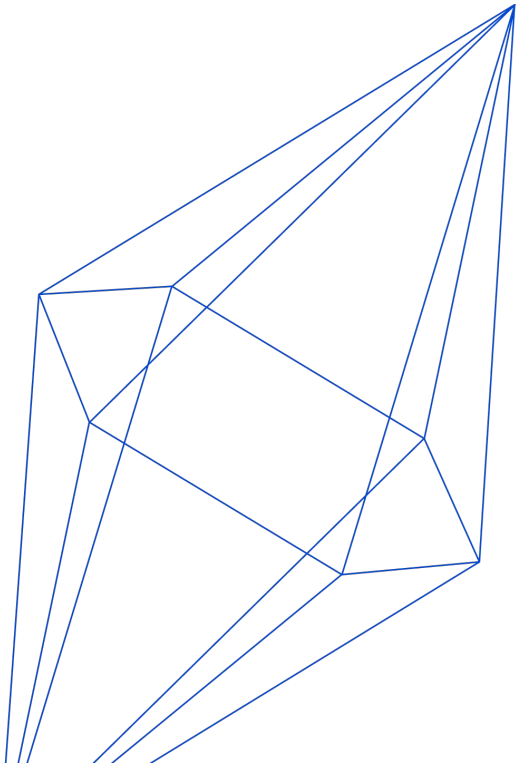
Per-Attribute Privacy Semantics for the Supplemental -Demographic and Housing Characteristics File (S -DHC)

David Pujol, Ashwin Machanavajjhala, William Sexton



David Pujol• Scientist • david.pujol@tmlt.io

FCSM Conference, October 2024

S-DHC Privacy Loss

- S-DHC uses zero Concentrated Differential Privacy (zCDP) for privacy protection

- zCDP measures:
  - the worst-case privacy loss for any possible individual, and
  - The worst-case privacy loss for any arbitrary change in a record

- We often, additionally, care about characteristics of an individual which are only represented by a subset of the overall record.

- We introduce the attribute privacy loss as a measure of the privacy loss of a subset of an individual's attributes and apply it to the S-DHC data product.

Outline

- Background and Motivation
- Per-attribute Bounded zCDP
- Example Analysis
- Full Results

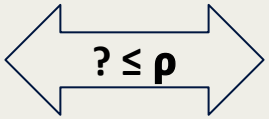
Bounded zCDP

A randomized mechanism **M** satisfies bounded **ρ**-zCDP, if for any two databases **D** and **D'** which only differ in the arbitrary change of one record the following holds for all values of $\alpha \in (0,1)$:

$$D_\alpha(M_D \| M_{D'}) = \frac{1}{\alpha - 1} \log \mathbb{E}_{x \sim M_{D'}} \left(\frac{M_D(x)}{M_{D'}(x)} \right)^\alpha \leq \rho$$

Bounded zCDP

Name	Age	Relation	Household Size
Alice	20	Biological Child	3
Bob	16	Biological Child	3
Carol	20	Householder	1
Dave	16	Householder	3
Elise	60	Householder	10



Name	Age	Relation	Household Size
Alice	50	Non-Biological Child	12
Bob	16	Biological Child	3
Carol	20	Householder	1
Dave	16	Householder	3
Elise	60	Householder	10

S-DHC Privacy Loss

PH1
Average Household Size By Age

ρ = 0.366

PH2
Household Type for the Population in Households

ρ = 0.038

PH3
Household Type by Relationship for the Population Under 18 Years

ρ = 1.358

PH4
Population in Families by Age

ρ = 0.364

PH5
Average Family Size by Age

ρ = 0.002

PH6
Family Type and Age for Own Children Under 18 Years

ρ = 0.028

PH7
Total Population in Occupied Housing Units by Tenure

ρ = 0.661

PH8
Average Household Size of Occupied Housing Units by Tenure

ρ = 0.002

Total Privacy Loss = 2.819

https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/supplemental-demographic-and-housing-characteristics-file/2020census-supplemental-dhc-techdoc.pdf

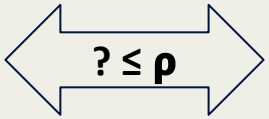
A-Attribute Privacy Loss

Let **A** be a set of attributes. A randomized mechanism M incurs an **A-**attribute privacy loss of **ρ** if for any two databases D and D' which only differ in one record, *which differs only in the attributes in* ***A,*** the following holds for all values of $\alpha \in (0,1)$:

$$D_\alpha(M_D \| M_{D'}) = \frac{1}{\alpha - 1} \log \mathbb{E}_{x \sim M_{D'}} \left(\frac{M_D(x)}{M_{D'}(x)} \right)^\alpha \leq \rho$$

{Relation} -Attribute Privacy Loss

Name	Age	Relation	Household Size
Alice	20	Biological Child	3
Bob	16	Biological Child	3
Carol	20	Householder	1
Dave	16	Householder	3
Elise	60	Householder	10



Name	Age	Relation	Household Size
Alice	20	Adopted Child	3
Bob	16	Biological Child	3
Carol	20	Householder	1
Dave	16	Householder	3
Elise	60	Householder	10

Alice's {Relation} -Attribute Privacy Loss

PH1
Average Household Size By Age

ρ = 0

PH2
Household Type for the Population in Households

ρ = 0.005

PH3
Household Type by Relationship for the Population Under 18 Years

ρ = 0

PH4
Population in Families by Age

ρ = 0.048

PH5
Average Family Size by Age

ρ = 0.002

PH6
Family Type and Age for Own Children Under 18 Years

ρ = 0

PH7
Total Population in Occupied Housing Units by Tenure

ρ = 0

PH8
Average Household Size of Occupied Housing Units by Tenure

ρ = 0

Total {Relation} Privacy Loss = 0.055

Bob's {Relation} -Attribute Privacy Loss

PH1
Average Household Size By Age

ρ =0

PH2
Household Type for the Population in Households

ρ = 0.005

PH3
Household Type by Relationship for the Population Under 18 Years

ρ = 0.443

PH4
Population in Families by Age

ρ = 0.048

PH5
Average Family Size by Age

ρ = 0.002

PH6
Family Type and Age for Own Children Under 18 Years

ρ = 0.009

PH7
Total Population in Occupied Housing Units by Tenure

ρ = 0

PH8
Average Household Size of Occupied Housing Units by Tenure

ρ =0

Total {Relation} Privacy Loss = 0.507

Measured Secrets

Secret	Attributes	Description
Sex	Sex, Relation	An individual's exact sex
Age	Age, Relation	An individual's exact age
Race-Ethnicity	Race, Ethnicity	An individual's racial ethnic makeup
Same-Sex Relationship	Sex, Relation	If an individual is currently in a same-sex relationship
Unmarried Partner	Relation	If an individual is in an unmarried partnership
Multi-racial Family	Race, Ethnicity, Relation	If an individual belongs to or lives in a household with a multiracial family
Over Rental Capacity	HouseholdID Tenure	If an individual lives in a rental household that is currently over capacity
Exact Household	HouseholdID	The exact location of an individual's household

Example Individuals

Name	Age	Relation	Household Size	Household Type
Alice	20	Biological Child	3	Married Couple, Household
Bob	16	Biological Child	3	Married Couple Household, with own Children < 18
Carol	20	Householder	1	Nonfamily: Female Householder
Dave	16	Householder	3	Married Couple Household, with own Children < 18
Elise	60	Householder	10	Married Couple Household, with own Children < 18

Results

Name	Sex	Age	Race Ethnicity	Same-Sex Relationship	Unmarried Partner	Multi-racial Family	Over Rental Capacity	Exact Household
Alice	0.055	0.558	0	0.055	0.055	0.055	1.433	1.433
Bob	0.508	0.558	0.443	0.508	0.508	0.508	2.819	2.819
Carol	0.126	0.126	0.038	0.126	0	0.039	1.067	1.067
Dave	0.508	0.558	0.633	0.508	0.508	0.647	2.819	2.819
Elise	0.404	2.156	1.395	0.404	1.79	2.819	1.433	1.433

Results: Large Households

Name	Sex	Age	Race Ethnicity	Same-Sex Relationship	Unmarried Partner	Multi-racial Family	Over Rental Capacity	Exact Household
Alice	0.055	0.558	0	0.055	0.055	0.055	1.433	1.433
Bob	0.508	0.558	0.443	0.508	0.508	0.508	2.819	2.819
Carol	0.126	0.126	0.038	0.126	0	0.039	1.067	1.067
Dave	0.508	0.558	0.633	0.508	0.508	0.647	2.819	2.819
Elise	0.404	2.156	1.395	0.404	1.79	2.819	1.433	1.433

Results: Age under 18

Name	Sex	Age	Race Ethnicity	Same-Sex Relationship	Unmarried Partner	Multi-racial Family	Over Rental Capacity	Exact Household
Alice	0.055	0.558	0	0.055	0.055	0.055	1.433	1.433
Bob	0.508	0.558	0.443	0.508	0.508	0.508	2.819	2.819
Carol	0.126	0.126	0.038	0.126	0	0.039	1.067	1.067
Dave	0.508	0.558	0.633	0.508	0.508	0.647	2.819	2.819
Elise	0.404	2.156	1.395	0.404	1.79	2.819	1.433	1.433

Results: Householders

Name	Sex	Age	Race Ethnicity	Same-Sex Relationship	Unmarried Partner	Multi-racial Family	Over Rental Capacity	Exact Household
Alice	0.055	0.558	0	0.055	0.055	0.055	1.433	1.433
Bob	0.508	0.558	0.443	0.508	0.508	0.508	2.819	2.819
Carol	0.126	0.126	0.038	0.126	0	0.039	1.067	1.067
Dave	0.508	0.558	0.633	0.508	0.508	0.647	2.819	2.819
Elise	0.404	2.156	1.395	0.404	1.79	2.819	1.433	1.433

Conclusions

- We introduced the A-privacy loss to measure the privacy loss of an individual's fine grained characteristics
- We demonstrated that the A-privacy loss is always no greater than the overall privacy loss and often leads a much lower privacy loss than the worst -case analysis.
- Data curators can use these values to inform them of the risk of particular secrets being revealed when privacy losses are high
- When applied to S -DHC we conclude that there are three classes of individuals who often incur significantly less A-privacy loss than others:
  - Individuals that live in small households
  - Adults above the age of 18
  - Non -Householders