Probabilistic Methods for Privacy and Secure Information Flow

Catuscia Palamidessi

Lecture 2

Resume of previous lecture

Privacy via anonymization.

- k-anonymity, based on the notion of quasi-identifier
- ℓ-diversity

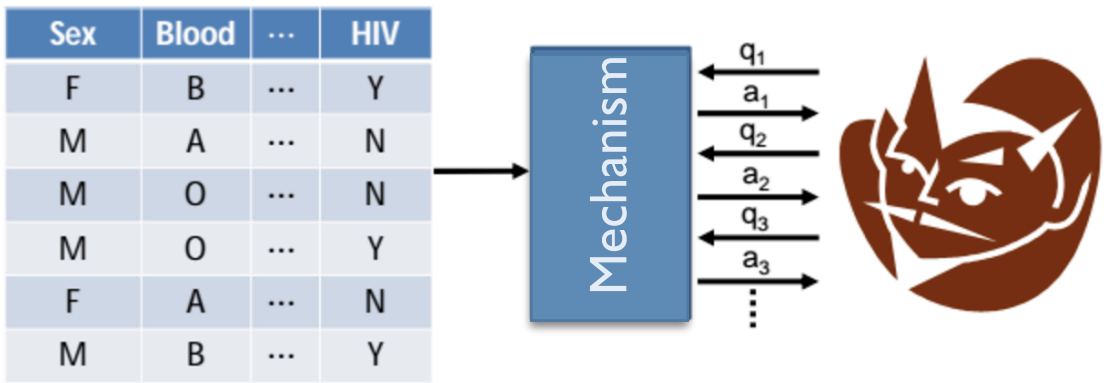
We saw that these methods are ineffective, due to the following:

1. The whole set of attributes can be a quasi-identifier.
   - Attacks on large sparse datasets. Example: Netfix prize attack
2. Vulnerability to composition attacks
   - example of combination of queries
   - general problem of deterministic methods

Solution: **randomization**

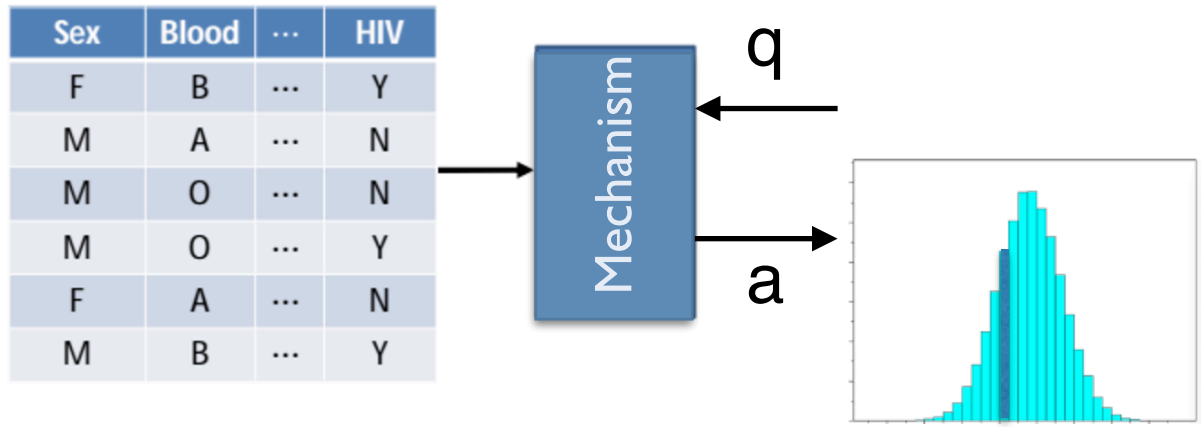
We assume the following setting:

- Centralized model (i.e., the data curator is *trusted*)
- Micro-data are not publicly accessible. The information can only be accessed by querying the DB



In this context, the access to the DB is via an interface (**mechanism**) which receives the queries, computes the answers and sanitises them before reporting them

Randomized mechanisms



- A randomized mechanism (for a certain query) reports an answer generated randomly according to some probability distribution
- We need to choose carefully the distribution, so to get the desired privacy guarantees, while maintaining a good utility for the query
- To find a good trade-off between privacy and utility, and to reason about them, we need formal, rigorous definitions of these notions.
- A definition of privacy that has become very popular: Differential Privacy [Cynthia Dwork, ICALP 2006]

Databases

- A record is an element v from some domain $\mathcal{V}$ of values. In general $\mathcal{V}$ is a structured domain, i.e., it is a product of domains corresponding to the attributes. But for our purposes the structure is not relevant and in general we will ignore it
- A **database** (or dataset) of n records is an element of $\mathcal{X} = \mathcal{V}^n$. We will represent the elements of $\mathcal{X}$ by $x, x_1, x_2, \ldots$.
- We will assume a probability distribution on distribution on $\mathcal{V}$ and $\mathcal{X}$ and indicate by V, X the respective random variables

Examples:

$\mathcal{V}$ = integers

x

20
14
51
75

$\mathcal{V}$ = names x integers

x

John	20
Mary	14
Dale	51
Anna	75

Adjacency

- Two databases x_1, x_2 are **adjacent** if they differ for exactly one record. We will indicate this property with the notation $x_1 \sim x_2$
  - $x_1 \sim x_2$ represent the fact that x_1 and x_2 differ for the information relative to an individual. Either this individual has been added to x_2, or he has been removed from x_2

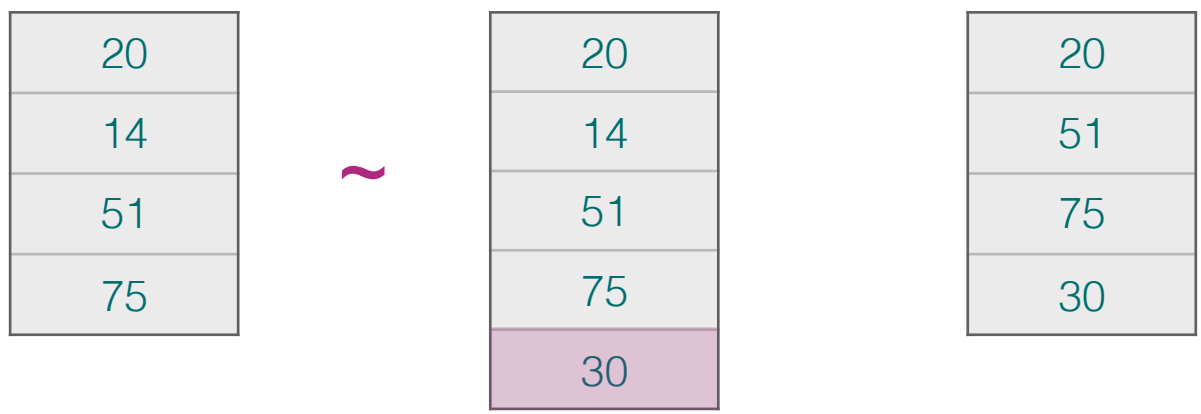
20
14
51
75

20
14
51
75
30

20
51
75
30

Adjacency

- Two databases x_1, x_2 are **adjacent** if they differ for exactly one record. We will indicate this property with the notation $x_1 \sim x_2$
  - $x_1 \sim x_2$ represent the fact that x_1 and x_2 differ for the information relative to an individual. Either this individual has been added to x_2, or he has been removed from x_2



Adjacency

- Two databases x_1, x_2 are **adjacent** if they differ for exactly one record. We will indicate this property with the notation $x_1 \sim x_2$
  - $x_1 \sim x_2$ represent the fact that x_1 and x_2 differ for the information relative to an individual. Either this individual has been added to x_2, or he has been removed from x_2

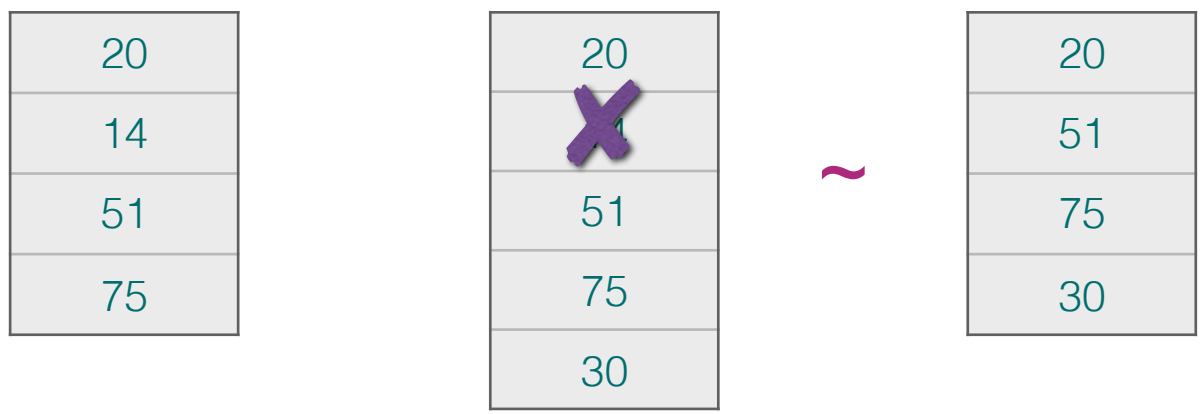
20
14
51
75

20
14
51
75
30

20
51
75
30

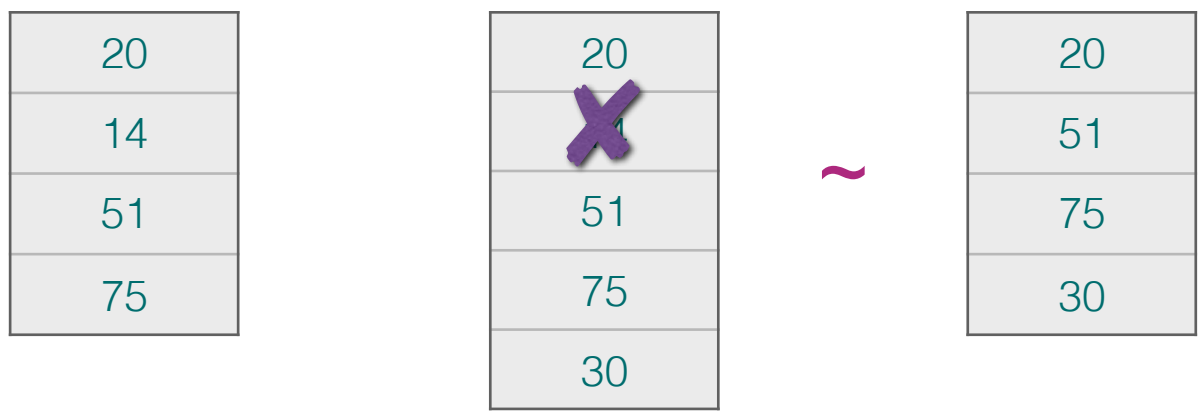
Adjacency

- Two databases x_1, x_2 are **adjacent** if they differ for exactly one record. We will indicate this property with the notation $x_1 \sim x_2$
  - $x_1 \sim x_2$ represent the fact that x_1 and x_2 differ for the information relative to an individual. Either this individual has been added to x_2, or he has been removed from x_2



Adjacency

- Two databases x_1, x_2 are **adjacent** if they differ for exactly one record. We will indicate this property with the notation $x_1 \sim x_2$
  - $x_1 \sim x_2$ represent the fact that x_1 and x_2 differ for the information relative to an individual. Either this individual has been added to x_2, or he has been removed from x_2



The adjacency relation is symmetric but not transitive

Queries

- (The answer to) a query f can be seen as a function from the set of databases $\mathcal{X} = V^n$ to a set of values $\mathcal{Y}$. Namely,

$$f : \mathcal{X} \to \mathcal{Y}$$

- $y = f(x)$ is the **true answer** of the query f on the database x.
- For a given f, the distribution π on $\mathcal{X}$ also induces a distribution on $\mathcal{Y}$. We will denote by Y the random variable associated to the distribution on $\mathcal{Y}$.

Example: f = average of all values in the DB

x
20
14
51
75

$f(x) = (20+14+51+75)/4 = 40$

Randomized mechanisms

- A randomized mechanism for the query f is any probabilistic function $\mathcal{K}$ from $\mathcal{X}$ to a set of values $\mathcal{Z}$. Namely,

$$\mathcal{K} : \mathcal{X} \to \mathcal{D}\mathcal{Z}$$

  where $\mathcal{D}\mathcal{Z}$ represents the set of probability distributions on $\mathcal{Z}$.

- $\mathcal{Z}$ does not necessarily coincide with $\mathcal{Y}$.

- z drawn from $\mathcal{K}(x)$ is a **reported answer** for the query on the DB x.

- Note that π and $\mathcal{K}$ induce a probability distribution also on $\mathcal{Z}$. We will denote by Z the random variable associated to this probability distribution

Differential Privacy

We are now ready to define differential privacy. We first consider the discrete case, i.e., when the reported answer is discrete

Definition (Differential Privacy) $\mathcal{K}$ is ε-differentially-private iff for every pair of databases $x_1, x_2 \in \mathcal{X}$ s.t. $x_1 \sim x_2$ and for every $z \in \mathcal{Z}$ we have

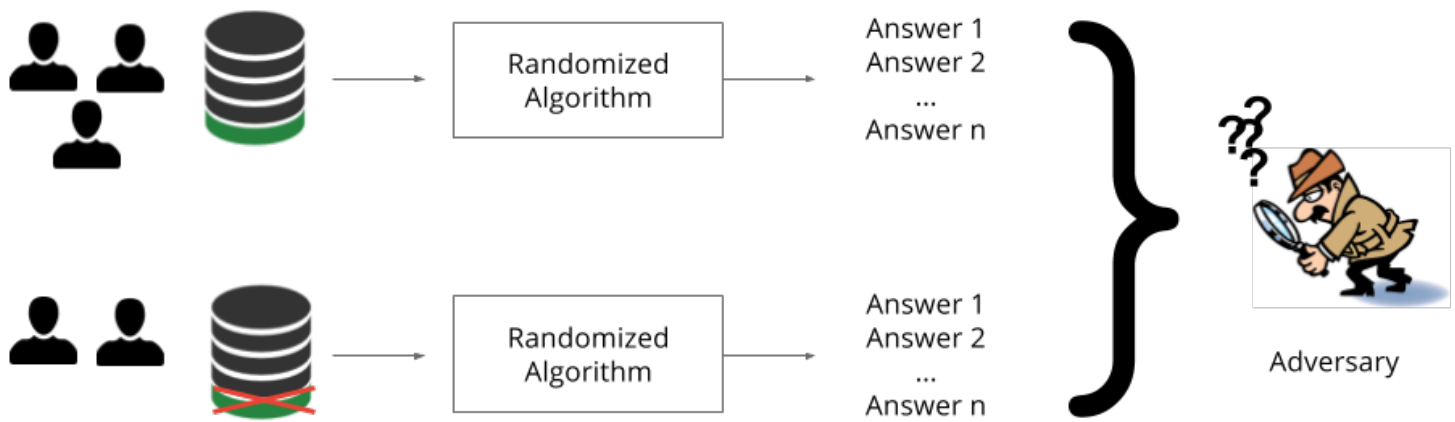
$$p(\mathcal{K}(x_1) = z) \leq e^{\varepsilon}\, p(\mathcal{K}(x_2) = z)$$

where $p(\mathcal{K}(x) = z)$ represents the probability that $\mathcal{K}$ applied to x reports the answer z

Note: $p(\mathcal{K}(x) = z)$ represents a conditional probability. We will write it as $p(Z = z|X = x)$ when we need to make this fact more explicit.

Meaning of Differential Privacy

Differential privacy essentially means that the presence or absence of an individual in a DB, does not make much difference for the information that the adversary acquires by querying the DB.



Hence an individual does not risks much by accepting that his data are collected in the DB

Is DP what we want?
Is DP the best we can do?

What we really would like to have is that by querying the DB the adversary cannot derive much information about the individual

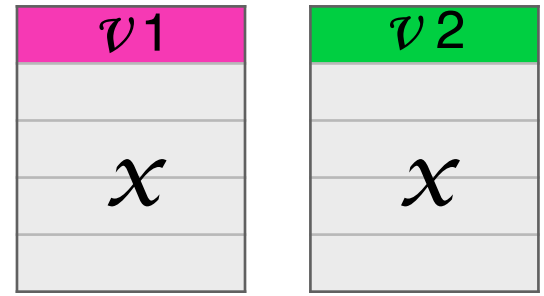
Unfortunately, this is not possible

Example: Assume that the adversary knows that Turing has the same height of the average height of the rest of the people in the DB. Then, by querying the DB, the adversary gains a lot of information about the heigh of Turing (if we want to preserve some utility)

Note that this happens whether or not John is in the DB

Other interpretations of DP

Consider two databases that differ only for the value of one individual record

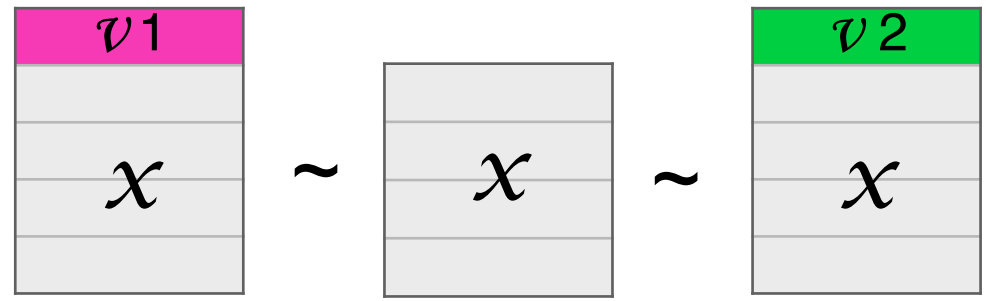


Theorem If $\mathcal{K}$ is ε-differentially-private then $\forall v_1, v_2 \in \mathcal{V}, \forall x \in \mathcal{X}, \forall z \in \mathcal{Z}$

$$p(\mathcal{K}(x \cup v_1) = z) \leq e^{2\varepsilon}\, p(\mathcal{K}(x \cup v_2) = z)$$

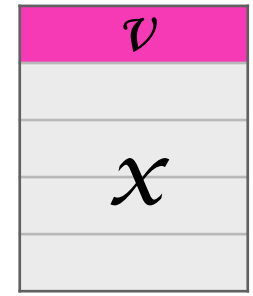
Note that also the reverse is true: it is sufficient to enrich $\mathcal{V}$ with an extra value $\perp$ that represents the absence of an individual

The proof of the theorem is immediate, just observe the relation with X, and then apply transitivity of $\leq$



Other interpretations of DP: Bayesian

Consider a database consisting of $\mathcal{X}$ plus a record $\mathcal{V}$



Theorem $\mathcal{K}$ is ε-differentially-private iff $\forall v \in \mathcal{V}, \forall x \in \mathcal{X}, \forall z \in \mathcal{Z}$

$$p(V = v|X = x, Z = z) \leq e^{2\varepsilon}\, p(V = v|X = x)$$

$$p(V = v|X = x) \leq e^{2\varepsilon}\, p(V = v|X = x, Z = z)$$

Proof: exercise.

This means that, if the adversary knows the value of all the other records of the database, then knowing the reported answer z does not improve much his knowledge of a given individual V

The assumption that the adversary knows the value of all the other records of the database is called **strong adversary model**

Question 1 Is the hypothesis of the strong adversary necessary for the result?

Question 2 How does this result reconcile with the example of the height of Turing ?

Examples of mechanisms

Let us assume that we have databases containing as values $\mathcal{V}$ the heights of people, in cm, ranging from $\mathbf{50}$ to $\mathbf{250}$ (integers). Let us assume that the query is: the average age of the people in the data base, rounded to the next integer

Examples of mechanisms

Let us assume that we have databases containing as values $\mathcal{V}$ the heights of people, in cm, ranging from $\mathbf{50}$ to $\mathbf{250}$ (integers). Let us assume that the query is: the average age of the people in the data base, rounded to the next integer

- Consider the mechanism that always reports the true answer. Is it differentially private ?

Examples of mechanisms

Let us assume that we have databases containing as values $\mathbf{V}$ the heights of people, in cm, ranging from $\mathbf{50}$ to $\mathbf{250}$ (integers). Let us assume that the query is: the average age of the people in the data base, rounded to the next integer

- Consider the mechanism that always reports the true answer. Is it differentially private ?

  No. It's not ε-DP for any ε

Examples of mechanisms

Let us assume that we have databases containing as values $\mathbf{V}$ the heights of people, in cm, ranging from $\mathbf{50}$ to $\mathbf{250}$ (integers). Let us assume that the query is: the average age of the people in the data base, rounded to the next integer.

- Consider the mechanism that always reports 150. Is it differentially private ?

Examples of mechanisms

Let us assume that we have databases containing as values $\mathbf{V}$ the heights of people, in cm, ranging from $\mathbf{50}$ to $\mathbf{250}$ (integers). Let us assume that the query is: the average age of the people in the data base, rounded to the next integer.

- Consider the mechanism that always reports 150. Is it differentially private ?

  Yes. It's ε-DP in the strong sense, i.e., for $\varepsilon = 0$.

Examples of mechanisms

Let us assume that we have databases containing as values $\mathbf{V}$ the heights of people, in cm, ranging from $\mathbf{50}$ to $\mathbf{250}$ (integers). Let us assume that the query is: the average age of the people in the data base, rounded to the next integer.

- Consider the mechanism that always reports 150. Is it differentially private ?

  Yes. It's ε-DP in the strong sense, i.e., for $\varepsilon = 0$

  However, it's totally useless !

Examples of mechanisms

Let us assume that we have databases containing as values $\mathbf{V}$ the heights of people, in cm, ranging from $\mathbf{50}$ to $\mathbf{250}$ (integers). Let us assume that the query is: the average age of the people in the data base, rounded to the next integer.

- Consider the mechanism that reports $\mathbf{100}$ if the true answer is less than $\mathbf{150}$, and $\mathbf{200}$ otherwise. Is it differentially private ?

Examples of mechanisms

Let us assume that we have databases containing as values $\mathbf{V}$ the heights of people, in cm, ranging from $\mathbf{50}$ to $\mathbf{250}$ (integers). Let us assume that the query is: the average age of the people in the data base, rounded to the next integer.

- Consider the mechanism that reports $\mathbf{100}$ if the true answer is less than $\mathbf{150}$, and $\mathbf{200}$ otherwise. Is it differentially private ?

  No. It's a bit more useful than the previous one, but it is not ε-DP for any ε

Examples of mechanisms

Let us assume that we have databases containing as values $\mathbf{V}$ the heights of people, in cm, ranging from $\mathbf{50}$ to $\mathbf{250}$ (integers). Let us assume that the query is: the average age of the people in the data base, rounded to the next integer.

Consider the mechanism that reports the true answer with probability $\frac{e^{\varepsilon}}{200+e^{\varepsilon}}$, and every other integer in $[50, 250]$ with probability $\frac{1}{200+e^{\varepsilon}}$. Is it differentially private ?

Examples of mechanisms

Let us assume that we have databases containing as values $\mathbf{V}$ the heights of people, in cm, ranging from $\mathbf{50}$ to $\mathbf{250}$ (integers). Let us assume that the query is: the average age of the people in the data base, rounded to the next integer.

Consider the mechanism that reports the true answer with probability $\frac{e^\varepsilon}{200+e^\varepsilon}$, and every other integer in $[50, 250]$ with probability $\frac{1}{200+e^\varepsilon}$. Is it differentially private ?

Yes. It's ε-DP

It is also relatively useful. We will study its utility later.

Properties of differential privacy

- Two important properties that have made differential privacy so successful:
  - Independence from the side knowledge of the adversary
  - Compositionality

Independence from the side knowledge of the adversary

- The distribution π on the databases is called prior, i.e., prior to the reported answer
- π represents the knowledge that a potential adversary has about the database (before knowing the answer of $\mathcal{K}$)
- We note that the definition of DP does not depend on π. This is a very good property, because it means that we can design mechanisms that satisfy DP without taking the knowledge of the adversary into account: the same mechanism will be good for all adversaries.

Compositionality

- Differential privacy is compositional, namely: given two mechanisms $\mathcal{K}_1$ and $\mathcal{K}_2$ on $\mathcal{X}$ that are respectively ε_1 and ε_2-differentially private, their composition $\mathcal{K}_1 \times \mathcal{K}_2$ is $(\varepsilon_1 + \varepsilon_2)$-differentially private.

  Note: $\mathcal{K}_1 \times \mathcal{K}_2$ is defined by the following property: if $\mathcal{K}_1(x)$ reports z_1 and $\mathcal{K}_2(x)$ reports z_2, then $(\mathcal{K}_1 \times \mathcal{K}_2)(x)$ reports (z_1, z_2).

  Proof: exercise

- **Privacy budget**: A DB is associated to an initial budget α. Each time a user asks a query, answered by ε-differentially private mechanism, his budget is decreased by ε. When his budget is exhausted, users are not allowed to ask queries anymore.
  Note that the budget is per DB and not per user because users may be colluded.