Introduction to cryptology
Final Exam

2024–05–03

Instructions

- No documents allowed.
- *Except indicated otherwise, answers must be carefully justified to get maximum credit.*
- Not all questions are independent, ***but you may admit a result from a previous question by clearly stating it.***
- You may answer in English or French.
- Duration: 2 hours.

Notation & definitions

We recall some notation and definitions.

- For any finite set $\mathcal{S}$, we write $X \twoheadleftarrow \mathcal{S}$ to mean that the random variable X is sampled uniformly from $\mathcal{S}$. Furthermore, in notation such as $X \twoheadleftarrow \mathcal{S}, Y \twoheadleftarrow \mathcal{S}$, the samplings of X and Y are independent (except specified otherwise).
- $\cdot \| \cdot$ denotes bitstring concatenation.

Definition 1 (IND-CPA)**.** We recall briefly and informally that an IND-CPA game is played in two phases. In a training phase, the Adversary has the possibility of sending query messages to the encryption scheme under analysis, and receives their encryption with some (fixed, a priori unknown, uniformly randomly picked) key. In a later challenge phase, the Adversary is tasked with deciding if an encrypted message c is an encryption of m_0 or an encryption of m_1, where m_0 and m_1 are two messages of its choosing of the same length; it wins the game if it makes a correct guess, and its advantage is $|2p - 1|$, with p the winning probability.

Definition 2 (PRF advantage)**.** Let $E : \mathcal{K} \times \mathcal{M} \to \mathcal{M}$ be a block cipher over the finite set $\mathcal{M}$. The *PRF advantage of E* is defined as:

$$\underset{E}{\overset{\mathrm{PRF}}{\mathbf{Adv}}}(q,t) = \max_{A_{q,t}} \left| \Pr[A^{\mathbb{O}}_{q,t}() = 1 \mid \mathbb{O} \twoheadleftarrow \mathrm{Funcs}(\mathcal{M})] - \Pr[A^{\mathbb{O}}_{q,t}() = 1 \mid \mathbb{O} = E(k,\cdot), k \twoheadleftarrow \mathcal{K}] \right|$$

where $\mathrm{Funcs}(\mathcal{M})$ denotes the set of all functions over the finite set $\mathcal{M}$, and $A^{\mathbb{O}}_{q,t}$ denotes an algorithm that runs in time t and makes q queries to the oracle $\mathbb{O}$ it is given access to.

Definition 3 (UP security)**.** Let $E : \mathcal{K} \times \mathcal{M} \to \mathcal{M}$ be a block cipher over the finite set $\mathcal{M}$. Define the game Forge^E as follows:

- The Adversary is an algorithm with oracle access to $\mathbb{O} = E(k,\cdot)$ for $k \twoheadleftarrow \{0,1\}^\kappa$
- The Adversary wins the game iff. it returns a couple (x, y) s.t.:
  1. x was not queried to $\mathbb{O}$
  2. $E(k, x) = y$

The UP security of E is then defined as:

$$\underset{E}{\overset{\mathrm{UP}}{\mathbf{Adv}}}(q,t) = \max_{A_{q,t}} \Pr[A^{\mathbb{O}}_{q,t}() \text{ wins } \mathrm{Forge}^E]$$

where $A_{q,t}$ runs in time t and makes q queries to its oracle.

Definition 4 (EUF-CMA advantage)**.** Let Sign be a signature algorithm, and Vrfy the corresponding verification algorithm. Define the game $\mathrm{Forge}^{\mathsf{Sign}}$ as follows:

— The Challenger generates a pair of keys $(\mathsf{sk}, \mathsf{pk})$
— The Adversary is given pk and oracle access to $\mathbb{O} = \mathsf{Sign}(\mathsf{sk}, \cdot)$
— The Adversary wins the game iff. it returns a pair (m, σ) s.t.:
1. m was not queried to $\mathbb{O}$
2. $\mathsf{Vrfy}_{\mathsf{pk}}(m, \sigma) = 1$

The *EUF-CMA advantage of* Sign is then defined as:

$$\underset{\mathsf{Sign}}{\overset{\text{EUF-CMA}}{\mathbf{Adv}}}(q, t) = \max_{A_{q,t}} \Pr[A^{\mathbb{O}}_{q,t}() \text{ wins Forge}^{\mathsf{Sign}}]$$

where $A_{q,t}$ runs in time t and makes q queries to its oracle.

Definition 5 (CDH advantage)**.** Let G be a cyclic group of order q and g be a generator of G. Define the CDH^G game as follows:
— The Challenger computes (g^a, g^b) where $a \twoheadleftarrow \{0, \ldots, q-1\}$ and $b \twoheadleftarrow \{0, \ldots, q-1\}$
— The Adversary is given (g^a, g^b) and wins iff. it outputs g^{ab}

The *CDH advantage in the group* G is then defined as

$$\underset{G}{\overset{\text{CDH}}{\mathbf{Adv}}}(t) = \max_{A_t} \Pr[A_t() \text{ outputs } g^{ab}]$$

where A_t runs in time t.

Definition 6 (DDH advantage)**.** Let G be a cyclic group of order q and g be a generator of G. Define the DDH^G game as follows:
— The Challenger computes (g^a, g^b) where $a \twoheadleftarrow \{0, \ldots, q-1\}$ and $b \twoheadleftarrow \{0, \ldots, q-1\}$
— The Challenger draws $x \twoheadleftarrow \{0, 1\}$ and computes g^c where $\begin{cases} c \twoheadleftarrow \{0, \ldots, q-1\} & \text{if } x = 0 \\ c = ab & \text{if } x = 1 \end{cases}$
— The Adversary is given (g^a, g^b, g^c) and outputs a bit y

The *DDH advantage in the group* G is then defined as

$$\underset{G}{\overset{\text{DDH}}{\mathbf{Adv}}}(t) = \max_{A_t} \Big| \Pr[A_t() \text{ outputs } 1 \mid x = 1] - \Pr[A_t() \text{ outputs } 1 \mid x = 0] \Big|$$

where A_t runs in time t.

Exercise 1: Short questions

All of those questions are independent and may be answered in any order.

Q.1: Let $H : \mathcal{M} \to \{0, 1\}^n$ be a hash function.
1. Give the definition of a collision for H.
2. Give the definition of a second preimage (problem) for H.

Suppose that for all $x \in \mathcal{M}$, the images $H(x)$ are drawn uniformly and independently at random from $\{0, 1\}^n$. In the two following questions, we consider a "generic" adversary that initially had no *a priori* knowledge about the outputs of H, and that then computed H on q inputs.
3. Without justification, give a non-trivial upper-bound on the probability that the adversary is able to find a collision for H.
4. Without justification, give a non-trivial upper-bound on the probability that the adversary is able to solve a second preimage problem for H.

Q.2: A certain version of the TLS protocol authenticates every packet of 384 bits using a MAC that has tags of bitlength 96. For every *session* of the protocol (what is a session is not important here, but in a typical day one expects much more than 2^{40} sessions to be created worldwide), an identifier that is expected to *uniquely* identify the session among all possible sessions (past and future) is taken to be the 96-bit tag of a designated packet that is part of the session.
1. Identify a problem in the above process.
2. Propose a simple solution to fix it.

Q.3: Let G be a cyclic group of order q, with generator g.

1. Define the *discrete logarithm problem* (DLP^{G}) in G (the inputs and outputs).
2. In the group G, CDH (resp. DDH, resp. DLP) is informally considered *hard* if any *efficient* adversary only has a *small* CDH advantage (resp. DDH advantage, resp. probability of success). What implications are there between CDH hardness, DDH hardness and DLP hardness? *Only a brief justification is required.*

Exercise 2: No confidentiality from unpredictability

Q.1: Let $E : \{0,1\}^{\kappa} \times \{0,1\}^{n} \to \{0,1\}^{n}$ be a block cipher whose unpredictability is "optimal", in the sense that for $q < 2^{n}$ and any t, $\mathbf{Adv}_{E}^{\mathrm{UP}}(q,t) = 1/(2^{n} - q)$. Further let $x\|b$ denote the bitstring of length $n+1$ formed by the concatenation of $x \in \{0,1\}^{n}$ and $b \in \{0,1\}$; then define $E' : \{0,1\}^{\kappa} \times \{0,1\}^{n+1} \to \{0,1\}^{n+1}$ as $E'(k, x\|b) = E(k,x)\|b$.

1. Show that E' is a block cipher, *i.e.*, that for all $k \in \{0,1\}^{\kappa}$, $E'(k,\cdot)$ is a permutation.
2. Show that:
$$\underset{E'}{\overset{\mathrm{UP}}{\mathbf{Adv}}}(q,t) = 1/(2^{n} - q)$$
by using a reduction.
3. Show that:
$$\underset{E'}{\overset{\mathrm{PRF}}{\mathbf{Adv}}}(1,1) \geqslant 1/2$$
by describing and analysing an explicit attack.

Q.2: Let $\mathsf{CTR}[E']$ denote the encryption scheme obtained by applying any instance of the CTR mode[1] to E' from the previous question.

1. Show that:
$$\underset{\mathsf{CTR}[E']}{\overset{\text{IND-CPA}}{\mathbf{Adv}}}(1,1) = 1$$
by describing and analysing an explicit attack.

Q.3: Let $F : \{0,1\}^{\kappa} \times \{0,1\}^{n} \to \{0,1\}^{n}$ be an arbitrary block cipher, and SECDEF be some security definition for block ciphers. We (informally) say that the IND-CPA security of the CTR mode *reduces tightly* to SECDEF security if one has:

$$\underset{\mathsf{CTR}[F]}{\overset{\text{IND-CPA}}{\mathbf{Adv}}}(q,t) \leqslant \underset{F}{\overset{\mathrm{SECDEF}}{\mathbf{Adv}}}(q,t) + \mathrm{small}_{\kappa,n}(q,t)$$

where $\mathrm{small}_{\kappa,n}$ informally represents any function of q, t, κ, n such that if q and t are both "much less" than both of 2^{κ} and 2^{n}, then $\mathrm{small}_{\kappa,n}(q,t)$ is "much less" than 1.

1. Deduce from the previous questions that the IND-CPA security of the CTR mode does not reduce tightly to UP security.
2. Justify the informal assertion: "unpredictability is not useful for encryption".
3. Give an example of application where unpredictability may be useful (no justification is necessary).

[1] You may for instance assume the simplified "one-way" mode for one-block messages of the lecture.

Exercise 3: BLS signature

In this exercise, we are given two cyclic groups G and Γ of the same prime order q, and generators g and γ of G and Γ respectively. We are also given a *pairing*, namely a function $e : \mathsf{G} \times \mathsf{G} \to \Gamma$ which is *non-degenerate*, i.e., $e(g, g) = \gamma$, and *bilinear*, i.e., $e(g^a, g^b) = \gamma^{ab}$ for all $a, b \in \{0, \dots, q-1\}$.

Q.1: We consider the following signature scheme (due to Boneh, Lynn and Shacham), where G, Γ, q and e are as above and $\mathsf{H} : \{0,1\}^* \to \mathsf{G}$ is a hash function, all publicly known:

— Gen samples $x \twoheadleftarrow \mathbb{Z}/q\mathbb{Z}$ and outputs $(pk, sk) = (g^x, x)$;

— $\mathsf{Sign}_{sk}(m) = \mathsf{H}(m)^x$ for a message $m \in \{0,1\}^*$;

— $\mathsf{Vrfy}_{pk}(m, \sigma) = 1$ if and only if $e(\sigma, g) = e(\mathsf{H}(m), pk)$.

1. Show that this signature scheme is correct.

We aim to show that the BLS signature scheme is EUF-CMA secure if CDH is hard in the group G, when $\mathsf{H}(\cdot)$ is modeled as a random oracle. *Reminder:* $\mathsf{H}(\cdot)$ *being a random oracle means that the only way to access a value* $\mathsf{H}(m)$ *is to ask the oracle, and that this value is uniform in* G*, independent from the other* $\mathsf{H}(m')$*.*

Q.2: Let $\mathcal{A}$ be an adversary in the game $\mathsf{Forge}^{\mathsf{Sign}}$, with running time T and advantage ϵ. Since H is modeled as a random oracle, $\mathcal{A}$ has also oracle access to $\mathsf{H}(\cdot)$. We make the following assumptions on $\mathcal{A}$:

— When it queries $\mathsf{Sign}_{sk}(m)$ for some m, it also queries $\mathsf{H}(m)$;

— Before returning (m, σ), it queries $\mathsf{H}(m)$;

— It does not query $\mathsf{H}(\cdot)$ twice on the same value;

— The total number of queries to $\mathsf{H}(\cdot)$ is t, denoted m_1 to m_t (in order).

1. Show that if $\mathcal{A}$ does not query $\mathsf{H}(m)$ before returning m, σ, its advantage is $1/q$.

Q.3: Given $\mathcal{A}$, we build an adversary $\mathcal{C}$ in the $\mathsf{CDH}^{\mathsf{G}}$ game, that *uses* $\mathcal{A}$: $\mathcal{C}$ plays the role of the challenger in the game $\mathsf{Forge}^{\mathsf{Sign}}$ and gets the result that $\mathcal{A}$ finally returns; to be the challenger, $\mathcal{C}$ has to answer the queries of $\mathcal{A}$. We first make a strong assumption on $\mathcal{A}$: we assume that if it returns (m, σ) at the end, m is actually the last query to $\mathsf{H}(\cdot)$, that is $m = m_t$.

To answer a query $\mathsf{H}(m_i)$, $i < t$, $\mathcal{C}$ samples $r_i \twoheadleftarrow \{0, \dots, q-1\}$ and sets $\mathsf{H}(m_i) = g^{r_i}$. And then to answer the related query $\mathsf{Sign}_{sk}(m_i)$, it outputs $\sigma_i = pk^{r_i}$. For the last query $\mathsf{H}(m_t)$, it outputs g^b. Finally, if $\mathcal{A}$ returns (m, σ) where $m = m_t$, $\mathcal{C}$ outputs σ.

1. Justify that $\mathsf{H}(m_i)$, $i < t$, is indeed uniform in G.
2. Prove that σ_i is a valid signature for m_i for all $i < t$.
3. Prove that if (m, σ) is a valid pair, then $\sigma = g^{ab}$.

Q.4: We now remove the strong assumption: m may be any m_i. Therefore, $\mathcal{C}$ first *guesses* i (that is, samples $i \twoheadleftarrow \{1, \dots, t\}$). It answers queries $\mathsf{H}(m_j)$, $j \neq i$, in the same way as before, and answers $\mathsf{H}(m_i)$ with g^b. If $\mathcal{A}$ queries $\mathsf{Sign}_{sk}(m_i)$, $\mathcal{C}$ stops with FAILURE.

1. What is the probability that the guess of $\mathcal{C}$ is correct?
2. Express the advantage and the running time of $\mathcal{C}$ in terms of ϵ and T.
3. Draw a conclusion: Why do the previous questions allow to conclude that if the CDH is hard in G and H is modeled as a random oracle, then the BLS signature scheme is EUF-CMA secure?