Homework 5

1. **Stretching PRG Output.** (10 points) Suppose we are given a length-doubling PRG G such that

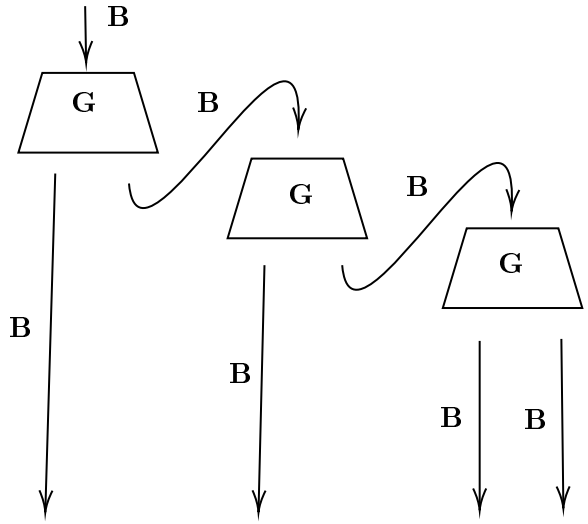
$$G : \{0,1\}^B \to \{0,1\}^{2B}$$

Using G, construct a new PRG G' such that

$$G' : \{0,1\}^B \to \{0,1\}^{2020B}$$

(Remark: We do not need a security proof. You should only use the PRG G to construct the new PRG G'. In particular, you should not use any other cryptographic primitive like one-way function etc.)

G' should start off by running a bit string B through G. From there, it should continually feed the lower B bits from the newly generated bit string of length 2B into G until the total length is 2020B. Executing G once would give you a bit string of length 2B. Twice would give you length 3B, three times would give you length 4B, etc. If you execute G a total of 2019 times, then the length of the final bit string would be 2020B. An illustration is provided below.



———- FINAL BIT STRING (length 4B) ———-

Basically, using the above method, after executing G 2019 times, you will have a bit string of length 2020B.

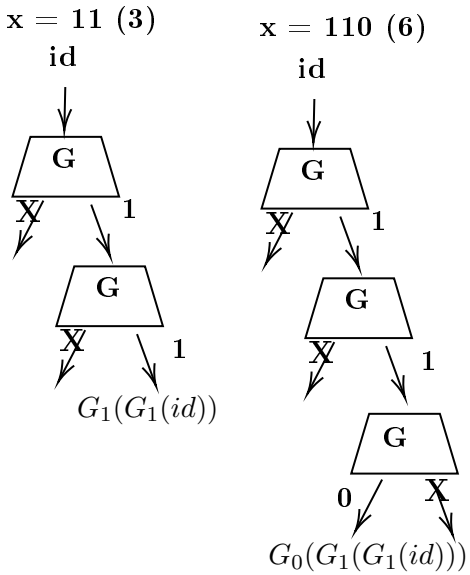
2. **New Pseudorandom Function Family.** (7+8+10) Let G be a length-doubling PRG $G\colon \{0,1\}^B \to \{0,1\}^{2B}$. Recall the basic GGM PRF construction presented below.

> - Define $G(x) = (G_0(x), G_1(x))$ where $G_0, G_1 : \{0,1\}^B \to \{0,1\}^B$
> - We define $g_{\mathsf{id}}(x_1, x_2, \ldots x_n)$ as $G_{x_n}(\ldots G_{x_2}(G_{x_1}(\mathsf{id}))\ldots)$ where $\mathsf{id} \xleftarrow{\$} \{0,1\}^B$.

Recall that in the class we studied that g_{id} is a PRF family for $\{0,1\}^n \to \{0,1\}^B$, for a fixed value of n when the key id is picked uniformly at random from the set $\{0,1\}^B$.

(a) (7 points) Why is the above-mentioned GGM construction not a pseudorandom function family from the domain $\{0,1\}^*$ to the range $\{0,1\}^B$? (Note that $\{0,1\}^*$ means that the length of the input to the PRF is arbitrary)

Let's consider the example inputs 3(11), 6(110), and 12(1100). Note that they all have a differing number of bits, but the first two bits are identical among the three. Executing GGM construction on the inputs result in the following:



As you can see, the beginning of these GGM constructions are identical (the first two bits 11). The output of $g_{id}(11) = G_1(G_1(id))$, and the output of $g_{id}(110) = G_0(G_1(G_1(id))$. Notice that the output of $g_{id}(110)$

is simply G_0 applied on the output of $g_{id}(11)$. This means that we can invert the PRG G, violating the security of the PRF Family. Since we know G and $g_{id}(11)$, we can calculate $g_{id}(111)$.

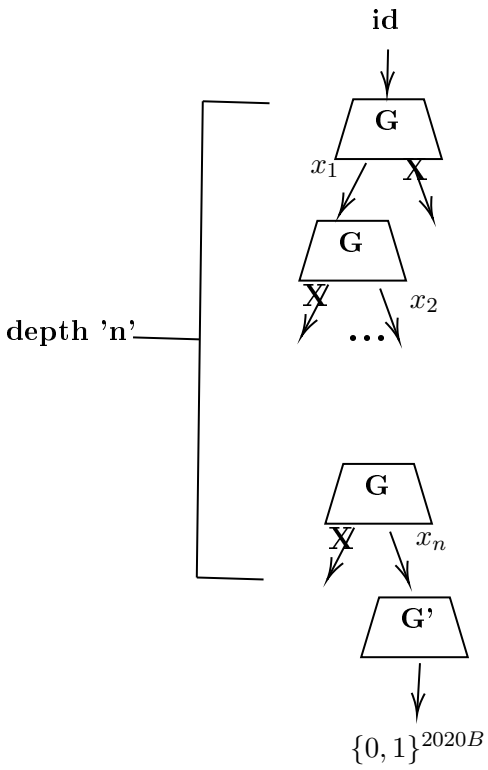
Since the output no longer appears uniformly random, and can be predicted, we can conclude that the mentioned GGM Construction is NOT a PRFF.

(b) (8 points) Given a length-doubling PRG G: $\{0,1\}^B \to \{0,1\}^{2B}$, construct a PRF family from the domain $\{0,1\}^n$ to the range $\{0,1\}^{2020B}$.
(Remark: Again, in this problem, do not use any other cryptographic primitive like one-way function etc. You should only use the PRG G in your proposed construction.)

To start, I will declare a few different variables:

- id **will be selected uniformly at random from the set** $\{0,1\}^B$
- x **will be selected selected uniformly at random from the domain** $\{0,1\}^n$
- $G' : \{0,1\}^B \to \{0,1\}^{2020B}$**, from Q1 (this only uses G).**

Now, the GGM construction for the PRFF is detailed below:



(c) (10 points) Consider the following function family $\{h_1, \ldots, h_\alpha\}$ from the domain $\{0,1\}^*$ to the range $\{0,1\}^B$. We define $h_{\mathsf{id}}(x) = g_{\mathsf{id}}(x, [\ |x|\ ]_2)$, for $\mathsf{id} \in \{1, 2, \ldots, \alpha\}$. Show that $\{h_1, \ldots, h_\alpha\}$ is not a secure PRF from $\{0,1\}^*$ to the range $\{0,1\}^B$.

(*Note*: The expression $[|x|\,]_2$ represents the length of x in n-bit binary expression. (n denotes the length of x))

For this, I will again use the example of $x = 3(11)$ vs $x = 6(110)$. Recall that the binary representation of 3 has 2 bits, and the binary representation of 6 has 3 bits.

$h_{id}(x) = g_{id}(x, [|x|\,]_2)$
$h_{id}(11) = g_{id}(11, 10)$
$h_{id}(110) = g_{id}(110, 11)$

The fact that the domain for this function family takes inputs of different lengths allows us to compromise the security of g_{id}, in a similar fashion to 2a and 3.

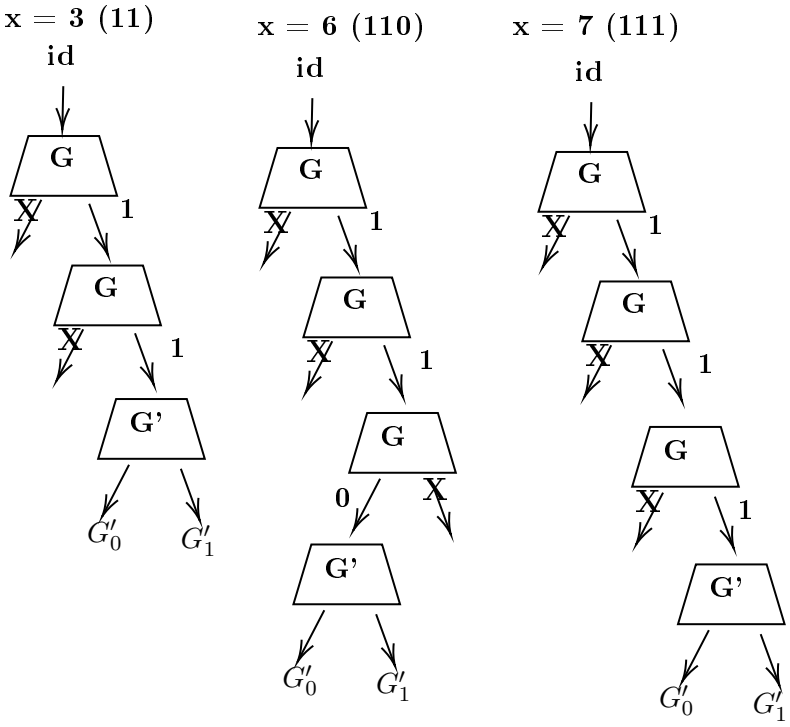
If you keep testing inputs, it becomes clear that, for all 2-bit numbers, the second argument of g_{id} will be 10, 11 for all 3-bit numbers, 100 for all 4-bit numbers, and so on. This means that the output is not completely independent of the input, which violates the definition of a PRFF.

3. **Variant of Pseudorandom Function Family.** (15 points) Let G be a length-doubling PRG G: $\{0,1\}^B \to \{0,1\}^{2B}$ and $G' : \{0,1\}^B \to \{0,1\}^T$ be a PRG where $T \geqslant B$. The following construction is suggested to construct a PRF family from $\{0,1\}^* \to \{0,1\}^T$. (Note that $\{0,1\}^*$ means that the length of the input to the PRF is arbitrary)

> - Define $G(x) = (G_0(x), G_1(x))$ where $G_0, G_1 : \{0,1\}^B \to \{0,1\}^B$
> - Let $G' : \{0,1\}^B \to \{0,1\}^T$ be a PRG.
> - We define $g_{\mathsf{id}}(x_1, x_2, \ldots x_n)$ as $G'(G_{x_n}(\ldots G_{x_2}(G_{x_1}(\mathsf{id}))\ldots))$ where $\mathsf{id} \xleftarrow{\$} \{0,1\}^B$.

Prove that the above-mentioned PRF construction is not secure when $G' = G$. (Note that when $G' = G$, then $T = 2B$).

Imagine inputs 3(11), 6(110), and 7(111). Note that three only has 2 bits, and all three numbers start with the bits 11. This means that the evaluation of GGM up to the first 2 bits deep will be identical for all three inputs. Also note that we are fixing the function *id* chosen from the PRFF. Evaluating GGM Construction results in the following:



Evaluating with $x = 11(3)$ will give us $g_{id}(11) = (G'_0(G_1(G_1(id))), G'_1(G_1(G_1(id))))$. For ease of explanation, I will refer to $a = G'_0(G_1(G_1(id)))$ and $b = G'_1(G_1(G_1(id)))$

Then, evaluating with $x = 110(6)$ will give us $g_{id}(110) = (G'_0(a), G'_1(a))$.
The Goldreich-Levin hardcore predicate theorem states that, if you have a OWP g_{id}, the output looks random, and is unpredictable. However, since we have the input and output of G' at an input a, we know what G' does. G' has been compromised. We can predict the output of the PRG G'.

Finally, let's evaluate with $x = 111(7)$. The GGM construction gives is $g_{id}(111) = (G_0(b), G'_1(b))$. Since we have compromised G' and know what it does, we can predict this output. This means that the PRFF is NOT SECURE.

4. **OWF.** (10 points) Let $f : \{0,1\}^n \to \{0,1\}^n$ be a one-way function. Define $g : \{0,1\}^{2n} \to \{0,1\}^{2n}$ as

$$g(x_1, x_2) = f(x_2)||x_1 \oplus x_2 \oplus 1^n$$

where $x_1 \in \{0,1\}^n$, $x_2 \in \{0,1\}^n$ and 1^n denotes a string of n bits. Show that g is also a one-way function.

Hint. Suppose there exists an efficient adversary $\mathcal{A}$ that inverts the function g . You should now construct a new efficient adversary $\mathcal{A}'$ that uses $\mathcal{A}$ as a subroutine to invert the function f.

Let's suppose that g is NOT a OWF, and the efficient adversary $\mathcal{A}$ knows how to invert g.

Now, let's say that $f(x_2)$ is given to the efficient adversary $\mathcal{A}'$. $\mathcal{A}'$ takes $f(x_2)$ and creates a new bit string y to give to the efficient adversary $\mathcal{A}$. y should be the output of $g(x_1, x_2)$. The bit string $y = f(x_2)||<$random n bits$>$. The first n bits of y will be $f(x_2)$, since that's how g is defined. The last n bits of y can be anything because $x_1 \oplus x_2 \oplus 1^n$ can result in anything, even with x_2 fixed at some value.

Now, we give $\mathcal{A}$ the bit string y, and since it knows how to invert g, it returns an (x_1, x_2) that satisfies $g(x_1, x_2) = y$.

Since $\mathcal{A}'$ now knows x_2, it returns that, successfully inverting f. Since $\mathcal{A}'$ has inverted f, this implies that f is not a OWF, which is a contradiction of what we were told. Therefore, by contradiction, g is also a OWF.

5. **Encryption using Random Functions.** (15+10 points) Let $\mathcal{F}$ be the set of all functions $\{0,1\}^n \to \{0,1\}^n$. Consider the following private-key encryption scheme.

> - $\mathsf{Gen}()$: Return $\mathsf{sk} = F$ uniformly at random from the set $\mathcal{F}$
> - $\mathsf{Enc}_{\mathsf{sk}}(m)$: Return (c, r), where r is chosen uniformly at random from $\{0,1\}^n$, $c = m \oplus F(r)$, and $\mathsf{sk} = F$.
> - $\mathsf{Dec}_{\mathsf{sk}}(\widetilde{c}, \widetilde{r})$: Return $\widetilde{c} \oplus F(\widetilde{r})$.

(a) (15 points) Suppose we want to ensure that even if we make 10^{10} calls to the encryption algorithm, all randomness r that are chosen are distinct with probability $1 - 2^{-201}$. What value of n shall you choose?
For a sample size S:
$S = \{0,1\}^n$**, and** $|S| = 2^n$

And if we pick a sample from the set S $K = 10^{10}$ times, we would like the probability that all samples are distinct with probability at least $1 - 2^{-201}$.

We first recall that, for a sample space, $\mathbb{P}$[K samples distinct]$\approx exp(\frac{-K^2}{2*|S|})$

It follows that:

$exp(\frac{-K^2}{2*|S|})$
$exp(\frac{-10^{20}}{2*2^n})$
$exp(\frac{-10^{20}}{2^{n+1}})$

Remembering that $1 - 2^{-201} \approx exp(2^{-201})$, We now must ensure that:
$exp(\frac{-10^{20}}{2^{n+1}}) \geqslant exp(-2^{-201})$
$\frac{-10^{20}}{2^{n+1}} \geqslant -2^{-201}$
$\frac{10^{20}}{2^{n+1}} \leqslant 2^{-201}$
$10^{20} \leqslant 2^{n-200}$
$n \geqslant \frac{20*\ln 10}{\ln 2} + 200$
$n \geqslant 266.438$

And since n must be an integer, we round it up to 267. Therefore, if we would like to make 10^{10} calls to the encryption algorithm, we can ensure that all randomness r that are chosen are distinct with probability at least $1 - 2^{-201}$ by using a value of $n = 267$.

(b) (10 points) Conditioned on the fact that all randomness r in the encryption schemes are distinct, prove that this scheme is secure.
This scheme is similar to one-time pad. We already know that one-time pad is perfectly secure with a secret key that cannot be computed by any adversary.

In order for us to say that this scheme is secure we must prove that $F(r)$ is completely unpredictable and random for every r.

The definition of a random function states the following:

Choose any $x_1 \in \mathbb{D}$.
$F(x_1)$ is uniformly random over $\mathbb{R}$, and now we know that $F(x_1) = y_1$.

Now, choose any $x_2 \in \mathbb{D}$.
If $x_2 = x_1$, then $F(x_1) = F(x_2)$. Otherwise, $F(x_2)$ is uniformly random over $\mathbb{R}$.

This means that, for any input given to F that you haven't seen the output, it is COMPLTELY IMPOSSIBLE to predict the output. Furthermore, since every randomness r given to F will be unique (for every input, we have never seen the output), we can confidently say that the output of F will always be completely unpredictable and random. No adversary will be able to compute the secret key.
Since one-time pad is perfectly secure with a completely random, uncomputable secret key, we can conclude that this scheme is secure.

6. **Birthday Paradox.** (10 points) Recall that the Birthday Paradox states that if we throw $m = c\sqrt{n}$ balls into n bins, then the probability that there exists a collision (i.e., a bin with at least two balls) is $\geqslant 0.99$, where $c > 0$ is an appropriate constant. An international university has 12 colleges. Moreover, the students of this university come from 100 different countries around the world. How many students (from the university) in a room will ensure with probability $\geqslant 0.99$ that there exists at least a pair of students such that they are from the same country, the same college, and they celebrate their birthday at the same month.

   We should think of the number of students m as balls, $m = c\sqrt{n}$.

   We should also think of the number of bins n as the possible combinations of colleges, countries, and months. Given that there are 12 different colleges, 100 different countries, and 12 different months, it follows that there are a total of $12 * 12 * 100 = 14,400$ possible combinations of those three, so there are 14,400 bins.

   So, $n = 14,400$
   and $m = c\sqrt{n} = c\sqrt{14,400} = 120c$

   Therefore, you just need $120c$ students in a room to ensure with probability $\geqslant 0.99$ that there exists at least one pair of students such that they are from the same country, the same college, and they celebrate their birthday at the same month.

7. **PRF.**(10 points) Suppose the set of functions $F_{\mathsf{id}}\colon \{0,1\}^n \to \{0,1\}^n$ forms a secure PRF when id is chosen uniformly at random from the set $\{0,1\}^n$.

   We are now constructing a new PRF family $G_{\mathsf{id}}\colon \{0,1\}^{2n} \to \{0,1\}^{2n}$, where $\mathsf{id} \in \{0,1\}^n$. This new function is defined as follows.

   $$G_{\mathsf{id}}(x_1, x_2) := \left(x_2 \oplus F_{\mathsf{id}}(x_1) \ , \ F_{\mathsf{id}}(x_2) \right)$$

   Is this new PRF secure or not?

   (If you think that it is secure, then prove that it is secure. If you think that it is insecure, then prove why this construction is insecure. You get no points for just writing Yes/No.)

   I do not think this PRF is secure. An adversary will notice that, whatever the value of x_2, the output of $G_{id}(x_1, x_2)$ ends in whatever $F_{id}(x_2)$ evaluates to (the last n bits).

   Using this, we can predict the output for some combination of (x_1, x_2) that has not been seen before. Here's a simple example:

   Input to $G_{id} = 001\ 010$
   $G_{id}(001, 010) = (010 \oplus F_{id}(001), F_{id}(010))$
   We know that the last n bits of $G_{id}(001, 010) = F_{id}(010)$.

   Input to $G_id = 010\ 010$
   We know $x_2 = 010$ because we chose it, and we know $F_{id}(x_1) = F_{id}(010)$ from the previous input we evaluated. Therefore, we can predict the following equation, since we know all parts of it.
   $G_{id}(010, 010) = (010 \oplus F_{id}(010), F_{id}(010))$

   Even though F_{id} itself is secure, this example shows a way that you can predict output for a PRF. This proves that G_{id} is NOT secure.

Collaborators :