Solution of Homework 5

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

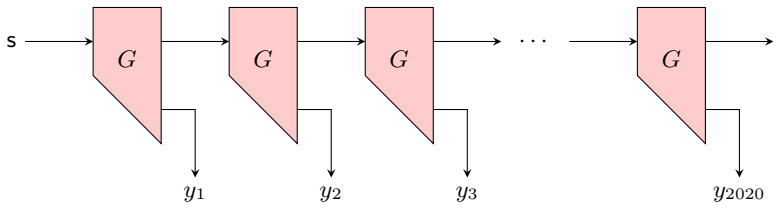
$$G : \{0,1\}^B \to \{0,1\}^{2B}$$

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.)

Solution.
We construct the PRG $G'\colon \{0,1\}^B \to \{0,1\}^{2020B}$ from the PRG $G\colon \{0,1\}^B \to \{0,1\}^{2B}$ using a construction similar to the construction of "arbitrary stretch PRG" from "one-bit extension PRG" constructed in the lectures. The output of $G'(s) = (y_1, y_2, \ldots, y_{2020})$ is computed as follows.



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)
Solution.
Consider the GGM construction of the PRF family. Given $z = g_{\mathsf{id}}(x)$ (where $x \in \{0,1\}$ is a bit), we can compute $z' = g_{\mathsf{id}}(x||y)$ (where $(x||y)$ represents the concatenation of the bit x with the bit y), for any x and y because $z' = g_{\mathsf{id}}(x||y) = G_y(G_x(id)) = G_y(z)$. On the other hand, we expect that for a random function F, the output of $F(x,y)$ to be uniformly at random and independent of $F(x)$. Therefore, the GGM construction is not a pseudo-random function family from $\{0,1\}^* \to \{0,1\}^B$.

(b) (8 points) Given a length-doubling PRG $G\colon \{0,1\}^B \to \{0,1\}^{2B}$, 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.)
Solution.
We can apply the PRG construction $G'\colon \{0,1\}^B \to \{0,1\}^{2020B}$ (From Problem 1) to the output of the GGM PRF family to get the PRF family for this problem.

(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))

Solution.

We query PRF on input $x = 1$ (whose length is 1). The binary representation of 1 is 1 and the binary representation of the length of $x = 1$ is also 1. It returns as output $h_{\mathsf{id}}(1) = g_{\mathsf{id}}(11) = G_1(G_1(\mathsf{id}))$. Note that 11 is the binary representation of $x' := 3$ with length 2 (whose binary representation is 10). Then, we have $h_{\mathsf{id}}(x') = g_{\mathsf{id}}(1110) = G_0(G_1(G_1(G_1(id)))) = G_0(G_1(h_{\mathsf{id}}(1)))$. Since the output of function G on different inputs is known, the output of functions G_0 and G_1 are also known. Therefore, after knowing $h_{\mathsf{id}}(1)$, we can predict $h_{\mathsf{id}}(x')$ by finding the value of $G_0(G_1(h_{\mathsf{id}}(1)))$ without knowing id. This shows that the given PRF can be distinguished from a truly random function and is not secure.

3. **Variant of Pseudorandom Function Family.** (15 points) Let G be a length-doubling PRG $G\colon \{0,1\}^B \to \{0,1\}^{2B}$ 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 $T = 2B$).

Solution.
Suppose $G' = G$ and id is chosen uniformly at random from $\{0,1\}^B$. Then, we can query PRF on some arbitrary input of arbitrary length n like $x_1, x_2, \ldots, x_n$. Then, it returns as output

$$G(G_{x_n}(\ldots G_{x_2}(G_{x_1}(\mathsf{id}))\ldots)) = \big(G_0(G_{x_n}(\ldots G_{x_2}(G_{x_1}(\mathsf{id}))\ldots)), G_1(G_{x_n}(\ldots G_{x_2}(G_{x_1}(\mathsf{id}))\ldots))\big)$$

This means that after such query, we find out both $G_0(G_{x_n}(\ldots G_{x_2}(G_{x_1}(\mathsf{id}))\ldots))$ and $G_1(G_{x_n}(\ldots G_{x_2}(G_{x_1}(\mathsf{id}))\ldots))$. So, we can apply G on them to find

$$G\left(G_0(G_{x_n}(\ldots G_{x_2}(G_{x_1}(\mathsf{id}))\ldots))\right)$$

and

$$G\left(G_1(G_{x_n}(\ldots G_{x_2}(G_{x_1}(\mathsf{id}))\ldots))\right)$$

which are respectively equal to the output of PRF on inputs $x_1, x_2, \ldots, x_n, 0$ and $x_1, x_2, \ldots, x_n, 1$. This shows that the given PRF is not secure because we can predict its output on new inputs in an efficient time and so we can distinguish it from a truly random function in an efficient time.

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.

Solution.

Assuming that we have an efficient adversary $\mathcal{A}$ that inverts the function g, we can construct an efficient adversary $\mathcal{A}'$ that inverts the function f in the following manner.

$\mathcal{A}'(y_f)$:

- Define $y_g := y_f \| 1^n$.
- Run $\mathcal{A}(y_g)$, denote the output of this algorithm as $x = (x_1, x_1) := \mathcal{A}(y_g)$.
- Output x_1.

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?
Solution.

Solution.
We have proved in class that the probability that all k samples chosen uniformly at random from a set S to be distinct is roughly equal to $\exp(\frac{-k^2}{2|S|})$. In this question, $|S| = 2^n$ and $k = 10^{10}$. We need to have:

$$\exp\left(-\frac{10^{20}}{2^{n+1}}\right) \geqslant 1 - 2^{-201} \approx \exp\left(-2^{-201}\right) \tag{1}$$

$$\iff \left(-\frac{10^{20}}{2^{n+1}}\right) \geqslant \left(-2^{-201}\right) \tag{2}$$

$$\iff 10^{20} \leqslant 2^{n-200} \tag{3}$$

$$\iff 20\log_2(10) \leqslant n - 200 \tag{4}$$

$$\iff n \geqslant 200 + 20\log_2(10) \approx 266.43 \tag{5}$$

$$\iff n \geqslant 267 \tag{6}$$

(b) (10 points) Conditioned on the fact that all randomness r in the encryption schemes are distinct, prove that this scheme is secure.

Solution.

We assume that for any i and j, $r_i \neq r_j$. Note that F is chosen uniformly at random from the set of all functions that map a string of n bits to a string of n bits. Therefore, the values $F(r_1), \ldots, F(r_q)$ are independent and have uniform distribution. This implies that $c_1 = m_1 \oplus F(r_1), \ldots, c_q = m_q \oplus F(r_q)$ are all uniform and independent of each other. This means that the encryption of $m_1, m_2, \ldots, m_{q-1}$ does not reveal any information about the encryption of m_q. This scheme is like one time pad. Because, it is equivalent to XORing the sequence $m_1 \ldots m_q$ with uniformly random sequence $F(r_1) \ldots F(r_q)$ to get $c_1 \ldots c_q$.

$$\begin{aligned}
&\Pr[C_1 = c_1, \ldots, C_q = c_q | M_1 = m_1, \ldots, M_q = m_q] && (7)\\
=&\Pr[m_1 \oplus F(r_1) = c_1, m_2 \oplus F(r_2) = c_2, \ldots, m_q \oplus F(r_q) = c_q] && (8)\\
=&\Pr[F(r_1) = m_1 \oplus c_1, F(r_2) = m_2 \oplus c_2, \ldots, F(r_q) = m_q \oplus c_q] && (9)\\
=&\Pr[F(r_1) = m_1 \oplus c_1] \times \cdots \times \Pr[F(r_q) = m_q \oplus c_q] = \left(\frac{1}{2^n}\right)^q && (10)
\end{aligned}$$

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.
Solution.
We first need to count the whole number of bins. Each bin denotes a tuple

$$\text{(college, country, month of birth).}$$

Since there are 12 colleges, 100 countries, and 12 months, there are

$$12 \times 100 \times 12 = 14400$$

in total. Therefore, according to Birthday Paradox, if we choose at least $\left\lceil c\sqrt{14400} \right\rceil = \lceil 120c \rceil$ [1] Purdue students, then with probability at least 0.99, there are at least two students among them who are from the same country and the same college and celebrate their birthday at the same month.

[1] $\lceil x \rceil$ denotes the least integer greater than or equal to x.

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.)

   Solution.
   We show that the given PRF is not secure. For a given $(x_1, x_2) \in \{0,1\}^n \times \{0,1\}^n$, we have:

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

   and for $(x_1', x_2) \in \{0,1\}^n \times \{0,1\}^n$ (where $x_1' \neq x_1$) we have:

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

   Suppose you know that the output of PRF on input (x_1, x_2) is (c, d) where $c = x_2 \oplus F_{\mathsf{id}}(x_1)$ and $d = F_{\mathsf{id}}(x_2) \in \{0,1\}^n$. Then, the second n bits of the output of G_{id} on input (x_1', x_2), will be d. This implies that for any id, the second n bits of the output of G_{id} on inputs (x_1, x_2) and (x_1', x_2) are the same; while the output of a truly random function on input (x_1', x_2) is independent of its output on input (x_1, x_2). So, the new PRF is not a secure PRF.

Collaborators :