On the Connections between Privacy Models Used in Statistical Disclosure Control

Josep Domingo-Ferrer

Universitat Rovira I Virgili

josep.domingo@urv.cat

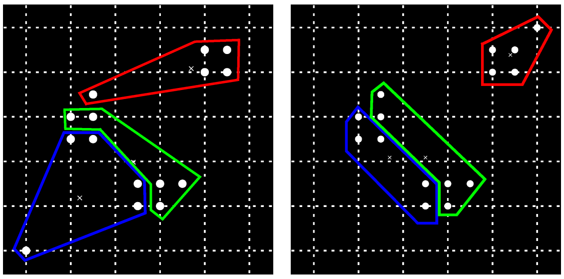
United Nations Educational, Scientific and Cultural Organization

Universitat Rovira i Virgili

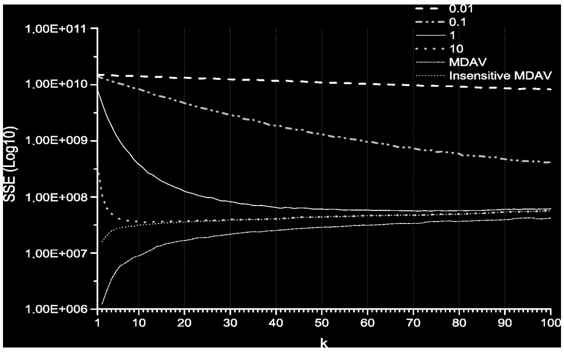
Chair in Data Privacy

Prior k-anonymity via insensitive microaggregation to reduce data utility loss when achieving ε-differential privacy in data releases

Insensitive microaggregation

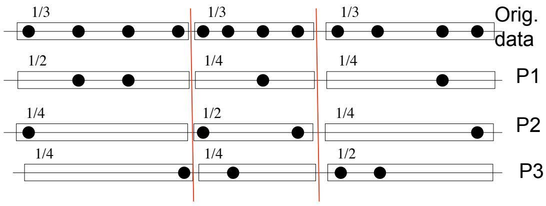


Information loss (k=1 is standard diff. privacy, ε from 0.01 to 10)



Construction to achieve t-closeness and ε-differential privacy

Partition of the data set into groups P1, P2, P3… by the quasi-identifiers and bucketization of the confidential attribute to achieve t-closeness



◆The granularity of confidential attribute is reduced, so t-closeness is achieved with distance

$$d(\mathcal{D}_1, \mathcal{D}_2) = \max_S \left\{ \frac{\Pr_{\mathcal{D}_1}(S)}{\Pr_{\mathcal{D}_2}(S)}, \frac{\Pr_{\mathcal{D}_2}(S)}{\Pr_{\mathcal{D}_1}(S)} \right\}$$

where D1 and D2 are two random distributions differing in one record and S is an arbitrary set.

◆ For uninformed intruders, such t-closeness implies ε-differential privacy with ε=ln(t):

$$\Pr_{\mathcal{D}_1}(S) \le \exp(\varepsilon) \times \Pr_{\mathcal{D}_2}(S)$$

where D1 is the distribution of the confidential attribute in the whole protected data set and D2 is the distribution of the confidential attribute in the group Pi containing a specific individual.

From ε-differential privacy to expected t-closeness

Let X be an original data set and X' be a corresponding anonymized data set such that its quasi-identifiers are k-anonymous and the projection of X' on the confidential attributes is ε-differentially private. Then X' satisfies expected t-closeness with

$$t = g^{-1}(\exp((N-k) \times \varepsilon))$$

Hence, a greedy way to achieve actual t-closeness is to keep generating ε-differentially private versions of the confidential attribute until a t-close version is found.

Conclusions

The k-anonymity, t-closeness and differential privacy models are connected. Using a prior k-anonymization step based on insensitive microaggregation allows achieving differential privacy in data set releases with less utility loss. Also, exp(ε)-closeness implies ε-differential privacy for uninformed intruders in data releases. Finally, k-Anonymity for quasi-identifiers combined with ε-differential privacy for confidential attributes yields t-closeness in expectation, with t=f(k,ε).

References

J. Domingo-Ferrer. On the connection between t-closeness and differential privacy for data releases. In *10th International Conference on Security and Cryptography-SECRYPT 2013*, Reykjavik, Iceland, July 29-31, 2013.

J. Soria-Comas, J. Domingo-Ferrer, D. Sánchez and S. Martínez. Improving the utility of differentially private releases via k-anonymity. In *12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications-IEEE TrustCom 2013*, IEEE Computer Society, Melbourne, Australia, July 16-18, 2013.

J. Soria-Comas and Josep Domingo-Ferrer. Differential privacy via t-closeness in data publishing. In *11th Annual Conference on Privacy, Security and Trust-PST 2013*, IEEE Computer Society, Tarragona, Catalonia, July 10-12, 2013.