Privacy-Preserving Data Analysis & Security by Design

Daniel Kraschewski

Big Techday 11

2018-05-18

Motivation

personal data is omnipresent

- internet browsing history
- cell phone movements
- smart metering, smart homes, IoT
- social media, cloud
- ...

Use Cases

TNG TECHNOLOGY CONSULTING

customer analytics

city planning

medical surveys

social surveys

Photos by @nordwood, @hellocolor, @rawpixel, @jaseess on Unsplash

GDPR (from 2018-05-25)

- strong notion of consent
  - informed
  - freely given
  - specific
  - unambiguous
  - clear affirmative act

- high fines (up to 4% of annual turnover) for data privacy violations

Privacy-Preserving Data Analysis

- derive large-scale statistical insights
- still preserve individual's privacy
- security by design
- provable security/privacy

TNG TECHNOLOGY CONSULTING

privacy-preserving results

- aggregated statistics must not reveal personal information
- proper formal notion of anonymity
- Differential Privacy, Laplace Mechanism

privacy-preserving computation

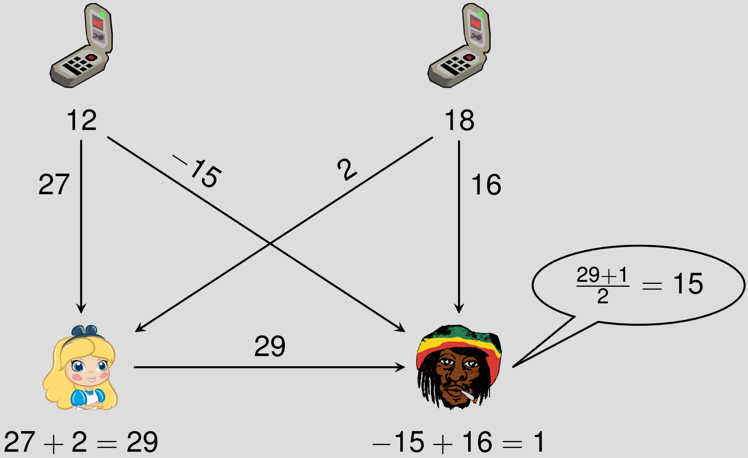
- personal data must be unaccessible even for system owner
- Cryptography
- Secret Sharing, Secure Multi-Party Computation

privacy-preserving environment

- no unauthorized third party may access any data
- IT-Security
- Access Control, encrypted & authenticated channels

Photo by @kristina on Unsplash

Example: Privacy-Preserving Averaging

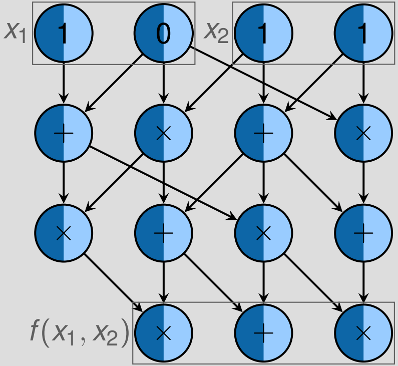


general setting

- mutually mistrusting parties $P_1, P_2, \ldots$
- secret inputs $x_1, x_2, \ldots$
- want to compute some agreed on function value $f(x_1, x_2, \ldots)$
- nothing but $f(x_1, x_2, \ldots)$ should be revealed about $x_1, x_2, \ldots$

a universal solution

1. write f as arithmetic circuit
2. transform each x_i into unintelligible *Secret Sharing*
3. evaluate f gate-by-gate, preserving *Secret Sharing*
4. recombine result



A Concrete Protocol for General MPC

Shared Addition

x_A, y_A $\quad x_B, y_B$ $\quad \rightsquigarrow \quad$ z_A $\quad z_B$ $\quad$ s.t. $z_A + z_B = x + y$

easy: $z_A = x_A + y_A$ and $z_B = x_B + y_B$

Shared Multiplication

x_A, y_A $\quad x_B, y_B$ $\quad \rightsquigarrow \quad$ z_A $\quad z_B$ $\quad$ s.t. $z_A + z_B = x \times y$

problematic: $z = x_A y_A + x_A y_B + x_B y_A + x_B y_B$

missing building block

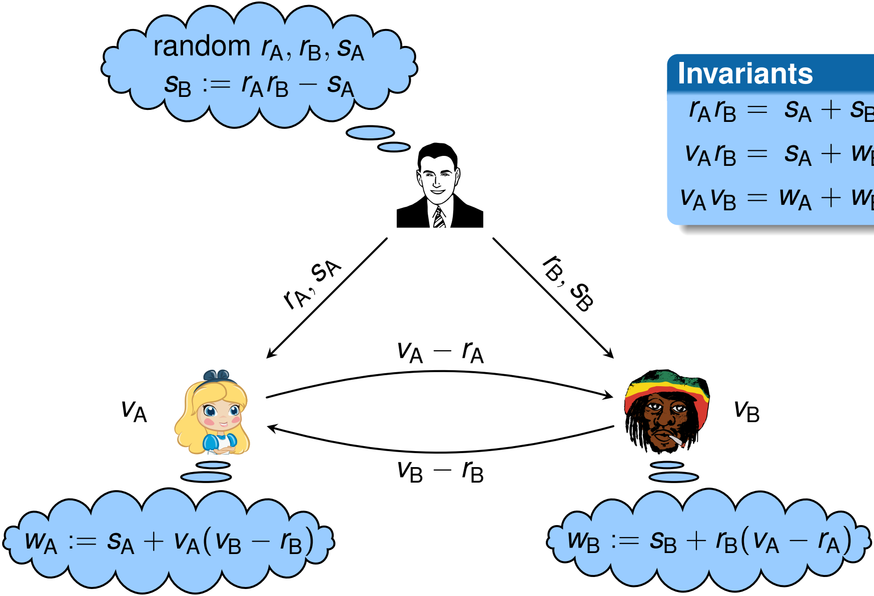
   

v_A $\quad v_B$ $\quad \rightsquigarrow \quad$ w_A $\quad w_B$ $\quad$ s.t. $w_A + w_B = v_A v_B$

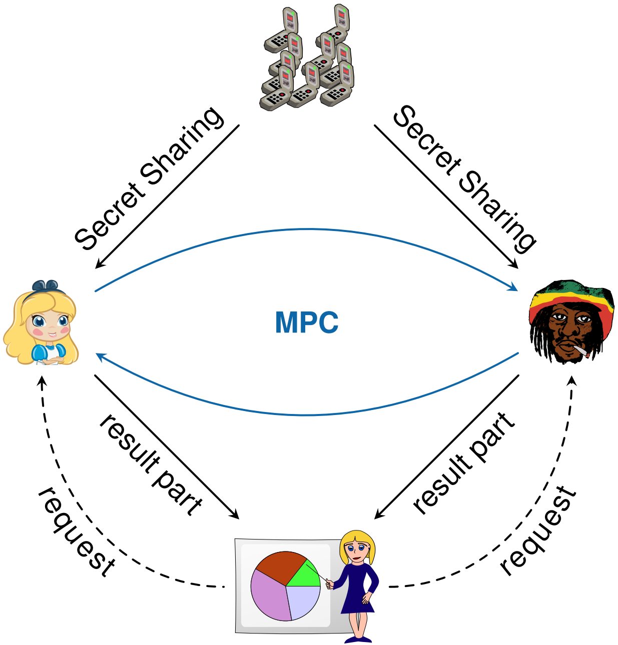
Building Block for Shared Multiplication

Invariants

$$r_A r_B = s_A + s_B$$
$$v_A r_B = s_A + w_B$$
$$v_A v_B = w_A + w_B$$



Application to Privacy-Preserving Data Analysis

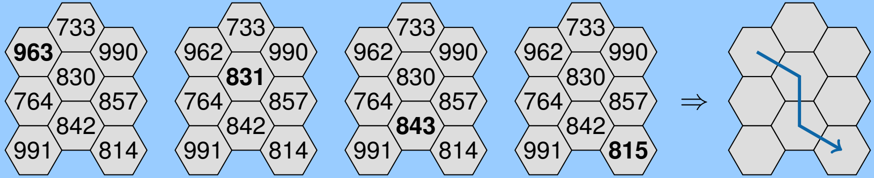


k-Anonymity

- published data must coincide with at lest *k* individuals

De-anonymization attack on correlated data[1]

- published data: number of people in mobile cell at time t_i



$t_1 = 02{:}30$ $t_2 = 03{:}00$ $t_3 = 03{:}30$ $t_4 = 04{:}00$

- trajectory recovery $\widehat{=}$ optimization problem
  - higher “costs” for sudden/far movements
  - higher “costs” for irregular movements and/or movements at night
- 50% – 91% accuracy, depending on space-time resolution

[1] Fengli Xu, Zhen Tu, Yong Li, Pengyu Zhang, Xiaoming Fu, Depeng Jin: Trajectory Recovery From Ash: User Privacy Is NOT Preserved in Aggregated Mobility Data, 26th International Conference on World Wide Web (WWW 2017)

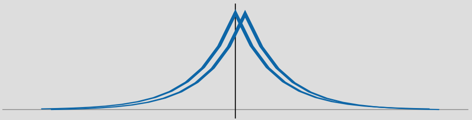
ε-Differential Privacy

- statistical similarity: κ(real data) $\approx$ κ(real data $\setminus$ me) up to factor e^{ε}

Laplace Mechanism

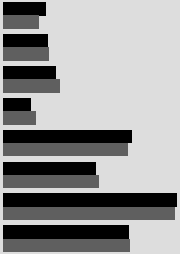
1. calculate histogram
2. add Laplace noise
3. output noisy group sizes

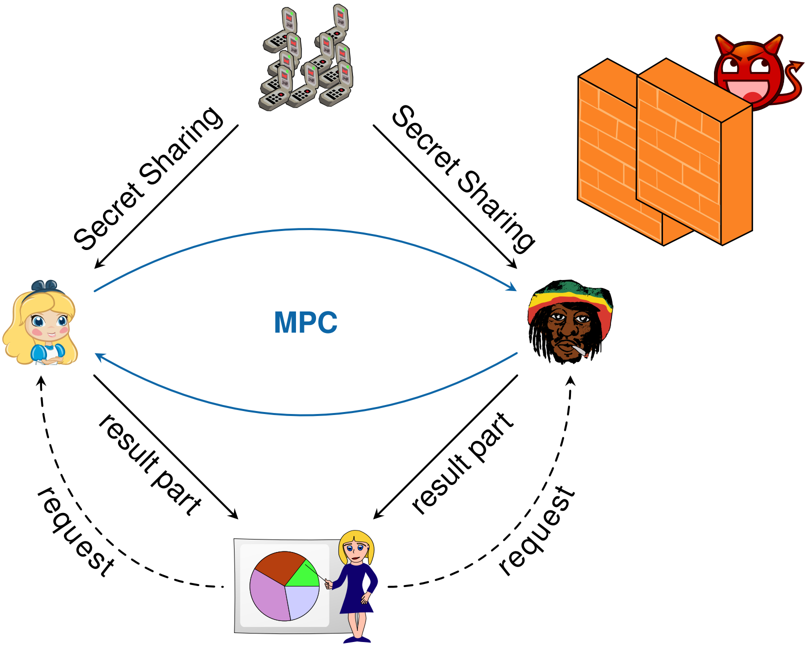
Laplace Distributions



Example histogram ($\frac{1}{10}$-Differential Privacy)

Age	Sex	Diagnosis	count	*noise*	result
< 35	f	infection	48	*9*	57
< 35	f	NCD	61	*-1*	60
< 35	m	infection	75	*-5*	70
< 35	m	NCD	44	*-7*	37
≥ 35	f	infection	165	*6*	171
≥ 35	f	NCD	127	*-4*	123
≥ 35	m	infection	228	*2*	230
≥ 35	m	NCD	168	*-2*	166




Secret Sharing
Secret Sharing
MPC
result part
result part
request
request

integration

Strictness

- fail-safe defaults
- need-to-know principle
- principle of least privilege

Robustness

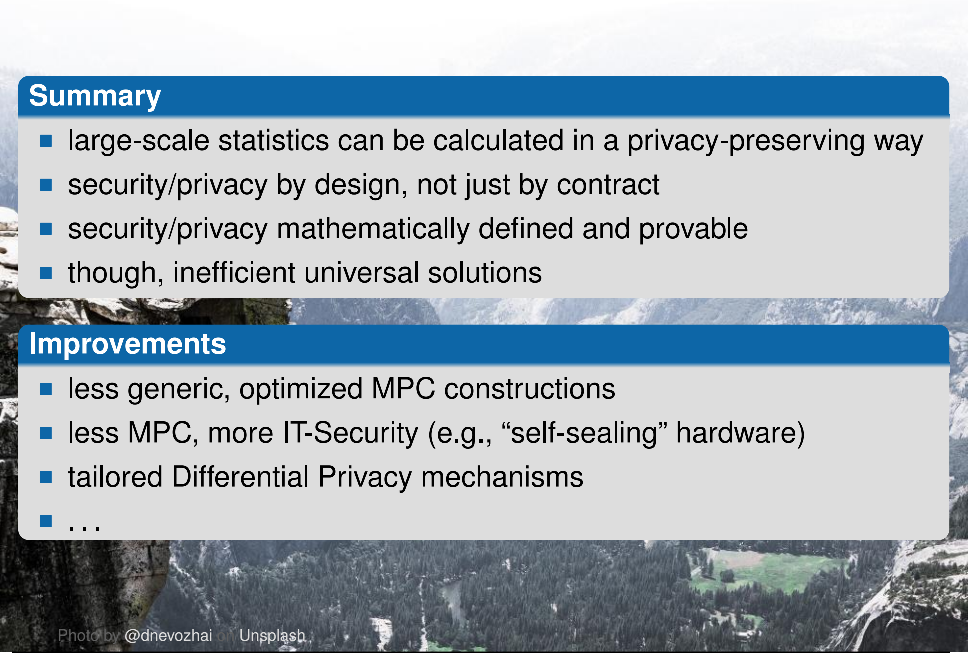
- separation of duties
- multi-factor/layered security
- forward secrecy

Consistency

- complete security-model
- defense in depth
- homogeneity/uniformity

simplicity

TNG TECHNOLOGY CONSULTING



Summary

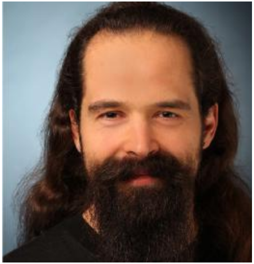
- large-scale statistics can be calculated in a privacy-preserving way
- security/privacy by design, not just by contract
- security/privacy mathematically defined and provable
- though, inefficient universal solutions

Improvements

- less generic, optimized MPC constructions
- less MPC, more IT-Security (e.g., “self-sealing” hardware)
- tailored Differential Privacy mechanisms
- …

Photo by @dnevozhai on Unsplash

Thank you for your attention!



Dr. Daniel Kraschewski
Senior Consultant

TNG Technology Consulting GmbH
Betastraße 13a
85774 Unterföhring

tel: +49 89 2158 9960
fax: +49 89 2158 9969

daniel.kraschewski@tngtech.com

Cliparts by AhNinniah, Anonymous, gswanson, j4p4n, qubodup, TikiGiki on openclipart, and succo on pixabay