Differentially Private Multi-Party Computation

Peter Kairouz

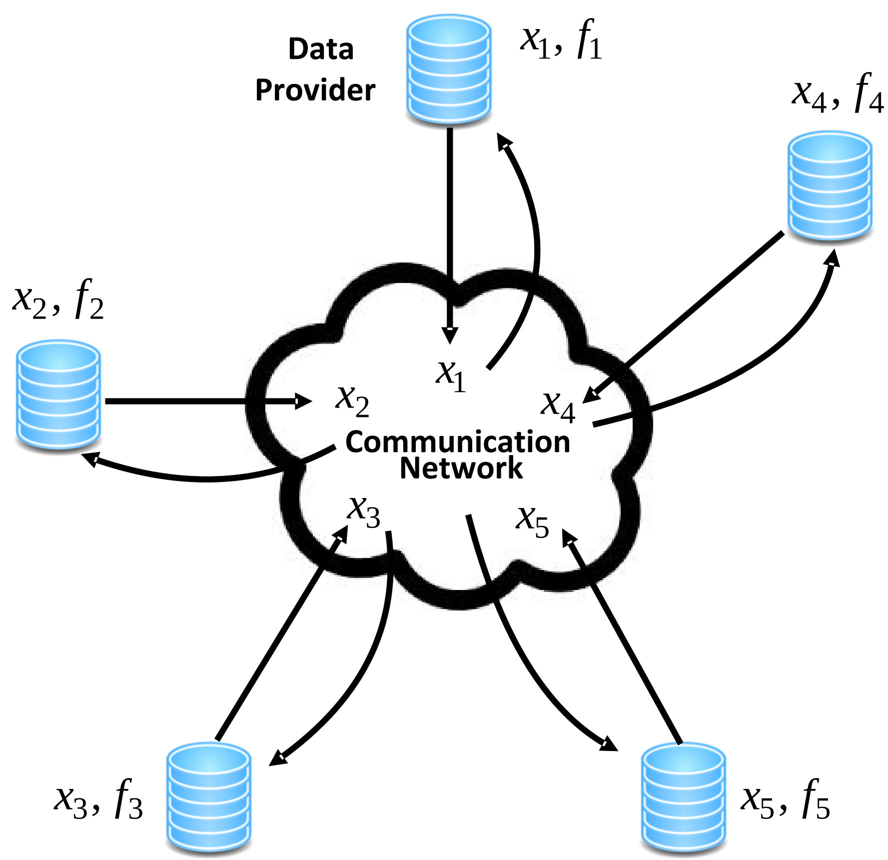
Sewoong Oh

Pramod Viswanath

ILLINOIS

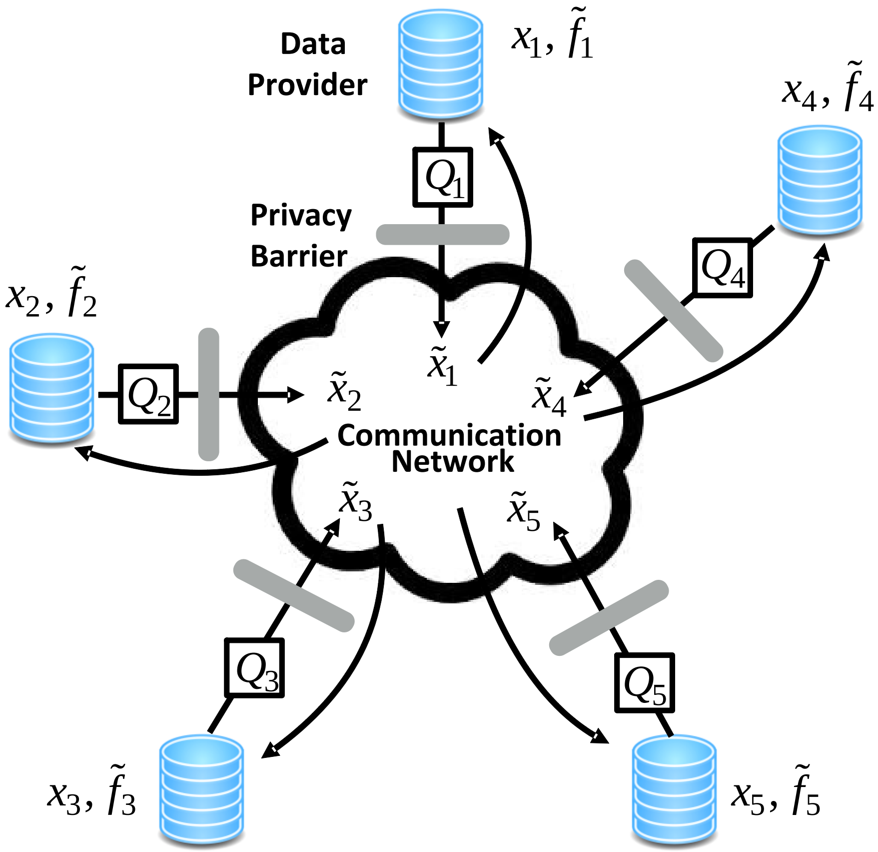
illinois.edu

Multi-party computation



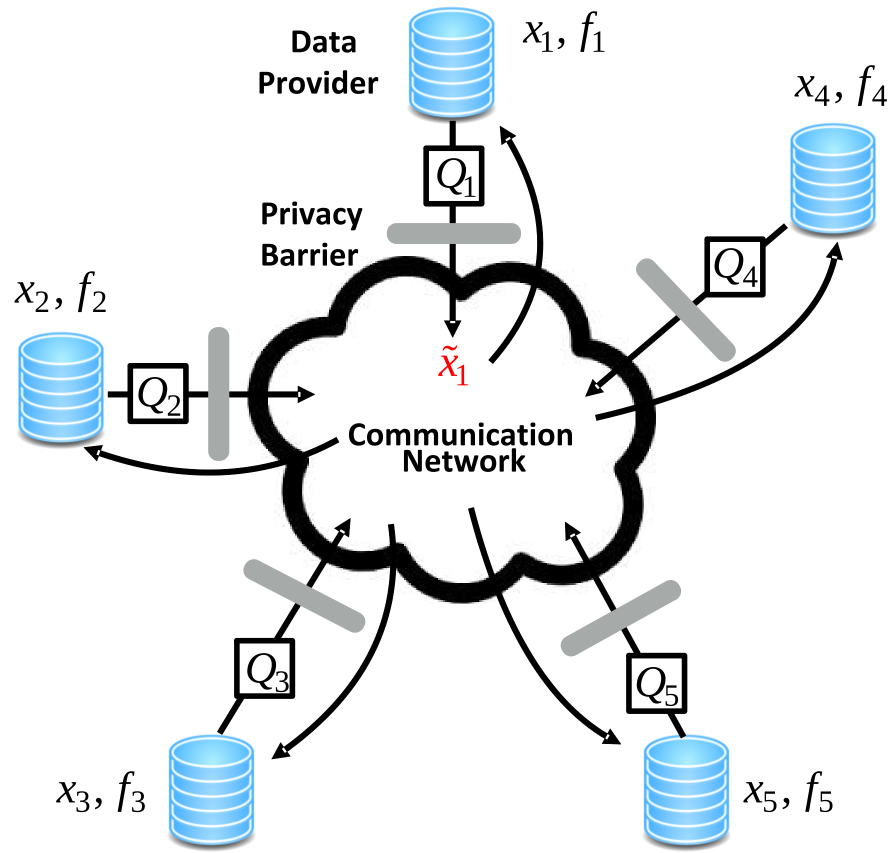
- important setting in distributed systems and cloud computing

Private multi-party computation

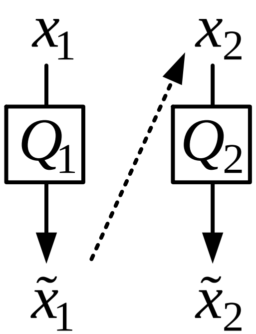


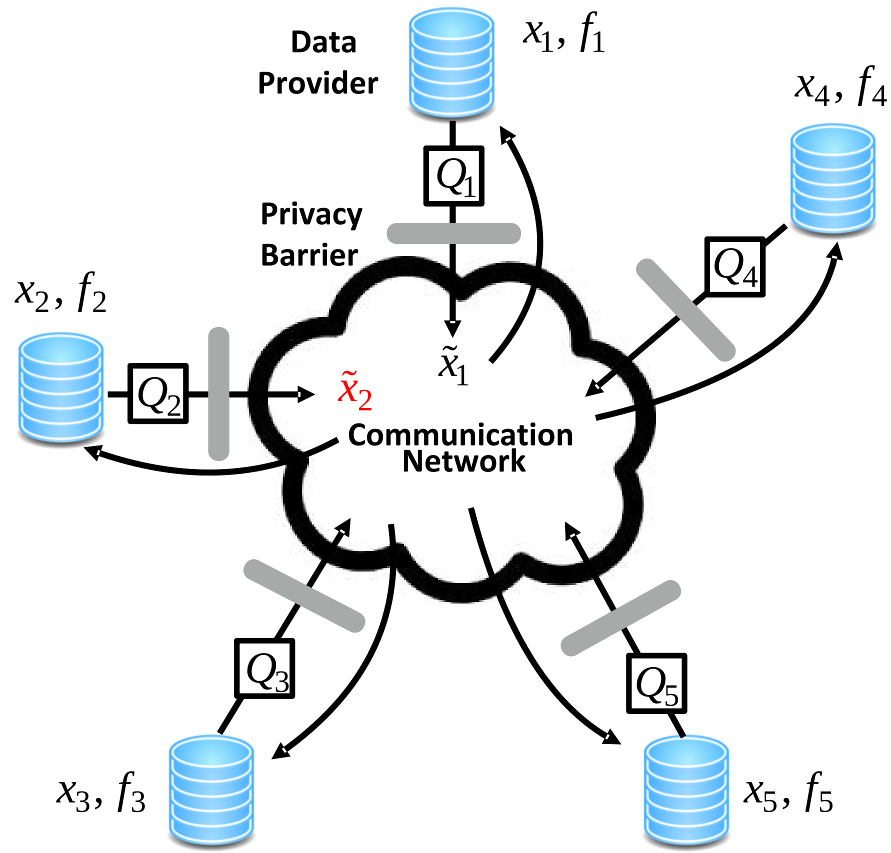
- each party shares a noisy version of its data

Interactive mechanisms

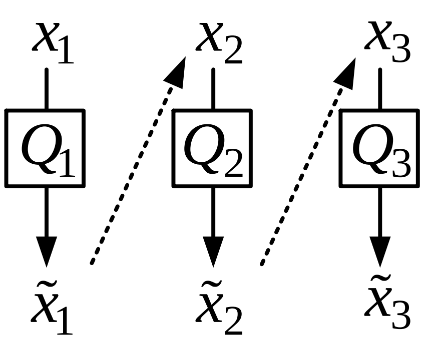


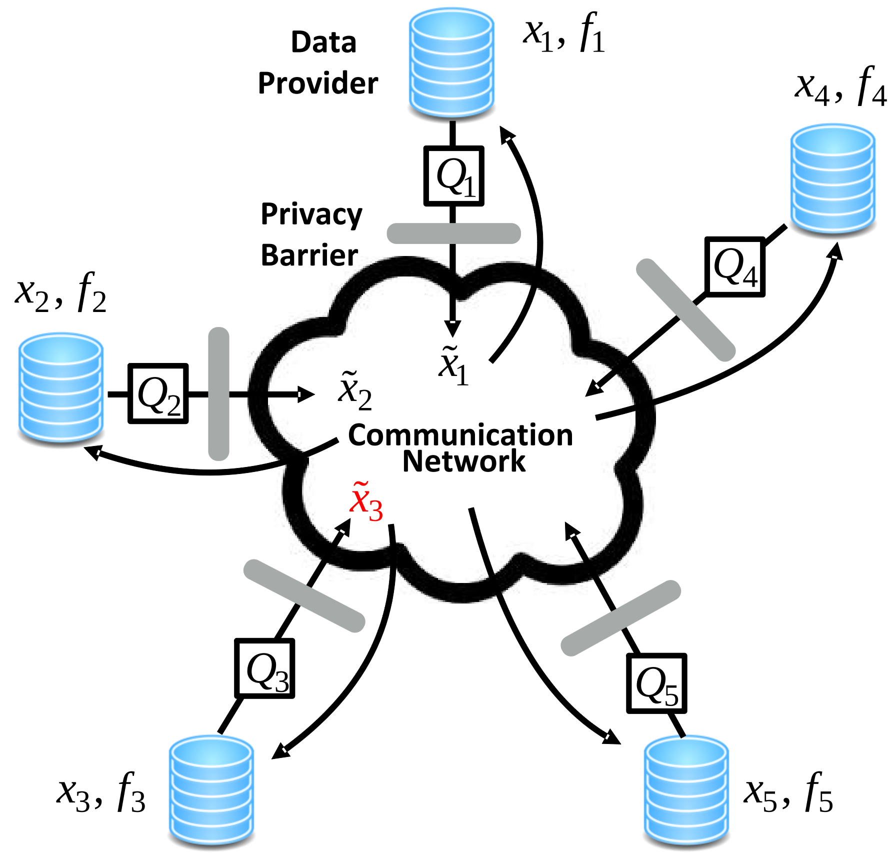
Interactive mechanisms



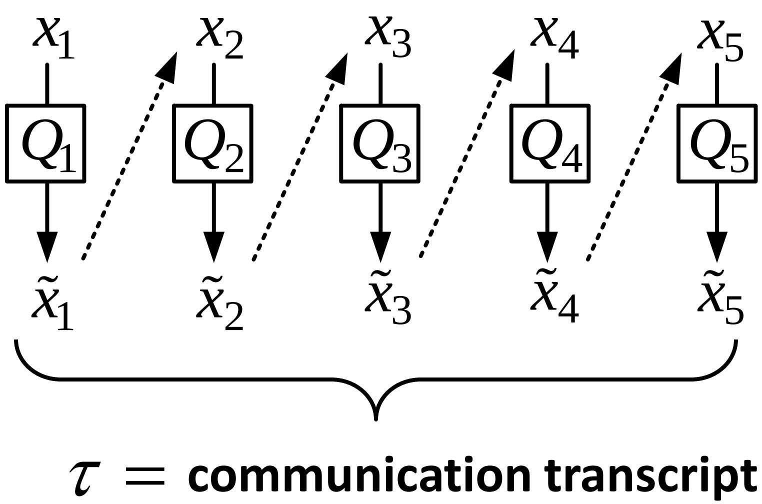


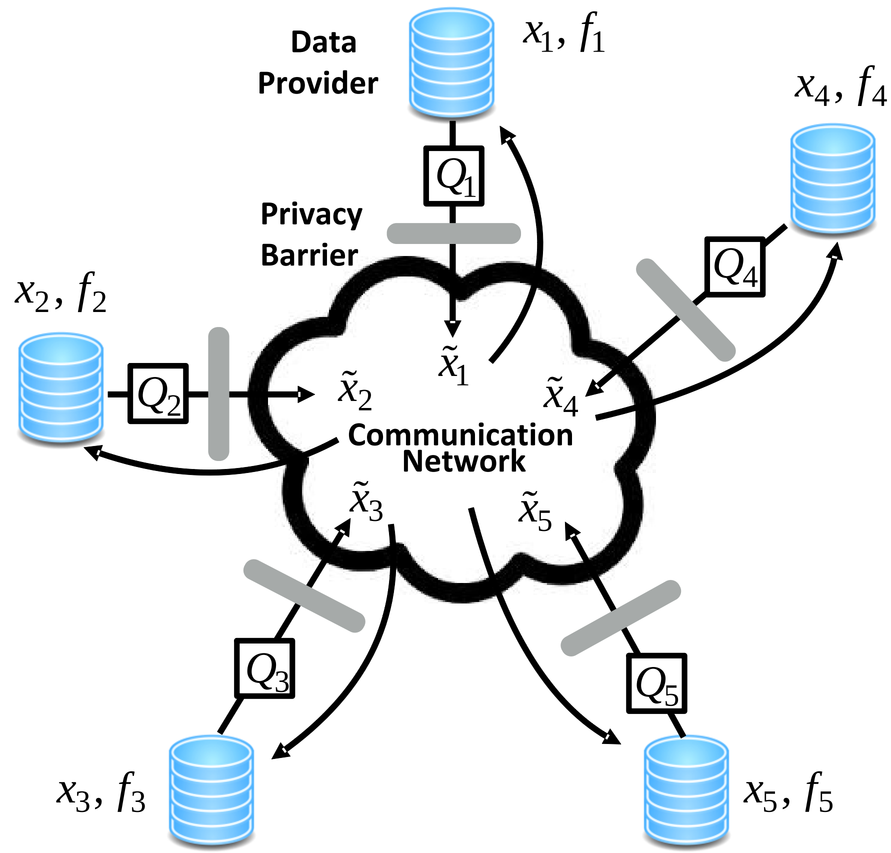
Interactive mechanisms



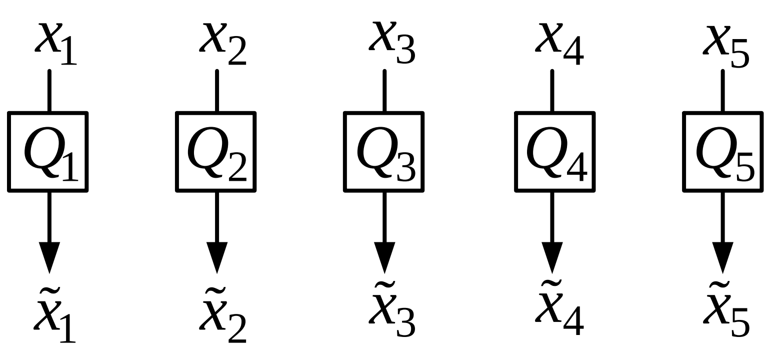


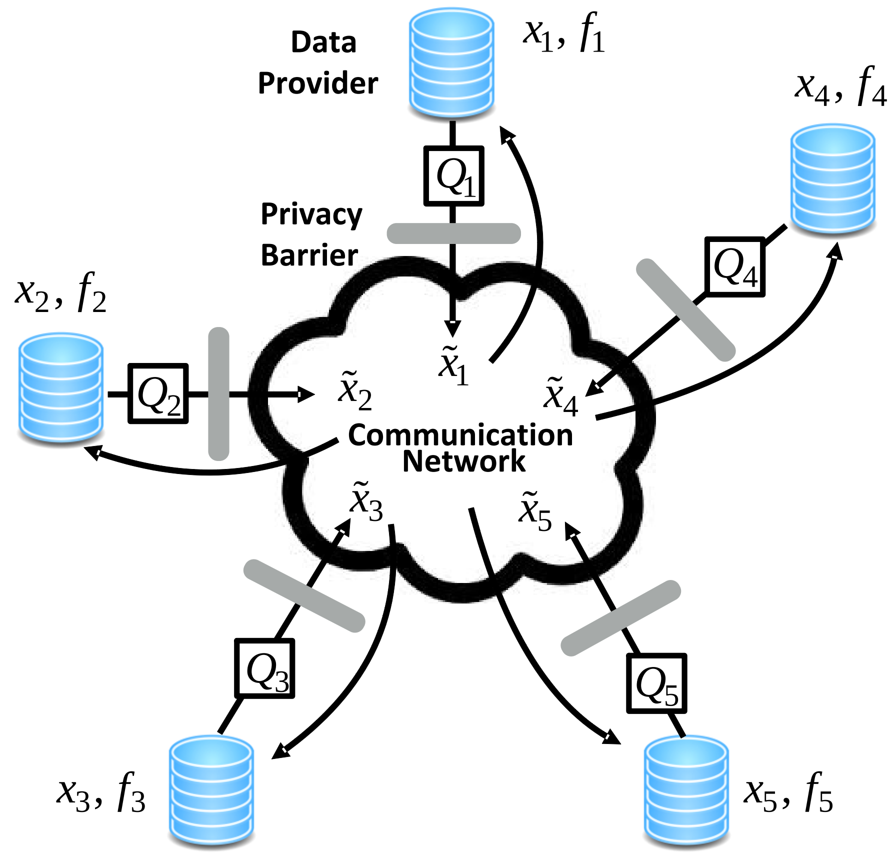
Interactive mechanisms



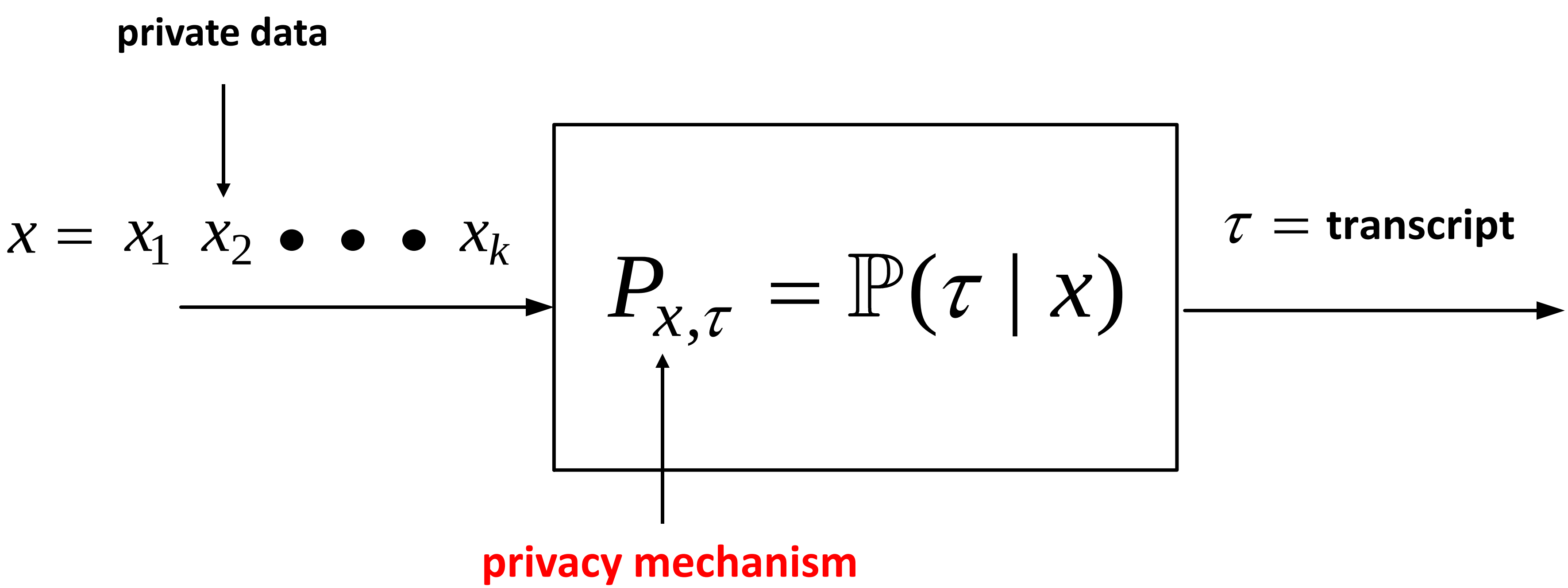


Non-interactive mechanisms





General representation



Multi-party differential privacy

$x = x_1 \; x_2 \bullet \bullet \bullet \; x_k$

$$P_{x,\tau} = \mathbb{P}(\tau \mid x)$$

$\tau =$ **transcript**

$$e^{-\varepsilon_i} \leq \frac{\mathbb{P}(\tau \mid x_i = 0, x_{-i})}{\mathbb{P}(\tau \mid x_i = 1, x_{-i})} \leq e^{\varepsilon_i}$$

$x_{-i} = (x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_k)$

- bounded likelihood even when all parties but one collude

Multi-party differential privacy

$x = x_1 \; x_2 \bullet \bullet \bullet \; x_k$

$$P_{x,\tau} = \mathbb{P}(\tau \mid x)$$

$\tau =$ **transcript**

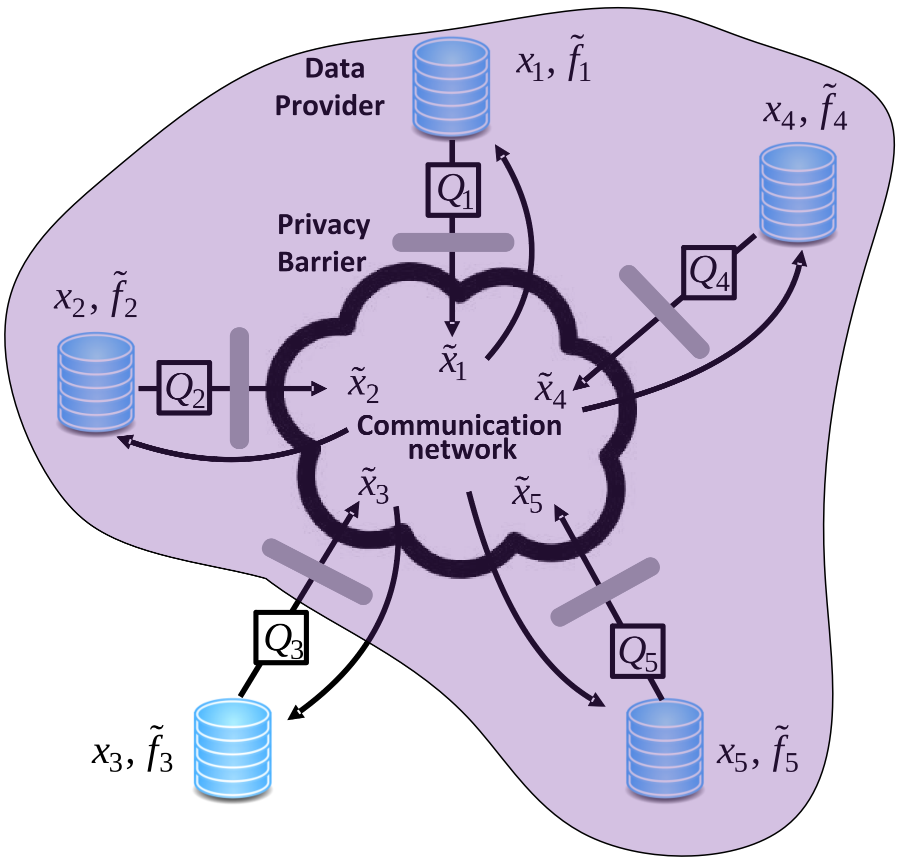
$$e^{-\varepsilon_i} \le \frac{\mathbb{P}(\tau \mid x_i = 0, x_{-i})}{\mathbb{P}(\tau \mid x_i = 1, x_{-i})} \le e^{\varepsilon_i}$$

ε_i **controls the level of privacy**

large ε_i, low privacy

small ε_i, high privacy

Can’t say much even if…



- all parties but one collude to figure out a party’s data

Approximate differential privacy

$x = x_1 \; x_2 \bullet \bullet \bullet \; x_k$

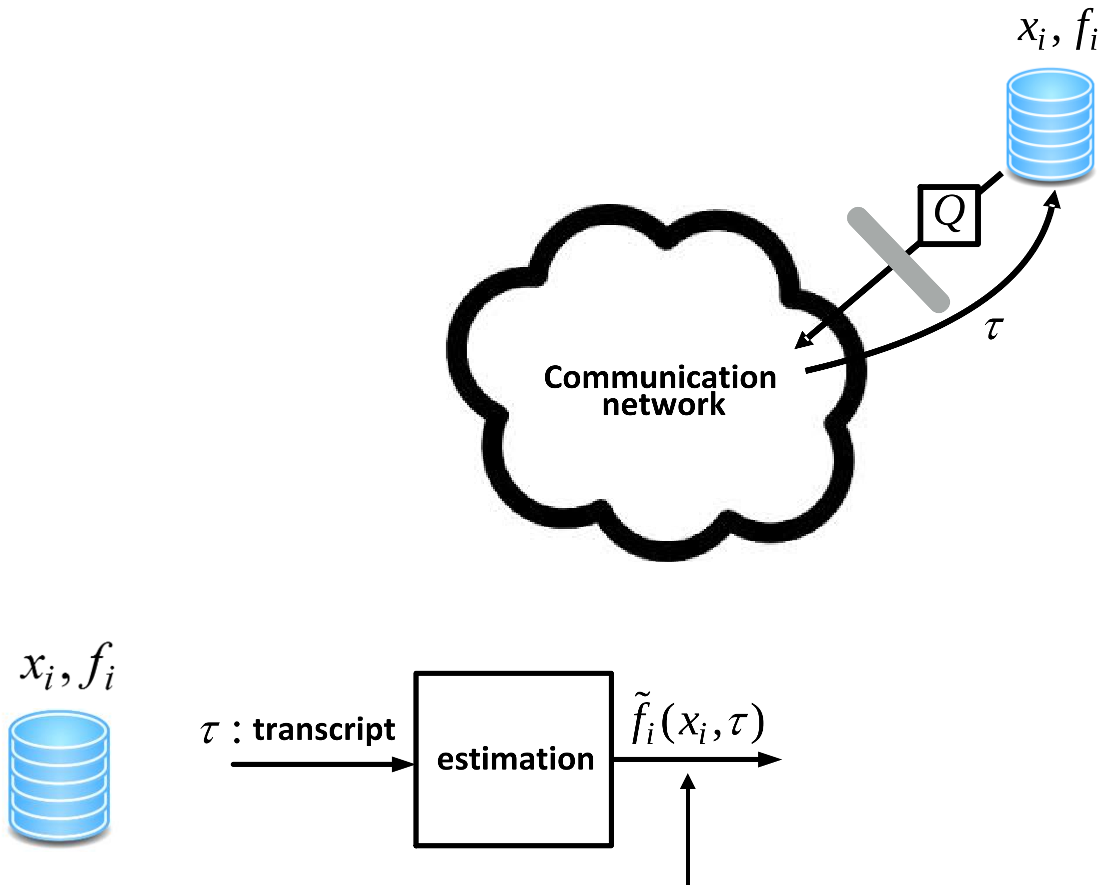
$$P_{x,\tau} = \mathbb{P}(\tau \mid x)$$

$\tau =$ **transcript**

[redacted] δ_i

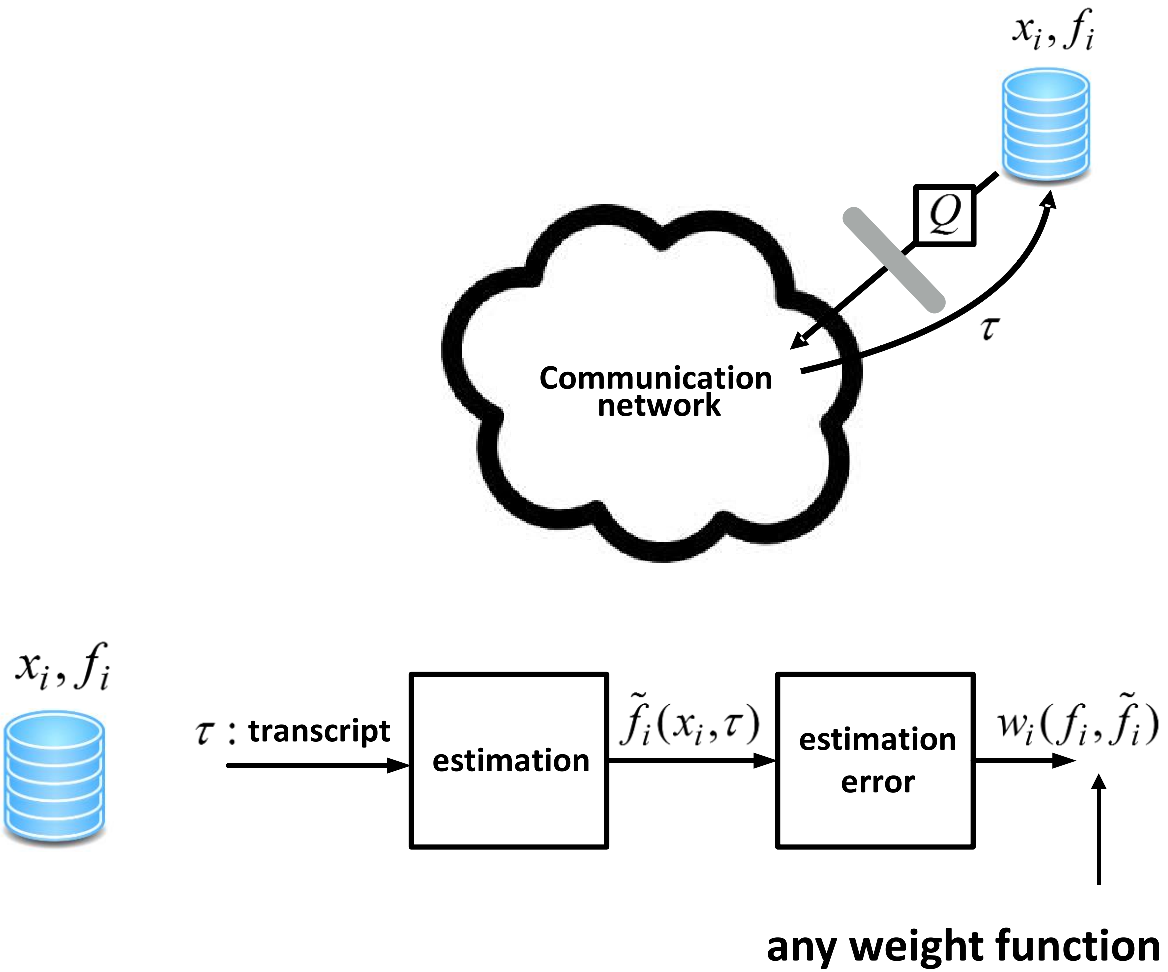
provides some slack

Function estimation

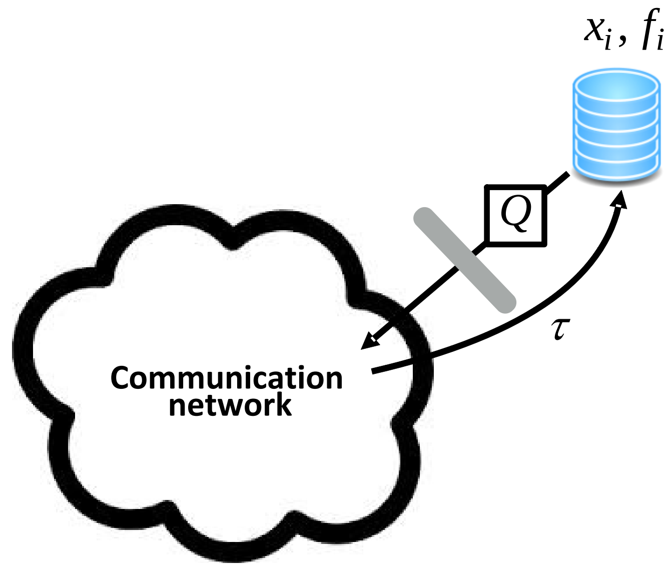


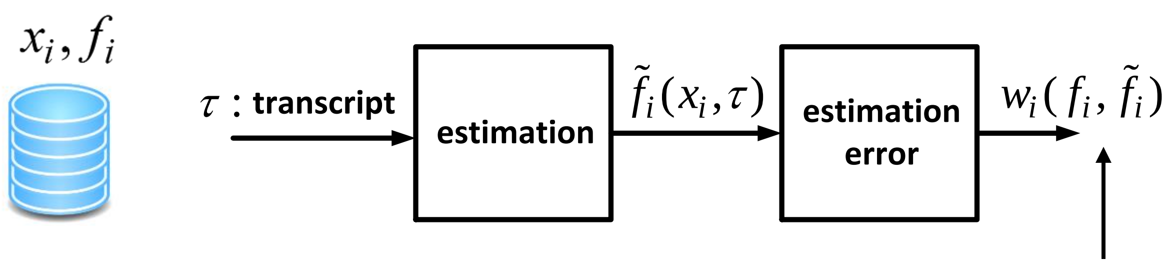
any estimation rule (potentially randomized)

Function estimation



Function estimation





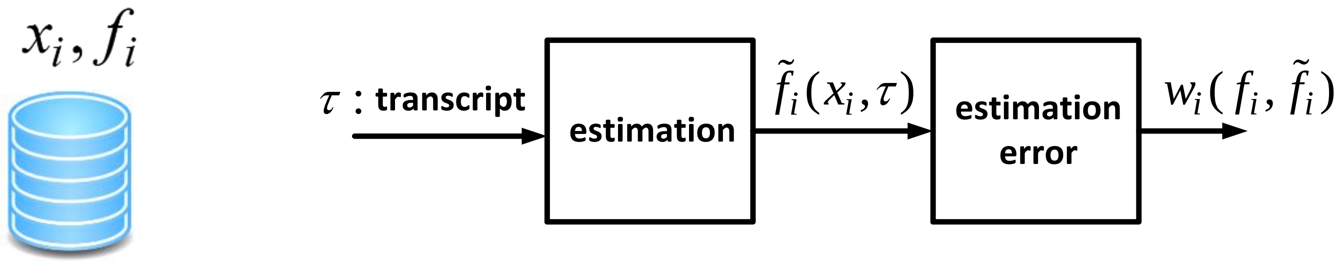
any weight function

examples:

$$w(f, \tilde{f}) = 1_{(f = \tilde{f})}$$

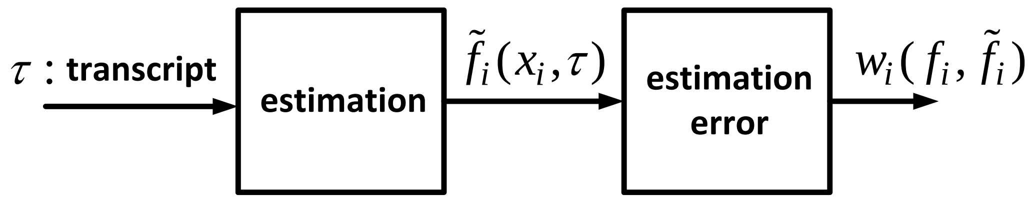
$$w(f, \tilde{f}) = | f - \tilde{f} |$$

Utility: average accuracy



Utility: average accuracy

x_i, f_i



$$\mathrm{ACC}_{\mathrm{ave}} \equiv \underbrace{\frac{1}{2^k} \sum_{x \in \{0,1\}^k}}_{\text{average over all possible inputs}} \mathbb{E}_{\hat{f}_i, P_{x,\tau}} [w_i(f_i(x), \tilde{f}_i(\tau, x_i))]$$

Privacy-utility tradeoff

$$
\begin{aligned}
\underset{P,\tilde{f}_i}{\text{maximize}} \quad & \mathrm{ACC}_{\mathrm{ave}}(P, w_i, f_i, \tilde{f}_i), \\
\text{subject to} \quad & P \text{ and } \tilde{f}_i \text{ are row-stochastic matrices} \\
& P_{(x_i, x_{-i}),\tau} \leq e^{\varepsilon_i} P_{(x_i', x_{-i}),\tau} + \delta_i \quad \forall i, x_i, x_i', x_{-i}, \tau
\end{aligned}
$$

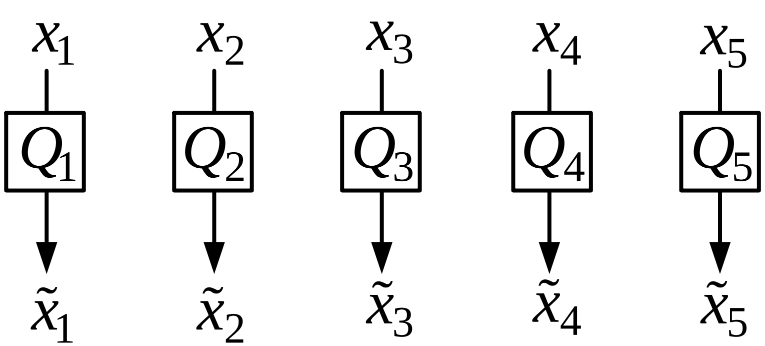
Privacy-utility tradeoff

$$
\begin{aligned}
\underset{P, \tilde{f}_i}{\text{maximize}} \quad & \mathrm{ACC}_{\mathrm{ave}}(P, w_i, f_i, \tilde{f}_i), \\
\text{subject to} \quad & P \text{ and } \tilde{f}_i \text{ are row-stochastic matrices} \\
& P_{(x_i, x_{-i}),\tau} \leq e^{\varepsilon_i} P_{(x_i', x_{-i}),\tau} + \delta_i \quad \forall i, x_i, x_i', x_{-i}, \tau
\end{aligned}
$$

- **heterogeneous privacy levels** across users
- each party possesses **a single bit**
- the **functions can vary** from one party to the other
- the **weight functions can vary** from one party to the other
- **interactive & non-interactive** mechanisms

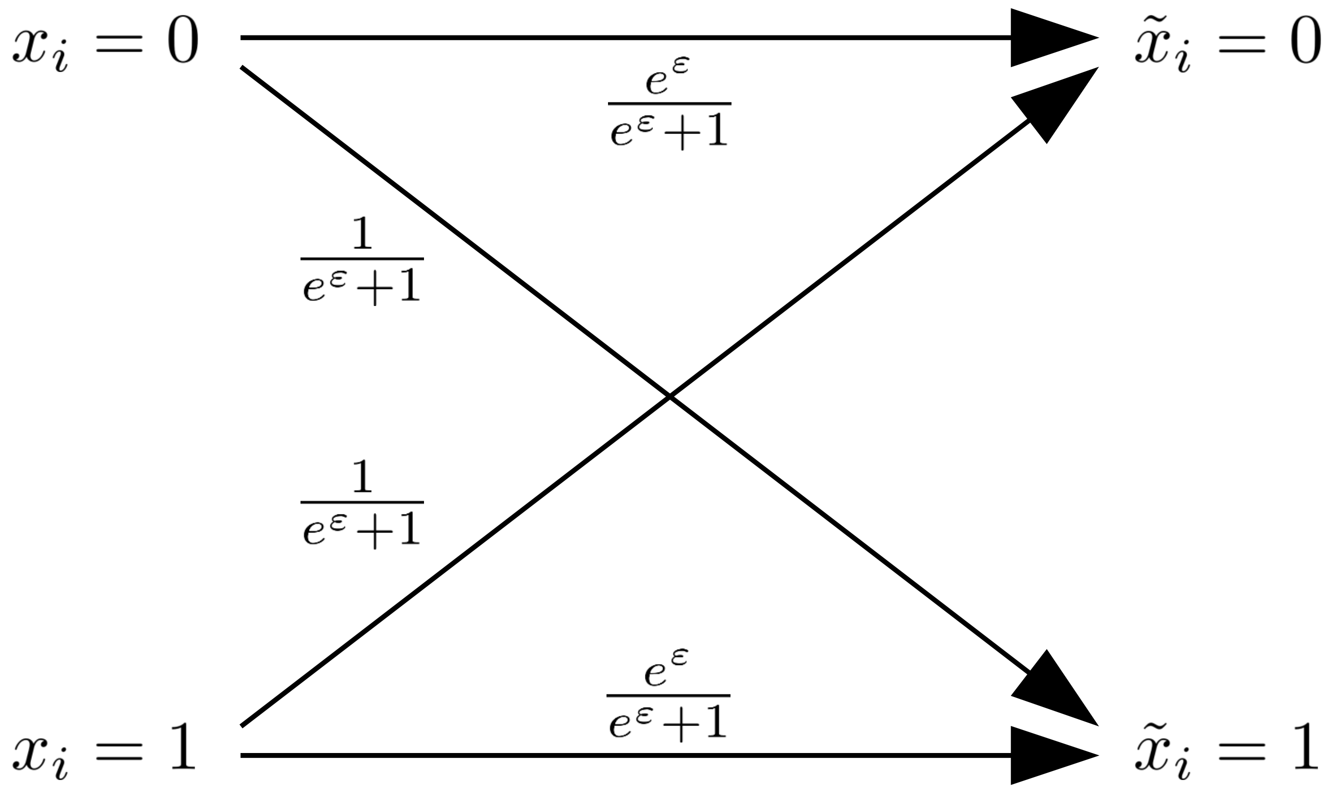
Main result: differential privacy

Non-interactive mechanisms are optimal



Main result: differential privacy

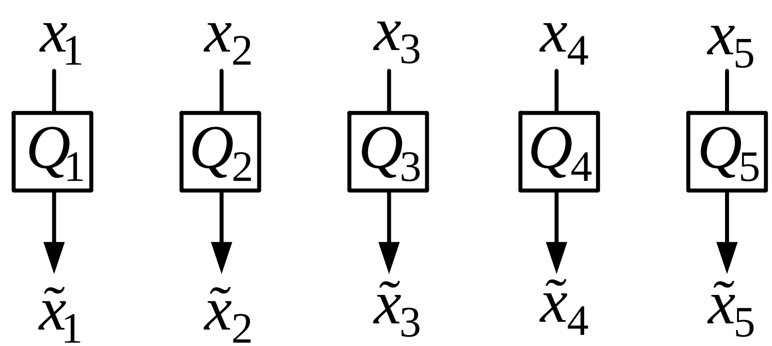
The randomized response is optimal!



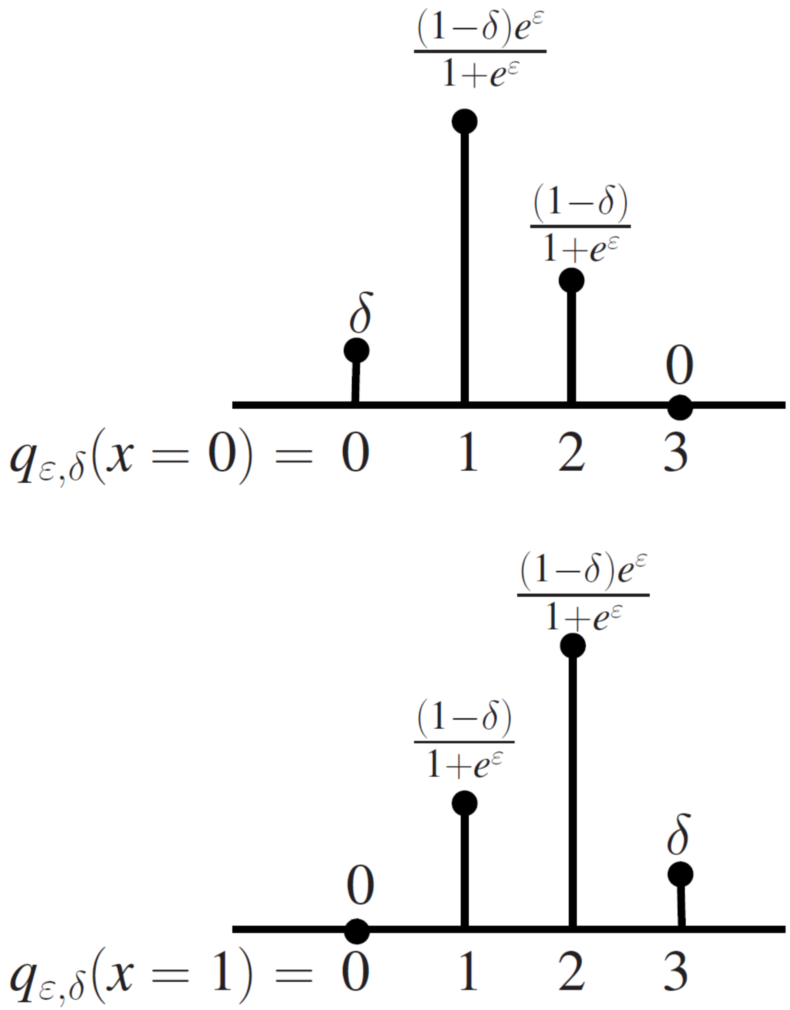
Approximate differential privacy?

Approximate differential privacy?

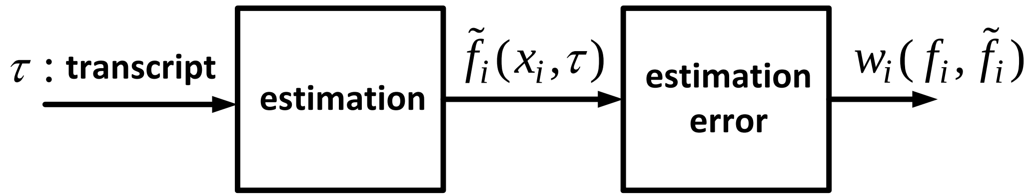
Non-interactive mechanisms are optimal



Approximate differential privacy



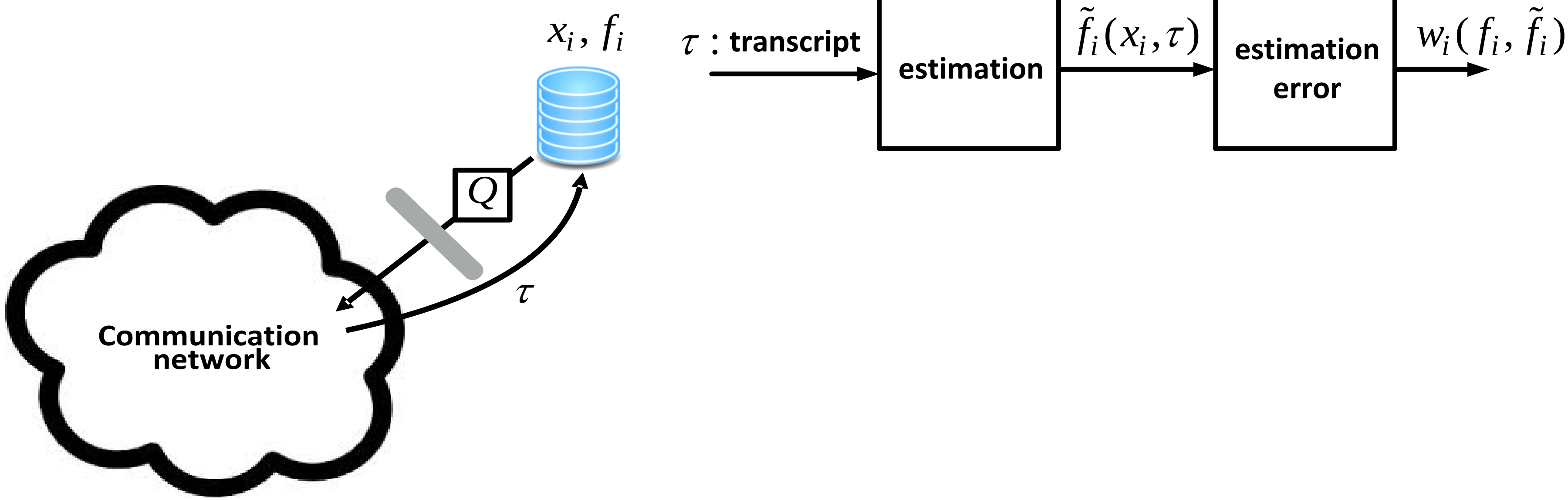
Optimal estimation rule



$$\mathrm{ACC}_{\mathrm{ave}} \equiv \underbrace{\frac{1}{2^k} \sum_{x \in \{0,1\}^k}}_{\textbf{average over all possible inputs}} \mathbb{E}_{\hat{f}_i, P_{x,\tau}} [w_i(f_i(x), \tilde{f}_i(\tau, x_i))]$$

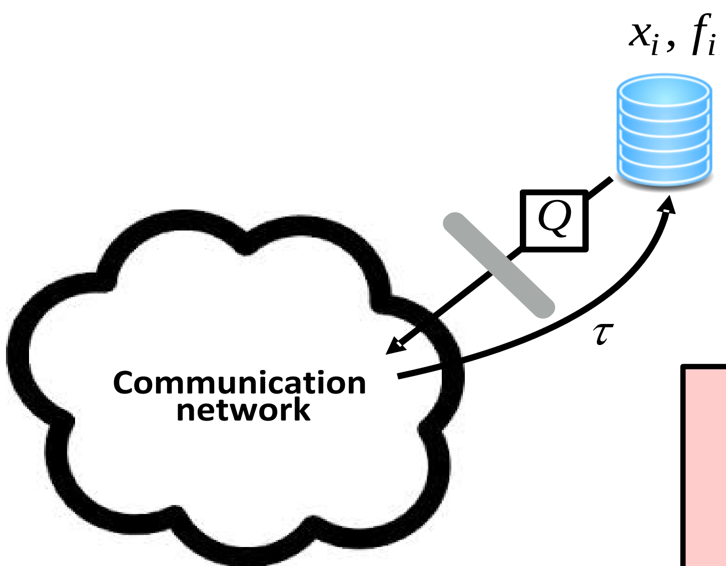
$$\boxed{\tilde{f}_{i,\mathrm{opt}}(\tau, x_i) = \arg\max_y \sum_{x_{-i} \in \{0,1\}^{k-1}} P_{x,\tau}\, w_i(f_i(x), y)}$$

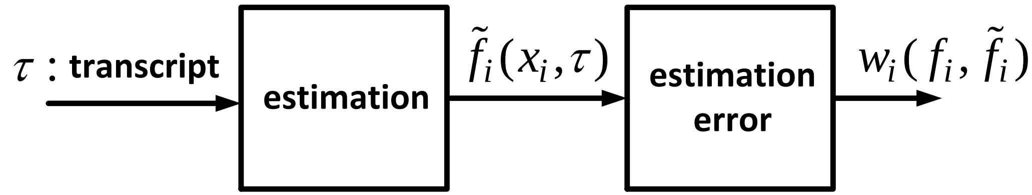
Worst case accuracy?



$$\mathrm{ACC}_{\mathrm{wc}} \equiv \underbrace{\min_{x \in \{0,1\}^k}}_{\text{worst case over all possible inputs}} \mathbb{E}_{\hat{f}_i, P_{x,\tau}} [w_i(f_i(x), \hat{f}_i(\tau, x_i))]$$

Worst case accuracy



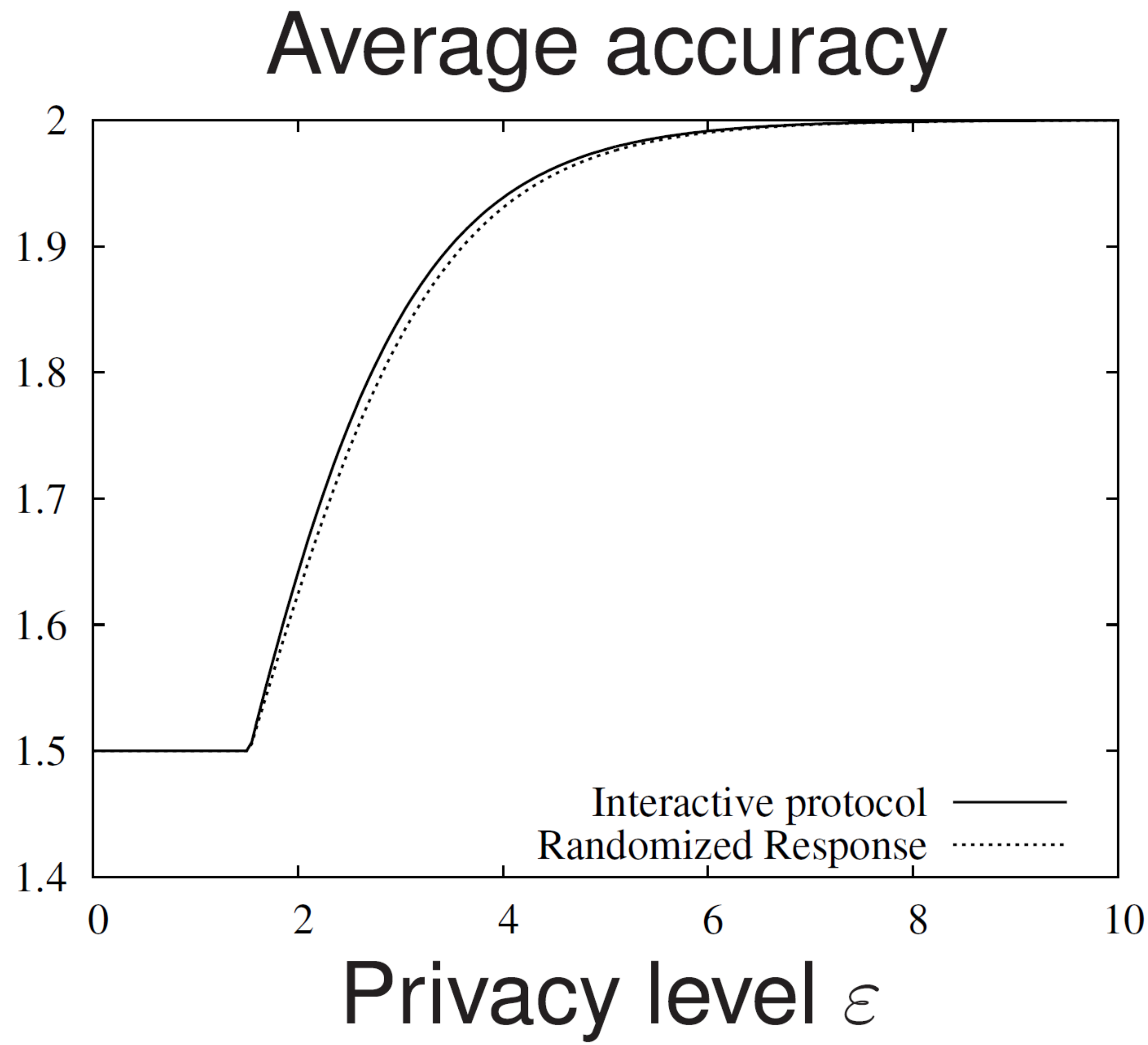


SAME RESULT!

$$\mathrm{ACC}_{\mathrm{wc}} \equiv \underbrace{\min_{x \in \{0,1\}^k}}_{\text{worst case over all possible inputs}} \mathbb{E}_{\hat{f}_i, P_{x,\tau}} [w_i(f_i(x), \tilde{f}_i(\tau, x_i))]$$

Non-binary data?

Non-binary data?



Acknowledgments

Pramod Viswanath

Sewoong Oh