TD 4: LWE and PRFs (corrected version)

Exercise 1. *PRG from LWE*

We recall the Learning with Errors assumption.

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

In this setting, the vector $\mathbf{s}$ is called the secret, and $\mathbf{e}$ the noise.
Remark. If q and B are powers of 2, we are manipulating bits, contrary to the DDH-based PRG from the lecture.
The *LWE assumption* states that, given suitable parameters q, B, m, n, it is computationally hard to distinguish D_{LWE} from the distribution $U(\mathbb{Z}_q^{m\times n} \times \mathbb{Z}_q^m)$.
Let us propose the following pseudo-random generator: $G(\mathbf{A}, \mathbf{s}, \mathbf{e}) = (\mathbf{A}, \mathbf{A}\cdot\mathbf{s} + \mathbf{e} \bmod q)$.

1. By definition, a PRG must have a bigger output size than input size. Give a bound on B that depends on the other parameters if we want G to satisfy this.

   ☞ We want the parameters to satisfy $q^{mn} \cdot q^n B^m \le q^{nm} \cdot q^m$ i.e. $B^m \le q^{m-n}$. Then the bound is $B \le q^{1-n/m}$.

2. Given suitable B, q, n, m such that the LWE assumption and previous bound hold, show that G is a secure pseudo-random generator.

   ☞ Let $\mathcal{A}$ be a PPT adversary that distinguishes with non negligible advantage the output of G from the uniform distribution. Let us use this adversary to solve the LWE problem.

   At the beginning of the game, the reduction $\mathcal{B}$ receives a LWE instance $(\mathbf{A}, \mathbf{b}) \in \mathbb{Z}_q^{m\times n} \times \mathbb{Z}_q^m$ of the LWE problem, the goal is to output `LWE` if it is a LWE instance, and `Unif` if it is uniform.

   The reduction sends $(\mathbf{A}, \mathbf{b})$ to the adversary $\mathcal{A}$ against the PRG. The adversary then returns a bit b' that the reduction returns to its challenger.

   Analysis. $\mathrm{Adv}^{\mathrm{LWE}}(\mathcal{B}) = |\Pr[\mathcal{B} \to 1 | b\ \mathrm{LWE}] - \Pr[\mathcal{B} \to 1 | b\ \mathrm{Unif}]| = |\Pr[\mathcal{A} \to 1 | b\ \mathrm{LWE}] - \Pr[\mathcal{A} \to 1 \mid b\ \mathrm{Unif}]| = \mathrm{Adv}^{\mathrm{PRG}}(\mathcal{A}) =$ non negl.

Exercise 2. *LWE with small secret*

We once more work in the setting of the LWE assumption. Let q, B, n, m such that the LWE assumption holds. Moreover, we assume that q is prime.

1. **(a)** What is the probability that $\mathbf{A}_1 \in \mathbf{Z}_q^{n\times n}$ is invertible where $\mathbf{A} =: [\mathbf{A}_1^\top | \mathbf{A}_2^\top]^\top$ is uniformly sampled?

   ☞ We have to compute $|GL_n(\mathbb{F}_q)|$, i.e. the number of invertibles matrices with coefficients in $\mathbb{F}_q$. We have $q^n - 1$ choice for the first vector (it can be any vector except the 0 vector), then $q^n - q^1$ for the second vector (anything except a vector collinear to the first one), then $q^n - q^2$ (anything that is not a linear combination of the first two vectors), etc. So we get

$$\Pr_{\mathbf{A}_1 \hookleftarrow U(\mathbb{F}_2^{m\times n})}[A_1 \in GL_n(\mathbb{F}_q)] = \frac{1}{q^{n^2}} \prod_{i=0}^{n-1} (q^n - q^i)$$
$$= \prod_{i=0}^{n-1} (1 - q^{i-n}),$$

   which is always $\ge \prod_{i=0}^{n-1}(1 - 2^{i-n}) \ge 0.288$.

   (b) Assume that $m \ge 2n$. Prove that there exists a subset of n lineraly independent rows of $\mathbf{A} \hookleftarrow U(\mathbb{Z}_q^{m\times n})$ with probability $\ge 1 - 1/2^{\Omega(n)}$ and that we can find them in polynomial time.

   ☞ If this is not the case, then there exists an hyperplane of $\mathbb{Z}_q^n$ in which each row is sampled. A hyperplane is given by a nonzero vector: there are at most $q^n - 1$ hyperplanes of the space and for a given hyperplane, the probability that each vector falls into it is $q^{(n-1)m}/q^{nm} = 1/q^m$. Then the union bound gives us that the probability is $\ge 1 - \frac{1}{q^{m-n}} \ge 1 - \frac{1}{q^n}$.

   To find such rows, the naive greedy algorithm works: select the first row. Then, repeat the following for $i = 2$ to m. If the i-th row is linearly independent from the selected rows, select it.

2. Let us define the distribution $D_B = U\left((-B, B] \cap \mathbb{Z}\right)$, and $m' = m - n$.

Show that under the $\mathrm{LWE}_{q,m,n,B}$ assumption, the distributions $(\mathbf{A}', \mathbf{A}'\mathbf{s}' + \mathbf{e}') \in \mathbb{Z}_q^{m' \times n} \times \mathbb{Z}_q^{m'}$, with $\mathbf{s}' \hookleftarrow D_B^n$ and $\mathbf{e}' \hookleftarrow D_B^{m'}$, and $(\mathbf{A}', \mathbf{b}')$ with $\mathbf{b}' \leftarrow U(\mathbb{Z}_q^{m'})$ are indistinguishable.

☞ We show how to reduce an instance of the decision problem $\mathrm{LWE}_{q,m,n,B}$ to an instance of this new decision problem. Let $(\mathbf{A}, \mathbf{b}) \in \mathbf{Z}_q^{m \times n} \times \mathbf{Z}_q^m$. With non negligible probability and up to permuting the rows of $\mathbf{A}$ (and $\mathbf{b}$), one can write $\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix}$, where $\mathbf{A}_1 \in \mathbf{Z}_q^{n \times n}$ is invertible.

Notice that in this case, $\mathbf{A}_2\mathbf{A}_1^{-1} \in \mathbb{Z}_q^{m' \times n}$ is still uniform because $\mathbf{A}_1$ is invertible, and $\mathbf{A}_2$ is uniformly sampled.

Assume that we are given a sample $(\mathbf{A}, \mathbf{A}\mathbf{s} + \mathbf{e})$ of the $\mathrm{LWE}_{q,m,n,B}$ distribution. Set $\mathbf{e} =: (\mathbf{e}_1^\top, -\mathbf{e}_2^\top)^\top$ Consider the following:

$$(\mathbf{A}_2\mathbf{A}_1^{-1}, \mathbf{A}_2\mathbf{A}_1^{-1}(\mathbf{A}_1\mathbf{s} + \mathbf{e}_1) - \mathbf{A}_2\mathbf{s} + \mathbf{e}_2) = (\mathbf{A}_2\mathbf{A}_1, \mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{e}_1 + \mathbf{e}_2).$$

This is exactly a sample from the new distribution, with secret $\mathbf{e}_1$ and noise $\mathbf{e}_2$.

Assume now that we are given a sample $(\mathbf{A}, \mathbf{b})$ where $\mathbf{b}$ is uniformly sampled. We write $\mathbf{b} =: (\mathbf{b}_1^\top, \mathbf{b}_2^\top)^\top$. With the previous transformation we get: $\mathbf{A}_2\mathbf{A}_1^{-1}, \mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{b}_1 - \mathbf{b}_2$. Whatever $\mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{b}_1$ is, since it is independent from $\mathbf{b}_2$, we get a uniform sample over $\mathbb{Z}_q^{m' \times n} \times \mathbb{Z}_q^{m'}$.

This means that any distinguisher for the new decision problem is a distinguisher for decision LWE. Under the LWE assumption, any efficient distinguisher has negligible advantage and this concludes the proof.

Exercise 3. *CTR Security*

Let $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$ be a PRF. To encrypt a message $M \in \{0,1\}^{d \cdot n}$, CTR proceeds as follows:

- Write $M = M_0 \| M_1 \| \ldots \| M_{d-1}$ with each $M_i \in \{0,1\}^n$.
- Sample IV uniformly in $\{0,1\}^n$.
- Return $IV \| C_0 \| C_1 \| \ldots \| C_{d-1}$ with $C_i = M_i \oplus F(k, IV + i \bmod 2^n)$ for all i.

The goal of this exercise is to prove the security of the CTR encryption mode against chosen plaintext attacks, when the PRF F is secure.

1. Recall the definition of security of an encryption scheme against chosen plaintext attacks.

☞ Let (KeyGen, Enc, Dec) be an encryption scheme. We consider the following experiments Exp_b for $b \in \{0,1\}$:

- Challenger samples $k \leftarrow$ KeyGen,
- Adversary makes q encryption queries on messages $(M_{i,0}, M_{i,1})$,
- Challenger sends back $Enc(k, M_{i,b})$ for each i,
- Adversary returns $b' \in \{0,1\}$.

We define the advantage of the adversary $\mathcal{A}$ against the encryption scheme as

$$\mathsf{Adv}^{\mathsf{CPA}}(\mathcal{A}) = \left|\Pr(\mathcal{A} \xrightarrow{\mathsf{Exp}_1} 1) - \Pr(\mathcal{A} \xrightarrow{\mathsf{Exp}_0} 1)\right|.$$

Then, the encryption scheme is said to be secure against chosen plaintext attacks if no probabilistic polynomial-time adversary has a non-negligible advantage with respect to n.

(Note in particular that since $\mathcal{A}$ runs in polynomial time, q must be polynomial in n.)

Remark: in another equivalent definition, there is only one experiment in which the challenger starts by choosing the bit b uniformly at random, and the advantage is defined as $\mathsf{Adv}^{\mathsf{CPA}}(\mathcal{A}) = |\Pr(\mathcal{A} \to 1 \mid b = 0) - \Pr(\mathcal{A} \to 1 \mid b = 1)|$.

2. Assume an attacker makes Q encryption queries. Let $IV_1, \ldots, IV_Q$ be the corresponding IV's. Let `Twice` denote the event "there exist $i, j \le Q$ and $k_i, k_j < d$ such that $IV_i + k_i = IV_j + k_j \bmod 2^n$ and $i \neq j$." Show that the probability of `Twice` is bounded from above by $Q^2 d / 2^{n-1}$.

☞ *Remark: the probability of* `Twice` *is obviously 1 if it is not required that i and j be distinct. Besides, considering the case $i = j$ is not interesting for our purpose.*

For $i, j \le Q$, let $\mathtt{Twice}_{i,j}$ be the event "$\exists k_i, k_j < d : \mathrm{IV}_i + k_i = \mathrm{IV}_j + k_j \pmod{2^n}$", which is equivalent to "$\exists k, |k| < d$ and $\mathrm{IV}_i - \mathrm{IV}_j = k \pmod{2^n}$). As the IVs are chosen uniformly and independently, $\mathrm{IV}_i - \mathrm{IV}_j$ is uniform modulo 2^n and $\Pr(\mathtt{Twice}_{i,j}) \le 2^{-n}(2d-1)$. (The inequality is strict when $2d - 1 > 2^n$, in which case $\Pr(\mathtt{Twice}_{i,j}) = 1$.) Then,

$$\Pr(\mathtt{Twice}) \le \sum_{1 \le i \neq j \le Q} \Pr(\mathtt{Twice}_{i,j}) = Q(Q-1)2^{-n}(2d-1) \le 2^{1-n}Q^2 d.$$

3. Assume the PRF F is replaced by a uniformly chosen function $f : \{0,1\}^n \to \{0,1\}^n$. Give an upper bound on the distinguishing advantage of an adversary $\mathcal{A}$ against this idealized version of CTR, as a function of d, n and the number of encryption queries Q.

☞ We write $M^{i,\beta} = M_0^{i,\beta} \| \ldots \| M_{d-1}^{i,\beta}$ with $1 \le i \le Q$ and $\beta \in \{0,1\}$ the encryption queries of the adversary $\mathcal{A}$ and $C^i = \mathrm{IV}_i \| C_0^i \| \ldots \| C_{d-1}^i$ with $1 \le i \le Q$ the replies. Given the value of $b \in \{0,1\}$ chosen by the challenger, we know that $C_j^i = M_j^{i,b} \oplus f(\mathrm{IV}_i + j \pmod{2^n})$ for all $1 \le i \le Q$ and $0 \le j < d$.

If $\texttt{Twice}$ does not occur, then all the $\mathrm{IV}_i + j \pmod{2^n}$ for $1 \le i \le Q$ and $0 \le j < d$ are pairwise distinct. Then the values of f at these points are independent and uniformly distributed, since $f : \{0,1\}^n \to \{0,1\}^n$ is chosen uniformly at random. Therefore, all the C_j^i are also independent and uniformly distributed regardless of the value of b, so that $\Pr(\neg\texttt{Twice} \wedge \mathcal{A} \to 1 \mid b = 0) = \Pr(\neg\texttt{Twice} \wedge \mathcal{A} \to 1 \mid b = 1)$. It follows that

$$\begin{aligned}
\mathsf{Adv}_{\mathcal{U}}^{\mathsf{CPA}}(\mathcal{A}) &= |\Pr(\texttt{Twice} \wedge \mathcal{A} \to 1 \mid b = 0) - \Pr(\texttt{Twice} \wedge \mathcal{A} \to 1 \mid b = 1)| \\
&= |\Pr(\mathcal{A} \to 1 \mid b = 0, \texttt{Twice}) - \Pr(\mathcal{A} \to 1 \mid b = 1, \texttt{Twice})| \Pr(\texttt{Twice}) \\
&\le \Pr(\texttt{Twice}) \le 2^{1-n} Q^2 d.
\end{aligned}$$

4. Show that if there exists a probabilistic polynomial-time adversary $\mathcal{A}$ against CTR based on PRF F, then there exists a probabilistic polynomial-time adversary $\mathcal{B}$ against the PRF F. Give a lower bound on the advantage degradation of the reduction.

☞ Assume that $\mathcal{A}$ is a PPT adversary against the encryption scheme with a non-negligible advantage for a chosen plaintext attack. We build an adversary $\mathcal{B}$ against the underlying PRF F as follows:

1. Choose $b \in \{0,1\}$ uniformly at random.
2. For each encryption query (M^0, M^1) from $\mathcal{A}$, encrypt M^b using the given scheme, that is,
   (a) Choose $\mathrm{IV} \in \{0,1\}^n$ uniformly at random.
   (b) For $j = 0$ to $d-1$, send a query for $\mathrm{IV} + j$ and with the reply f_j compute $C_j = M_j^b \oplus f_j$.
   (c) Send $\mathrm{IV} \| C_0 \| \ldots \| C_{d-1}$ back to $\mathcal{A}$.
3. When $\mathcal{A}$ finally outputs a bit $b' \in \{0,1\}$, output 1 if $b' = b$ and 0 otherwise.

The advantage of $\mathcal{B}$ against the PRF F is

$$\mathsf{Adv}_F^{\mathsf{PRF}}(\mathcal{B}) = |\Pr(\mathcal{B} \to 1 \mid \mathsf{PRF}) - \Pr(\mathcal{B} \to 1 \mid \mathsf{Unif})|$$

where PRF is the experiment in which replies to $\mathcal{B}$ are computed by calling F and Unif is the one in which replies to $\mathcal{B}$ are computed from a uniformly chosen random function f.

Considering the two terms separately gives

$$\begin{aligned}
\Pr(\mathcal{B} \to 1 \mid E) &= \frac{1}{2}\left(\Pr(b' = 0 \mid E, b = 0) + \Pr(b' = 1 \mid E, b = 1)\right) \\
&= \frac{1}{2}\left(1 + \Pr(\mathcal{A} \to 1 \mid E, b = 1) - \Pr(\mathcal{A} \to 0 \mid E, b = 0)\right)
\end{aligned}$$

where E is either PRF or Unif. Therefore

$$\mathsf{Adv}_F^{\mathsf{PRF}}(\mathcal{B}) \ge \frac{1}{2}\left(\mathsf{Adv}^{\mathsf{CPA}}(\mathcal{A}) - \mathsf{Adv}_{\mathcal{U}}^{\mathsf{CPA}}(\mathcal{A})\right) \ge \frac{1}{2}\mathsf{Adv}^{\mathsf{CPA}}(\mathcal{A}) - 2^{1-n}Q^2 d$$

using the previous question. Thus, if $\mathsf{Adv}^{\mathsf{CPA}}(\mathcal{A})$ is non-negligible then so is $\mathsf{Adv}_F^{\mathsf{PRF}}(\mathcal{B})$, which is then about a half of $\mathsf{Adv}^{\mathsf{CPA}}(\mathcal{A})$.