CS208: Applied Privacy for Data Science
The Local Model: Foundations

James Honaker, Priyanka Nanayakkara, Salil Vadhan

School of Engineering & Applied Sciences
Harvard University

April 6, 2025

Housekeeping

- Detailed project descriptions due this Friday!
  - You can still change your topic, eg based on the feedback we gave.
  - Come to OH to discuss!

- No pset due this week, hw8b due Fri 4/18.

- Other project deadlines:
  - Full project paper: Wed 4/30
  - Revision of paper: Thu 5/8
  - Poster session: Thu 5/8, 9am-12pm in the SEC.

Class-wide exercise

- Privately:
  - Write down your preference: vanilla (1) or chocolate (0)
  - Choose a random number from 1-4 using Google, www.random.org, or by tossing a coin twice.

- Class Poll: Salil will ask everyone to report their preference
  - If your random number is 1,2,3: report truthfully
  - If your random number is 4: report falsely

Group Exercise

1. Use the reported counts for vanilla and chocolate to compute an unbiased estimator $\hat{\mu}$ of the fraction of people in the class who prefer vanilla.

2. What is the standard deviation of your estimator?

3. For what ε is this method ε-DP? (Consider the release to be collection of everyone's "noisy" reports.)

Individual Survey

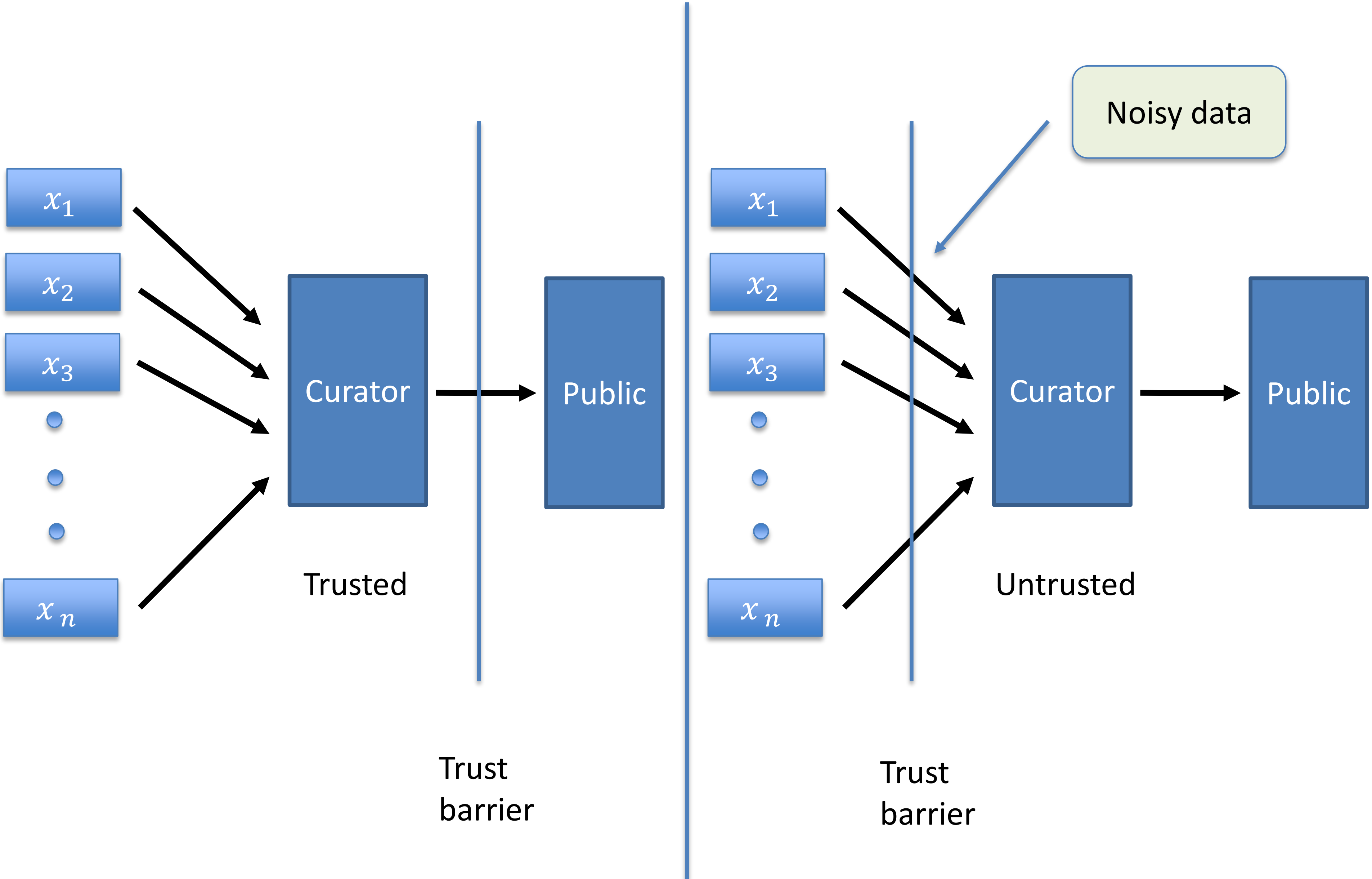
Compare the method we just saw for doing a DP count to a standard noise-addition mechanism (e.g. the Laplace mechanism).

1. What is an advantage of the method we just used?

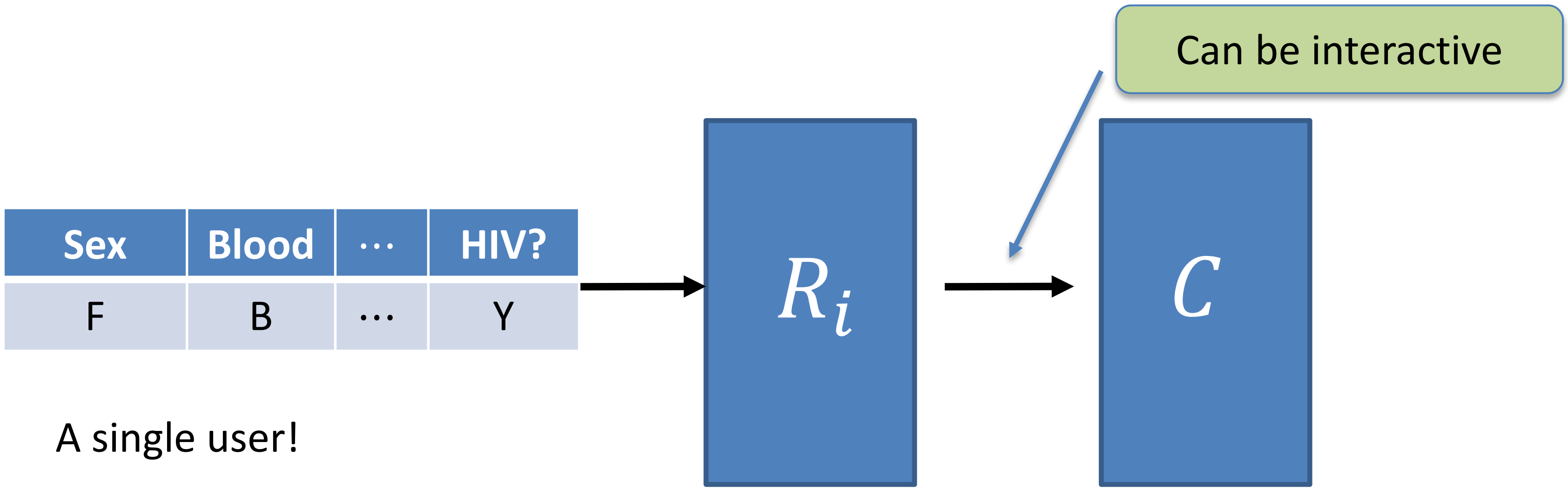
2. What is a disadvantage of the method we just used?

In either case, if you don’t think there’s an advantage or disadvantage, give your intuition.

Central Model vs Local Model



Local Differential Privacy



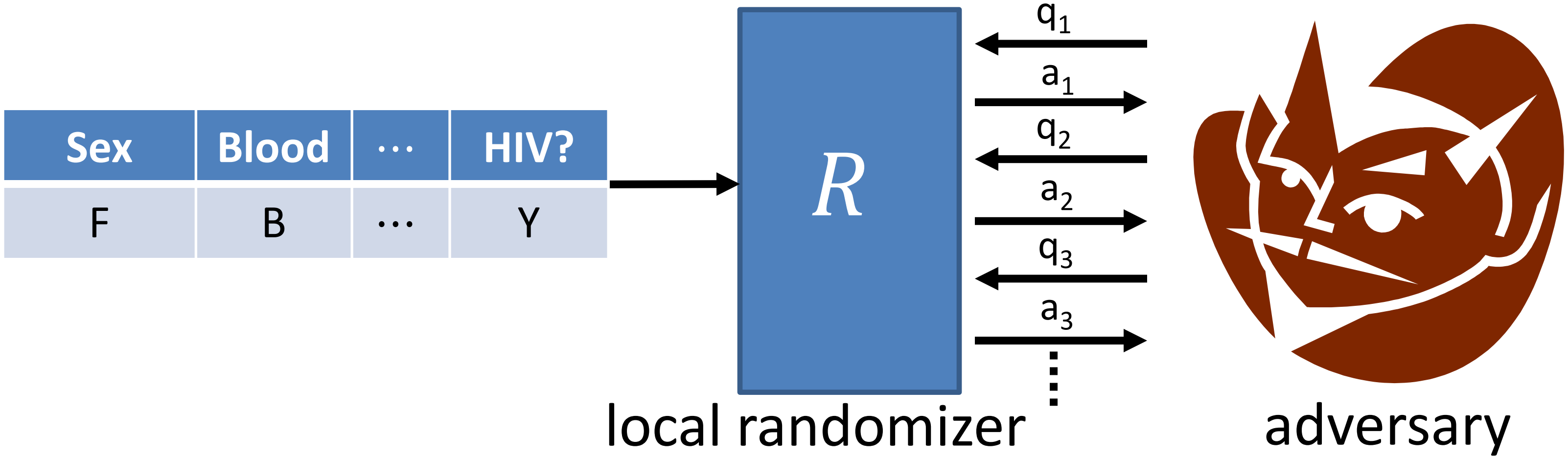
local randomizer

$R: \mathcal{X} \to \mathcal{Y}$ is (ε, δ)-locally differentially private (LDP) if for all $x, x' \in \mathcal{X}, S \subseteq \mathcal{Y}$

$$\Pr[R(x) \in S] \leq e^{\varepsilon} \cdot \Pr[R(x') \in S] + \delta$$

That is, a protocol is ε-LDP if each party's local randomizer R_i is an ε-DP mechanism for *1-row databases*.

Interactive Local DP



Require: for all x, x', all adversarial strategies A

$$\underbrace{\mathrm{View}_A\big(A \leftrightarrow M(x)\big)}\approx_\varepsilon \underbrace{\mathrm{View}_A\big(A \leftrightarrow M(x')\big)}$$

Everything A sees (its internal randomness & query answers)

Equivalently: $\forall\, A \;\; \Pr[A \text{ outputs "In" after interacting w}/M(x)]$
$\leq e^{\varepsilon} \cdot \Pr[A \text{ outputs "In" after interacting w}/M(x')]$

Randomized Response

[Warner'65]

For $x_i \in \{0,1\}$, $\mathrm{RR}_\varepsilon(x_i) = \begin{cases} x_i & \text{w.p. } \frac{e^\varepsilon}{1+e^\varepsilon} \\ 1 - x_i & \text{w.p. } \frac{1}{1+e^\varepsilon} \end{cases}$

Theorem: RR_ε is ε-LDP

Unbiased estimator of the mean μ given $y_i = \mathrm{RR}_\varepsilon(x_i)$ for $i = 1, \ldots, n$:

$$\hat{\mu} =$$

Standard deviation: $O\left(\frac{1}{\varepsilon\sqrt{n}}\right)$ for $\varepsilon \leq 1$.

Randomized Response

RR gives an ε-locally DP protocol that

- Estimates "statistical queries" (means/avgs) to $\pm O\left(\frac{1}{\varepsilon\sqrt{n}}\right)$.
  - Q: how to use RR for fractional-valued functions?
- Estimates count/sum of a bounded function to $\pm O\left(\frac{\sqrt{n}}{\varepsilon}\right)$.
- Worse than centralized DP by a factor of $\sqrt{n}$, but still useful.
- Fact: The above privacy-accuracy tradeoff is the best possible for ε-local DP.

Local DP Histograms

$x_{1,} \ldots, x_n \in [D]$ (D bins). Use a 1-hot encoding:

x_i = 0 0 0 … 1 0 … 0 …. 0 Length D

$\mathrm{RR}_{\varepsilon/2}$ on every coordinate

y_i = 1 0 1 … 1 0 … 0…. 1

$$\hat{h} =$$

Local DP Histograms

- Expected error on each bin is $\pm O\left(\frac{\sqrt{n}}{\varepsilon}\right)$.
- Expected max error over all D bins is $\pm O\left(\frac{\sqrt{n \cdot \log D}}{\varepsilon}\right)$.

- We need to communicate D bits from each user.
There are protocols using some sophisticated algorithmic ideas to get communication complexity sublinear in D.

Local vs. Centralized DP

Central Model

- Central curator collects the data from all users, then performs privatization
- Requires the users to trust the curator with their private data
- Most DP algorithms are in this model

Local Model

- Each user privatizes their own data then sends it to a central curator
- Requires less trust from users
- Worse accuracy

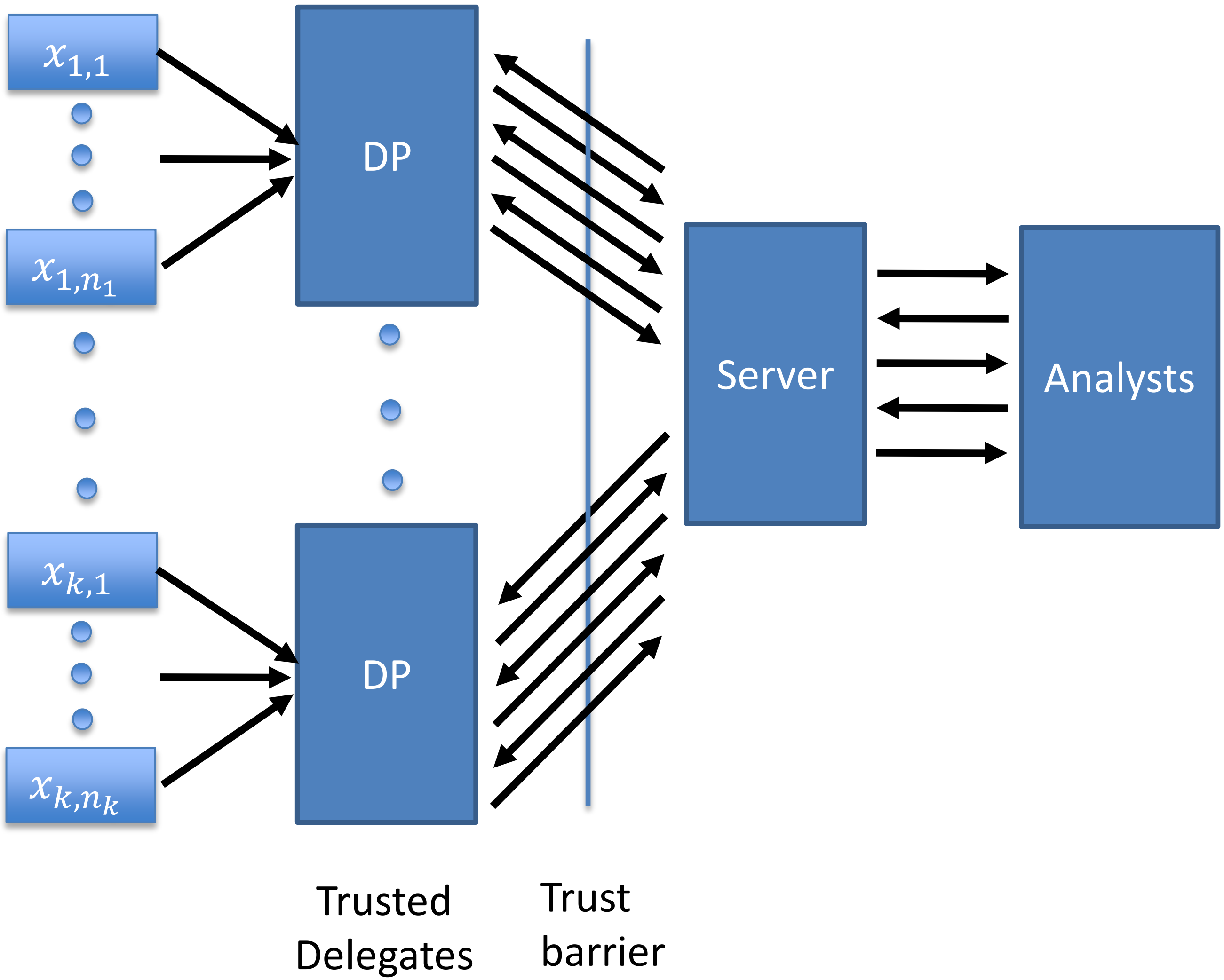
Slide based on one from Brendan Avent's presentation

Local vs. Centralized DP

- Local DP protocols provably have lower accuracy for counts/averages than centralized DP protocols.
  - $\Theta(1/\varepsilon\sqrt{n})$ error vs. $\Theta(1/\varepsilon n)$.
  - Successful deployments have very large n (Google, Apple).

- Next class: Gap can be closed by relaxing adversarial model (e.g. anonymous participants, computationally bounded adversaries) and using crypto/infrastructure (secure MPC, mix-nets).

Federated DP



Comparing the Models

- Federated DP with k delegates, $n = n_1 + \cdots + n_k$
  - "horizontally partitioned" data
  - $k = 1$: central DP
  - $k = n$: local DP

- Error for sum of bounded values (like in DP-SGD) = $\Theta\left(\frac{\sqrt{k}}{\varepsilon}\right)$.
  - Interpolates between local & central model

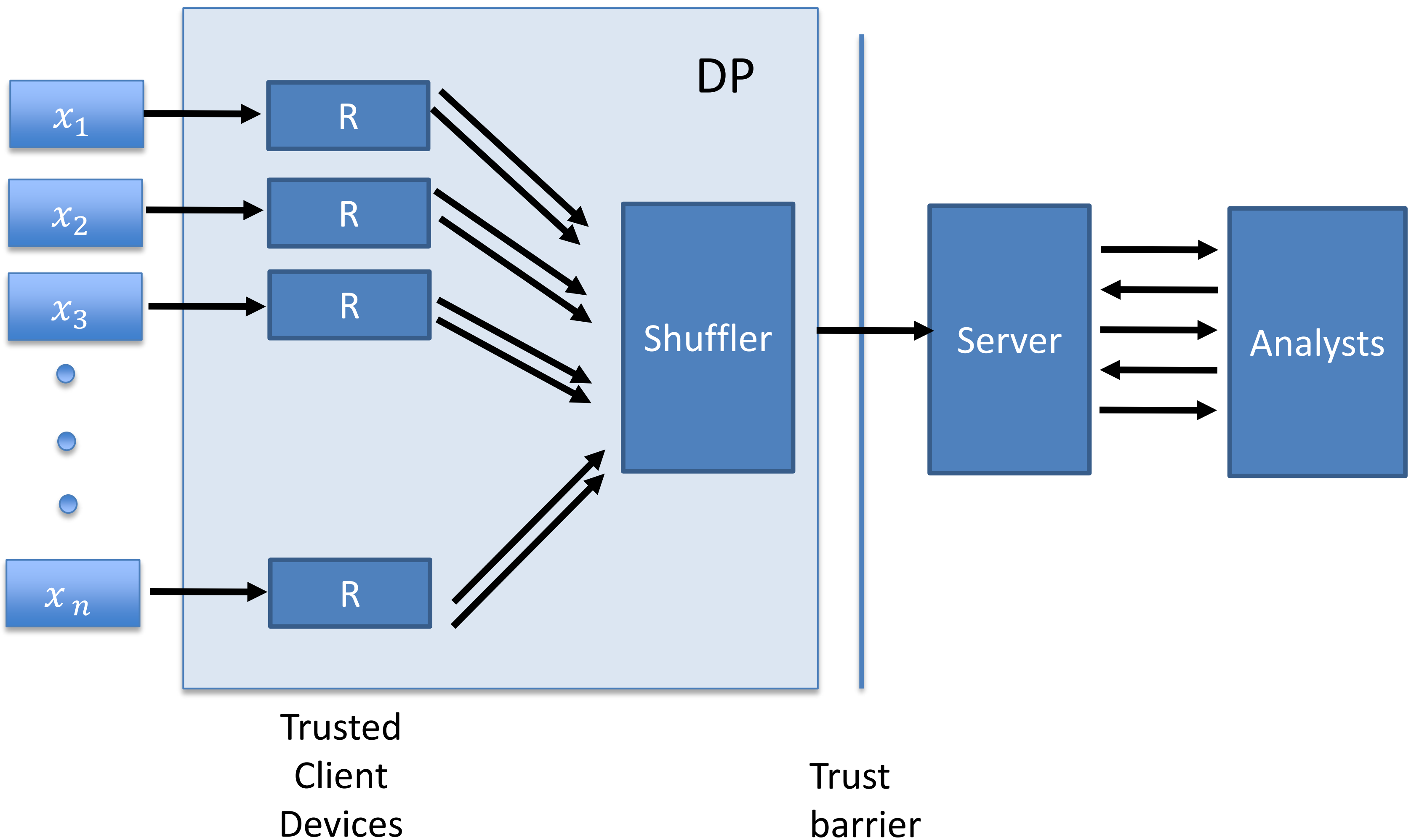
- Error for set intersection when $k = 2$: $\Theta\left(\frac{\sqrt{n}}{\varepsilon}\right)$
  - No better than local model!

Other Models

- Can we get the “best of both worlds”?
  - Privacy protections like the local model
  - Accuracy like the central model

- Two approaches
  - The shuffle model
  - Using cryptography (secure multiparty computation)

Shuffle DP


x_1
x_2
x_3
x_n
R
R
R
R
DP
Shuffler
Server
Analysts
Trusted Client Devices
Trust barrier