k-Anonymität

Thomas Maier, Kai Sonnenwald, Tom Petersen

Universität Hamburg
Fachbereich Informatik

Agenda

1. Motivation & Abgrenzung
2. k-Anonymität
   - Generalisierung
   - Unterdrückung
3. Schwächen der k-Anonymität
4. l-Diversity
5. Schwächen der l-Diversity
6. t-Closeness
7. Literaturverzeichnis

Motivation

Einführendes Beispiel, dass das Setting und in die Begriffe einführt

Abgrenzung

Abgrenzung zu anderen Konzepten (statistische Datenbanken, Authentifikation, ...)

k-Anonymität

Notizen

identifier, quasi-identifier, sensitive attributes

k-anonymity

Begriffe

Explicit identifier Attribut, das ein Individuum (nahezu) eindeutig identifiziert. Bsp: Name, Adresse, Steuernummer, ...

Quasi identifier Attributmenge, die ein Individuum in Kombination identifizieren kann. *Formal in [Swe02] p. 7*, auch [MKGV07] p. 3

Sensitive attribute Attribut, dessen Wert für ein Individuum in einer Datenmenge nicht herausgefunden werden darf.

k-Anonymität

Informell: Eine Tabelle (Datensatz?) erfüllt k-Anonymität, wenn jede Zeile (jeder Eintrag) ununterscheidbar von $k-1$ anderen Zeilen im Bezug auf jeden “quasi identifier“-Menge ist.

Formal: Sei $T(A_1, \ldots, A_n)$ eine Tabelle und $Q_T = \{A_i, \ldots, A_j\}$ der zugehörige quasi identifier. T erfüllt k-Anonymität genau dann, wenn jede Belegung von Werten in $T[Q_T]$ mindestens k mal auftritt, wobei $T[Q_T]$ die duplikatenerhaltende Projektion von T auf die Attribute des quasi identifiers beschreibt.

Generalisierung

s. Samarati, Sweeney Kapitel 3

domain, ground domain, generalization (partial ordering on domains)
generalized table

Suppression

...

in combination with generalization

Schwächen der k-Anonymität

Unsorted matching attack Veröffentlichung mehrerer $k-$anonymer Tabellen mit derselben Sortierung ausgehend von einer nicht-öffentlichen Tabelle. [Swe02] p.10

Complementary release attack Veröffentlichung mehrerer $k-$anonymer Tabellen unterschiedlicher Generalisierung, die zusammengeführt die $k-$Anonymität verletzen. [Swe02] p.11

Temporal attack Dynamische Tabellen können $k-$Anonymität verletzen. [Swe02] p.12

Homogeneity attack Gleichheit der sensitive attributes einer Gruppe, die sich in den Werten des quasi identifiers gleicht, leakt das sensitive attribute eines Individuums. [MKGV07] p. 2

Background knowledge attack Nutzen von Hintergrundwissen, um mit hoher Wahrscheinlichkeit auf den Wert des sensitive attributes eines Individuumsin einer Gruppe

I-Diversity

Schwächen der l-Diversity

Skewness attack

similarity attack

t-Closeness

Literaturverzeichnis


MACHANAVAJJHALA, Ashwin ; KIFER, Daniel ; GEHRKE, Johannes ; VENKITASUBRAMANIAM, Muthuramakrishnan:
l-diversity: Privacy beyond k-anonymity.
In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 1 (2007), Nr. 1, S. 3


SAMARATI, Pierangela ; SWEENEY, Latanya:
Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression / Technical report, SRI International.
1998. –
Forschungsbericht


SWEENEY, Latanya:
k-anonymity: A model for protecting privacy.
In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10 (2002), Nr. 05, S. 557–570