Advanced Cryptography — Midterm Exam

Serge Vaudenay

28.4.2015

- 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

1 Blind Computing with a DH Oracle

The goal of this exercise is to look at what happens when the discrete logarithm problem is hard but the Diffie-Hellman problem is easy. Let g be an element of a group of prime order q. Computing the discrete logarithm in the group generated by g is assumed to be hard. We assume that we have an oracle function $\mathsf{DH}(X,Y)$ such that when queried with $X = g^x$ and $Y = g^y$ for integers x and y, it returns $\mathsf{DH}(g^x, g^y) = g^{xy}$ with one unit of time complexity.

In this exercise we construct series of algorithms using the oracle DH. These algorithms also know g and q. They can perform a group multiplication and a group inversion within one unit of time complexity. In each question of this exercise except the last one, we define a function by a property based on values which are not always computable. For instance, a blind multiplication defined by $f_0(g^x, g^y) = g^{xy}$ is implemented by the algorithm

```
$f_0(X,Y)$:
  $Z \leftarrow \mathsf{DH}(X,Y)$
  return $Z$
```

without being able to compute the logarithms x or y of X and Y.

For each of these questions, define an efficient algorithm to implement the computation of the function and give its complexity. When studying the complexity, separate the number of queries, the number of group multiplications/inversions, and the usual asymptotic complexity of other operations.

Q.1 (blind addition) $f_1(g^x, g^y) = g^{x+y}$.

Q.2 (blind scalar multiplication) $f_2(a, g^x) = g^{ax}$ when a is an integer (positive or negative).

Q.3 (blind power) $f_3(e, g^x) = g^{x^e}$ when e is a positive integer.

Q.4 (blind sparse polynomial) $f_4(a_1, e_1, \ldots, a_n, e_n, g^x) = g^{\sum_{i=1}^{n} a_i x^{e_i}}$ when the e_i's are positive integers and the a_i's are nonzero integers.

Q.5 (blind inversion) $f_5(g^x) = g^{\frac{1}{x} \bmod q}$ when $g^x \neq 1$.

Q.6 (blind e-th root when e is invertible) $f_6(e, g^{y^e}) = g^y$ when e is a positive integer which is coprime with $q-1$.

Q.7 (blind square root) $f_7(g^{y^2}) \in \{g^y, g^{-y}\}$. (For simplicity, we assume $q \bmod 4 = 3$.)

Q.8 With the same notations and assumptions, construct a commitment scheme which is deterministic computationally hiding and perfectly binding on $\mathbf{Z}_q$, with the property that given a rational function $f(x_1, \ldots, x_n)$ and some commitments on x_1, ..., and x_n, it is easy to deduce a commitment to $f(x_1, \ldots, x_n)$ without knowing $x_1, \ldots, x_n$.

2 On the Necessary Number of Samples to Distinguish a Biased Coin

We are given a source of independent random bits following one given distribution. We want to distinguish a given distribution from the uniform one. The goal of this exercise is to prove that if ε is the statistical distance between the two distributions, then ε^{-2} is a necessary and sufficient order of magnitude of number of samples which is needed to reach an advantage of $\frac{1}{2}$ or higher.

Given two random variables X and Y with the same support $\mathcal{Z}$, we define

$$L(X,Y) = \sum_{z \in \mathcal{Z}} \left|\Pr[X=z] - \Pr[Y=z]\right|$$

$$D(X\|Y) = \sum_{z \in \mathcal{Z}} \Pr[X=z] \log_2 \frac{\Pr[X=z]}{\Pr[Y=z]}$$

In what follows, some questions are more related to calculus than cryptography. Their results are necessary for the exercise but left as "bonus questions".

Q.1 Let $p = \frac{1}{2}(1+\varepsilon)$ for some $\varepsilon \in [-1,+1]$ and the X_i's be independent boolean random variables with expected value p. Let the Y_i's be independent uniformly distributed boolean random variables. Given a number of samples n, we want to distinguish $X = (X_1, \ldots, X_n)$ from $Y = (Y_1, \ldots, Y_n)$. We assume that the value of ε is known.

Q.1a Given a threshold λ, we propose the distinguisher

```
get the samples $z_1, \ldots, z_n$
compute $s = z_1 + \cdots + z_n$
if $\frac{s}{n} \geq \lambda$ then
    return 1
else
    return 0
end if
```

Show that for some value λ^* (give the formula) of λ, this distinguisher is optimal among those using n samples.

Q.1b (Bonus question) Show that λ^* is close to $\frac{1}{2} + \frac{\varepsilon}{4}$ when $|\varepsilon|$ is small.
HINT: for θ close to 0, $\ln(1+\theta) = \theta - \frac{\theta^2}{2} + o(\theta^2)$.

Q.1c For $n = 12\varepsilon^{-2}$, show that the advantage of the above distinguisher for $\lambda = \frac{1}{2} + \frac{\varepsilon}{4}$ is greater than $\frac{1}{2}$.
HINT: if $Z_1, \ldots, Z_n$ are i.i.d. boolean random variables of expected value μ, the Chernoff-Hoeffding bound says that $\Pr[Z_1 + \cdots + Z_n < n(\mu - t)] \leq e^{-2nt^2}$.

The goal of the next questions is to show that for $n \ll \varepsilon^{-2}$, the best advantage is negligible.

Q.2 If X_1 and X_2 are independent and Y_1 and Y_2 are independent, for $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$, show that $D(X\|Y) = D(X_1\|Y_1) + D(X_2\|Y_2)$.

Q.3 Given two random boolean variables X' and Y', show that $\frac{L(X',Y')^2}{2\ln 2} \leq D(X'\|Y')$.
HINT: express $g(t) = (2\ln 2) \cdot D(X'\|Y') - L(X',Y')^2$ in terms of $t = \Pr[X' = 1]$ then derivate $g(t)$ to study the variations of this function.

Q.4 (Bonus question) Let $p = \frac{1}{2}(1+\varepsilon)$ for some $\varepsilon \in [-1, +1]$ and X_1 be a boolean random variable with expected value p. Let Y_1 be a uniformly distributed boolean random variable. Show that $D(X_1\|Y_1) \leq \frac{\varepsilon^2}{(2\ln 2)\cdot(1-\varepsilon^2)}$.
HINT: the Taylor-Lagrange Theorem states that there exists some t_2 between t_0 and t such that $g(t) = g(t_0) + g'(t_0)(t-t_0) + \frac{1}{2}g''(t_2)(t-t_0)^2$.

Q.5 The aim of the next sub-questions is to show that for all function f, $D(f(X)\|f(Y)) \leq D(X\|Y)$.

Q.5a Show $D(g(X)\|g(Y)) = D(X\|Y)$ for all 1-to-1 mapping g.

Q.5b We say that m is a merging function if every input x except one is a fixed point of m, i.e. $m(x) = x$. Show that an arbitrary f can be written as a composition $f = g \circ m_n \circ \cdots \circ m_1$ of merging functions m_i and a 1-to-1 function g, for some integer n.
HINT: make a proof by induction based on the number of collisions.

Q.5c (Bonus question) Show that for all positive $\alpha, \beta, \alpha', \beta'$ real numbers,

$$(\alpha+\beta)\ln\frac{\alpha+\beta}{\alpha'+\beta'} \leq \alpha\ln\frac{\alpha}{\alpha'} + \beta\ln\frac{\beta}{\beta'}$$

Deduce that $D(m_i(X)\|m_i(Y)) \leq D(X\|Y)$ for all merging functions m_i (as defined in Q.5b).
HINT: use the convexity of $x \mapsto x\ln x$ on the two points $\frac{\alpha}{\alpha'}$ and $\frac{\beta}{\beta'}$ and their weighted average $\frac{\alpha+\beta}{\alpha'+\beta'}$.

Q.5d Show that $D(f(X)\|f(Y)) \leq D(X\|Y)$ for all functions f.

Q.6 Show that the best advantage Adv to distinguish the boolean random variables X_i and Y_i, for $E(X_i) = p = \frac{1}{2}(1+\varepsilon)$ and $E(Y_i) = \frac{1}{2}$ satisfies $\mathsf{Adv} \leq \frac{1}{2} \times \sqrt{\frac{n\varepsilon^2}{1-\varepsilon^2}}$. Assuming that $|\varepsilon| \leq \frac{1}{2}$, deduce that for $n \ll \varepsilon^{-2}$, the best advantage is negligible.
HINT: define f the function mapping the vector of n sample bits to the outcome of the distinguisher. Given $X' = f(X_1, \ldots, X_n)$ and $Y' = f(Y_1, \ldots, Y_n)$ for some independent uniformly distributed bits $Y_1, \ldots, Y_n$, express the advantage in terms of $L(X', Y')$ and bound it in terms of $D(X'\|Y')$.