Homework 3

(38 points)

Due on 11:59pm Dec. 9th.
Solution must be typed, preferably using LaTeX.
It can be submitted via email to rachel.lin@cs.ucsb.edu or in class.
You can collaborate with one other student in class. Please acknowledge your collaborator and all public resources that you use.

Part 1 — Message Authentication Codes (MAC)

(6 points)

Task (a) — Message Authentication Codes (MAC) for Broadcasting (6 points)

Suppose user A is broadcasting packets to n recipients $B_1, \cdots B_n$. Each of the recipient B_i wants to be assured that the message he receives is indeed from A and not from anyone else. They decided to use a MAC scheme $(\mathsf{Gen}, \mathsf{Tag}, \mathsf{Ver})$ to ensure this.

(i) **(2 pts)** Suppose that A shares a key k with all the recipients $B_1, \cdots, B_n$. For every message m that A broadcases, he also sends a MAC of that message $\sigma \xleftarrow{\$} \mathsf{Tag}(k, m)$; each receipient can then verify the MAC. Explain in two sentences, why this scheme is insecure, that is, a recipient B_i cannot be sure that a message and tag pair she receives is indeed from A?

(ii) **(2 pts)** Suppose user A has a set $S = \{k_1, \cdots, k_\ell\}$ of ℓ secret keys. Each user B_i has some subset $S_i \subseteq S$ of the keys. When A broadcasts a message m, it also broadcasts ℓ MACs, $\sigma_1, \cdots, \sigma_\ell$, where $\sigma_i \xleftarrow{\$} \mathsf{Tag}(k_i, m)$ is the MAC generated using the i'th key. A recipient B_i now accepts a message if all the MACs correspoinding to the keys she holds in S_i accepts. So what property should the set $S_1 \cdots S_n$ satisfy so that the attack from (i) above no longer applies? (We assume that different recipients are far apart and do not collude.)

(iii) **(2 pts)** Show that when $n = 10$, that is, there are 10 recipients, the broadcaster A only needs to append 5 MACs to every message to satisfy the condition of part (ii). Describe the sets $S_1, \cdots, S_{10} \subseteq \{k_1, \cdots, k_5\}$ you would use.

Part 2 — Collision Resistant Hash Function (CRHF)

(12 points)

Task (a) — The Relation between CRH and OWF. (4 points)

Let $\mathcal{F} = \{f_i : D_i \to R_i\}_{i \in I}$ be a family of CRHFs with a function sampling algorithm Gen, such that, for every function f_i in the support of $\mathsf{Gen}(1^n)$ (i.e., the probability that $\mathsf{Gen}(1^n)$ outputs i is non-zero), $D_i = \{0,1\}^{2n}$ and $R_i = \{0,1\}^n$. That is, $\mathcal{F}$ is a family of CRHFs that **halves** the length of the inputs. Show that $\mathcal{F}$ is a collection of OWFs.

- **(2 pts)** Argue informally why this is the case.

- **(2 pts)** Prove this fomally.

  Hint: The statement can be shown by contra-positive. That is, if there is an adversary A that can invert $y = f_i(x)$ with polynomial probability $1/p(n)$, when the function f_i and input x are sampled at random (i.e., $i \xleftarrow{\$} \mathsf{Gen}(1^n)$ and $x \xleftarrow{\$} \{0,1\}^{2n}$), then there is an adversary B that can find a collision of f_i with polynomial probability $1/q(n)$, when the function f_i is sampled at random. Note that to invert y, A succeeds as long as it finds any x', such that $y = f_i(x')$; in particular, x' does not necessarily equal to x. Can you exploit this fact to construct B that finds collisions?

Task (b) — Merkle Hash Trees. (8 points)

Merkle suggested a parallelizable way of constructing a hash function that compresses inputs by many bits, from a hash function that halves the length of the inputs. More specifically, let's say that f is a compressing function that maps 2 n-bit strings, or blocks, into one n-bit block. Then, consider the following hash function h for compress 2^k n-bit blocks into one n-bit long block, in a tree-fashion as described below.

To compute $h(M)$ with $M = m_0, \cdots m_{2^k-1}$, consider a binary tree with 2^k leaves; its depth is exactly k, with the root at level 0 and leaves at level k. Each tree node is labelled with a n-bit block. For the leaves, the i'th leaf is labelled with the i'th block m_i; for an intermediate node, its label is calculated from the labels l_0, l_1 of it's two children, as $f(l_1, l_2)$. According to this definition, the labels of all nodes on level $k-1$ can be calculated from M, then one can compute labels of all nodes on level $k-2$, so on and so forth, until the label of the root is computed. The label of the root defines the value of $h(M)$.

(i) **(2 pts)** Show that if an adversary can find a collision for hash function h (i.e., $M_1 \neq M_2$ such that $h(M_1) = h(M_2)$). Then one can use M_1 and M_2 to find a collision of f (i.e., $m \neq m'$ and $f(m) = f(m')$).

(ii) **(2 pts)** Consider extending the hash function h to handle messages of different lengths; for simplicity, consider only message consisting of a power of two number of blocks (i.e., $|M| = 2^k \cdot n$ for some k, rather than a fixed k). More precisely, the output of function h' for an input M of length $2^k \cdot n$ is defined exactly as above; its output for an input M' of length $2^l \cdot n$ with $l \neq k$ is defined as above except that the tree will have depth l instead of k.

Now is this new hash function h' that accepts inputs of different lengths collision resistant?

Hint: Try to find two messages M and M' with different lengths that match to the same value.

(iii) **(4 pts)** How can we fix function h' described above so that the property of (i) is true again, i.e., any collision (M, M') of h' can be used to find a collision of h.

Part 3 — Public Key Encryption (PKE) (10 points)

Task (a) — Stronger Security of PKE (10 points)

In class, we introduced many message security, that is,

Definition 1. *We say that a public key encryption scheme* $\Pi = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ *for* n*-bit messages is many message secure if for every polynomial* q*, and every two sequences* $\{\vec{m}_0\}_n$ *and* $\{\vec{m}_1\}_n$*, where* $q = q(n)$*,* $\vec{m}_0 = m_0[1], \cdots m_0[q]$ *and* $\vec{m}_1 = m_1[1], \cdots m_1[q]$*, with* $|m_b[i]| = n$ *for all* b, i*, the following two ensembles are indistinguishable.*

- $\left\{(\mathsf{pk}, \mathsf{sk}) \xleftarrow{\$} Gen(1^n) \ : \ pk, \mathsf{Enc}(\mathsf{pk}, m_0[1]), \cdots, \mathsf{Enc}(\mathsf{pk}, m_0[q])\right\}_n$
- $\left\{(\mathsf{pk}, \mathsf{sk}) \xleftarrow{\$} Gen(1^n) \ : \ pk, \mathsf{Enc}(\mathsf{pk}, m_1[1]), \cdots, \mathsf{Enc}(\mathsf{pk}, m_1[q])\right\}_n$

Furthermore, we say that Π *is single message secure if the above condition holds w.r.t.* $q(n) = 1$ *for all* $n \in \mathbb{N}$*.*

In class we showed that the following theorem holds.

Theorem 1. *Any encryption scheme that is single message secure is also multi-message secure.*

Let us now consider another definition of security called IND-CPA security, which stands for INDistinguishability based, Chosen Plaintext Attack security. This definition considers an adversary who chooses what messages he wants to see encryption of adaptively. More precisely, consider the following experiment $\mathsf{Exp}^b(n)$ defined by an adversary A, an encryption scheme Π, and a parameter n.

Experiment $\mathsf{Exp}^b(n)$**:**

1. $(\mathsf{pk}, \mathsf{sk}) \xleftarrow{\$} \mathsf{Gen}(1^n)$;
2. A receives pk and can issue many queries (as many as he wants) to an encryption oracle $\mathcal{O}$, which on input query $q_i = (m_i^0, m_i^1)$, returns an encryption c of message m_i^b (i.e., $c \xleftarrow{\$} \mathsf{Enc}(\mathsf{sk}, m_i^b)$).
3. Finally A outputs a bit b'.

Define the advantage of A to be the difference in the probabilities that A outputs 1 in experiments $\mathsf{Exp}^0(n)$ and $\mathsf{Exp}^1(n)$, that is,

$$\mathsf{Adv}_A(n) = |\mathbf{Pr}[A \text{ outputs 1 in } \mathsf{Exp}^0(n)] - \mathbf{Pr}[A \text{ outputs 1 in } \mathsf{Exp}^1(n)]|$$

Definition 2. *We say that a public key encryption scheme* $\Pi = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ *for* n*-bit messages is IND-CPA secure if for all non-uniform PPT adversary* A*, there is a negligible function* ε*, such that, the advantage of* A $\mathsf{Adv}_A(n) \leq \varepsilon(n)$ *for all* $n \in \mathbb{N}$*.*

Furthermore, we say that Π *is one-time secure if the above condition holds for all non-uniform PPT adversary* A *that asks for at most one query to the encryption oracle in experiment* $\mathsf{Exp}^b(n)$*, for all* $b \in \{0, 1\}$ *and all* $n \in \mathbb{N}$*.*

Consider a theorem similar to Theorem 1.

Theorem 2. *Any encryption scheme that is one-time secure is IND-CPA secure.*

(i) **(2 pts)** Argue informally why Theorem 2 is true.

(ii) **(2 pts)** Prove Theorem 2 formally.

Hint: For both (i) and (ii), the argument and proof is similar to that for Theorem 1. It needs to use a sequence of hybrid experiments. The hybrid experiments here are just slightly more complicated, since they will be interactive similar to $\mathsf{Exp}^b(n)$.

(iii) **(2 pts)** Argue informally that IND-CPA security implies many-message security.

(vi) **(4 pts)** Is IND-CPA security strictly stronger than many-message security? If your answer is Yes, you need to show a counter-example—an encryption scheme that is many-message secure, but not IND-CPA secure. If you answer is No, you need to show that if an encryption is many-message secure, then it is also IND-CPA secure.

Part 4 — Digital Signatures (10 points)

Task (a) — Improving efficiency of Lamport's signature scheme (10 points)

In class, we gave a construction of a one-time secure signature scheme from one-way function. This construction is due to Lamport, so we refer to it as the Lamport's signature scheme below.

Let $f : \{0,1\}^k \to \{0,1\}^k$ be a one-way function, and let ℓ be an even number. Recall that when using f to instantiate Lamport's signature scheme for messages of length $\ell/2$ (thus allowing us to sign $2^{\ell/2}$ messages), the signing and verification keys consist of ℓ k-bit strings.

In this task, we are going to construct a signature scheme $\Pi = (\mathsf{Gen}, \mathsf{Sign}, \mathsf{Ver})$ which is meant to achieve an efficiency improvement over Lamport's scheme presented in class. For keys consisting of k ℓ-bit strings, we can sign a larger set of messages consisting of roughly $2^\ell/\sqrt{\ell}$ messages.

c) [4 pts] Let S_ℓ be the set of ℓ-bit strings with exactly $\ell/2$ bits equal one. (E.g., for $\ell = 4$, strings like 1100, 0101, etc are elements of S_ℓ.) Also, note that S_ℓ has $\binom{\ell}{\ell/2}$ elements, which is approximately $2^\ell/\sqrt{\ell}$ elements.

Argue that for all distinct ℓ-bit strings $X, X' \in S_\ell$, i.e., $X \neq X'$, there exist $i \neq j$ such that $X[i] = 1$, $X'[i] = 0$, and $X[j] = 0$, $X'[j] = 1$.

c) [6 pts] Show how to use f to build a signature scheme $\Pi = (\mathsf{Gen}, \mathsf{Sign}, \mathsf{Ver})$ with message space $\{M\} = S_\ell$ and with both verification and signing keys consisting of ℓ k-bit strings. Argue (informally) that the resulting scheme is *one-time* secure.

Hint: Use the same ideas behind Lamport's signature scheme.