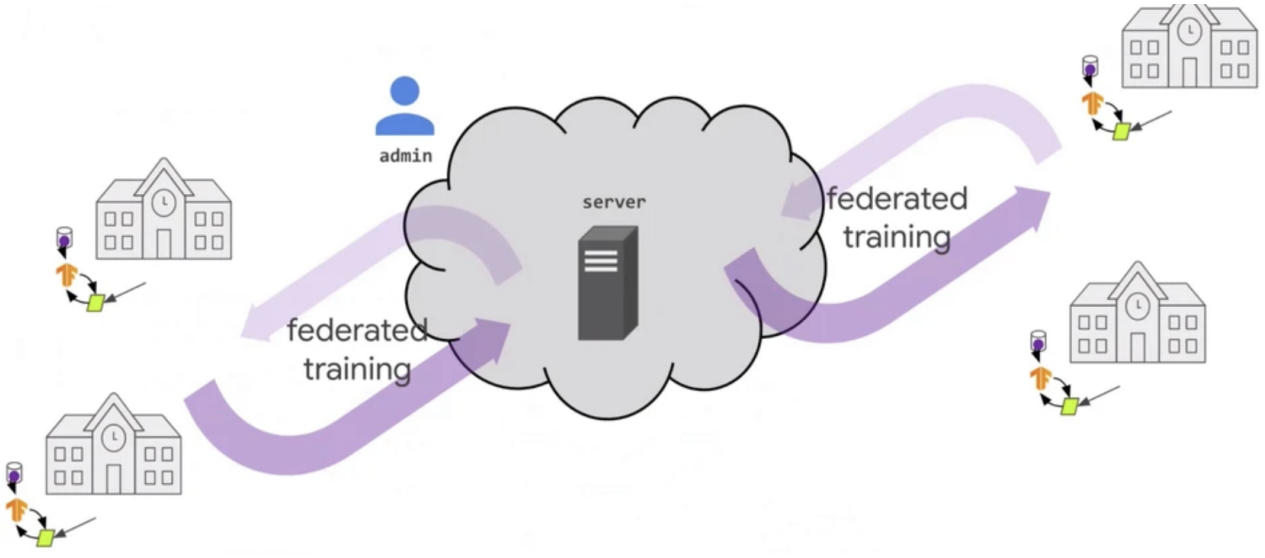
Cross-Device Federated Learning



See slideslive.com/38935813/federated-learning-tutorial

- Central engineering unit provides models to individual users
- Users train model locally with their data and return trained version
- Globally trained models used to derive individual conclusions

Cross-Silo Federated Learning



See slideslive.com/38935813/federated-learning-tutorial

- Individual institutions (clinics) store data collections
- Institutional data is used to train centrally administered models
- Institutions use globally trained models to derive conclusions

Blockchains
–
Motivation

Major Applications

- Management of individual medical records
- Insurance claim processes
- Clinical / biomedical research / studies
- Biomedical / health care data ledger

Central Benefits

- Immutability: once deposited, data cannot be changed
- Transparency: every participant can see data
- Anonymity / Security: real identities not revealed
- Robustness: Data resistant to blackouts / technical failures
- Decentralization: Nobody ”owns” the data

Example
–
Electronic health records (EHR)

EHRs - Immutability

Use case - Bob visits a doctor

- Bob has a stomach ache and visits doctor Alice
- Alice assumes Bob ate too much and isn't really sick
- Alice prescribes chamomile tea and puts the case to her files

EHRs - Immutability

Use case - Bob gets misdiagnosed

- However, Bob has a severe infection and has to go to the hospital
- Alice is afraid that she is going to face repercussions because of her mistake
- Alice would like to access Bob's file to fake the evidence and change Bob's diagnosis

Databased management systems (DBMSs) versus Blockchains

- *Database management systems (DBMSs)* have "delete" and "modify" functionalities, so that's possible
- *Blockchains support immutability:* no record can be altered retroactively

EHRs - Privacy / Transparency

Use case - Accessing Bob's files

- Independent authorities
  - get access to Bob's files to evaluate the situation
  - should not be able to identify Bob's identity
  - should nevertheless be sure it's from the right patient
  - should be able to make sure that records are consistent

DBMSs vs Blockchains

- *DBMSs:* Records contain names, addresses etc, to identify ownership of records; records could not be approved by patients
- *Blockchains:*
  - Privacy through anonymized identifiers, while still assignable to real people when necessary
  - Enhanced transparency, everyone can check validity of records without discovering Bob's real identity

EHRs - Decentralization

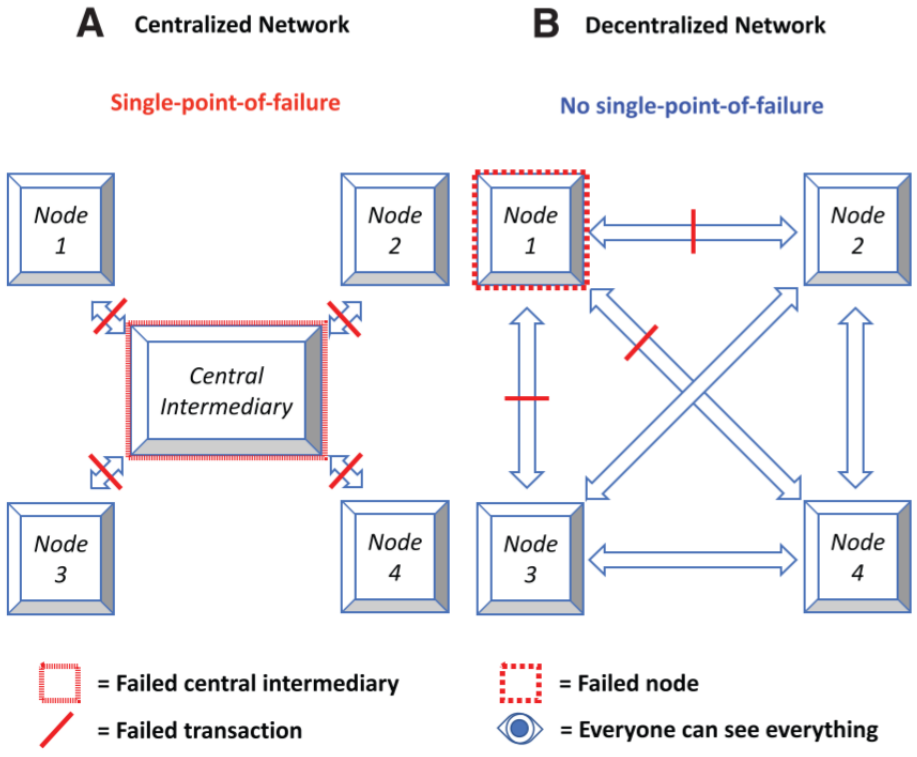
Use case - Bob goes to the hospital

- Bob does not trust Alice any longer and goes to the hospital instead
- At the hospital, he receives treatment against the infection
- However, the hospital was subject to a hack and all data got lost, which prevents Walter, the new doctor, to treat Bob
- Bob has to undergo a series of test, so that the doctors can continue his treatment

DBMSs vs blockchains

- *DBMSs:* Centralized storage, so no remote backups available
- *Blockchains:* Build on decentralized network.
  - Records are stored ”everywhere in the network”
  - This avoids ”single points of failure”

Motivation - Decentralization



A: Central authority (e.g. running a DBMS), single point of failure

B: Cluster / cloud: no single point of failure. However, no transparency, anonymity, immutability

☞ We'll get to all of that!

Other Prominent Applications

Insurance Claim Processes

- *Immutability:* No party involved can tamper with relevant records / evidence; audits facilitation and fraud detection
- *Transparency:* All records that support decisions verifiable by anyone involved
- *Anonymity / Security:* No hacking of medical / financial information
- *Robustness:* Patient data accessible from multiple silos
- *Decentralization:* No intermediaries who could have own interests necessary

Clinical / Biomedical Studies / Research

- *Immutability:* Trackable, timestamped patient-generated data
- *Transparency:* Continuous access to real-time data and information on provenance, overall verifiability. Relevant cross-study insights can be gained
- *Anonymity / Security:* No real-world identities to be maintained other than with the participating patients themselves
- *Robustness:* No broken real-time data records.
- *Decentralization:* Each institution keeps control of their own resources, while allowing for full collaboration on shared data

Health Care Data Ledger

Internet of Things, Mobile Devices

- *Immutability:* Providing ordered (timestamped), continuously updated data
- *Transparency:* Forged, poor quality or stolen data easily identified
- *Anonymity / Security:* Patients can provide access to data using cryptographic protocols
- *Robustness:* Drug counterfeiting in drug supply chains impossible
- *Decentralization:* Data "pooled", so central authorities do not prevent individual usage

Bitcoin
–
Online Cash

Offline Cash

Disadvantages

- Needs to be ”bootstrapped”: initial distribution of cash to participants necessary
- Physical presence required for transactions

Advantages

- Full anonymity: no spending records, no identities
- Offline transactions, no involvement of third parties

Electronic Banking

Credit Cards

- Buyer sends credit card details to seller; seller contacts ”system”
- The ”system” involves various third parties: banks, processors, credit card intermediaries, and so on
- *Disadvantages:*
  - Seller has credit card details
  - Third parties, even if trustworthy, can exploit records for legal things

PayPal

- Buyer and seller communicate via PayPal
- Seller does not receive credit card details
- *Disadvantages:*
  - PayPal has access to personal data
  - Buyer and seller need account with PayPal

Online Buying / Selling

Situation before Bitcoin

ACC	CyberCents	iKP	MPTP	Proton
Agora	CyberCoin	IMB-MP	Net900	Redi-Charge
AIMP	CyberGold	InterCoin	NetBill	S/PAY
Allopass	DigiGold	Ipin	NetCard	Sandia Lab E-Cash
b-money	Digital Silk Road	Javien	NetCash	Secure Courier
BankNet	e-Comm	Karma	NetCheque	Semopo
Bitbit	E-Gold	LotteryTickets	NetFare	SET
Bitgold	Ecash	Lucre	No3rd	SET2Go
Bitpass	eCharge	MagicMoney	One Click Charge	SubScrip
C-SET	eCoin	Mandate	PayMe	Trivnet
CAFÉ	Edd	MicroMint	PayNet	TUB
CheckFree	eVend	Micromoney	PayPal	Twitpay
ClickandBuy	First Virtual	MilliCent	PaySafeCard	VeriFone
ClickShare	FSTC Electronic Check	Mini-Pay	PayTrust	VisaCash
CommerceNet	Geldkarte	Minitix	PayWord	Wallie
CommercePOINT	Globe Left	MobileMoney	Peppercoin	Way2Pay
CommerceSTAGE	Hashcash	Mojo	PhoneTicks	WorldPay
Cybank	HINDE	Mollie	Playspan	X-Pay
CyberCash	iBill	Mondex	Polling	

Many more have tried without success

From `https://bitcoinbook.cs.princeton.edu`

Bitcoin Electronic Cash

Bitcoins versus Cash

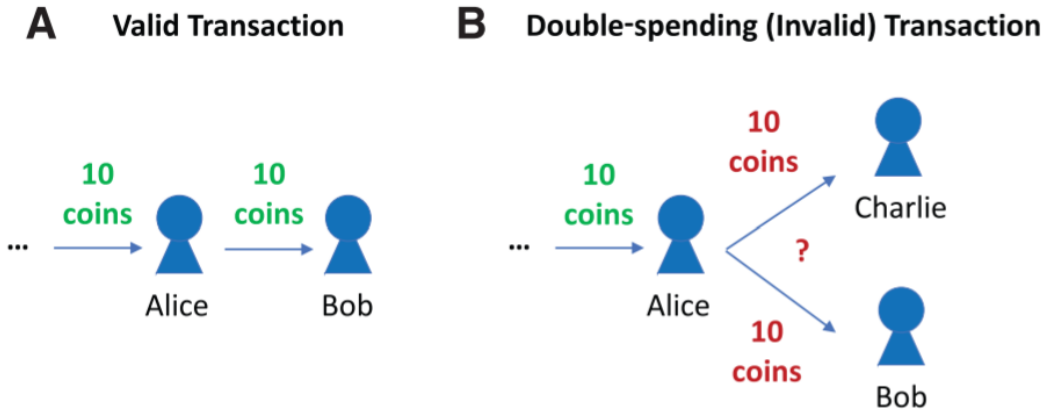
- Bitcoin does not reach full anonymity
- Bitcoin does not reach no involvement of third parties
- *However:* Bitcoin comes very close using cryptographic principles

Bitcoins: Principle and Major Issue

- Money is a piece of data
- *Caveat:* Copy piece of data, and spend it twice

"Double Spending"

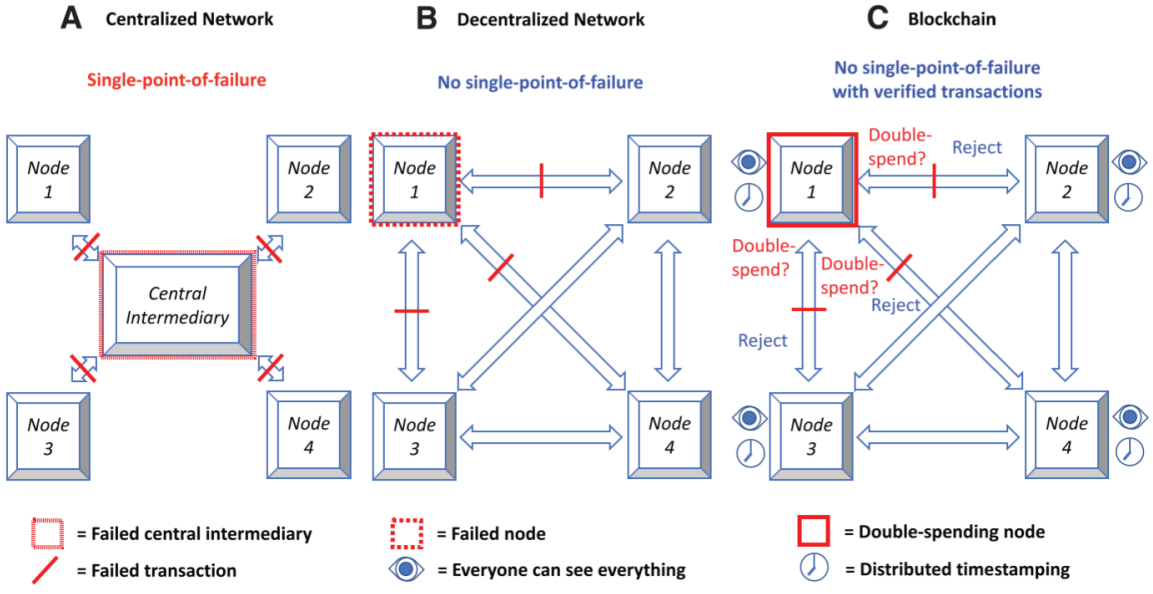
Double Spending



From Kuo et al., 2018

- As of today, no solution without central authority conceivable
- *Issue:* Adding unique identifiers to pieces of data (= coins!) requires central server to keep track of identities of coins
- *Bitcoin:* Don't worry – let double spending happen, detect it afterwards, and reverse it in the shortest amount of time possible

Decentralization



From Kuo et al., 2018

Advantages of Blockchains

- No single "point of failure"
- No central authority
- Everyone observing everything suppresses "double spending"

Creating Bitcoin I

- The *creator of Bitcoin* adopted the pseudonym *Satoshi Nakamoto*.
- Female or male, one or several people? Nobody knows.
- Started coding in May 2007; claimed domain `bitcoin.org` in August 2008
- Released white paper in October 2008; soon thereafter released the code
- By December 2010, others had taken over maintenance

Creating Bitcoin II

- *Fun fact:* Wikipedia planned to dismiss Bitcoin mid 2010 because of missing relevance
- Bitcoin was the first decentralized platform to work; many concepts were entirely new, circumventing various patents for electronic cash systems released by others
- Reasons for anonymity:
  - Just for fun...
  - Legal worries: founders of ”Liberty” and ”e-Gold” accused for money laundering, guilty plea shortly before spring 2008
  - Satoshi, likely, is stinking rich, as possessing lots of bitcoins...

Materials / Outlook

- See *Bitcoin and Cryptocurrency Technologies*, Preface
- See `https://bitcoinbook.cs.princeton.edu/` for further resources
- Further: T. Kuo, H.Kim and L. Ohno-Machado (2017): *Blockchain ditributed ledger technologies for biomedical and health care applications*
- Next lecture: "Cryptography I"
  - See *Bitcoin and Cryptocurrency Technologies* 1.2–1.4, 2.1
  - The Internet Society (2006). `https://www.rfc-editor.org/rfc/rfc4634`, page 6