# Advanced Cryptography — Final Exam
## Solution

Serge Vaudenay

3.8.2019

- duration: 3h
- any document allowed
- a pocket calculator is allowed
- communication devices are not allowed
- the exam invigilators will **not** answer any technical question during the exam
- readability and style of writing will be part of the grade

*The exam grade follows a linear scale in which each question has the same weight.*

## 1 Security of Key Agreement

We consider a key agreement scheme defined by

- one PPT algorithm $\mathsf{setup}(1^s) \to \mathsf{pp}$ which generates public parameters $\mathsf{pp}$;
- two probabilistic polynomially bounded interactive machines $A$ and $B$ with input $\mathsf{pp}$ and producing a secret output $K$ (denoted by $K_A$ for $A$ and by $K_B$ for $B$).

Correctness implies that the following game outputs 1 with probability 1.

1: $\mathsf{setup}(1^s) \to \mathsf{pp}$
2: make $A(\mathsf{pp})$ and $B(\mathsf{pp})$ interact with each other and output $K_A$ and $K_B$
3: output $1_{K_A=K_B}$

**Q.1** Give a formal definition for the security against key recovery under passive attacks.

*Given an adversary $\mathcal{A}$, we consider the following game with security parameter $s$.*

*1:* $\mathsf{setup}(1^s) \to \mathsf{pp}$
*2: make $A(\mathsf{pp})$ and $B(\mathsf{pp})$ interact with each other and output $K_A$ and $K_B$; define* $\mathsf{transcript}$ *as the list of exchanged messages*
*3: run $\mathcal{A}(\mathsf{pp}, \mathsf{transcript}) \to K$*
*4: output $1_{K=K_A=K_B}$*

*The protocol is secure against key recovery under passive attack if for any PPT adversary $\mathcal{A}$, the above game returns 1 with negligible probability.*

**Q.2** Formalize how to define the Diffie-Hellman protocol under this setting.

*In the Diffie-Hellman protocol, we assume that* $\mathsf{pp}$ *is of form* $\mathsf{pp} = (q, g)$ *where* $g$ *generates a (multiplicatively denoted) group of order* $q$*. The algorithm* $A$ *works as follows:*

```
pick a ∈ Z_q^* at random
pk_A ← g^a
send pk_A
receive pk_B
if pk_B ∉ ⟨g⟩ − {1} then abort
K ← pk_B^a
return K (private output)
```

*The algorithm* $B$ *works as follows:*

```
pick b ∈ Z_q^* at random
pk_B ← g^b
receive pk_A
if pk_A ∉ ⟨g⟩ − {1} then abort
send pk_B
K ← pk_A^b
return K (private output)
```

**Q.3** Formally prove that the Diffie-Hellman protocol is secure in the sense of the previous question if and only if the computational Diffie-Hellman problem is hard.

*By plugging the algorithms $A$ and $B$ in the security game, we obtain*

$\mathsf{setup}(1^s) \to (q, g)$
*pick* $a \in \mathbf{Z}_q^*$ *at random*
$\mathsf{pk}_A \leftarrow g^a$
*pick* $b \in \mathbf{Z}_q^*$ *at random*
$\mathsf{pk}_B \leftarrow g^b$
***if*** $\mathsf{pk}_A \notin \langle g \rangle - \{1\}$ ***then*** *abort*
$K_B \leftarrow \mathsf{pk}_A^b$
***if*** $\mathsf{pk}_B \notin \langle g \rangle - \{1\}$ ***then*** *abort*
$K_A \leftarrow \mathsf{pk}_B^a$
*run* $\mathcal{A}(\mathsf{pp}, \mathsf{pk}_A, \mathsf{pk}_B) \to K$
*output* $1_{K=K_A=K_B}$

*Clearly, the two **if** are useless and we always have $K_A = K_B = g^{ab}$. Hence, the game simplifies to*

$\mathsf{setup}(1^s) \to (q, g)$
*pick* $a \in \mathbf{Z}_q^*$ *at random*
*pick* $b \in \mathbf{Z}_q^*$ *at random*
*run* $\mathcal{A}(q, g, g^a, g^b) \to K$
*output* $1_{K=g^{ab}}$

*which is the computational Diffie-Hellman problem (CDH). An adversary answers 1 in the security game with the same probability as in the CDH game.*

**Q.4** We now consider security against Alice's key recovery under active attacks as defined by the following game:

$\mathsf{setup}(1^s) \to \mathsf{pp}$
$\mathsf{st}_A \leftarrow \mathsf{pp}$, $\mathsf{finished}_A \leftarrow \mathsf{false}$
$\mathsf{st}_B \leftarrow \mathsf{pp}$, $\mathsf{finished}_B \leftarrow \mathsf{false}$
run $\mathcal{A}^{\mathsf{OA},\mathsf{OB}}(\mathsf{pp}) \to K$
output $1_{K=K_A \text{ and } \mathsf{finished}_A}$

$\mathsf{OA}(x)$:
**if** $\mathsf{finished}_A$ **then return**
$\mathsf{st}_A \leftarrow (\mathsf{st}_A, x)$
run $A(\mathsf{st}_A)$ to get private output $\mathsf{st}_A$ and next message $y$
**if** $y$ non-final **then return** $y$
$\mathsf{finished}_A \leftarrow \mathsf{true}$
$K_A \leftarrow \mathsf{st}_A$
**return** $y$

And the same for oracle $\mathsf{OB}$. Prove that the Diffie-Hellman protocol is insecure in this sense.

*The man-in-the-middle attack is breaking the protocol. We consider the adversary:*

*Input:* $(q, g)$

1: *pick* $c \in \mathbf{Z}_q^*$
2: $\mathsf{OA}() \to \mathsf{pk}_A$
3: $\mathsf{OA}(g^c)$
4: ***return*** $\mathsf{pk}_A^c$

*(Note that the interaction with Bob is useless in this security model.)*

**Q.5** Based on some attacks seen in the course, formalize security against key recovery under *active* attacks making $K_A = K_B$. Prove that Diffie-Hellman is secure by assuming that the problem defined by the following game is hard:

1: $\mathsf{setup}(1^s) \to \mathsf{pp} = (q, g)$
2: pick $x, y \in \mathbf{Z}_q^*$
3: $\mathcal{B}(\mathsf{pp}, g^x, g^y) \to (u, v, w)$
4: **return** $1_{u^x = v^y = w \text{ and } u,v,w \in \langle g \rangle \text{ and } w \neq 1}$

where $g$ generates $\langle g \rangle$ of order $q$, with neutral element 1.

*The output of the security game is now* $1_{K = K_A = K_B \text{ and } \mathsf{finished}_A \text{ and } \mathsf{finished}_B}$.
*We want to prove that the protocol is secure. Let $\mathcal{A}$ by an adversary against the protocol. We define $\mathcal{B}$ as follows:*

$\mathcal{B}(\mathsf{pp}, X, Y)$*:*

1: *run* $\mathcal{A}^{\mathsf{OA},\mathsf{OB}}(\mathsf{pp}) \to w$ *and simulate the oracles as follows:*
   $\mathsf{OA}()$*: simulate $A$ choosing $X$*
   *next* $\mathsf{OA}(x)$*: set $v \leftarrow x$*
   $\mathsf{OB}(x)$*: set $u \leftarrow x$ and simulate $B$ choosing $Y$*
2: ***return*** $(u, v, w)$

*When $\mathcal{B}$ is put in its game, the simulation of the selection of the public keys of $A$ and $B$ are perfect. It is also clear that the winning conditions in both games are equivalent. So, they have the same advantage. If the game that $\mathcal{B}$ plays is hard, then it must be the case that $\mathcal{A}$ has a negligible advantage.*

## 2 Advantage Amplification

Let $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ be $2n$ independent Boolean variables. We assume that $X_1, \ldots, X_n$ are identically distributed and that $Y_1, \ldots, Y_n$ are identically distributed. We assume that the statistical distance between the distributions of $X_i$ and $Y_j$ is $\varepsilon$. Given distinguisher, i.e. a Boolean algorithm $\mathcal{A}$ (with unbounded complexity), we define $X = \mathcal{A}(X_1, \ldots, X_n)$ and $Y = \mathcal{A}(Y_1, \ldots, Y_n)$. We are interested in $\mathcal{A}$ which maximizes the statistical distance between the distributions of $X$ and $Y$. We denote by $d$ the statistical distance and we identify random variables by their distributions when computing distances, by abuse of notation.

**Q.1** Prove that $d(X, Y) = d((X_1, \ldots, X_n), (Y_1, \ldots, Y_n))$.

*We know from the course that for any* $\mathcal{A}$

$$d(X, Y) \leq d((X_1, \ldots, X_n), (Y_1, \ldots, Y_n))$$

*and equality can be reached by using the likelihood ratio. We actually known that*

$$\mathcal{A}(z_1, \ldots, z_n) = 1_{\Pr[X_1=z_1,\ldots,X_n=z_n]<\Pr[Y_1=z_1,\ldots,Y_n=z_n]}$$

*reaches the equality case.*

**Q.2** Assume that $\Pr[X_i = 1] = 0$.

**Q.2a** Give the distributions of $X_i$ and $Y_j$.

*We have* $\Pr[X_i = 1] = 0$ *and* $\Pr[X_i = 0] = 1$*. Due to the statistical distance of* $\varepsilon$*, we have* $\Pr[Y_j = 1] = \varepsilon$ *and* $\Pr[Y_j = 0] = 1 - \varepsilon$*.*

**Q.2b** Compute $d(X, Y)$ in terms of $\varepsilon$ and $n$.

*We compute the statistical distance by regrouping all* $(z_1, \ldots, z_n)$ *by their Hamming weight.*

$$\begin{aligned} d(X, Y) &= \frac{1}{2} \sum_{z_1,\ldots,z_n} |\Pr[X_1 = z_1, \ldots, X_n = z_n] - \Pr[Y_1 = z_1, \ldots, Y_n = z_n]| \\ &= \frac{1}{2}\left(1 - (1-\varepsilon)^n\right) + \frac{1}{2} \sum_{h=1}^{n} \binom{n}{h} \varepsilon^h (1-\varepsilon)^{n-h} \\ &= 1 - (1-\varepsilon)^n \end{aligned}$$

*In the sum, only the* $(0, \ldots, 0)$ *case makes the first probability nonzero. This is the* $h = 0$ *case.*

**Q.2c** Give an asymptotic equivalent of the minimal $n$ such that $d(X, Y) \geq \frac{1}{2}$ in terms of $\varepsilon$, when $\varepsilon \to 0$.

$1 - (1-\varepsilon)^n \geq \frac{1}{2}$ *is equivalent to* $n \geq -\frac{\ln 2}{\ln(1-\varepsilon)}$*. So, the minimal* $n$ *is* $n \sim \frac{\ln 2}{\varepsilon}$*.*

**Q.3** Assume now that $\Pr[X_i = 1] = \frac{1}{2}(1-\varepsilon)$ and $\Pr[Y_i = 1] = \frac{1}{2}(1+\varepsilon)$.

**Q.3a** Show that $\mathcal{A}(z_1, \ldots, z_n) = 1_{z_1+\cdots+z_n<\frac{n}{2}}$ makes $d(X,Y)$ maximal.

> *Let $h = z_1 + \cdots + z_n$. We have*
>
> $$\Pr[X_1 = z_1, \ldots, X_n = z_n] = 2^{-n}(1-\varepsilon)^h(1+\varepsilon)^{n-h}$$
> $$\Pr[Y_1 = z_1, \ldots, Y_n = z_n] = 2^{-n}(1+\varepsilon)^h(1-\varepsilon)^{n-h}$$
>
> *So, $\Pr[X_1 = z_1, \ldots, X_n = z_n] < \Pr[Y_1 = z_1, \ldots, Y_n = z_n]$ is equivalent to $(1-\varepsilon)^h(1+\varepsilon)^{n-h} < (1+\varepsilon)^h(1-\varepsilon)^{n-h}$, which is equivalent to $(1+\varepsilon)^{n-2h} < (1-\varepsilon)^{n-2h}$, which is equivalent to $h < \frac{n}{2}$. Hence, the suggested $\mathcal{A}$ is actually equivalent to the optimal algorithm based on the likelihood ratio. We know it makes $d(X,Y)$ maximal.*

**Q.3b** Given that $\Pr[X_1 + \cdots + X_n < \frac{n}{2}] = \Pr[Y_1 + \cdots + Y_n > \frac{n}{2}]$, prove that for $n$ odd, we have $d(X,Y) = |1 - 2\Pr[X_1 + \cdots + X_n < \frac{n}{2}]|$.

> *Actually, $d(X,Y)$ is the advantage which is $d(X,Y) = |\Pr[Y_1 + \cdots + Y_n < \frac{n}{2}] - \Pr[X_1 + \cdots + X_n < \frac{n}{2}]|$. For $n$ odd, we have $\Pr[Y_1 + \cdots + Y_n < \frac{n}{2}] = 1 - \Pr[Y_1 + \cdots + Y_n > \frac{n}{2}]$ which gives the answer.*

**Q.3c** Compute the expected value and the variance of $X_1 + \cdots + X_n$.

> *We have*
> $$E(X_1 + \cdots + X_n) = n \cdot E(X_i) = \frac{n}{2}(1-\varepsilon)$$
> *and*
> $$V(X_1 + \cdots + X_n) = n \cdot V(X_i) = \frac{n}{4}(1-\varepsilon^2)$$
> *because $V(X_i) = E(X_i)(1 - E(X_i))$.*

**Q.3d** By approximating $X_1 + \cdots + X_n$ to a normal distribution, give an asymptotic equivalent to $n$ so that $d(X,Y)$ is a constant.

> *For $\Pr[X_1 + \cdots + X_n < \frac{n}{2}]$ to be constant, we need $\frac{n}{2}\varepsilon$ and $\sqrt{\frac{n}{4}(1-\varepsilon^2)}$ of same order of magnitude. This means $n \sim \frac{\mathsf{cste}}{\varepsilon^2}$.*
> *It is interesting to observe that to amplify the statistical distance with close-to-unbiased distributions, it is harder than for close-by distributions which are heavily biased.*
> *Nice solution from a student: we apply the upper-tail Chernoff bound with $\delta = \frac{\varepsilon}{1-\varepsilon}$ which says*
>
> $$\Pr[X_1 + \cdots + X_n > (1+\delta)\mu] \leq e^{-\frac{\delta^2}{2+\delta}\mu}$$
>
> *hence $\Pr[X_1 + \cdots + X_n > \frac{n}{2}] \leq e^{-\frac{\varepsilon^2}{2(2-\varepsilon)}n}$. So, with $n > \frac{4}{\varepsilon^2}$, we get $\Pr[X_1 + \cdots + X_n > \frac{n}{2}] \leq e^{-1}$.*