Differential Privacy with Imperfect Randomness

Yevgeniy Dodis
New York University

Ilya Mironov
Microsoft Research

Adriana López-Alt
New York University

Salil Vadhan
Harvard University
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.

```c
int getRandomNumber()
{
    return 4;  // chosen by fair dice roll.
    // guaranteed to be random.
}
```
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.

Main Question: Can we base cryptography on (realistic) imperfect randomness?
Imperfect Sources

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

**Main Question (Restated):** What imperfect sources are enough for cryptography?
Extractable Sources

- Sources permitting (deterministic) extraction of nearly perfect randomness [vNeu, Eli’72, Blum’85 ...]
- 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

- **What about “entropy” (weak) sources?**
  - Generally non-extractable [SV85,CG89] 😞
  - Simplest example: $\gamma$ - Santha-Vazirani sources – $SV(\gamma)$
    - Produces bits $b_1, b_2, \ldots$, each having bias at most $\gamma$ (possibly dependent on prior bits).
    
    \[
    \frac{1}{2} \cdot (1 - \gamma) \leq \Pr[b_i = 0 \mid b_1 b_2 \ldots b_{i-1}] \leq \frac{1}{2} \cdot (1 + \gamma)
    \]

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

Cryptography is Impossible

No Entropy (Deterministic)

Cryptography is Possible

Extractable Sources

General (Weak) Entropy Sources?

( Depends on Application)
BPP Simulation

Impossible

No Entropy (Deterministic)

Weak Sources

Extractable Sources

Possible

VV’85, SV’86, CG’88, Zuc’96, ACRT’99

Same good news for Crypto?
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 $R$, 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 $f(X) \approx_c g(X)$, then $\Pr_{x \leftarrow U}[f(x) \neq g(x)] = \text{negl}(k)$

- We require adversary to have a negligible advantage in distinguishing (e.g. $\text{Enc}(0) \approx_c \text{Enc}(1)$)

- 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.
  - Neighboring databases - \( D_1 \) \( D_2 \) differ in 1 entry.
- Queries \( f(D) \rightarrow Z \)
  - Low sensitivity queries – answer does not change by much on neighboring databases.

A mechanism **M** is \( \varepsilon \)-differentially private w.r.t. source **S** if for all neighboring databases \( D_1 \) \( D_2 \), all distributions \( R \in S \), and all possible outcomes \( z \):

\[
\frac{\Pr_{r \leftarrow R} [M(D_1, f; r) = z]}{\Pr_{r \leftarrow R} [M(D_2, f; r) = z]} \leq e^\varepsilon \approx 1 + \varepsilon
\]
Differential Privacy [Dwork’06, DMNS’06]

- Notice, $\epsilon$ cannot be negligible
  - Implies output of mechanism is negligibly close on any two different databases – not useful.
  - Hope to overcome impossibility result of DOPS’04.

A mechanism $M$ is $\epsilon$-differentially private w.r.t. source $S$ if for all neighboring databases $D_1$ $D_2$, all distributions $R \in S$, and all possible outcomes $z$:

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

A mechanism $M$ has $\rho$-utility w.r.t. source $S$ if for all databases $D$ and all distributions $R \in S$:

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

A mechanism $M$ is $\varepsilon$-differentially private w.r.t. source $S$ if for all neighboring databases $D_1, D_2$, all distributions $R \in S$, and all possible outcomes $z$:

$$\frac{\Pr_{r \leftarrow R}[M(D_1, f; r) = z]}{\Pr_{r \leftarrow R}[M(D_2, f; r) = z]} \leq e^\varepsilon \approx 1 + \varepsilon$$
Accurate and Private Mechanisms

Can we achieve a good **tradeoff** between privacy and utility?

Family of mechanisms is **accurate and private** w.r.t. source $S$ if for all $\epsilon > 0$ there is $M_\epsilon$ that is $\epsilon$-DP and has $g(\epsilon)$ utility w.r.t $S$, for some $g(.)$.
Additive-Noise Mechanisms (ANM)

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

- \([DN’03, DN’04, BDMN’05, DMNS’06, GRS’09, HT’10]\)
- E.g. Add \textbf{Laplacian} noise[DMNS’06]

\[ M(D, f) = f(D) + \text{Lap}(1/\varepsilon) \]
\[ M(D, f; r) = f(D) \pm \log(r)/\varepsilon \]

- \(\varepsilon\)-differentially private and has \(\Theta(1/\varepsilon)\)-utility w.r.t. \(U\)
- Hence, “non-trivial” w.r.t. \(U\)
Our Question

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

- **Impossible**
  - No Entropy (Deterministic)

- **Possible**
  - $\gamma$-SV Sources
  - Extractable Sources

- **Negative** result
  - Additive-noise mechanisms cannot be “non-trivial” w.r.t. $SV(\gamma)$

- Most surprising, **positive** result
  - “Non-trivial” “SV-robust” mechanisms for low-sensitivity functions

- **Separation** between traditional and differential privacy
A General Lower Bound

First, a useful **Lemma**:

- Sets $T_1, T_2 \subset \{0,1\}^n$ s.t. $|T_1| \geq |T_2| > 0$
- Define
  $$\sigma = \frac{|T_2 \setminus T_1|}{|T_2|}$$
  **Degree of disjointness**
  - Disjoint: $\sigma = 1$
  - Contained: $\sigma = 0$

- There exists distribution $SV(\gamma)$ s.t.

  $$\frac{\Pr_{r \leftarrow SV(\gamma)}[r \in T_1]}{\Pr_{r \leftarrow SV(\gamma)}[r \in T_2]} \geq (1 + \gamma \sigma) \cdot \frac{|T_1|}{|T_2|} \geq 1 + \gamma \sigma$$

  **Factor by which $SV(\gamma)$ can increase ratio.**
A General Lower Bound

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

$$\frac{\Pr_{r \leftarrow SV(\gamma)}[M(D_1, f ; r) = z]}{\Pr_{r \leftarrow SV(\gamma)}[M(D_2, f ; r) = z]} = \frac{\Pr_{r \leftarrow SV(\gamma)}[r \in T_1]}{\Pr_{r \leftarrow SV(\gamma)}[r \in T_2]} \geq (1 + \gamma \sigma)$$

By lemma

In additive-noise mechanisms:
- $T_1, T_2$ disjoint, so $\sigma = 1$
- Explains why cannot have $\varepsilon$-DP for $\varepsilon < \gamma$
A General Lower Bound

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

\[
\frac{\Pr_{r \leftarrow SV(\gamma)}[M(D_1, f; r) = z]}{\Pr_{r \leftarrow SV(\gamma)}[M(D_2, f; r) = z]} = \frac{\Pr_{r \leftarrow SV(\gamma)}[r \in T_1]}{\Pr_{r \leftarrow SV(\gamma)}[r \in T_2]} \geq (1 + \gamma \sigma)
\]

By lemma

Conclusion:
- $\epsilon$-DP w.r.t. $SV(\gamma)$ requires $\sigma \leq \epsilon/\gamma = O(\epsilon)$
- $T_1 \cap T_2$ must be “big” – a $1 - \epsilon$ fraction of $T_2$. 
Consistent Sampling (Man’94, Hol’07, MMP+’10)

A mechanism $M$ has **$\varepsilon$-consistent sampling** if for all queries $f \in F$, all neighboring databases $D_1, D_2$, and all possible outcomes $z$:

$$\frac{|T_1 \setminus T_2|}{|T_2|} \leq \varepsilon$$

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

**Proof:**

$$\frac{\Pr_{r \leftarrow U_n}[M(D_1,f;r) = z]}{\Pr_{r \leftarrow U_n}[M(D_2,f;r) = z]} = \frac{\Pr_{r \leftarrow U_n}[r \in T_1]}{\Pr_{r \leftarrow U_n}[r \in T_2]} = \frac{|T_1|}{|T_2|} = \frac{|T_1 \cap T_2|}{|T_2|} + \frac{|T_1 \setminus T_2|}{|T_2|} \leq 1 + \varepsilon$$
A New Mechanism

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

- Round outcome to nearest multiple of \( 1/\varepsilon \)
  - Utility is conserved (asymptotically): still \( \Theta(1/\varepsilon) \)-utility
A New Mechanism

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

- Round outcome to nearest multiple of \(1/\varepsilon\)
  - Utility is conserved (asymptotically): still \(\Theta(1/\varepsilon)\)-utility
A New Mechanism

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

- Round outcome to nearest multiple of \( 1/\varepsilon \)
  - Utility is conserved (asymptotically): still \( \Theta(1/\varepsilon) \)-utility
- Guarantees \( T_1, T_2 \) will intersect on a large fraction of coins, as required for \( \varepsilon \)-consistent sampling.
- **Overcomes our lower bound.**
A New Mechanism

\[ M(D,f) = [f(D) + \text{Lap}(1/\varepsilon)]^{1/\varepsilon} \]

Can we implement it in a “SV-robust” manner?

- **Yes!** But **non-trivial**
  - Not every implementation is “SV-robust”
  - \( \varepsilon \)-consistent sampling is **necessary** but not **sufficient**

- Define **\( \varepsilon \)-SV-consistent sampling**
  - Natural definition, does not reference \( SV(\gamma) \)
  - **Sufficient** for “SV robustness”

- Use **arithmetic coding** to ensure SV-consistency
  - Need to be careful with **finite precision**
Differential Privacy – Our Results

- Differential privacy is**possible** with $SV(\gamma)$ sources.
- Separation between **traditional** (Enc/Com/ZK) and **differential** privacy.
Differential privacy is possible with $SV(\gamma)$ sources.

- Separation between traditional (Enc/Com/ZK) and differential privacy.

- Motivate consistent sampling as a design paradigm.
  - Useful applications in upcoming CCS paper [Mir’12].

Thank you!