1 Statistical distance of $\mathcal{D}'_0$ and $\mathcal{D}'_1$

Here are some facts about statistical distance:

Fact 1.1. For any three distributions $\mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2$,

$$\mathsf{SD}(\mathcal{D}_0, \mathcal{D}_2) \le \mathsf{SD}(\mathcal{D}_0, \mathcal{D}_1) + \mathsf{SD}(\mathcal{D}_1, \mathcal{D}_2).$$

Fact 1.2. For any two distributions $\mathcal{D}_0, \mathcal{D}_1$, $\mathsf{SD}(\mathcal{D}_0, \mathcal{D}_1) = \epsilon$ if and only if there exists an adversary $\mathcal{A}$ (not necessarily polynomial time) that can win the following game with probability $1/2 + \epsilon/2$:

- Challenger picks a bit $b \leftarrow \{0,1\}$, samples $u \leftarrow \mathcal{D}_b$ and sends u to $\mathcal{A}$.
- $\mathcal{A}$ sends its guess b' and wins if $b = b'$.

We will use the above facts to prove the following theorem.

Theorem 1.3. Suppose $\mathsf{SD}(\mathcal{D}_0, \mathcal{D}_1) \le \epsilon$. Then $\mathsf{SD}(\mathcal{D}'_0, \mathcal{D}'_1) \le t \cdot \epsilon$.

Proof. The proof follows via a sequence of hybrid distributions $\mathcal{H}_0, \ldots, \mathcal{H}_t$, where $\mathcal{H}_0 \equiv \mathcal{D}'_0$, $\mathcal{H}_t \equiv \mathcal{D}'_1$, and $\mathcal{H}_i$ is defined as follows:

$\mathcal{H}_i$ = {output (t-i) samples chosen independently from $\mathcal{D}_0$ and output i samples chosen independently from $\mathcal{D}_1$ }

Claim 1.4. For any $i \le t$, $\mathsf{SD}(\mathcal{H}_{i-1}, \mathcal{H}_i) \le \epsilon$.

Proof. Let Ω be the Distribution of all sized sampling from $\mathcal{D}_0$ and $\mathcal{D}_1$.
Ω = { output (t-i) samples from $\mathcal{D}_0$ and i samples from $\mathcal{D}_1$ chosen independently from $\mathcal{D}_0$, $\mathcal{D}_1$ respectively where t $\in$ N and i $\in$ [0,t] }

Let define events on distribution Ω :

E_i^0 = { output i samples chosen independently from $\mathcal{D}_0$ }.

E_i^1 = { output i samples chosen independently from $\mathcal{D}_1$ }.

$\mathsf{SD}(\mathcal{H}_{i-1}, \mathcal{H}_i) = \frac{1}{2} * (\sum_{j\in\Omega} |\Pr_{z\leftarrow\mathcal{H}_{i-1}}[z = j] - \Pr_{z\leftarrow\mathcal{H}_i}[z = j]|) = \frac{1}{2} *(\sum_{\substack{E^0_{t-i+1}, E^1_{i-1}\in\Omega \\ E^0_{t-i}, E^1_i \in\Omega}} |\Pr_{z\leftarrow\mathcal{H}_{i-1}}[z = \{E^0_{t-i+1} \bigwedge E^1_{i-1}\}] - \Pr_{z\leftarrow\mathcal{H}_i}[z = \{E^0_{t-i} \bigwedge E^1_i\}]|) = \frac{1}{2} *(\sum_{\substack{E^0_{t-i+1}, E^1_{i-1}\in\Omega \\ E^0_{t-i}, E^1_i \in\Omega}} |Pr[\{E^0_{t-i+1} \bigwedge E^1_{i-1}\}] - Pr[\{E^0_{t-i} \bigwedge E^1_i\}]|)$.

$$\mathsf{SD}(\mathcal{H}_{i-1}, \mathcal{H}_i) = \frac{1}{2} * (\sum_{\substack{E^0_{t-i+1}, E^1_{i-1}\in\Omega \\ E^0_{t-i}, E^1_i \in\Omega}} |Pr[\{E^0_{t-i+1} \bigwedge E^1_{i-1}\}] - Pr[\{E^0_{t-i} \bigwedge E^1_i\}]|) \qquad -(1).$$

Now let use the fact that sampling from $\mathcal{D}_0$ and $\mathcal{D}_1$ are independent. $\Pr[E_i^0 \bigwedge E_j^1]=\Pr[E_i^0]*\Pr[E_j^1]$ ...(2)

As chosen element are independent during sampling than we can use probability of independent event.

$\Pr[E_1^0 | E_i^0]= \Pr[E_1^0]$ $\qquad$ – (3)

$\Pr[E^0_{i+1}] = \Pr[E_1^0 \bigwedge E_i^0]=\Pr[E_1^0 | E_i^0]*\Pr[E_i^0]=\Pr[E_1^0]*\Pr[E_i^0]$ (using 3rd Equation) –(4) .

Using 2nd and 4th Equation we can simplify the 1st Equation.

$\mathsf{SD}(\mathcal{H}_{i-1}, \mathcal{H}_i) = \frac{1}{2} * (\sum_{\substack{E^0_{t-i+1}, E^1_{i-1} \in \Omega \\ E^0_{t-i}, E^1_i \in \Omega}} |Pr[E^0_{t-i+1}] * Pr[E^1_{i-1}] - Pr[E^0_{t-i}] * [E^1_i]| \ ) = \frac{1}{2} *(\sum_{\substack{E^0_1, E^1_1 \in \Omega \\ E^0_{t-i}, E^1_{i-1} \in \Omega}} |Pr[E^0_{t-i}] * Pr[E^0_1] * Pr[E^1_{i-1}] - Pr[E^0_{t-i}] * [E^1_{i-1}] * Pr[E^1_1]| \ ) = \frac{1}{2} *(\sum_{\substack{E^0_1, E^1_1 \in \Omega \\ E^0_{t-i}, E^1_{i-1} \in \Omega}} |Pr[E^0_{t-i}] * Pr[E^1_{i-1}](Pr[E^0_1] - Pr[E^1_1])| \ )$
$\leq \frac{1}{2} *(\sum_{E^0_1, E^1_1 \in \Omega} |1 * 1(Pr[E^0_1] - Pr[E^1_1])| \ ) = \frac{1}{2} *(\sum_{E^0_1, E^1_1 \in \Omega} |Pr[E^0_1] - Pr[E^1_1]| \ )$. –(5)

Let re-visit the definition of E^0_i and E^1_i.

For i = 1 :

E^0_1 = { output 1 sample chosen independently from $\mathcal{D}_0$ } and E^1_1 = { output 1 sample chosen independently from $\mathcal{D}_1$ }

E^0_1 is same as distribution $\mathcal{D}_0$ and E^1_1 is same as distribution $\mathcal{D}_1$.

Hence we can re-write 5th Equation in term of $\mathcal{D}_0$ and $\mathcal{D}_1$ distribution.

$\mathsf{SD}(\mathcal{H}_{i-1}, \mathcal{H}_i) \leq \frac{1}{2} *(\sum_{E^0_1, E^1_1 \in \Omega} |Pr[E^0_1] - Pr[E^1_1]| \ ) = \frac{1}{2} *(\sum_{E^0_1, E^1_1 \in \Omega} | \Pr_{z \leftarrow \mathcal{D}_0}[z = E^0_1] - \Pr_{z \leftarrow \mathcal{D}_1}[z = E^1_1]| \ )$
$\leq \frac{1}{2} *(\sum_{j \in \Omega} | \Pr_{z \leftarrow \mathcal{D}_0}[z = j] - \Pr_{z \leftarrow \mathcal{D}_1}[z = j]| \ ) = \mathsf{SD}(\mathcal{D}_0, \mathcal{D}_1) \leq \epsilon.$ – –(6)

Hence from 6th Equation we have got that

$\mathsf{SD}(\mathcal{H}_{i-1}, \mathcal{H}_i) \leq \epsilon$ for i $\in$ [1,t]. –(7)

□

[**TODO**: conclude proof of theorem using the above claim]

□

2 Weak PRPs

Theorem 2.1. Assuming F is a secure PRF, the construction described in the assignment is a weak PRP.

Proof. We will prove that the construction is a weak PRP via a sequence of hybrid worlds. We first present the hybrid worlds below, then show that they are indistinguishable.

World 0: In this world, the challenger uses two PRF keys k_1, k_2. For every query, the challenger picks (x_i, y_i) uniformly at random, and sends the output of the PRP construction.

- The challenger chooses two uniformly random PRF keys k_1, k_2.
- For the i^{th} query, the challenger chooses uniformly randomly x_i, y_i and then computes $v_i = y_i \oplus F(x_i, k_1)$. It then sends (x_i, y_i) together with $(v_i, x_i \oplus F(v_i, k_2))$.
- Adversary sends b'.

Hybrid 1:

- The challenger chooses $\underline{\text{a uniformly random function } f \leftarrow \mathsf{Func}[\mathcal{X}, \mathcal{X}]}$ and PRF key k_2.
- For the i^{th} query, the challenger chooses uniformly random x_i, y_i and then computes $\underline{v_i = y_i \oplus f(x_i)}$. It then sends (x_i, y_i) together with $(v_i, x_i \oplus F(v_i, k_2))$.
- Adversary sends b'.

Hybrid 2:

- The challenger chooses $\underline{\text{two uniformly random functions } f_1, f_2.}$
- For the i^{th} query, the challenger chooses uniformly randomly x_i, y_i and then computes $v_i = y_i \oplus f_1(x_i)$. It then sends (x_i, y_i) together with $\underline{(v_i, x_i \oplus f_2(v_i))}$.
- Adversary sends b'.

World 1:

- The challenger chooses $\underline{\text{uniformly random permutation } P \leftarrow \mathsf{Perm}[\mathcal{X}^2].}$
- For the i^{th} query, the challenger chooses uniformly randomly x_i, y_i and then sends (x_i, y_i) together with $\underline{P(x_i, y_i)}$.
- Adversary sends b'.

We will now prove that the hybrids are indistinguishable.

Claim 2.2. Suppose there exists a p.p.t adversary $\mathcal{A}$ such that $p_0 - p_{\text{hyb},1} = \epsilon$. Then there exists a **p.p.t** reduction algorithm $\mathcal{B}$ that breaks the PRF security of F with probability $1/2 + \epsilon/2$.

Proof. [**TODO**: Describe **p.p.t** reduction algorithm. You can skip the success probability analysis.] □

Claim 2.3. Suppose there exists a p.p.t adversary $\mathcal{A}$ such that $p_{\text{hyb},1} - p_{\text{hyb},2} = \epsilon$. Then there exists a **p.p.t** reduction algorithm $\mathcal{B}$ that breaks the PRF security of F with probability $1/2 + \epsilon/2$.

Proof. This proof is very similar to the proof of previous claim. ☐

Claim 2.4. Hybrid 2 is indistinguishable from world 1.

Proof. [**TODO**: Fill in the proof of this claim.] ☐

☐

3 Composing PRGs and PRFs

3.1

Theorem 3.1. Assuming F is a secure PRF and G is a secure PRG, F' is a secure PRF.

Proof. We will prove this theorem via a sequence of hybrid experiments, where world-0 (= hybrid-0) corresponds to the challenger choosing a PRF key, and world-1 (= final hybrid) corresponds to the challenger choosing a uniformly random function. Let t denote the total number of PRF queries made by the adversary $\mathcal{A}$.

Description of hybrids: [**TODO**: Define the hybrid worlds you will use for the proof in the next question.]

Next, we show that the consecutive hybrids are computationally indistinguishable.

Analysis: Let $p_{\text{hyb},i}$ denote the probability of $\mathcal{A}$ outputting 0 in Hybrid-i.

[**TODO**: Show that if $p_{\text{hyb},i}$ and $p_{\text{hyb},i+1}$ are far-apart, there exists a p.p.t. reduction algorithm that breaks the security of ....] $\square$

3.2

3.2.1 Construction of $\mathcal{G}'$

[**TODO**: describe construction of $\mathcal{G}'$. Hint: what happens if $\mathcal{G}'$ is not injective?]

Claim 3.2. Suppose there exists a p.p.t adversary $\mathcal{A}$ that breaks the security of the PRG $\mathcal{G}'$ then there exists a p.p.t adversary $\mathcal{B}$ that breaks the PRG security of $\mathcal{G}$.

Proof. [**TODO**: Show a reduction, followed by an analysis of the reduction algorithm's success probability.] $\square$

Claim 3.3. F' is not a secure PRF.

Proof. [**TODO**: Show a p.p.t. adversary $\mathcal{A}$ that breaks PRF security.] $\square$

4 CBC mode

Theorem 4.1. Assuming F is a secure PRP, and $|\mathcal{X}|$ is super-polynomial in the security parameter, the CBC mode of encryption satisfies No-Query-Semantic-Security.

Proof. As discussed in class (Lecture 11, Section 2), this proof goes through a sequence of hybrids.

World 0:

- Adversary $\mathcal{A}$ sends two messages m_0,m_1 s.t $|m_0| = |m_1| = n \cdot \ell$. Let $m_b = (m_{b,1} \mid\mid \ldots \mid\mid m_{b,\ell})$.
- Challenger chooses PRP key $k \leftarrow \mathcal{K}$. It computes $\mathsf{ct}_1 = F(m_{0,1}, k)$. For all $i > 1$, it computes $\mathsf{ct}_i = F(m_{0,i} \oplus \mathsf{ct}_{i-1}, k)$.

  Finally, it sends $(\mathsf{ct}_1, \ldots, \mathsf{ct}_\ell)$ to $\mathcal{A}$.
- Adversary sends b'

Hybrid 1:

- Adversary $\mathcal{A}$ sends two messages m_0,m_1 s.t $|m_0| = |m_1| = n \cdot \ell$. Let $m_b = (m_{b,1} \mid\mid \ldots \mid\mid m_{b,\ell})$.
- Challenger chooses $\underline{f \leftarrow \mathsf{Perm}[\mathcal{X}]}$. It computes $\underline{\mathsf{ct}_1 = f(m_{0,1})}$. For all $i > 1$, it computes $\underline{\mathsf{ct}_i = f(m_{0,i} \oplus \mathsf{ct}_{i-1})}$.

  Finally, it sends $(\mathsf{ct}_1, \ldots, \mathsf{ct}_\ell)$ to $\mathcal{A}$.
- Adversary sends b'

Hybrid 2:

- Adversary $\mathcal{A}$ sends two messages m_0,m_1 s.t $|m_0| = |m_1| = n \cdot \ell$. Let $m_b = (m_{b,1} \mid\mid \ldots \mid\mid m_{b,\ell})$.
- Challenger chooses $f \leftarrow \mathsf{Perm}[\mathcal{X}]$. It computes $\underline{\mathsf{ct}_1 = f(m_{1,1})}$. For all $i > 1$, it computes $\underline{\mathsf{ct}_i = f(m_{1,i} \oplus \mathsf{ct}_{i-1})}$.

  Finally, it sends $(\mathsf{ct}_1, \ldots, \mathsf{ct}_\ell)$ to $\mathcal{A}$.
- Adversary sends b'

World 1:

1. A sends two messages m_0,m_1 s.t $|m_0| = |m_1| = n \cdot \ell$.
2. Challenger chooses $\underline{\text{PRF key } k \leftarrow \mathcal{K}}$ and computes $\underline{\mathsf{ct}_1 = F(m_{1,1}, k)}$. For all $i > 1$, it computes $\underline{\mathsf{ct}_i = F(m_{1,i} \oplus \mathsf{ct}_{i-1}, k)}$.

   Finally, it sends $(\mathsf{ct}_1, \ldots, \mathsf{ct}_\ell)$ to $\mathcal{A}$.
3. Adversary sends b'.

Let $p_0, p_1, p_{\text{hyb},1}$ and $p_{\text{hyb},2}$ denote the probability of adversary $\mathcal{A}$ outputting 0 in world-0, world-1, hybrid-1 and hybrid-2 respectively.

Claim 4.2. Assuming F is a secure PRP, $p_0 \approx p_{\text{hyb},1}$.

Proof. This follows from the PRP security — for a uniformly random PRP key, $F(\cdot, k)$ is indistinguishable from a uniformly random permutation. $\square$

Claim 4.3. For any adversary $\mathcal{A}$, $p_{\text{hyb},1} - p_{\text{hyb},2} \leq$ ....

Proof. [**TODO**: Complete proof]

□

Claim 4.4. Assuming F is a secure PRP, $p_{\mathrm{hyb},2} \approx p_1$.

Proof. This proof is similar to the proof of Claim 4.2. □

Putting together the above claims, it follows that the CBC mode of encryption satisfies No-Query-Semantic-Security.

□