

Theory of Differential Privacy

Lecture 12:

Binary Tree Mechanism

EN.601.438/638 Fall 2026

Lydia Zakynthinou

Johns Hopkins University

Oct 12, 2026

Range queries

Dataset $x = (x_1, \dots, x_n) \in [D]^n$

Range query: $f_{s,t}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{s \leq x_i \leq t\}$ $1 \leq s \leq t \leq D$

Prefix queries

$$f_t(x) = f_{1,t}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{x_i \leq t\}$$

$$f_0(x) = 0, \quad f_D(x) = 1.$$

Goal:

return answers $b_{s,t}$, $\forall 1 \leq s \leq t \leq D$:

$$\Pr \left[\max_{1 \leq s \leq t \leq D} |f_{s,t}(x) - b_{s,t}| \leq \alpha \right] \geq 1 - \theta.$$

Naïve approach:

$$f_{s,t}, \quad 1 \leq s \leq t \leq D$$

$$k = \binom{D}{2} \approx D^2.$$

$$\ell_1 \text{ sensitivity: } \max_{x \sim x'} \left\| \left(f_{s,t}(x) \right)_{1 \leq s \leq t \leq D} - \left(f_{s,t}(x') \right)_{1 \leq s \leq t \leq D} \right\|_1 \leq \frac{D^2}{n}$$

$$\text{Output } b_{s,t} = f_{s,t}(x) + Z_{s,t}, \quad Z_{s,t} \sim \text{Lap}\left(\frac{D^2}{\epsilon n}\right)$$

$$\max_{1 \leq s \leq t \leq D} |f_{s,t}(x) - b_{s,t}| \leq \frac{D^2}{\epsilon n} \log \frac{D}{\delta} \quad \text{w.p. } 1 - \delta.$$

Less Naïve approach:

$$\text{Observe } f_{s,t}(x) = f_t(x) - f_{s-1}(x).$$

$$k = D - 1.$$

$$\ell_1 \text{ sensitivity : } \max_{x \sim x'} \left\| \left(f_t(x) \right)_{1 \leq t \leq D} - \left(f_t(x') \right)_{1 \leq t \leq D} \right\|_1 = \frac{D-1}{n}.$$

$$\text{Output } C_t = f_t(x) + \text{Lap}\left(\frac{D-1}{\epsilon n}\right)$$

$$\max_t |C_t - f_t(x)| \leq \underbrace{\frac{D}{\epsilon n} \log \frac{D}{\delta}}_{\substack{\downarrow \\ \text{w.p. } 1-\delta.}} \quad \text{w.p. } 1-\delta.$$

$$\text{Answer } a_{s,t} = b_t - b_{s-1}$$

$$\max_{s,t} |a_{s,t} - f_{s,t}(x)| \leq 2 \cdot \frac{D}{\epsilon n} \log \frac{D}{\delta}$$

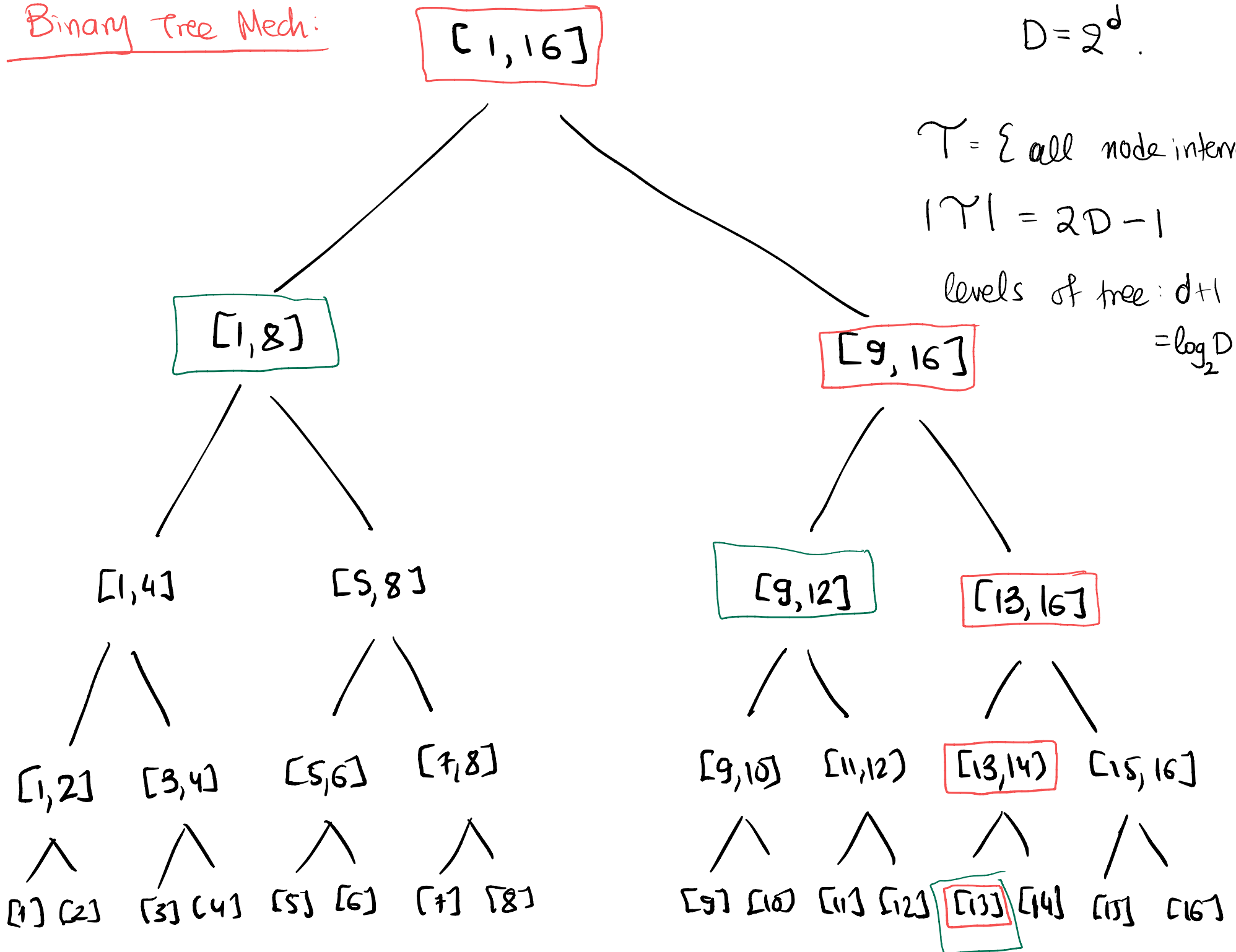
Binary Tree Mech:

$$D = 2^d.$$

$\mathcal{T} = \{\text{all node intervals}\}$

$$|\mathcal{T}| = 2D - 1$$

levels of tree: $d+1$
 $= \log_2 D + 1.$



for every node interval $I \in \mathcal{Y}$, let $f_I(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{x_i \in I\}$

Denote these by vector $G(x) = (f_I(x))_{I \in \mathcal{Y}}$.

for each $I \in \mathcal{Y}$: $f_I(x) - f_I(x') = \frac{1}{n} (\mathbf{1}\{a \in I\} - \mathbf{1}\{b \in I\})$.

if I change one
record from a to b

ℓ_1 sensitivity: $\|G(x) - G(x')\|_1 \leq \frac{2d}{n} = \frac{2 \log_2 D}{n}$.

let $\lambda = \frac{2d}{\epsilon n}$.

1 Release for every $I \in \mathcal{Y}$: $a_I = f_I(x) + Z_I$, $Z_I \sim \text{Lap}(\lambda)$.

Vector $a = (a_I)_{I \in \mathcal{Y}}$ is ϵ -DP.

Claim. for every $t \in [D]$, a set S_t of node intervals $I \in \mathcal{I}$
 $S_t \subseteq \mathcal{I}$ s.t

$$[1, t] = \bigcup_{I \in S_t} I, \quad |S_t| \leq d.$$

Reconstruct prefix queries $f_t(x) = \sum_{I \in S_t} f_I(x)$

[2] $b_t = \sum_{I \in S_t} a_I$

[3] Return $b_{s,t} = b_t - b_{s-1}, \quad \forall 1 \leq s \leq t \leq D.$

Accuracy analysis

$$\Pr \left[\max_{I \in \mathcal{I}} |Z_I| \leq \lambda \log \frac{|\mathcal{I}|^{2D-1}}{\delta} \right] \geq 1 - \delta.$$

$\Rightarrow$ w.p. $1 - \delta$ $\underbrace{\max_{I \in \mathcal{I}} |Z_I| \leq \frac{2d}{\epsilon n} \log \frac{D}{\delta}}_{\text{event } E}.$

$$\begin{aligned} \text{Conditioning on } E, \max_t |b_t - f_t(x)| &= \left| \sum_{I \in S_t} z_I \right| \leq \sum_{I \in S_t} |z_I| \leq \frac{2d}{\epsilon n} \log \frac{D}{\theta} \cdot |S_t| \\ &\leq \frac{2d^2}{\epsilon n} \log \left(\frac{D}{\theta} \right) \end{aligned}$$

$$\Rightarrow \forall 1 \leq s \leq t \leq D : b_{s,t} = b_t - b_{s-1}$$

$$\begin{aligned} \text{I have } \max_{1 \leq s \leq t \leq D} |b_{s,t} - f_{s,t}(x)| &\leq 2 \cdot \frac{2d^2}{\epsilon n} \log \frac{D}{\theta} = O \left(\frac{\log^2 D}{\epsilon n} \cdot \log \frac{D}{\theta} \right) \\ &\approx \tilde{O} \left(\frac{\log^3 D}{\epsilon n} \right). \end{aligned}$$

Actually, with
better analysis $\nearrow \frac{\log^2 D}{\epsilon n}$