Anonymisieren, Pseudonymisieren, Maskieren:

Daten-Cocktails richtig mixen!

AGENDA

→ Worum geht‘s eigentlich?

→ Was ist Anonymisierung & Pseudonymisierung?

→ Welche Methoden gibt‘s?

Anonymisieren & Pseudonymisieren

Worum geht‘s?

Nice Facts zur Digitalisierung:

→ Immer mehr Daten werden generiert & verarbeitet.

→ ~ 21 Milliarden IoT-Geräte bereits im Internet
(~ 4 Milliarden Menschen haben Internetzugang)

→ Weltweiter Umsatz durch Big-Data
in 2016: ~ 1,7 Milliarden $
in 2020: ~ 9,4 Milliarden $

→ „Daten sind das Öl des 21. Jahrhunderts.“

Sad Facts:

→ Daten werden zu wenig geschützt.

→ Immer mehr Data Leaks & Data Breaches passieren.

→ Weltweiter Schaden durch Cyberkriminalität in 2018: ~ 600 Milliarden $

→ Nur ~ 55% der österreichischen Unternehmen waren im Mai 2018 „GDPR-fit“

Collections #1-5

https://blog.fefe.de/?ts=a2b06ef1

Fefes Blog

Wer schöne Verschwörungslinks für mich hat: ab an felix-bloginput (at) fefe.de!

Mon Jan 28 2019

- [1] Sagten wir 700 Mio Accounts? Wir meinten eher so 2,2 Mrd Accounts!

  Old and busted: Daten sind das Öl des 21. Jahrhunderts.

  New hotness: Daten sind die Ölpest des 21. Jahrhunderts. (Danke, Matthias)

https://www.heise.de/security/meldung/Neue-Passwort-Leaks-Insgesamt-2-2-Milliarden-Accounts-betroffen-4287538.html

heise online heise+ Anmelden

IT Mobiles Entertainment Wissen Netzpolitik Wirtschaft Journal

TOPTHEMEN: DSGVO WINDOWS 10 ANDROID AMAZON KI ELEKTROAUTOS ANZEIGE: ZUKUNI

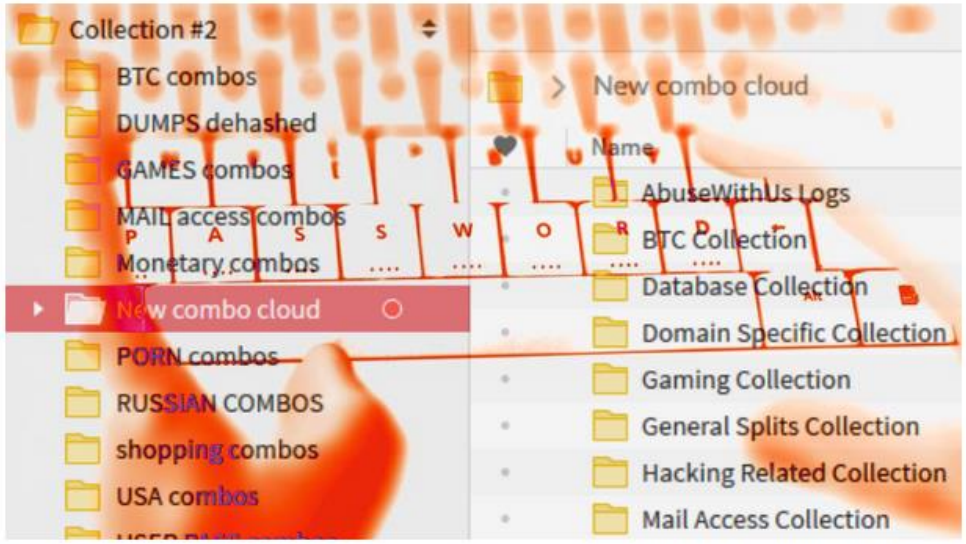
Security 〉7-Tage-News 〉01/2019 〉Neue Passwort-Leaks: Insgesamt 2,2 Milliarden Accounts betroffen

25.01.2019 12:51 Uhr | Security

Neue Passwort-Leaks: Insgesamt 2,2 Milliarden Accounts betroffen

Nach der Passwort-Sammlung Collection #1 kursieren nun auch die riesigen Collections #2-5 im Netz. So überprüfen Sie, ob Ihre Accounts betroffen sind.

Von Ronald Eikenberg 577



(Bild: plantic\Shutterstock.com, Montage: heise Online)

Uber Data Breach

https://www.heise.de/newsticker/meldung/Uber-verschwieg-Daten-Diebstahl-bei-50-Millionen-Kunden-3897154.html

heise online heise+ Anmelden Suchen

IT Mobiles Entertainment Wissen Netzpolitik Wirtschaft Journal News

TOPTHEMEN: DSGVO WINDOWS 10 ANDROID AMAZON KI ELEKTROAUTOS ANZEIGE: ZUKUNFTSMACHER

heise online › News › 11/2017 › Uber verschwieg Daten-Diebstahl bei 50 Millionen Kunden

22.11.2017 08:14 Uhr

Uber verschwieg Daten-Diebstahl bei 50 Millionen Kunden

An Skandale um den Fahrdienst-Vermittler Uber konnte man sich schon gewöhnen, doch eine neue Enthüllung offenbart schockierende Verantwortungslosigkeit. Hacker stahlen schon 2016 Daten von 57 Millionen Fahrgästen und Fahrern, Uber verschwieg das.

dpa 86

(Bild: dpa, Christoph Dernbach)

https://businessnewswales.com/uber-data-breach-lesson-using-cloud-technology/

Business NewsWales HOME SECTORS REGIONS EVENTS BUSINESS ZONE

"Hackers accessed Uber's private development area within GitHub, an online resource for developers – they essentially went in through the tradesman's entrance.

"From here they were able to obtain authentication and login details for Uber's Amazon Web Service (AWS) account, a cloud computing service used by Uber to store data for back-office software development. Once into AWS the hackers accessed a large cache of hosted driver and customer data and then blackmailed Uber with the threat to release this data.

"There is a huge issue there and this particular method of hacking – a back office hack – is an important lesson in the dangers of using Cloud computing for IT development. There are two glaringly obvious questions – why was it possible to access Uber's AWS account at all via its GitHub, and why was development apparently being carried out using 'live' rather than dummy data?

"While the common weakness in most hacks is the human factor, it's tempting to think of this as unsophisticated users falling vulnerable to people with much greater technical knowledge. This does not seem to have been the case here. It seems more likely to be a case of Uber's IT developers being careless and making use of short cuts which exposed the company to the kind of security risks which occurred here.

"I imagine Amazon Web Service may be looking at enforcing its own terms of use against Uber: typically a hosting services provider puts the onus on its customers to safeguard its

Grundsätze des Datenschutzes:

→ Verbotsprinzip

→ Prinzip der Datensparsamkeit

→ Prinzip der engen Zweckbindung

Personenbezogene Daten dürfen nicht für
Test- & Entwicklungszwecke verarbeitet werden!

Anonymisierung / Pseudonymisierung notwendig!

Worum geht‘s:

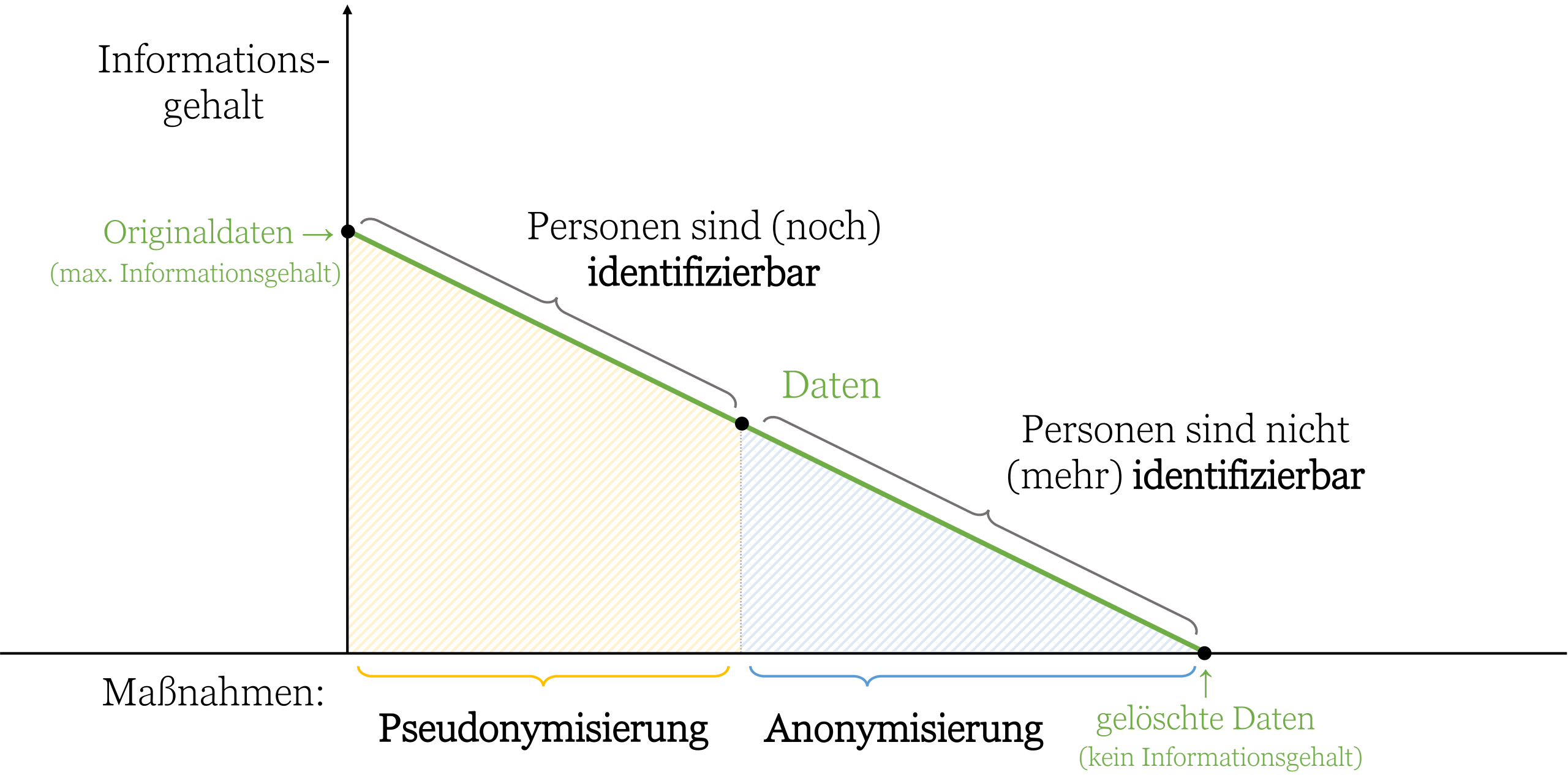
→ Die Angriffsfläche kann präventiv minimiert werden.

→ Es gibt Daten, die nicht als Original gespeichert werden sollten (z.B.: Passwörter).

→ Bei Preproductive Environments sollten verarbeitete Daten anonymisiert bzw. pseudonymisiert sein.

Anonymisieren & Pseudonymisieren

Was ist das eigentlich?


Informations-
gehalt
Originaldaten →
(max. Informationsgehalt)
Personen sind (noch)
identifizierbar
Daten
Personen sind nicht
(mehr) identifizierbar
Maßnahmen:
Pseudonymisierung
Anonymisierung
gelöschte Daten
(kein Informationsgehalt)

Anonymisierung:

→ **Formale Anonymisierung:**
nur direkte Identifikatoren werden entfernt

→ **Faktische Anonymisierung:**
Deanonymisierung kann nicht ausgeschlossen werden

→ **Absolute Anonymisierung:**
Deanonymisierung ist unmöglich

Anonymisierung:

→ **Formale Anonymisierung:**
= rechtlich nicht ausreichend!

→ **Faktische Anonymisierung:**
= Pseudonymisierung

→ **Absolute Anonymisierung:**
= Anonymisierung

Anonymisierung:

→ **GDPR & DSG:**
Anonyme Daten = „absolut anonymisiert“
Anonyme Daten = out of Scope

→ **TKG:**
Absolute Anonymisierung ≙ Löschung

Anonymisierung:

Ab wann gilt eine Person als nicht mehr identifizierbar?

Wenn trotz Berücksichtigung aller Mittel (Kosten, Zeitaufwand, verfügbare Technologie & technologische Entwicklungen) kein Rückschluss auf die Person gezogen werden kann.

Pseudonymisierung:

→ **GDPR & DSG:**
Pseudonymisierte Daten = mit Hinzuziehung weiterer Informationen könnte Person identifiziert werden
Pseudonymisierte Daten = im Scope

Achtung:
Informationen, die das Identifizieren ermöglichen, müssen gesondert aufbewahrt und geschützt werden!

Data Masking:

→ Oberbegriff für spezifische Methoden der Anonymisierung / Teilanonymisierung

Verfremdung:

→ Oberbegriff,
keine klare Definition bekannt

How to:

Pseudonymisieren

Pseudonymisieren:

ID	Full Name	Sex	Age	Favorite Drink
001	Mimi Musterfrau	W	35	Gin Gimlet

→ **Pseudonymisieren durch Trennen:**
Der Personenbezug wird gesondert gespeichert.

ID	Favorite Drink
001	Gin Gimlet

ID	Full Name	Sex	Age
001	Mimi Musterfrau	W	35

Pseudonymisieren:

ID	Full Name	Sex	Age	Favorite Drink
001	Mimi Musterfrau	W	35	Gin Gimlet

→ **Pseudonymisieren durch Verschlüsseln:**
Die Keys werden gesondert gespeichert.

ID	Key
001	00 01 02 03 04 05 06 07 08 09 0a 0b 0c 0d 0e 0f 10 11 12 13 14 15 16 17 18 19 1a 1b 1c 1d 1e 1f

ID	Full Name	Sex	Age	Favorite Drink
001	IaPvSPNBEU9BjC8hs+XKOA==	Rn…	Ax…	Gin Gimlet

How to:

Anonymisieren

Anonymisieren:

ID	Full Name	Sex	Age	Favorite Drink
001	~~Mimi Musterfrau~~	~~W~~	~~35~~	Gin Gimlet

→ **Anonymisieren durch Löschen:**
Der Personenbezug wird entfernt.

ID	Favorite Drink
001	Gin Gimlet

Anonymisieren:

ID	Full Name	Sex	Age	Favorite Drink
001	Mimi Musterfrau	W	35	Gin Gimlet

→ **Anonymisieren durch Maskieren:**
Maskiert wird durch „aus-X-en“

ID	Full Name	Sex	Age	Favorite Drink
001	XXXXXXXXXXXX	X	XX	Gin Gimlet

Anonymisieren:

ID	Full Name	Sex	Age	Favorite Drink
001	Mimi Musterfrau	W	35	Gin Gimlet

→ **Anonymisieren durch Maskieren:**
Maskiert wird durch Ersetzen mit Random-Werten

ID	Full Name	Sex	Age	Favorite Drink
001	**John Doe**	**M**	**52**	Gin Gimlet

Anonymisieren:

ID	Full Name	Sex	Age	Favorite Drink
001	Mimi Musterfrau	W	35	Gin Gimlet

→ **Anonymisieren durch Verschlüsseln:**
Die Keys werden gelöscht!

ID	Key
001	00 01 02 03 04 05 06 07 08 09 0a 0b 0c 0d 0e 0f 10 11 12 13 14 15 16 17 18 19 1a 1b 1c 1d 1e 1f

ID	Full Name	Sex	Age	Favorite Drink
001	IaPvSPNBEU9BjC8hs+XKOA==	Rn…	Ax…	Gin Gimlet

k-Anonymisierung:

ID	Full Name	Sex	Age	Favorite Drink
001	Mimi Musterfrau	W	35	Gin Gimlet
002	Frodo Beutlin	M	33	Cosmopolitan
003	Jolly Jumper	M	22	Mojito
004	Luky Luke	M	40	Bier
005	John Doe	W	51	White Russian
006	Hermine Granger	W	20	Butterbier
007	Homer Simpson	M	41	Bier
008	Michael Häupl	M	69	Spritzwein
009	Lara Croft	W	43	Vanilla on Ice
010	Marilyn Monroe	W	30	Dom Perignon

k-Anonymisierung:

ID	Full Name	Sex	Age	Favorite Drink
*	*	W	21–40	Gin Gimlet
*	*	M	21–40	Cosmopolitan
*	*	M	21–40	Mojito
*	*	M	21–40	Bier
*	*	W	41–70	White Russian
*	*	W	21–40	Butterbier
*	*	M	41–70	Bier
*	*	M	41–70	Spritzwein
*	*	W	41–70	Vanilla on Ice
*	*	W	21–40	Dom Perignon

k-Anonymisierung:

Äquivalenzklasse	Full Name	Sex	Age	Favorite Drink
A	*	W	21–40	Gin Gimlet
B	*	M	21–40	Cosmopolitan
B	*	M	21–40	Mojito
B	*	M	21–40	Bier
C	*	W	41–70	White Russian
A	*	W	21–40	Butterbier
D	*	M	41–70	Bier
D	*	M	41–70	Spritzwein
C	*	W	41–70	Vanilla on Ice
A	*	W	21–40	Dom Perignon

k-Anonymisierung:

Äquivalenzklasse	Full Name	Sex	Age	Favorite Drink
A	*	W	21–40	Gin Gimlet
	*	W	21–40	Butterbier
	*	W	21–40	Dom Perignon
B	*	M	21–40	Cosmopolitan
	*	M	21–40	Mojito
	*	M	21–40	Bier
C	*	W	41–70	White Russian
	*	W	41–70	Vanilla on Ice
D	*	M	41–70	Bier
	*	M	41–70	Spritzwein

k-Anonymisierung:

→ **k = 2**
Die kleinsten Äquivalenzklassen (C, D)
haben jeweils 2 Datensätze.
Je höher k, desto höher die Anonymisierung.

→ **Schwächen:**
Homogeneity-Attacks
Background-Knowledge-Attacks

k-Anonymisierung:

→ **Homogeneity-Attack:**
Die Attribute in den k-Gruppen sind zu wenig divers.

Beispiel:

Mimi lässt sich im Krankenhaus behandeln.
Der Angreifer kennt Mimi, ihr Geschlecht und ihr ungefähres Alter.
Der Angreifer erlangt Zugriff auf die k-anonymisierten Daten des Krankenhauses.

k-Anonymisierung:

→ **Homogeneity-Attack:**
Die Attribute in den k-Gruppen sind zu wenig divers.

Beispiel:

Äquivalenzklasse	Full Name	Sex	Age	Diagnosis
A	*	W	21-40	Nagelpilz
	*	W	21-40	Nagelpilz
	*	W	21-40	Nagelpilz
B	*	W	41-70	Borreliose
	*	W	41-70	Borreliose

k-Anonymisierung:

→ **Homogeneity-Attack:**
Die Attribute in den k-Gruppen sind zu wenig divers.

Beispiel:

Der Angreifer erfährt, dass Mimi Nagelpilz hat.

l-Diversität:

Äquivalenzklasse	Full Name	Sex	Age	Favorite Drink
A	*	W	21–40	Gin Gimlet
	*	W	21–40	Butterbier
	*	W	21–40	Dom Perignon
B	*	M	21–40	Cosmopolitan
	*	M	21–40	Mojito
	*	M	21–40	Bier
C	*	W	41–70	White Russian
	*	W	41–70	Vanilla on Ice
D	*	M	41–70	Bier
	*	M	41–70	Spritzwein

l-Diversität:

→ Eine k-Gruppe ist l-divers, wenn das betroffene Attribut zumindest l verschiedene Ausprägungen aufweist.

→ l = 2
Die Äquivalenzklassen weisen mindestens 2 unterschiedliche Datensätze auf.
Je höher l, desto höher die Diversität in der k-Anonymisierung.

Hash-Verfahren:

→ **Einweg-Eigenschaft:**
Der Klartext kann nicht mehr bestimmt werden.
Hashing = Form der Anonymisierung

→ Niemals auf den Salt vergessen!

Last, but not least...

Viele von uns wenden **k-Anonymisierung** täglich an!

Q&A

maha.sounble@gmail.com