

Secure Data Processing



Prof. Dr. E. Rahm
und Mitarbeiter

Seminar, WS 2018/19

NEW CHALLENGES

- Cloud
- Social media
- CryptoCurrencies



ScaDS DRESDEN LEIPZIG **UNIVERSITÄT LEIPZIG**

OUTSOURCED PRIVATE DATA

Alice: read (a), write (c, data), read (b) → source → ACK → amazon cloud drive, Dropbox, Google Drive, OneDrive

Security Concerns?
Confidentiality of Data

Solution:
Encryption

© Amr El Abbadi 3

ScaDS DRESDEN LEIPZIG **UNIVERSITÄT LEIPZIG**

BUILDING TOOLS: ENCRYPTION


Plaintext: Hello World! → Encryption → Ciphertext: f559e6da5e9efb90c34cf27170 1fad34ba5952f9

Ciphertext: f559e6da5e9efb90c34cf27170 1fad34ba5952f9 → Decryption → Plaintext: Hello World!

Key: K

<p>Deterministic</p> <p>AES + EBC Electronic Codebook Mode</p> <p>$ENC_K(X) = ENC_K(X)$</p>	<p>Non-deterministic</p> <p>AES + CBC Cipher Block Chaining Mode</p> <p>$ENC_K(X) \neq ENC_K(X)$</p>
---	--

© Amr El Abbadi 4



SECURE SQL?

UNIVERSITÄT
LEIPZIG

GOAL: Developing algorithms that can answer queries over securely outsourced data without fetching all data


```

SELECT SUM(price) AS total
FROM orders
WHERE 10 <= price AND city = 'Vienna'
GROUP BY order_id
HAVING total > 20

```

comparison of entries for equality
keyword search: search for pattern
range query: comparison of numerical value
aggregation

© Amr El Abbadi 5



ENCRYPTION

UNIVERSITÄT
LEIPZIG

- **Homomorphic encryption:** allows some computations on ciphertext without decrypting
 - partially homomorphic (e.g. additive / multiplicative)
 - fully homomorphic (quite inefficient)

- **Order-preserving encryption** (Agrawal et al, Sigmod2004)
 - standard database indexes can be used
 - vulnerable to statistical attacks

6

ScaDS
DRESDEN LEIPZIG

FULL-FLEDGED SYSTEMS

UNIVERSITÄT
LEIPZIG

CryptDB, SOSP'11

MONOMI,
VLDB'13

TrustedDB,
SIGMOD'11

hardware
assisted

Cipherbase, CIDR'13

© Amr El Abbadi

- Microsoft „Always Encrypted“ database (based on Cipherbase)

7

ScaDS
DRESDEN LEIPZIG

DATA PRIVACY

UNIVERSITÄT
LEIPZIG

Privacy

- right of individuals to determine by themselves when, how and to what extent information about them is communicated to others (Agrawal 2002)

Privacy threats

- extensive **collection of personal/private information** / surveillance
- Information dissemination: **disclosure** of sensitive/confidential information
- Invasions of privacy: **intrusion attacks** to obtain access to private information
- Information aggregation: **combining data**, e.g., to enhance personal profiles or identify persons (de-anonymization)

Challenge:

- preserve privacy despite need to use person-related data for improved data analysis

8




PRIVACY FOR BIG DATA

UNIVERSITÄT
LEIPZIG

- need for comprehensive privacy support (“privacy by design”)
- privacy-preserving publishing of datasets
 - anonymization of datasets
- privacy-preserving data mining
 - analysis of anonymized data without re-identification
- privacy-preserving record linkage
 - object matching with encoded data to preserve privacy
 - prerequisite for privacy-preserving data mining

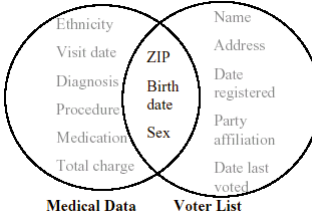
9



RE-IDENTIFICATION OF „ANONYMOUS DATA“ (SWEENEY 2001)

UNIVERSITÄT
LEIPZIG

- US voter registration data
 - 69% unique on postal code (ZIP) and birth date
 - 87% US-wide with sex, postal code and birth data



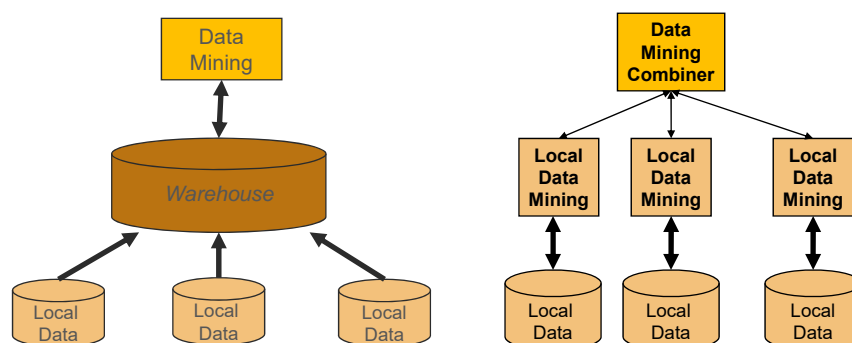
- solution approach: **K-Anonymity**
 - any combination of values appears at least k times
 - generalize values, e.g., on ZIP or birth date

10

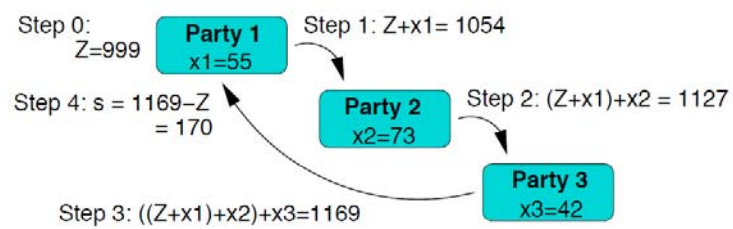
- statistical approach to derive accurate analysis results despite systematic changes to data/query answers, e.g. randomized response
- Example: randomize Yes/No answer (e.g. „Have you ever used illegal drugs?“)
 - throw coin - if head: answer correctly, if tail: throw coin again and answer correctly if head
- most approaches assume **trusted central party**
 - has access to raw data and performs data perturbation for query answers
- **Local differential privacy**: eliminate trusted party
 - Google Rappor prototype for browser accesses
 - Apple smartphones (since IOS 10)

11

- physically integrated data (e.g. data warehouse) about persons entails greatest privacy risks
- data mining over distributed data can better protect personal data by limiting data exchange, e.g., using SMC (secure multiparty computation) methods



- compute a function across several parties, such as no party learns the information from the other parties, but all receive the final results
- example 1: millionaire problem
 - two millionaires, Alice and Bob, are interested in knowing which of them is richer but without revealing their actual wealth.
- example 2: secure summation




13

Bitcoin & Blockchain



14




BITCOIN

UNIVERSITÄT
LEIPZIG

- decentralized, peer-to-peer electronic cash system
- digital identities/signatures
 - public/private key pair
- ledger
 - the balance of each identity – saved in a blockchain (instead of a central bank database)
- transactions
 - move bitcoins from one to another
 - concurrency control to serialize transactions
 - typically backed by transaction log (blockchain)
 - log is persistent (replicated across all network nodes), immutable and tamper-free

L



© Amr El Abbadi 15

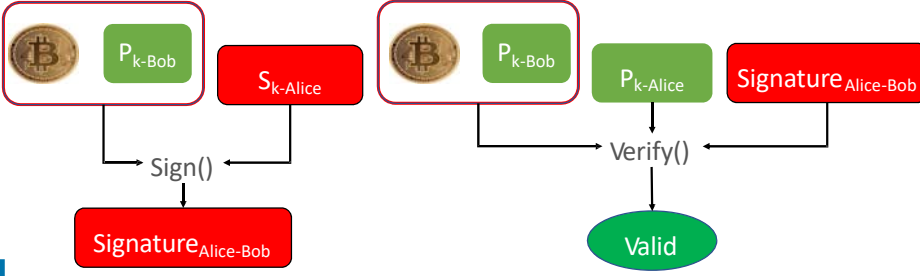


DIGITAL SIGNATURES AND BITCOIN

UNIVERSITÄT
LEIPZIG


- a bitcoin is a chain of digital signatures
 - coin owners digitally sign their coins to transfer them to other recipients
 - Alice wants to move a bitcoin to Bob

 key S_k
 public key P_k
 to verify signature S_k
 (authentication)



L

© Amr El Abbadi 16



HASHING H(X)

UNIVERSITÄT
LEIPZIG

- signatures and public keys are combined using hashing
 - takes any string x of any length as input
 - fixed output size (e.g., 256 bits)
 - efficiently computable
- bitcoin uses SHA-256


$$\text{SHA256} (\text{Signature}_{\text{Alice-Bob}} \ P_{k\text{-Diana}}) =$$

256-bit (32-byte) unique string

- satisfies:
 - *collision free*: no two x, y s.t. $H(x) = H(y)$
 - *hiding*: Given $H(x)$ infeasible to find x (one-way hash function)


L

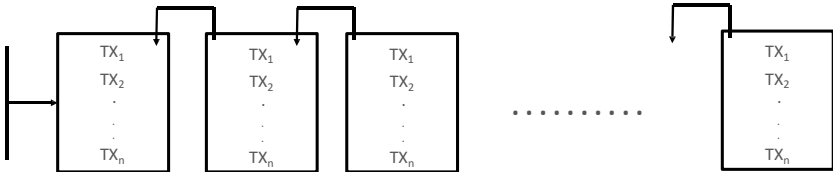
© Amr El Abbadi 17



WHAT IS THE LEDGER?

UNIVERSITÄT
LEIPZIG

- Blockchain 
- transactions are grouped into blocks
 - blocks are chained to each other through pointers (hence blockchain)



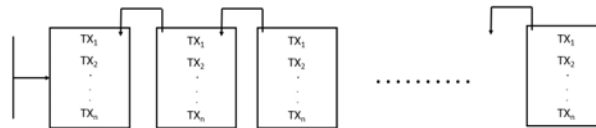
L

© Amr El Abbadi

- Where is the ledger stored?
 - each network node maintains its copy of the ledger
- How is the ledger tamper-free?

blocks are connected through **hash-pointers**

 - each block contains the hash of the previous block
 - this hash gives each block its location in the blockchain
 - tampering with the content of any block can easily be detected



© Amr El Abbadi


SEMINAR

ScaDS
DRESDEN LEIPZIG

SEMINAR GOALS

UNIVERSITÄT
LEIPZIG

- Learn to know about a new topic of scientific and practical importance
 - can be basis for bachelor/master thesis
- student tasks
 - study scientific literature to prepare presentation and written summary on 1 topic
 - presentation
 - discussion
 - summarizing article
- mentoring co-worker provide help and feedback





ScaDS
DRESDEN LEIPZIG

SEMINAR

UNIVERSITÄT
LEIPZIG

- **Master Informatik**
 - part of module „Modern database technologies“
 - seminar module
- **Bachelor Informatik**
 - seminar module






SEMINAR DETAILS

UNIVERSITÄT
LEIPZIG

- **presentation with discussion (45 minutes)**
 - **slides should be in English**
 - talk in German or English
 - discuss slides with mentor beforehand
- **article/report (ca. 15 pages)**
 - discuss/iterate with mentor
 - final deadline March 31, 2019
- **active participation in all presentations**
 - module workload: 30 h presence, 120 h self study
- **successful seminar requires both: talk/discussion + report**



SEMINAR (3)

UNIVERSITÄT
LEIPZIG

- **Topic assignment**
 - **meet mentor within two weeks, i.e., until Nov. 9th, 2018**
 - otherwise seminar registration will become void
 - voluntary leave also until Nov. 9th, 2018
- **Presentation dates**
 - fridays, Ritterstr, starting Jan. 11th 2019
 - max. 4 presentations starting at 1:30 pm

Themen	Betreuer	#Themen	Termin	Studenten
Introduction <ul style="list-style-type: none"> • intro security/encryption • public key & one-way encryption 	Kricke Christen	2	17.1.	Furke Wegler
Secure knowledge bases <ul style="list-style-type: none"> • searchable encryption /secure indexes • privacy-preserving fuzzy search • priv.-pres. search of chemical compounds • encryption on labeled graphs • encrypted dbs / Cipherbase 	Sehili/F S/Franke Franke Wilke Sehili	5		Kühnleitz Möslin
Privacy-preserving data mining <ul style="list-style-type: none"> • methods, metrics, applications • RAPPOR (randomized aggregation) • calibrating noise in private data analysis • LinkMirage: pr.-pres. an. of social relations • Cryptonets: neural networks on encr. data • Differential privacy for SQL • privacy-preserving deep learning 	Christen Christen Christen Rost Rostami Zschache Zschache	7	18.1. 25.1.	Meincke Elksthew Kobold Klein Turke Sager
Intrusion detection <ul style="list-style-type: none"> • Host Intrusion Detection Systems • HIDS: sequence / argument analysis • secure data processing in the cloud & SGX • SgxPectre attacks via speculative execution 	Grimmer/K G/Kricke Gomez Nentwig	4	1-2.	Kreißel Sutayev Nirsberger Bihler
Blockchain <ul style="list-style-type: none"> • introduction cryptocurrencies/blockchain • scalability of blockchains • using blockchains for smart contracts • risks of blockchain technology 	Peukert Saeedi Alkhouri Wilke	4	8.2.	Musiol Kriegel Hu