Pseudorandom Generators from Regular One-way Functions: New Constructions with Improved Parameters

Yu Yu

清華大學

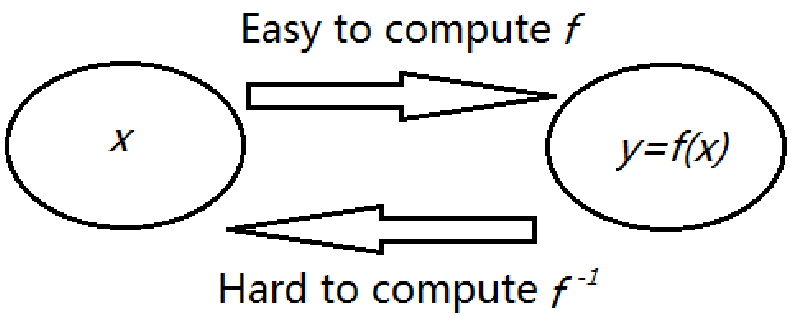
Tsinghua University

Joint work with Xiangxue Li and Jian Weng

Asiacrypt 2013

One-way Functions

One-way functions are an ensemble of functions $\{f_n : \{0,1\}^n \to \{0,1\}^{l(n)}\}_{n\in N}$ that are

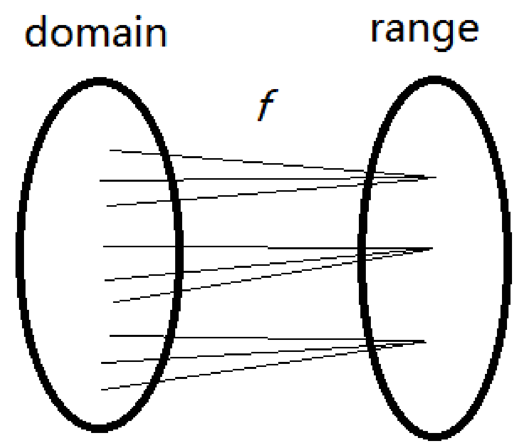


- Simplifying notation : $f : \{0,1\}^n \to \{0,1\}^{l(n)}$
- Definition: f is a (t,ε)-one-way function (OWF) if for all adversaries A of running time t, $\Pr_{y \leftarrow f(U_n)}[A(y) \in f^{-1}(y)] \le \varepsilon$
- Standard OWF: $t \in$ super-poly , $\varepsilon \in$ negl
- Folklore: OWFs can be assumed to be length-preserving, i.e., $l(n)=n$.

Regular Functions

- f is a regular function if for any n the preimage size

$\alpha = |f^{-1}(y)|$ is fixed (independent of y).



- Known-regular function: a regular function f whose regularity α is polynomial-time computable from security parameter n.
- Unknown-regular function: a regular function f whose regularity α is inefficient to approximate from security parameter n.

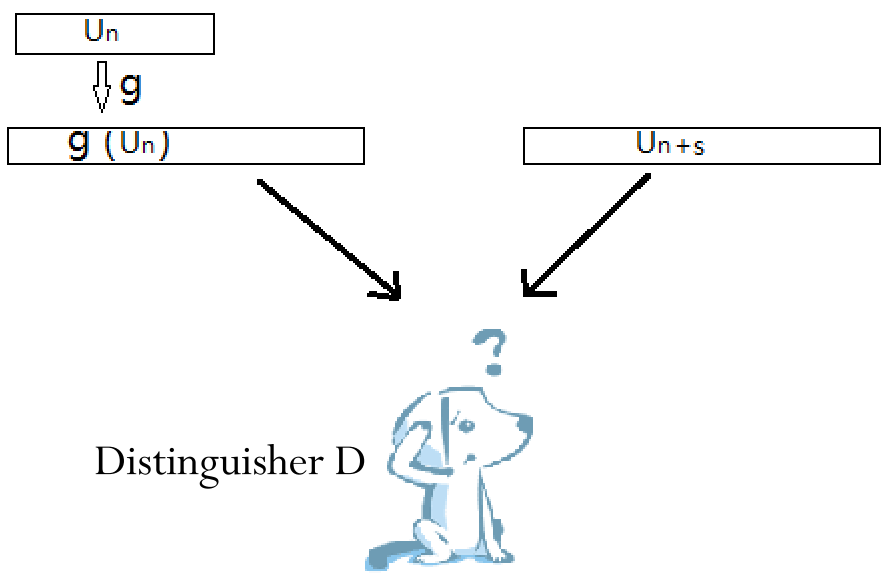
Note: one-way permutation is a special known-regular function.

Pseudorandom Generators

$g:\{0,1\}^n \rightarrow \{0,1\}^{n+s}$ is a (t,ε) -pseudorandom generator (PRG) with stretch s if for all distinguishers D of running time t,

$$|\Pr[D(g(U_n))=1]-\Pr[D(U_{n+s})=1]|\leq \varepsilon$$

$t \in$ super-poly , $\varepsilon \in$ negl, U_n is uniform distribution over $\{0,1\}^n$



Entropies, computational and statistical distance

collision entropy $\mathbf{H}_2(X) \stackrel{\text{def}}{=} -\log \sum_x \Pr[X = x]^2$

min-entropy $\mathbf{H}_\infty(X) \stackrel{\text{def}}{=} -\log\left(\max_x \Pr[X = x]\right)$

conditional collision entropy $\mathbf{H}_2(X|Z) \stackrel{\text{def}}{=} -\log\left(\mathbb{E}_{z\leftarrow Z}\left[\sum_x \Pr[X = x|Z = z]^2 \right] \right)$

conditional min-entropy $\mathbf{H}_\infty(X|Z) \stackrel{\text{def}}{=} -\log\left(\mathbb{E}_{z\leftarrow Z}\left[\max_x \Pr[X = x|Z = z] \right] \right)$

computational distance between X and Y

X and Y are (t, ε)-close, denoted by $\mathsf{CD}_t(X,Y) \le \varepsilon$,

if for every probabilistic distinguisher D of running time up to t it holds that

$$|\ \Pr[\mathsf{D}(X) = 1] - \Pr[\mathsf{D}(Y) = 1]\ | \le \varepsilon$$

statistical distance between X and Y, denoted by $\mathsf{SD}(X,Y)$, is defined by

$$\mathsf{SD}(X,Y) \stackrel{\text{def}}{=} \frac{1}{2}\sum_x |\Pr[X = x] - \Pr[Y = x]| = \mathsf{CD}_\infty(X,Y)$$

shorthand: $\mathsf{SD}(X,Y|Z) = \mathsf{SD}((X,Z),(Y,Z))$

$\mathsf{CD}_t(X,Y|Z) = \mathsf{CD}_t((X,Z),(Y,Z))$

Leftover Hash Lemma

leftover hash lemma *For any integers $d<k\leq n$, there exists a (polynomial-time computable) universal hash function family $\mathcal{H} \stackrel{\mathsf{def}}{=} \{h : \{0,1\}^n \rightarrow \{0,1\}^{k-d}\}$ such that for any joint distribution (X,Z) where $X \in \{0,1\}^n$ and $\mathbf{H}_2(X|Z) \geq k$, we have*

$$\mathsf{SD}(H(X),\ U_{k-d} \mid H, Z) \ \leq\ 2^{-\frac{d}{2}}$$

where H is uniformly distributed over the members of $\mathcal{H}$, the description size of H is called seed length, and d is called entropy loss, *i.e., the difference between the entropy of X (given Z) and the number of bits that were extracted from X.*

Informally: universal hash functions are good randomness extractors

Unpredictability Pseudoentropy (UP)

Definition 2.5 (unpredictability pseudo-entropy) *For distribution ensemble* (X, Z), *we say that* X *has* k *bits of pseudo-entropy conditioned on* Z *for all* t *-time adversaries , denoted by* $\mathbf{H}_t(X|Z) \geq k$ *, if for any* $n \in \mathbb{N}$ *and any probabilistic adversary* A *of running time* t

$$\Pr_{(x,z)\leftarrow(X,Z)} [\mathsf{A}(z) = x] \;\leq\; 2^{-k}$$

Alternatively, we say that X *is* 2^{-k} *-hard to predict given* Z *for all* t *-time adversaries.*

$$\boxed{\begin{array}{lcc}
\text{Challenger } \mathsf{C} & & \text{Adversary } \mathsf{A} \\
x \leftarrow U_n;\; y := f(x) & \xrightarrow{\quad y \quad} & \\
 & & x' := \mathsf{A}(y) \\
 & \xleftarrow{\quad x' \quad} & \\
\mathsf{A} \text{ wins iff } x' = x & &
\end{array}}$$

The interactive game between A and C that defines unpredictability pseudo-entropy, where $x \leftarrow U_n$ denotes sampling a random $x \in \{0,1\}^n$.

Goldreich-Levin Theorem

Goldreich-Levin Theorem: *For* $(X,Y) \in \{0,1\}^n \times \{0,1\}^*$ *, and for any integer* $m \leq n$*, there exists a function family* $\mathcal{H}_C \stackrel{\text{def}}{=} \{h_c : \{0,1\}^n \to \{0,1\}^m\}$ *of description size* $\Theta(n)$*, such that*

- *If* $Y = f(X)$ *for any* (t,ε)*-OWF* f *and* X *uniform over* $\{0,1\}^n$*, then we have*

$$\mathsf{CD}_{t'}(\, H_C(X) \,,\, U_m \mid Y, H_C) \in O(2^m \cdot \varepsilon) \;. \tag{1}$$

- *If* X *is* ε*-hard to predict given* Y *for all* t*-time adversaries, namely,* $\mathbf{H}_t(X|Y) \geq \log(1/\varepsilon)$*, then we have*

$$\mathsf{CD}_{t'}(\, H_C(X) \,,\, U_m \mid Y, H_C) \in O(2^m \cdot (n \cdot \varepsilon)^{\frac{1}{3}}) \;. \tag{2}$$

where $t' = t \cdot (\varepsilon/n)^{O(1)}$ *and function* H_C *is uniformly distributed over the members of* $\mathcal{H}_C$.

A Key Oberservation about Unpredictability Pseudoentropy

- Unpredictability Pseudoentropy (UP) : X has m bits of UP given f(X) for t-time adversaries if every A of running time t wins the following game with probability no greater than 2^{-m}

Challenger C

$x \leftarrow X; y := f(x)$ $\xrightarrow{\quad y \quad}$

Adversary A

$x' \leftarrow A(y)$

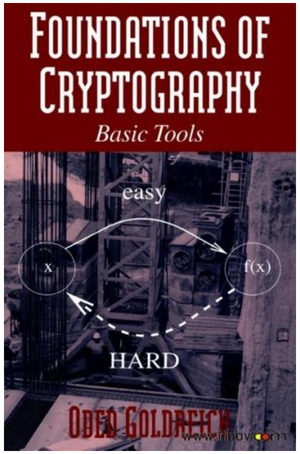
A wins iff $x = x'$ $\xleftarrow{\quad x' \quad}$

- Question: what's the UP of X given $f(X)$ if f is a (t, ε) - regular OWF with $|f^{-1}(y)| = 2^k$?
- Observation: X given f(X) has $k + \log(1/\varepsilon)$ bits of UP.
- Rationale: $\Pr[A(f(X)) \in f^{-1}(f(X))] \le \varepsilon \Rightarrow \Pr[A(f(X)) = X] \le 2^{-k} \cdot \varepsilon$

The FIRST CONSTRUCTION (from known-regular OWF)

- $g(X, h_1, h_2, h_c) = (h_1(f(X_1)), h_2(X_1), h_c(X_1), h_1, h_2, h_c)$

A complicated proof by Goldreich in Section 3.5.2 of



PRGs from Known-Regular OWFs by three extractions (a three-line proof)

- Assumption: f is (t,ε) -one-way and 2^k-regular, i.e. $|f^{-1}(y)|=2^k$
- Construction and Proof.

1. $H_\infty(f(X))=n-k$ $\Rightarrow$ extract $(n-k)$ bits using h_1
2. $H_\infty(X\mid f(X))=k$ $\Rightarrow$ extract k bits using h_2
3. chain rule:

$$H^t_{\mathrm{up}}(X\mid f(X))=k+\log(1/\varepsilon) \Rightarrow H^t_{\mathrm{up}}(X\mid f(X),h_2(X))\geq\log(1/\varepsilon)$$

extract $O(\log(1/\varepsilon))$ bits using hard-core function h_c

- This completes the proof for the folklore construction, i.e.

  $g\,(X,\,h_1,\,h_2,\,h_c) = (h_1(f(X_1)),\,h_2(X_1),\,h_c(X_1),\,h_1,\,h_2,\,h_c)$ is a PRG.
- Parameters: seed length linear in n, and a single call to f .

Tightening the security bounds

- $g(x, h_1, h_2, h_c) = (h_1(f(x)), h_2(x), h_c(x), h_1, h_2, h_c)$

The proof for 3rd extraction: consider $f'(x,h_2)=(f(x), h_2(x), h_2)$

$\because x$ is ε-hard to predict given $f'(x,h_2)$, i.e. $\mathrm{H}_{up}^{t}(X|f'(X,H_2)) \geq \log(1/\varepsilon)$

$\therefore$ by Goldreich-Levin Thm, $h_c(x)$ is $2^m(n\cdot\varepsilon)^{1/3}$ -close to U_m given $f'(x,h_2)$

- A tighter approach (use the tight version of Goldreich-Levin)?

if f' is an ε'-hard OWF, then $h_c(x)$ is $(2^m\cdot\varepsilon')$ -close to U_m given $f'(x,h_2)$

1. Goldreich show $\varepsilon' = O(\varepsilon^{1/5})$ in [Gol01,vol-1]
2. We show $\varepsilon' = 3\sqrt{\varepsilon}$ against t-time adversaries

the idea: show f' is almost 1-to-1, i.e. $\mathrm{H}_2(f'(X,H_2)|H_2) \geq n-1$

The Second Construction
(NEW, improving the Randomized Iterate)

The Randomized Iterate

- Goldreich, Krawczyk and Luby (SICOMP 93) :

  PRGs from known regular OWFs with seed length $O(n^3)$

- Haitner, Harnik and Reingold (CRYPTO 2006):

PRGs from unknown regular OWFs with seed length $O(n \cdot \log n)$

$$\begin{tikzcd}
\boldsymbol{x} \arrow[r, "f"] & y_1 = f(x) \arrow[r, "h_1"] \arrow[d] & x_1 = h_1(y_1) \arrow[r, "f"] & y_2 = f(x_1) \arrow[r, "h_2"] \arrow[d] & x_2 = h_2(y_2) \arrow[r, "f"] & y_3 = f(x_2) \arrow[d] \\
\text{output} & h_c(x) & & h_c(x_1) & & h_c(x_2) \quad \cdots
\end{tikzcd}$$

$h_1, h_2, \ldots$ are random pairwise independent hash, h_c is hard-core function

Lower bounds by Holenstein and Sinha (FOCS12)

- Asymptotic setting: Any black-box construction of PRG must make $\Omega(n/\log n)$ calls to an arbitrary (including unknown regular) OWF.
- Concrete setting : Any black-box construction of PRG must make $\Omega(n/\log(1/\varepsilon))$ calls to an arbitrary (including unknown regular) $(\varepsilon^{-1},\varepsilon)$ -secure OWF.

PRGs from unknown-regular OWFs: a new construction

- Assumption: f is (t,ε)-one-way and 2^k-regular (k is unknown).
- The goal: a PRG construction oblivious of k.
- The idea: transform f into a known-regular OWF $\bar{f}$

$$f:\{0,1\}^n \to \mathcal{Y}, \text{ where } \mathcal{Y} \subseteq \{0,1\}^n$$

$$\text{define } \bar{f}:\mathcal{Y}\times\{0,1\}^n \to \mathcal{Y}$$

$$\bar{f}(y,r) = f(y \oplus r)$$

$$\text{where } \oplus: \text{"bitwise XOR"}, y \leftarrow f(U_n),\ r \leftarrow U'_n$$

1. $\bar{f}$ is also a (t,ε)-one-way function
2. $\bar{f}$ is a 2^n-regular function, i.e. $|\bar{f}^{-1}(y,r)| = 2^n$ regardless of k

PRGs from unknown-regular OWFs: a new construction (cont'd)

- Given a one-way function with known pre-image size 2^n

$$\bar{f} : \mathcal{Y} \times \{0,1\}^n \to \mathcal{Y}$$

- Similarly, (Y, R) has $n + \log(1/\varepsilon)$ bits of UP given $\bar{f}(Y, R)$.
- We get a special PRG $\bar{g} : \mathcal{Y} \times \{0,1\}^n \to \mathcal{Y} \times \{0,1\}^{n+\Theta(\log(1/\varepsilon))}$
- Done?

No, n bits needed to sample from $\mathcal{Y}$ (i.e. $f(U_n)$)

stretch : $-n + \Theta(\log(1/\varepsilon))$ $+\Theta(\log(1/\varepsilon))$ $+\Theta(\log(1/\varepsilon))\cdots\cdots$

To make it positive: iterate $\bar{g}$

- In summary: a PRG from unknown regular OWF with linear seed length (hybrid argument) and $\Theta(n/\log(1/\varepsilon))$ OWF calls.
- Tight (Holenstein and Sinha, FOCS 2012): BB construction of PRG requires $\Omega(n/\log(1/\varepsilon))$ OWF calls, and $\Omega(n/\log n)$ calls in general.

Summary

- PRG from any known-regular ε-hard OWF:

seed length $\tilde{O}(n)$ and a single call to the underlying OWF

- PRG from any unknown-regular ε-hard OWF:

seed length $\tilde{O}(n)$ and $\tilde{O}(n/\log(1/\varepsilon))$ OWF calls

Question: remove the dependency on ε?

Yes, by paying a factor $\omega(1)$ in seed length and number of calls.

Why? Due to the entropy loss of the Leftover Hash Lemma.

Given 1-to-1 OWF $f:\{0,1\}^n \to \{0,1\}^{n+1}$ (without knowing ε)

Run $q=\omega(1)$ copies of f, extracting $2\log n$ hardcore bits per copy,
followed by a single extraction with entropy loss set to $q \cdot \log n$.

More details

Full version at eprint http://eprint.iacr.org/2013/270

Thank you!