

# CSCI 1515 Applied Cryptography

## This Lecture:

- Bootstrapping SWHE to FHE
- Secure Hardware: Intel SGX, HSM
- Differential Privacy

# FHE Constructions

Step 1: Somewhat Homomorphic Encryption (SWHE)

- over Integers

- from RLWE (BFV)

Step 2: Bootstrapping

## Step 2: Bootstrapping

$ct_1 \quad ct_2 \quad \dots \quad ct_n$



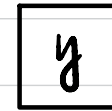
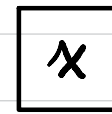
$ct_f \leftarrow$  too much noise!



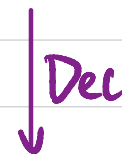
$y$



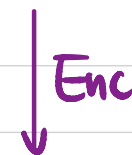
$ct_y \leftarrow$  fresh noise!



$y = f(x)$



$y$



# Levelled FHE

$(pk_1, sk_1)$     $ct_1$     $ct_2$     $\dots$     $ct_n$     $\boxed{x}_{pk_1}$

$\downarrow f$

too much noise!  $\rightarrow$   $ct_f$     $\boxed{y}_{pk_1}$

$\parallel$

$1001011 \dots 0$

$l$

$sk_1$

$\parallel$

$01101 \dots 1$

$k$

$(pk_2, sk_2)$

$\boxed{y}_{pk_1}$

$pk_2$

$Enc_{pk_2}$

$ct_1^{(2)}$     $ct_2^{(2)}$     $\dots$     $ct_l^{(2)}$

$Enc_{pk_2}$

$\tilde{ct}_1^{(2)}$     $\dots$     $\tilde{ct}_k^{(2)}$

$\boxed{sk_1}_{pk_2}$

$\boxed{\cancel{y}_{pk_1}}$

$sk_1$

$pk_2$

$\downarrow f' = Dec(sk_1, ct_f)$

$ct_{f'} = Enc_{pk_2}(y)$     $\boxed{y}_{pk_2}$

One more operation ADD & MULT

## Step 2: Bootstrapping

Leveled FHE:  $pk_1, pk_2, pk_3, \dots, pk_n$   
 $Enc_{pk_2}(sk_1), Enc_{pk_3}(sk_2), \dots, Enc_{pk_n}(sk_{n-1})$

FHE:  $pk, Enc_{pk}(sk)$

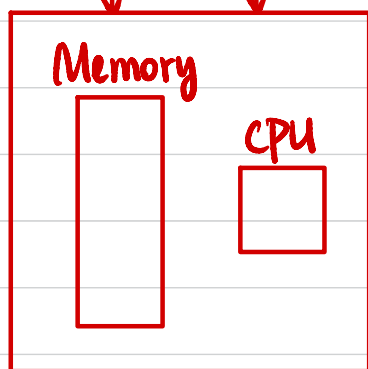
"circular secure" assumption

# Outsourcing Computation by FHE

Server



ct    f



Public Eval

Client



Data  $x$

Key  $sk$

$$ct \leftarrow \text{Enc}(x)$$

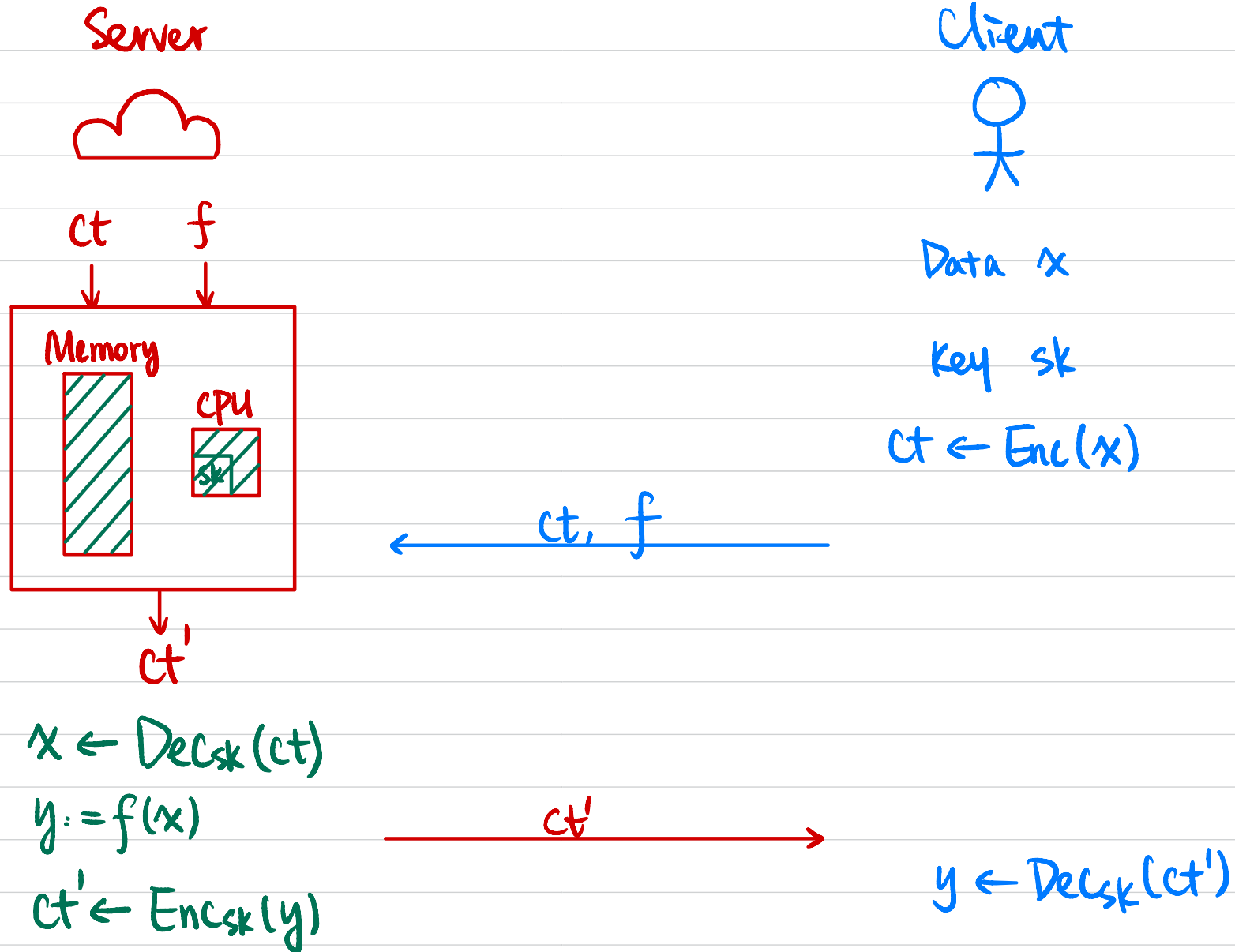


$$ct' \leftarrow \text{Eval}(f, ct)$$



$$f(x) \leftarrow \text{Dec}_{sk}(ct')$$

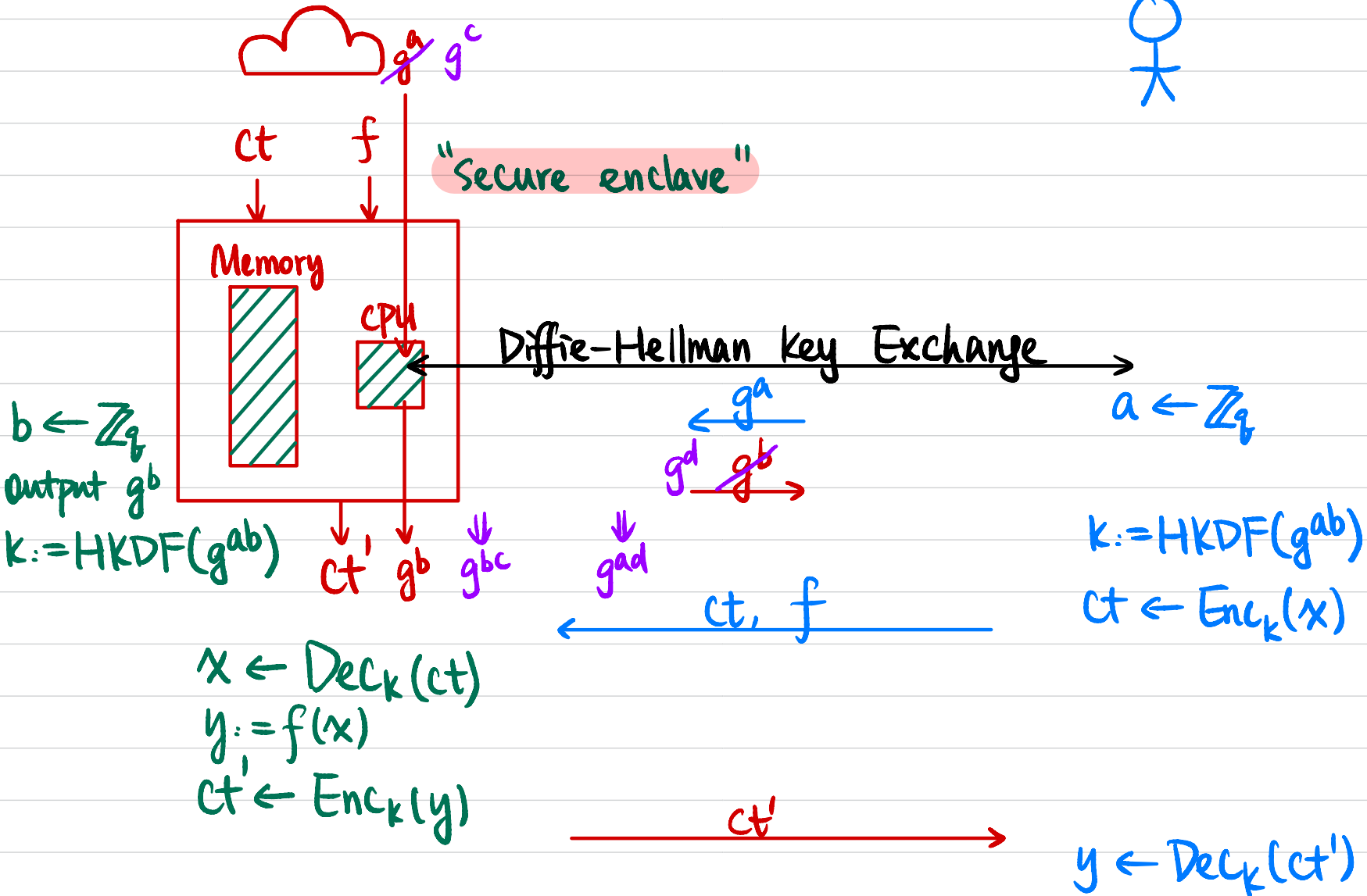
# Outsourcing Computation by Secure Hardware



# Intel Software Guard Extension (SGX)

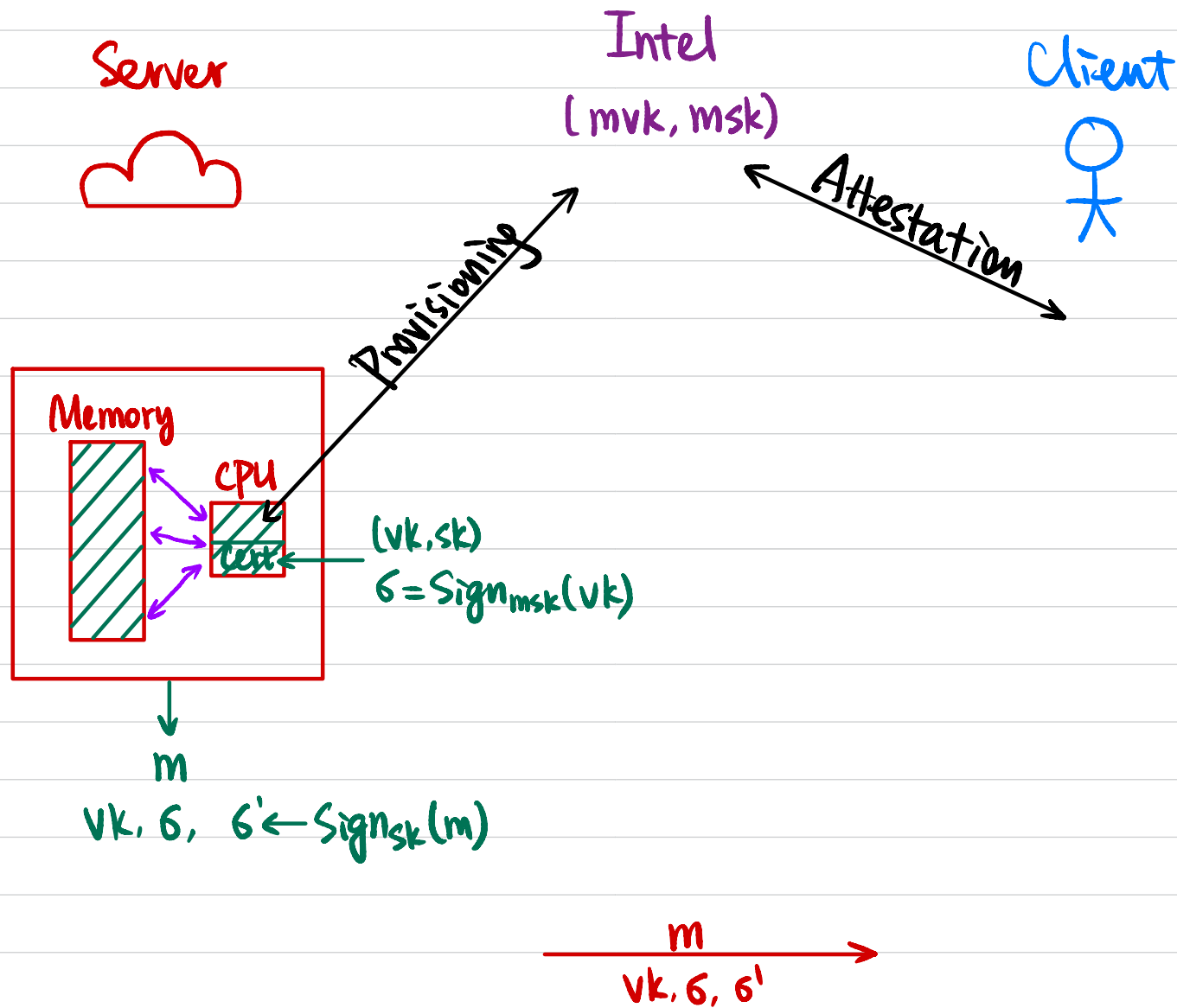
Server

Client



What could go wrong?

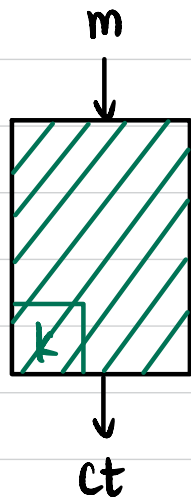
# Intel Software Guard Extension (SGX)



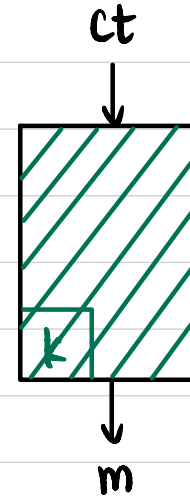
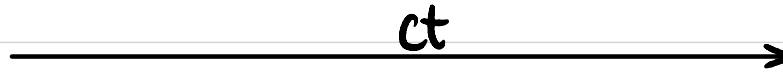
# Constraints & Attacks

- Trust in hardware
- Trust in Intel
- Limited memory size
- Replay attacks
- Side-channel attacks : memory access pattern
  - ↳ fix: Oblivious RAM (ORAM)
  - overhead  $\Theta(\log N)$

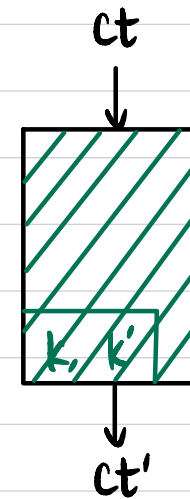
# Hardware Secure Module (HSM)



$$ct \leftarrow Enc_k(m)$$

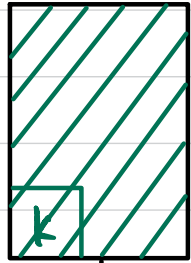


$$m \leftarrow Dec_k(ct)$$



$$m \leftarrow Dec_k(ct)$$
$$ct' \leftarrow Enc_{k'}(m)$$

# Key Agreement



Sample  $k_1, k_2, k_3$  s.t.

$$k_1 \oplus k_2 \oplus k_3 = k$$



$k_1$

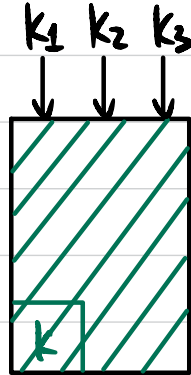
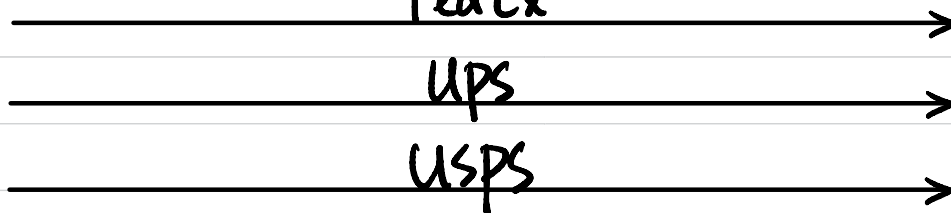
$k_2$

$k_3$

FedEx

UPS

USPS



$$k := k_1 \oplus k_2 \oplus k_3$$

# Differential Privacy

Name	Age	Gender	Race	Weight	ZIP	Disease
Alice						
Bob						
Charlie						
David						
Emily						
Fiona						

Want to make the (sensitive) data public / available to others  
(e.g. for medical study).

Attempt 1: "Anonymize" the data.

Delete personally identifiable information (PII): name, DOB, ...

Attempt 2: Only answer aggregate statistics queries.

## Privacy Guarantee?

Access to the output shouldn't enable one to learn anything about an individual compared to one without access.

Is this possible?

Privacy / Utility tradeoff

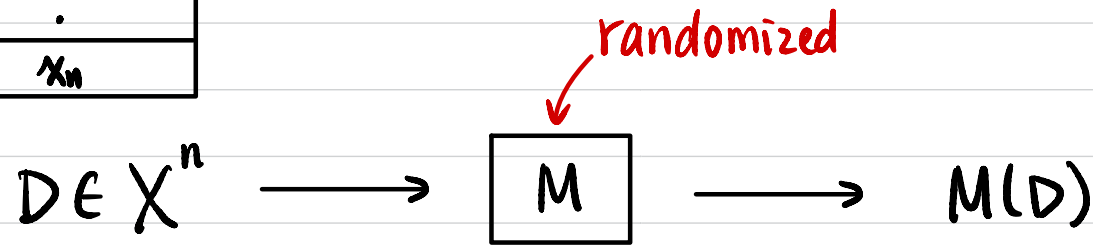
## Privacy Guarantee?

Access to the output shouldn't enable one to learn <sup>much more</sup> ~~anything~~ about an individual compared to one ~~without access~~.

with access to the output computed on a database without the individual.

# Differential Privacy

$x_1$
$x_2$
$\vdots$
$x_n$

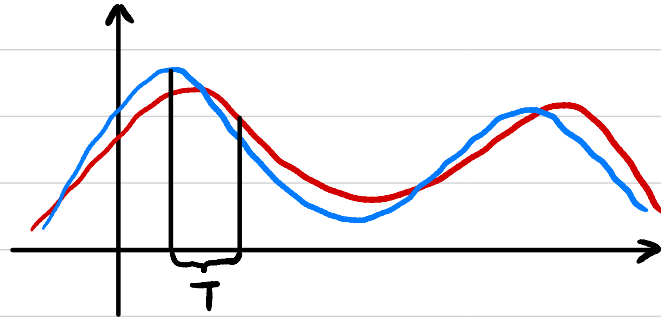


Def  $\epsilon$ -Differential Privacy for a randomized mechanism:

$\forall$  neighboring datasets  $D_1$  &  $D_2$  (differing in one row),

$\forall T \subseteq \text{range}(M)$ ,

$$\Pr[M(D_1) \in T] \leq e^\epsilon \cdot \Pr[M(D_2) \in T]$$



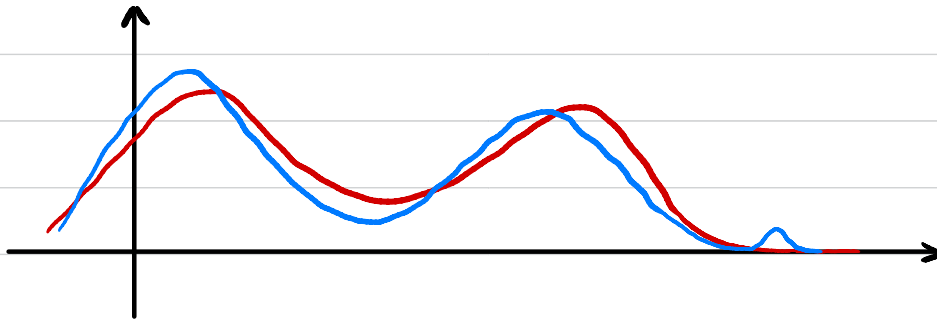
# Differential Privacy

Def  $(\epsilon, \delta)$  - Differential Privacy for a randomized mechanism:

$\forall$  neighboring datasets  $D_1$  &  $D_2$  (differing in one row),

$\forall T \subseteq \text{range}(M)$ ,

$$\Pr[M(D_1) \in T] \leq e^\epsilon \cdot \Pr[M(D_2) \in T] + \delta$$



Is a bigger  $\epsilon$  better for privacy, or worse? Worse

Is a bigger  $\delta$  better for privacy, or worse? Worse

# Randomized Response

Counting query: What percentage of individuals satisfy predicate  $P$ ?  $\alpha$

For each row  $x_i$ :

① Sample  $b \leftarrow \{0, 1\}$

② If  $b=0$ , then  $y_i := P(x_i) \rightarrow \frac{1}{2} \cdot \alpha$

Otherwise,  $y_i \leftarrow \{0, 1\} \rightarrow \frac{1}{2} \cdot \frac{1}{2}$

$$\frac{1}{2} \cdot \alpha + \frac{1}{2} \cdot \frac{1}{2} = \beta$$

$$\Rightarrow \alpha = (\beta - \frac{1}{4}) \cdot 2$$

$M(D) := (y_1, y_2, \dots, y_n) \leftarrow$  fraction of 1s =  $\beta$

Thm Randomized Response is  $\ln 3$ -DP.

How to estimate the query output?

$$\alpha = (\beta - \frac{1}{4}) \cdot 2$$

How to make the mechanism more private?

$b \rightarrow$  honestly 10%  
 $b \rightarrow$  randomly 90%

# Laplace Mechanism

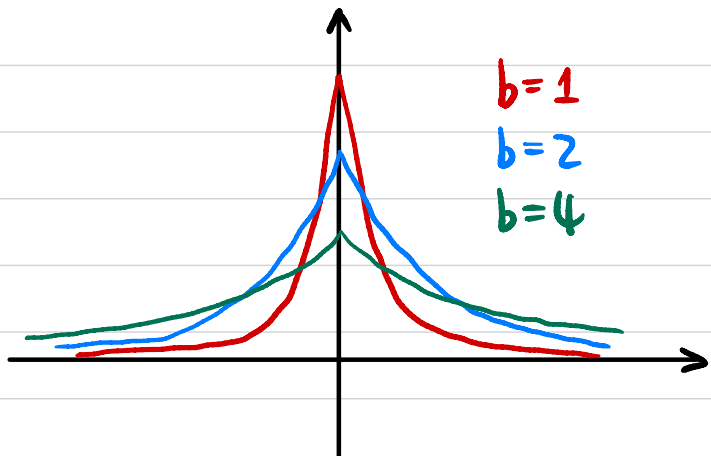
Def Sensitivity of a function  $f: X^n \rightarrow \mathbb{R}$

$$\Delta f := \max_{D_1 \sim D_2} |f(D_1) - f(D_2)|$$

Laplace Mechanism:  $M(D) = f(D) + \text{Lap}(\Delta f / \epsilon)$

Thm The Laplace Mechanism is  $\epsilon$ -DP.

Laplace distribution:



probability distribution function

$$\text{PDF}(x) = \frac{1}{2b} \cdot \exp\left(-\frac{|x|}{b}\right)$$

For  $X \sim \text{Lap}(b)$ ,  $\Pr[|X| \geq bt] \leq \exp(-t)$

Is a bigger  $b$  better for privacy, or worse?