CSE660
Differential Privacy

November 13, 2017

Marco Gaboardi
Room: 338-B
gaboardi@buffalo.edu
http://www.buffalo.edu/~gaboardi

Differential privacy

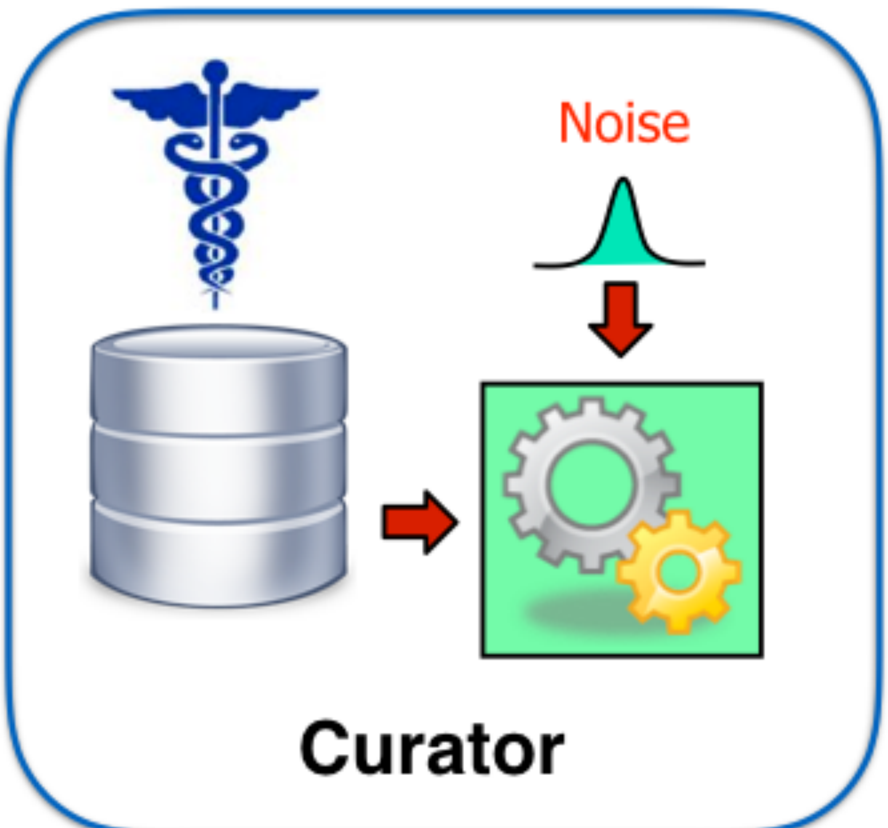
Definition

Given $\varepsilon,\delta \geq 0$, a probabilistic query $Q: X^n \rightarrow R$ is (ε,δ)-differentially private iff for all adjacent database b_1, b_2 and for every $S \subseteq R$:

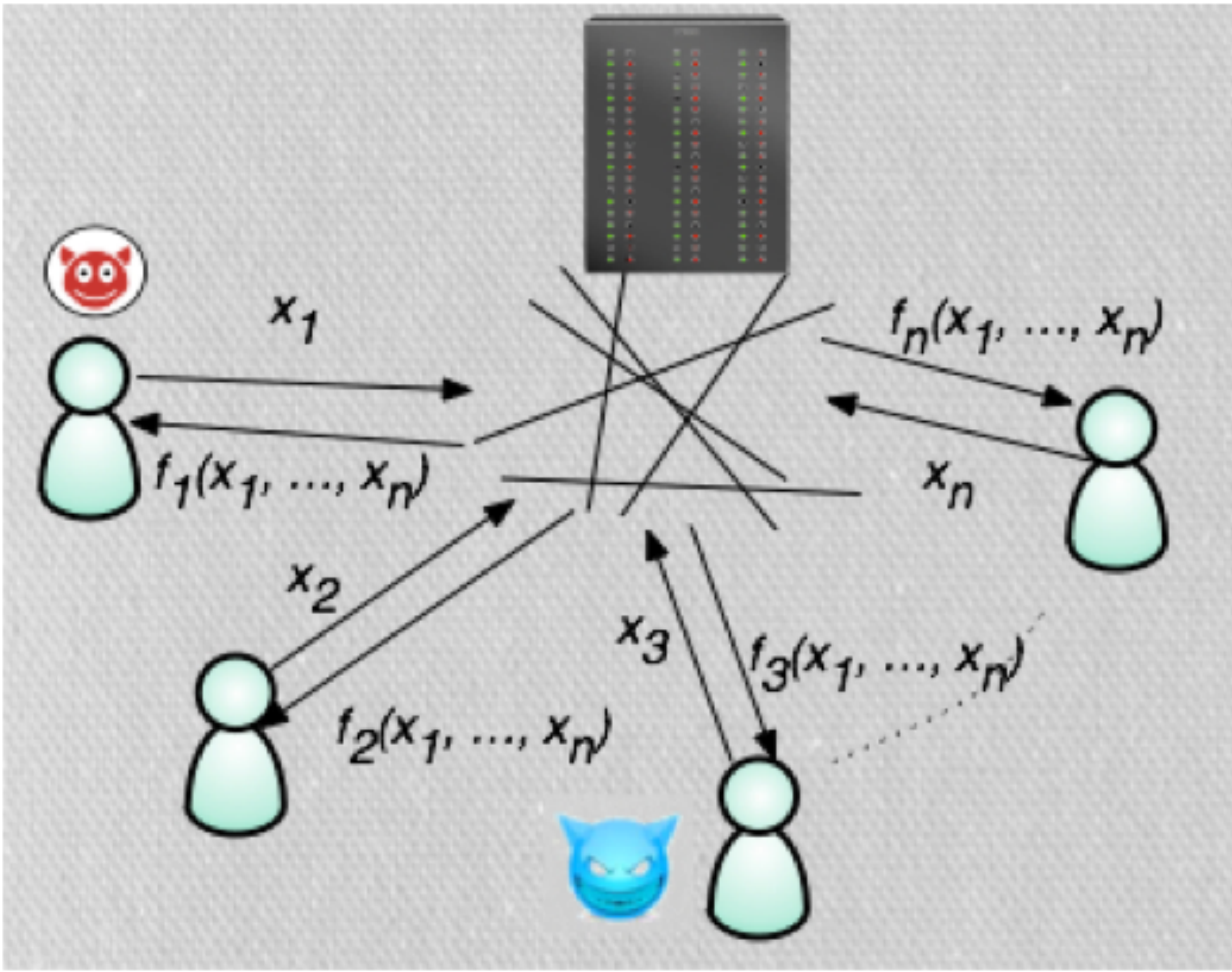
$$\Pr[Q(b_1) \in S] \leq \exp(\varepsilon)\Pr[Q(b_2) \in S] + \delta$$

Differential privacy

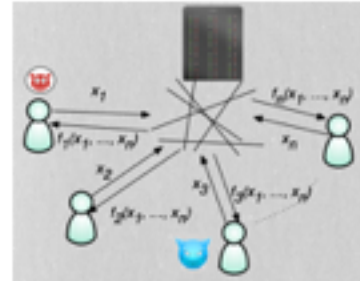
So far, we have considered a **curator model**: a model where there is a trusted centralized party that holds the data and to which we can ask our queries.



Multiparty differential privacy



Multiparty Setting



We now consider a model where the data is distributed among m parties $P_1,\ldots,P_m$.

We assume that the data is evenly split among the parties, each party P_i has n/m rows of the dataset.

Each party P_i want to guarantee privacy for its data against an adversary that may control the other parties.

We will study protocols to compute statistics over the data.

Adversaries

We assume that the adversaries are:

- passive (honest-but-curious): they follow the specified protocol but try to extract information from what they see,
- computationally unbounded: we will not restrict the capacity of the adversary,
- control several parties: an adversary can control $t \leq m-1$ parties. We will focus on $t=m-1$.

Protocol

$$(P_1, \ldots, P_m)(x)$$

We consider a protocol as a sequence of rounds where:

- every party P_i selects a message to be broadcast based on its input (a part of x), internal coin tosses, and all messages received in previous rounds,
- the output of the protocol is specified by a deterministic function of the transcript of messages exchanged,

Adversary view

$$\mathrm{View}_{P_{-k}}(P_{-k} \leftrightarrow (P_1, \ldots, P_m)(x)) \in T$$

We are interested in a protection against an adversary that controls all the parties except the k-th one.

The view of the adversary is then determined by the inputs and coin tosses of all parties other than P_k as well as the messages sent by P_k.

Multiparty differential privacy

Definition 9.1 (multiparty differential privacy [7]). For a protocol $P = (P_1, \ldots, P_m)$ taking as input datasets $(x_1, \ldots, x_m) \in (\mathcal{X}^{n/m})^m$, we say that P is (ε, δ) differentially private (for passive adversaries) if for every $k \in [m]$ and every two dataset $x, x' \in (\mathcal{X}^{n/m})^m$ that differ on one row of P_k's input (and are equal otherwise), the following holds for every set T:

$$\Pr[\mathrm{View}_{P_{-k}}(P_{-k} \leftrightarrow (P_1, \ldots, P_m)(x)) \in T] \leq e^{\varepsilon} \cdot \Pr[\mathrm{View}_{P_{-k}}(P_{-k} \leftrightarrow (P_1, \ldots, P_m)(x')) \in T] + \delta.$$

Randomized Response is optimal in the local model

Theorem 9.3 (randomized response is optimal in the local model [25]). *For every* nonconstant *counting query* $q : \mathcal{X} \to \{0, 1\}$, *and* $n \in \mathbb{N}$, *and* $(1, 0)$*-differentially private* n*-party protocol* P *for approximating* q, *there is an input data set* $x \in \mathcal{X}^n$ *on which* P *has error* $\alpha = \Omega(1/\sqrt{n})$ *with high probability.*

Randomized Response vs Laplace

Accuracy for counting queries in the local model

Using RR

$$\left|q(D) - r\right| = \Omega(\frac{1}{\sqrt{n}})$$

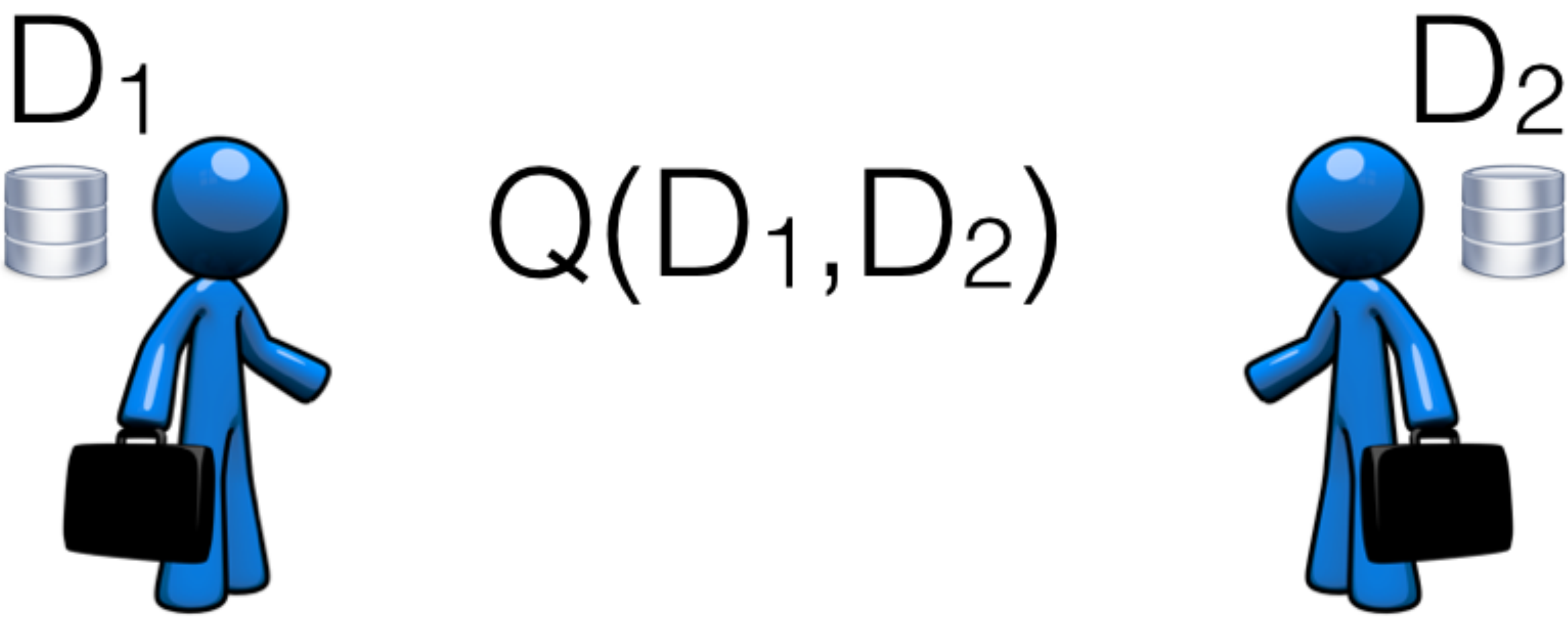
Accuracy for counting queries in the curator model

Using Laplace

$$\left|q(D) - r\right| \leq O\left(\frac{1}{n}\right)$$

Two party differential privacy

We now consider the case of two parties that want to compute a common statistics.
Each party has a database of size n/2.



Counting queries in the 2-party model

How can we compute efficiently a counting query q in the 2-party model?

Protocol:

-each party P_i computes $a_i=q(D_i)+Lap(2/\varepsilon n)$ and shares it,

-we collect the results and compute $a=(a_1+a_2)/2$

Accuracy for counting queries in the 2-parties model

$$\Big|q(D) - r\Big| \leq O\Big(\frac{1}{n}\Big)$$

Counting queries

Accuracy for counting queries in the local model

Using RR

$$\left|q(D) - r\right| = \Omega(\frac{1}{\sqrt{n}})$$

Accuracy for counting queries in the 2-party model

Using Laplace

$$\left|q(D) - r\right| \leq O\left(\frac{1}{n}\right)$$

Accuracy for counting queries in the curator model

Using Laplace

$$\left|q(D) - r\right| \leq O\left(\frac{1}{n}\right)$$

How about other statistics?

Let's consider the normalized Inner Product:

$$\mathrm{IP} : \{0,1\}^{n/2} \times \{0,1\}^{n/2} \to [0,1]$$

$$\mathrm{IP}(D_1, D_2) = \frac{2\langle D_1, D_2 \rangle}{n}$$

In the curator model we can compute $r = \mathrm{IP}(D_1, D_2) + \mathrm{Lap}(2/\varepsilon n)$ and so we have:

$$\left| \mathrm{IP}(D_1, D_2) - r \right| \leq O\left(\frac{1}{n}\right)$$

How can we compute IP in the 2-parties model?

Inner product

Theorem 9.4 (2-party DP protocols for inner product [80, 77]).

1. *There is a two-party differentially private protocol that estimates* IP *to within error* $O(1/\varepsilon \cdot \sqrt{n})$ *with high probability, and*

2. *Every two party* $(1, 0)$*-differentially private protocol for* IP *incurs error* $\tilde{\Omega}(1/\sqrt{n})$ *with high probability on some dataset.*

Proof sketch. For the upper bound, we again use randomized response:

1. On input $x \in \{0,1\}^{n/2}$, Alice uses randomized response to send a noisy version $\hat{x}$ of x to Bob.

2. Upon receiving $\hat{x}$ and his input $y \in \{0,1\}^{n/2}$, Bob computes

$$z = \frac{2}{n} \sum_{i=1}^{n/2} \frac{y_i}{\varepsilon} \cdot \left(\hat{x}_i - \frac{(1-\varepsilon)}{2} \right),$$

   which will approximate $\mathrm{IP}(x, y)$ to within $O(1/\varepsilon\sqrt{n})$.

3. Bob sends the output $z + \mathrm{Lap}(O(1/\varepsilon^2 n))$ to Alice, where this Laplace noise is to protect the privacy of y, since z has global sensitivity $O(1/\varepsilon n)$ as a function of y.

Inner product

Theorem 9.4 (2-party DP protocols for inner product [80, 77]).

1. *There is a two-party differentially private protocol that estimates* IP *to within error* $O(1/\varepsilon \cdot \sqrt{n})$ *with high probability, and*

2. *Every two party* $(1, 0)$*-differentially private protocol for* IP *incurs error* $\tilde{\Omega}(1/\sqrt{n})$ *with high probability on some dataset.*

For the lower bound, we follow the same outline as Theorem 9.3. Let $X = (X_1, \ldots, X_{n/2})$ and $Y = (Y_1, \ldots, Y_{n/2})$ each be uniformly distributed over $\{0,1\}^{n/2}$ and independent of each other. Then, conditioned on a transcript t of an $(\varepsilon, 0)$-differentially private protocol, we have:

1. X and Y are independent, and

2. For every $i \in [n/2]$, $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$,

$$\Pr[X_i = 1 | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, \ldots, X_n = x_n] \in (1/4, 3/4),$$

   and similarly for Y.

Item 2 again follows from differential privacy and Bayes' Rule. (Consider the two neighboring datasets $(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n)$ and $(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n)$.)

Inner Product

Accuracy for inner product in the local model

Using RR

$$\left| q(D) - r \right| = \Omega(\frac{1}{\sqrt{n}})$$

Accuracy for inner product in the 2-party model

Using RR

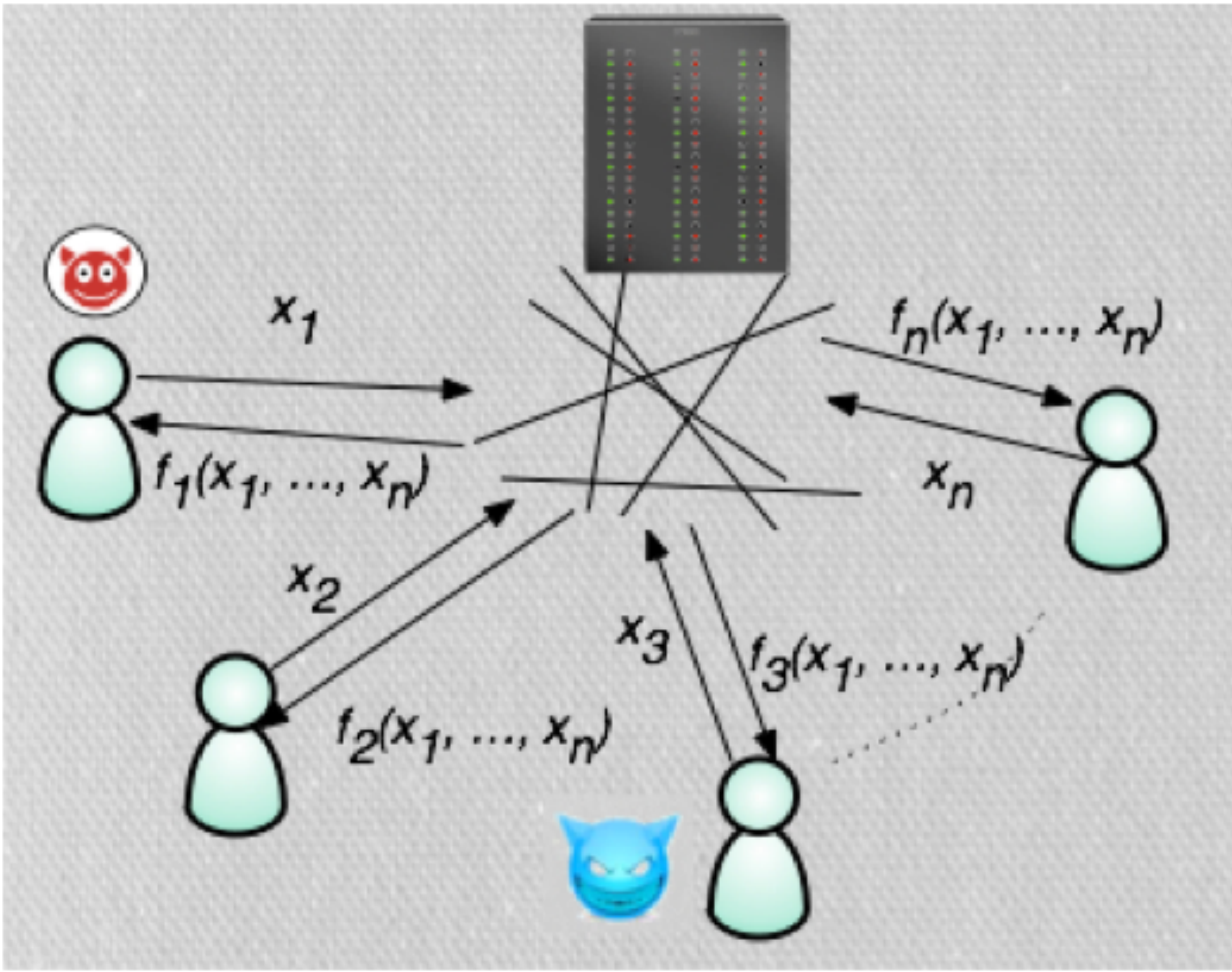
$$\left| q(D) - r \right| = \Omega(\frac{1}{\sqrt{n}})$$

Accuracy for inner product in the curator model

Using Laplace

$$\left| q(D) - r \right| \leq O\left(\frac{1}{n}\right)$$

Multiparty differential privacy



DP in IOS

https://images.apple.com/au/privacy/docs/Differential_Privacy_Overview.pdf