CSCI 1515 Applied Cryptography

This Lecture:

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

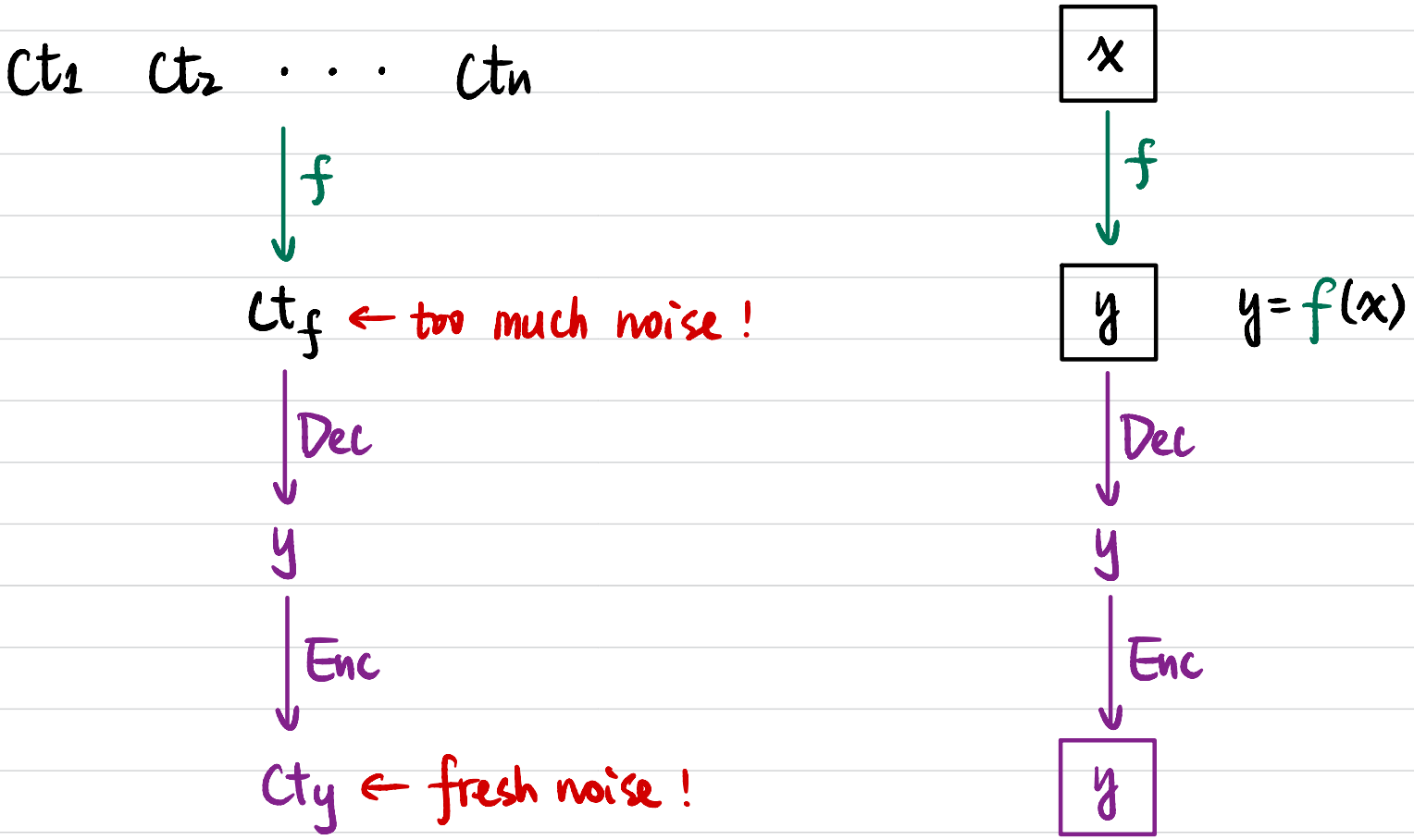
FHE Constructions

Step 1: Somewhat Homomorphic Encryption (SWHE)

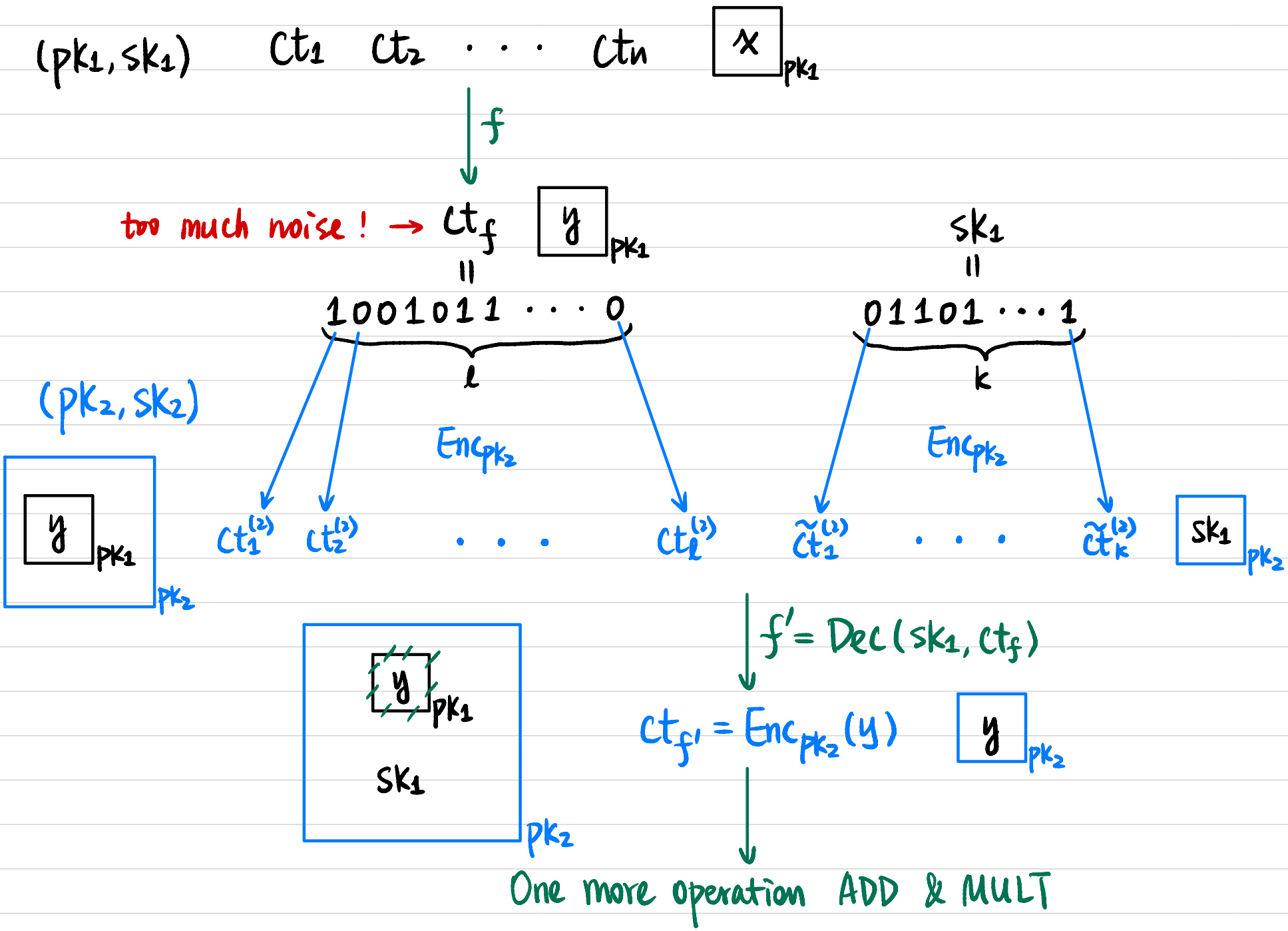
- over Integers
- from RLWE (BFV)

Step 2: Bootstrapping

Step 2: Bootstrapping



Leveled FHE


(pk_1, sk_1)
Ct_1 Ct_2 · · · Ct_n
x_{pk_1}
f
too much noise ! → Ct_f
y_{pk_1}
sk_1
=
1001011 · · · 0
l
01101 · · · 1
k
(pk_2, sk_2)
Enc_{pk_2}
Enc_{pk_2}
y_{pk_1} pk_2
$Ct_1^{(2)}$ $Ct_2^{(2)}$ · · · $Ct_l^{(2)}$
$\tilde{Ct}_1^{(2)}$ · · · $\tilde{Ct}_k^{(2)}$
sk_1 pk_2
y_{pk_1}
sk_1
pk_2
$f' = Dec(sk_1, Ct_f)$
$Ct_{f'} = Enc_{pk_2}(y)$
y_{pk_2}
One more operation ADD & MULT

Step 2: Bootstrapping

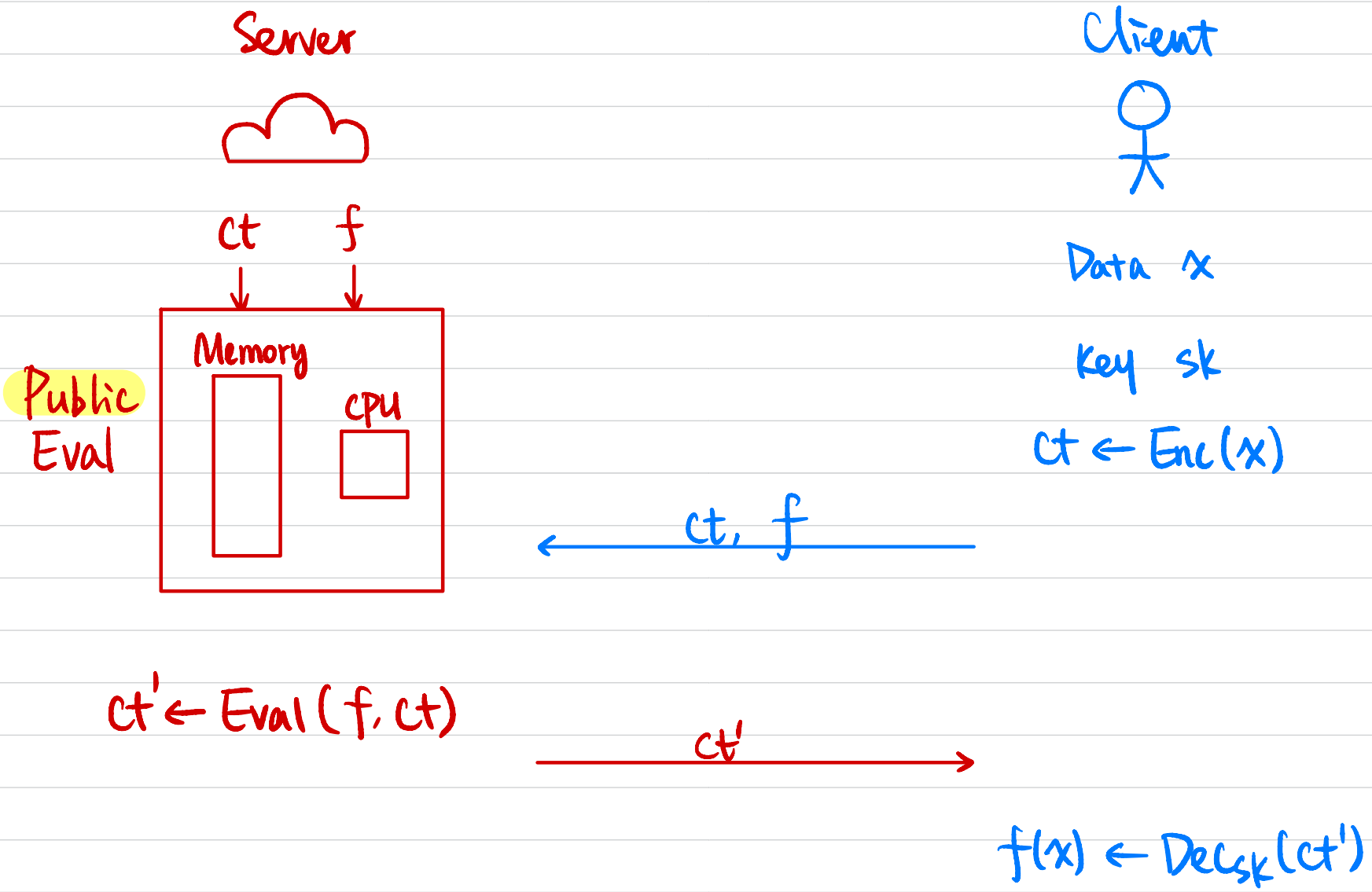
Leveled FHE: pk_1, pk_2, pk_3, . . . , pk_n

$Enc_{pk_2}(sk_1)$ $Enc_{pk_3}(sk_2)$ $Enc_{pk_n}(sk_{n-1})$

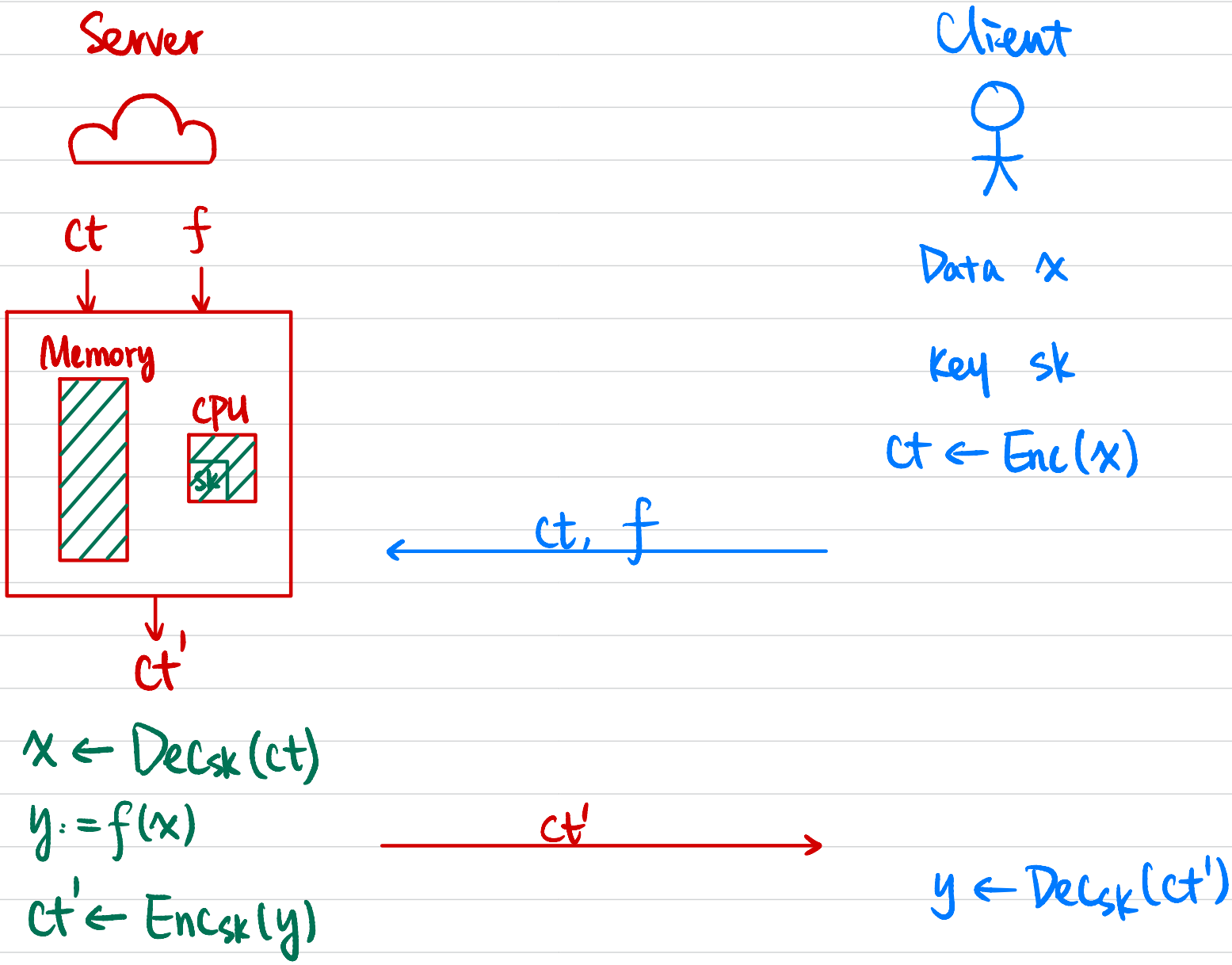
FHE: pk. $Enc_{pk}(sk)$

"circular secure" assumption

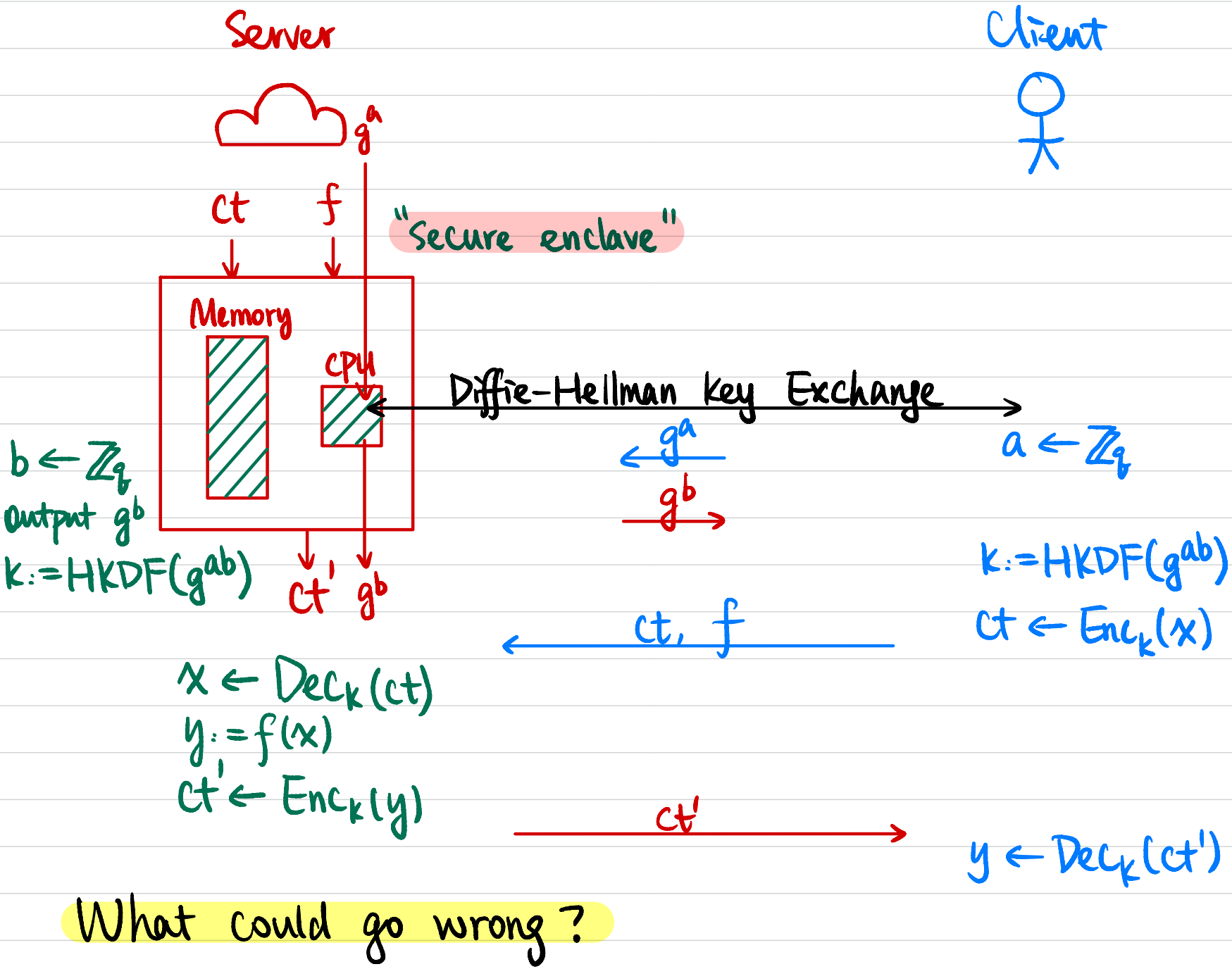
Outsourcing Computation by FHE



Outsourcing Computation by Secure Hardware

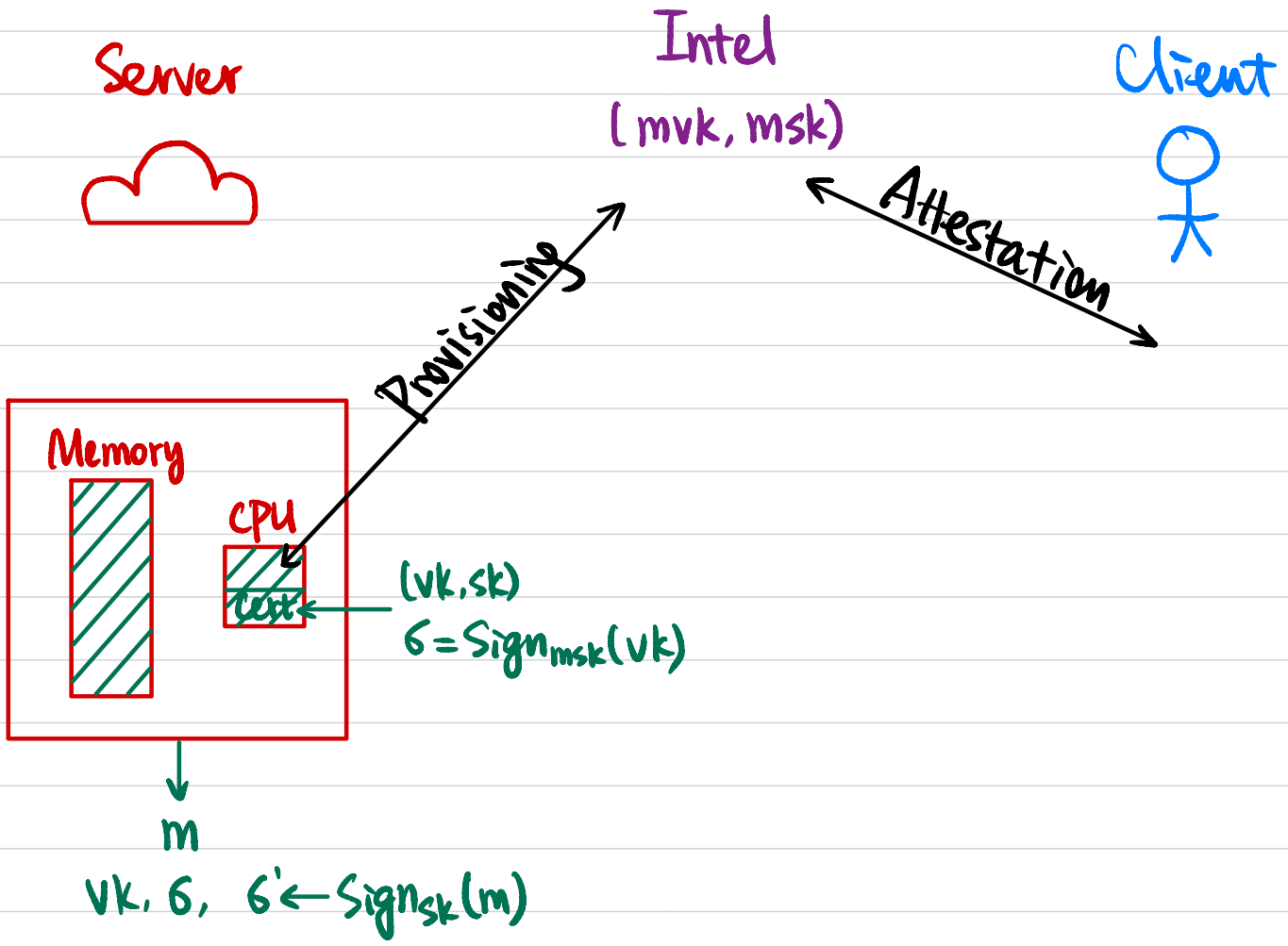


Intel Software Guard Extension (SGX)



What could go wrong?

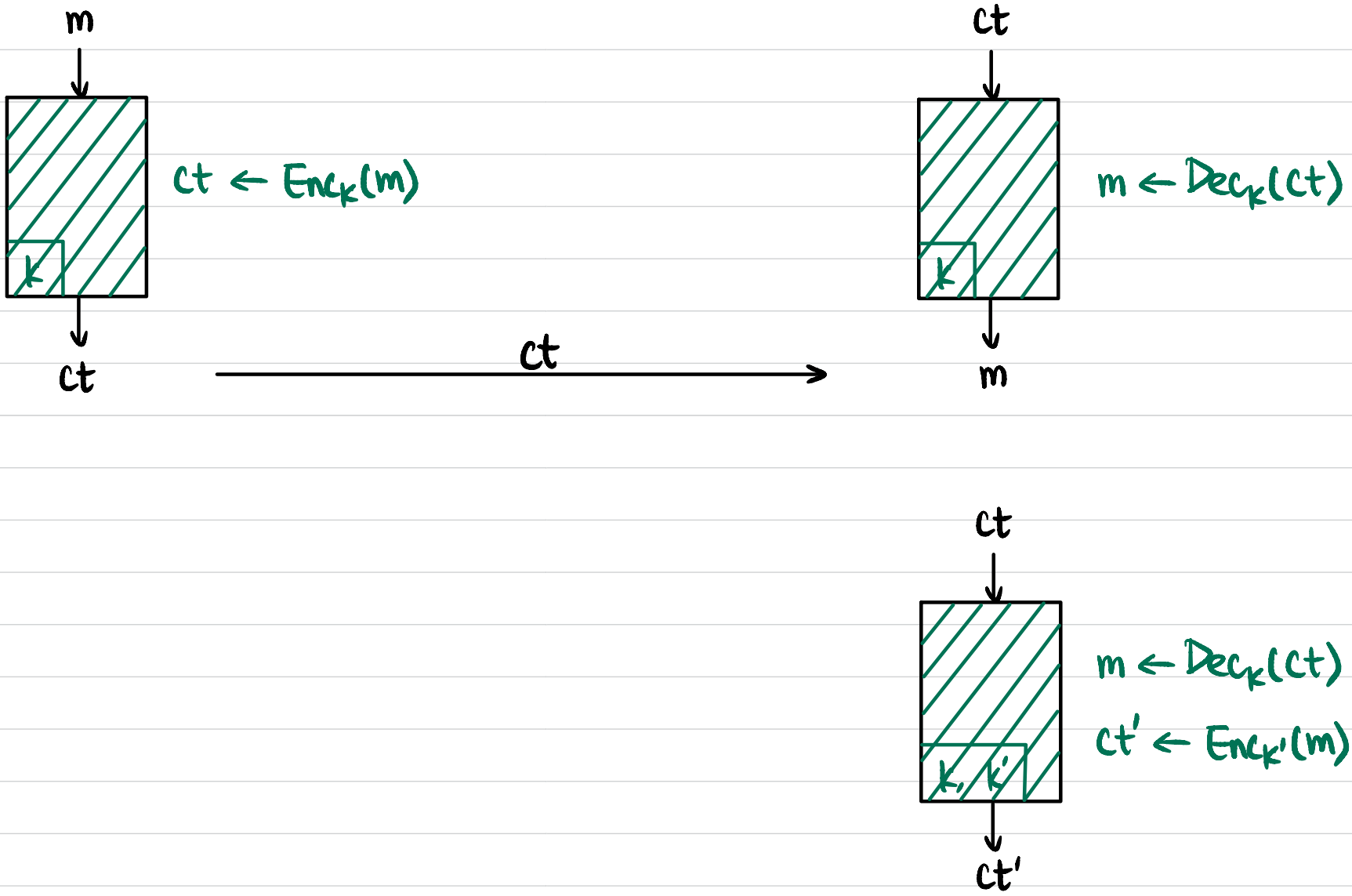
Intel Software Guard Extension (SGX)



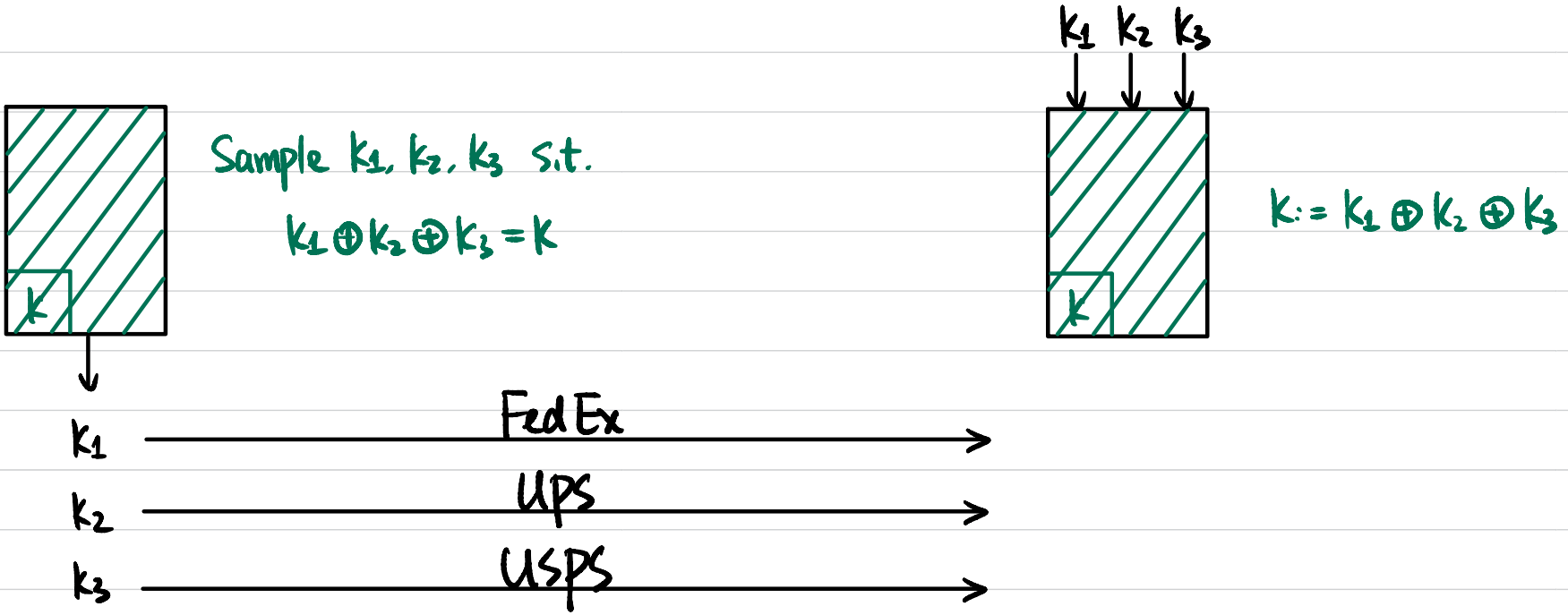
Constraints & Attacks

- 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)



Key Agreement



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.

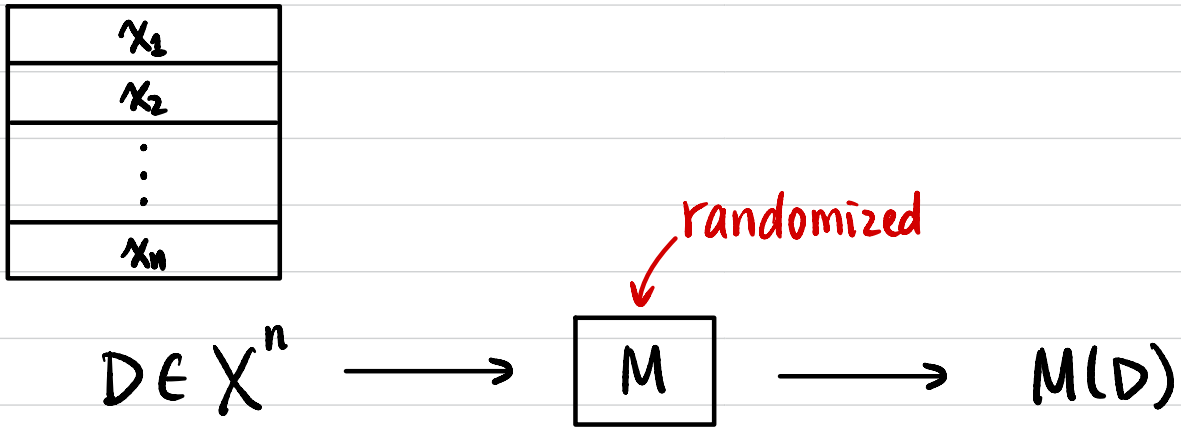
Is this possible?

Privacy Guarantee?

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

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

Differential Privacy

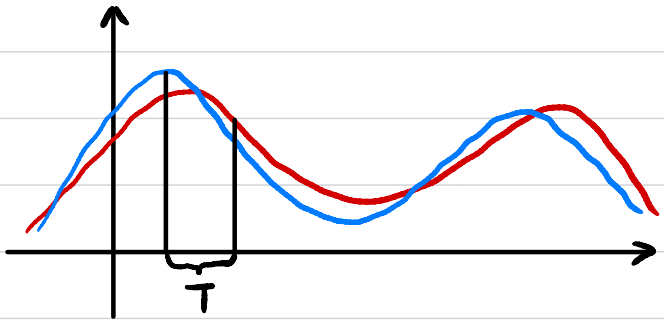


Def ε-Differential Privacy for a randomized mechanism:

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

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

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



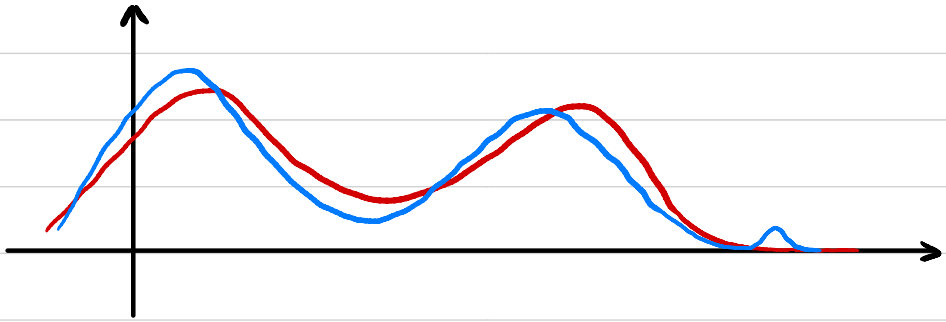
Differential Privacy

Def (ε, δ) - Differential Privacy for a randomized mechanism:

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

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

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



Is a bigger ε better for privacy, or worse?

Is a bigger δ better for privacy, or worse?

Randomized Response

Counting query: What percentage (α) of individuals satisfy predicate P?

For each row x_i:

① Sample $b \xleftarrow{\$} \{0, 1\}$

② If $b = 0$, then $y_i := P(x_i)$

Otherwise, $y_i \xleftarrow{\$} \{0, 1\}$

$M(D) := (y_1, y_2, \cdots, y_n)$ ← fraction of 1s = p

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

How to estimate the query output?

$\alpha = ?$

How to make the mechanism more private?

Laplace Mechanism

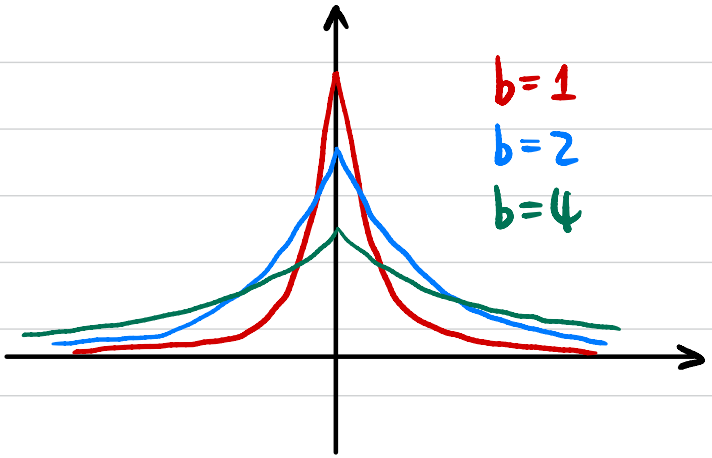
Def Sensitivity of a function $f: X^n \to \mathbb{R}$

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

Laplace Mechanism: $M(D) = f(D) + Lap(\Delta f / \varepsilon)$

Thm The Laplace Mechanism is ε-DP.

Laplace distribution:



probability distribution function

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

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

Is a bigger b better for privacy, or worse?

Composition Theorems

Thm (post-processing) If $M: X^n \to Y$ is (ε, δ)-DP.

$f: Y \to Z$ is an arbitrary randomized function,

then $f \circ M: X^n \to Z$ is also (ε, δ)-DP.

Thm (group privacy) If $M: X^n \to Y$ is $(\varepsilon, 0)$-DP.

then M is $(k\cdot\varepsilon, 0)$-DP for groups of size k.

Thm (composition) If $M_i: X^n \to Y$ is $(\varepsilon_i, \delta_i)$-DP $\quad \forall i \in [k]$,

then $M(D) := (M_1(D), \cdots, M_k(D))$ is $\left(\sum_{i\in[k]} \varepsilon_i, \sum_{i\in[k]} \delta_i\right)$-DP.