1 Motivation

Protect information contained in statistical database, i.e, if a database is a representative sample of an underlying, the goal of a privacy preserving statistical database is to enable the user to learn properties of the population as a whole , while protecting the privacy of the individuals in the sample.
What we want to achieve is something close to semantic security. Dalenius in 1977 articulated a desired property for database: nothing abouth an individual should be learnable from a database that cannot be learned without access to the database. This kind of privacy cannot be achieved: the obstacle is *auxiliary information.* Here an example:
suppose one's exact height were considered sensitive information and revealing it were a privacy breach. Assume that the database yields the average heights of women of different nationalities. AN adversary who access the database and the auxiliary information "Terry Gross is two inches shorter than the average lithuanian woman" learns Terry Gross height while anyone learning only the auxiliary information without access to the database ,learns relatively little.
There are to remarks to do wrt the impossibility results:

- it applies regardless of whether or not Terry Gross is in the database
- Dalenius goal cannot be achieved while semantic security for crypto yes.

The first of these leads naturally to a new approach to formulating privacy goals: the risk to ones privacy, or in general, any type of risk, such as the risk of being denied automobile insurance, should not substantially increase as a result of participating in a statistical database. This is captured by *differential privacy*

2 Differential Privacy: the setting

Given the impossibility result just showed we need to relax the definition of privacy: we move from absolute guarantees about disclosures to relative ones; any given disclosure will be, within a small multiplicative factor.

Definition 2.1. A randomized algrotihm A gives ϵ-differential privacy if for all data sets D_1 and D_2 differing on at most one element, and all S $\subseteq$ Range(A),

$$Pr[A(D_1) \in S] \leq exp(\epsilon) \times Pr[A(D_2) \in S] \tag{1}$$

A mechanism K satisfying this definition addresses concerns that any participant might have about the leakage of her personal information x: even if the participant removed her data from the data set, no outputs (and thus consequences of outputs) would become significantly more or less likely. For example, if the database were to be consulted by an insurance provider before deciding whether or not to insure Terry Gross, then the presence or absence of Terry Gross in the database will not significantly affect her chance of receiving coverage.

3 Achieving DP