VORLESUNG DATENSCHUTZ

SOMMERSEMESTER 2024

Durchführung:

Benjamin Bremert und

Beschäftigte des Unabhängigen Landeszentrums

für Datenschutz Schleswig-Holstein (ULD), Kiel

Ansprechpartner: Benjamin Bremert <benjamin@bremert.de>

Rechtsauffassungen sind solche der jeweiligen Referent:innen.

Plattform Privatheit

Datenschutz und Technik IV
– Datenschutzfördernde Technik –

CAU-Vorlesung, 03.06.2024

Dr. h.c. Marit Hansen
Landesbeauftragte für Datenschutz Schleswig-Holstein

Hinweis:
Rechtsauffassungen sind solche der jeweiligen Referent:innen.

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Begriffe in der DSGVO: personenbezogen

Art. 4 Nr. 1 DSGVO:

> **„personenbezogene Daten" alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person [...] beziehen;**
>
> als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind, identifiziert werden kann;

„nicht personenbezogen" = „anonym"

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Begriffe in der DSGVO: Pseudonymisierung

Art. 4 Nr. 5 DSGVO:

„Pseudonymisierung" die Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten **ohne Hinzuziehung zusätzlicher Informationen nicht mehr** einer spezifischen betroffenen Person zugeordnet werden können,

sofern diese zusätzlichen In ↓ gezielte Nichtverfügbarkeit bewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden;

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Begriffe in der DSGVO: Pseudonymisierung

Art. 4 Nr. 5 DSGVO (visualisiert)

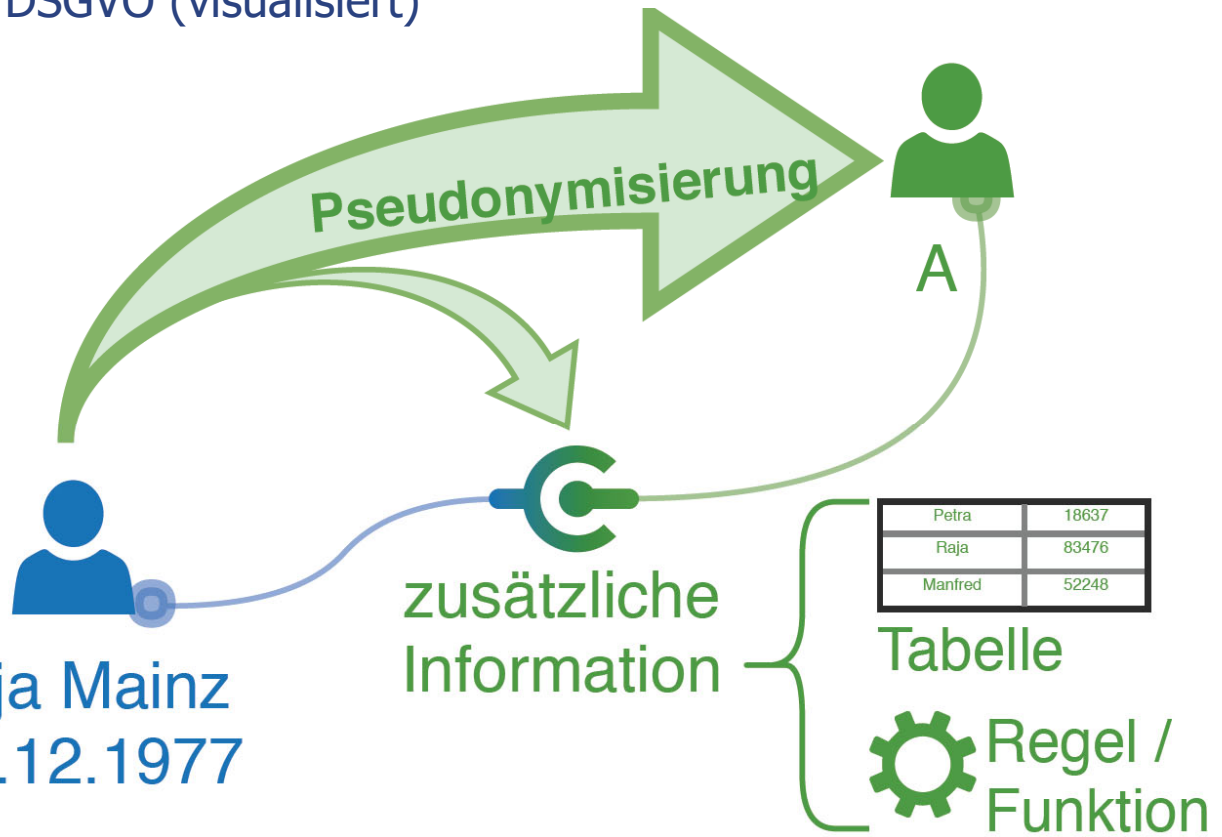
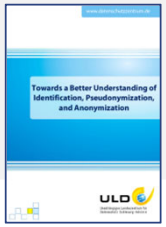


Bild: Benjamin Walczak (ULD)

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion



https://uldsh.de/PseudoAnon

Verschiedene Arten von „zusätzlichen Informationen"

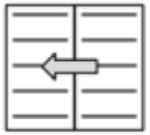
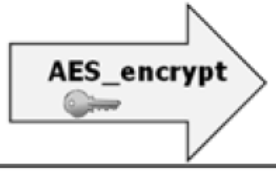
type of additional information	forward function identified ⇨ pseudonymous	inverse function pseudonymous ⇨ identified
lookup-based bi-directional	 lookup table	 lookup table
formula-based bi-directional		
formula-based one-directional		X

Figure 26: Examples of different types of additional information.

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Unterschiede Pseudonymisierung ↔ Anonymisierung

- **Pseudonymisierung**
  Verarbeitung dergestalt, dass

  aus personenbezogenen Daten [Input]

  veränderte Daten
  (pseudonymisierte Daten) [Output]

  werden,
  die dann nur mit Hilfe „zusätzlicher Informationen" einer spezifischen Person zugeordnet werden können.

  (Pseudonymisierung ist selbst Verarbeitung i.S.d. DSGVO.)

- **Anonymisierung**
  Verarbeitung dergestalt, dass

  aus personenbezogenen Daten [Input]

  veränderte Daten ohne Personenbezug
  (anonymisierte Daten) [Output]

  werden.

  (Anonymisierung ist selbst Verarbeitung i.S.d. DSGVO.)

weiterhin personenbezogen!

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Art. 29-Datenschutzgruppe 2014 (DSGVO-kompatibles Update ca. 2024)

- Art. 29 Data Protection Working Party:
  Opinion 05/2014 on "Anonymisation Techniques" (WP 216)
  https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

- Drei Risiken, die zur Identifizierung von Personen führen können:
  - Singling out
  - Linkability
  - Inference

- Anonymisierung möglich durch „randomization" oder „generalization"
- Vielfache Möglichkeiten der Pseudonymisierung

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Beispiele für „Bad Anonymization"

https://www.nytimes.com/2006/08/09/technology/09aol.html

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr. AUG. 9, 2006

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

No. 4417749 conducted hundreds of searches o
month period on topics ranging from "numb fi
"60 single men" to "dog that urinates on everyt

And search by search, click by click, the identit
user No. 4417749 became easier to discern. Th

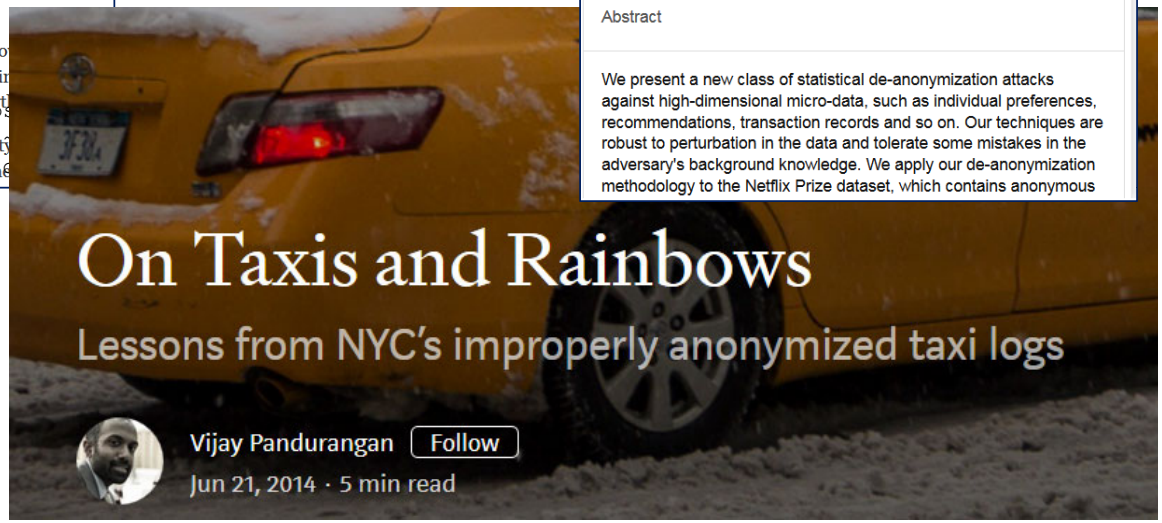
Robust De-anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset)

Article · November 2006 **Arvind Narayanan** **Vitaly Shmatikov**

Abstract

We present a new class of statistical de-anonymization attacks against high-dimensional micro-data, such as individual preferences, recommendations, transaction records and so on. Our techniques are robust to perturbation in the data and tolerate some mistakes in the adversary's background knowledge. We apply our de-anonymization methodology to the Netflix Prize dataset, which contains anonymous

https://hackaday.com/2018/01/28/opt-out-fitness-data-sharing-leads-to-massive-military-locations-leak/



https://tech.vijayp.ca/of-taxis-and-rainbows-f6bc289679a1?gi=6f79683ccb1d

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Somewhat pseudonymised
Example: 2006: AOL publishes ~~anonymised~~ search engine requests of 3 months

116874	thompson water seal	2006-05-24 11:31:36	1	http://www.thompsonswaterseal.com
116874	express-scripts.com	2006-05-30 07:56:03	1	http://www.express-scripts.com
116874	express-scripts.com	2006-05-30 07:56:03	2	https://member.express-scripts.com/
116874	knbt	2006-05-31 07:57:28		
116874	knbt.com	2006-05-31 08:09:30	1	http://www.knbt.com
117020	naughty thoughts	2006-03-01 08:33:07	2	http://www.naughtythoughts.com
117020	really eighteen	2006-03-01 15:49:55	2	http://www.reallyeighteen.com
117020	texas penal code	2006-03-03 17:57:38	1	http://www.capitol.state.tx.us
117020	hooks texas	2006-03-08 09:47:08		
117020	homicide in hooks texas	2006-03-08 09:47:35		
117020	homicide in bowie county	2006-03-08 09:48:25	6	http://www.tdcj.state.tx.us
117020	texarkana gazette	2006-03-08 09:50:20	1	http://www.texarkanagazette.com
117020	tdcj	2006-03-08 09:52:36	1	http://www.tdcj.state.tx.us
117020	naughty thoughts	2006-03-11 00:04:40	1	http://www.naughtythoughts.com
117020	cupid.com	2006-03-11 00:08:50		

Quelle: http://www.lunchoverip.com/2006/08/being_user_4417.html

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Number 4417749

Mrs Arnold said she was shocked that her search queries had been recorded and released to the public by AOL.

"My goodness, it's my whole personal life," she said.
"I had no idea somebody was looking over my shoulder."

school supplies for Iraq children

the best season to visit Italy

termites

safest place to live

mature living

tea for good health

nicotine effects on the body

hand tremors

dry mouth

bipolar

60 single men

numb fingers

dog that urinates on everything

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Netflix-Wettbewerb

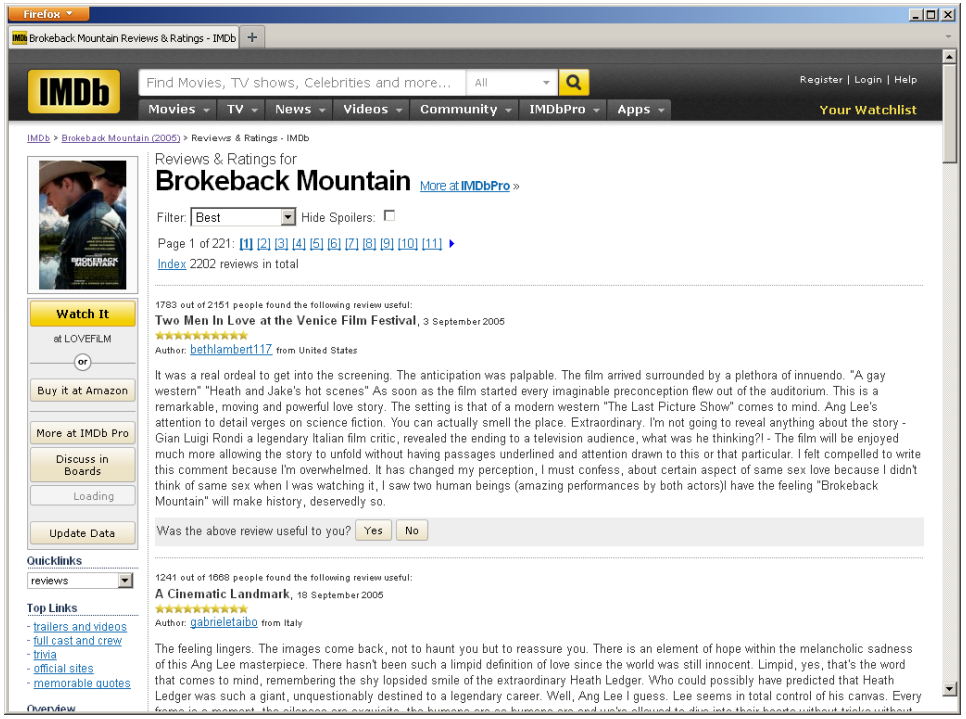
- 2006: 10 Mio. Film-Bewertungen von 500.000 Kunden
- Ziel: besseres Empfehlungssystem
- "Anonymity of the study data is comparable to the strictest Federal standards for anonymizing personal health information."

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Netflix-Wettbewerb

- Arvind Narayanan und Vitaly Shmatikov: Reidentifizierung durch Abgleich mit Internet Movie Database (IMDb) über Verkettung per Ranking und Timestamp
- "Robust De-anonymization of Large Sparse Datasets (How To Break Anonymity of the Netflix Prize Dataset)"
- Z.B. mit 8 Reviews



Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Erwägungsgrund 26 DSGVO (S. 1-2)

„sollten“: Formulierung für Aussagen in den ErwGr.

Die Grundsätze des Datenschutzes sollten für alle Informationen gelten, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen.
Einer Pseudonymisierung unterzogene personenbezogene Daten, die durch Heranziehung zusätzlicher Informationen einer natürlichen Person zugeordnet werden könnten, sollten als Informationen über eine identifizierbare natürliche Person betrachtet werden.
[…]

Zuordenbarkeit zu einer natürlichen Person: personenbezogen

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiel Diskussion

Erwägungsgrund 26 DSGVO (S. 3-4)

„alle Mittel", „nach allgemeinem Ermessen wahrscheinlich"

[…]
Um festzustellen, ob eine natürliche Person identifizierbar ist, sollten alle Mittel berücksichtigt werden, die von dem Verantwortlichen oder einer anderen Person nach allgemeinem Ermessen wahrscheinlich genutzt werden, um die natürliche Person direkt oder indirekt zu identifizieren, wie beispielsweise das Aussondern.

Aussondern = „Singling out"

Bei der Feststellung, ob Mittel nach allgemeinem Ermessen wahrscheinlich zur Identifizierung der natürlichen Person genutzt werden, sollten alle objektiven Faktoren, wie die Kosten der Identifizierung und der dafür erforderliche Zeitaufwand, herangezogen werden, wobei die zum Zeitpunkt der Verarbeitung verfügbare Technologie und technologische Entwicklungen zu berücksichtigen sind.

Heute und in Zukunft

[…]

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Erwägungsgrund 26 DSGVO (S. 5-6)

Anonyme Informationen – außerhalb der DSGVO

[…]
Die Grundsätze des Datenschutzes sollten daher nicht für anonyme Informationen gelten,
d. h. für Informationen, die sich nicht auf eine identifizierte oder identifizierbare natürliche Person beziehen,
oder personenbezogene Daten, die in einer Weise anonymisiert worden sind, dass die betroffene Person nicht oder nicht mehr identifiziert werden kann.
Diese Verordnung betrifft somit nicht die Verarbeitung solcher anonymer Daten, auch für statistische oder für Forschungszwecke.

anonym

anonymisiert

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Faktoren für das Risiko einer Identifizierung

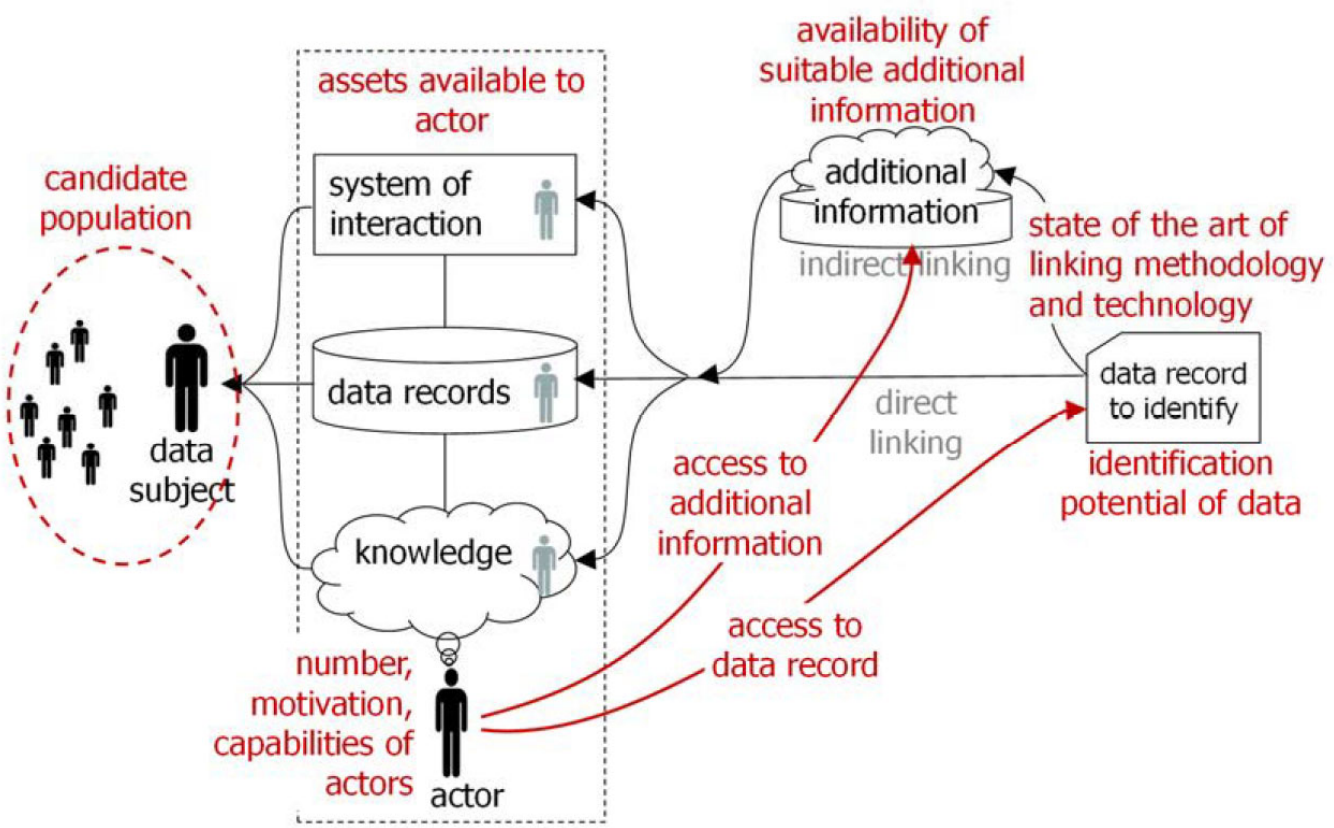
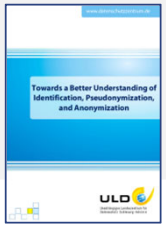


Figure 6: Factors that influence the risk of identification by a given actor.

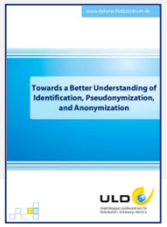


https://uldsh.de/PseudoAnon

Möglichkeiten der Identifizierung verändern sich über die Zeit

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion



https://uldsh.de/PseudoAnon

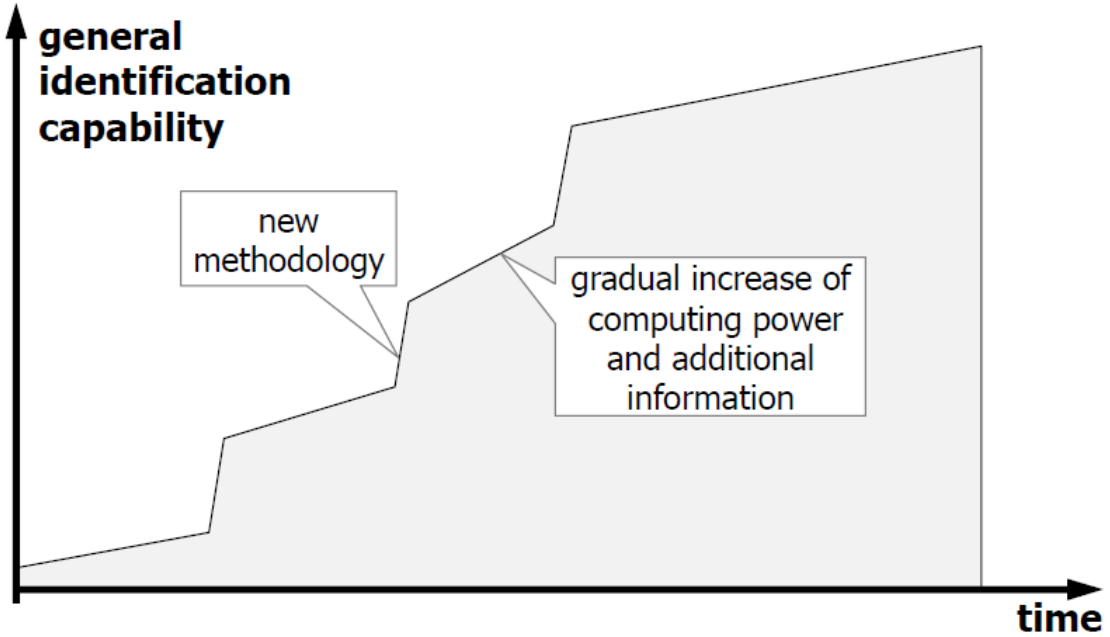
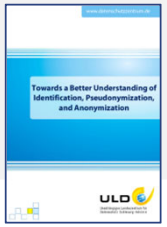


Figure 42: Example of identification capability over time.

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion



https://uldsh.de/PseudoAnon

Kontinuierlich identifizierbare Daten

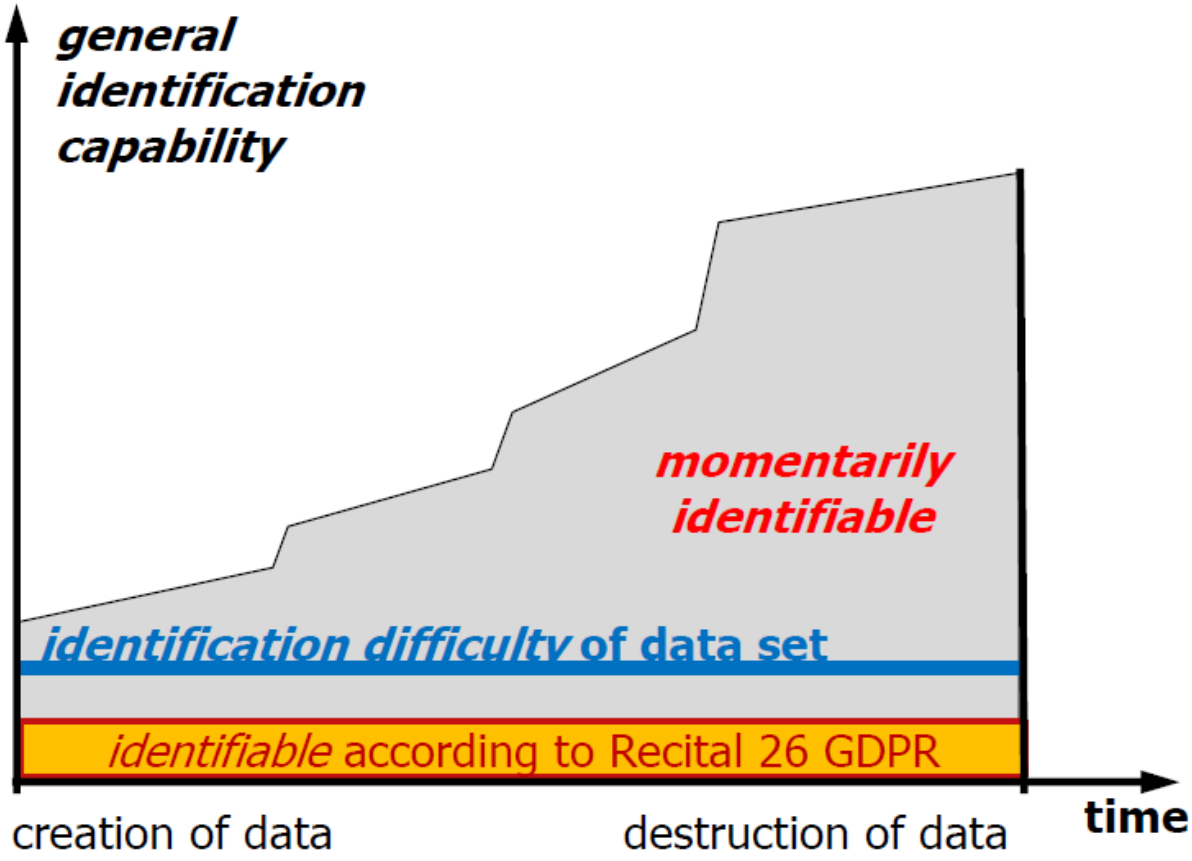
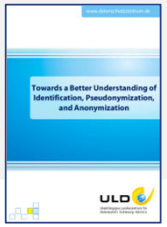


Figure 43: Continuously identifiable data.

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion



https://uldsh.de/PseudoAnon

Künftig identifizierbare Daten

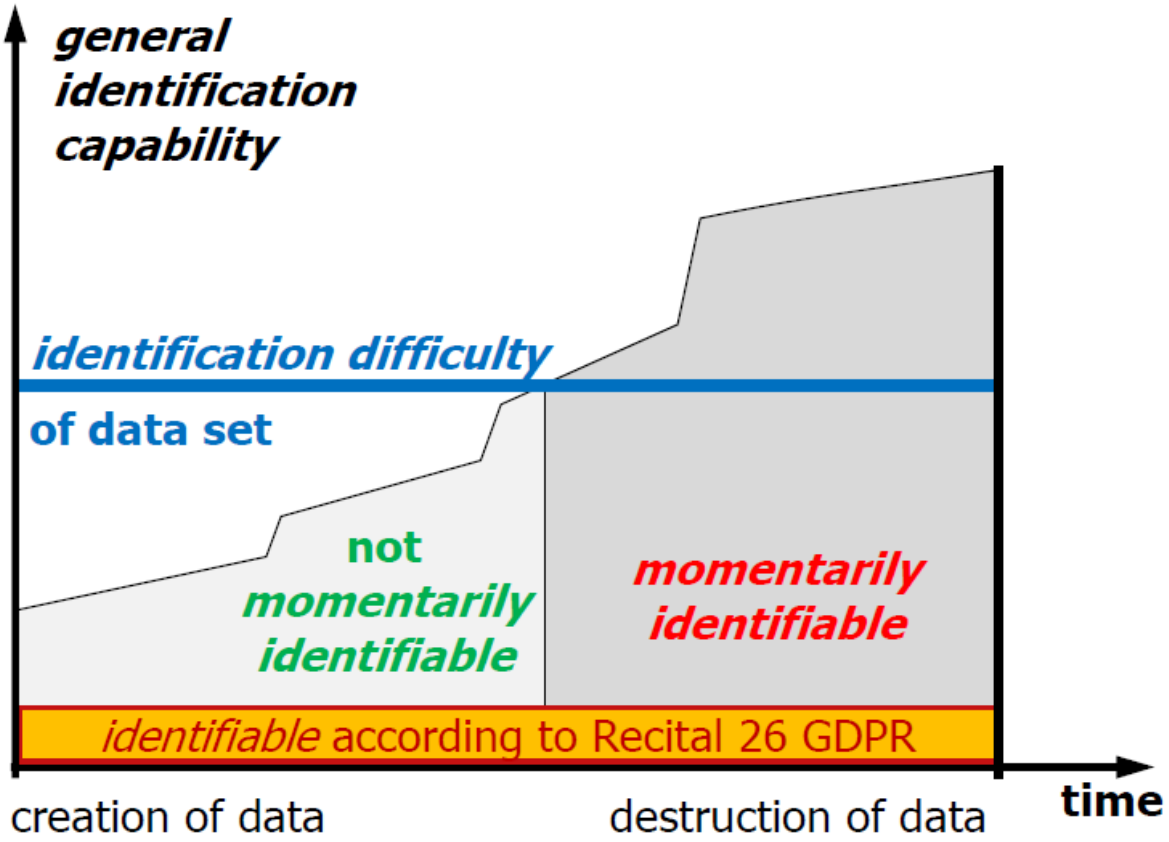
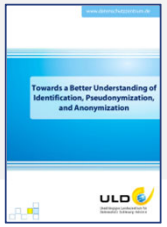


Figure 44: Eventually identifiable data.

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion



https://uldsh.de/PseudoAnon

Anonyme Daten

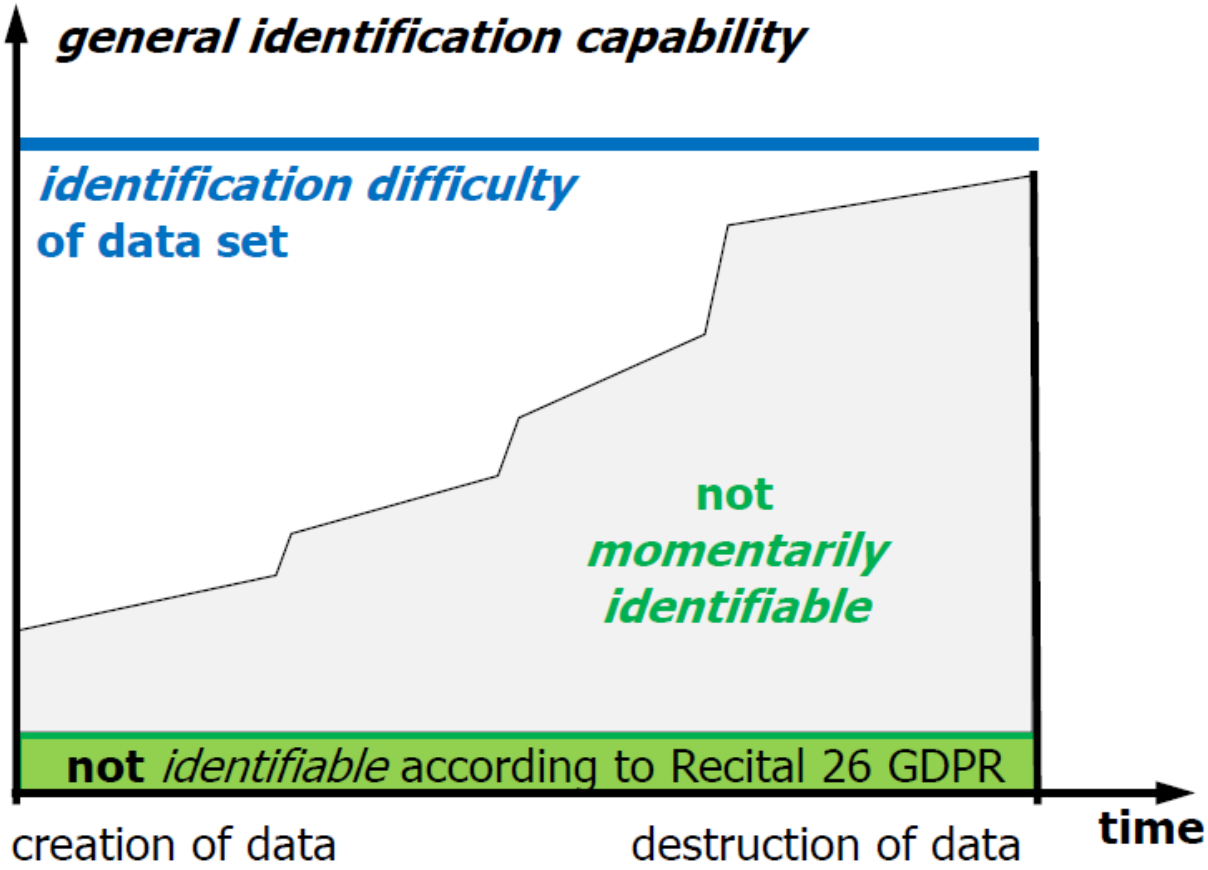


Figure 45: Anonymous data.

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Art. 11 DSGVO

Artikel 11 – Verarbeitung, für die eine Identifizierung der betroffenen Person nicht erforderlich ist

(1) Ist für die Zwecke, für die ein Verantwortlicher personenbezogene Daten verarbeitet, die Identifizierung der betroffenen Person durch den Verantwortlichen nicht oder nicht mehr erforderlich, so ist dieser nicht verpflichtet, zur bloßen Einhaltung dieser Verordnung zusätzliche Informationen aufzubewahren, einzuholen oder zu verarbeiten, um die betroffene Person zu identifizieren.

Mit „Einhaltung dieser Verordnung" insbes. gemeint: Betroffenenrechte wie Auskunft

Denn: Es ist nicht gewährleistet, dass der Verantwortliche dann den Personenbezug herstellen kann.

Bsp.: Cookies, IP-Adressen

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Art. 11 DSGVO

Artikel 11 – Verarbeitung, für die eine Identifizierung der betroffenen Person nicht erforderlich ist

(1) Ist für die Zwecke, für die ein Verantwortlicher personenbezogene Daten verarbeitet, die Identifizierung der betroffenen Person durch den Verantwortlichen nicht oder nicht mehr erforderlich, so ist dieser nicht verpflichtet, zur bloßen Einhaltung dieser Verordnung zusätzliche Informationen aufzubewahren, einzuholen oder zu verarbeiten, um die betroffene Person zu identifizieren.

(2) Kann der Verantwortliche in Fällen gemäß Absatz 1 des vorliegenden Artikels nachweisen, dass er nicht in der Lage ist, die betroffene Person zu identifizieren, so unterrichtet er die betroffene Person hierüber, sofern möglich. In diesen Fällen finden die Artikel 15 bis 20 keine Anwendung, […]

Wie unterrichten? Ginge ein Widerspruch?

Artt. 15-20: Betroffenenrechte außer Widerspruch (Art. 21) und automatisierte Entscheidung (Art. 22)

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Art. 11 DSGVO

Artikel 11 – Verarbeitung, für die eine Identifizierung der betroffenen Person nicht erforderlich ist

(1) Ist für die Zwecke, für die ein Verantwortlicher personenbezogene Daten verarbeitet, die Identifizierung der betroffenen Person durch den Verantwortlichen nicht oder nicht mehr erforderlich, so ist dieser nicht verpflichtet, zur bloßen Einhaltung dieser Verordnung zusätzliche Informationen aufzubewahren, einzuholen oder zu verarbeiten, um die betroffene Person zu identifizieren.

(2) Kann der Verantwortliche in Fällen gemäß Absatz 1 des vorliegenden Artikels nachweisen, dass er nicht in der Lage ist, die betroffene Person zu identifizieren, so unterrichtet er die betroffene Person hierüber, sofern möglich. In diesen Fällen finden die Artikel 15 bis 20 keine Anwendung, es sei denn, die betroffene Person stellt zur Ausübung ihrer in diesen Artikeln niedergelegten Rechte zusätzliche Informationen bereit, die ihre Identifizierung ermöglichen.

Welche Fälle denkbar?

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Datenminimierende Verfahren: Differential Privacy

- Hinzufügen von Unschärfen/Rauschen bei der Nutzung von Daten, so dass keine Identifizierbarkeit gegeben ist, so dass allgemeine, statistische Informationen zugänglich sind
- Besondere kryptographische Verfahren
- Parametrisierbar (Rauschen/Nutzen-Rate)
- Einsatz auch bei Google + Apple

Cynthia Dwork: Differential Privacy, in: 33rd International Colloquium on Automata, Languages and Programming, part II (ICALP 2006), Springer, Juli 2006, S. 1-12

> Starting with iOS 10, Apple is using Differential Privacy technology to help discover the usage patterns of a large number of users without compromising individual privacy. To obscure an individual's identity, Differential Privacy adds mathematical noise to a small sample of the individual's usage pattern. As more people share the same pattern, general patterns begin to emerge, which can inform and enhance the user experience. In iOS 10, this technology will help improve QuickType and emoji suggestions, Spotlight deep link suggestions and Lookup Hints in Notes.

https://www.quora.com/What-are-the-limitations-of-Apple%E2%80%99s-%E2%80%9Cdifferential-privacy%E2%80%9D-approach-to-collecting-user-data

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Beispiel Apple

Differential Privacy Team, Apple: Learning with Privacy at Scale, https://machinelearning.apple.com/docs/learning-with-privacy-at-scale/appledifferentialprivacysystem.pdf

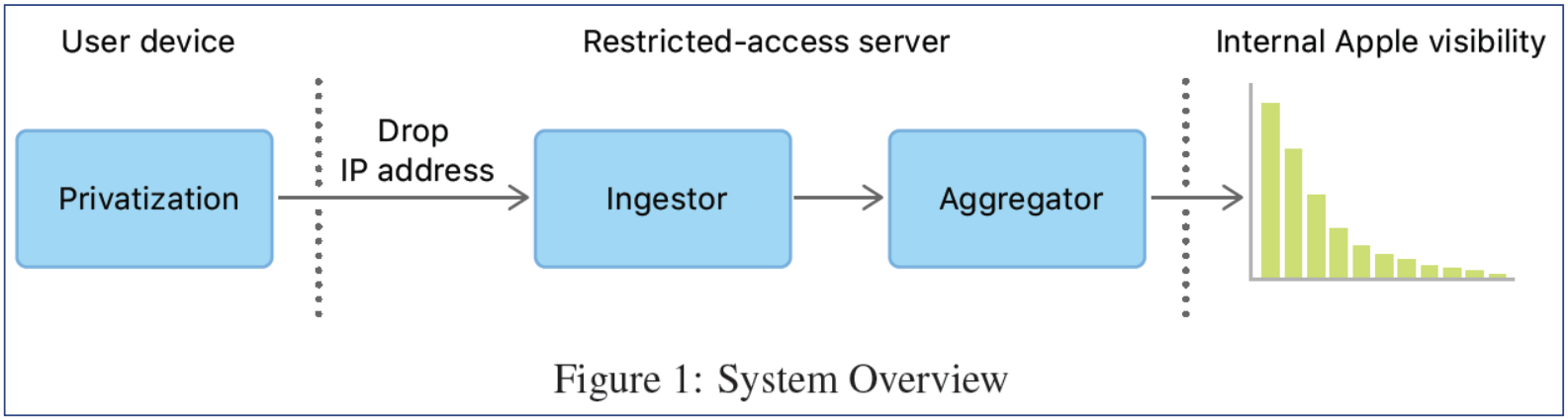


Figure 1: System Overview

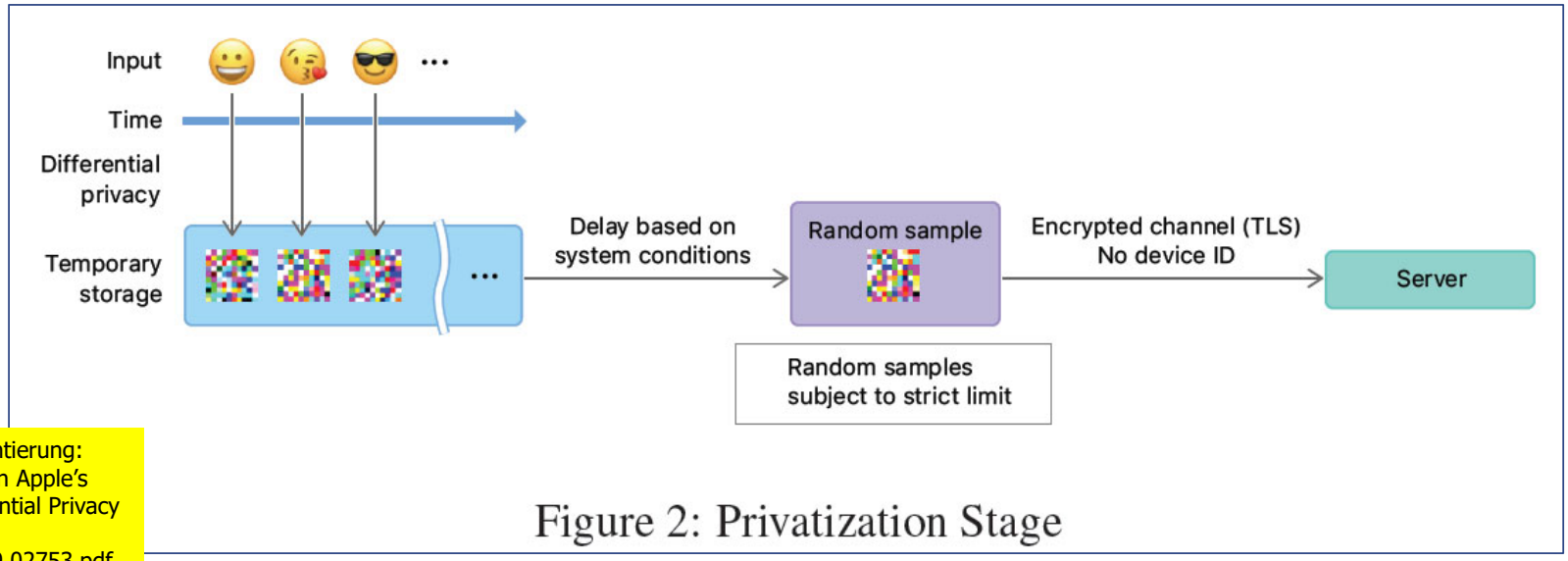


Figure 2: Privatization Stage

Kritik an Apples Implementierung: *Tang et al.,* Privacy Loss in Apple's Implementation of Differential Privacy on MacOS 10.12, 2017, https://arxiv.org/pdf/1709.02753.pdf

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Beispiel Apple

Differential Privacy Team, Apple: Learning with Privacy at Scale, https://machinelearning.apple.com/docs/learning-with-privacy-at-scale/appledifferentialprivacysystem.pdf

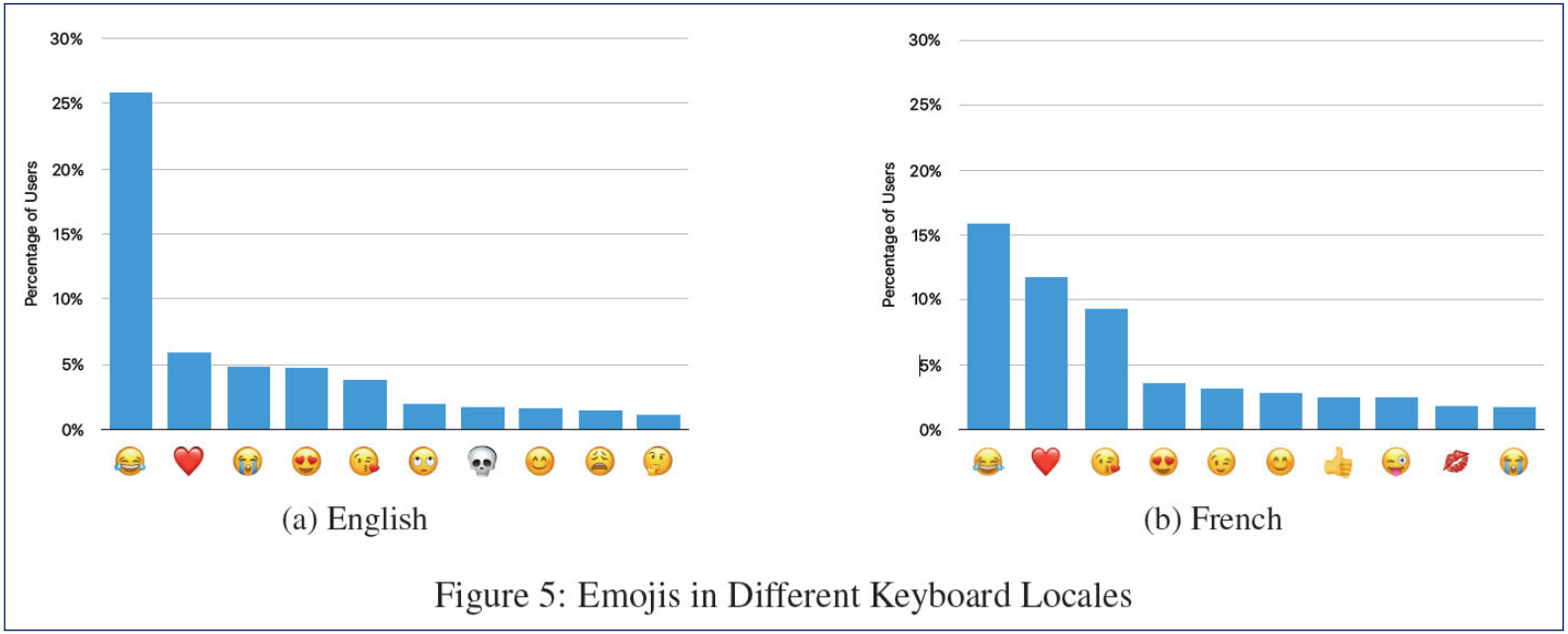


Figure 5: Emojis in Different Keyboard Locales

Kritik an Apples Implementierung: *Tang et al.,* Privacy Loss in Apple's Implementation of Differential Privacy on MacOS 10.12, 2017, https://arxiv.org/pdf/1709.02753.pdf

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Mechanismen, die für Statistiken interessant sein können

- Perturbative Ansätze
  - Hinzufügen von Rauschen, Vertauschen von Daten, Runden, Aggregation
  - Verfälschung der Daten, auch wenn Mittelwerte o.ä. bewahrt werden können
- Nicht-perturbative Ansätze
  - Generalisierung:
    PLZ (24103 → 241**), Geschlecht (M → *), Alter (24 → [20–29])
  - Unterdrückung (Ausreißer)
  - Keine Verfälschung der Daten

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Ausreißer vermeiden

- Auffällige Einzelwerte entfernen
  - Vollständig
  - Durch Angabe des möglichen Wertebereichs

	braun	blau	Σ
blond	2	10	12
braun	12	6	18
Σ	14	16	

	braun	blau	Σ
blond	*	*	12
braun	*	*	18
Σ	14	16	

	braun	blau	Σ
blond	[0,12]	[0,12]	12
braun	[0,14]	[0,16]	18
Σ	14	16	

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

k-Anonymität

Identifikator	Quasi-Identifikatoren			Sensibles Attribut
Name	**Alter**	**Geschlecht**	**PLZ**	**Krankheit**
Anna	21	Weiblich	76189	Grippe
Louis	35	Männlich	77021	Krebs
Holger	39	Männlich	63092	Haarausfall
Frederic	23	Männlich	63331	Muskelzerrung
Anika	25	Weiblich	76121	Grippe
Peter	31	Männlich	77462	Vergiftung
Tobias	38	Männlich	77109	Demenz
Charlotte	19	Weiblich	83133	Karies
Sarah	27	Weiblich	89777	Akne

Name beseitigen + Generalisierung durch Abstraktion (Alter, PLZ)

Identifikator	Quasi-Identifikatoren			Sensibles Attribut
Name	**Alter**	**Geschlecht**	**PLZ**	**Krankheit**
*	20 < Alter < 25	Weiblich	76*	Grippe
*	30 < Alter < 40	Männlich	77*	Krebs
*	20 < Alter < 40	Männlich	63*	Haarausfall
*	20 < Alter < 40	Männlich	63*	Muskelzerrung
*	20 < Alter < 25	Weiblich	76*	Grippe
*	30 < Alter < 40	Männlich	77*	Vergiftung
*	30 < Alter < 40	Männlich	77*	Demenz
*	18 < Alter < 28	Weiblich	8*	Karies
*	18 < Alter < 28	Weiblich	8*	Akne

https://de.wikipedia.org/wiki/K-Anonymität

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

k-Anonymität: Äquivalenzklassen mit mind. k Elementen

Andere Darstellung: Äquivalenzklassen mit mind. 2 Elementen: 2-Anonymität

	Identifikator	Quasi-Identifikatoren			Sensibles Attribut
Äquivalenzklasse	**Name**	**Alter**	**Geschlecht**	**PLZ**	**Krankheit**
A	*	20 < Alter < 25	Weiblich	76*	Grippe
	*	20 < Alter < 25	Weiblich	76*	Grippe

	Identifikator	Quasi-Identifikatoren			Sensibles Attribut
Äquivalenzklasse	**Name**	**Alter**	**Geschlecht**	**PLZ**	**Krankheit**
B	*	30 < Alter < 40	Männlich	77*	Krebs
	*	30 < Alter < 40	Männlich	77*	Vergiftung
	*	30 < Alter < 40	Männlich	77*	Demenz

	Identifikator	Quasi-Identifikatoren			Sensibles Attribut
Äquivalenzklasse	**Name**	**Alter**	**Geschlecht**	**PLZ**	**Krankheit**
C	*	20 < Alter < 40	Männlich	63*	Haarausfall
	*	20 < Alter < 40	Männlich	63*	Muskelzerrung

	Identifikator	Quasi-Identifikatoren			Sensibles Attribut
Äquivalenzklasse	**Name**	**Alter**	**Geschlecht**	**PLZ**	**Krankheit**
D	*	18 < Alter < 28	Weiblich	8*	Karies
	*	18 < Alter < 28	Weiblich	8*	Akne

https://de.wikipedia.org/wiki/K-Anonymität

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Ausnutzen von Homogenität

Homogeneity Attack [Bearbeiten | Quelltext bearbeiten]

Bei der Homogenitätsattacke wird ausgenutzt, dass unter Umständen alle k Datensätze einer Äquivalenzklasse identische sensible Attribute vorweisen. Weiß der Angreifer über die Existenz einer Person in einer Datenbank und kann er diese Person der korrekten Äquivalenzklasse zuweisen, erfährt er deren sensible Attribute.

Veranschaulichung[3] [Bearbeiten | Quelltext bearbeiten]

Alice ist eine sehr neugierige Nachbarin von Bob. Als Bob eines Tages mit dem Krankenwagen abgeholt wird, möchte Alice herausfinden, woran Bob erkrankt ist. Sie entdeckt die 4-anonyme Tabelle mit aktuellen Patientendaten, die vom Krankenhaus veröffentlicht wird. Sie weiß, dass Bob in der Tabelle enthalten sein muss und kennt sein Alter, Geschlecht sowie Postleitzahl. Dadurch schließt sie darauf, dass sein Datensatz in der Äquivalenzklasse C enthalten sein muss. Da alle Patienten dieser Äquivalenzklasse an derselben Krankheit leiden, erfährt Alice auch Bobs Krankheit.

	Identifikator	Quasi-Identifikatoren			Sensibles Attribut
Äquivalenzklasse	**Name**	**Alter**	**Geschlecht**	**PLZ**	**Krankheit**
B	*	25 < Alter < 30	Weiblich	13*	...
					Herzerkrankung
C	*	40 < Alter < 50	Männlich	13*	Krebs
					Krebs
					Krebs
					Krebs
D	*	20 < Alter < 35	Weiblich	12*	Grippe
					...

Lösung: *l*-Diversity

https://de.wikipedia.org/wiki/K-Anonymität

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Beispiel für l-Diversity

IBM Research – Zurich

IBM

l-Diversity

Jeder q*-Block enthält mindestens l "wohl-vertretene" Werte des sensitiven Attributes s.

Alter	Geschlecht	Krankheit
[26 – 27]	M	Grippe
[26 – 27]	M	Grippe
[23 – 25]	*	Erkältung
[23 – 25]	*	Diabetes
22	M	Grippe
22	M	Krebs

$k = 2$

Alter	Geschlecht	Krankheit
[25 – 27]	M	Grippe
[25 – 27]	M	Grippe
[25 – 27]	M	Erkältung
[22 – 23]	*	Diabetes
[22 – 23]	*	Grippe
[22 – 23]	*	Krebs

k = 3, $E \geq log(1.9)$

- Datenveröffentlicher benötigt weniger Information als der Angreifer
- berücksichtigt *instance-level knowledge* ("mein Nachbar hat keine Diabetes")

13 / 19 Dr. Günter Karjoth | Technische Datenschutzlösungen bei der Analyse großer Datenmengen | 26. August 2013 © 2013 IBM Corporation

Folien von Günter Karjoth, IBM, 2013, ULD-Sommerakademie 2013, Kiel

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Ausnutzen von Zusatzwissen

Background Knowledge Attack [Bearbeiten | Quelltext bearbeiten]

Durch den Einsatz von Zusatzwissen kann es möglich sein, Personen trotz k-Anonymität eindeutig zuzuordnen. Weiß der Angreifer über die Existenz einer Person in einer Datenbank und kann er diese Person der korrekten Äquivalenzklasse zuweisen, kann er gegebenenfalls durch das Zusatzwissen manche sensible Attribute für die Person ausschließen.

Veranschaulichung [3] [Bearbeiten | Quelltext bearbeiten]

Alice hat eine Brieffreundin namens Yui, die in ein Krankenhaus eingeliefert und deren Patientendaten in einer 4-anonymen Tabelle enthalten sind, die vom Krankenhaus regelmäßig veröffentlicht wird. Alice weiß, dass Yui eine 21 Jahre alte Japanerin ist, die momentan unter der PLZ 12345 gemeldet ist. Ausgehend von diesen Informationen kann Alice darauf schließen, dass Yuis Datensatz in der Äquivalenzklasse B enthalten sein muss. Ohne zusätzliche Informationen kann sich Alice nicht sicher sein, ob Yui an einer Viruserkrankung oder an einer Herzerkrankung leidet. Jedoch ist hinlänglich bekannt, dass Japaner sehr selten an Herzerkrankungen leiden. Dadurch kann Alice darauf schließen, dass bei Yui wohl eine Viruserkrankung vorliegt.

	Identifikator	Quasi-Identifikatoren			Sensibles Attribut
Äquivalenzklasse	**Name**	**Alter**	**Geschlecht**	**PLZ**	**Krankheit**
A	*	30 < Alter < 35	Männlich	14*	...
					Grippe
B	*	20 < Alter < 30	Weiblich	12*	Herzerkrankung
					Viruserkrankung
					Viruserkrankung
					Herzerkrankung
C	*	30 < Alter < 35	Weiblich	12*	Krebs
					...

https://de.wikipedia.org/wiki/K-Anonymität

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Beispiel für t-Plausibility

IBM Research – Zurich

IBM

Text De-identification

Ein Einwohner von Kiel kaufte Marihuana gegen lumbale Schmerzen, verursacht durch Leberkrebs.

Ein Einwohner von ~~Kiel~~ kaufte ~~Marihuana~~ gegen ~~lumbale Schmerzen~~, verursacht durch ~~Leberkrebs~~.

t-Plausibility verallgemeinert sensitive Terme zu semantisch ähnlichen Termen, z. B. “Tuberkulose” $\rightarrow$ “Infektion”.

Ist eine Wortontologie und ein Grenzwert t gegeben, kann der gesäuberte Text mindestens $t-1$ anderen Texten zugeordnet werden.

Ein Einwohner von Landeshauptstadt kaufte Droge gegen Schmerzen, verursacht durch Karzinom.

18 / 19 Dr. Günter Karjoth | Technische Datenschutzlösungen bei der Analyse großer Datenmengen | 26. August 2013 © 2013 IBM Coorporation

Folien von Günter Karjoth, IBM, 2013, ULD-Sommerakademie 2013, Kiel

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Übersicht in WP 216 der Art. 29-Datenschutzgruppe

	Is Singling out still a risk?	Is Linkability still a risk?	Is Inference still a risk?
Pseudonymisation	Yes	Yes	Yes
Noise addition	Yes	May not	May not
Substitution	Yes	Yes	May not
Aggregation or K-anonymity	No	Yes	Yes
L-diversity	No	Yes	May not
Differential privacy	May not	May not	May not
Hashing/Tokenization	Yes	Yes	May not

Table 6. Strengths and Weaknesses of the Techniques Considered

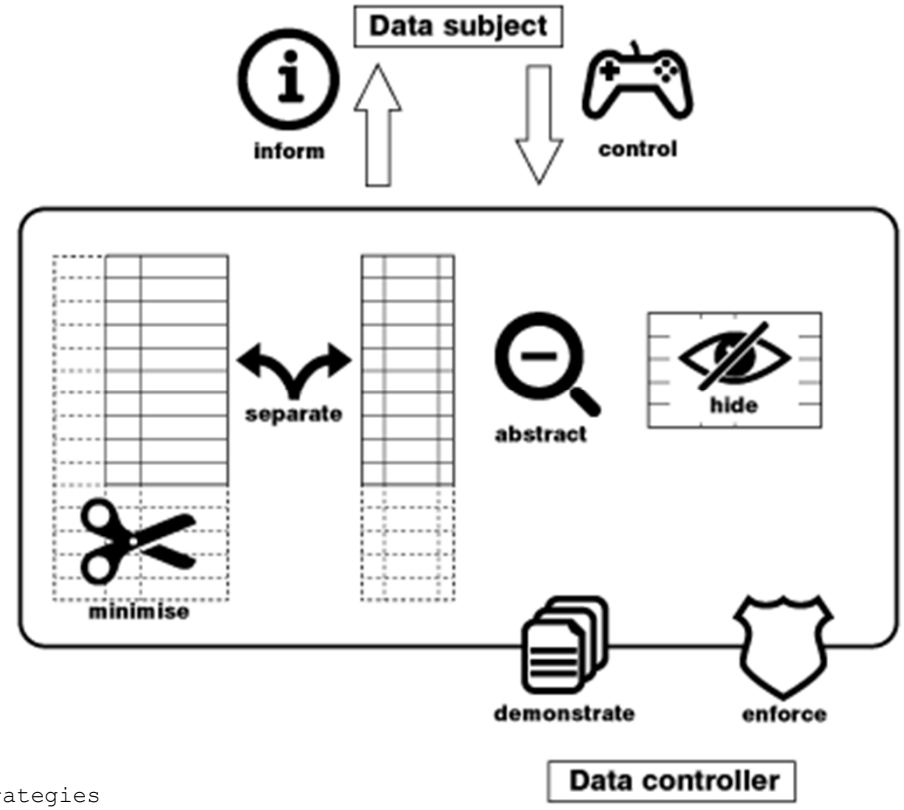
Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Vortrag bisher: Datenbankzentrierte Sicht

Z.B. in den
Privacy Design Strategies

- Minimise
- Separate
- Abstract
- Hide



https://wiki.privacy.cs.ru.nl/Privacy_design_strategies

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Anders: Nutzungszentrierte Sicht

- Agieren unter (wechselnden) Pseudonymen
- Änderungen der Bedingungen
  - Wer ist Verantwortlicher?
  - Wie werden die Pseudonyme erzeugt?
  - Wer kümmert sich um die Sicherheit?
  - Wie kann der Nutzende realistisch das Risiko einschätzen?
  - Auskunft und andere Betroffenenrechte unter Pseudonym?

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Best Practice „Datenminimierung": Authentifikation ohne Identifikation

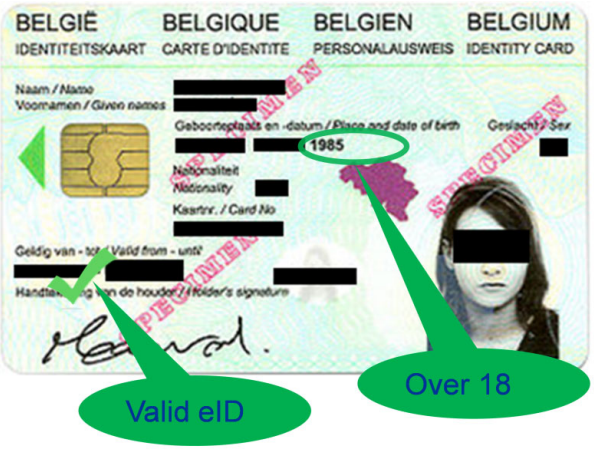
Vollständige Daten:

Vorab Prüfen der Anforderungen:
Welche Daten sind wirklich erforderlich?

Information auch ohne pb Daten

Oft sind nicht alle Daten erforderlich

Minimale Daten:



Attribute-based Credentials

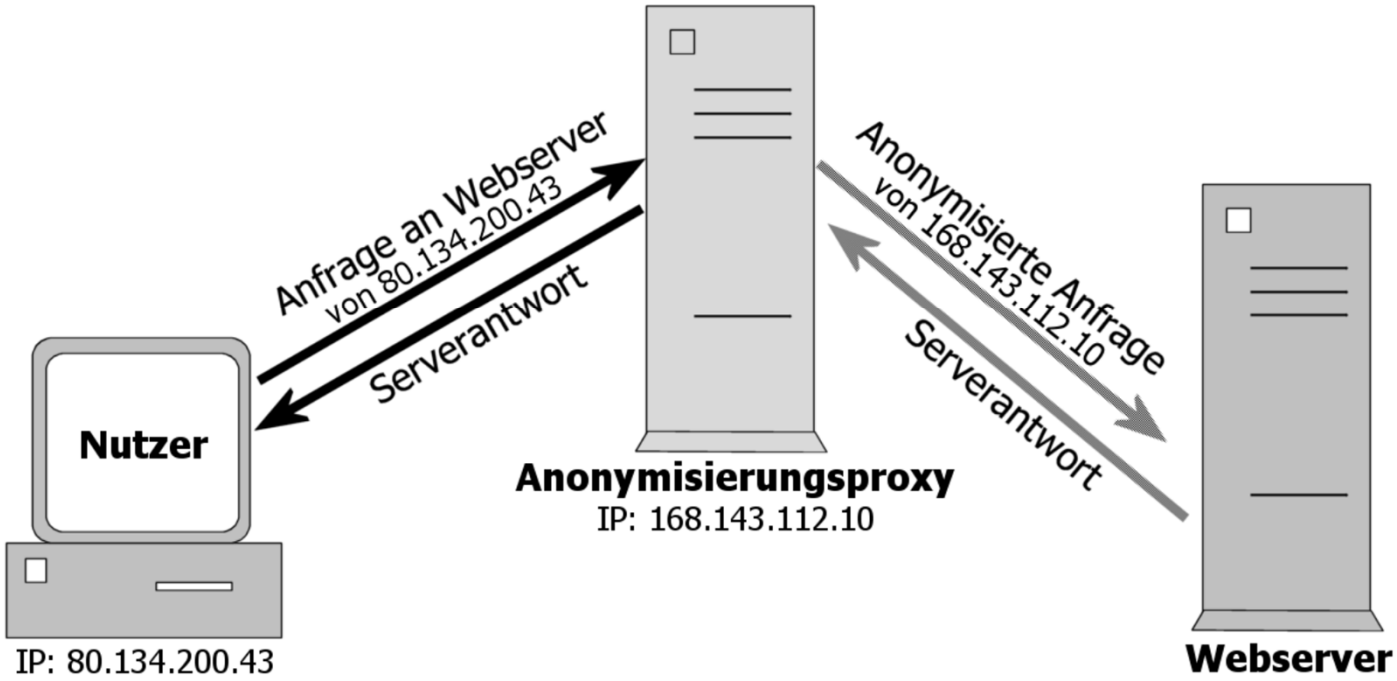
Symbolbild (nicht belgischer Ausweis)

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Der einfache „Anonymisierungs"-Proxy

... weiß alles!

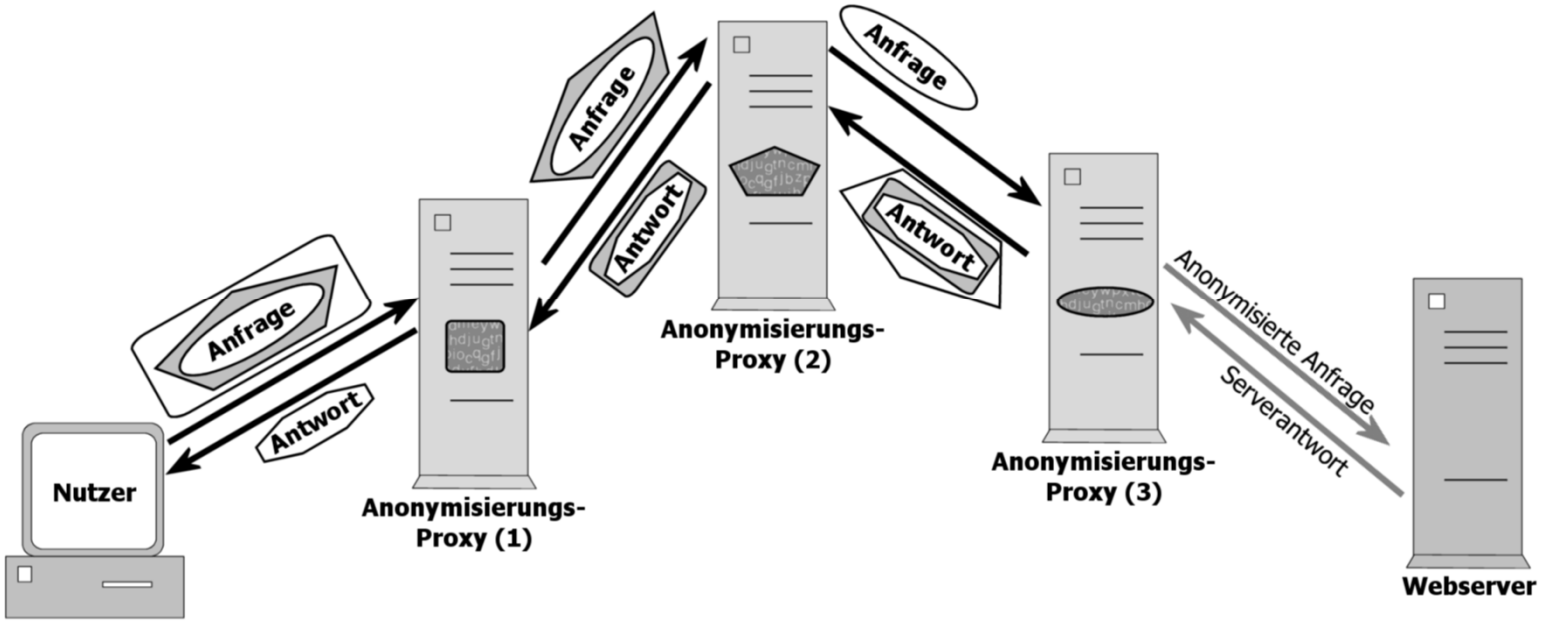


Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Hintereinanderschaltung von „Anonymisierungs"-Proxies

… mit Verschlüsselung bewirkt Aufteilung der Information: keiner weiß alles!

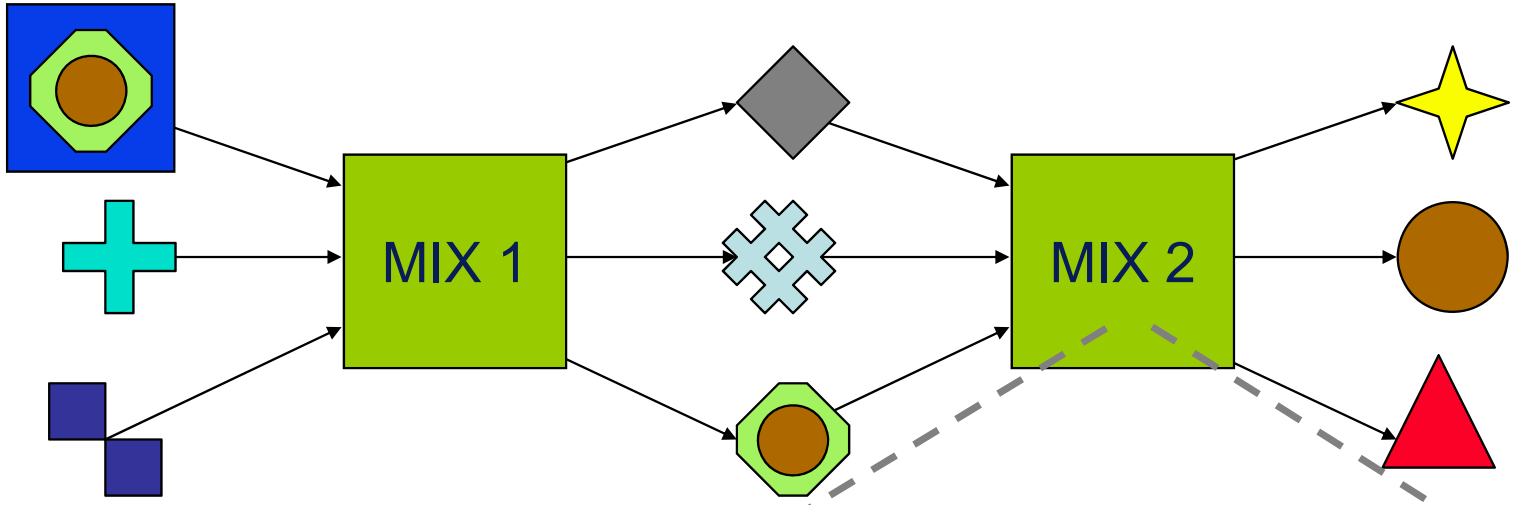


Überblick

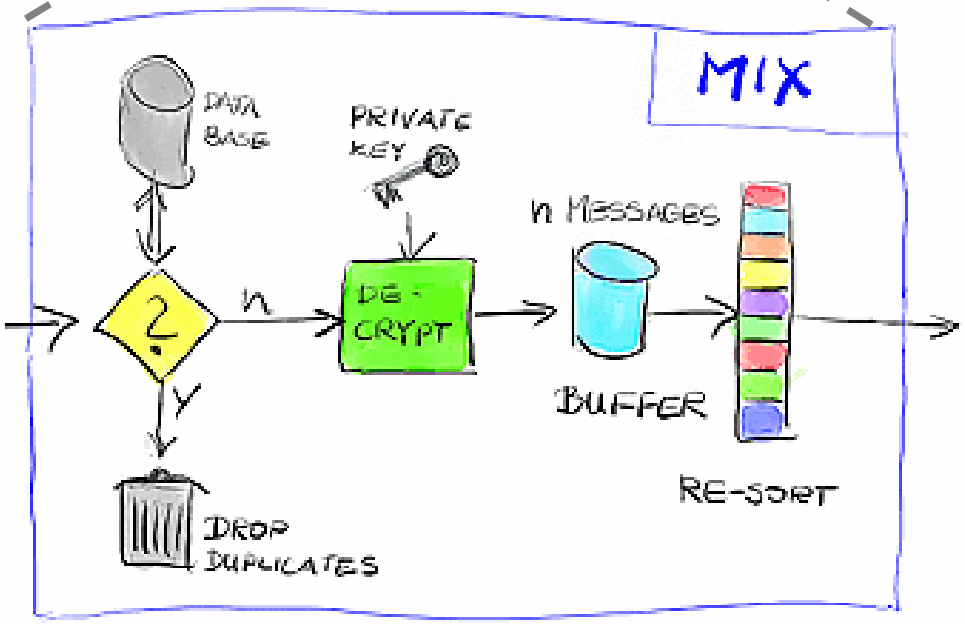
1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung

2. Exkurs: Art. 11 DSGVO

3. Beispiele und Diskussion



„Anonymität" mit Mix-Rechnern (David Chaum 1981)



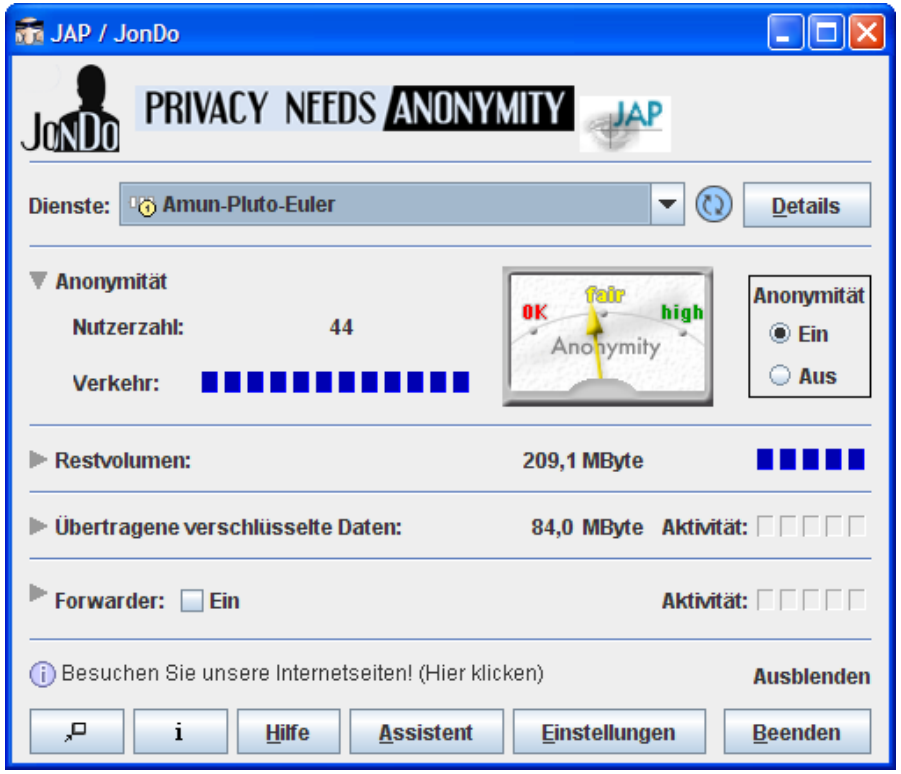
Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

„Anonymizer"

Anwendungsfall:
Infrastruktur? Betrieb durch wen?

- Z.B. TOR oder AN.ON auf Ebene der IP-Adressen
- Eine technische Anonymitätsmethode: „Gleichmacherei", denn viele Nutzer teilen sich dieselben Proxies (und Ketten von Proxies („Mix-Kaskaden")))



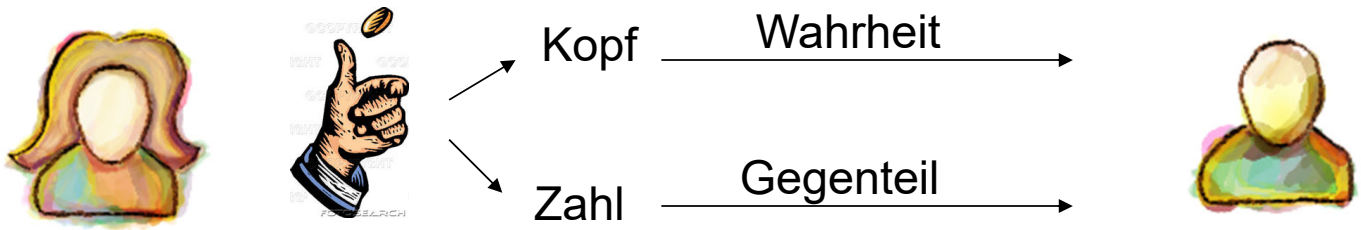
Überblick

Anwendungsfall:
Lagebild IT-Sicherheit?

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

„Randomized Response"

- Ergebnisse werden mit einer bekannten Wahrscheinlichkeit ($\neq$ 0,5) verfälscht
- Herausrechnen dieses bekannten Fehlers liefert brauchbaren statistischen Wert ohne individuelle Zurechenbarkeit



Für Datenbanken: W. Du: Using Randomized Response Techniques for Privacy-Preserving Data Mining, 2003

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Bloom-Filter: Feststellen von Ähnlichkeiten

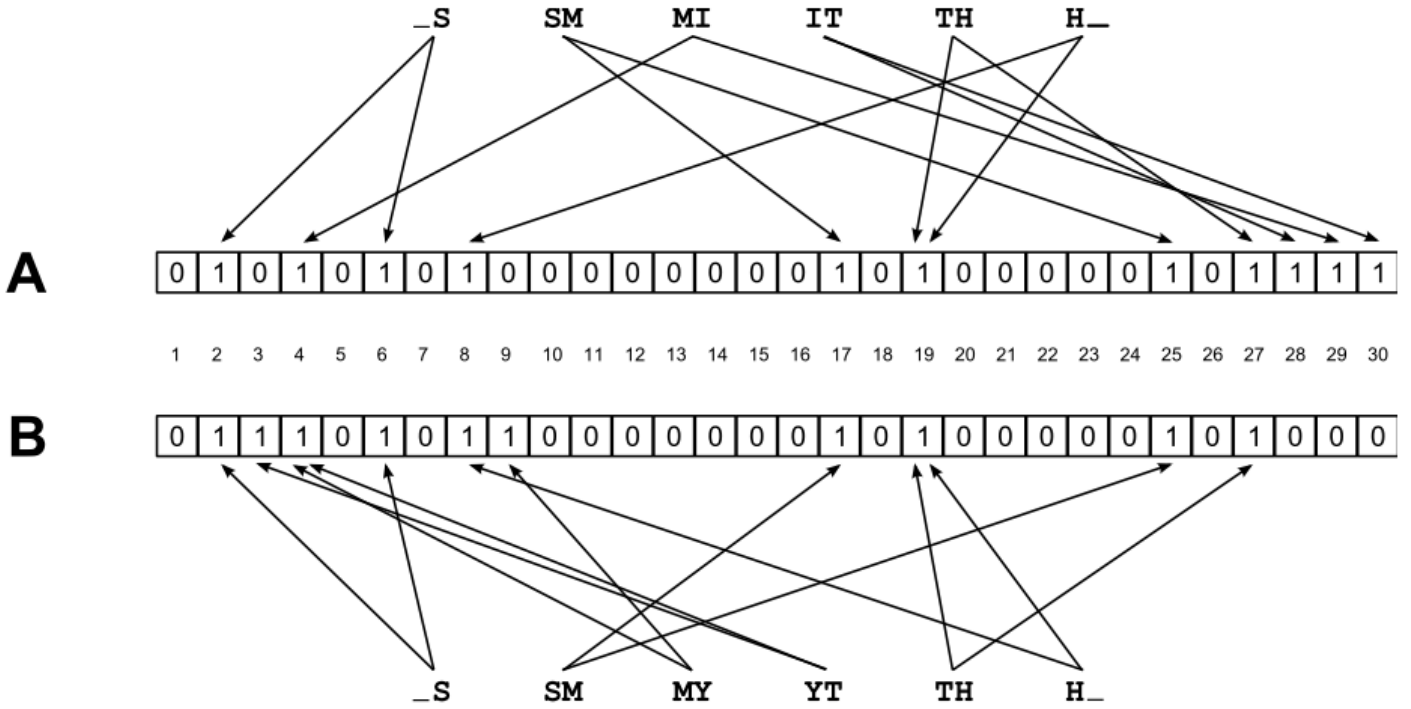


Abbildung 3: Beispiel für die Verschlüsselung der Namen „SMITH“ und „SMYTH“ in zwei binäre Vektoren

Quelle: R. Schnell: Getting Big Data but avoiding Big Brother, German Record Linkage Center (2013)

Überblick

1. Methoden zur Datenminimierung: Pseudonymisierung und Anonymisierung
2. Exkurs: Art. 11 DSGVO
3. Beispiele und Diskussion

Bloom-Filter: In Menge nicht enthalten?

Ähnlich:
fiu.net / ma³tch: Anonymous Autonomous Analysis