Instructions

- You are allowed to work in groups of size at most 2.
- The assignments must be typed in Latex, and the resulting pdf must be submitted on Gradescope.
- The bonus questions are somewhat challenging, and you are recommended to attempt them only after solving all the other problems.
- **Plagiarism policy**: You should not discuss your solutions with other group members. Sharing your solutions with other group members is strictly not allowed, and if there are significant similarities in two or more submissions, all relevant group members will be penalized.

  You can refer to resources online, but you should write your solutions in your own words (and also cite the resources used).

Notations

For any positive integer $i < 2^n$, let $[\mathsf{bin}(i)]_n$ denote the binary representation of i using n bits.

Questions

1. (10+5+5 marks)

   Counter-mode MAC, long messages and small signatures

   Let $F : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}^n$ be a secure pseudorandom function with input space, key space and output space all equal to $\{0,1\}^n$. Consider the following MAC scheme $\mathsf{MAC} = (\mathsf{Sign}, \mathsf{Verify})$ with message space $(\{0,1\}^n)^*$ (that is, any message is of $n \cdot k$ bits for some positive integer k) and key space $\{0,1\}^n$.

   - $\mathsf{Sign}(m, k)$: Let $m = (m_1, m_2, \ldots, m_\ell)$. The signing algorithm chooses a uniformly random string $r \leftarrow \{0,1\}^{n/4}$. Next, it sets $x_i = [\mathsf{bin}(\ell)]_{n/4} \mid\mid [\mathsf{bin}(i)]_{n/4} \mid\mid r \mid\mid m_i$, computes $y_i = F(x_i, k)$ and outputs $\sigma = (r, \oplus_i\, y_i)$.
   - $\mathsf{Verify}(m, \sigma, k)$: Let $m = (m_1, m_2, \ldots, m_\ell)$ and $\sigma = (r, z)$. The verification algorithm sets $x_i = [\mathsf{bin}(\ell)]_{n/4} \mid\mid [\mathsf{bin}(i)]_{n/4} \mid\mid r \mid\mid m_i$, computes $y_i = F(x_i, k)$ and checks if $z = \oplus_i y_i$.

   1. Show that the above MAC scheme is strongly unforgeable, assuming that F is a secure pseudorandom function. For the security analysis, you must define appropriate hybrid games, and formally state why the consecutive hybrids are indistinguishable. Finally, you must argue why the adversary's probability of success in the final game is negligible.
   2. **Concrete security:** Suppose we are using AES-128 for the PRF. The input space, key space and output space are all $\{0,1\}^{128}$. Additionally, you are given that any algorithm that sees at most 2^{64} AES evaluations (on inputs of its choice, using the same randomly chosen key) has at most $1/2^{64}$ advantage in the PRP security game. Propose the best possible attack on the above MAC scheme.
   3. Suppose we wish to support arbitrary bit-strings, instead of bit strings whose length is a multiple of n. Propose a modification of the above scheme that can support message space $\mathcal{M} = \{0,1\}^*$. Argue informally why you think the modification is a secure MAC scheme (no formal proof of security needed here).

2. (10 marks) **CBC-MAC and its variants**

Recall the CBC-based MAC scheme discussed in class. This construction, for fixed block-length messages, uses a PRF $F : \mathcal{X} \times \mathcal{K} \to \mathcal{X}$. Let $\mathcal{M} = \mathcal{X}^\ell$ be the message space of our MAC scheme $\mathsf{MAC}_\ell = (\mathsf{Sign}_\ell, \mathsf{Verify}_\ell)$, where Sign_ℓ and Verify_ℓ are defined below.

- $\mathsf{Sign}_\ell(m = (m_1, \ldots, m_\ell) \in \mathcal{X}^\ell, k \in \mathcal{K})$: Let $t_1 = F(m_1, k)$. For all $i \in [2, \ell]$, compute $t_i = F(m_i \oplus t_{i-1}, k)$. Output t_ℓ as the final signature.
- $\mathsf{Verify}_\ell(m = (m_1, \ldots, m_\ell), \sigma, k)$: Let $t_1 = F(m_1, k)$. For all $i \in [2, \ell]$, compute $t_i = F(m_i \oplus t_{i-1}, k)$. Output 1 iff $t_\ell = \sigma$.

We discussed that the above scheme is secure for fixed block-length messages.

Theorem A3.01. *Assuming F is a secure PRF scheme, for every fixed ℓ the above MAC scheme MAC_ℓ is a strongly unforgeable MAC scheme for message space $\mathcal{X}^\ell$.*

1. **A randomized variant of the above scheme:** Suppose we alter the scheme above, and make the signing algorithm randomized.

   - $\mathsf{Sign}'_\ell(m = (m_1, \ldots, m_\ell) \in \mathcal{X}^\ell, k \in \mathcal{K})$: Choose a random string $x \leftarrow \mathcal{X}$. Let $t_1 = F(m_1 \oplus x, k)$. For all $i \in [2, \ell]$, compute $t_i = F(m_i \oplus t_{i-1}, k)$. Output (x, t_ℓ) as the final signature.

   The verification algorithm can be appropriately defined.

   Show that the above scheme is NOT strongly unforgeable, even for fixed length messages.

2. **Handling unbounded length messages:** There are a few easy modifications for handling unbounded block-length messages. One of them is described below. It gives us a MAC scheme with message space $\mathcal{X}^*$.

   - $\mathsf{Sign}^*(m = (m_1, \ldots, m_r), k)$: Let $[r]_\mathcal{X}$ denote some canonical representation of the length r as an element in $\mathcal{X}$. For instance, if $\mathcal{X} = \{0, 1\}^n$, then this would simply be the binary representation of r. Let $m_0 = [r]_\mathcal{X}$, and $m^* = (m_0, m_1, \ldots, m_r)$. Output $\sigma \leftarrow \mathsf{Sign}_{r+1}(m^*, k)$ as the signature.

   Verification can be defined appropriately, and this gives us a secure MAC scheme for message space $\mathcal{X}^*$. It is crucial that the message block-length is **prepended** before signing. Consider the following variant where we **append** the block-length:

   - $\mathsf{Sign}'(m = (m_1, \ldots, m_r), k)$: Let $[r]_\mathcal{X}$ denote some canonical representation of the length r as an element in $\mathcal{X}$. Let $m_{r+1} = [r]_\mathcal{X}$, and $m' = (m_1, \ldots, m_r, m_{r+1})$. Output $\sigma' \leftarrow \mathsf{Sign}_{r+1}(m', k)$ as the signature.

   Show that the resulting MAC scheme is **not** strongly unforgeable.

3. (20 marks) **Semantic Security: Equivalent Definitions**

In Quiz 3, we saw the following security game (Figure 1) for defining secure encryption. Let us refer to this as the *pre-challenge query-based semantic security game.*

We can define a strengthening of this security game (see Figure 2), where the adversary is also allowed **post-challenge** encryption queries. We call this *query-based semantic security.*

1. (10 marks) In the quiz, you had shown that semantic security implies pre-challenge query-based semantic security. A similar reduction can be used to show that semantic security also implies query-based semantic security.

   In this exercise, we will show that query-based semantic security is equivalent to semantic security. In particular, show that if an encryption scheme $\mathcal{E} = (\mathsf{KeyGen}, \mathsf{Enc}, \mathsf{Dec})$ satisfies **query-based semantic security**, then it also satisfies **semantic security** (Definition 15.01 in Lecture 15).

Pre-Challenge Query-based semantic security

1. **Setup:** Challenger chooses an encryption key $k \leftarrow \mathcal{K}$ and a bit $b \leftarrow \{0,1\}$.
2. **Pre-Challenge Encryption queries:** Adversary sends polynomially many encryption queries (adaptively). The i^{th} query consists of a message m_i. The challenger sends $\mathsf{ct}_i = \mathsf{Enc}(m_i, k)$ to the adversary.

   Note that the key k was chosen during setup, and the same key is used for all queries. The bit b is not used here.
3. **Challenge:** After the encryption queries, the adversary sends two messages (m_0^*, m_1^*) and receives $\mathsf{ct}^* = \mathsf{Enc}(m_b^*, k)$.

   Note that the key k and bit b were chosen during setup phase.
4. **Guess:** The adversary sends its guess b', and wins the security game if $b = b'$.

Figure 1: Pre-Challenge Query-based semantic security game

Query-based semantic security

1. **Setup:** Challenger chooses an encryption key $k \leftarrow \mathcal{K}$ and a bit $b \leftarrow \{0,1\}$.
2. **Pre-Challenge Encryption queries:** Adversary sends polynomially many encryption queries (adaptively). The i^{th} query consists of a message m_i. The challenger sends $\mathsf{ct}_i = \mathsf{Enc}(m_i, k)$ to the adversary.

   Note that the key k was chosen during setup, and the same key is used for all queries. The bit b is not used here.
3. **Challenge:** After the encryption queries, the adversary sends two messages (m_0^*, m_1^*) and receives $\mathsf{ct}^* = \mathsf{Enc}(m_b^*, k)$.

   Note that the key k and bit b were chosen during setup phase.
4. **Post-Challenge Encryption queries:** Adversary sends polynomially many encryption queries (adaptively). The i^{th} query consists of a message m_i'. The challenger sends $\mathsf{ct}_i' = \mathsf{Enc}(m_i', k)$ to the adversary.
5. **Guess:** The adversary sends its guess b', and wins the security game if $b = b'$.

Figure 2: Query-based semantic security game

In order to prove this, you will need a sequence of hybrid experiments. First, define the hybrid experiments formally. Then show that the consecutive hybrids are indistinguishable, assuming the encryption scheme satisfies query-based semantic security.

2. (10 marks) Somewhat surprisingly, we can show that the post-challenge queries are not very useful. Pre-challenge query-based semantic security, query-based semantic security and semantic security are all equivalent security definitions!

   Show that if there exists a p.p.t. adversary $\mathcal{A}$ that breaks query-based semantic security, then there exists a p.p.t reduction algorithm $\mathcal{B}$ that breaks pre-challenge query-based security.

   Note tha the reduction algorithm is not allowed to make any queries to the challenger after it receives the challenge ciphertext. However, it must somehow respond to the adversary's post-challenge queries.

   For simplicity, you can make the following assumptions:

   - The message space is $\{0,1\}^n$. However, you must not assume that the encryption scheme encrypts the message bit-by-bit.
   - The adversary makes at most q post-challenge queries.

   This question is a bit challenging, and I would recommend trying it only after finishing the other problems.

Definition A3.01. *An encryption scheme* $\mathcal{E} = (\mathsf{KeyGen}, \mathsf{Enc}, \mathsf{Dec})$ *is said to satisfy query-based semantic security if, for any p.p.t. adversaries* $\mathcal{A}$*, there exists a negligible function* $\mu(\cdot)$ *such that for all* n*,*

$$\Pr\left[\mathcal{A} \textit{ wins the query-based semantic security game } \right] \leq 1/2 + \mu(n)$$

where the probability is over the choice of key k*, randomness used in* Enc*, and the adversary's randomness.*

Definition A3.02. *An encryption scheme* $\mathcal{E} = (\mathsf{KeyGen}, \mathsf{Enc}, \mathsf{Dec})$ *is said to satisfy pre-challenge query-based semantic security if, for any p.p.t. adversaries* $\mathcal{A}$*, there exists a negligible function* $\mu(\cdot)$ *such that for all* n*,*

$$\Pr\left[\mathcal{A} \textit{ wins the pre-challenge query-based semantic security game } \right] \leq 1/2 + \mu(n)$$

where the probability is over the choice of key k*, randomness used in* Enc*, and the adversary's randomness.*