

BU CAS CS 538: Cryptography
Lecture Notes. Fall 2005.
<http://www.cs.bu.edu/~itkis/538/>

Gene Itkis*

Boston University Computer Science Dept.

Notes for Lectures 6–7: PRG Constructions

1 On the Necessity of Assumptions

When one tries to build pseudorandom generators, one runs into the following problem: predicting the next bit is in NP. Because we don't know that P is not equal to NP, if P=NP, then one can always predict the next bit in polynomial time, thus making pseudorandom generators impossible to build.

More formally, suppose the adversary is given bits $y_1 \dots y_{i-1}$ and has to predict y_i . Since the adversary knows G , the adversary can build a circuit C that is satisfied only by inputs x for which the first $i-1$ bits of $G(x)$ are $y_1 \dots y_{i-1}$. Furthermore, the adversary can modify that circuit to make C_1 , which is satisfied only by inputs x for which the first $i-1$ bits of $G(x)$ are $y_1 \dots y_{i-1}$ and for which the i -th bit of $G(x)$ is 1. Similarly, the adversary can modify C to make C_0 , which is satisfied only by inputs x for which the first $i-1$ bits of $G(x)$ are $y_1 \dots y_{i-1}$ and for which the i -th bit of $G(x)$ is 0. (Thus, $C(x)$ is the “or” of $C_0(x)$ and $C_1(x)$.) Then, if C_0 is satisfiable and C_1 is not, $y_i = 0$ is the correct prediction; else if C_1 is satisfiable and C_0 is not, $y_i = 1$ is the correct prediction (it could be that both are satisfiable, which means x is not unique and both predictions are possible).

If P=NP, then circuit satisfiability can be decided in polynomial time, and we can build a bit predictor as follows. Request bits until exactly one of C_0 and C_1 is satisfiable (this must happen, because the length $l(k)$ of the pseudorandom string y is greater than the length k of the seed x , so not all strings y are possible, so at some point only a single possibility for the next bit must remain), and then predict the bit according to whichever circuit is satisfiable. This predictor will be always correct!

Therefore, at the very least we must assume that P≠NP in order to build pseudorandom generators. In fact, we will have to make much stronger assumptions about intractability of particular problems in the *average* case (whereas P vs. NP question is about only the worst case).

1.1 Discrete Logarithm Assumption

Notation for the purposes of this course: $\mathbb{Z}_p = \{0, 1, \dots, p-1\}$ and $\mathbb{Z}_p^* = \{1, 2, \dots, p-1\}$. \mathbb{Z}_p^* is generally understood to mean that we are interested in the multiplication modulo p operation.

It is a fact from number theory that modulo every prime p , there exists a generator g of \mathbb{Z}_p^* , i.e., g such that $g, g^2 \bmod p, g^3 \bmod p, \dots, g^{p-1} \bmod p$ actually covers the whole set $\{1, 2, \dots, p-1\}$. Moreover, p and g can be found efficiently: there exists an algorithm that, for any k , finds a random k -bit prime p and some generator g of \mathbb{Z}_p^* in time polynomial in k .

The function $f(x) : \mathbb{Z}_p^* \rightarrow \mathbb{Z}_p^*$ that is computed as $f(x) = g^x \bmod p$ is a bijection (a.k.a. permutation) of \mathbb{Z}_p^* . Its inverse is known as the “discrete logarithm”: when working modulo p , we will denote by $\log_g y$ the unique value x , $1 \leq x \leq p-1$, such that $g^x \equiv y \pmod{p}$.

* These notes are heavily based on the notes by Leo Reyzin (see <http://www.cs.bu.edu/fac/reyzin/teaching/f04cs538/index.html>), with later modifications by Gene Itkis (who is deeply grateful to Leo for kindly making the latex source for the notes available).

As we know, f is easy to evaluate. It is believed (because many have tried and failed) that f is hard to invert. Namely, the following assumption is widely believed to hold.

Assumption 1 (The Discrete Logarithm (DL) Assumption) *For any poly-time algorithm A , there exists a negligible function η such that, if you generate random k -bit p and its generator g (according to the algorithm described above) and select a random $x \in \mathbb{Z}_p^*$, $\Pr[A(p, g, g^x \bmod p) = x] \leq \eta(k)$.*

2 First attempt at a PRG

Now, let's try to build a generator based on that assumption. Select a random p of length k , generator g of \mathbb{Z}_p^* , and $x \in \mathbb{Z}_p^*$. For some n (greater than the number of random bits you used to generate p, g, x), compute $x_1 = x, x_2 = g^{x_1}, x_3 = g^{x_2}, \dots, x_n = g^{x_{n-1}}$ (all modulo p). Output, in reverse order, $x_n, x_{n-1}, x_{n-2}, \dots, x_1$. Seems like should be hard to predict (because you need to take DL of x_i to get x_{i-1}). However, all that DL assumption says is that it is hard to predict *values*, not *individual bits*. In fact, we know from PS2 that given g^x , you can tell whether x has last bit 0 or 1 (i.e., x is a square or not), so clearly you can predict some bit of x_{i-1} given x_i . Thus, at least some bits are predictable, and hence this is not a proper PRG.

To build a PRG, we need to condense the hardness of predicting x_{i-1} down to the hardness of predicting a single bit. We do this below.

3 Blum-Micali PRG

The following construction was proposed together with the above definition of unpredictability in [BM84].

Define $B(x) = \{0, \text{ if } x < p/2, \text{ and } 1 \text{ otherwise } \}$.

Lemma 1. *Given g^x , $B(x)$ is unpredictable. I.e., for every polynomial-time algorithm D there exists a negligible function $\eta(k)$ such that for all k ,*

$$\Pr[D(p, g, g^x \bmod p) = B(x)] \leq 1/2 + \eta(k),$$

where p is a random k -bit prime, g generator, $x \in_R \mathbb{Z}_p^*$.

Using this lemma, we can prove the following theorem.

Theorem 1. *If, instead of outputting $x_n, x_{n-1}, x_{n-2}, \dots, x_1$ we output $B(x_n), B(x_{n-1}), B(x_{n-2}), \dots, B(x_1)$, we get a PRG as defined above, as long as the DL assumption holds.*

We will prove the theorem first, then the lemma.

Proof of Theorem 1 Suppose it's not a PRG. Then there exists a bit predictor A .

How will we prove the lemma? By *reduction* (note that this is the first reduction proof we will see!). That is, from such an evil A that violates pseudorandomness, we will build D that violates claim 1. Hence, it will be a contradiction.

So, now, let's build.

First of all, as input we are given $p, g, y = g^x$, and we have to predict $B(x)$. We can use A to help.

Notice that A has to succeed in predicting some bit: first, second, \dots , i -th, \dots , n -th. Pick i at random between 1 and n .

Set $x_{n-i+2} = y$ and compute $x_{n-i+3} = g^{x_{n-i+2}}, x_{n-i+4} = g^{x_{n-i+3}}, \dots, x_n = g^{x_{n-1}}$. When A asks for the first bit, give it $B(x_n)$; for the next bit, give it $B(x_{n-1})$, and so on. Note that A is getting the same answers as it would be getting when G is run on $x_1 = \log_g(\log_g(\log_g(\dots(x)\dots)))$ (there are $n - i$ logs here). Note also that if x is distributed uniformly, so is x_1 , because modular exponentiation (and hence its inverse, discrete logarithm) is a permutation \mathbb{Z}_p^* . Hence, what A observes is the same as the output of the PRG on a

completely random input x_1 . So the distribution of the inputs to A is the same as in `experiment-predict`, so A would have the same probability of success.

Recall that A guesses some bit of the string. Because we chose i at random, A has $1/n$ probability of trying guessing exactly the i -th bit, which is $B(x_{n-i+1}) = B(x)$, which is what we need. So in $1/n$ of the cases, we will use A 's guess. Else (if A guesses before the i -th bit, or asks for the i -th bit which we don't know), we simply guess a bit b at random.

Suppose A has probability of $1/2 + \epsilon$ of being correct. Then what is the probability of D being correct? It's $(1/2)(n-1)/n + (1/n)(1/2 + \epsilon) = 1/2 + \epsilon/n$. So if A 's advantage over $1/2$ is not negligible, neither is D 's. Contradiction.

Proof of Lemma 1 Suppose there is a predictor. Then we'll build a machine that takes discrete logs. Recall from PS2, there are two square roots of g^x , if x is even. One is $r_1 = g^{x/2}$, and the other is $r_2 = g^{x/2+(p-1)/2}$. Note that $B(x/2) = 0$ and $B(x/2 + (p-1)/2) = 1$, so if we have a predictor D , then we can distinguish r_1 from r_2 . So here's the algorithm for taking discrete logarithms using the bit predictor: consider $y = g^x$. Find the last bit of x (by simply seeing if y is a square or not by raising it to $(p-1)/2$). If it's 1, divide y by g to make it zero. Now take the square root of y modulo p (there are efficient algorithms for it). You have two roots r_1 and r_2 . Get r_1 (by using D). Thus, r_1 is the same as y , except x is right-shifted by one bit. And you know the bit that came out. Repeat, until you get all of x bit-by-bit.

What's wrong with this proof? It only works if D is perfect. It's more complicated if D only succeeds sometimes. We won't do it in class; see [BM84].

Discussion

This pseudorandom generator requires one modular exponentiation for each output bit. Naive algorithms for modular exponentiation take $\Theta(k^3)$ bit operations (where k the length of the modulus and the exponent); more sophisticated algorithms can do only slightly better. Common values for $|p|$ given the strength of today's discrete logarithm algorithms are somewhere in the range of 1,024–2,048 bits; a single pseudorandom bit would take a few milliseconds to compute on today's PCs.

Besides its relatively low speed, another disadvantage of this PRG is that one needs to know the number n of output bits in advance (because bits are output backwards). We will fix it in the next couple of lectures.

Finally, recall that pseudorandom generators were defined to take truly random strings as inputs, rather than primes, generators, etc. This minor technical point can be rectified in one of two ways. First, we can redefine the PRG to work with "public values" (in this case, p and g) that are known to everyone, including the bit-predictor. Then x , which is a truly random seed, is the only input remaining. Alternatively, we can redefine the PRG to take a truly random string as input, and use it for generating a random p , a random g and a random x before starting the actual bit generation.

References

- [BM84] M. Blum and S. Micali. How to generate cryptographically strong sequences of pseudo-random bits. *SIAM Journal on Computing*, 13(4):850–863, November 1984.