A PSEUDORANDOM GENERATOR FROM ANY ONE-WAY FUNCTION

JOHAN HÅSTAD , RUSSELL IMPAGLIAZZO , LEONID A. LEVIN , AND MICHAEL LUBY

3.2. One-way function.

DEFINITION 3.2 (one-way function). *Let* $f : \{0,1\}^{t_n} \to \{0,1\}^{\ell_n}$ *be a* $\mathbf{P}$*-time function ensemble and let* $X \in_{\mathcal{U}} \{0,1\}^{t_n}$. *The* success probability *of adversary* A *for inverting* f *is*

$$sp_n(A) = \Pr[f(A(f(X))) = f(X)].$$

Then f *is an* $\mathbf{R}$*-secure one-way function if there is no* $\mathbf{R}$*-breaking adversary for* f.

DEFINITION 3.3 (computationally indistinguishable). *Let* $\mathcal{D} : \{0,1\}^{\ell_n}$ *and* $\mathcal{E} : \{0,1\}^{\ell_n}$ *be probability ensembles. The* success probability *of adversary* A *for distinguishing* $\mathcal{D}$ *and* $\mathcal{E}$ *is*

$$sp_n(A) = |\Pr[A(X) = 1] - \Pr[A(Y) = 1]|,$$

where X *has distribution* $\mathcal{D}$ *and* Y *has distribution* $\mathcal{E}$. $\mathcal{D}$ *and* $\mathcal{E}$ *are* $\mathbf{R}$*-secure* computationally indistinguishable *if there is no* $\mathbf{R}$*-breaking adversary for distinguishing* $\mathcal{D}$ *and* $\mathcal{E}$.

Definition 3.5 (pseudorandom generator). *Let $g : \{0,1\}^{t_n} \to \{0,1\}^{\ell_n}$ be a* **P***-time function ensemble where $\ell_n > t_n$. Then g is an* **R***-secure pseudorandom generator if the probability ensembles $g(\mathcal{U}_{t_n})$ and $\mathcal{U}_{\ell_n}$ are* **R***-secure computationally indistinguishable.*

PROPOSITION 3.6. *Suppose* $g : \{0,1\}^n \to \{0,1\}^{n+1}$ *is a pseudorandom generator that stretches by one bit. Define* $g^{(1)}(x) = g(x)$*, and inductively, for all* $i \geq 1$*,*

$$g^{(i+1)}(x) = \langle g(g^{(i)}(x)_{\{1,\ldots,n\}}), g^{(i)}(x)_{\{n+1,\ldots,n+i\}} \rangle.$$

Let k_n *be an integer-valued* $\mathbf{P}$*-time polynomial parameter. Then* $g^{(k_n)}$ *is a pseudorandom generator.*

7. A direct construction. We have shown how to construct a false-entropy generator from an arbitrary one-way function, a pseudoentropy generator from a false-entropy generator, and finally a pseudorandom generator from a pseudoentropy generator. The combinations of these constructions give a pseudorandom generator from an arbitrary one-way function as stated in Theorem 6.3. By literally composing the reductions given in the preceding parts of this paper, we construct a pseudorandom generator with inputs of length n^{34} from a one-way function with inputs of length n. This is obviously not a suitable reduction for practical applications. In this subsection, we use the concepts developed in the rest of this paper, but we provide a more direct and efficient construction. However, this construction still produces a pseudorandom generator with inputs of length n^{10}, which is clearly still not suitable for practical applications. (A sharper analysis can reduce this to n^8, which is the best we could find using the ideas developed in this paper.) The result could only be considered practical if the pseudorandom generator had inputs of length n^2, or perhaps even close to n. (However, in many special cases of one-way functions, the ideas from this paper are practical; see, e.g., [Luby96].)

DEFINITION 2.3 (information and entropy). *Let $\mathcal{D}$ be a distribution on a set S. For each $x \in S$, define the* information *of x with respect to $\mathcal{D}$ to be* $\mathbf{I}_{\mathcal{D}}(x) = -\log(\mathcal{D}(x))$.

The following definition characterizes how much entropy is lost by the application of a function f to the uniform distribution.

DEFINITION 2.7 (degeneracy of f). *Let $f : \{0,1\}^n \to \{0,1\}^{\ell_n}$ and let $X \in_{\mathcal{U}} \{0,1\}^n$. The* degeneracy *of f is* $\mathbf{D}_n(f) = \mathbf{H}(X|f(X)) = \mathbf{H}(X) - \mathbf{H}(f(X))$.

DEFINITION 2.13 ($\tilde{\mathbf{D}}_f$). *Let $f : \{0,1\}^n \to \{0,1\}^{\ell_n}$ be a $\mathbf{P}$-time function ensemble. For $z \in \mathrm{range}_f$, define the* approximate degeneracy of z *as*

$$\tilde{\mathbf{D}}_f(z) = \left\lceil \log(\sharp \mathrm{pre}_f(z)) \right\rceil .$$

Notice that $\tilde{\mathbf{D}}_f(z)$ is an approximation to within an additive factor of 1 of the quantity $n - \mathbf{I}_{f(X)}(z)$. Furthermore, $\mathrm{E}[\tilde{\mathbf{D}}_f(f(X))]$ is within an additive factor of 1 of the degeneracy of f. If f is a σ_n-regular function then, for each $z \in \mathrm{range}_f$, $\tilde{\mathbf{D}}_f(z)$ is within an additive factor of 1 of $\log(\sigma_n)$, which is the degeneracy of f.

Let $\mathbf{p}_n$ be the probability that $I \leq \tilde{\mathbf{D}}_f(f(X))$.

Let

$$f : \{0,1\}^n \to \{0,1\}^{\ell_n} \tag{6.1}$$

be a one-way function and let

$$h : \{0,1\}^{p_n} \times \{0,1\}^n \to \{0,1\}^{n+\lceil \log(2n) \rceil} \tag{6.2}$$

be a universal hash function. Similar to Construction 5.1, for $x \in \{0,1\}^n$, $i \in \{0,\ldots,n-1\}$, and $r \in \{0,1\}^{p_n}$, define **P**-time function ensemble

$$f'(x,i,r) = \langle f(x), h_r(x)_{\{1,\ldots,i+\lceil \log(2n) \rceil\}}, i, r \rangle. \tag{6.3}$$

Let $\mathcal{X} = \langle X, I, R \rangle$ represent the input distribution to f', and let c_n be the length of $\mathcal{X}$ and c'_n the length of $f'(\mathcal{X})$.

Let $\mathbf{e}_n = \mathbf{H}(f'(\mathcal{X}))$.

Let $b(x,y) = x \odot y$.

Set $k_n = 2000n^6$.

Intuitively, we generate pseudorandom bits as follows: let $\mathcal{X}' = \mathcal{X}^{k_n}$ and $Y' = Y^{k_n}$. We first compute $f'^{k_n}(\mathcal{X}')$ and $b^{k_n}(\mathcal{X}', Y')$. Intuitively, we are entitled to recapture

$$k_n c_n - \mathbf{H}\langle f'^{k_n}(\mathcal{X}'), b^{k_n}(\mathcal{X}', Y')\rangle$$

bits from $\mathcal{X}'$, because this is the conditional entropy left after we have computed f'^{k_n} and b^{k_n}.

We are entitled to recapture $k_n \mathbf{p}_n$ bits from the $b^{k_n}(\mathcal{X}', Y')$ (since we get a hidden bit out of each copy whenever $I \leq \tilde{\mathbf{D}}_f(f(X))$). Finally, we should be able to extract $\mathbf{e}_n k_n$ bits from ${f'}^{k_n}(\mathcal{X}')$, since $\mathbf{e}_n$ is the entropy of $f'(\mathcal{X})$. Since $b(\mathcal{X}, Y)$ is almost totally predictable for almost all inputs where $I \geq \tilde{\mathbf{D}}_f(f(X))$,

$$\mathbf{H}\langle f'(\mathcal{X}), b(\mathcal{X}, Y)\rangle \leq \mathbf{e}_n + \mathbf{p}_n - 1/n + 1/(2n).$$

Let $Z \in_{\mathcal{U}} \{0,1\}^{m_n}$, and let

$$\mathcal{D} = \langle {h'}_U(\langle X'_1 \odot Y'_1, \ldots, X'_{k_n} \odot Y'_{k_n}\rangle), {f'}^{k_n}(X', I', R'), U, Y'\rangle,$$

$$\mathcal{E} = \langle Z, {f'}^{k_n}(X', I', R'), U, Y'\rangle.$$

LEMMA 6.4. $\mathbf{H}(\mathcal{E}) \geq \mathbf{H}(\mathcal{D}) + 10n^2$.

Proof. The entropy of $\mathcal{D}$ and $\mathcal{E}$ excluding the first m_n bits is exactly the same. The additional entropy in the first m_n bits of $\mathcal{E}$ is equal to m_n. An upper bound on the additional entropy in the first m_n bits of $\mathcal{D}$ is the additional entropy in $\langle X'_1 \odot Y'_1, \ldots, X'_{k_n} \odot Y'_{k_n}\rangle$. For each $j \in \{1, \ldots, k_n\}$ where $I'_j < \tilde{\mathbf{D}}_f(f(X'_j))$, the amount of entropy added by $X'_j \odot Y'_j$ is at most 1. On the other hand, under the condition that $I'_j \geq \tilde{\mathbf{D}}_f(f(X'_j))$, $X'_j \odot Y'_j$ is determined by $\langle f'(X'_j, I'_j, R'_j), Y'_j\rangle$ with probability at least $1 - 1/2n$, and thus the additional entropy under this condition is at most $1/2n$. Since $I'_j < \tilde{\mathbf{D}}_f(f(X'_j))$ with probability $\mathbf{p}_n - 1/n$, it follows that the additional entropy added by $X'_j \odot Y'_j$ is at most $\mathbf{p}_n - 1/2n$. Therefore, the additional entropy in the first m_n bits of $\mathcal{D}$ is at most $k_n(\mathbf{p}_n - 1/2n) = m_n + 2k_n^{2/3} - k_n/2n < m_n - 10n^2$ by choice of k_n. ∎

Thus, if we add up all the output bits, we are entitled to $k_n(c_n+1/(2n))$, or $k_n/(2n)$ more bits than the input to ${f'}^{k_n}$. However, our methods of extracting entropy are not perfect, so we need to sacrifice some bits at each stage; to use Corollary 4.10, we need to sacrifice $2nk_n^{2/3}$ at each stage, so we chose k_n to satisfy $k_n/(2n) > 6n{k_n}^{2/3}$.

COROLLARY 4.10. *Let k_n be an integer-valued* **P***-time polynomial parameter.*

- *Let $\mathcal{D} : \{0,1\}^n$ be a probability ensemble, let $m_n = k_n\mathbf{H}(\mathcal{D}) - 2nk_n^{2/3}$, and let $h : \{0,1\}^{p_n} \times \{0,1\}^{nk_n} \to \{0,1\}^{m_n}$ be a universal hash function. Let $X' \in_{\mathcal{D}^{k_n}} \{0,1\}^{k_n \times n}$ and let $Y \in_{\mathcal{U}} \{0,1\}^{p_n}$. Then*

$$\mathbf{L_1}(\langle h_Y(X'), Y\rangle, \mathcal{U}_{m_n+p_n}) \leq 2^{1-k_n^{1/3}}.$$

- *Let $\mathcal{D}_1 : \{0,1\}^n$ and $\mathcal{D}_2 : \{0,1\}^n$ be not necessarily independent probability ensembles, and let $\mathcal{D} = \langle \mathcal{D}_1, \mathcal{D}_2\rangle$. Let $m_n = k_n\mathbf{H}(\mathcal{D}_2|\mathcal{D}_1) - 2nk_n^{2/3}$. Let $h : \{0,1\}^{p_n} \times \{0,1\}^{nk_n} \to \{0,1\}^{m_n}$ be a universal hash function. Let $\langle X_1', X_2'\rangle \in_{\mathcal{D}^{k_n}} \{0,1\}^{k_n \times 2n}$ and let $Y \in_{\mathcal{U}} \{0,1\}^{p_n}$. Then*

$$\mathbf{L_1}(\langle h_Y(X_2'), Y, X_1'\rangle, \langle \mathcal{U}_{m_n+p_n}, X_1'\rangle) \leq 2^{1-k_n^{1/3}}.$$

let $m_n = k_n(c_n - \mathbf{e}_n - \mathbf{p}_n + 1/(2n)) - 2nk_n{}^{2/3}$,
$m'_n = k_n\mathbf{p}_n - 2nk_n{}^{2/3}$,
and $m''_n = k_n\mathbf{e}_n - 2nk_n{}^{2/3}$.

Let R_1, R_2, and R_3 be indices of hash functions so that
h_{R_1} maps k_nc_n bits to m_n bits,
h_{R_2} maps k_n bits to m'_n bits, and
h_{R_3} maps $k_nc'_n$ bits to m''_n bits.

CONSTRUCTION 7.1.

$$g(\mathcal{X}', Y', R_1, R_2, R_3) = \langle h_{R_1}(\mathcal{X}'), h_{R_2}(b^{k_n}(\mathcal{X}', Y')), h_{R_3}(f'^{k_n}(\mathcal{X}')), Y', R_1, R_2, R_3\rangle.$$

THEOREM 7.2. *If f is a one-way function and g is as in Construction* 7.1*, then g is a mildly nonuniform pseudorandom generator.*

Proof. It is easy to check that g outputs more bits than it inputs.

As noted above, the conditional entropy of $\mathcal{X}$ given $f'(\mathcal{X})$ and $b(\mathcal{X}, Y)$ is at least $c_n - \mathbf{e}_n - \mathbf{p}_n + (1/2n)$. Thus, from Corollary 4.10, we have that $\langle h_{R_1}(\mathcal{X}'), R_1\rangle$ is statistically indistinguishable from random bits given $\langle {f'}^{k_n}(\mathcal{X}'), b^{k_n}(\mathcal{X}', Y'), Y'\rangle$.

Hence, $g(\mathcal{X}', Y', R_1, R_2, R_3)$ is statistically indistinguishable from

$$\langle Z_1, h_{R_2}(b^{k_n}(\mathcal{X}', Y')), h_{R_3}({f'}^{k_n}(\mathcal{X}')), Y', R_1, R_2, R_3\rangle,$$

where $Z_1 \in_{\mathcal{U}} \{0,1\}^{m_n}$.

Now, from Lemmas 6.5 and 6.1, it follows that $h_{R_2}(b^{k_n}(\mathcal{X}', Y'))$ is computationally indistinguishable from random bits given $\langle {f'}^{k_n}(\mathcal{X}'), R_2, Y' \rangle$. Thus, $g(\mathcal{X}', Y', R_1, R_2, R_3)$ is computationally indistinguishable from

$$\langle Z_1, Z_2, h_{R_3}({f'}^{k_n}(\mathcal{X}')), Y', R_1, R_2, R_3 \rangle,$$

where $Z_2 \in_{\mathcal{U}} \{0,1\}^{m'_n}$.

Finally, from Corollary 4.10, $\langle h_{R_3}({f'}^{k_n}(\mathcal{X}')), R_3 \rangle$ is statistically indistinguishable from $\langle Z_3, R_3 \rangle$, where $Z_3 \in_{\mathcal{U}} \{0,1\}^{m''_n}$. Thus, the output of g is computationally indistinguishable from a truly random output of the same length. □

We still need to use Proposition 4.17 to get rid of the mild nonuniformity. From the arguments above, it is clear that an approximation of both $\mathbf{e}_n$ and $\mathbf{p}_n$ that is within $1/(8n)$ of their true values is sufficient. Since $0 \leq \mathbf{e}_n \leq n$, and $0 \leq \mathbf{p}_n < 1$, there are at most $\mathcal{O}(n^3)$ cases of pairs to consider. This means that we get a total of $\mathcal{O}(n^3)$ generators, each needing an input of length $\mathcal{O}(n^7)$. Thus the total input size to the pseudorandom generator is $\mathcal{O}(n^{10})$, as claimed.

PROPOSITION 4.17. *Let $\mathbf{a}_n$ be any value in $\{0, \ldots, k_n\}$, where k_n is an integer-valued* **P***-time polynomial parameter. Let $g : \{0,1\}^{\lceil \log(k_n) \rceil} \times \{0,1\}^n \to \{0,1\}^{\ell_n}$ be a* **P***-time function ensemble, where $\ell_n > nk_n$. Let $x' \in \{0,1\}^{k_n \times n}$ and define* **P***-time function ensemble $g'(x') = \oplus_{i=1}^{k_n} g(i, x'_i)$. Let g be a mildly nonuniform pseudorandom generator when the first input is set to $\mathbf{a}_n$. Then g' is a pseudorandom generator.*

A PSEUDORANDOM GENERATOR FROM ANY ONE-WAY FUNCTION

JOHAN HÅSTAD , RUSSELL IMPAGLIAZZO , LEONID A. LEVIN , AND MICHAEL LUBY

JOHAN HÅSTAD , RUSSELL IMPAGLIAZZO , LEONID A. LEVIN , AND MICHAEL LUBY