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

$$G' : \{0,1\}^B \to \{0,1\}^{2024B}$$

(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.
Given $G : \{0,1\}^B \to \{0,1\}^{2B}$, we can construct the new PRG $G' = \{0,1\}^B \to \{0,1\}^{2024B}$ by using a similar construct shown in lecture as shown in the figure below, regarding the construction of l-stretched PRGs using one bit extended PRGs. Given input x, let y denote the output of the PRG from index $B+1$ to $2B$, which is of length B. Thus, the output of $G'(X) = (y_1, y_2, y_3, ..., y_{2024})$.

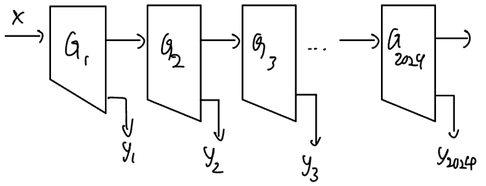

Figure 1: PRG extension as shown in lecture

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.

Reusing the GGM PRF construction above, let us consider two bits $(a, b) \in \{0,1\}$. Let's also define $z = g_{id}(a)$ and $z' = g_{id}(a||b)$, where $||$ represents the concatenation operator, and $(a||b)$ represents concatenation of bit a with bit b.

Suppose that the GGM construction is a pseudorandom function family from the domain $\{0,1\}^* \to \{0,1\}^B$. We now have $z' = g_{id}(a||b) = G_b(G_a(id)) = G_b(z)$ (By definition above of GGM PRF construction).

This means that there exists a function F, where outout of $F(a, b)$ is not uniformly random and independent of $F(a)$. This means that the function will definitely not appear to be random, and thus not pseudorandom in nature.

This forms a contradiction that the GGM construction is a pseudorandom function family from domain $\{0,1\}^* \to \{0,1\}^B$ and thus, it is not a pseudorandom function family.

(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 $\{0,1\}^{2024B}$.

(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.

Using the PRG construction from question 1, we can obtain $G' : \{0,1\}^B \to \{0,1\}^{2024B}$. Then, we can apply G' to the output of the GGM PRF family above(Suppose output $= z = g_{id}(x)$, we just need to compute $G'(z)$), and we obtained the PRF family from the domain $\{0,1\}^n \to \{0,1\}^{2024B}$.

(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.

From lecture, we know that id(secret key sk) is private and not known to adversaries.

First, an adversary can send input $a = 1$ which is 1 in binary. Hence, $|a| = 1$ and thus $h_{id}(1) = g_{id}(11) = G_1(G_1(id))$, by definition.

Then, for input $b = 3$, its corresponding binary representation is 11 and thus, $|b| = 2$. Then, $h_{id}(3) = g_{id}(1110) = G_0(G_1(G_1(G_1(id)))) = G_0(G_1(h_1(id)))$.

Similarly, for input $c = 6$, its corresponding binary representation is 110 and thus, $|c| = 3$. Then, $h_{id}(6) = g_{id}(11011) = G_1(G_1(G_0(G_1(G_1(id))))) = G_1(G_1(G_0(h_1(id))))$.

Then, as adversaries are able to obtain the input of G on different inputs, and also the output of G_0 and G_1 correspondingly, hence, upon obtaining $h_{id}(1)$, an adversary is able to predict $h_{id}(a)$ and $h_{id}(b)$ without obtaining id at all. Thus, this shows that $\{h_1, ..., h_\alpha\}$ produces outputs that do not appear random in nature and we are able to distinguish it from a truly random function, making it 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 <u>not</u> secure when $G' = G$. (Note that when $G' = G$, then $T = 2B$).

Solution.
When $G = G'$, and assuming that id is chosen uniformly at random from $\{0,1\}^B$, we can query the above PRF on an arbitrary input of arbitrary length n, where the input is $x_1, x_2, ..., x_n$.
Next, since $G = G'$, querying the PRF using the arbitrary input $x_1, x_2, ..., x_n$ would produce $G_0(G_{x_n}(...G_{x_2}(G_{x_1}(id))))$ and $G_1(G_{x_n}(...G_{x_2}(G_{x_1}(id))))$, which is equivalent to the output of the PRF on the inputs of $x_1, x_2, ..., x_n, 0$ and $x_1, x_2, ..., x_n, 1$ respectively. This means that we are able to predict outputs on new inputs and thus, the output does not appear to be random and hence, the given PRF 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

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

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.

Using the hint, suppose that we have an efficient adversary $\mathcal{A}$ that inverts the function g. We can construct a new efficient adversary $\mathcal{A}'$ that uses $\mathcal{A}$ as a subroutine to invert the function f, as follows:

y_f := output upon applying f on $x_1 \oplus 1^n : f(x_1 \oplus 1^n)$
$y_g := y_f||1^n$ (to ensure that $y_g \in \{0,1\}^{2n}$)
Compute $\mathcal{A}(y_g)$, and denote output $x = (x_1, x_2)$.
Then output $(x_1 \oplus 1^n)$.
Thus, this shows that g is also a one-way function.

5. **Encryption using Random Functions.** (15+10 points) Let $\mathcal{F}$ be the set of all functions $\{0,1\}^n \rightarrow \{0,1\}^n$. 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^{30} calls to the encryption algorithm, all randomness r that are chosen are distinct with probability $1 - 2^{-401}$. What value of n shall you choose?
Solution.
From Lecture 19 on Random Functions and Encrypting Long Messages,
Probability that all k samples are all distinct, given that all k samples are chosen uniformly at random from set S, is estimated to be equal to $exp(\frac{-k^2}{2|S|})$.
Hence, given context of the question, $k = 10^{30}$ and $|S| = 2^n$, the inequality is as follows:

$$exp(\frac{-(10^{30})^2}{2(2^n}) \geqslant 1 - 2^{-401}$$

$$exp(\frac{-10^{60}}{2^{n+1}}) \geqslant exp(-2^{-401})$$

(Since 1-x $\approx exp(-x)$)
$\Leftrightarrow$

$$\frac{10^{60}}{2^{n+1}} \leqslant 2^{-401}$$

(Taking natural log on both sides)
$\Leftrightarrow$

$$10^{60} \leqslant 2^{n-400}$$

$\Leftrightarrow$

$$60log_2(10) \leqslant n - 400$$

(Taking log_2 on both sides
$\Leftrightarrow$

$$n \geqslant 400 + 60log_2(10)$$

$\Leftrightarrow$

$$n \geqslant 599.316$$

$\Leftrightarrow$

$$n \geqslant 600$$

(Nearest integer)
Hence, value of n that I will choose: $n = 600$.

(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 first note that F is chosen uniformly at random from the set $\mathcal{F}$, of all functions $\{0,1\}^n \rightarrow \{0,1\}^n$. This means that the set of all values $F(r_i), ..., F(r_j)$, $\forall i \neq j$, $r_i, r_j \in \{0,1\}^n$, are independent and uniformly distributed. This also means that for $Enc_{sk}(m)$, $c_1 = m_1 \oplus F(r_1), c_2 = m_2 \oplus F(r_2), ..., c_j = m_j \oplus F(r_j)$ are all independent of one another and uniformly distributed.

Then, this means that an adversary who observes $m_1, m_2, ...m_{j-1}$ would not have any advantage to predict output of encryption of m_j, as the encryption of $m_1, m_2, ...m_{j-1}$ does not reveal information about the encryption of m_j.
In probability terms:
$Pr[C_1 = c_1, C_2 = c_2, ..., C_j = c_j | M_1 = m_1, M_2 = m_2, ..., M_j = m_j]$
$= Pr[m_1 \oplus F(r_1) = c_1, m_2 \oplus F(r_2) = c_2, ..., m_j \oplus F(r_j) = c_j]$ (By definition of $Enc_{sk}(m)$)
$= Pr[m_1 \oplus F(r_1) = c_1] \times Pr[m_2 \oplus F(r_2) = c_2] \times ... \times Pr[m_j \oplus F(r_j) = c_j]$ (By independence)
$= (\frac{1}{2^n})^j$
Hence, 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 121 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.
   Using the birthday paradox, the number of bins required $= 12 \times 121 \times 12$(12 colleges, 121 different countries, 12 months) = 17424.
   Thus, by birthday paradox, number of students in a room, m
   $= \lceil c \cdot \sqrt{17424} \rceil, \forall c > 0$
   $= \lceil 132 \cdot c \rceil$, will ensure with probability $\geqslant 0.99$ that there exists at least a pair of students such that they are from the same country, college, and 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.

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

   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.
   This given PRF is not secure.
   Suppose we have $(x_1, x_2) \in \{0,1\}^n \times \{0,1\}^n$, and $(x_1', x_2) \in \{0,1\}^n \times \{0,1\}^n$, such that $x_1 \neq x_1'$.

   Then, we have the following:
   $G_{id}(x_1, x_2) := (x_2 \oplus F_{id}(x_1), F_{id}(x_2))$ and
   $G_{id}(x_1', x_2) := (x_2 \oplus F_{id}(x_1'), F_{id}(x_2))$.

   Let us denote the output of PRF on input (x_1, x_2) be (a, b), where $a = x_2 \oplus F_{id}(x_1)$, $b = F_{id}(x_2)$. This means that the output of the PRF on input $(x_1', x_2) = (a', b)$. This means that an adversary does not need to know id, but would be able to recognise that the second n bits of the output of G_{id} on inputs (x_1, x_2) and (x_1', x_2) are equal. This shows that the outputs of the PRF is not independent of one another, making the PRF not a truly random function. Hence, the new PRF is not a secure PRF.

8. **One Way Function (12 points)** In this problem, we will show that a one-way function cannot have a small range.

(a) (**6 points**) First, we need an inequality. Prove that

$$p_1^2 + p_2^2 + \cdots + p_k^2 \geqslant \frac{1}{k},$$

where $p_1, \ldots, p_k \geqslant 0$ such that $p_1 + \cdots + p_k = 1$.

Hint: The Cauchy-Schwarz inequality says that for non-negative x_i and y_is, the following inequality holds.

$$\sum_{i=1}^{k} x_i y_i \leqslant \left(\sum_{i=1}^{k} x_i^2\right)^{1/2} \left(\sum_{i=1}^{k} y_i^2\right)^{1/2}$$

First, let's define $x_i = p_i$ and $y_i = 1, \forall i \in \{1, 2, ..., k\}$.
Then, by Cauchy-Schwarz inequality,

$$\sum_{i=1}^{k} p_i \cdot 1 \leqslant (\sum_{i=1}^{k} p_i^2)^{\frac{1}{2}} \cdot (\sum_{i=1}^{k} 1^2)^{\frac{1}{2}}$$

$\Leftrightarrow$

$$(\sum_{i=1}^{k} p_i)^2 \leqslant (\sum_{i=1}^{k} p_i^2) \cdot (\sum_{i=1}^{k} 1^2)$$

(Taking square on both sides)
$\Leftrightarrow$

$$1^2 \leqslant \sum_{i=1}^{k} p_i^2 \cdot k$$

(Since from question above, $\sum_{i=1}^{k} p_i = 1$, and $\sum_{i=1}^{k} 1^2 = k$)
$\Leftrightarrow$

$$\sum_{i=1}^{k} p_i^2 \geqslant \frac{1}{k}$$

(Algebraic manipulation)
Thus, since $p_1^2 + p_2^2 + ... + p_k^2 = \sum_{i=1}^{k} p_i^2$,
$p_1^2 + p_2^2 + ... + p_k^2 \geqslant \frac{1}{k}$(proven).

(b) (**6 points**) Suppose $f\colon \{0,1\}^n \to \{0,1\}^\ell$ be a function. The honest challenger samples $x \xleftarrow{\$} \{0,1\}^n$ and sends the challenge $y = f(x)$ to us. We, on input y, run the following algorithm

```
Sample $r \xleftarrow{\$} \{0,1\}^n$
Compute $t = f(r)$
If $y == t$: return $\widetilde{x} = r$
Else: return $\widetilde{x} = 0$.
```

Prove that the probability of our algorithm successfully inverting the one-way function f is $\geqslant \frac{1}{2^\ell}$ (over the random choice of $x \xleftarrow{\$} \{0,1\}^n$ by the honest challenger). Equivalently, prove that

$$\Pr_{x,r \xleftarrow{\$} \{0,1\}^n} \left[f(\widetilde{x}) = f(x) \right] \geqslant \frac{1}{2^\ell}.$$

(Remark: It is not necessary that the size of the preimage of $y \in \{0,1\}^\ell$ is same as the size of the preimage of another $y' \in \{0,1\}^\ell$.)

First, we note that for a given y, there are 2^n possible preimages that the challenger could have chosen. Since the function f maps from $\{0,1\}^n \to \{0,1\}^\ell$, and the output length is ℓ, this means that there are 2^n possible inputs and 2^ℓ possible outputs.

Assuming that f is surjective, each output has at least $2^n/2^\ell$ preimages. This is so because the total number of possible inputs must be distributed across 2^ℓ outputs.

Hence, given that the algorithm samples one of the 2^n inputs uniformly at random, the probability that it picks the correct preimage of y(or one of the correct preimages if there are multiple) is at least $\frac{1}{2^\ell}$.

Collaborators :