Tutorial 2: PRGs and one time pad

Exercise 1. *Introduction to Computational Hardness Assumptions*

Definition 1 (Decisional Diffie-Hellman distribution). *Let $\mathbb{G}$ be a cyclic group of prime order q, and let g be a publicly known generator of $\mathbb{G}$. The* decisional Diffie-Hellman *distribution (DDH) is, $D_{\mathrm{DDH}} = (g^a, g^b, g^{ab}) \in \mathbb{G}^3$ with a, b sampled independently and uniformly at random in $\mathbb{Z}_q$.*

Definition 2 (Decisional Diffie-Hellman assumption). *The decisional Diffie-Hellman assumption states that there exists no probabilistic polynomial-time distinguisher between D_{DDH} and (g^a, g^b, g^c) with a, b, c sampled independently and uniformly at random in $\mathbb{Z}_q$.*

1. Does the DDH assumption hold in $\mathbb{G} = (\mathbb{Z}_p, +)$ for $p = \mathcal{O}(2^\lambda)$ prime?
2. Same question for $\mathbb{G} = (\mathbb{Z}_p^\star, \times)$ of order $p-1$.
3. Now we take $\mathbb{Z}_p$ such that $p = 2q+1$ with q prime (also called a *safe-prime*). Let us work in a subgroup $\mathbb{G}$ of order q in $(\mathbb{Z}_p^\star, \times)$.
   - **(a)** Given a generator g of $\mathbb{G}$, propose a construction for a function $\hat{G} : \mathbb{Z}_q \to \mathbb{G} \times \mathbb{G}$ (which may depend on public parameters) such that $\hat{G}(U(\mathbb{Z}_q))$ is computationally indistinguishable from $U(\mathbb{G} \times \mathbb{G})$ based on the DDH assumption on $\mathbb{G}$ (where, in $\hat{G}(U(\mathbb{Z}_q))$, the probability is also taken over the public parameters of $\hat{G}$).
   - **(b)** What is the size of the output of $\hat{G}$ given the size of its input?
   - **(c)** Why is it not a pseudo-random generator from $\{0,1\}^\ell$ to $\{0,1\}^{2\ell}$ for $\ell = \lceil \lg q \rceil$?

Exercise 2. *Let us go post-quantum!*

Definition 3 (Learning with Errors). *Let $\ell < k \in \mathbb{N}$, $n < m \in \mathbb{N}$, $q = 2^k$, $B = 2^\ell$, $\mathbf{A} \hookleftarrow U(\mathbb{Z}_q^{m \times n})$. The Learning with Errors (*LWE*) distribution is defined as follows: $D_{\mathrm{LWE},\mathbf{A}} = (\mathbf{A}, \mathbf{A} \cdot \mathbf{s} + \mathbf{e} \bmod q)$ for $\mathbf{s} \hookleftarrow U(\mathbb{Z}_q^n)$ and $\mathbf{e} \hookleftarrow U\left(\left[-\frac{B}{2}, \frac{B}{2} - 1\right]^m \cap \mathbb{Z}^m\right)$.*

Note. In this setting, the vector $\mathbf{s}$ is called the secret, and $\mathbf{e}$ the noise.
The *LWE assumption* states that, given suitable parameters k, ℓ, m, n, it is computationally hard to distinguish $D_{\mathrm{LWE},\mathbf{A}}$ from the distribution $(\mathbf{A}, U(\mathbb{Z}_q^m))$.
Let us propose the following generator: $G_{\mathbf{A}}(\mathbf{s}, \mathbf{e}) = \mathbf{A} \cdot \mathbf{s} + \mathbf{e} \bmod q$.

1. Given the binary representation of $\mathbf{s}$, $\mathbf{e}$, compute the bitsize of the input and the output of the function G with respect to k, ℓ, m, n.
2. Evaluate the cost of a bruteforce attack to retrieve the input $\mathbf{s}, \mathbf{e}$ in terms of arithmetic operations in $\mathbb{Z}_q$.
3. What happens if $B = 0$? ☞ *This bound can prove useful:* $\prod_{i=1}^n (1 - 2^{-i}) > 0.288$.
4. Given the previous question, refine the bruteforce attack of question 2. What does it mean for the security of the generator G?
5. What happens if $\ell = k$?
6. Given suitable ℓ, k, n, m such that the LWE problem holds in this setting, show that $G_{\mathbf{A}}$ is a pseudo-random generator.

Exercise 3. *One-time pad is semantically secure.*
Let us recall the one-time pad scheme to encrypt a message $m \in \{0,1\}^\ell$ for $\ell \in \mathbb{N}$.

$\textsf{Keygen}(1^\ell)$**:** Outputs $k \leftarrow U(\{0,1\}^\ell)$

$\textsf{Enc}_k(m)$**:** Outputs $c = m \oplus k$

$\textsf{Dec}_k(c)$**:** Outputs $m' = c \oplus k$

1. Recall the definition of semantic security for a symmetric encryption scheme (for one-time key and chosen plaintext attack).

2. Prove that one-time pad is semantically secure.

Exercise 4. *Sub-bits of a Generator.*
Let $G : \{0,1\}^s \to \{0,1\}^n$ be a pseudo-random generator, $S \subseteq [1,n] \cap \mathbb{Z}$ of size ℓ. Let us define the function $G' : \{0,1\}^s \to \{0,1\}^\ell$ as $x \to G(x)_{|S} = \|_{i \in S} G(x)_i$, where $\|$ denotes the concatenation.

1. Given that G is secure, prove that the distribution defined by the output of G' on $x \leftarrow U(\{0,1\}^s)$ is indistinguishable from the uniform distribution over $\{0,1\}^\ell$.

Exercise 5. *Increasing the expansion factor of a PRG.*
We recall that the advantage $\mathrm{Adv}^{PRG}_{\mathcal{A}}[G]$ of an algorithm $\mathcal{A}$ against a PRG (pseudo-random generator) $G : \{0,1\}^n \to \{0,1\}^m$ is the difference of the probabilities that $\mathcal{A}$ returns 1 when it is given $G(x) \in \{0,1\}^m$ for x uniformly sampled in $\{0,1\}^n$, and when it is given u uniformly sampled in $\{0,1\}^m$. We say that G is a secure PRG if, for any probabilistic polynomial-time $\mathcal{A}$, the advantage of $\mathcal{A}$ is negligible in n, i.e., $\mathrm{Adv}^{PRG}_{\mathcal{A}}[G] \leq n^{-\omega(1)}$.

We assume that we have a pseudo-random generator $G : \{0,1\}^n \to \{0,1\}^{n+1}$.

1. Consider $G' : \{0,1\}^n \to \{0,1\}^{n+2}$ defined as follows. On input $x \in \{0,1\}^n$, G' first evaluates $G(x)$ and obtains $(x', y') \in \{0,1\}^n \times \{0,1\}$ such that $G(x) = x' \| y'$. It then evaluates G on x' and eventually returns $G(x') \| y'$. Show that if G is a secure PRG, then so is G'.

An arbitrary-length PRG is a function G taking as inputs $x \in \{0,1\}^n$ and $\ell \geq 1$ in unary, and returning an element of $\{0,1\}^\ell$. It is said to be secure if for all ℓ polynomially bounded with respect to n, the distributions $G(U(\{0,1\}^n), 1^\ell)$ and $U(\{0,1\})^\ell$ are computationally indistinguishable.

2. Let $n \geq 1$. Propose a construction of an arbitrary-length PRG G^* based on G. Show that if G is a secure PRG, then so is G^*.

Exercise 6. *Increasing the advantage of an attacker.*
Let G be a pseudo-random generator from $\{0,1\}^s$ to $\{0,1\}^n$ for some integers s and n. Let $i \in \{1, \cdots, n\}$ and let $\mathcal{A}$ be a PPT algorithm such that, for all $k \in \{0,1\}^s$, we have

$$\Pr[\mathcal{A}(G(k)_{1\cdots i-1}) = G(k)_i] \geq \frac{1}{2} + \varepsilon,$$

where the probability runs over the randomness of $\mathcal{A}$. Note that unlike the definition of the advantage seen in class, here we consider only the probability over the randomness of $\mathcal{A}$ and not over the random choice of k (we will see why later).
Our objective is to construct a new attacker $\mathcal{A}'$ with an advantage arbitrarily close to 1 (for instance $\Pr[\mathcal{A}(G(k)_{1\cdots i-1}) = G(k)_i] \geq 0.999$ for all $k \in \{0,1\}^s$).

1. Propose a method to improve the success probability of $\mathcal{A}$.

   Let m be some integer to be determined. Let $\mathcal{A}'$ be an algorithm that evaluates $\mathcal{A}$ on $G(k)_{1\cdots i-1}$ $2m+1$ times, to obtain $2m+1$ bits $b_1, \cdots, b_{2m+1}$ and then outputs the bit that appeared the most (i.e. at least $m+1$ times).

2. Give a lower bound on $\Pr[\mathcal{A}'(G(k)_{1\cdots i-1}) = G(k)_i]$, for all $k \in \{0,1\}^s$. We recall Hoeffding's inequality for Bernoulli variables: let $X_1, \cdots, X_{2m+1}$ be independent Bernoulli random variables, with $\Pr(X_i = 1) = 1 - \Pr(X_i = 0) = p$ for all i, and let $S = X_1 + \cdots + X_{2m+1}$. Then, for all $x > 0$, we have
$$\Pr[|S - \mathbb{E}(S)| \geq x\sqrt{2m+1}] \leq 2e^{-2x^2}.$$

3. What should be the value of m (depending on ε) if we want that $\Pr[\mathcal{A}'(G(k)_{1\cdots i-1}) = G(k)_i] \geq 0.999$ for all k? It may be useful to know that $e^{-8} \leq 0.0005$.

4. Do we have $\mathrm{Adv}_{\mathrm{unpredictability}}(\mathcal{A}') \geq 0.999$ if $\Pr[\mathcal{A}'(G(k)_{1\cdots i-1}) = G(k)_i] \geq 0.999$ for all k?

5. What condition on ε do we need to ensure that $\mathcal{A}'$ runs in polynomial time?

   Let now $\mathcal{A}$ be an attacker such that
$$\mathrm{Adv}(\mathcal{A}) = \Pr_{k \leftarrow U(\{0,1\}^s)}[\mathcal{A}(G(k)_{1\cdots i-1}) = G(k)_i] \geq \frac{1}{2} + \varepsilon.$$

   Note that we are now looking at the definition of advantage given in class, where the probability also depends on the uniform choice of k. We want to show that in this case, we cannot always amplify the success probability of the attacker by repeating the computation.

   In the following, we write $\Pr[\mathcal{A}(G(k)_{1\cdots i-1}) = G(k)_i]$ when we only consider the probability over the internal randomness of $\mathcal{A}$ (and k is fixed) and $\Pr_{k \leftarrow U(\{0,1\}^s)}[\mathcal{A}(G(k)_{1\cdots i-1}) = G(k)_i]$ when we consider the probability over the choice of k and the internal randomness of $\mathcal{A}$.

   Suppose that $s \geq 2$ and define
$$G(k) = \begin{cases} 00\cdots 0 & \text{if } k_0 = k_1 = 0 \\ G_0(k) & \text{otherwise,} \end{cases}$$

   where G_0 is a secure PRG from $\{0,1\}^s$ to $\{0,1\}^n$.

6. Show that there exists a PPT attacker $\mathcal{A}$ with non negligible advantage (for the unpredictability definition) against G.

7. Show on the contrary that there is no PPT attacker $\mathcal{A}$ with $\mathrm{Adv}(\mathcal{A}) \geq 7/8$ (assuming that G_0 is a secure PRG).