Introduction to cryptology
Examen terminal

2024–05–03

Instructions

— Aucun document n'est autorisé.
— *Sauf mention contraire, les réponses doivent être soigneusement justifiées pour tous les points.*
— Toutes les questions ne sont pas indépendantes, ***mais vous pouvez admettre le résultat d'une question précédente en l'indiquant clairement.***
— Vous pouvez composer en français ou en anglais.
— Durée : 2 heures.

Notation & définitions

On rappelle quelques notations et définitions.
— Pour un ensemble fini $\mathcal{S}$, on écrit $X \twoheadleftarrow \mathcal{S}$ pour signifier que la variable aléatoire X est tirée uniformément dans $\mathcal{S}$. De plus, dans une notation telle que $X \twoheadleftarrow \mathcal{S}, Y \twoheadleftarrow \mathcal{S}$, les tirages de X et Y sont indépendants (sauf mention contraire).
— $\cdot\|\cdot$ dénote la concaténation de chaînes binaires.

Definition 1 (IND-CPA)**.** On rappelle brièvement et informellement qu'un jeu IND-CPA se déroule en deux phases. Dans une première phase d'entraînement, l'Adversaire a la possibilité d'envoyer des messages de requêtes au schéma en train d'être analysé, et reçoit leur chiffrement avec une certaine clef (fixée, a priori inconnue, tirée aléatoirement et uniformément). Dans une seconde phase de challenge, l'Adversaire doit décider si un message chiffré c est un chiffrement de m_0 ou de m_1, avec m_0 et m_1 deux messages de son choix de la même longueur ; il gagne le jeu s'il devine correctement, et son avantage est défini comme $|2p - 1|$, où p est la probabilité de gain.

Definition 2 (Avantage PRF)**.** Soit $E : \mathcal{K} \times \mathcal{M} \to \mathcal{M}$ un chiffre par bloc sur l'ensemble fini $\mathcal{M}$, l'*avantage PRF de* E est défini comme :

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

où $\mathrm{Funcs}(\mathcal{M})$ dénote l'ensemble de toutes les fonctions sur l'ensemble fini $\mathcal{M}$, et $A^{\mathbb{O}}_{q,t}$ dénote un algorithme qui s'exécute en temps t et effectue q requêtes à l'oracle $\mathbb{O}$ auquel il a accès.

Definition 3 (Sécurité UP)**.** Soit $E : \mathcal{K} \times \mathcal{M} \to \mathcal{M}$ un chiffre par bloc sur l'ensemble fini $\mathcal{M}$, on définit le jeu Forge^E comme suit :
— L'Adversaire est un algorithme ayant accès à un oracle $\mathbb{O} = E(k, \cdot)$ pour $k \twoheadleftarrow \{0,1\}^\kappa$.
— L'Adversaire gagne le jeu ssi. il renvoie un couple (x, y) t.q. :
  1. x n'est pas une requête à $\mathbb{O}$;
  2. $E(k, x) = y$.

La sécurité UP de E est alors définie comme :

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

où $A_{q,t}$ s'exécute en temps t et effectue q requêtes à son oracle.

Definition 4 (Avantage EUF-CMA)**.** Soit Sign un algorithme de signature, et Vrfy l'algorithme de vérification correspondant. On définit le jeu $\text{Forge}^{\text{Sign}}$ comme suit :

— le Challenger engendre une paire de clés (sk, pk) ;
— on donne à l'Adversaire pk et un accès à l'oracle $\mathbb{O} = \text{Sign}(\text{sk}, \cdot)$;
— l'Adversaire remporte le jeu ssi. il renvoie une pair (m, σ) telle que :
  1. $\mathbb{O}$ n'a pas été interrogé sur m,
  2. $\text{Vrfy}_{\text{pk}}(m, \sigma) = 1$.

L'*avantage EUF-CMA de* Sign est alors défini comme

$$\mathop{\mathbf{Adv}}_{\text{Sign}}^{\text{EUF-CMA}}(q, t) = \max_{A_{q,t}} \Pr[A_{q,t}^{\mathbb{O}}() \text{ remporte } \text{Forge}^{\text{Sign}}]$$

où $A_{q,t}$ s'exécute en temps t et effectue q requête à son oracle.

Definition 5 (Avantage CDH)**.** Soit G un groupe cyclique d'ordre q et g un générateur de G. On définit le jeu CDH^G comme suit :

— le Challenger calcule (g^a, g^b) où $a \twoheadleftarrow \{0, \dots, q-1\}$ et $b \twoheadleftarrow \{0, \dots, q-1\}$;
— l'Adversaire reçoit (g^a, g^b) et remporte le jeu ssi. il renvoie g^{ab}.

L'*avantage CDH dans le groupe* G est alors défini comme

$$\mathop{\mathbf{Adv}}_{G}^{\text{CDH}}(t) = \max_{A_t} \Pr[A_t() \text{ renvoie } g^{ab}]$$

où A_t s'exécute en temps t.

Definition 6 (Avantage DDH)**.** Soit G un groupe cyclique d'ordre q et g un générateur de G. On définit le jeu DDH^G comme suit :

— le Challenger calcule (g^a, g^b) où $a \twoheadleftarrow \{0, \dots, q-1\}$ et $b \twoheadleftarrow \{0, \dots, q-1\}$;
— le Challenger tire $x \twoheadleftarrow \{0, 1\}$ et calcule g^c où $\begin{cases} c \twoheadleftarrow \{0, \dots, q-1\} & \text{si } x = 0, \\ c = ab & \text{si } x = 1 \,; \end{cases}$
— l'Adversaire reçoit (g^a, g^b, g^c) et renvoie un bit y.

L'*avantage DDH dans le groupe* G est alors défini comme

$$\mathop{\mathbf{Adv}}_{G}^{\text{DDH}}(t) = \max_{A_t} \Big| \Pr[A_t() \text{ renvoie } 1 \mid x = 1] - \Pr[A_t() \text{ renvoie } 1 \mid x = 0] \Big|$$

où A_t s'exécute en temps t.

Exercice 1 : Questions courtes

Toutes ces questions sont indépendantes et peuvent être traitées dans n'importe quel ordre.

Q.1 : Soit $H : \mathcal{M} \to \{0,1\}^n$ une fonction de hachage.

1. Donnez la définition d'une collision pour H.
2. Donnez la définition d'un (problème de) seconde préimage pour H.

On suppose que pour tout $x \in \mathcal{M}$, les images $H(x)$ sont tirées aléatoirement uniformément et indépendamment dans $\{0,1\}^n$. Dans les deux questions suivantes, on considère un adversaire « générique » qui n'a initialement aucune connaissance *a priori* sur les sorties de H, et qui calcule ensuite H sur q entrées.

3. Sans justification, donnez un majorant non trivial de la probabilité que l'adversaire trouve une collision pour H.
4. Sans justification, donnez un majorant non trivial de la probabilité que l'adversaire résolve un problème de seconde préimage pour H.

Q.2 : Une certaine version du protocole TLS authentifie chaque paquet de 384 bits avec un MAC dont les tags font 96 bits. Pour chaque *session* du protocole (ce en quoi consiste une session n'est pas important ici, mais on s'attend à ce que bien plus que 2^{40} sessions soient mondialement créées en une journée), un identifiant qui est supposé uniquement identifier la session parmi toutes les autres sessions possibles (passées et futures) est pris comme le tag de 96 bits d'un certain paquet faisant partie de la session.

1. Identifiez un problème dans le processus décrit ci-dessus.
2. Proposez une solution simple pour le résoudre.

Q.3 : Soit G un groupe cyclique d'ordre q et g un générateur de G.

1. Définir le *problème du logarithme discret* (DLP) dans G (les entrées et sorties).
2. Dans le groupe G, CDH (resp. DDH, resp. DLP) est informellement considéré *difficile* si tout adversaire *efficace* a un *faible* avantage CDH (resp. avantage DDH, resp. probabilité de succès). Quelles implications existent entre la difficulté de CDH, celle de DDH et celle du DLP ? *Seule une brève justification est demandée.*

Exercice 2 : Pas de confidentialité par l'imprédictabilité

Q.1 : Soit $E : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^n$ un chiffre par bloc dont l'imprédictabilité est « optimale », dans le sens où pour $q < 2^n$ et tout t, $\mathbf{Adv}^{\mathrm{UP}}_E(q,t) = 1/(2^n - q)$. Soit $x\|b$ la chaîne binaire de longueur $n+1$ formée par la concaténation de $x \in \{0,1\}^n$ et $b \in \{0,1\}$, on définit $E' : \{0,1\}^\kappa \times \{0,1\}^{n+1} \to \{0,1\}^{n+1}$ comme $E'(k, x\|b) = E(k,x)\|b$.

1. Montrez que E' est un chiffre par bloc, c-à-d que pour tout $k \in \{0,1\}^\kappa$, $E'(k,\cdot)$ est une permutation.
2. Montrez que :
$$\mathop{\mathbf{Adv}}^{\mathrm{UP}}_{E'}(q,t) = 1/(2^n - q)$$
en utilisant une réduction.
3. Montrez que :
$$\mathop{\mathbf{Adv}}^{\mathrm{PRF}}_{E'}(1,1) \geqslant 1/2$$
en décrivant et analysant une attaque explicite.

Q.2 : Soit $\mathsf{CTR}[E']$ le chiffrement symétrique obtenu en appliquant n'importe quelle instance du mode CTR[1] à E'.

1. Montrez que :
$$\mathop{\mathbf{Adv}}^{\text{IND-CPA}}_{\mathsf{CTR}[E']}(1,1) = 1$$
en décrivant et analysant une attaque explicite.

Q.3 : Soit $F : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^n$ un chiffre par bloc arbitraire et SECDEF une définition de sécurité pour chiffres par bloc. On dit (informellement) que la sécurité IND-CPA du mode CTR se réduit de façon *tight* à la sécurité SECDEF si on a que :

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

où $\mathrm{small}_{\kappa,n}$ représente informellement toute fonction de q, t, κ, n telle que si q et t sont tous deux « très inférieurs » à à la fois 2^κ et 2^n, alors $\mathrm{small}_{\kappa,n}(q,t)$ est « très inférieur » à 1.

1. Déduisez des questions précédentes que la sécurité IND-CPA du mode CTR ne peut pas se réduire de façon tight à la sécurité UP.
2. Justifiez la proposition informelle : « l'imprédictabilité n'est pas utile pour le chiffrement ».
3. Donnez (sans justification) un exemple d'application pour laquelle l'imprédictabilité peut être utile.

1. Vous pouvez par exemple supposer le mode simplifié « unidirectionnel » et pour messages d'un seul bloc vu en cours.

Exercice 3 : Signature BLS

Dans cet exercice, on nous donne deux groupes cycliques G et Γ de même ordre premier q, et des générateurs g et γ de G et Γ respectivement. On nous donne également un *couplage*, à savoir une fonction $e : G \times G \to \Gamma$ qui est *non-dégénérée*, i.e., $e(g, g) = \gamma$, et *bilinéaire*, i.e., $e(g^a, g^b) = \gamma^{ab}$ pour tout a, $b \in \{0, \ldots, q-1\}$.

Q.1 : On considère le schéma de signature suivant (dû à Boneh, Lynn et Shacham), où G, Γ, q et e sont comme ci-dessus et $H : \{0,1\}^* \to G$ est une fonction de hachage, tous connus publiquement :

— Gen échantillonne $x \leftarrow \mathbb{Z}/q\mathbb{Z}$ et renvoie $(pk, sk) = (g^x, x)$;
— $\mathsf{Sign}_{sk}(m) = H(m)^x$ pour un message $m \in \{0,1\}^*$;
— $\mathsf{Vrfy}_{pk}(m, \sigma) = 1$ si et seulement si $e(\sigma, g) = e(H(m), pk)$.

1. Montrer que ce schéma de signature est correct.

On souhaite montrer que le schéma de signature BLS est EUF-CMA sûr si CDH est difficile dans le groupe G, lorsque $H(\cdot)$ est modélisée comme un oracle aléatoire. *Rappel : dire qu'*$H(\cdot)$ *est un oracle aléatoire signifie que la seule manière d'accéder à une valeur* $H(m)$ *est d'interroger l'oracle, et que cette valeur est uniforme dans* G*, indépendante des autres* $H(m')$.

Q.2 : Soit $\mathcal{A}$ un adversaire dans le jeu $\mathsf{Forge}^{\mathsf{Sign}}$, de temps d'exécution T et avantage ϵ. Puisque H est modélisée par un oracle aléatoire, $\mathcal{A}$ a également un accès par oracle à $H(\cdot)$. On fait les hypothèses suivantes sur $\mathcal{A}$:

— Quand il effectue une requête $\mathsf{Sign}_{sk}(m)$ pour un certain m, il fait également la requête $H(m)$;
— Avant de renvoyer (m, σ), il effectue la requête $H(m)$;
— Il ne fait jamais deux requêtes à $H(\cdot)$ sur la même entrée ;
— Le nombre total de requêtes à $H(\cdot)$ est t, dénotées m_1 à m_t (dans l'ordre).

1. Montrer que si $\mathcal{A}$ ne fait pas la requête $H(m)$ avant de renvoyer m, σ, son avantage est $1/q$.

Q.3 : Étant donné $\mathcal{A}$, on construit un adversaire $\mathcal{C}$ dans le jeu CDH^G, qui *utilise* $\mathcal{A}$: $\mathcal{C}$ joue le rôle du challenger dans le jeu $\mathsf{Forge}^{\mathsf{Sign}}$ et obtient le résultat que $\mathcal{A}$ renvoie à la fin ; pour être le challenger, $\mathcal{C}$ doit répondre aux requêtes de $\mathcal{A}$. On commence avec une hypothèse forte sur $\mathcal{A}$: on suppose que s'il renvoie (m, σ) à la fin, m est sa dernière requête à $H(\cdot)$, c'est-à-dire $m = m_t$.

Pour répondre à une requête $H(m_i)$, $i < t$, $\mathcal{C}$ échantillonne $r_i \leftarrow \{0, \ldots, q-1\}$ et définit $H(m_i) = g^{r_i}$. Puis pour répondre à la requête associée $\mathsf{Sign}_{sk}(m_i)$, il renvoie $\sigma_i = pk^{r_i}$. Pour la dernière requête $H(m_t)$, il renvoie g^b. Finalement, si $\mathcal{A}$ renvoie (m, σ) où $m = m_t$, $\mathcal{C}$ renvoie σ.

1. Justifier que $H(m_i)$, $i < t$, est bien uniforme dans G.
2. Démontrer que σ_i est une signature valide pour m_i pour tout $i < t$.
3. Démontrer que si (m, σ) est une paire valide, alors $\sigma = g^{ab}$.

Q.4 : On supprime maintenant l'hypothèse forte : m peut être n'importe quel m_i. Par conséquent, $\mathcal{C}$ commence par *deviner* i (c'est-à-dire, échantillonne $i \leftarrow \{1, \ldots, t\}$). Il répond aux requêtes $H(m_j)$, $j \neq i$, de la même manière qu'avant, et répond à $H(m_i)$ avec g^b. Si $\mathcal{A}$ effectue la requête $\mathsf{Sign}_{sk}(m_i)$, $\mathcal{C}$ s'arrête avec ÉCHEC.

1. Quelle est la probabilité que $\mathcal{C}$ ait correctement deviné ?
2. Exprimer l'avantage et le temps de calcul de $\mathcal{C}$ en fonction d'ϵ et T.
3. Tirer une conclusion : pourquoi les questions précédentes permettent de conclure que si CDH est difficile dans G et si H est modélisée par un oracle aléatoire, alors le schéma de signature BLS est EUF-CAM sûr ?