Privacy-Enhancing Technologies

Module 2: Measuring Privacy – Metrics

Dr. Patricia Arias-Cabarcos, Thorsten Strufe

10.05.2021 – KIT and TU Dresden – the pandemic continues..

Disclaimer: This lecture was prepared in cooperation with Dr. Javier Parra-Arnau

Competence Center for Applied Security Technology

Outline

- The importance of privacy metrics
- Privacy domains
- Aspects of privacy metrics
- Classification of privacy metrics[1]

[1] Isabel Wagner and David Eckhoff, "Technical Privacy Metrics: A Systematic Survey", ACM Comput. Surv. 51, 3, Article 57, June 2018.

Outline

- The importance of privacy metrics
- Privacy domains
- Aspects of privacy metrics
- Classification of privacy metrics

Privacy and human rights

- Data privacy is the adaptation to the Information Society of the fundamental right to privacy and private life.
- It is included by the United Nations in the ***Universal Declaration of Human Rights*** (1948), in Article 12:

> No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation. Everyone has the right to the protection of the law against such interference or attacks.

The need for privacy and utility metrics

- Data privacy technologies are about technically enforcing that right in the information society
  - Anonymous-communication networks, anonymous credentials, multiparty computation and oblivious transfer protocols are some examples of general-purpose PETs

- The use of these technologies is not widespread yet
  - are seen as an expensive innovation with unclear benefits
  - frequently come at the expense of system functionality and data distortion (a.k.a. utility)

Privacy: What is promised and really achieved?

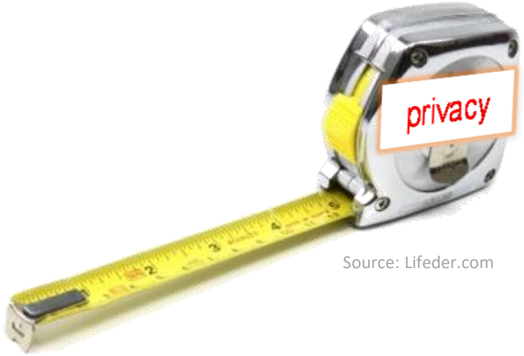
- PETS hide or distort PII
  - (Hopefully) impact on privacy! *(which?)*
  - Impact on utility (in some cases) -> Privacy – utility tradeoff *(s.t.: cost)*

Source: resp. Web pages

urce: census.gov

Privacy and Utility `Metrics‘?

- Quantifiable measures of privacy and utility enable us to
  - assess, compare,
  - improve and optimize privacy-enhancing mechanisms



Source: Lifeder.com

- What is a ‚metric‘?
  - A measure of the extent of inequality
  - Math requires: non-negativity, identity of indiscernibles, symmetry, triangle ineq.
  - Privacy metrics often just measure, and not metrics in the mathematical sense!

- Spectrum of expression
  - Pessimistic / worst-case metrics *(the conventional security view)*
  - Average case
  - Optimistic / best-case metrics

Outline

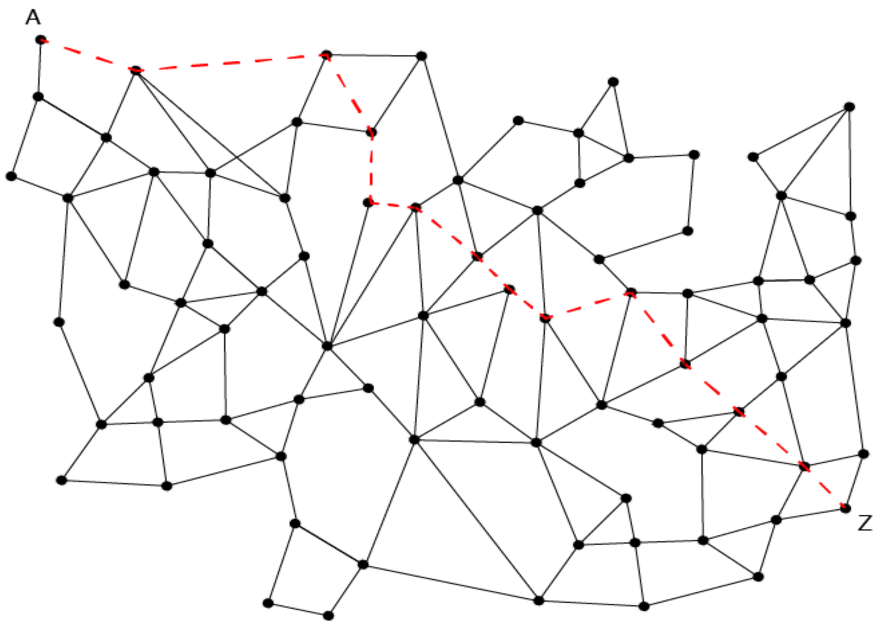
- The importance of privacy metrics
- Privacy domains
- Aspects of privacy metrics
- Classification of privacy metrics

Privacy domains

- Privacy domains are areas where PETs can be applied
- Common privacy domains:
  - Anonymous-communication systems
  - Databases
  - Personalized information systems
  - Location-based services
  - Interaction graph privacy
  - Genome privacy

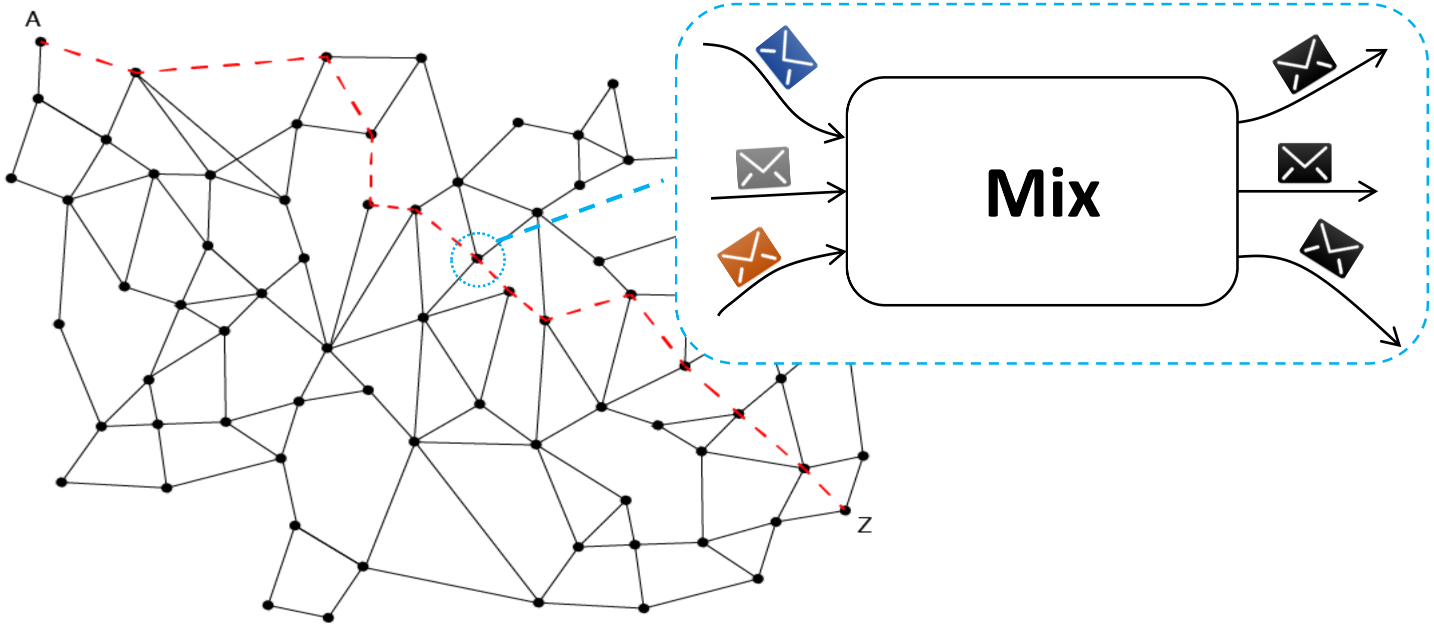
Privacy domains

- Anonymous-communication systems
  - The goal is to prevent an adversary from linking an outgoing message to its corresponding input message



Privacy domains

- Anonymous-communication systems[2]
  - The goal is to prevent an adversary from linking an outgoing message to its corresponding input message



[2] D. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms", Commun. ACM, vol. 24, no. 2, pp. 84-88, 1981.

Privacy domains

- Database anonymization
  - E.g., microdata

Identifiers	Key Attributes: Height	Key Attributes: Weight	Confidential Attributes: High Cholesterol
John Smith	5’4”	158	Y
Tang Lee	5’3”	162	Y
Luis Melo	5’6”	161	Y
Anna Frank	5’8”	157	N

Microdata

Privacy domains

- Database anonymization[3]
  - E.g., microdata

Key Attributes: Height	Key Attributes: Weight	Confidential Attributes: High Cholesterol
5’4”	158	Y
5’3”	162	Y
5’6”	161	Y
5’8”	157	N

De-identified microdata

[3] L. Sweeney, Uniqueness of Simple Demographics in the U.S. Population, LIDAPWP4. Carnegie Mellon University, Laboratory for International Data Privacy, Pittsburgh, PA, 2000.

Privacy domains

- Database anonymization
  - E.g., microdata

Key Attributes		Confidential Attributes
Height	Weight	High Cholesterol
5’4”	158	Y
5’3”	162	Y
5’6”	161	Y
5’8”	157	N

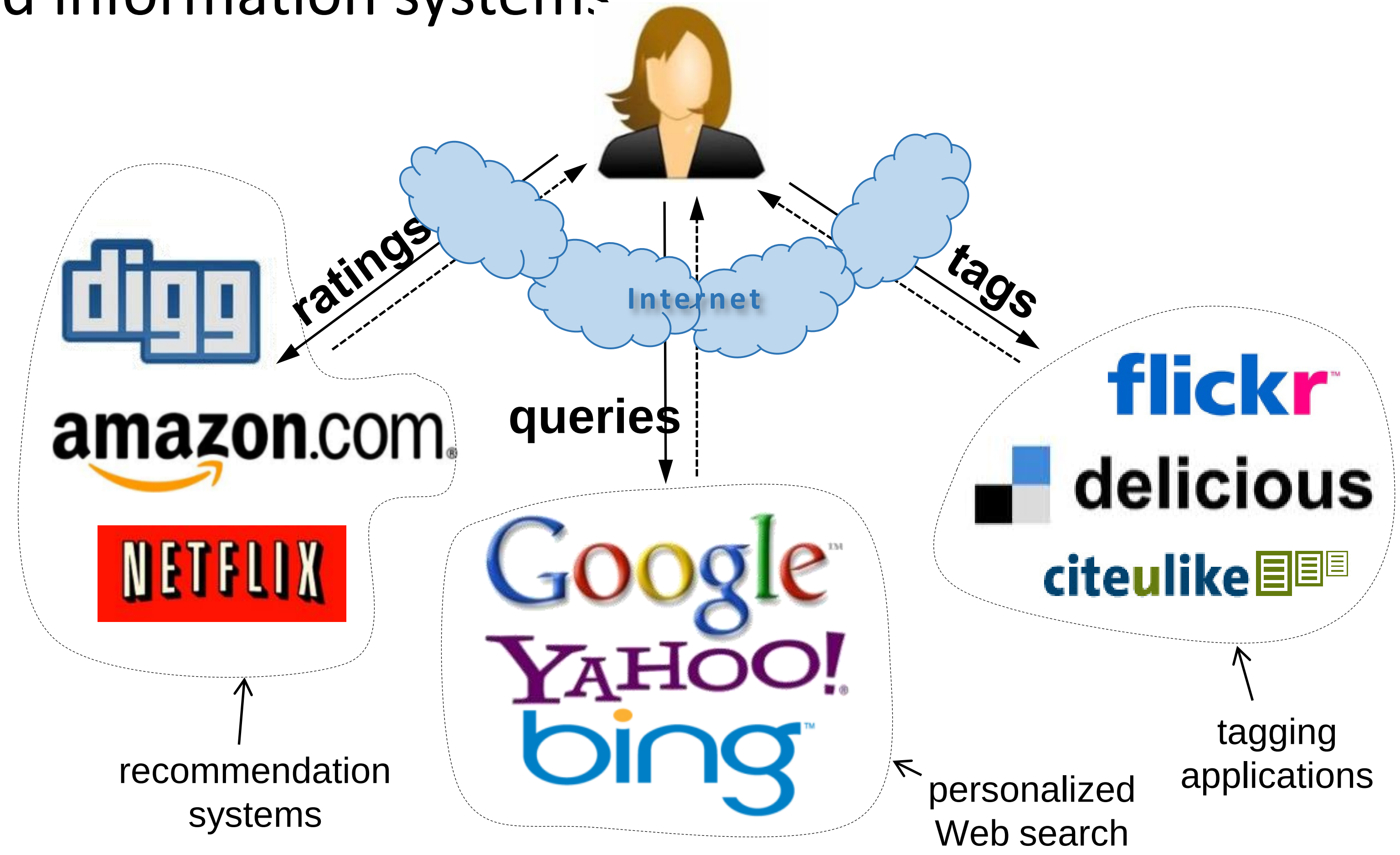
De-identified microdata

	Perturbed Key Attributes		Confidential Attributes
	Height	Weight	High Cholesterol
Aggregated Records	5’5”	160	Y
	5’5”	160	Y
	5’5”	160	Y
	6’0”	155	N

Protected microdata

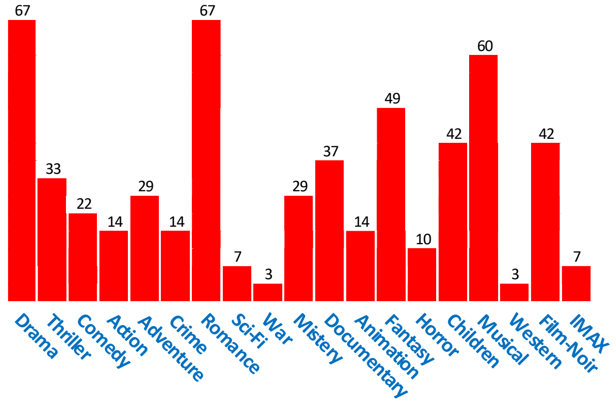
Privacy domains

- Personalized information systems



Privacy domains

- Personalized information systems



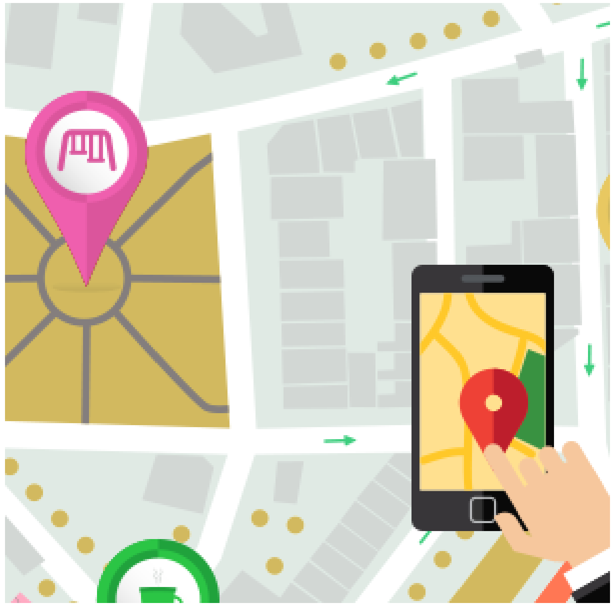
Your categories

Below you can edit the interests and inferred demographics that Google has associated with your cookie:

Category	
Beauty & Fitness – Fitness – Yoga & Pilates	Remove
Hobbies & Leisure – Water Activities – Surf & Swim	Remove
Home & Garden – Home Improvement – House Painting & Finishing	Remove
News – Health News	Remove
People & Society – Family & Relationships – Family – Baby Names	Remove
People & Society – Family & Relationships – Family – Parenting – Baby Care	Remove
Sports – Individual Sports - Cycling	Remove
Sports – Individual Sports – Gymnastics	Remove
Demographics – Age – 25-34	Remove
Demographics – Gender – Female	Remove

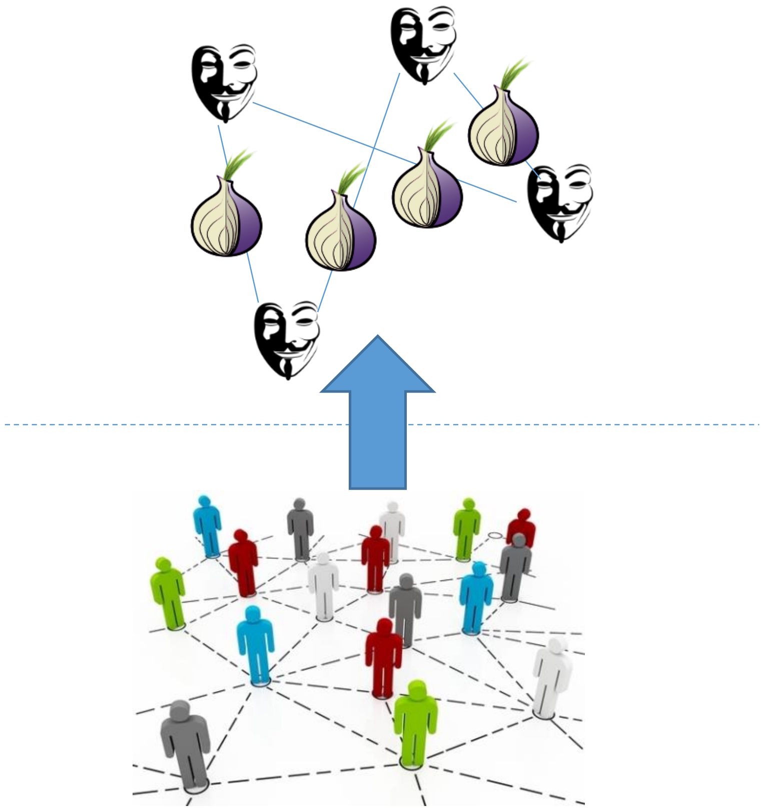
Privacy domains

- Location-based services

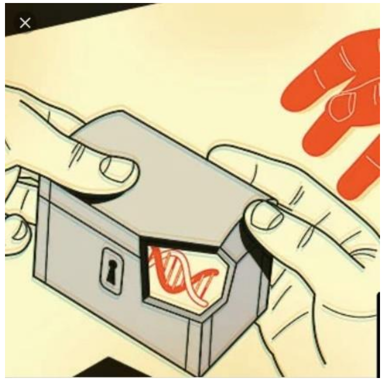


Source: Geospatial World

Interaction Graphs



Genomic Privacy



Source: Scientific American

Outline

- The importance of privacy metrics
- Privacy domains
- Aspects of privacy metrics
- Classification of privacy metrics

Aspects of privacy metrics

- Although there is a wide variety of privacy metrics, they all share some common features:
  - Adversary goals
  - Adversary capabilities
  - Data sources
  - Input of metric
  - Output measures

Aspects of privacy metrics

- Although there is a wide variety of privacy metrics, they all share some common features:
  - Adversary goals
  - Adversary capabilities
  - Data sources
  - Input of metric
  - Output measures

- Metrics are defined for a specific adversary
- Goals include
  - **identifying** a user
  - user **properties** (interests, preferences, location, etc.)
- Metrics need to be chosen according to that goal

Aspects of privacy metrics

- Although there is a wide variety of privacy metrics, they all share some common features:
  - Adversary goals
  - Adversary capabilities
  - Data sources
  - Input of metric
  - Output measures

- Attacker's success depends on its capabilities
- Metrics can only be employed to compare two PETs if they rely on the **same adversary** capabilities
- Taxonomy
  - Local-global
  - Passive-active
  - Internal-External
  - Prior knowledge
  - Resources

Aspects of privacy metrics

- Although there is a wide variety of privacy metrics, they all share some common features:
  - Adversary goals
  - Adversary capabilities
  - Data sources
  - Input of metric
  - Output measures

- Which **data** is to be protected? How does the adversary **gain access** to them?
  - Published data
  - Observable data
  - Repurposed data
  - All other data

Aspects of privacy metrics

Karlsruhe Institute of Technology

- Although there is a wide variety of privacy metrics, they all share some common features:
  - Adversary goals
  - Adversary capabilities
  - Data sources
  - Input of metric
  - Output measures

- What are **assumptions** about the adversary, protection requirements?
  - Prior knowledge of the adversary
  - Adversary’s resources
  - Adversary’s estimate
  - Ground truth/true outcome
  - Parameters

Aspects of privacy metrics

- Although there is a wide variety of privacy metrics, they all share some common features:
  - Adversary goals
  - Adversary capabilities
  - Data sources
  - Input of metric
  - Output measures

- Which **property** is the metric **measuring**?
  - Uncertainty
  - Information gain/loss
  - Data similarity/dissimilarity
  - Indistinguishable
  - Error-based metrics
  - Time-based metrics

Outline

- The importance of privacy metrics
- Privacy domains
- Aspects of privacy metrics
- Privacy metrics by class (output)
  - Uncertainty-based
  - Information-gain/loss
  - Estimation error
  - Time-based metrics
  - Data-similarity
  - Indistinguishability-based

1) Uncertainty-based privacy metrics

Karlsruhe Institute of Technology

- Assume that low uncertainty in the adversary's estimate correlates with low privacy
- The majority of these privacy metrics rely upon information-theoretic quantities (e.g., entropy)
- Origin in anonymous-communication systems
- Examples
  - Anonymity set size[4]
  - Shannon's entropy[5]
  - Normalized Shannon's entropy[5]
  - Inherent privacy[6]
  - Rényi entropy[7]

[4] D. Chaum, "The dining cryptographers problem: unconditional sender and recipient untraceability. J. Cryptol. vol. 1, no. 1, pp. 65–75, March 1988.

[5] C. Diaz, S. Seys, J. Claessens, B. Preneel, "Towards measuring anonymity", Privacy Enhancing Technologies (PET'02). LNCS 2482, pp. 54-68, 2002.

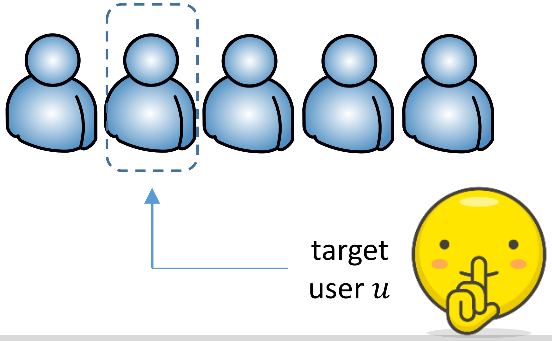
[6] C. Andersson, R. Lundin, "On the fundamentals of anonymity metrics", In Proc. IFIP Int. Summer School on the Future of Identity in the Information Society. Karlstad, Sweden, pp. 325–341, 2008.

[7] S. Clauß, S. Schiffner, "Structuring anonymity metrics", In Proc. ACM Workshop on Digital Identity Management (DIM'06), pp. 55-62, 2006.

Anonymity set (size)

- Given a target member u, it is defined as the (size of the) set of members the adversary cannot distinguish from u
- The larger the anonymity set, the more anonymity a member is enjoying
- Widely used metric, not only in ACSs
- Simplicity, tractability are positive properties of this metric
- However: it only depends on the number of members in the system

the metric assumes the attacker cannot distinguish any of them



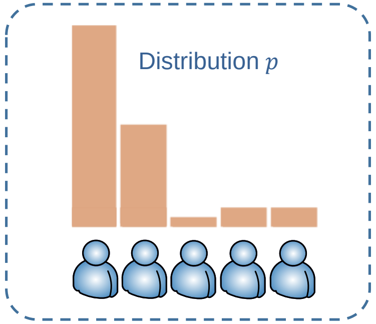
Figures sources: blog.yellowoctopus.com.au, iconscout.com

Shannon's entropy

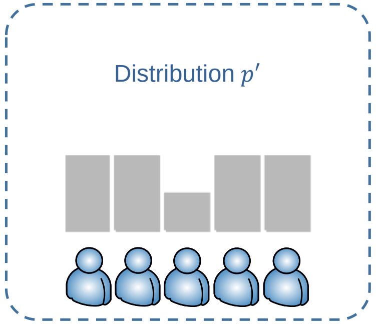
- From information-theory, it measures the uncertainty associated with predicting the outcome of a random variable (r.v.)
- As a privacy metric
  - An adversary aims to learn which member of an anonymity set (or: group of suspects) performed a certain action (e.g., sent a message)
  - Let $\{x_1, x_2, \ldots, x_n\}$ be the anonymity set and $p(x_i)$ the probability estimated by the adversary of x_i being the user who performed such action
  - Attacker's aim: predict the outcome of an r.v. X distributed according to p (identify victim)
  - Defined as

$$H(p) = -\sum_i p(x_i) \log p(x_i)$$

System #1



$$H(p) < H(p')$$

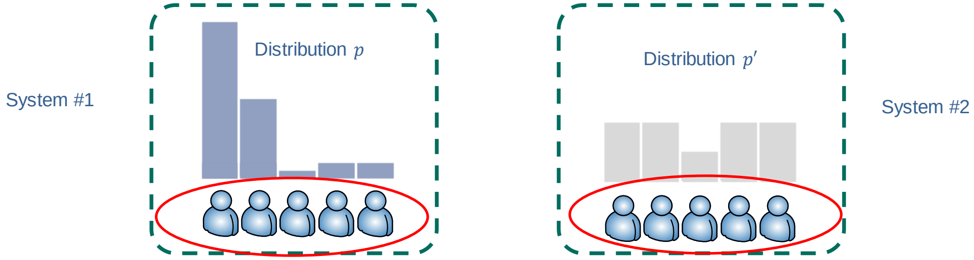


System #2

Normalized Shannon's entropy

Karlsruhe Institute of Technology

- SE is useful if the **size** of the anonymity sets of both systems coincide
- Normalized Shannon's entropy allows comparison also otherwise



- What if I tell you that the Shannon's entropy of a system
  - is 4 bits?
  - is 8 bits?

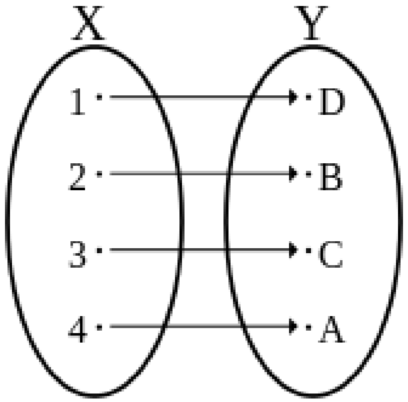
$$H(p) = -\sum_i p(x_i) \log p(x_i) \longrightarrow \frac{H(p)}{\log n}$$

Prove it!

- NSE yields output in (0,1)

Inherent privacy and bijections

- Based on the same concept. Privacy is defined as $2^{H(p)}$
- Is it really different from normalized or Shannon's entropy?
- Using a metric or a bijection of this metric is essentially the same, both in terms of comparison and optimization



$$2^{H(p)},\ H(p)/\log n,\ H(p)$$

Figure source: Wikipedia

Rényi's entropy

- Rényi's entropy is a family of functions widely used in information theory as a measure of uncertainty
- More specifically, Rényi's entropy of order α is defined as

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_{i=1}^{n} p(x_i)^\alpha,$$

support set

$$\begin{cases} H_0(X) = \log \boxed{|\{x \in \mathcal{X} : p(x) > 0\}|} & \text{Hartley} \\ H_1(X) = -\sum_i p(x_i) \log p(x_i) & \text{Shannon} \\ H_\infty(X) = \min_i -\log p(x_i) = -\log \max_i p(x_i) & \text{min-entropy} \end{cases}$$

Interpretation of several entropy measures

- Again, the attacker's aim is to predict the outcome of an r.v. X distributed according to p

$$H_\infty(X) = \min_i -\log p(x_i) = -\log \max_i p(x_i)$$ — worst-case

$$\text{|}\wedge$$

$$H_1(X) = -\sum_i p(x_i)\, \log p(x_i)$$ — average-case

$$\text{|}\wedge$$

$$H_0(X) = \log |\{x \in \mathcal{X} : p(x) > 0\}|$$ — best-case

Prove it!

measurement of privacy

Cross-Entropy

- Measurement of the number of bits needed to identify an event *x* drawn from a set *X* if the original data are coded according to the model‘s distribution *P*, not their true distribution *Q*.

$$H(p, q) = -\sum_{x \in \mathcal{X}} p(x)\, \log q(x)$$

- Originated in privacy-preserving ML

2) Information gain/loss-based privacy metrics

- Measure how much information is gained by an adversary after the attack
- Originate from information theory
- Applied to a variety of information, although mostly in anonymous communications and database
- Well-known examples include
  - KL divergence[9]
  - Mutual information[10]
  - Loss of anonymity[11]
  - Information privacy assessment metric (IPAM) [12]

[9] J. Parra-Arnau, D. Rebollo-Monedero, J. Forné, "Measuring the Privacy of User Profiles in Personalized Information Systems", Future Gen. Comput. Syst. (FGCS), vol. 33, pp. 53-63, Apr. 2014.

[10] D. Rebollo-Monedero, J. Forné, J. Domingo-Ferrer, "From t-Closeness-Like Privacy to Postrandomization via Information Theory", IEEE Trans. Knowl., Data Eng., vol. 22, no. 11, pp. 1623-1636, Nov. 2010.

[11] K. Chatzikokolakis, C. Palamidessi, P. Panangaden, "Anonymity protocols as noisy channels", Inf. Comput. 206, 2-4, pp.378-401, Feb. 2008.

[12] S. Oukemeni, H. Rifà-Pous and J. M. Marquès Puig, "IPAM: Information Privacy Assessment Metric in Microblogging Online Social Networks," in IEEE Access, vol. 7, pp. 114817-114836, 2019.

Relative entropy

- Given two probability distributions $p(x)$ and $q(x)$ over the same alphabet, the Kullback-Leibler (KL) divergence or relative entropy is defined as

$$D(p \,\|\, q) = E_p \log \frac{p(X)}{q(X)} = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

- Let u denote the uniform distribution on an alphabet of size n. Shannon's entropy is a special case of KL divergence as per

$$D(p \,\|u) = \log n - H(p)$$

- Gives a measure of discrepancy between distributions

$$D(p \,\|\, q) \geq 0, \quad \text{with equality if, and only if, } p = q$$

- Input: prior and posterior distribution of adversary, comp. to true distribution

Interpretation of relative entropy

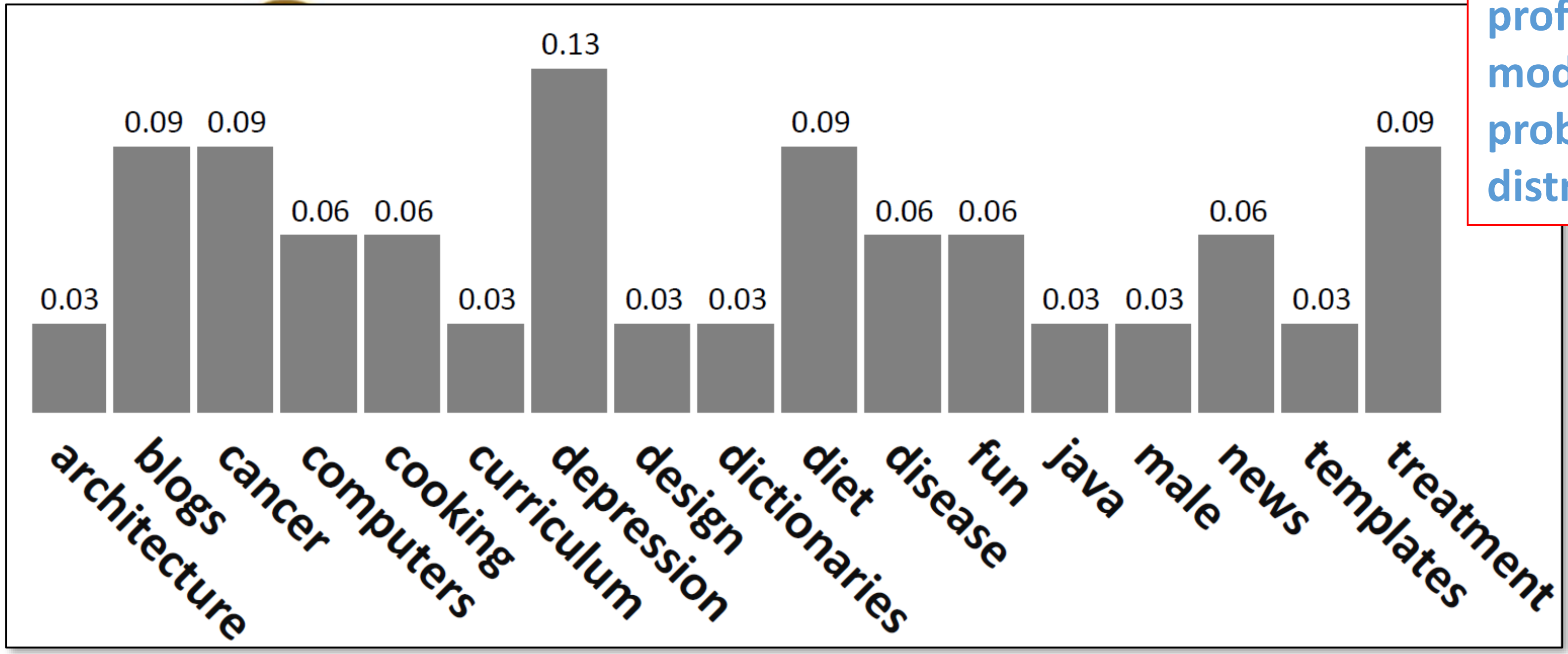
- We interpret KL divergence as privacy metric in the application of personalized information systems under two different adversary goals
  - Individuation
  - Classification
- Users counter the adversary by distorting their private data

user’s actual profile q

user’s apparent profile t

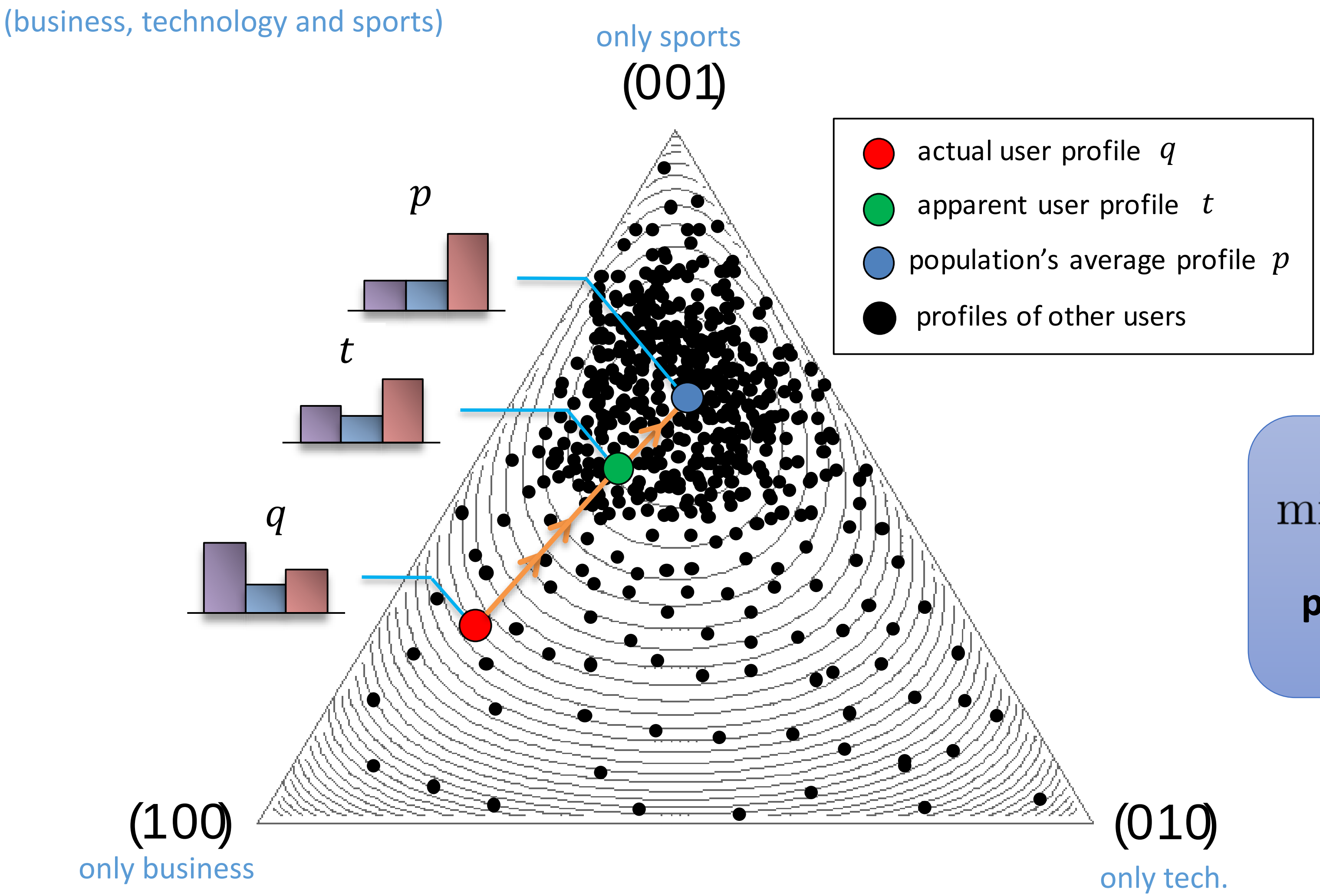
Interpretation of relative entropy

- We interpret KL divergence as privacy metric in the application of personalized information systems under two different adversary goals
  - Individuation
  - Classification
- Users counter the adversary by distorting their private data



profiles are modeled as probability distributions

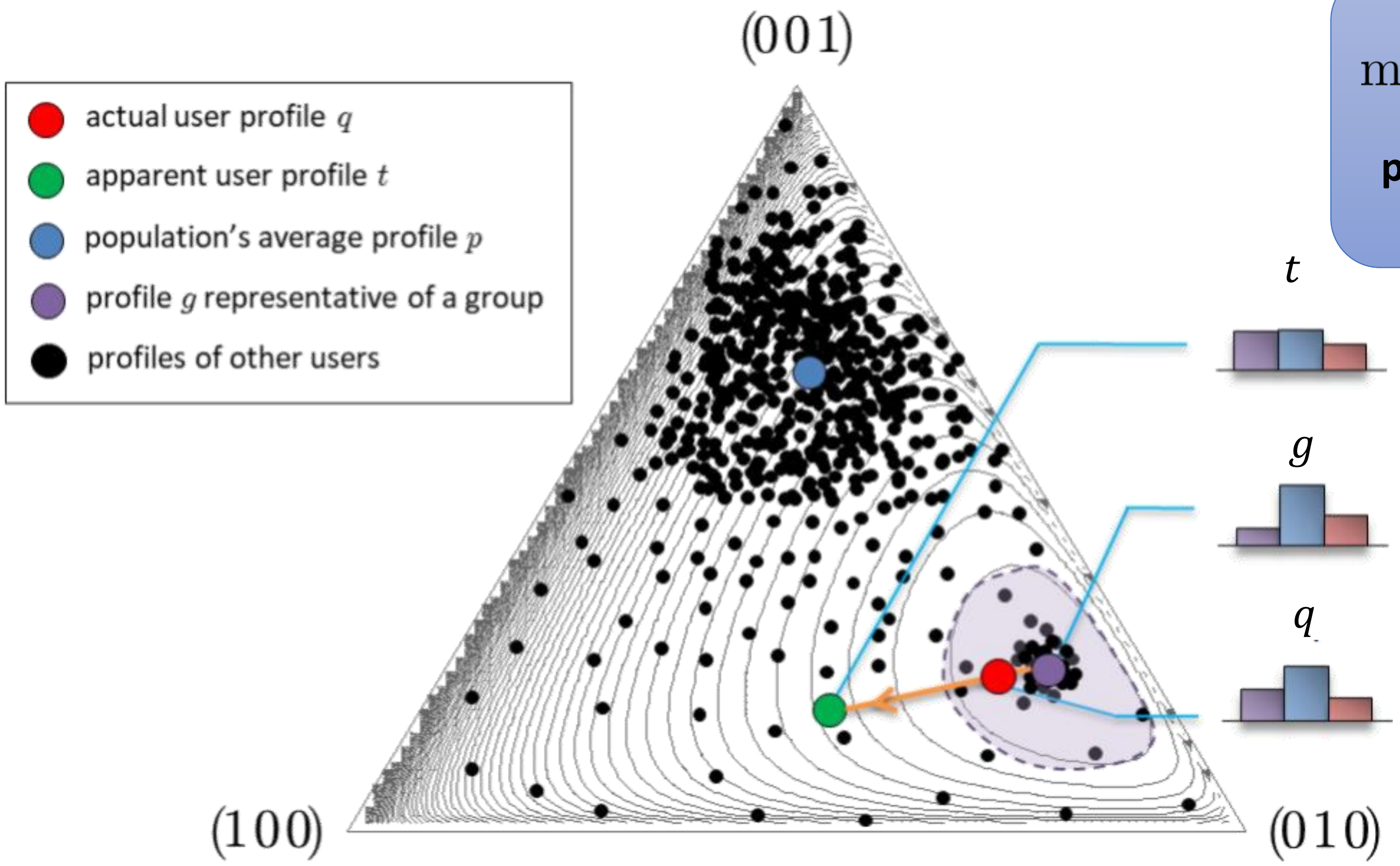
Interpretation of relative entropy



$$\min D(t \,\|\, p)$$

privacy risk

Interpretation of relative entropy



$\max D(t \,\|\, g)$

privacy gain

Mutual information

- Consider two random variables X and Y with a joint probability mass function $p(x, y)$ and marginal probability mass functions $p(x)$ and $p(y)$. The mutual information $I(X;\ Y)$ is defined as

$$I(X;Y) = E_{X,Y} \log \frac{p_{X|Y}(X|Y)}{p_X(X)} = D(p_{X,Y} \,\|\, p_X p_Y)$$

- Nonnegativity of mutual information

$I(X;Y) \geq 0,$ with equality if and only if X and Y are independent

- Typical use
  - X, sensitive unknown user data
  - Y, data observed by the adversary, accompanied possibly with background-knowledge information; or information disclosed by the user

Information privacy assessment metric (IPAM)

- Framework that calculates a privacy score in microblogging social networks
- Assessment questions measure the score based on privacy and security requirements (e.g., accessibility, information extraction)
- Examples of questions

- Compute these variables

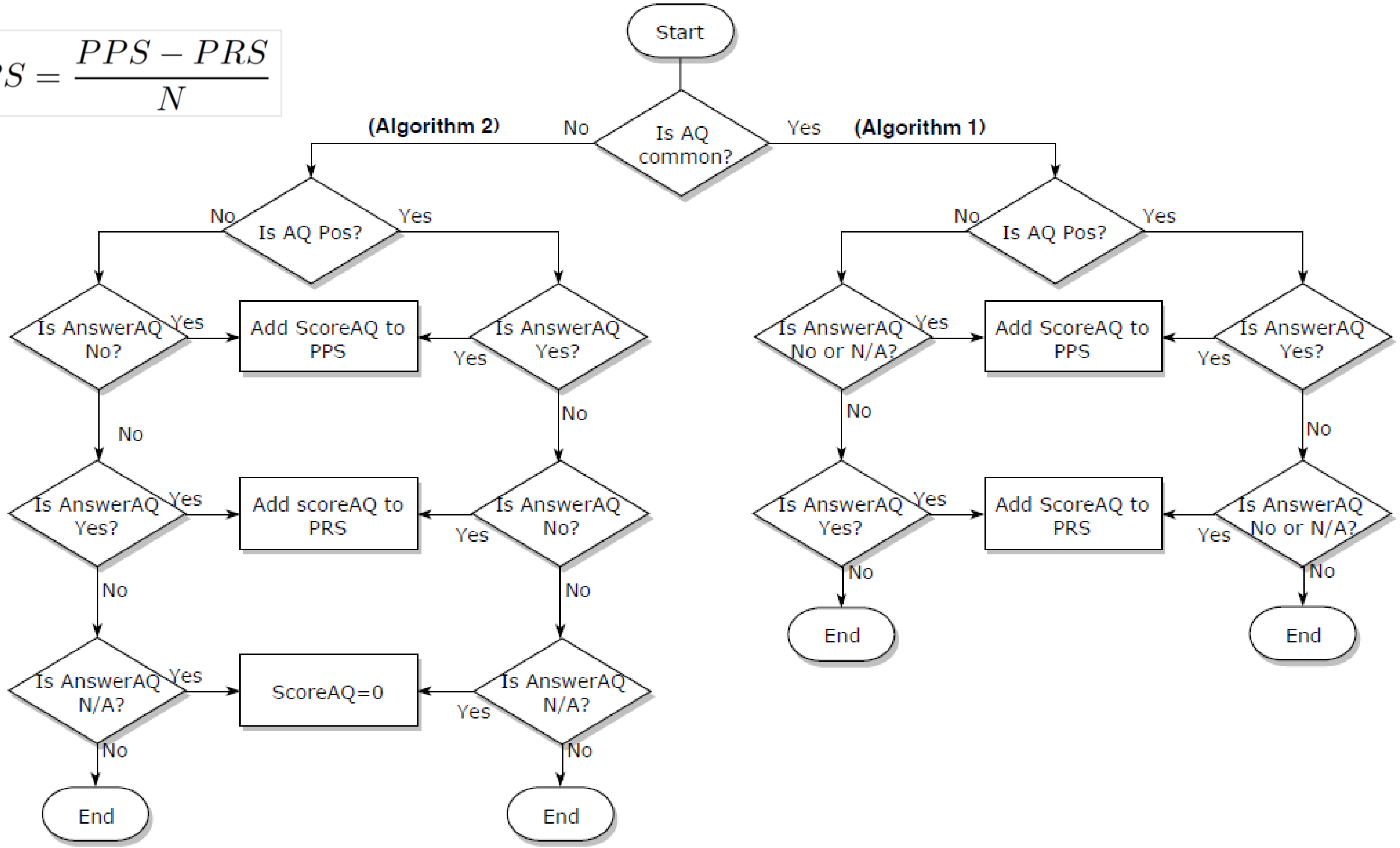
$$TPS = \frac{PPS - PRS}{N}$$

Notation	Description
TPS	Total Privacy Score
PPS	Privacy Protection Score
PRS	Privacy Risk Score
N	Total number of questions applicable to the SUI
N_{Common}	Number of answered questions in case of common set
$N_{Specific}$	Number of answered questions in case of specific set
N_{PP}	Number of answered privacy protection questions
N_{PR}	Number of answered privacy risk questions
N_{NA}	Number of answered N/A questions
$ScoreAQ$	Privacy score calculated for a question
Imp_{Priv}	Privacy Impact score
Imp_{Sec}	Security Impact score
AV	Accessibility Value
$Diff$	Data extraction difficulty

Source: original paper

Information privacy assessment metric (IPAM)

$$TPS = \frac{PPS - PRS}{N}$$



Source: original paper

3) Error-based privacy metrics

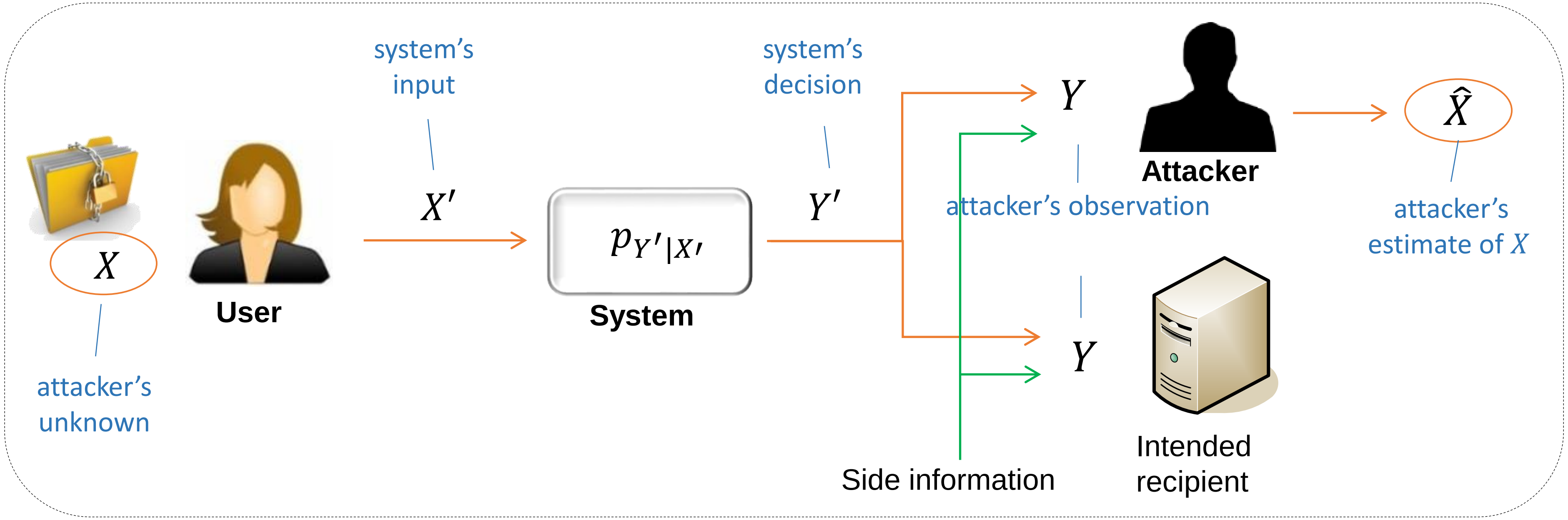
- Measure the error an adversary may make in their attempt to estimate unknown private information
- Examples include
  - Bayes risk – attacker's estimation error by Rebollo et al[13]
  - Correctness, by Shokri et al[14]
  - Mean squared error[15]

[13] D. Rebollo-Monedero, J. Parra-Arnau, C. Diaz, J. Forné, "On the measurement of privacy as an attacker's estimation error", Int. Journal Inform. Secur., vol. 12, no. 2, Apr. 2013, pp. 129-149.

[14] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, J.-P. Hubaux, "Quantifying location privacy", In Proc. IEEE Symp. on Security and Privacy, pp. 247–262, 2011.

[15] S. Oya, C. Troncoso, F. Pérez-González, "Do dummies pay off? Limits of dummy traffic protection in anonymous communications", In Proc. Privacy Enhancing Technologies (PETS), pp. 204–223, 2014.

Attacker's estimation error [13]



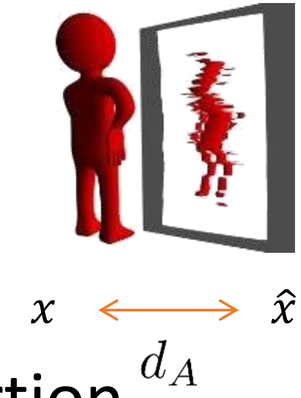
- Probabilistic formulation
  - Confidential information X, unknown to the attacker
  - User's data X' required by the system to make a decision
  - Information disclosed by the system, Y'
  - Information observed by the attacker, Y
  - Attacker's estimate $\hat{X}$ of the confidential information, from observation

Attacker's estimation error[13]

- The attacker's distortion (or error) measure $d_A(x, \hat{x})$ represents the (instantaneous) privacy attained when the unknown confidential information takes on the value $X = x$ but the attacker's estimate is $\hat{X} = \hat{x}$

- We measure privacy as the (expected) privacy attained, also known as Bayes risk,

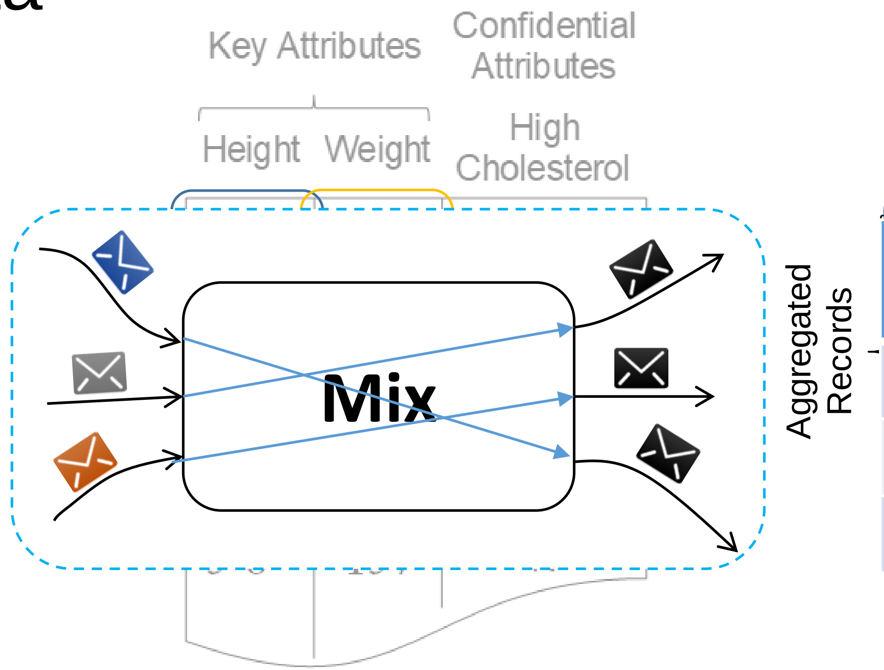
$$P = E\, d_A(X, \hat{X})$$



- Analogously, we measure (expected) utility by using a utility distortion measure $d_S(x', y')$ defined by the system, $D = E\, d_S(X', Y')$

Mean squared error

- What is the most popular measure of utility?
- In microdata



Perturbed Key Attributes
Confidential Attributes
Height Weight
High Cholesterol

Actual sender	Attacker decision	Squarede rror
2	3	1^2
3	1	2^2
1	2	1^2

MSE = (1+4+1)/3 = 2

De-identified microdata

Protected microdata

[16] S. Oya, C. Troncoso, F. Pérez-González, "Do dummies pay off? Limits of dummy traffic protection in anonymous communications", In Proc. Privacy Enhancing Technologies (PETS), pp. 204-223, 2014.

4) Metrics based on adversary's success probability

Karlsruhe Institute of Technology

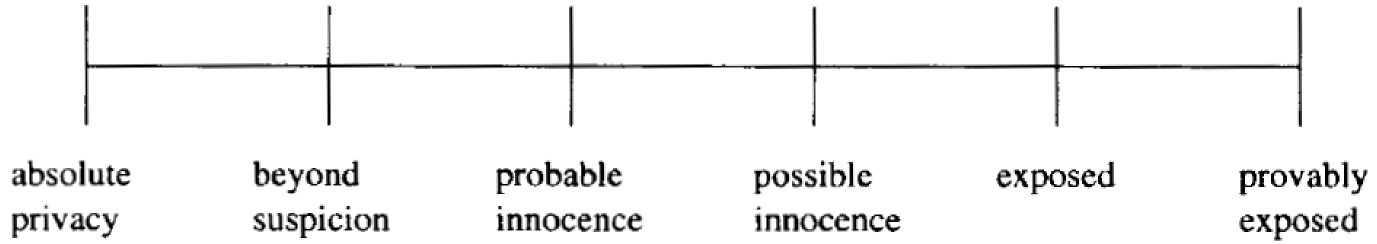
- Capture how likely the adversary will be to compromise our privacy in one or several attacks
- High privacy correlates with low success probability
- Examples include
  - Degrees of anonymity [17]
  - Sender anonymity [18]

[17] M. K. Reiter and A. D. Rubin, "Crowds: Anonymity for Web transactions", ACM Trans. Inform. Syst. Secur., vol. 1, no. 1, pp. 66-92, 1998.

[18] C. Tripp Barba, L. Urquiza Aguiar, M. Aguilar, J. Parra-Arnau, D. Rebollo-Monedero, J. Forné, E. Pallarès, "A Collaborative Protocol for Anonymous Reporting in Vehicular Ad Hoc Networks", Computer Standards & Interfaces, vol. 36, no. 1, Nov. 2013, pp. 188-197.

Degree of anonymity

- Defined in the context of anonymous communications, with respect to **sender anonymity**



- **Provably exposed**: the attacker can identify (and prove to others) the sender of a message. Formally, $p_1 = 1$.
- **Absolute privacy**: sending a message produces no observable effects on the attacker. Formally, $p_1 = 0$.
- **Beyond suspicion**: the sender appears no more likely to be the originator than others. Formally, $p_1 \leq \mathrm{p}_2, \ldots, \mathrm{p}_\mathrm{n}$
- **Probable innocent**: the sender appears no more likely to be the originator than to not be the originator. Formally, $p_1 \leq 0.5$.
- **Possible innocent**: there is a non-negligible probability that the real sender is someone else. Formally, $p_1 \leq 1 - \delta$, with $\delta \leq 0.5$
- **Exposed:** the adversary's probability is above a threshold τ (e.g., $\tau = 0.9$)

Source: original paper

5) Time-based privacy metrics

- The output is time, an important resource for adversaries to compromise user privacy
- Pessimistically assume the adversary will succeed at some point
  - Time until adversary's success[19]
  - Maximum tracking time[20]

[19] M. Wright, M. Adler, B. N. Levine, C. Shields, "An analysis of the degradation of anonymous protocols", In Proc. Network and Distributed System Security Symp. (NDSS), vol. 2. pp. 39–50, 2002.

[20] K. Sampigethaya, L. Huang, M. Li, R. Poovendran, K. Matsuura, K. Sezaki, "CARAVAN: Providing location privacy for VANET", In Embedded Security in Cars (ESCAR), pp. 29–37, 2005.

Time until adversary’s success

- In the context of ACSs
- Measure privacy as the time required for attackers to degrade the anonymity of a particular initiator with high probability
- Define “success”
  - Able to identify n out of N of the target’s possible communication peers

Maximum tracking time

- Privacy defined as the cumulative time the attacker tracks a user
- Assumes tracking is carried out only if the size of the anonymity set is 1
- Optimistic or pessimist privacy metric?

6) Data-similarity-based privacy metrics

- Arise in the context of database anonymity
- Measure properties of observable or published data
- Derive the privacy level based on the features of disclosed data
- Well-known examples include
  - k-anonymity[21]
  - p-sensitive k-anonymity[22]
  - l-diversity[23]
  - t-closeness[24]
  - stochastic t-closeness[25]

[21] L. Sweeney, "k-Anonymity: A model for protecting privacy", Int. J. Uncertain., Fuzz., Knowl.-Based Syst., vol. 10, no. 5, pp. 557-570, 2002.
[22] T. M. Truta and B. Vinay, "Privacy protection: p-sensitive k-anonymity property", in Proc. Int. Workshop Priv. Data Manage. (PDM), Atlanta, GA, 2006.
[23] A. Machanavajjhala, J. Gehrke, D. Kiefer, M. Venkitasubramanian, "l-Diversity: Privacy beyond k-anonymity", in Proc. IEEE Int. Conf. Data Eng. (ICDE), Atlanta, GA, Apr. 2006.
[24] N. Li, T. Li, S. Venkatasubramanian, "t-Closeness: Privacy beyond k-anonymity and l-diversity", in Proc. IEEE Int. Conf. Data Eng. (ICDE), Istanbul, Turkey, Apr. 2007, pp. 106-115.
[25] J. Domingo-Ferrer, J. Soria-Comas, "From t-closeness to differential privacy and vice versa in data anonymization", Know.-Based Syst. 74, 1, pp. 151–158, 2015.

k-Anonymity

Identifying Attribute

Quasi-identifier

Sensitive attribute

Name	DOB	Gender	Zipcode	Disease
Andre	1/21/76	Female	53715	Heart Disease
Beth	4/13/86	Female	53715	Hepatitis
Carol	2/28/76	Male	53703	Brochitis
Dan	1/21/76	Male	53703	Broken Arm
Ellen	4/13/86	Female	53806	Flu
Eric	2/28/76	Female	53806	Hang Nail

a tuple

- The information for each respondent contained in the released data set cannot be distinguished from at least $k - 1$ individuals
- Each tuple of quasi-identifier values in the released table must appear in at least k records

k-Anonymity

date of birth

original table

Name	DOB	Gender	Zipcode	Disease
Andre	1/21/76	Female	53715	Heart Disease
Beth	4/13/86	Female	53715	Hepatitis
Carol	2/28/76	Male	53703	Brochitis
Dan	1/21/76	Male	53703	Broken Arm
Ellen	4/13/86	Female	53806	Flu
Eric	2/28/76	Female	53806	Hang Nail

2-anonymous table

DOB	Gender	Zipcode	Disease
*	Female	5371*	Heart Disease
*	Female	5371*	Hepatitis
*	Male	5370*	Brochitis
*	Male	5370*	Broken Arm
*	Female	538**	Flu
*	Female	538**	Hang Nail

Limitations of k-anonymity

Original microdata

Alice → 47676, 27, F
Naroto → 47905, 43, M

QID			SA
Zipcode	Age	Sex	Disease
47676	27	F	Ovarian Cancer
47602	22	F	Ovarian Cancer
47678	27	M	Ovarian Cancer
47905	43	M	Heart disease
47909	52	F	Cancer
47906	47	M	Cancer

3-anonymous table

QID			SA
Zipcode	Age	Sex	Disease
476**	2*	*	Ovarian Cancer
476**	2*	*	Ovarian Cancer
476**	2*	*	Ovarian Cancer
4790*	[43,52]	*	Heart disease
4790*	[43,52]	*	Cancer
4790*	[43,52]	*	Cancer

- Suppose that the adversary knows Alice's combination of quasi-identifier attributes is (47676, 27, F). The attacker does not know which of the first 3 records corresponds to Alice's record, but learns her health condition is cancer
  - **Homogeneity attack**

- Suppose that the adversary knows Naroto's combination of quasi-identifier attributes is (47905, 47, M). The attacker learns the last record is probably Naroto's as Japanese people have low incidence of heart attacks
  - **Background knowledge attack**

Limitations of k-anonymity

- It provides identity disclosure
  - The attacker cannot find out which record corresponds to a given respondent
  - however, from the previous examples, it is prone to homogeneity and background-knowledge attacks - **no privacy at all**
- But not (sensitive or confidential) attribute disclosure
  - The adversary cannot tell that a given person has a certain sensitive attribute
- Assumes which information is available for linkage or which not

p-Sensitive, k-anonymity

3-sensitive, 6-anonymous table

Caucas	787XX	Flu
Caucas	787XX	Shingles
Caucas	787XX	Acne
Caucas	787XX	Flu
Caucas	787XX	Acne
Caucas	787XX	Flu
Asian/AfrAm	78XXX	Flu
Asian/AfrAm	78XXX	Flu
Asian/AfrAm	78XXX	Acne
Asian/AfrAm	78XXX	Shingles
Asian/AfrAm	78XXX	Acne
Asian/AfrAm	78XXX	Flu

at least 3 different values
of the confidential attribute

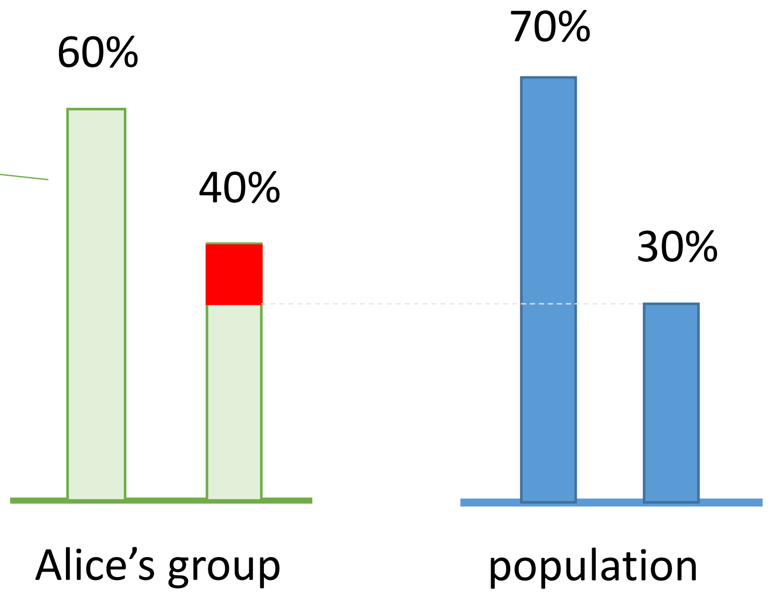
- Aimed to protect against confidential attribute disclosure
- The idea is to have at least p different sensitive values of the confidential attribute within each k-anonymous class

Limitations of p-sensitive, k-anonymity

- Prone to skewness attacks

Alice

ZIP	Age	Disease
089**	<30	N
089**	<30	Y
089**	<30	N
089**	<30	N
089**	<30	Y
283**	[30-40]	N
283**	[30-40]	N
283**	[30-40]	Y
283**	[30-40]	N
283**	[30-40]	N



l-Diversity

- The idea is that the sensitive attributes are “diverse” within each k-anonymous group
- Each equivalence class has at least l well-represented sensitive values
- Different meanings of “well-represented” values, in addition to distinct l-diversity
  - Entropy l-diversity. The entropy of the distribution of sensitive values in each equivalence class is at least log l

$$H(Z|X=x) = -\sum_z p_{Z|X}(z|x) \log p_{Z|X}(z|x) \geq \log l \quad \text{for all class } x$$

entropy of the confidential attribute Z on the equivalent class x

parameter

Limitations of l-diversity

- Still vulnerable to skewness attacks
- And similarity attacks…

3-diverse, 3-anonymous table

QID			SA
Zipcode	Age	Sex	Disease
476**	2*	*	Lung Cancer
476**	2*	*	Prostate Cancer
476**	2*	*	Bladder Cancer
4790*	[43,52]	*	Heart disease
4790*	[43,52]	*	Flu
4790*	[43,52]	*	Diabetes

t-Closeness

overall distribution

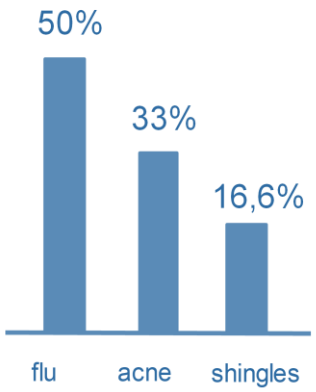
Caucas	787XX	Flu
Caucas	787XX	Shingles
Caucas	787XX	Acne
Caucas	787XX	Flu
Caucas	787XX	Acne
Caucas	787XX	Flu
Asian/AfrAm	78XXX	Flu
Asian/AfrAm	78XXX	Flu
Asian/AfrAm	78XXX	Acne
Asian/AfrAm	78XXX	Shingles
Asian/AfrAm	78XXX	Acne
Asian/AfrAm	78XXX	Flu

- The idea is that the distribution of confidential attributes given perturbed key attributes observed must be close to the entire distribution of the confidential attribute

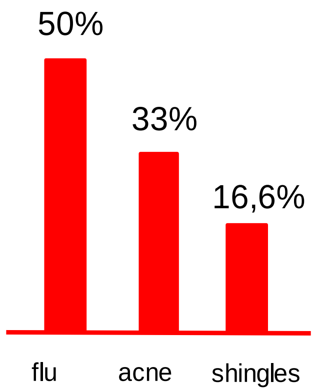
$$\left| d(p_{Z|X}(z|x), p_Z(z)) \right| \leq t$$

confidential

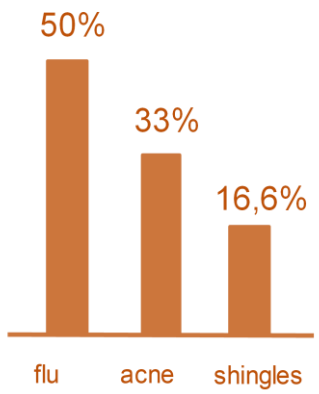
group or equivalence class



distribution of group 1

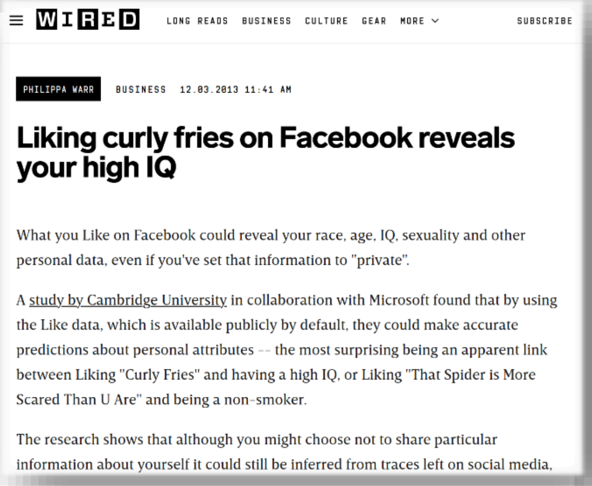


distribution of group 2



overall distribution

The Inference Privacy Fallacy



WIRED LONG READS BUSINESS CULTURE GEAR MORE SUBSCRIBE

PHILIPPA WARR BUSINESS 12.03.2013 11:41 AM

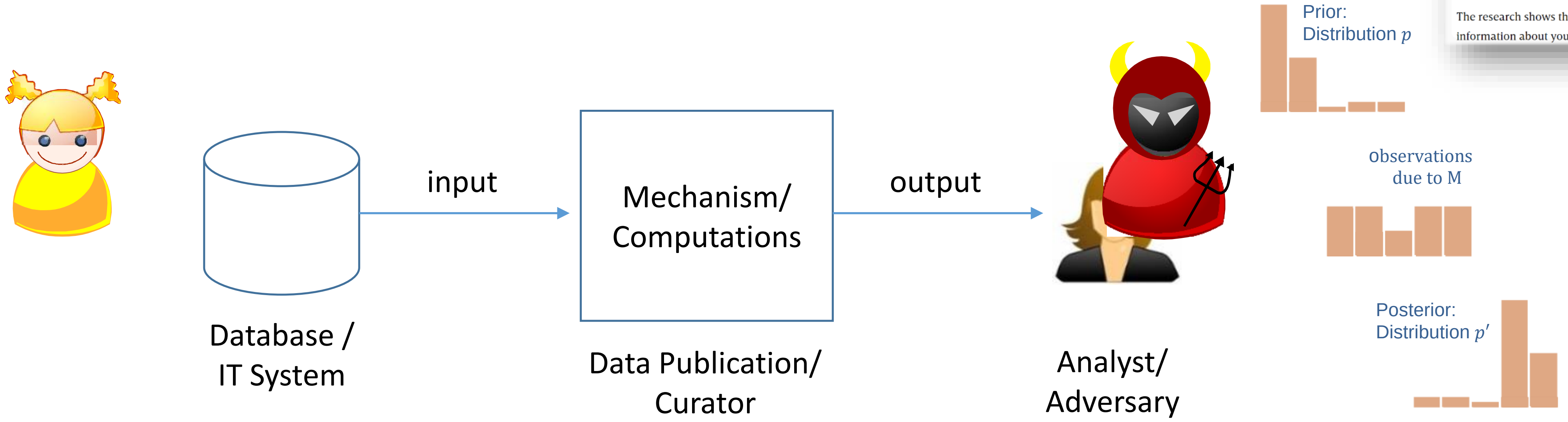
Liking curly fries on Facebook reveals your high IQ

What you Like on Facebook could reveal your race, age, IQ, sexuality and other personal data, even if you've set that information to "private".

A study by Cambridge University in collaboration with Microsoft found that by using the Like data, which is available publicly by default, they could make accurate predictions about personal attributes -- the most surprising being an apparent link between Liking "Curly Fries" and having a high IQ, or Liking "That Spider is More Scared Than U Are" and being a non-smoker.

The research shows that although you might choose not to share particular information about yourself it could still be inferred from traces left on social media.

- We measure the privacy of the data release mechanism



- We cannot measure (or: protect) adaptation to the prior (and corresponding inference)
- If statistics are revealed, they are useless or help improve the prior

6) Indistinguishability-based privacy metrics

Karlsruhe Institute of Technology

- Is the adversary able to distinguish between two outcomes of a PET?
- The harder for the adversary to distinguish any pair of outcomes, the higher the privacy provided by the PET
- Typically binary metrics
- Examples include
  - Differential privacy[27]
  - Individual differential privacy[28]

[27] C. Dwork, "Differential privacy," in Proc. Int. Colloq. Automata, Lang., Program. Springer-Verlag, 2006, pp. 1-12.

[28] J. Soria-Comas, J. Domingo-Ferrer, D. Snchez, and D. Megas, "Individual differential privacy: a utility-preserving formulation of differential privacy guarantees," IEEE Transactions on Information Forensics and Security, vol. 12, no. 6, pp. 1418-1429, Jun. 2017.

Differential privacy

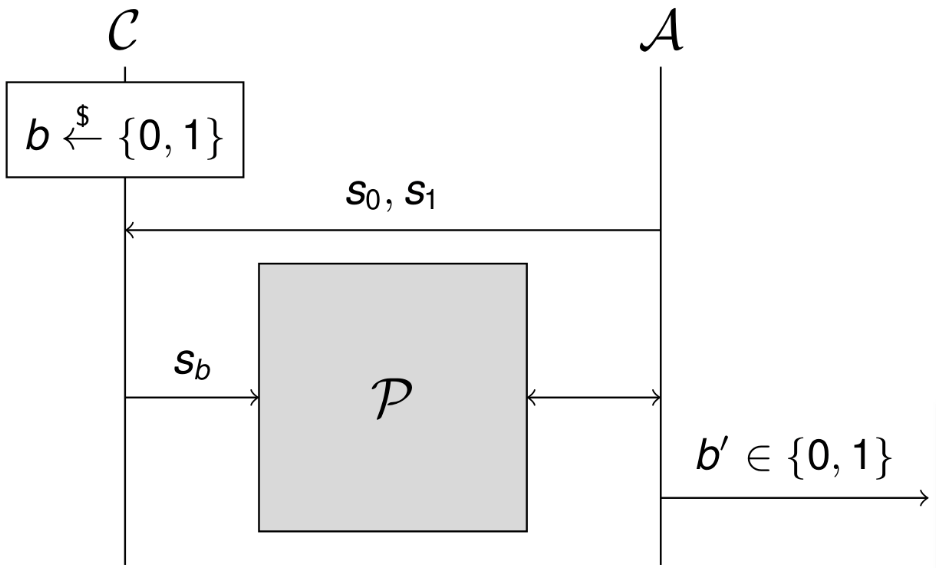
To be discussed
a bit later

- Setting
  - **Database:** composed of individual
  - **Curator:** aimed to protect individu
  - **Analyst or data user:** wishes to perform computations on the database
- A computation protects the privacy of individuals in the data if its output does not reveal any information that is **specific to any individual data subject**
- Differential privacy formalizes this intuition as a mathematical definition

Back to ACS: Indistinguishability Games

Karlsruhe Institute of Technology

- Recall IND-CPA game from crypto…



[29] Kuhn et al., "On Privacy Notions in Anonymous Communication", PoPETS (2) 2019: 105-125

- Communication properties
  - U and U' – Which senders/receivers are active?
  - |U| and |U'| – How many senders/receivers are active?
  - Q and Q' – Which user sends/receives how many messages?
  - H and H' – How many users sends/receives how many messages?
  - P and P' – Which messages are send/received by the

Example: Sender Notions

- All disclose receiver-message relation, but hide who sends which message
- **Sender Unobservability ($S\overline{O}$)** additionally discloses number of
- communications
- **Sender-Frequency Unlinkability** ($SF\overline{L}$) additionally discloses number of communications and set of active users
- **Sender-Messages Unlinkability** ($SM\overline{L}$) additionally discloses number of communications, set of active users, and number of messages per sender

Summary

- Selection of over 25 privacy metrics across four privacy domains
- Followed the structure proposed by[1] based on metrics' outputs
  - Uncertainty
  - Information gain/loss
  - Data similarity/dissimilarity
  - Indistinguishable
  - Error-based metrics
  - Time-based metrics
- Best-case, average-case and worst-case
- Connections among them (e.g., stochastic t-closeness and DP)

[1] Isabel Wagner and David Eckhoff, "Technical Privacy Metrics: A Systematic Survey", ACM Comput. Surv. 51, 3, Article 57, June 2018.