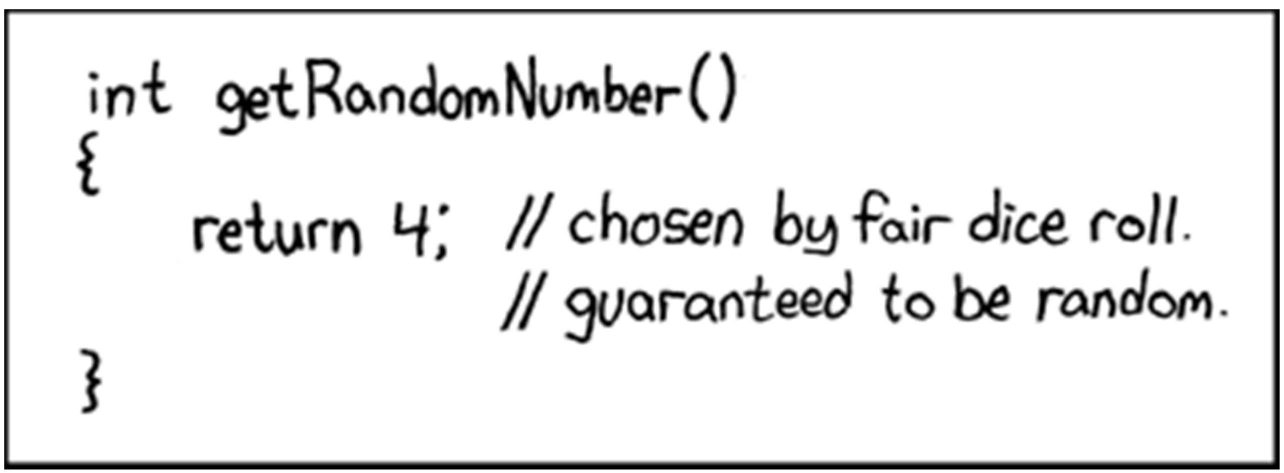
Differential Privacy with Imperfect Randomness

Yevgeniy Dodis
Adriana López-Alt
Ilya Mironov
Salil Vadhan

Randomness in Cryptography

- Cryptographic algorithms **require** randomness.
  - Secret keys must have entropy
  - Many primitives must be randomized (Enc, Com, ZK, etc.)
- Common to assume **perfect** randomness is available
- But real-world randomness is **imperfect**.



Randomness in Cryptography

- Cryptographic algorithms **require** randomness.
  - Secret keys must have entropy
  - Many primitives must be randomized (Enc, Com, ZK, etc.)
- Common to assume **perfect** randomness is available
- But real-world randomness is **imperfect**.
  - $\in \subset \subseteq \cap \cup \supset \supseteq \emptyset \pm \varepsilon \gamma$
- Main Question: Can we base cryptography on (realistic) **imperfect** randomness?

Imperfect Sources

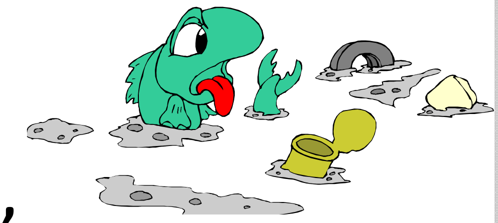
- **Imperfect source S**: family of distributions **R** satisfying some property (i.e., entropy)
- "Tolerate" imperfect source: have **one** scheme correctly working for **any** **R** in the source **S**

Main Question: **(restated) Which imperfect sources are enough for cryptography?**

Extractable Sources

- Sources permitting (deterministic) extraction of nearly perfect randomness [vNeu, Eli, …]
- **Example: von Neumann's extractor**
  - Independent coins, all with (unknown) bias **p**.
  - Obtain uniform distribution by:
    - **HT → 0**
    - **TH → 1**
- Suffice for (almost) anything possible with perfect randomness
- **Bad news**: many sources are non-extractable ☹

Non-Extractable Sources



- Obvious: sources with no "entropy"
  - Clearly, cannot do crypto as well
- **What about "entropy" (weak) sources?**
  - Generally non-extractable [SV85,CG89] ☹
  - Simplest example: γ-Santha-Vazirani sources – **SV(γ)**
    - Produces bits **b_1, b_2, …** , each having bias at most **γ** (possibly dependent on prior bits).

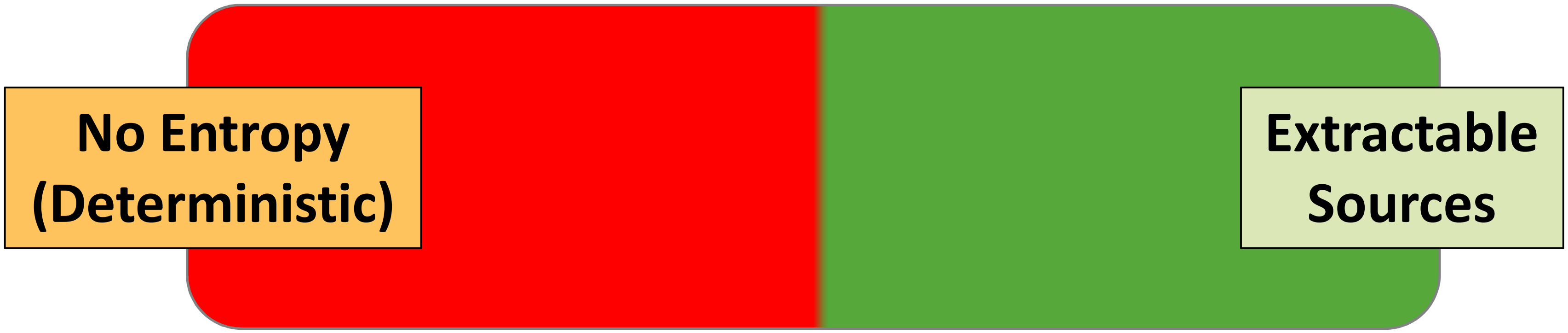
$$\frac{1}{2}\cdot(1-\gamma)\le \Pr[b_i=0\mid b_1b_2\ldots b_{i-1}]\le\frac{1}{2}\cdot(1+\gamma)$$

    - **Non-extractable**: for any **f: $\{0,1\}^n \to \{0,1\}$**, there exists a **SV(γ)** source s.t. **f(SV(γ))** has bias at least **γ**.

Randomness in Cryptography

Cryptography is Impossible

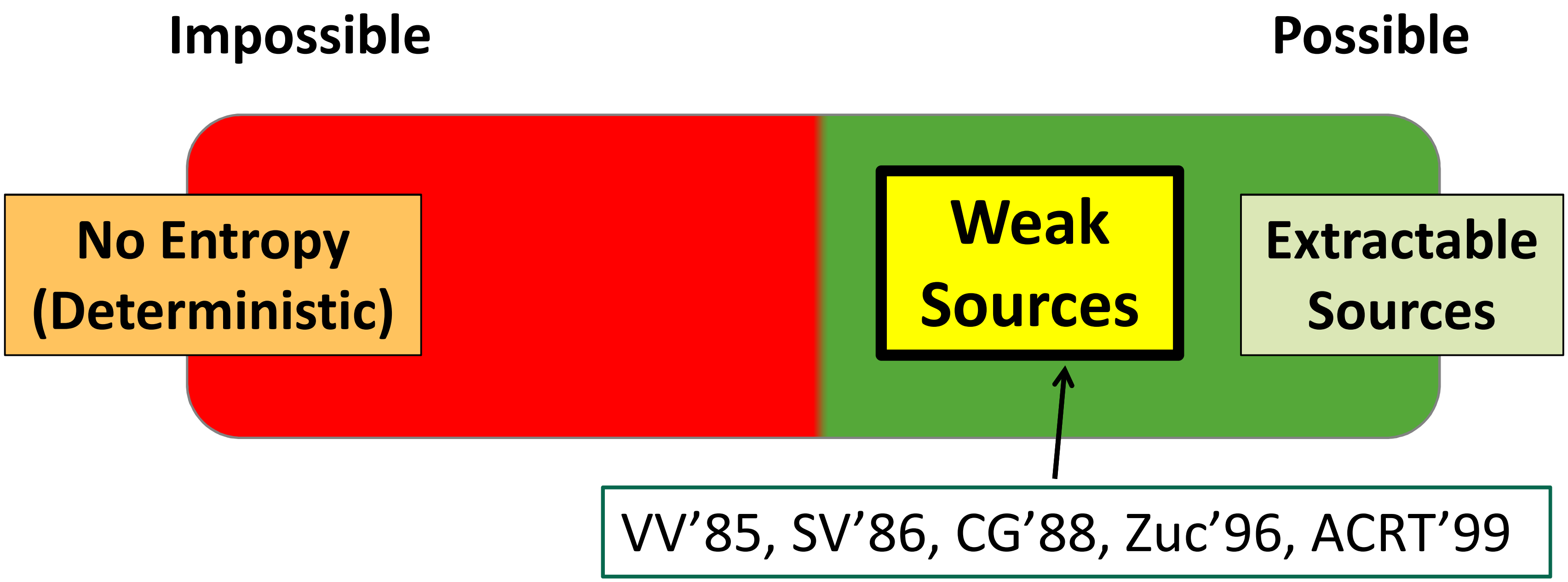
Cryptography is Possible



General (Weak) Entropy Sources?

(Depends on Application)

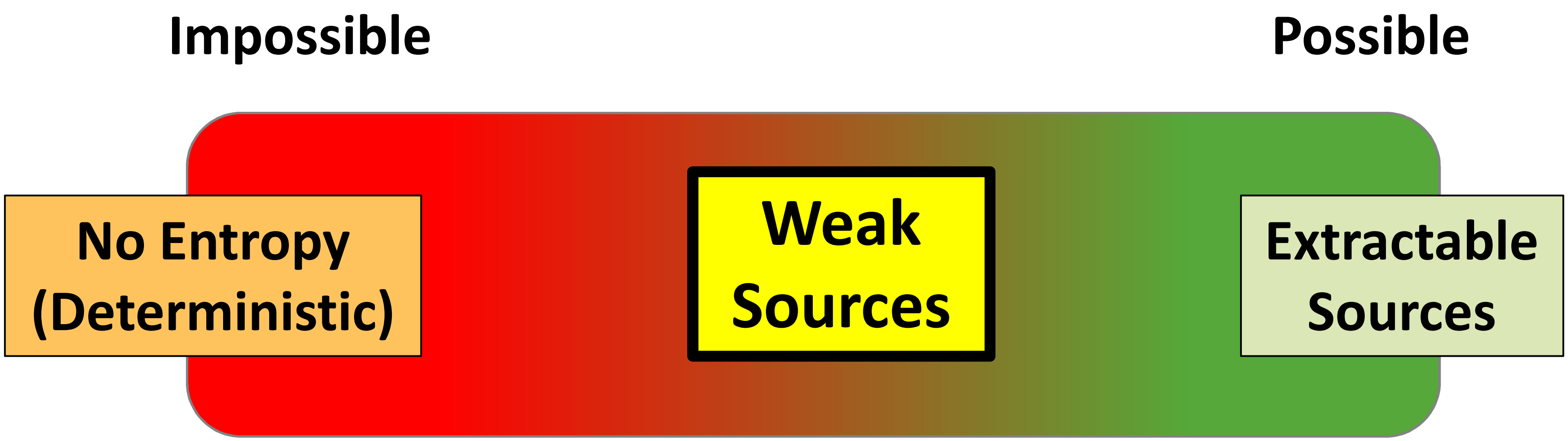
BPP Simulation



Same good news for Crypto?

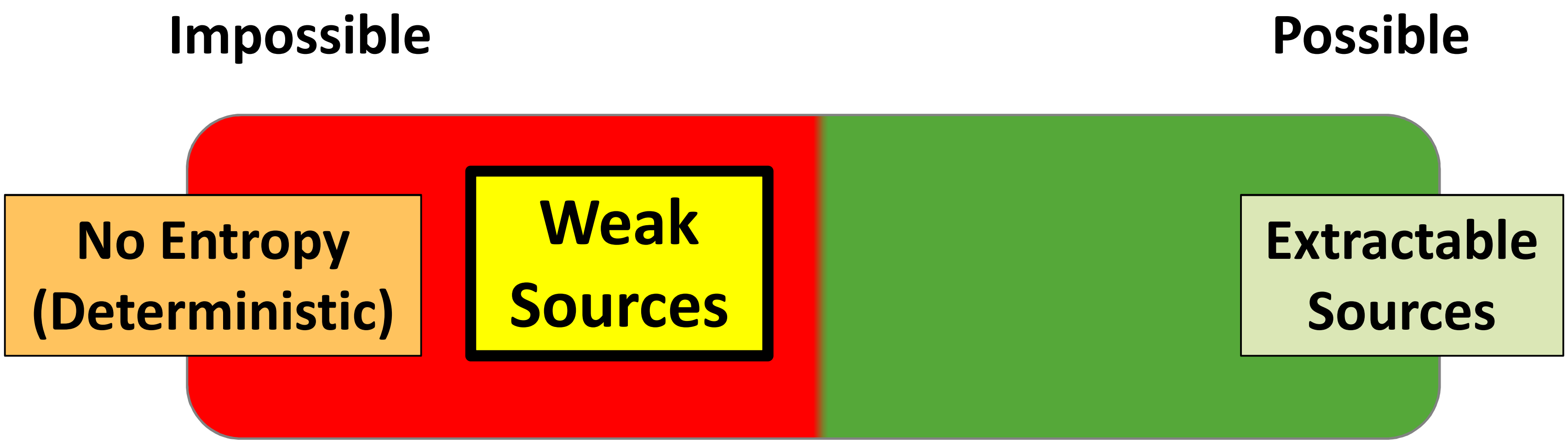
- Authentication (MACs, Sig)
- Privacy/Secrecy (Enc, Com, ZK)

Authentication (MACs, Sigs)



- Many (but not all [DS02]) weak sources are **sufficient** for:
  - **MACs** [MW'97, DKRS'06]
  - **Signature Schemes** [DOPS'04] – under appropriate hardness assumptions.
- **Intuition:** only require that it is hard to guess ("forge") a long string, so having (min-)entropy suffices

Privacy/Secrecy (Enc, Com, ZK)



- **SV(γ)** *not* sufficient for:
  - Unconditionally-secure encryption (MP'90)
  - Computationally-secure encryption (DOPS'04)
  - Commitment, Zero-Knowledge, Secret-Sharing (DOPS'04)
- **BD'07**: If can generate **k**-bit SK from source **R**, can extract **k** almost uniform bits from **R**.
  - **Traditional privacy requires an extractable source.**

Privacy/Secrecy (Enc, Com, ZK)

DOPS'04 Main Lemma: Let **X** be a "weak source".
If $\mathbf{f(X) \approx_c g(X)}$, then $\mathbf{Pr_{x \leftarrow U}[f(x) \neq g(x)] = negl(k)}$

- Reason: We require adversary to have a **negligible** advantage in distinguishing (e.g. $\mathbf{Enc(0) \approx_c Enc(1)}$)
- **Can privacy/secrecy be based on weak (e.g., SV) sources if we (naturally) relax the security definition?**
  - E.g. consider Differential Privacy

Differential Privacy (Dwork'06, DMNS'06)

- Database **D**: Array of rows.
- Queries **f(D) → Z**
  - Low sensitivity queries – answer does not change by much on neighboring databases.

D_1 D_2 differ in **1** entry.

A mechanism **M** is **ε-differentially private** for **F** w.r.t. source **S** if for all queries **f** ∈ **F**, all neighboring databases **D_1 D_2**, all distributions **R** ∈ **S**, and all possible outcomes **z**:

$$\frac{\Pr_{r\leftarrow R}[M(D_1, f; r) = z]}{\Pr_{r\leftarrow R}[M(D_2, f; r) = z]} \le 1 + \varepsilon$$

Differential Privacy (Dwork'06, DMNS'06)

- Notice, ε <u>cannot</u> be negligible
  - Implies output of mechanism is negligibly close on <u>any</u> two **different** databases – **not useful**.
  - Hope to overcome impossibility result of DOPS'04.

A mechanism **M** is **ε-differentially private** for **F** w.r.t. source **S** if for all queries **f** ∈ **F**, all neighboring databases D_1 D_2, all distributions **R** ∈ **S**, and all possible outcomes **z**:

$$\frac{\Pr_{r \leftarrow R}[M(D_1, f; r) = z]}{\Pr_{r \leftarrow R}[M(D_2, f; r) = z]} \leq 1 + \varepsilon$$

Utility

A mechanism **M** has **ρ-utility** for **F** w.r.t. **S** if for all databases **D**, all queries **f** ∈ **F**, all distributions **R** ∈ **S**:

$$E_{r \leftarrow R}\left[\left|f(D) - M(D, f; r)\right|\right] \leq \rho$$

A mechanism **M** is **ε-differentially private** for **F** w.r.t. source **S** if for all queries **f** ∈ **F**, all neighboring databases $\mathbf{D_1}$ $\mathbf{D_2}$, all distributions **R** ∈ **S**, and all possible outcomes **z**:

$$\frac{\Pr_{r \leftarrow R}[M(D_1, f; r) = z]}{\Pr_{r \leftarrow R}[M(D_2, f; r) = z]} \leq 1 + \varepsilon$$

Accurate and Private Mechanisms

Can we achieve a good **<u>tradeoff</u>** between privacy and utility?

"non-trivial"

F admits ~~**accurate and private**~~ mechanisms w.r.t. **S** if for all $\varepsilon > 0$ there is M_ε that is ε-DP and has $g(\varepsilon)$ utility w.r.t **S**, for some $g(.)$

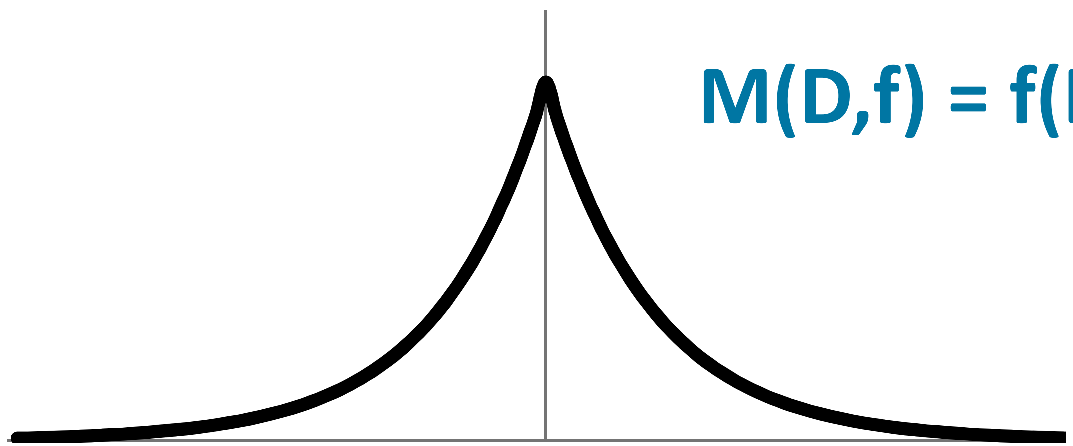
Additive-Noise Mechanisms (ANM)

Not too "sensitive" on neighboring **D**

$$M(D,f\,;\,r) = f(D) + X_\varepsilon(r)$$

appropriate "noise" distribution

- (DN'03, DN'04, BDMN'05, DMNS'06, GRS'09, HT'10)
- E.g. Add **Laplacian** noise (DMNS'06)



$$M(D,f) = f(D) + \mathrm{Lap}(1/\varepsilon)$$

- **ε**-differentially private and has **Θ(1/ε)**-utility w.r.t. **U**
  - Hence, "non-trivial" w.r.t. **U**

Our Question

Are weak entropy sources sufficient to achieve “non-trivial” mechanisms?

- Most surprising, **positive** result
  - “Non-trivial” “SV-robust” mechanisms for low-sensitivity functions
- **<u>Separation</u>** between **traditional** and **differential** privacy

First Attempt

Hope: Any class of "non-trivial" mechanisms w.r.t. **U** is also "non-trivial" w.r.t. **SV(γ)**.

Too optimistic:

- See paper for a "dramatic" (but artificial) example.
- Natural example: additive-noise, $M(D,f\,;R) = f(D) + X(R)$
  - Can show if any ANM **M** is ε-DP then $X'(R) = X(R) \bmod 2$ is a ε-biased one-bit extractor for **R**.
  - **SV(γ)** is "non-extractable" – i.e. cannot extract ε-biased bit for $\varepsilon < \gamma$
  - Thus, **no ANMs** can be "non-trivial" w.r.t. **SV(γ)**

Second Attempt

Hope: Any class of "non-trivial" mechanisms w.r.t. **U** is also "non-trivial" w.r.t. **SV(γ)** *if we first run an "extractor" on the randomness.*

Also doesn't work:

- Applying **Ext** to ANM is still ANM
  - **M'(D,f ; R) = f(D) + X(Ext(R))**
- ANMs are **<u>not</u>** "SV-robust".

Conclusion:

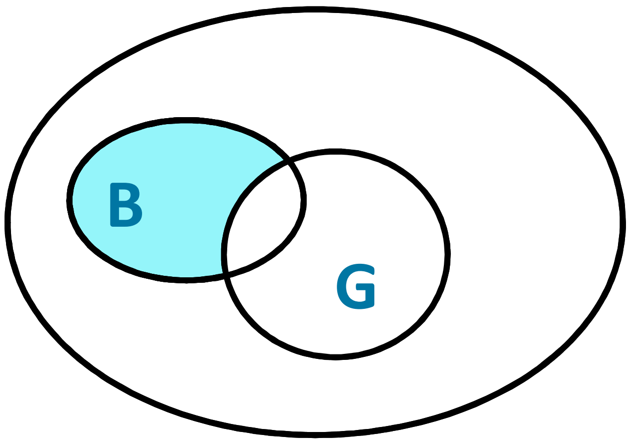
- Need a **<u>non-</u>**additive-noise mechanism.

A General Lower Bound

First, a useful **Lemma**:

- Sets **G, B ⊏ $\{0,1\}^n$** s.t. **|G| ≥ |B| > 0**
- Define $$\sigma = \frac{|B \setminus G|}{|B|}$$
- There exists distribution **SV(γ)** s.t.

$$\boxed{\frac{\Pr_{r \leftarrow SV(\gamma)}[r \in G]}{\Pr_{r \leftarrow SV(\gamma)}[r \in B]} \geq (1 + \gamma\sigma)}$$



A General Lower Bound

- Fix neighboring databases D_1,D_2, query **f** and outcome **z**
- Define S_b= {r | M(D_b,f ;r) = z}
  (i.e., set of coins that make **M** output **z** on D_b)

$$\frac{\Pr_{r \leftarrow SV(\gamma)}[M(D_1, f; r) = z]}{\Pr_{r \leftarrow SV(\gamma)}[M(D_2, f; r) = z]} = \frac{\Pr_{r \leftarrow SV(\gamma)}[r \in S_1]}{\Pr_{r \leftarrow SV(\gamma)}[r \in S_2]} \geq (1 + \gamma\sigma)$$

By lemma

$$\sigma = \frac{|S_2 \setminus S_1|}{|S_2|}$$

Conclusion:

- **ε**-DP w.r.t. **SV(γ)** **requires** **σ ≤ ε/γ = O(ε)**
- $S_1 \sqcap S_2$ must be “big” – a **1 – ε** fraction of S_1.

Consistent Sampling (Man'94, Hol'07, MMP+'10)

A mechanism **M** has **ε-consistent sampling** if for all queries **f** $\in$ **F**, all neighboring databases **D_1 D_2**, and all possible outcomes **z**: $\frac{|S_1 \setminus S_2|}{|S_2|} \le \varepsilon$

Lemma: If M is ε-consistent, then **M** is ε-DP w.r.t. **U**

Proof:

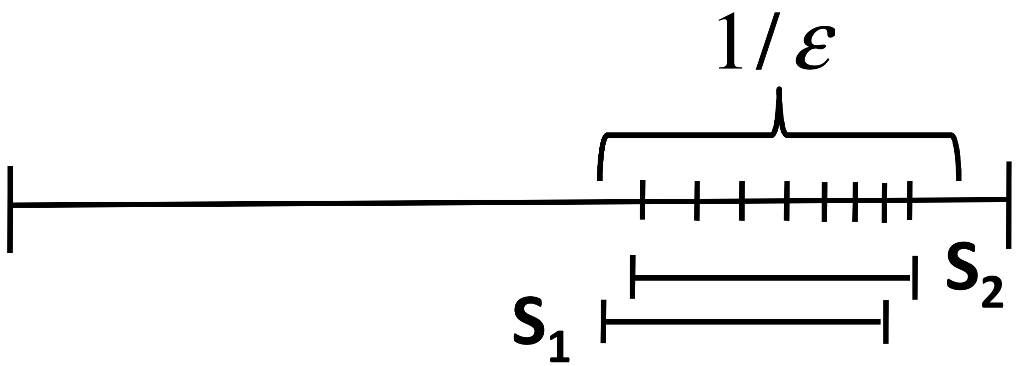
$$\frac{\Pr_{r \leftarrow U_n}[M(D_1, f; r) = z]}{\Pr_{r \leftarrow U_n}[M(D_2, f; r) = z]} = \frac{\Pr_{r \leftarrow U_n}[r \in S_1]}{\Pr_{r \leftarrow U_n}[r \in S_2]}$$

$$= \frac{|S_1|}{|S_2|} = \frac{|S_1 \cap S_2|}{|S_2|} + \frac{|S_1 \setminus S_2|}{|S_2|} \le 1 + \varepsilon$$

A New Mechanism

$$\boxed{M(D,f) = [f(D) + \mathrm{Lap}(1/\varepsilon)]_{1/\varepsilon}}$$

- Round outcome to nearest multiple of **$1/\varepsilon$**
  - Utility is conserved (asymptotically): still **$\Theta(1/\varepsilon)$**-utility
- Guarantees **S_1, S_2** will intersect on a large fraction of coins, as required for **ε**-consistent sampling.



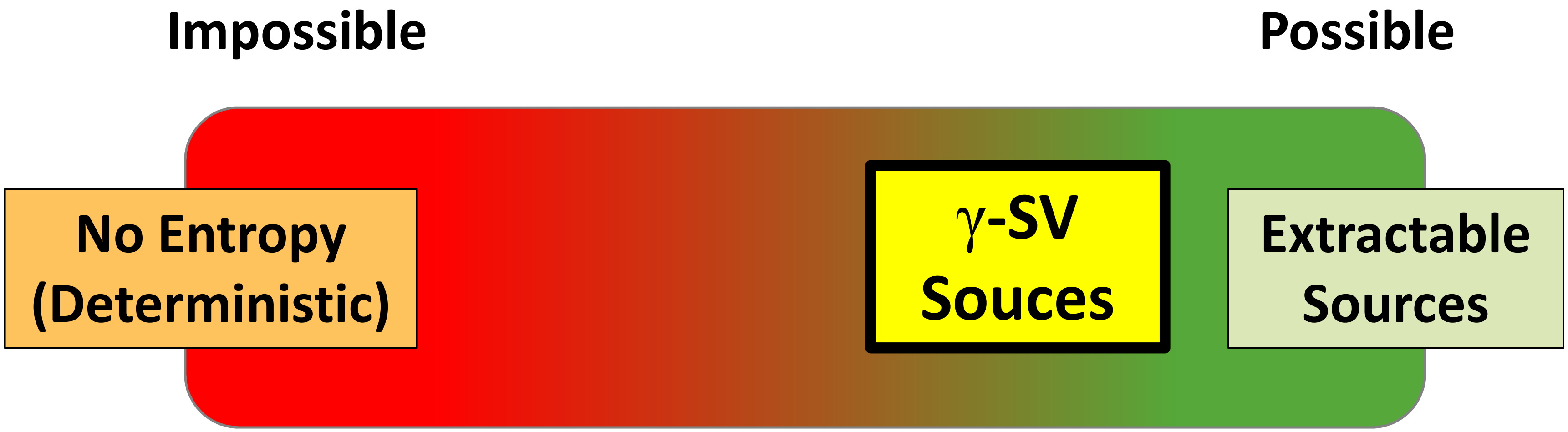
A New Mechanism $\boxed{M(D,f) = [f(D) + Lap(1/\varepsilon)]_{1/\varepsilon}}$

- Satisfies **ε**-consistent sampling.
- Overcomes our lower bound.

Can we implement it in a "SV-robust" manner?

- Yes! But **non-trivial** (no pun intended ☺)
  - Not every implementation is "SV-robust"
  - **ε**-consistent sampling is **necessary** but **not sufficient**
- Define **ε-SV-consistent sampling**
  - Natural definition, does not reference **SV(γ)**
  - **Sufficient** for "SV robustness"
  - Use **arithmetic coding** to ensure SV-consistency
    - Need to be careful with **finite precision**

Differential Privacy – Our Results



- Any “SV-robust” **ε**-DP mechanism:
  - **Must** satisfy **ε**-consistent sampling
  - **Enough** to satisfy **ε**-SV-consistent sampling
- We show a “non-trivial” (accurate and private) “SV-robust” family of mechanisms for low sensitivity queries.

Thank you!

Weak Sources **?**