Computational Differential Privacy

Ilya Mironov	(MICROSOFT)
Omkant Pandey	(UCLA)
Omer Reingold	(MICROSOFT)
Salil Vadhan	(HARVARD)

Focus of the Talk

- Relaxations of differential privacy for computational adversaries
- How they relate to one another and other existing notions
- Natural protocols demonstrating their benefits

Motivation

- **Achieve better utility**
- **Standard MPC does not prevent what is leaked by the output**
  - Can we combine computational MPC protocols with DP-functions [DKMMN'06,BNO'08]?
- **Nontrivial differentially private mechanisms must be randomized**
  - Applications typically use pseudorandom sources. What are the formal privacy guarantees achieved?

Differential Privacy

[Dwork'06]

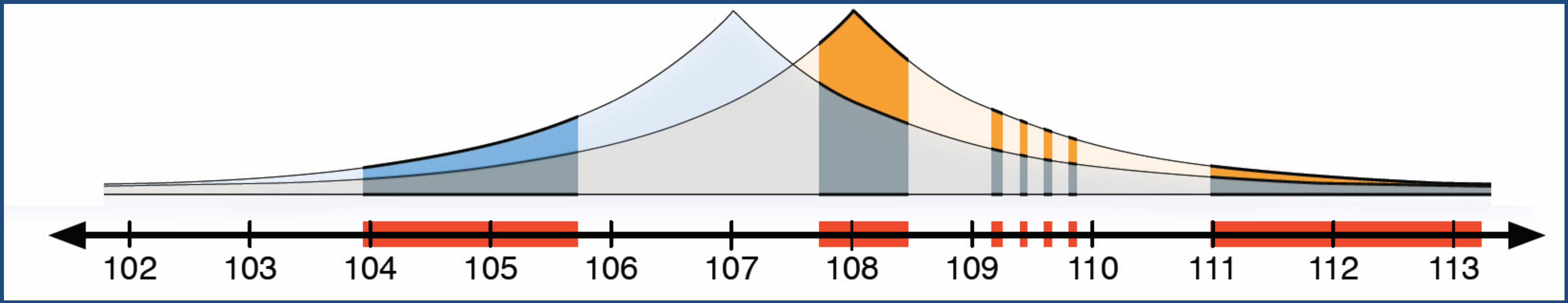
"adjacent" means "differ in one individual's entry"

- Mechanism ***K*** provides privacy to individual's data effects the output of "little"

K: ***D*** → ***R*** ensures **ε**-DP if for all adjacent datasets D_1, D_2 and for all subsets S of ***R***:

$$\frac{\Pr[K(D1) \in S]}{\Pr[K(D2) \in S]} \le e^{\varepsilon}$$

Pictorial Representation



— bad outcome

— probability with record *x*

— probability without record *x*

Towards Computational Notions

$$\Pr[K(D_1) \in S] \leq e^{\varepsilon} \Pr[K(D_2) \in S]$$

Equivalently,

$$\Pr[\mathrm{A}(K(D_1)) = 1] \leq e^{\varepsilon} \Pr[\mathrm{A}(K(D_2)) = 1]$$

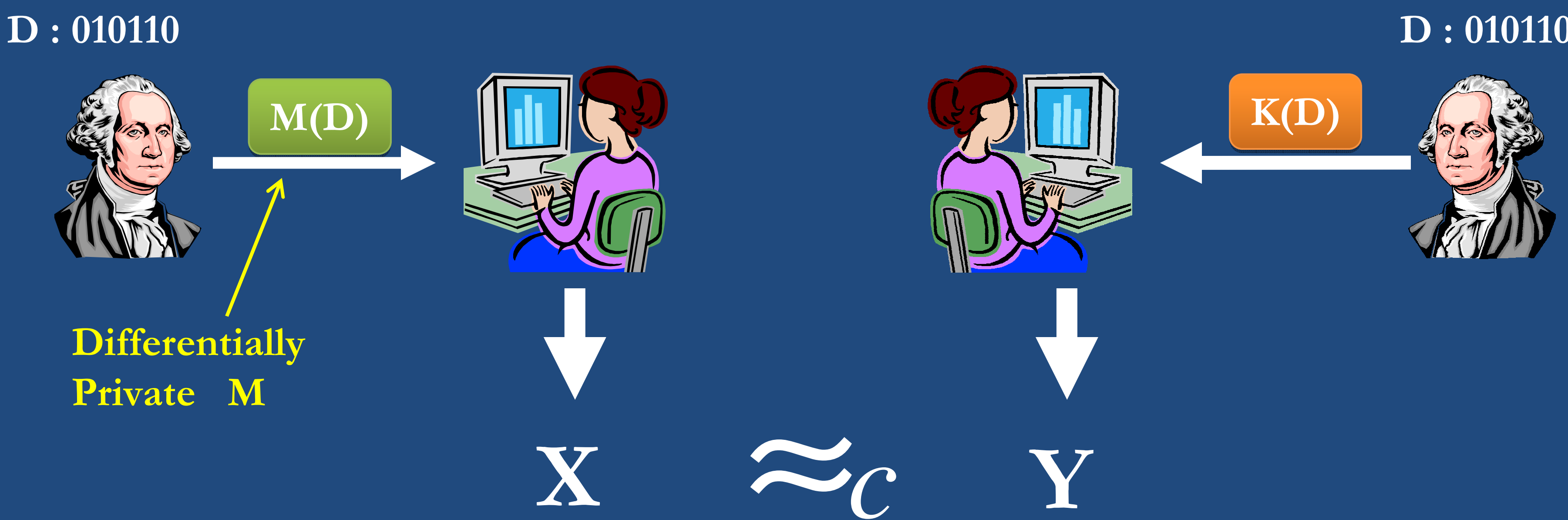
First Definition: IND-CDP

ε-IND-CDP : **Mechanism *K* is ε-IND-CDP if for all adjacent D_1, D_2, for all polynomial sized circuits A, and for all large enough λ, it holds that,**

$$\Pr[A(K(D_1)) = 1] \leq e^{\varepsilon} \Pr[A(K(D_2)) = 1] + \mathrm{negl}(\lambda)$$

Necessary

Simulation-based Approach



Second Definition: SIM-CDP

ε-SIM-CDP : Mechanism *K* is ε-SIM-CDP if there exists an ε-differentially-private mechanism *M* such that f… all *D*, distributions *M(D)* and *K(D)* …nally indistinguishable.

$\exists M, \forall(D_1, D_2)$

- M is not necessarily a PPT mechanism
- Reversing the order of quantifiers yields another definition, $\text{SIM}_{\forall\exists}$-CDP:

$$\forall (D_1, D_2), \exists M$$

Immediate Questions

- Are these definitions equivalent?
- Not hard to see that

$$\text{SIM-CDP} \Longrightarrow \text{IND-CDP}$$

- Main question:

$$\text{IND-CDP} \Longrightarrow \text{SIM-CDP}?$$

Connection with Dense Models

[RTTV'08, Imp'08]

- **Distribution X is α-dense in Y if for all tests T,**

$$\Pr[T(X)=1] \le \frac{1}{\alpha}\Pr[T(Y)=1]$$

- **X is α-pseudodense in Y if for all PPT tests T,**

$$\Pr[T(X)=1] \le \frac{1}{\alpha}\Pr[T(Y)=1] + \text{negl}$$

[RTTV'08] : Reingold, O., Trevisan, L., Tulsiani, M., Vadhan, S. "Dense subsets of Pseudorandom Sets", FOCS 2008.

Connection with Dense Models

[RTTV'08, Imp'08]

- **Differential Privacy:**
  - $\Pr[K(D_1) \in S] \leq e^{\varepsilon} \Pr[K(D_2) \in S]$
  - $\Pr[K(D_2) \in S] \leq e^{\varepsilon} \Pr[K(D_1) \in S]$
- **In the language of dense models**
  - $K(D_1)$ **is** e^{ε}**-dense in** $K(D_2)$
  - $K(D_2)$ **is** e^{ε}**-dense in** $K(D_1)$

ε-DP: $K(D_1)$ and $K(D_2)$ are mutually e^{ε}-dense

Connection with Dense Models

[RTTV '08, Imp'08]

- **ε - IND-CDP:**
  - $\Pr[A(K(D1)) = 1] \leq e^{\varepsilon} \Pr[A(K(D2) = 1] + \text{negl}$
  - $\Pr[A(K(D2)) = 1] \leq e^{\varepsilon} \Pr[A(K(D1) = 1] + \text{negl}$
- **In the language of dense models**
  - $K(D_1)$ **is e^{ε}-pseudodense in** $K(D_2)$
  - $K(D_2)$ **is e^{ε}-pseudodense in** $K(D_1)$

ε-IND-CDP: $K(D_1)$ and $K(D_2)$ are mutually e^{ε}-pseudodense

Some Notation

$X \dashleftarrow Y$ (X is pseudodense in Y)

$X \dashleftarrow\!\dashrightarrow Y$ (X,Y are mutually pseudodense)

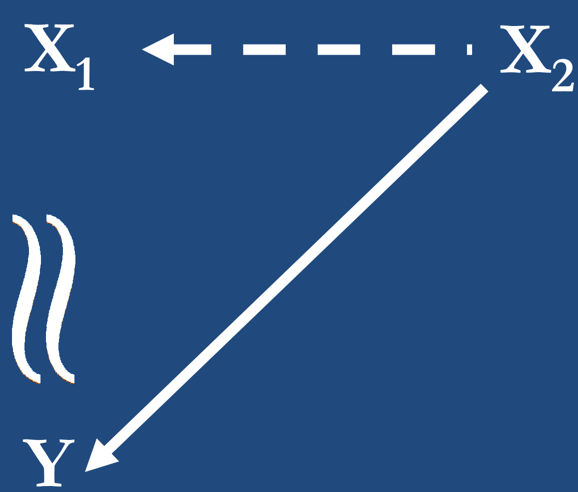
$X \longleftarrow Y$ (X is dense in Y)

$X \longleftrightarrow Y$ (X,Y are mutually dense)

$X \approx Y$ (X,Y comp. indistinguishable)

The Dense Model Theorem

[RTTV'08]



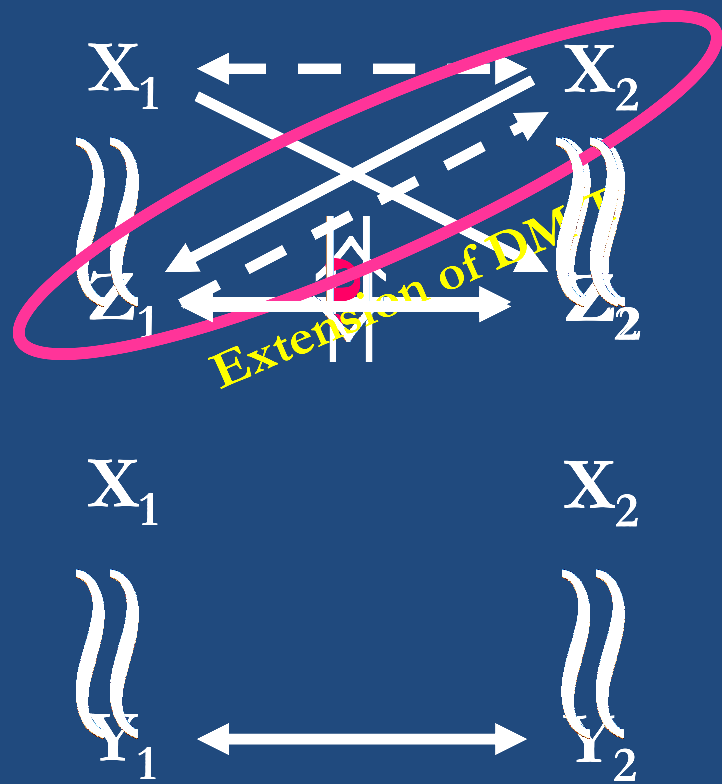
Thm : If X_1 is pseudodense in X_2, there exists a model Y (truly) dense in X_2 such that X_1 is computationally indistinguishable from Y.

Proof Ideas

(IND-CDP)

$\Updownarrow$

(SIM$_{\cancel{\forall\exists}}$-CDP)



X_1=K(D_1)

X_2=K(D_2)

Y_1=M(D_1)

Y_2=M(D_2)

$\cancel{\forall(D_1, D_2), \exists M}$

$\exists M, \forall(D_1, D_2)$

X $\longleftarrow$ Y: X dense in Y,
X $\leftarrow$- - - - - Y: X pseudo-dense in Y,

X $\longleftrightarrow$ Y: X,Y mutually dense
X $\leftarrow$- - -$\rightarrow$ Y: X,Y mutually pseudo-dense

To Recap

- We prove an extension of "The Dense Model Theorem" of [RTTV'08].

- Sufficient to establish: **IND-CDP** $\Leftrightarrow$ **SIM$_{\forall\exists}$-CDP**

- Still OPEN: **IND-CDP** $\overset{?}{\Longrightarrow}$ **SIM-CDP**

Benefits: Better Utility

Alice **Bob**

$x_1 \longrightarrow$ **SFE** $\longleftarrow y_1$

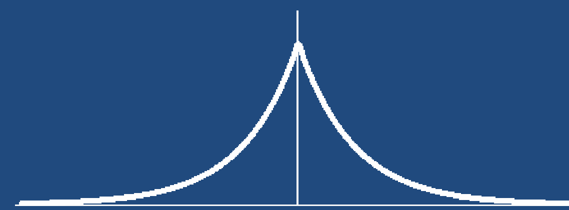
x_2 $\downarrow$ y_2

$\ldots$ H(x,y) $\ldots$

x_n y_n

Protocol:

~~Trusted Party:~~ H(x,y)+Lap(1/ε)



DP : Requires $\tilde{\Omega}(n^{1/2})$ error ! **[Reingold-Vadhan]**

CDP : Easily get $\Theta(1/\varepsilon)$ error w/ constant probability.

Other Results

- A new protocol for Hamming Distance:
  - Differentially private (standard)
  - Constant multiplicative error
- Differentially Private Two-Party Computation

Thank you for your attention!