Homework 2 — Due date: 28 April 2023, 23.59pm

Chameleon hashing and static security of signatures

A chameleon hash function is a regular hash function with an additional algorithm `Trap_Coll` that computes collisions when given as input a trapdoor information. More formally, a chameleon hash function is a triple of probabilistic polynomial-time algorithms (`Gen, Hash, Trap_Coll`) with the following specifications:

- `Gen` takes as input a security parameter and returns a public key pk and a trapdoor $trap$.
- `Hash` is deterministic: it takes as inputs a public key pk, a message M and an r that can be viewed as a random string, and returns $\texttt{Hash}(pk, M, r)$.
- `Trap_Coll` takes as inputs pk, $trap$, a pair (M_1, r_1) and a message M_2, and returns r_2 such that $\texttt{Hash}(pk, M_1, r_1) = \texttt{Hash}(pk, M_2, r_2)$. Intuitively, it finds a collision by modifying the random string used to hash.
- **Collision resistance**: Given pk (but not $trap$), it must be hard to find $(M_1, r_1) \neq (M_2, r_2)$ such that $\texttt{Hash}(pk, M_1, r_1) = \texttt{Hash}(pk, M_2, r_2)$.
- **Uniformity**: For any two messages M_1, M_2, the distributions $\texttt{Hash}(pk, M_1, r)$ and $\texttt{Hash}(pk, M_2, r)$ for r uniform must be identical.

We consider the following chameleon hash function H_{cham}:

- Given a security parameter λ, algorithm `Gen` samples (G, g, p) where $G = \langle g \rangle$ is a cyclic group of known prime order p. It samples x uniformly in $(\mathbb{Z}/p\mathbb{Z}) \setminus \{0\}$ and computes $h = g^x$. It returns $pk = (G, p, g, h)$ and $trap = x$.
- To hash $M \in \mathbb{Z}/p\mathbb{Z}$, it samples r uniformly in $\mathbb{Z}/p\mathbb{Z}$ and returns $H_{cham}(pk, M, r) = g^M \cdot h^r$.

1. Show that H_{cham} is collision-resistant, under the assumption that the Discrete Logarithm Problem (DLP) is hard for G.

2. Describe a correct algorithm `Trap_Coll`.

3. Show that h is a generator of G. Derive that H_{cham} satisfies the uniformity property.

Chameleon hashing is used to transform a signature scheme that is existentially unforgeable for static chosen messages (stat-EU-CMA) into a signature scheme that is existentially unforgeable for adaptive chosen messages (EU-CMA). Stat-EU-CMA security of a signature scheme $(\texttt{KeyGen}, \texttt{Sign}, \texttt{Verify})$ is defined by the following game:

- The adversary gives to the challenger the messages $(M_1, \ldots, M_q)$ it wants to query (before anything else);
- The challenger replies with a verification key vk and valid signatures $(S_1, \ldots, S_q)$, i.e., satisfying $\texttt{Verify}(vk, M_i, S_i) = 1$ for all i;
- The adversary sends a pair (M^*, S^*) to the challenger;

- The adversary wins the game if $M^* \notin \{M_1, \ldots, M_q\}$ and $\texttt{Verify}(vk, M^*, S^*) = 1$.

The scheme is stat-EU-CMA-secure if no probabilistic polynomial-time adversary wins this game with non-negligible probability. We recall that in the EU-CMA security game, the message queries are sent from the adversary to the challenger **after** the challenger has made the verification key *vk* available to the adversary.

We now assume that we have a stat-EU-CMA-secure signature scheme (`KeyGen`, `Sign`, `Verify`) and a secure chameleon hash (`Gen, Hash, Trap_Coll`). Our goal is to build a signature scheme (`KeyGen`′, `Sign`′, `Verify`′) that is EU-CMA-secure. We define:

- `KeyGen`′: Run `KeyGen` to get a verification key *vk* and a secret key *sk*. Run `Gen` to get a public key *pk* and a trapdoor *trap*. Return $vk' = (vk, pk)$ and $sk' = sk$.

- `Sign`′: To sign M using $sk' = sk$, sample a uniform r, compute $h = \texttt{Hash}(pk, M, r)$, and return $S = (r, \texttt{Sign}(sk, h))$.

4. Give a (non-trivial) polynomial-time algorithm `Verify`′ that accepts properly generated signatures.

5. Show that if (`KeyGen`, `Sign`, `Verify`) is stat-EU-CMA-secure and (`Gen, Hash, Trap_Coll`) is a secure chameleon hash function, then (`KeyGen`′, `Sign`′, `Verify`′) is EU-CMA-secure.

Regev public-key encryption

We work over $\mathbb{Z}_q := \mathbb{Z}/q\mathbb{Z}$ for a prime $q \in \mathbb{Z}$, so computations are always modulo q. We denote by Δ the statistical distance, defined for two random variables over $\mathcal{X}$ as:

$$\Delta(X, Y) := \frac{1}{2} \sum_{x \in \mathcal{X}} |\Pr[X = x] - \Pr[Y = x]| \ .$$

Part 1: Leftover Hash Lemma

We say that a family of hash function $(h_k)_{k \in \mathcal{K}}$ with $h_k : \mathcal{X} \to \mathcal{Y}$ for all $k \in \mathcal{K}$, is 2-universal if for all $x, x' \in \mathcal{X}, x \neq x'$, we have:

$$\Pr_{k \leftarrow U(\mathcal{K})}[h_k(x) = h_k(x')] = \frac{1}{|\mathcal{Y}|} \ .$$

6. For $m > n$, we consider the family of hash functions $(h_{\mathbf{A}})_{\mathbf{A} \in \mathbb{Z}_q^{m \times n}}$ with $h_{\mathbf{A}} : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$, defined as $h_{\mathbf{A}} : \mathbf{r} \mapsto \mathbf{r}^T \mathbf{A}$. Show that this family is 2-universal.

For a distribution $\mathcal{D}$ over $\mathcal{X}$, we define its min-entropy as:

$$H_\infty(\mathcal{D}) := -\log(\max_{x \in \mathcal{X}} \Pr_{x' \leftarrow \mathcal{D}}[x' = x]) \ .$$

That is, if $\mathcal{D}$ has min-entropy H, then for any $x \in \mathcal{X}$, $\Pr_{x' \leftarrow \mathcal{D}}[x' = x] \leq \frac{1}{2^H}$.

We admit the following lemma, termed Leftover Hash Lemma (Impagliazzo-Levin-Luby, 1990), which states that for a 2-universal hash function family, the evaluation of h_k on some secret input x is statistically close to a uniform value over $\mathcal{Y}$, even when k is public, as long as x is sampled from a distribution with high enough min-entropy.

Lemma: Let $(h_k)_{k \in \mathcal{K}}$ be a 2-universal family of hash functions with $h_k : \mathcal{X} \to \mathcal{Y}$ for all $k \in \mathcal{K}$. Let $\mathcal{D}$ be a distribution over $\mathcal{X}$ with min-entropy H. Then, we have:

$$\Delta(\{(k, h_k(x))\}, \{(k, y)\}) \leq \sqrt{\frac{|\mathcal{Y}|}{2^H}} \ ,$$

where the distributions are over $k \leftarrow U(\mathcal{K})$, $x \leftarrow \mathcal{D}$, and $y \leftarrow U(\mathcal{Y})$.

7. Let $\mathcal{D} = U(\{0,1\}^m)$. Applying the above lemma, show that if $m \geq 3n \log q$, then we have:

$$\Delta(\{(\mathbf{A}, h_{\mathbf{A}}(\mathbf{r}))\}, \{(\mathbf{A}, \mathbf{u})\}) \leq \frac{1}{q^n} \ ,$$

where the distributions are over $\mathbf{A} \leftarrow U(\mathbb{Z}_q^{m \times n})$, $\mathbf{r} \leftarrow \mathcal{D}$, and $\mathbf{u} \leftarrow U(\mathbb{Z}_q^n)$.

Part 2: Standard IND-CPA security

Consider the following public-key encryption scheme $(\texttt{Gen}, \texttt{Enc}, \texttt{Dec})$ with message space $\{0,1\}$ (so $\beta \in \{0,1\}$ below):

- $\texttt{Gen}(1^\lambda)$: sample $\mathbf{A} \leftarrow U(\mathbb{Z}_q^{m \times n})$, $\mathbf{s} \leftarrow U(\mathbb{Z}_q^n)$, $\mathbf{e} \leftarrow U(\{-\eta, \ldots, \eta\}^m)$, let $\mathbf{b} = \mathbf{A}\mathbf{s} + \mathbf{e}$. Return $pk = (\mathbf{A}, \mathbf{b})$ and $sk = \mathbf{s}$;
- $\texttt{Enc}(pk, \beta)$: Parse pk as $(\mathbf{A}, \mathbf{b})$. Sample $\mathbf{r} \leftarrow \{0,1\}^m$, compute $ct_1 \leftarrow \mathbf{r}^T \mathbf{A}$ and $ct_2 \leftarrow \mathbf{r}^T \mathbf{b} + \beta \lceil q/2 \rfloor$, return (ct_1, ct_2);
- $\texttt{Dec}(sk, (ct_1, ct_2))$: return 0 if $|ct_2 - ct_1 \cdot s| \leq q/4$, else return 1.

The goal of this exercise is to show that this PKE scheme achieves IND-CPA security.

8. Show that the scheme is correct as long as $m\eta \leq q/4$.

9. Show that the distribution of pk is computationally indistinguishable from the uniform distribution, under LWE.

10. Show that, when pk is uniformly random, the distribution of ciphertexts is statistically close to the uniform distribution over $\mathbb{Z}_q^{n+1}$, assuming $m \geq 3n \log q$

11. Conclude about IND-CPA security of the original scheme.

We say that the ℓ-secret $\mathrm{LWE}_{n,m,q,\eta}$ assumption holds if distributions $\{(\mathbf{A}, \mathbf{AS} + \mathbf{E})\}$ and $\{(\mathbf{A}, \mathbf{U})\}$ are computationally indistinguishable, where the distributions are over $\mathbf{A} \leftarrow U(\mathbb{Z}_q^{m \times n})$, $\mathbf{S} \leftarrow U(\mathbb{Z}_q^{n \times \ell})$, $\mathbf{E} \leftarrow U(\{-\eta, \ldots, \eta\}^{m \times \ell})$, and $\mathbf{U} \leftarrow U(\mathbb{Z}_q^{m \times \ell})$.

12. Propose a variant of the above scheme which allows to encrypt ℓ-bit messages (and which is more compact that encrypting each bit of the messages with the previous scheme) and whose security relies on the above assumption.

LWE with small secret

Consider an $\mathrm{LWE}_{n,q,\eta}$ instance $(\mathbf{A}, \mathbf{A}\mathbf{s} + \mathbf{e})$ with small secret, that is, with $\mathbf{A} \leftarrow U(\mathbb{Z}_q^{m\times n})$, $\mathbf{s} \leftarrow U(\{0,1\}^n)$, $\mathbf{e} \leftarrow U(\{-\eta, \ldots, \eta\}^m)$. We aim to show that such instances are also pseudorandom assuming the standard LWE assumption (with parameters to be specified). Note that the exercise also uses the multi-secret LWE introduced just above (question 12).

13. Show that the above instance $(\mathbf{A}, \mathbf{A}\mathbf{s} + \mathbf{e})$ is computationally indistinguishable from instance $(\mathbf{B}\mathbf{C} + \mathbf{N}, (\mathbf{B}\mathbf{C} + \mathbf{N}) \cdot \mathbf{s} + \mathbf{e})$, where $\mathbf{B}, \mathbf{C}, \mathbf{N}$ are respectively uniformly sampled from $\mathbb{Z}_q^{m\times k}, \mathbb{Z}_q^{k\times n}$, and $\{-\nu, \ldots, \nu\}^{m\times n}$ for some $\nu > 0$ to be specified later.

14. Show that the new distribution of $(\mathbf{B}\mathbf{C} + \mathbf{N}, (\mathbf{B}\mathbf{C} + \mathbf{N}) \cdot \mathbf{s} + \mathbf{e})$ described in the previous question is statistically close to the distribution $(\mathbf{B}\mathbf{C} + \mathbf{N}, (\mathbf{B}\mathbf{C}) \cdot \mathbf{s} + \mathbf{e})$ if $\eta \gg n\nu$ (e.g., $\eta > 2^\lambda n\nu$).

15. Using the Leftover Hash Lemma, show that the latter distribution $(\mathbf{B}\mathbf{C} + \mathbf{N}, \mathbf{B}\mathbf{C} \cdot \mathbf{s} + \mathbf{e})$ described in the previous question is statistically close to the distribution $(\mathbf{B}\mathbf{C} + \mathbf{N}, \mathbf{B}\mathbf{t} + \mathbf{e})$ if $n \geq 3k \log q$, where $\mathbf{t}$ is uniform over $\mathbb{Z}_q^k$.

16. Finally show that the latter distribution $(\mathbf{B}\mathbf{C} + \mathbf{N}, \mathbf{B}\mathbf{t} + \mathbf{e})$ described in the previous question is computationally indistinguishable from the distribution $(\mathbf{A}, \mathbf{A}\mathbf{s} + \mathbf{e})$, where $\mathbf{A}$ and $\mathbf{s}$ is uniform over $\mathbb{Z}_q^{m\times n}$ and $\mathbb{Z}_q^n$ respectively. Conclude.