Sharp lower bounds on the extractable randomness from non-uniform sources

Boris Škorić, Chibuzo Obi, Evgeny Verbitskiy and Berry Schoenmakers

WISSEC 2008
13&14 Nov, Eindhoven

Outline

- Randomness extraction
- Motivation: RNGs, PUFs and Biometrics
- Lower bound based on Leftover Hash Lemma & "smooth min-entropy"
- Leftover Hash Lemma revisited
- Improved lower bound

Randomness extraction: terminology

- Privacy amplification:
  Given a non-uniform source X, derive an L-bit string f(X) as uniformly distributed as possible on $\{0,1\}^L$.
  - true random number generators

- Information reconciliation:
  If the source is noisy, then some redundancy data W(X) must be given before privacy amplification is possible.
  - biometrics
  - PUFs

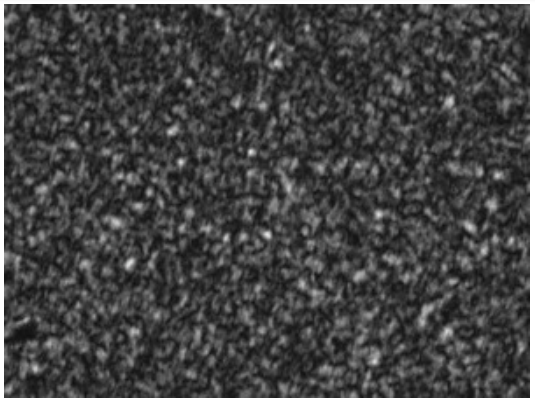
- Fuzzy extractor:
  - Does both information reconciliation and privacy amplification.
  - Extracts secret key K from noisy source.
  - Aims for high entropy H(K | W).

Randomness extraction: Examples

Biometrics:

***Privacy preserving** data.*
Hash of biometric stored in database.
Uniformity ⇒ efficiency.

PUFs:



Secret key K.
Uniformity ⇒ security.

Measure of uniformity: $\Delta(K,U) := \sum_k \left| \Pr[K = k] - \frac{1}{2^L} \right|$

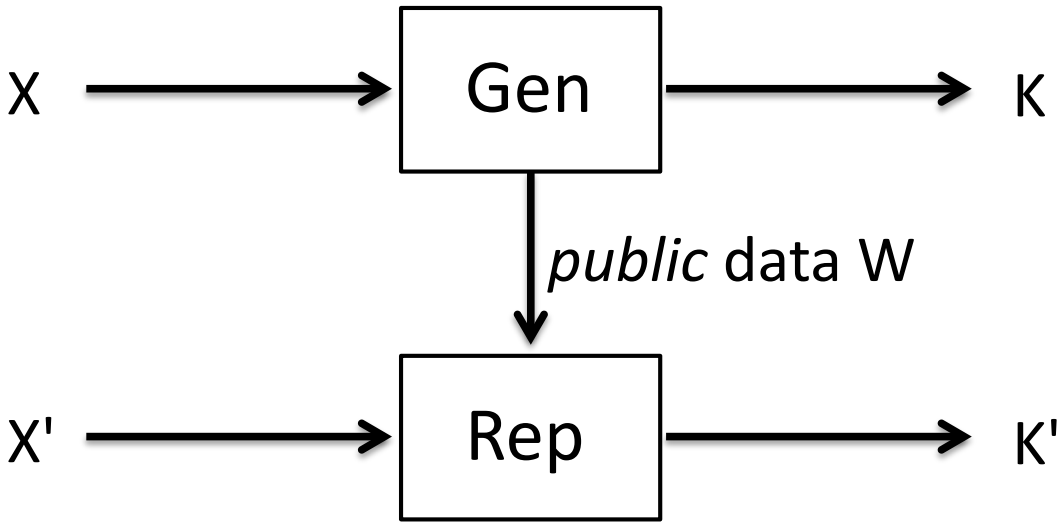
Statistical distance between K and uniform distribution U

Randomness extraction

L bits; $\Delta(K,U) \leq \varepsilon$.
L depends on "H"(X).

Case A: Without noise correction (RNG)

Case B: Noise correction (PUFs, biometrics)



conditioned on W

L bits; $\Delta(KW,UW) \leq \varepsilon$.
L depends on "H"(X|W).

Randomness extraction

What is achievable?

- If prob. distribution of X,W is known precisely:
  Use dedicated compression scheme.
  Extracts close to L = H(X | W).
- If not, resort to universal hash functions and Leftover Hash Lemma.

universal hash: $$\mathrm{Prob}[\Phi_R(x) = \Phi_R(x')] \le 1/2^L$$

leftover hash lemma: $$\Delta(R\Phi_R(X), RU) \le \frac{1}{2}\sqrt{2^L \sum_x P_x^2}$$

Known bounds

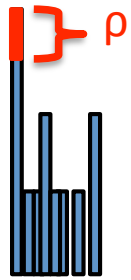
Renner & Wolf 2005

Def. **Extractable randomness** $L^{\varepsilon}_{ext}(X|W)$:
Given XW, maximum length of K=f(X), given W, such that $\Delta(KW,UW) \leq \varepsilon$.

$$L^{\varepsilon}_{\text{ext}}(X|W) \geq \max_{\rho}\left[H^{\rho}_{\infty}(X|W) - 2\log\frac{1}{\varepsilon-\rho}\right]$$

Smooth min-entropy

$$H^{\rho}_{\infty}(P) := \max_{Q\ \text{"}\rho\text{-close"}\ \text{to}\ P} -\log\max_{x} Q(x)$$



Leftover Hash Lemma revisited

- Generalization to **almost universal** hash functions Φ_r.
  Much more efficient implementation than **universal**.

$$\text{Prob}[\Phi_R(x) = \Phi_R(x')] \leq 2^{-L}(1+\delta)$$

- Use fact that smoothening violates normalization by amount ρ.

without conditioning

$$\Delta(R\Phi_R(X), RU) \leq \frac{1}{2}\sqrt{\rho^2 + \delta + 2^L \sum_x Q_x^2}$$

conditioned on W

$$\Delta(RW\Phi_R(X), RWU) \leq \frac{1}{2}\sqrt{1-\rho}\sqrt{\delta + 2^L \sum_w Q_w \sum_x Q_{x|w}^2}$$

average over w

Improved lower bound on ♯extractable bits

More general derivation

- Distribution not normalized
- Use Renyi H_2 instead of H_∞.
- Average-case conditioning vs worst-case
- (Almost universal hash functions vs universal)

$$L^{\varepsilon}_{\text{ext}}(X) \geq \max_{\rho} \left[H_2^{\rho}(X) + 2 - \log \frac{1}{\varepsilon(\varepsilon-\rho) - \delta/4} \right]$$

$$L^{\varepsilon}_{\text{ext}}(X \mid W) \geq \max_{\rho} \left[\tilde{H}_2^{\rho}(X \mid W) + 2 - \log \frac{1}{(\varepsilon-\rho)^2 - \delta/4} \right]$$

RW2005:

$$L^{\varepsilon}_{\text{ext}}(X \mid W) \geq \max_{\rho} \left[H_\infty^{\rho}(X \mid W) - \log \frac{1}{(\varepsilon-\rho)^2} \right]$$

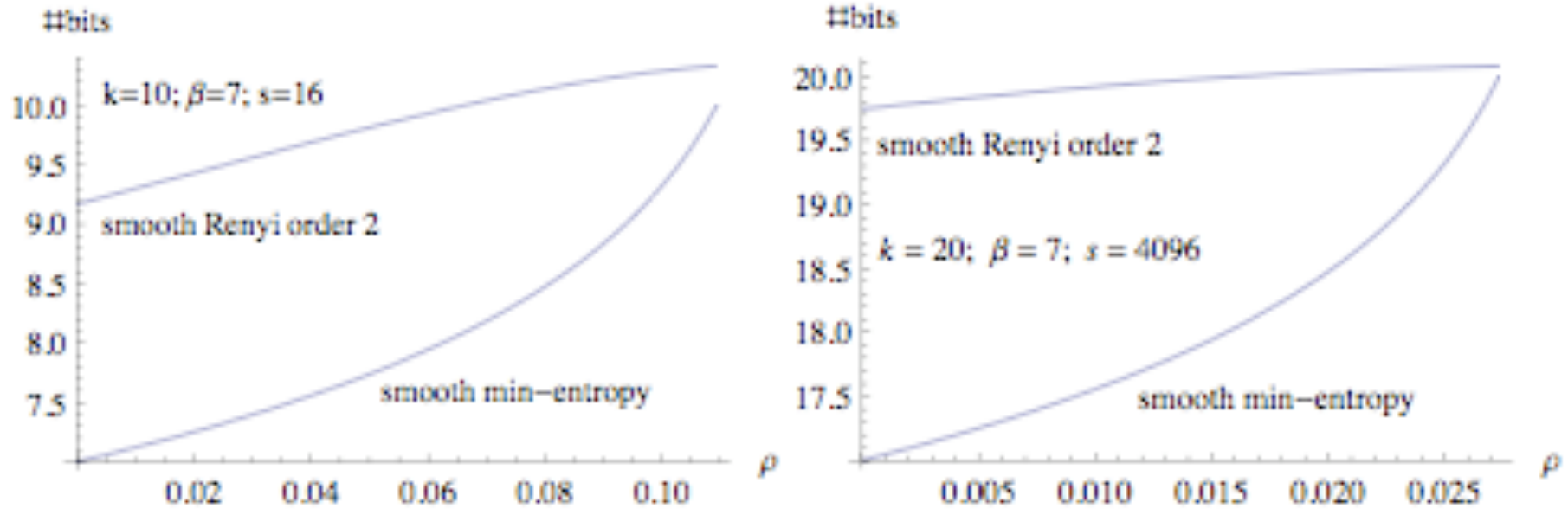
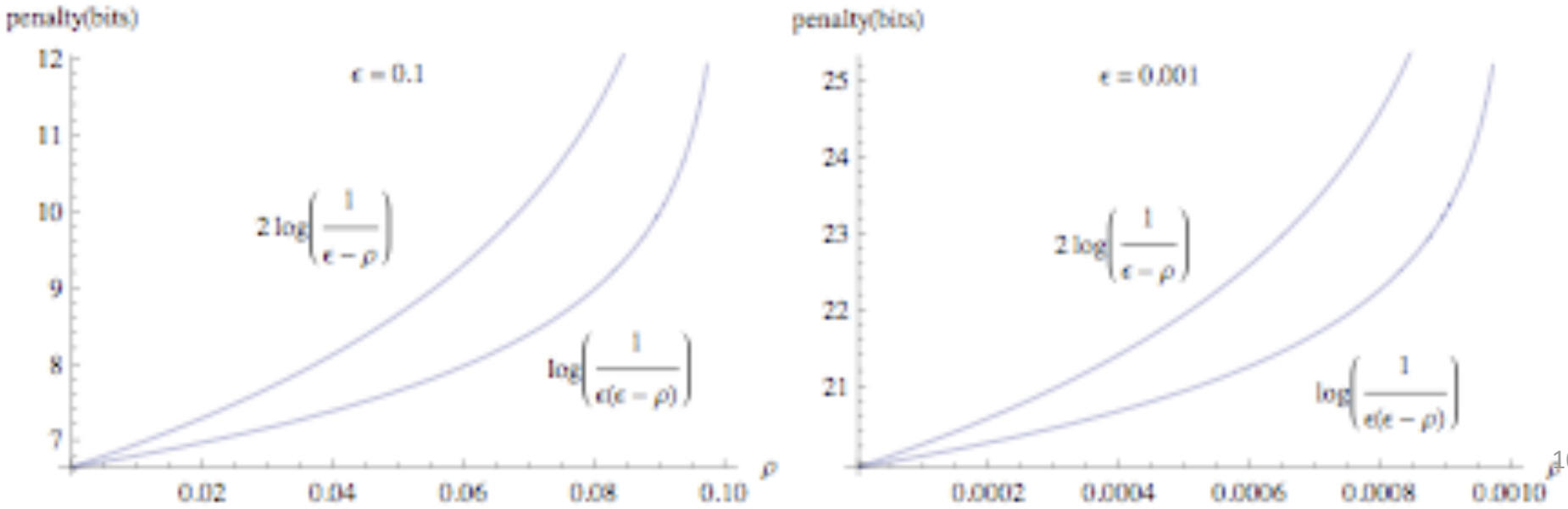


Figure 1: *Comparison of* $\mathrm{H}_\infty^{\rho,\mathrm{strict}}(X)$ *and* $\mathrm{H}_2^{\rho,\mathrm{strict}}(X)$ *in a toy example:* $X \in \{0,1\}^k$, s *elements have probability* $p_1 = 2^{-k}(1+\beta)$, *and the remaining* $2^k - s$ *elements all have probability* $p_2 = (1 - sp_1)/(2^k - s)$.



Conclusions

"When all else fails, extract with universal hash functions."

Results

- Unconditional case: Reduced smoothening penalty
- Explicitly taken into account the "almost" parameter in expression for L^{ε}_{ext}
- Difference between H_∞ and H_2 can be several bits
- Not spectacular, but useful for low-entropy sources

Future work

- See if smoothening penalty can be reduced for conditional case