

ISO/IEC 20889:2018 (E)

Privacy enhancing data de-identification terminology and classification of techniques

Contents

	Foreword
	Introduction
1	Scope
2	Normative references
3	Terms and definitions
4	Symbols and abbreviated terms
5	Overview
6	Technical model and terminology
7	Re-identification
7.1	General
7.2	Re-identification attacks
8	Usefulness of de-identified data
9	De-identification techniques
9.1	Statistical tools
9.1.1	General
9.1.2	Sampling
9.1.3	Aggregation
9.2	Cryptographic tools
9.2.1	General
9.2.2	Deterministic encryption
9.2.3	Order-preserving encryption
9.2.4	Format-preserving encryption
9.2.5	Homomorphic encryption
9.2.6	Homomorphic secret sharing
9.3	Suppression techniques
9.3.1	General
9.3.2	Masking
9.3.3	Local suppression
9.3.4	Record suppression
9.4	Pseudonymization techniques
9.4.1	General
9.4.2	Selection of attributes
9.4.3	Creation of pseudonyms
9.4.3.1	General
9.4.3.2	Pseudonyms independent of identifying attributes
9.4.3.3	Pseudonyms derived from identifying attributes using cryptography
9.4.3.3.1	Overview of cryptography use for pseudonymization
9.4.3.3.2	Encryption
9.4.3.3.3	Hashing
9.5	Anatomization
9.6	Generalization techniques
9.6.1	General
9.6.2	Rounding

- 9.6.3 Top and bottom coding
- 9.6.4 Combining a set of attributes into a single attribute
- 9.6.5 Local generalization
- 9.7 Randomization techniques
 - 9.7.1 General
 - 9.7.2 Noise addition
 - 9.7.3 Permutation
 - 9.7.4 Microaggregation
- 9.8 Synthetic data
- 10 Formal privacy measurement models
 - 10.1 General
 - 10.2 K-anonymity model
 - 10.2.1 General
 - 10.2.2 L-diversity
 - 10.2.3 T-closeness
 - 10.3 Differential privacy model
 - 10.3.1 General
 - 10.3.2 Server model
 - 10.3.3 Local model
 - 10.3.4 Key considerations for a Differentially Private System
 - 10.3.4.1 Probability distribution
 - 10.3.4.2 Sensitivity
 - 10.3.4.3 Privacy budget, ϵ
 - 10.3.4.4 Cumulative privacy loss
 - 10.4 Linear sensitivity model
 - 10.4.1 General
 - 10.4.2 Threshold rule
 - 10.4.3 Dominance rule
 - 10.4.4 Ambiguity rule
- 11 General principles for application of de-identification techniques
 - 11.1 General
 - 11.2 Sampling considerations
 - 11.3 Aggregated vs. microdata
 - 11.4 Classification of attributes
 - 11.5 Handling of direct identifiers
 - 11.6 Handling of remaining attributes
 - 11.7 Privacy guarantee models
- 12 Additional technical or organizational measures
 - 12.1 General
 - 12.2 Data flow scenarios
 - 12.3 Access to de-identified data
 - 12.4 Controlled re-identification
- Annex A (informative) Summary of de-identification tools and techniques
- Annex B (informative) Prior art terminology
 - B.1 General
 - B.2 Definitions of individual terms
 - B.3 Relationship to ISO/IEC 19944
 - B.4 Relationship to Statistical Disclosure Control terminology
- Annex C (informative) De-identification of free-form text
 - C.1 General
 - C.2 Annotating
 - C.3 Conversion of data to the form of a table
 - C.4 Scrubbing
 - C.5 Segmentation
 - C.6 Aggregation
- Annex D (informative) Normalization of structured data

Annex E (informative) Overview of approaches to formal privacy measurement models

- E.1 General**
- E.2 Record linkage**
- E.2.1 General**
- E.2.2 (X,Y)-anonymity**
- E.2.3 Multirelational K-anonymity**
- E.3 Attribute linkage**
- E.3.1 General**
- E.3.2 Confidence bounding**
- E.3.3 (X,Y)-privacy**
- E.3.4 (a,k)-anonymity**
- E.3.5 (k,e)-anonymity**
- E.3.6 Personalized privacy**
- E.3.7 FF-anonymity**
- E.3.8 m-invariance**
- E.4 Table linkage**
- E.4.1 General**
- E.4.2 Delta-presence**
- E.5 Probabilistic attacks**
- E.5.1 General**
- E.5.2 (d,y)-privacy**
- E.6 Variants of differential privacy**
- E.6.1 General**
- E.6.2 Approximate differential privacy**
- E.6.3 Random differential privacy**
- E.6.4 Probabilistic differential privacy**
- E.6.5 Concentrated differential privacy**
- E.6.6 Approximate concentrated differential privacy**
- E.6.7 Multiparty differential privacy**
- E.6.8 Computational differential privacy**
- E.7 Variants of L-diversity**

Page count: 46