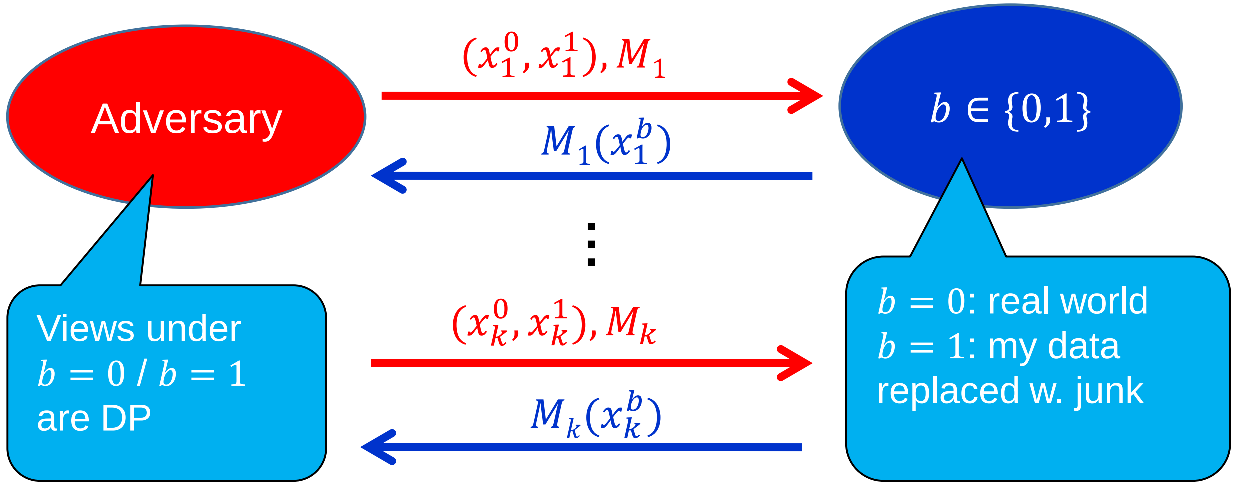
Using your privacy budget: Tree algorithm and Advanced Composition



Kobbi Nissim, Georgetown University

Bar-Ilan Winter School on Differential Privacy

February 2017

Includes slides taken from Guy Rothblum

Important properties of differential privacy

- Post processing:
  - If A is ε-dp then $B \circ A$ is also ε-dp for all B

Special case of composition

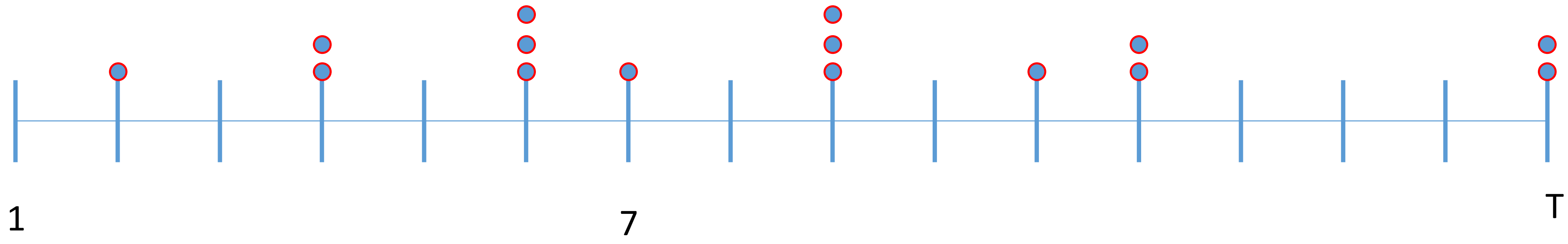
- Composition:
  - Adaptive executions of differentially private mechanisms results in differential privacy [DMNS06, …]
- Why do we care?
  - For privacy: A definition that does not post process/compose is (to the least) problematic
  - For DP algorithm design: Allows a modular design of an analysis from simpler analyses
  - For data analysis (even when privacy is not a goal): Statistical validity under adaptive querying [DFHPRR'15, …]

Basic composition

- Setting:
  - M_i be (ϵ_i, δ_i)-differentially private
  - M applies $M_1, \dots, M_t$ on its input (the inner $M_1, \dots, M_t$ use independent randomness).
- Basic composition theorem [DMNS06, DL09]:
  - M is $(\sum_i \epsilon_i, \sum_i \delta_i)$-differentially private
- Basic composition suggests that ϵ (and to a lesser account δ) can be treated as a 'privacy budget':
  - Split 'privacy budget' ϵ into smaller budget $\sum_i \epsilon_i$; allocate portion ϵ_i to mechanism M_i
    - Spend your budget carefully!
- More refined theorems (later):
  - Advanced composition [DRV10]
  - Optimal composition [KOV15, MV15]

Answering all threshold queries

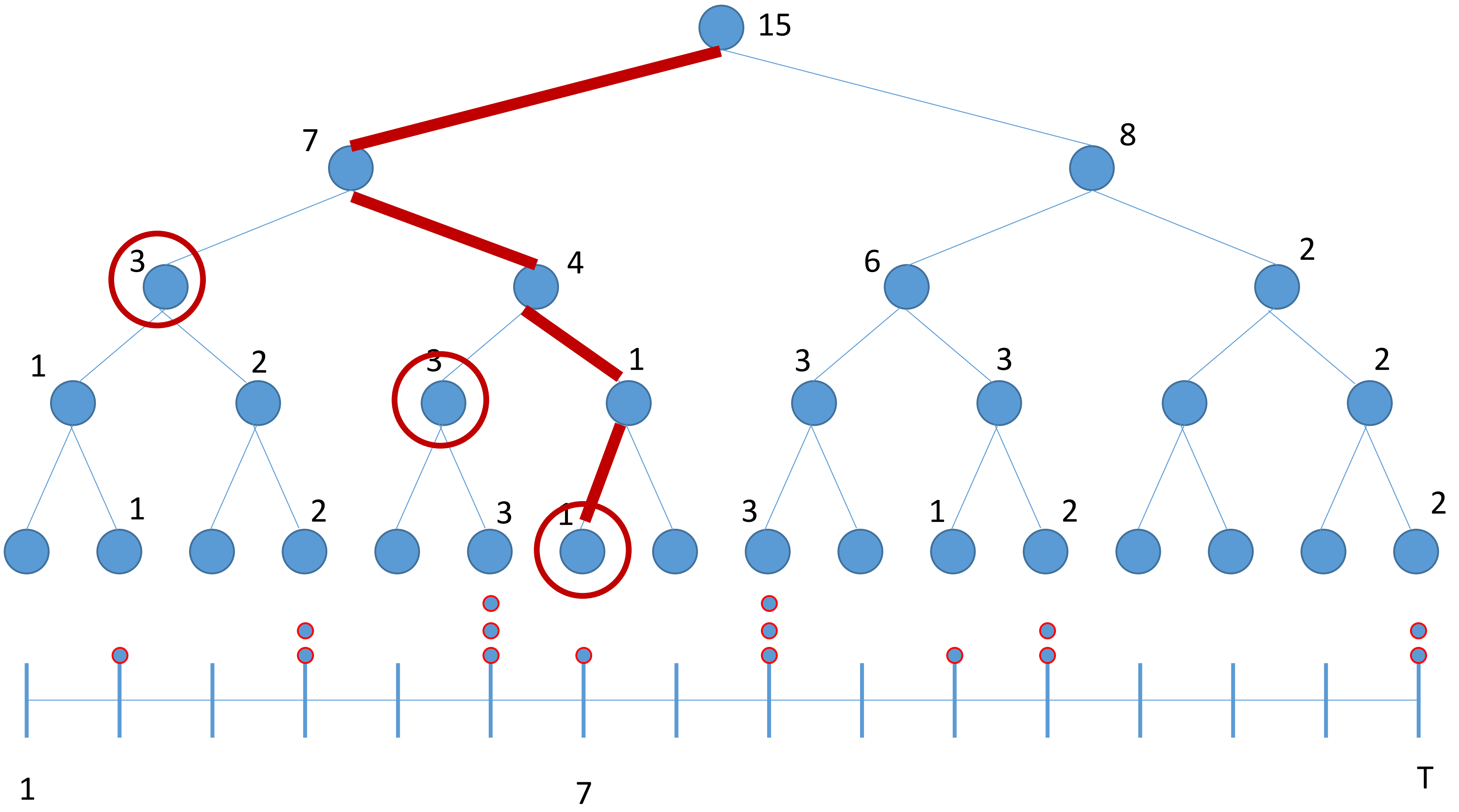
- Data domain: $X = \{1, \ldots, T\}$ (ordered domain with T elements)
- Database: $d \in X^n$
- Want (approx.) answers to all queries of the form: $q_t(d) = \frac{|\{i \,:\, 1 \le x_i \le t\}|}{n}$
- $GS(q_t) = \frac{1}{\mathrm{n}}$ (changing a data point in d can increase/decrease $q_t(d)$ by at most one)
- Idea: answer all T queries by adding noise $Lap(\frac{1}{\epsilon'})$ where $\epsilon' = \frac{\epsilon}{T}$
  - Using (simple) composition, this provides ϵ-differential privacy
  - Problem: noise magnitude linear in T; can we do better?



$n = 15$

$q_7(d) = \frac{7}{15}$

Answering all threshold queries

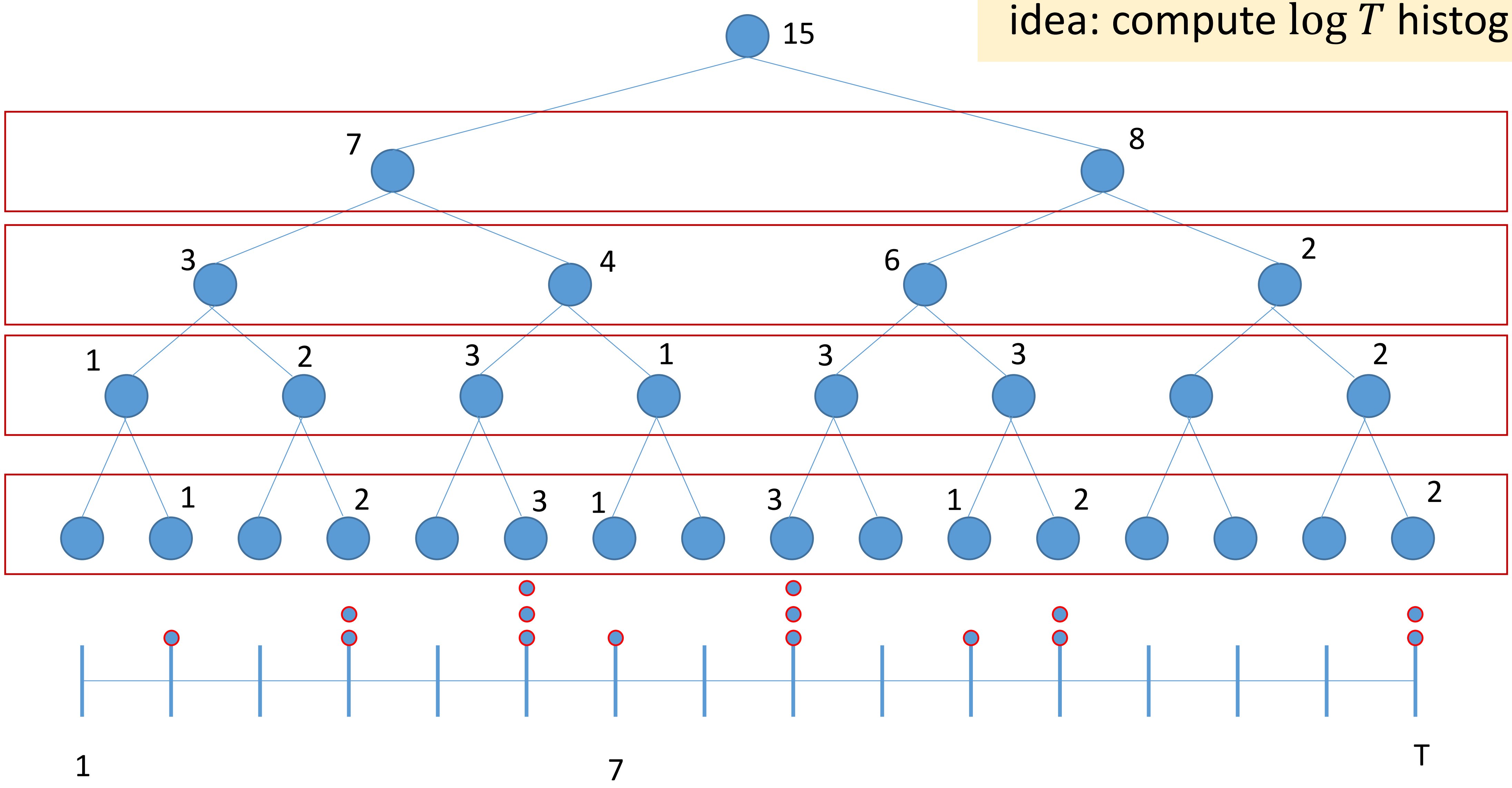


$n = 15$

$q_7(d) = \frac{7}{15}$

Answering all threshold queries

idea: compute $\log T$ histograms

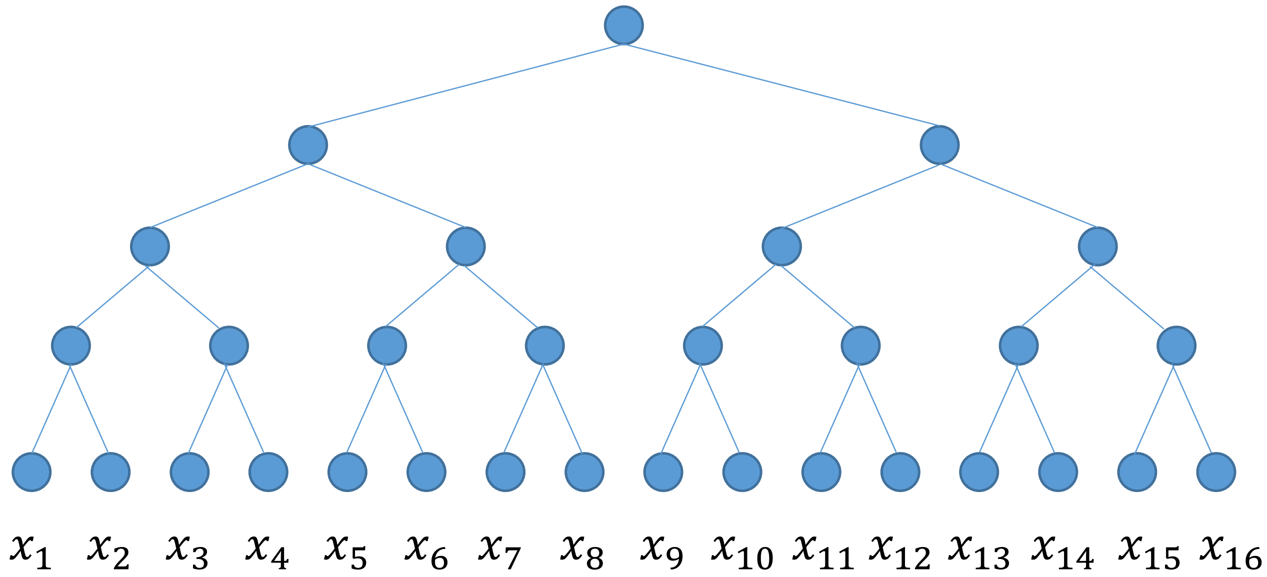


Answering all threshold queries

- What we get (using basic composition):
  - Computing $\log T$ histograms, each with $\epsilon' = \frac{\epsilon}{\log T}$
    - E.g., add noise $\mathrm{Lap}(2\epsilon / \log T)$ to each count
      - Noise variance $\sim \left(\frac{\log T}{\epsilon}\right)^2$
  - Each answer to threshold query is sum of (at most) $\log T$ noisy estimates
  - Overall noise variance $\sim \log T \left(\frac{\log T}{\epsilon}\right)^2$
    - Whp noise magnitude = $\frac{polylog(T)}{\epsilon}$

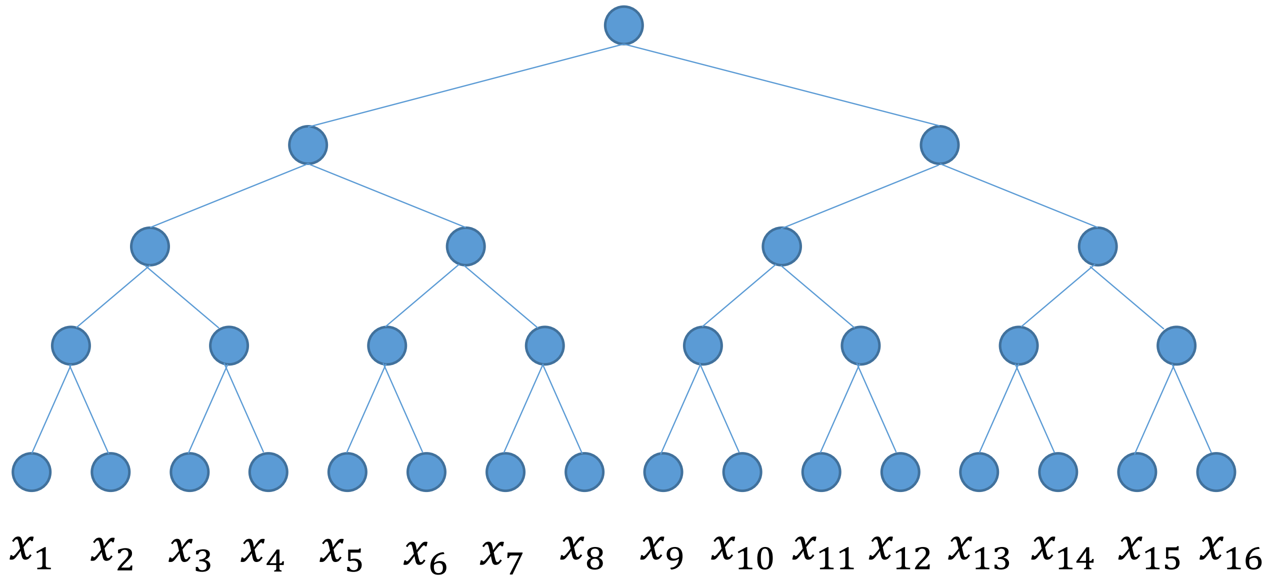
Application: online counting

- Individual values $x_1, x_2, \dots, x_T$ appear in an online manner; $x_i \in \{0,1\}$
  - Goal: online estimation of $s(t) = \sum_{i=1}^{t} x_i$
    - Observation: ($\equiv$ threshold queries) → use tree algorithm!
  - Assign individual values to tree leaves as they arrive



Application: online counting [DNPR10, CSS10]

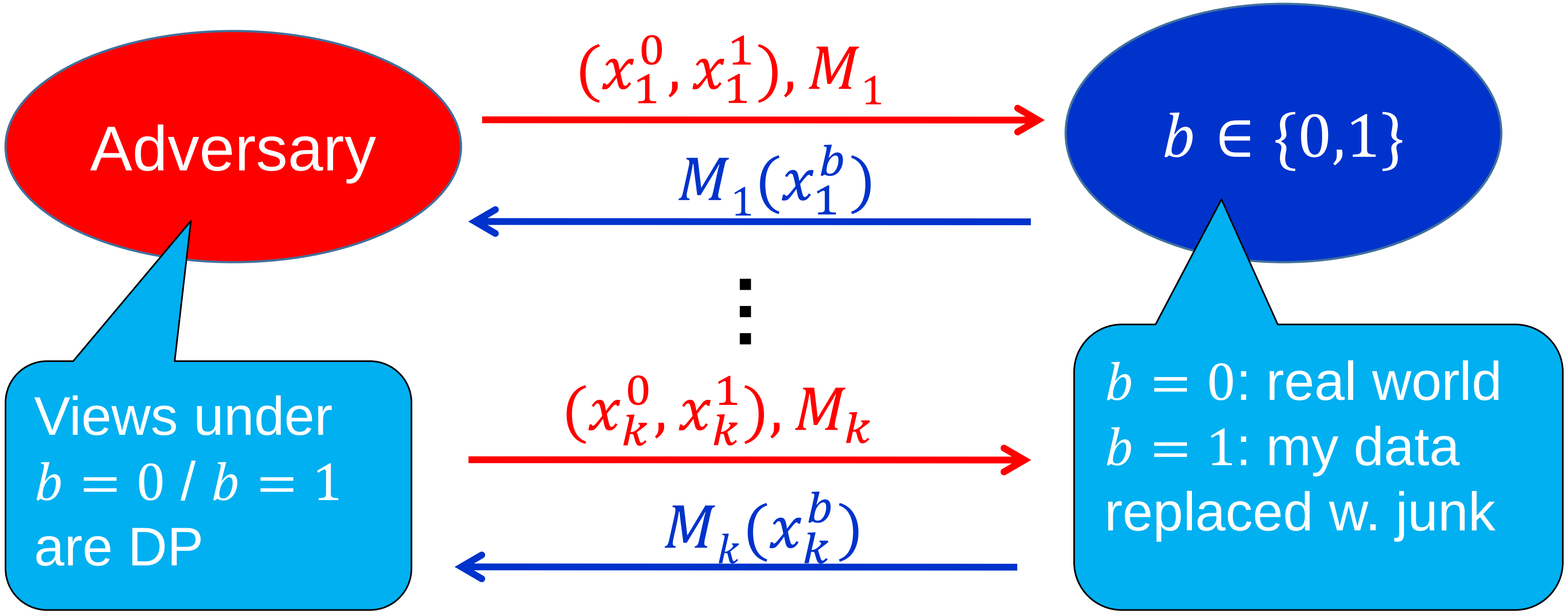
- Think: which databases are neighboring in this setting?
- Observation: Nodes ‘fill up’ before they need to be used
- Suffices to hold O(log T) counts
- Add Laplace noise once a node fills up



Advanced composition

Composition in differential privacy

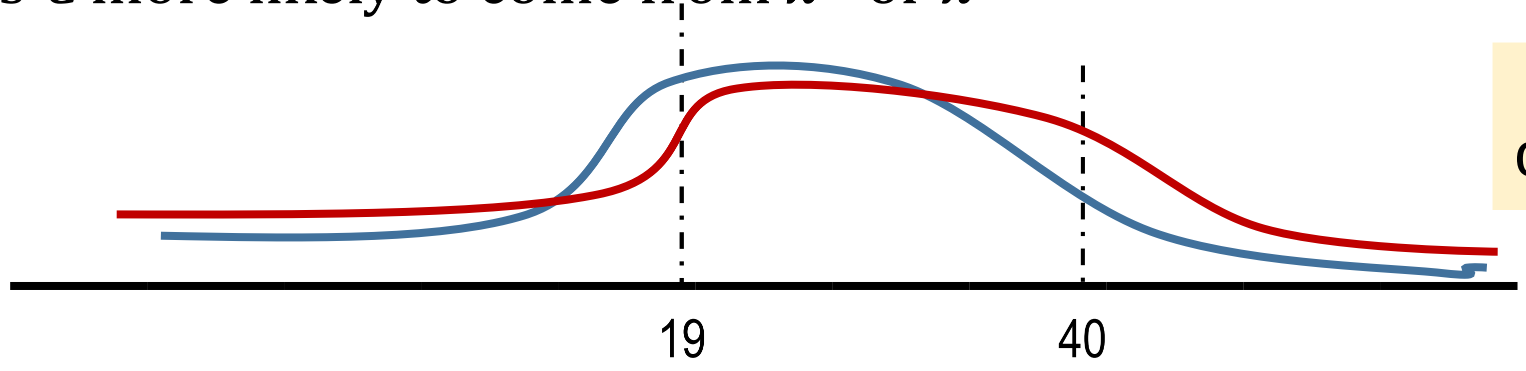
- How do we define it?
  - Both choice of databases and algorithms is adaptive and adversarial [DRV10]



Thx: Guy Rothblum

What is privacy loss?

- Measured by the 'privacy loss' parameter ϵ
- Fix adjacent x^0, x^1, draw $C \leftarrow M(x_0)$
  - Is C more likely to come from x^0 or x^1



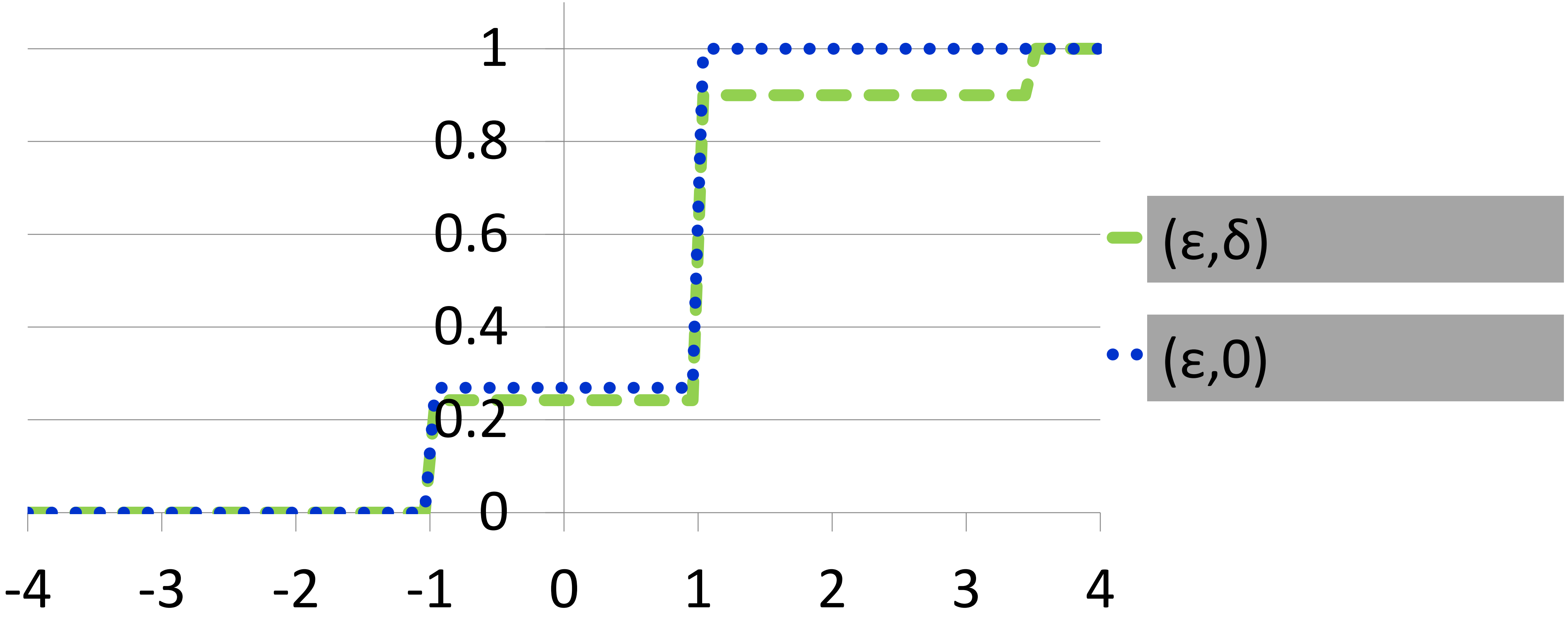
"19" more likely as output on x^0 than on x^1

"40" more likely as output on x^1 than on x^0

- Define $Loss(C) = \ln\left[\frac{\Pr[M(x^0)=C]}{\Pr[M(x^1)=C]}\right]$
  - $(\varepsilon, 0) - DP$: $w.p. 1$ $over$ C, $|Loss(C)| \leq \varepsilon$
  - $(\varepsilon, \delta) - DP^*$: $w.p. 1 - \delta$ $over$ C, $|Loss(C)| \leq \varepsilon$

Log of likelihood ratio

Comparison: Privacy Loss (cdf)



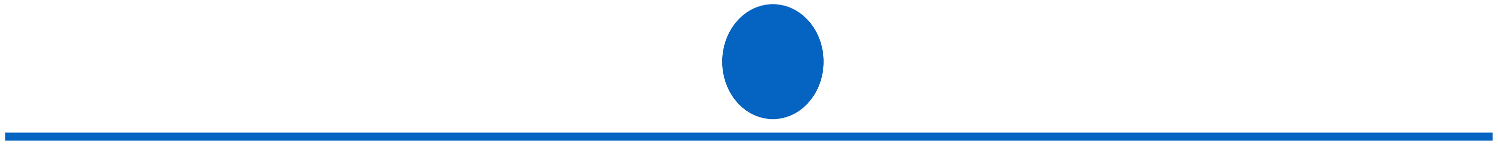
Thx: Guy Rothblum

What is privacy loss?

- Fix adjacent x^0, x^1, draw $C \leftarrow M(x_0)$

$$Loss(C) = \ln\left[\frac{\Pr\left[M\left(x^0\right) = C\right]}{\Pr[M(x^1) = C]}\right]$$

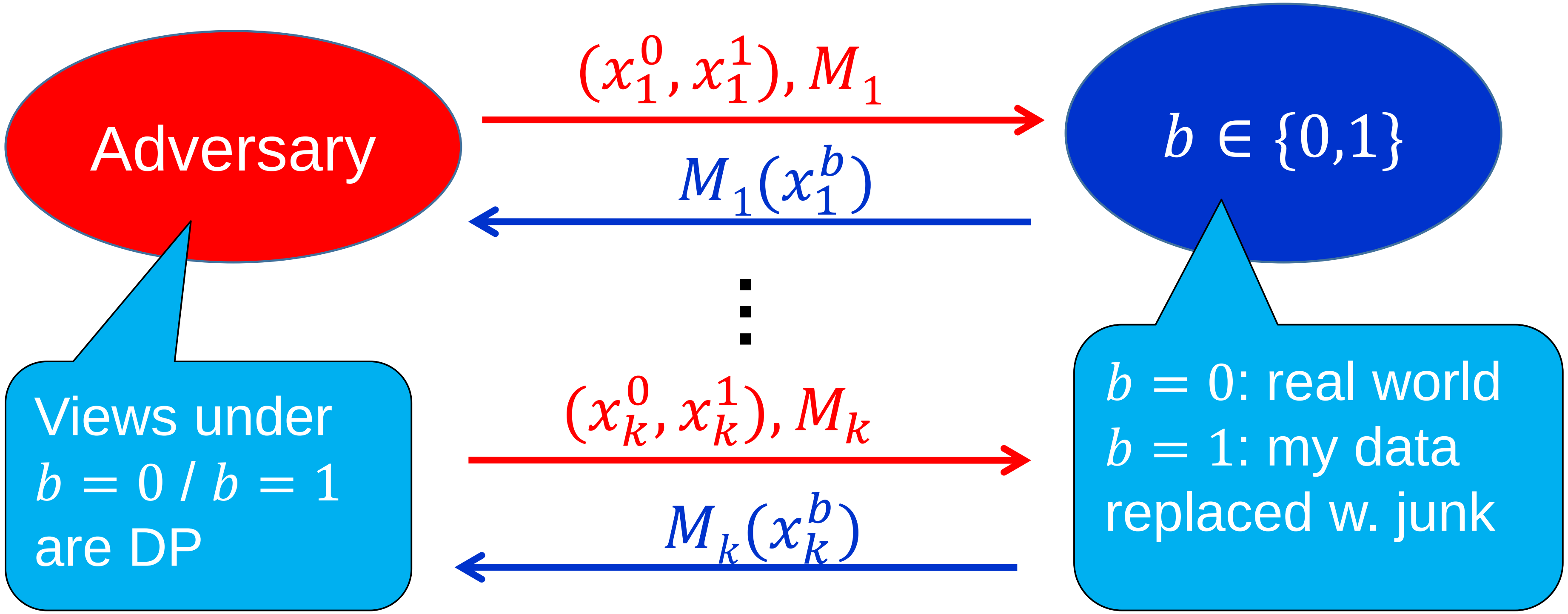
- In multiple independent executions *loss* accumulates
  - Worst case: *Loss*= ε for every execution (as in analysis of basic composition)
  - This is pessimistic: *Loss* can be positive, negative → cancellations
  - Random variable, has a mean ([DDN03, DRV10]…)



Composition in differential privacy

- Challenger has a bit *b*
- In every round *i*, Adversary specifies a differentially private encoding of *b:*
  - If *b=0* : send me $M_i(x_i^0)$
  - If b=1 : send me $M_i(x_i^1)$

Adversary would use "best" differentially private encoding of *b*



Privacy loss in randomized response

- Enough to understand how randomized response composes [KOV15, MV16]:

- $RR_\epsilon(x) = \begin{cases} x & w.p. \frac{e^\epsilon}{e^\epsilon+1} \\ \neg x & w.p. \frac{1}{e^\epsilon+1} \end{cases}$

$Loss(x) = \epsilon$

$Loss(\neg x) = -\epsilon$

With almost equal probability (if ϵ small)

- Expected Loss $= \epsilon \frac{e^\epsilon}{e^\epsilon+1} - \epsilon \frac{1}{e^\epsilon+1} = \epsilon \frac{e^\epsilon-1}{e^\epsilon+1} \approx \epsilon \frac{1+\epsilon-1}{2+\epsilon} \approx \frac{\epsilon^2}{2}$

Advanced composition - proof idea

- If M is ϵ-DP, then the Loss random variable has:
  - $E[Loss(C)] = O(\varepsilon^2)$ (down to $\varepsilon^2/2$ [DR15])
  - $|Loss(C)| \leq \varepsilon$

- Model cumulative loss from $M_1 \dots M_k$ as Martingale

$$\Pr\left[\left(\sum_{i=1}^{k} Loss(C_i)\right) > k\varepsilon^2 + \sqrt{k}\varepsilon \cdot t\right] \leq \exp(-t^2/2)$$

- Choosing $t \sim \sqrt{\log \frac{1}{\delta}}$ results in $(k\epsilon^2 + \sqrt{k \log \frac{1}{\delta}}\,\varepsilon, \delta)$-DP*

Advanced Composition [DRV10]

Composing k pure-DP algorithms (each ε_0-DP):

$$\varepsilon_g = O\left(\sqrt{k \cdot \ln \frac{1}{\delta_g}} \cdot \varepsilon_0 + k \cdot \varepsilon_0^2\right)$$ with all but δ_g probability.

For all δ_g simultaneously

Dominant if $k \ll \frac{1}{\epsilon_0^2}$

Dominant if $k \gg \frac{1}{\epsilon_0^2}$

Advanced Composition [DRV10]

Composing k algorithms, each ε_0-DP:

$$\varepsilon_g = O\left(\sqrt{k \cdot \ln\frac{1}{\delta_g}} \cdot \varepsilon_0 + k \cdot \varepsilon_0^2\right)$$ with all but δ_g probability.

For all δ_g simultaneously

- Compare with: $\varepsilon_g = k \cdot \varepsilon_0$ (basic composition)
- Better composition, better DP algorithms:
  - Answer n count queries, error $\tilde{O}(\sqrt{n \cdot \ln(1/\delta_g)})$ (independent Laplace noise)

Almost tight: Reconstruction attacks [DN03]: Must have error $\Omega(\sqrt{n})$

Composing k algorithms, each $(\varepsilon_0, \delta_0)$-DP:

$$\varepsilon_g = O\left(\sqrt{k \cdot \ln\frac{1}{\delta_{err}}} \cdot \varepsilon_0 + k \cdot \varepsilon_0^2\right)$$ with all but $\delta_g = \delta_{err} + k \cdot \delta_0$ probability.

δ grows linearly in k

Can we do better? optimal DP composition

Goal: Find best $(\varepsilon_g, \delta_g)$ for given $((\varepsilon_1, \delta_1), \ldots, (\varepsilon_k, \delta_k))$

Best worst-case result

I.e., best result– over all mechanisms, databases, events

- Homogeneous case [KOV15]
  - Tight bounds when $\forall i, \varepsilon_i = \varepsilon, \delta_i = \delta$
- Heterogeneous case [MV16]
  - Tight bounds for general ε_i, δ_i
  - Exactly computing ε_g is $\#P$-complete (unlikely to take less than $\exp(k)$ time)
  - Approximate ε_g up to additive η in time $poly(k, 1/\eta)$

Improves over [DRV10] (may be of practical significance)

Concentrated Differential Privacy [DR15,BS16]

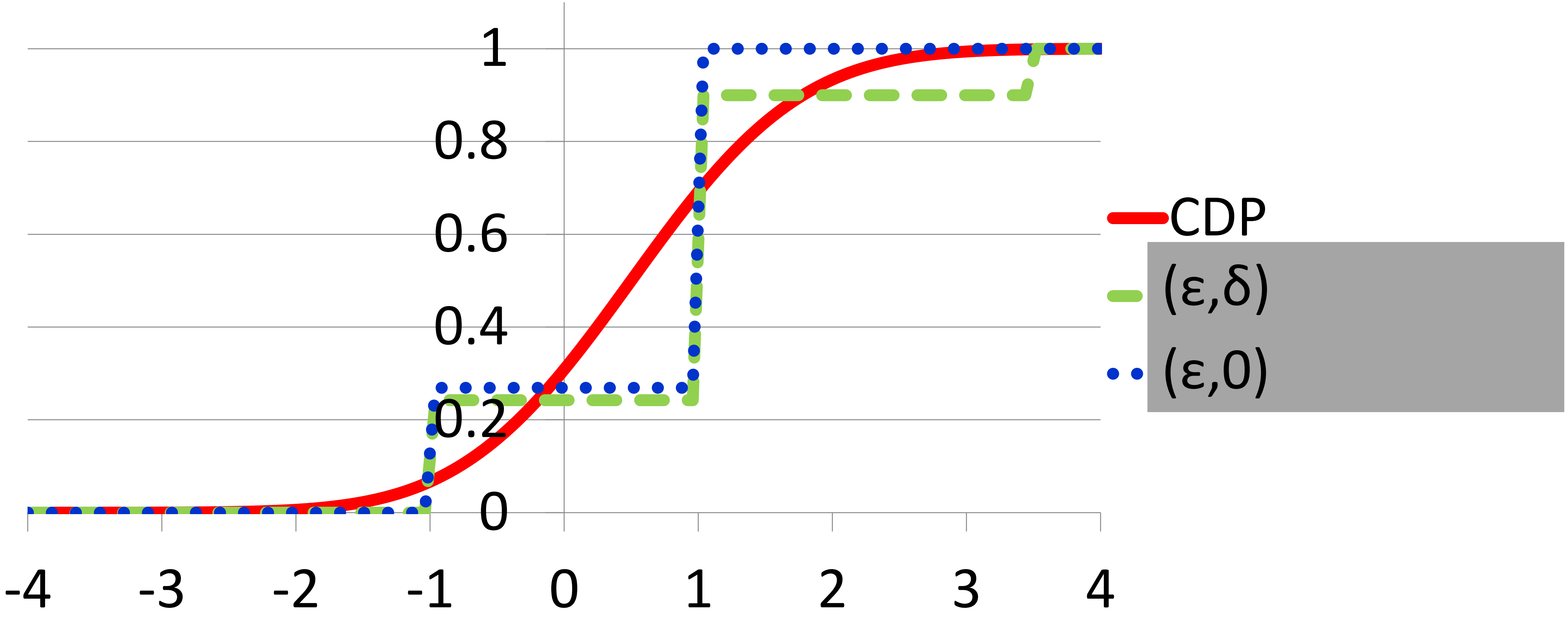
- Fix adjacent x^0, x^1, draw $C \leftarrow M(x_0)$

$$Loss(C) = \ln\left[\frac{\Pr\left[M\left(x^0\right) = C\right]}{\Pr[M(x^1) = C]}\right]$$

- (μ, τ^2)-concentrated differential privacy [DR15]
  - Intuition: $Loss(C)$ is concentrated
  - $\mathrm{E}_{C \leftarrow M(D)}[Loss(C)] \leq \mu$
  - concentration “no worse than” Gaussian$\left(\mu, \tau^2\right)$

Alternative : bound Renyi divergences [BS16]

Comparison: Privacy Loss (cdf)



Thx: Guy Rothblum

Concentrated Differential Privacy

Intuition: privacy loss "no worse than" $N(\mu, \tau^2)$

Formally: privacy loss is *Subgaussian* random variable, rich theory to draw on

- $\mathrm{E}_{C \leftarrow M(D)}[Loss(C)] \leq \mu$
- $(Loss(C) - \mu)$ is "Subgaussian"
- $\Pr[|Loss(C) - \mu| \geq t \cdot \tau] \leq e^{-t^2/2}$

Maintains many advantages of differential privacy:

- Composes automatically
  Addition of Gaussians is Gaussian: μ and τ^2 add up
- Handles linkage / auxiliary data
  (similarly to standard differential privacy)

Concentrated Differential Privacy Summary: Improved Utility, Relaxed Privacy

Privacy (CDP vs. $(\boldsymbol{\varepsilon}, \boldsymbol{\delta})$-DP)

- Per study: somewhat weaker/relaxed guarantee
- Composition over many studies: (roughly) identical behavior!

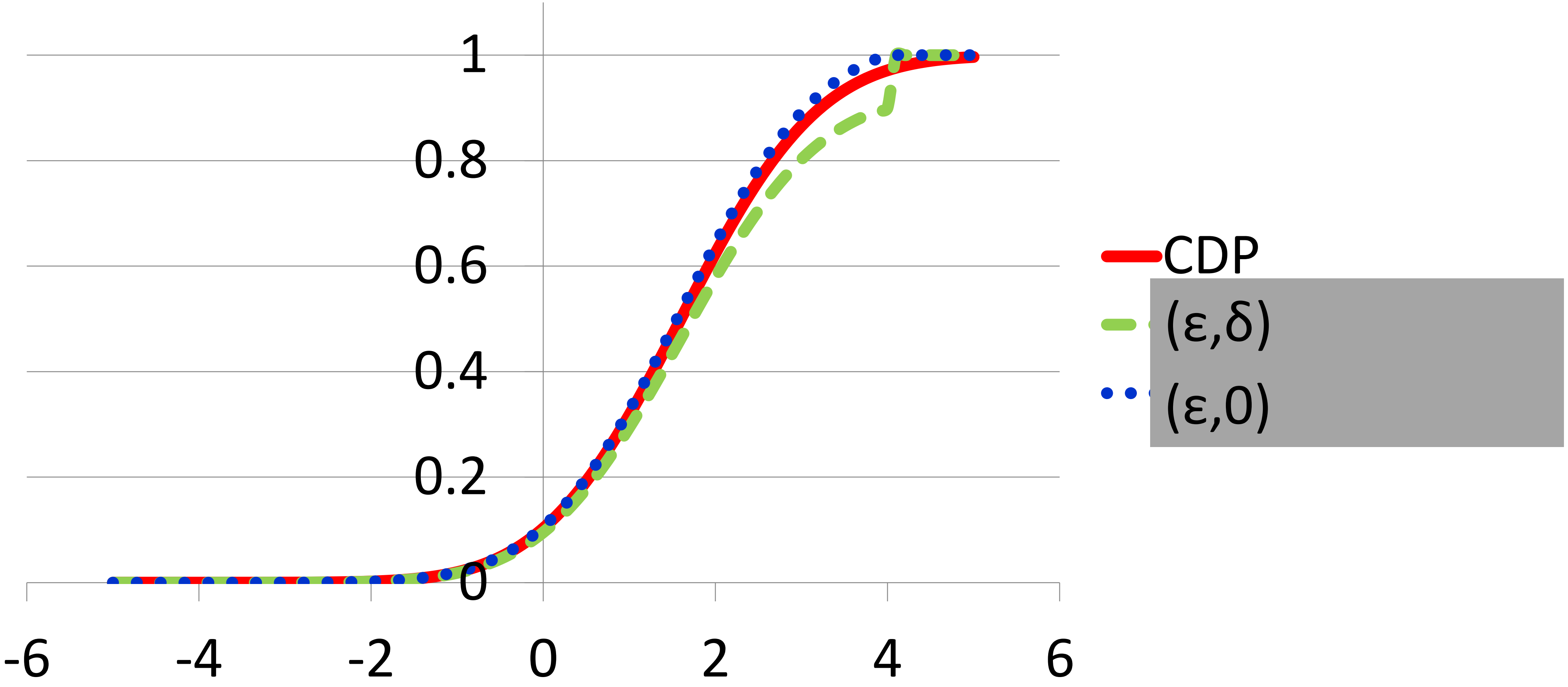
Accuracy (answering k queries, $\varepsilon = 1$)

- $(\varepsilon, 0)$-DP: noise $\approx k$
- (ε, δ)-DP: noise $\approx \sqrt{k \cdot \ln(1/\delta)}$
- $\left(\varepsilon^2/2, \varepsilon^2\right)$ −CDP: noise $\approx \sqrt{k}$

Reconstruction attacks [DN03]: Must have error $\Omega(\sqrt{n})$

Factor of $\sqrt{\ln 1/\delta}$ can be significant in applications

Comparison: Composed Privacy Loss (cdf)

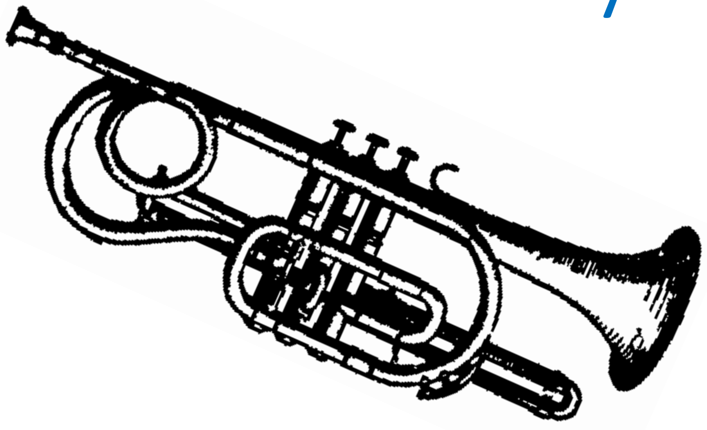


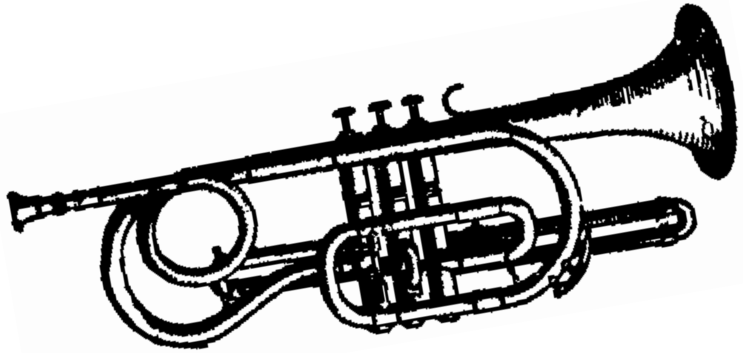
Thx: Guy Rothblum

Summary

- Adaptive composition important for privacy, algorithm design, data analysis

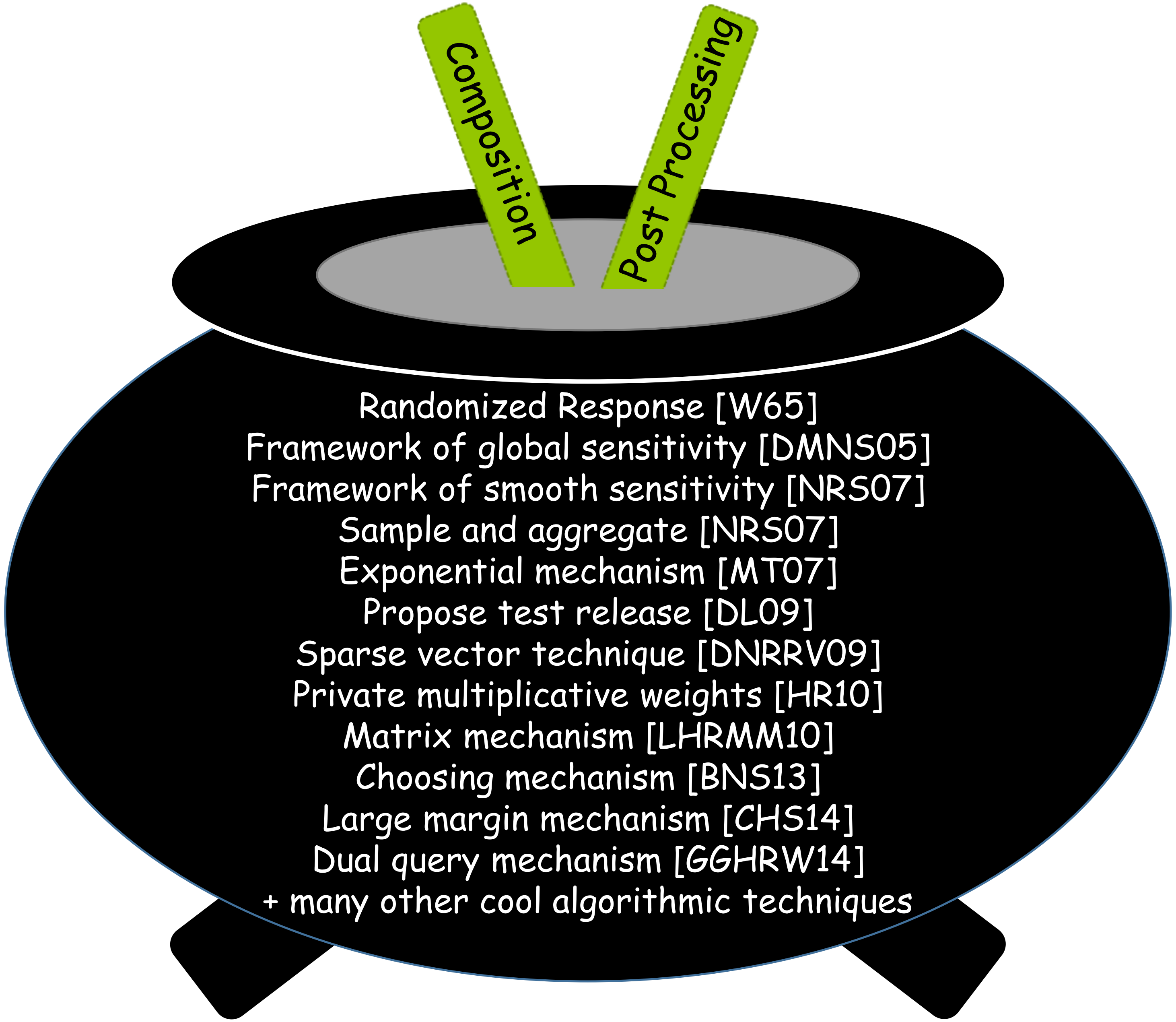
Many Ways of Making (Less) Noise





Randomized Response [W65]
Framework of global sensitivity [DMNS06]
Framework of smooth sensitivity [NRS07]
Sample and aggregate [NRS07]
Exponential mechanism [MT07]
Propose test release [DL09]
Sparse vector technique [DNRRV09]
Private multiplicative weights [HR10]
Matrix mechanism [LHRMM10]
Choosing mechanism [BNS13]
Large margin mechanism [CHS14]
Dual query mechanism [GGHRW14]
+ many other cool algorithmic techniques

A Programmable Framework:



Summary

- Adaptive composition important for privacy, algorithm design, data analysis
- Variety of composition theorems
  - Basic composition
  - Advances composition
  - Optimal composition
- ϵ treated as a “privacy budget”
- Concentrated differential privacy

References

- Dwork, Naor, Pitassi, Rothblum: Differential privacy under continual observation. 2010
- Chan, Shi, Song: Private and Continual Release of Statistics. 2010
- Dwork, Rothblum, Vadhan: Boosting and dierential privacy. 2010
- Dwork, Rothblum: Concentrated Differential Privacy. 2016
- Kairouz, Oh, Viswanath: The Composition Theorem for Differential Privacy. 2015
- Murtagh, Vadhan: The Complexity of Computing the Optimal Composition of Differential Privacy. 2016
- Bun, Steinke: Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds. 2016