LASEC
SECURITY AND CRYPTOGRAPHY LABORATORY

Family Name: ....

First Name: ....

Section: ....

Advanced Cryptography

Midterm Exam
Solution

April 28th, 2009

Duration: 3 hours 45 minutes

This document consists of 14 pages.

Instructions

Electronic devices are *not* allowed.

All printed documents are permitted.

Answers must be written on the exercises sheet.

This exam contains 4 *independent* exercises.

Answers can be either in French or English.

Questions of any kind will certainly *not* be answered.
Potential errors in these sheets are part of the exam.

You have to put your full name on the first page and have all pages *stapled*.

1 RSA Public-Key Recovery

Given an integer e and a few (x_i, y_i) pairs such that $y_i = x_i^e \bmod N$ for some unknown common N of known bit-length ℓ, we consider the problem of recovering N. We assume that $0 \le x_i, y_i < N$ and that i ranges from 1 to n.

1. Using Buffon's needle problem we can show that the probability that two independent uniformly distributed integers in $\{0, 1, \ldots, 2^\ell - 1\}$ are coprime tends towards $\frac{6}{\pi^2}$ as ℓ goes to infinity.

   We take independent uniformly distributed integers $X_1, \ldots, X_n$ in $\{0, 1, \ldots, 2^\ell - 1\}$. Show that the probability that $\mathsf{gcd}(X_1, \ldots, X_n) > 1$ is less than $\left(1 - \frac{6}{\pi^2}\right)^{\frac{n}{2}}$ as ℓ goes to infinity.

   Hint: consider $\frac{n}{2}$ disjoint pairs of form (X_{2i-1}, X_{2i}).

Let $\mathcal{S} = \{0, 1, \ldots, 2^\ell - 1\}$

$$\lim_{\ell \to \infty} \Pr_{X_1, X_2 \in \mathcal{S}}[\mathsf{gcd}(X_1, X_2) = 1] = \frac{6}{\pi^2}$$

$$\iff \lim_{\ell \to \infty} \Pr_{X_1, X_2 \in \mathcal{S}}[\mathsf{gcd}(X_1, X_2) > 1] = 1 - \frac{6}{\pi^2}$$

$$\lim_{\ell \to \infty} \Pr_{X_1, \ldots, X_n \in \mathcal{S}}[\mathsf{gcd}(X_1, X_2, \ldots, X_n) > 1]$$

$$\leqslant \lim_{\ell \to \infty} \Pr_{X_{2i-1}, X_{2i} \in \mathcal{S}}[\mathsf{gcd}(X_1, X_2) > 1, \mathsf{gcd}(X_3, X_4) > 1, \ldots, gcd(X_{n-1}, X_n) > 1]$$

$$\leqslant \lim_{\ell \to \infty} \prod_{i=1}^{n/2} \Pr_{X_{2i-1}, X_{2i} \in \mathcal{S}}[\mathsf{gcd}(X_{2i-1}, X_{2i}) > 1]$$

$$\leqslant \prod_{i=1}^{n/2} \lim_{\ell \to \infty} \Pr_{X_{2i-1}, X_{2i} \in \mathcal{S}}[\mathsf{gcd}(X_{2i-1}, X_{2i}) > 1] \leqslant \prod_{i=1}^{n/2} \left(1 - \frac{6}{\pi^2}\right) = \left(1 - \frac{6}{\pi^2}\right)^{n/2}$$

2. We now take independent random integers $X_1, \ldots, X_n$ which are uniformly distributed among the set of all multiples of N in $\{0, 1, \ldots, 2^\ell - 1\}$. Show that $\mathsf{gcd}(X_1, \ldots, X_n) = N$ except with negligible probability as n increases.

$X_i = k_i N,\ k_i \in \mathcal{S} = \{0, 1, \ldots, 2^\ell - 1\}$

$\mathsf{gcd}(X_1, \ldots, X_n) = \mathsf{gcd}(k_1 N, \ldots, k_n N) = N \cdot \mathsf{gcd}(k_1, \ldots, k_n)$

From the previous question we know that:

$$\lim_{\ell \to \infty} \Pr_{k_1, \ldots, k_n \in \mathcal{S}}[\mathsf{gcd}(k_1, \ldots, k_n) > 1] \leqslant \left(1 - \frac{6}{\pi^2}\right)^{n/2}$$

When n increases this probability decreases. Hence $\mathsf{gcd}(k_1, \ldots, k_n) = 1$ except with negligible probability, which means that $\mathsf{gcd}(X_1, \ldots, X_n) = N$ except with negligible probability as n increases.

3. Deduce that we can recover N by computing $\mathsf{gcd}(x_1^e - y_1, \ldots, x_n^e - y_n)$. What is its complexity in terms of ℓ, e, and n?

$x_i^e \mod N = y_i$ and $0 \leqslant x_i, y_i < N$

Hence we can write $x_i^e - y_i = k_i N$

$\mathsf{gcd}(x_1^e - y_1, \ldots, x_n^e - y_n) = \mathsf{gcd}(k_1 N, \ldots, k_n N) = N \cdot \mathsf{gcd}(k_1, \ldots, k_n)$

From the previous question we can conclude that $\mathsf{gcd}(x_1^e - y_1, \ldots, x_n^e - y_n) = N$ except with negligible probability.

The complexity to compute this gcd can be devided in three parts:
First there is n exponentiation to compute: $O(n \cdot e \cdot l^2 \cdot \log e)$
Then there is n substraction to do: $O(n \cdot l)$
And finaly there is the gcd to do between n elements: $O(n \cdot e^2 \cdot l^2)$

Hence the overall computation is $O(n \cdot e^2 \cdot l^2 \cdot \log e)$

2 DP and LP Tricks

Consider a function f from $A = \{0,1\}^p$ to $B = \{0,1\}^q$. We define DP^f and LP^f as functions from $A \times B$ to $\mathbf{R}$ as usual by

$$\begin{aligned}\mathsf{DP}^f(a,b) &= \Pr_X[f(a \oplus X) \oplus f(X) = b] \\ \mathsf{LP}^f(a,b) &= \left(2\Pr_X[a \cdot X = b \cdot f(X)] - 1\right)^2\end{aligned}$$

1. Show that for any $b \neq 0$ we have $\mathsf{DP}^f(0,b) = 0$. Give a necessary and sufficient condition about f so that

$$\forall a \neq 0 \quad \mathsf{DP}^f(a,0) = 0$$

$\forall b \neq 0,\ \mathsf{DP}^f(0,b) = \Pr_X[f(X) \oplus f(X) = b] = \Pr_X[b = 0] = 0$

Note that $\mathsf{DP}^f(a,0) = \Pr_X[f(a \oplus X) \oplus f(X) = 0] = \Pr_X[f(a \oplus X) = f(X)]$

$\forall a \neq 0 \quad \mathsf{DP}^f(a,0) = \Pr_X[f(a \oplus X) = f(X)] = 0 \Longleftrightarrow \forall X,\ \forall a \neq 0 \quad f(a \oplus X) \neq f(X)$

Moreover $a \neq 0 \Longleftrightarrow a \oplus X \neq X \Longleftrightarrow \forall X, Y \quad X \neq Y$

Hence f is injective.

2. Show that for any $a \neq 0$ we have $\mathsf{LP}^f(a,0) = 0$.

$\forall a \neq 0 \quad \mathsf{LP}^f(a,0) = \left(2\Pr_X[a \cdot X = 0] - 1\right)^2$

$\Pr_X[a \cdot X = 0] = \Pr_{X_1,\ldots,X_p}[a_1 X_1 \oplus \ldots \oplus a_p X_p = 0] = \frac{1}{2}$

$\Rightarrow \forall a \neq 0 \quad \mathsf{LP}^f(a,0) = \left(2 \cdot \frac{1}{2} - 1\right)^2 = 0$

3. We define a function g from B to $\mathbf{R}$ by $g(y) = \Pr[f(X) = y]$ for all $y \in B$ where X is uniformly distributed in A. Show that for any function h we have

$$E(h(f(X))) = 2^q \cdot E(g(Y)h(Y))$$

where Y is uniformly distributed in B and where $E(X)$ is the expected value of the random variable X.

Recall: $E(X) = \sum_x x \Pr[X = x]$; and $E(f(X)) = \sum_x f(x) \Pr[X = x]$

$$\begin{aligned} E(h(f(X))) &= \sum_x h(f(x)) \Pr[X = x] \\ &= 2^q \sum_y h(y) \Pr[f(X) = y] \Pr[Y = y] \\ &= 2^q \sum_y h(y) g(y) \Pr[Y = y] \\ &= 2^q E(g(Y)h(Y)) \end{aligned}$$

4. Deduce that

$$\mathsf{LP}^f(0, b) = 2^{2q} \left(E\left(g(Y)(-1)^{b \cdot Y} \right) \right)^2$$

where Y is uniformly distributed in B.

Recall: $\mathsf{LP}^f(a, b) = \left(E\left((-1)^{a \cdot X \oplus b \cdot f(x)} \right) \right)^2$

Hence $\mathsf{LP}^f(0, b) = \left(E\left((-1)^{b \cdot f(x)} \right) \right)^2$

Let $h(y) = (-1)^{b \cdot y}$ then from the previous question we have:

$$\begin{aligned} \mathsf{LP}^f(0, b) &= \left(E\left(h(f(X)) \right) \right)^2 \\ &= \left(2^q E\left(g(Y)h(Y) \right) \right)^2 \\ &= 2^{2q} \left(E\left(g(Y)(-1)^{b \cdot Y} \right) \right)^2 \end{aligned}$$

5. Show that
$$g(y) = 2^{-q} \sum_{b \in B} (-1)^{b \cdot y} E\left((-1)^{b \cdot f(X)}\right)$$
where X is uniformly distributed in A.

$$\begin{aligned}
2^{-q} \sum_{b \in B} (-1)^{b \cdot y} E\left((-1)^{b \cdot f(X)}\right) &= \sum_{b \in B} (-1)^{b \cdot y} E\left(g(Y)(-1)^{bY}\right) \\
&= E\left(\sum_{b \in B} (-1)^{b \cdot y} g(Y)(-1)^{bY}\right) \\
&= E\left(g(Y) \sum_{b \in B} (-1)^{b \cdot (y+Y)}\right) \\
&= E\left(g(Y) 2^q 1_{Y=y}\right) \\
&= 2^q E\left(g(y) 1_{Y=y}\right) = g(y)
\end{aligned}$$

6. Deduce that
$$\forall b \neq 0 \quad \mathsf{LP}^f(0,b) = 0$$
if and only if $g(y) = 2^{-q}$ for all $y \in B$.

$$\begin{aligned}
\text{If } \forall b \neq 0 \quad \mathsf{LP}^f(0,b) = 0 \quad &\Rightarrow \quad \forall b \neq 0 \quad E\left((-1)^{bf(x)}\right) = 0 \text{ (from question 4)} \\
&\Rightarrow \quad g(y) = 2^{-q} \text{ for all } y \in B \text{ (from question 5)}
\end{aligned}$$

$$\begin{aligned}
\text{Conversely, if } \forall y \in B \quad g(y) = 2^{-q} \quad &\Rightarrow \quad \forall b \neq 0 \quad E\left(g(Y)(-1)^{bY}\right) = 0 \\
&\Rightarrow \quad \forall b \neq 0 \quad \mathsf{LP}^f(0,b) = 0 \text{ (from question 4)}
\end{aligned}$$

7. Deduce that

$$\forall b \neq 0 \quad \mathsf{LP}^f(0,b) = 0$$

if and only if f is balanced, i.e. all elements in B are equally taken as images by f.

$$\begin{aligned} \forall b \neq 0 \quad & \mathsf{LP}^f(0,b) = 0 \\ & \iff \forall y \in B \quad g(y) = \Pr[f(X) = y] = \frac{1}{2^q} \text{ (from question 6)} \\ & \iff \text{ All elements in } B \text{ are equally taken as images by } f \\ & \iff f \text{ is balanced.} \end{aligned}$$

3 Applied Crypto-polymorphism

The CONFIKER worm is permanently updating itself by looking for updates over the Internet. Once it has found the update, it checks if the update code has a correct RSA signature with modulus N and public exponent e. One problem is that the value of N in the code of the worm is large enough to be used by anti-virus software to detect the presence of the worm. The worm conceptor attended to a lecture on cryptography and would like to obfuscate N using cryptographic tricks.

1. Recall how the RSA signature verification works for a message m with signature σ.

   (Assume for example PKCS#1v1.5 with deterministic formatting rules for m.)

   RSA signature public key: (e, N)
   RSA signature secret key: (d, N)
   $\sigma = \mathsf{format}(m)^d \mod N$, where $\mathsf{format}(m) = 0100\mathsf{FF}\cdots\mathsf{FF}00\|h(m)$
   $\Rightarrow \sigma^e \mod N = \mathsf{format}m$

2. Once the worm installs, it picks a random prime number p, computes $N' = pN$ and discards p and N. The value of N' remains in the worm code. Show that a signature σ of an update code m can still be verified using e and N' instead of e and N.

   Can an anti-virus software detect the presence of the RSA key?

   The standard method for signature verification is to verify the following equality: $(\sigma^e - \mathsf{format}(m)) \mod N = 0$. However the value of N has been discarded.

   We know that:
   $\sigma^e \mod N = \mathsf{format}(m)$
   $\Rightarrow N \mid ((\sigma^e \mod N') - \mathsf{format}(m))$
   $\Rightarrow \mathsf{gcd}((\sigma^e \mod N') - \mathsf{format}(m), N') > 1$
   and this latter property is exceptionnal as N' only has 3 big prime factors.

3. Assume that the anti-virus software conceptor has analyzed the code of the worm on two independent infected machines and extracted N'_1 and N'_2. Show that he can deduce the value of N.

   With the value of N, show that we can still detect the presence of the worm based on the value of N' in the code. (Assume that N' can easily be extracted from the code.)

$N'_1 = p_1 N, \quad N'_2 = p_2 N$ If $\mathsf{gcd}(p_1, p_2) = 1 \Rightarrow \mathsf{gcd}(N'_1, N'_2) = N$

Detection is done by verifying the following equality: $\mathsf{gcd}(N', N) = N$.

4 Distinguishing Sources

We consider a source producing iid random variables $X_i \in \{0, 1, \ldots, 2^\ell - 1\}$ for $i = 1, \ldots, q$. For this, we consider two distributions:

- the uniform distribution P_0
- the distribution P_1 induced by $X_i = Y_i \bmod 2^\ell$ where Y_i is uniformly distributed in $\{0, 1, \ldots, p-1\}$ and $p > 2^\ell$. (Note that P_0 can be considered as a particular case of P_1 with $p = 2^\ell$.)

We assume that ℓ is large, e.g. $\ell \geq 80$ and we let $r = p \bmod 2^\ell$.

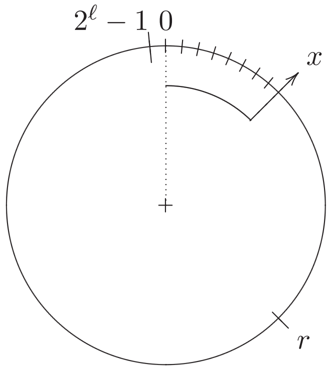
1. Given $x \in \{0, 1, \ldots, 2^\ell - 1\}$, show that

$$P_1(x) = \begin{cases} \left(1 - \frac{r}{p}\right) 2^{-\ell} + \frac{1}{p} & \text{if } x < r \\ \left(1 - \frac{r}{p}\right) 2^{-\ell} & \text{if } x \geq r. \end{cases}$$

P_0 uniform over $\{0, 1, \ldots, 2^\ell - 1\}$
$P_1 = ($ uniform over $\{0, 1, \ldots, p-1\})\ \bmod 2^\ell$

$p \bmod 2^\ell = r \Rightarrow p = r + n2^\ell \Longleftrightarrow n = (p - r)2^{-\ell}$

$$P_1(x) = \begin{cases} \frac{n+1}{p} = \left(1 - \frac{r}{p}\right) 2^{-\ell} + \frac{1}{p} & \text{if } x < r \\ \frac{n}{p} = \left(1 - \frac{r}{p}\right) 2^{-\ell} & \text{if } x \geq r. \end{cases}$$



2. Describe a distinguisher using $q = 1$ which achieves the optimal advantage.

Recall that $P_0(x) = 2^{-\ell}$

$$\frac{P_0(x)}{P_1(x)} = \begin{cases} < 1 & \text{if } x < r \\ \geq 1 & \text{if } x \geq r. \end{cases}$$

Hence if $x < r$ answer 1, else answer 0.

3. For $q = 1$, what is the best advantage for distinguishing P_0 from P_1? Express it as a formula in terms of ℓ, p, and r.

$$\begin{aligned}
\text{Adv} &= \frac{1}{2}\sum_x |P_0(x) - P_1(x)| \\
&= \frac{1}{2}\left(\sum_{x<r}\left|\frac{r}{p}2^{-\ell} - \frac{1}{p}\right| + \sum_{x\geqslant r}\left|\frac{r}{p}2^{-\ell}\right|\right) \\
&= \frac{1}{2p}\left(r(1 - r2^{-\ell}) + (2^\ell - r)r2^{-\ell}\right) \\
&= \frac{r}{2p}\left((1 - r2^{-\ell}) + (2^\ell - r)2^{-\ell}\right) \\
&= \frac{r}{2p}\left((1 - r2^{-\ell}) + (1 - r2^{-\ell})\right) \\
&= \frac{r}{p}\left(1 - r2^{-\ell}\right)
\end{aligned}$$

4. Deduce that for $p \leq c2^\ell$ with c small and $r2^{-\ell}$ neither too small nor too close to 1, then P_0 and P_1 can be distinguished using a single sample.

Adv$= \frac{r}{p}\left(1 - r2^{-\ell}\right) = \frac{2^\ell}{p} \cdot r2^{-\ell} \cdot \left(1 - r2^{-\ell}\right)$

As $\frac{2^\ell}{p} \geq \frac{1}{c}$ is non negligeable, and so are $r2^{-\ell}$ and $\left(1 - r2^{-\ell}\right)$, a single sample can be used to distinguish P_0 from P_1.

5. Describe a distinguisher using an arbitrarily fixed q which achieves the optimal advantage.

The best distinguisher using multiple samples (q samples) will have as parameter R the following:

$$R = \frac{2^{-q\ell}}{\left(\left(1-\frac{r}{p}\right)2^{-\ell}+\frac{1}{p}\right)^{\#\{i:x_i<r\}}\left(\left(1-\frac{r}{p}\right)2^{-\ell}\right)^{q-\#\{i:x_i<r\}}}$$

The best distinguisher will thus simply compare R and 1.

6. Compute the squared Euclidean distance between P_0 and P_1.

$$\begin{aligned}
\mathrm{SEI}(P_0,P_1) &= 2^\ell \sum_x |P_0(x)-P_1(x)|^2 \\
&= 2^\ell \left(\sum_{x<r}\left(\frac{r}{p}2^{-\ell}-\frac{1}{p}\right)^2 + \sum_{x\geqslant r}\left(\frac{r}{p}2^{-\ell}\right)^2\right) \\
&= 2^\ell \left(r\left(\frac{r}{p}2^{-\ell}\right)^2 - 2\frac{r^2}{p^2}2^{-\ell} + \frac{r}{p^2} + \left(2^\ell - r\right)\left(\frac{r}{p}2^{-\ell}\right)^2\right) \\
&= 2^\ell \left(\frac{-r^2}{p^2}2^{-\ell} + \frac{r}{p^2}\right) \\
&= 2^\ell \frac{r}{p^2}\left(1-r2^{-\ell}\right)
\end{aligned}$$

7. Assuming that P_1 is close to P_0, approximate the Chernoff information between P_0 and P_1. Deduce that $C(P_0, P_1) \leq \frac{2^\ell}{2p\ln 2}$ whatever r.

$$\begin{aligned} C(P_0, P_1) &= \frac{\mathrm{SEI}(P_0, P_1)}{8\ln 2} = \frac{1}{8\ln 2} \cdot \frac{r}{p^2} 2^\ell (1 - r2^{-\ell}) \\ &= \frac{2^\ell}{2p\ln 2} \cdot \frac{r}{4p} \cdot (1 - r2^{-\ell}) \\ &\leqslant \frac{2^\ell}{2p\ln 2} \end{aligned}$$

8. Deduce that for p larger than $2^{2\ell}$ the two distributions are indistinguishable in practice.

$$p > 2^{2\ell} \Rightarrow C(P_0, P_1) \leqslant \frac{2^\ell}{2p\ln 2} < \frac{2^\ell}{2 \cdot 2^{2\ell} \ln 2} \Rightarrow C(P_0, P_1) < \frac{1}{2^{\ell+1}\ln 2} < \frac{1}{2^\ell}$$

We need at least $2^{\ell+1}\ln 2$ samples wich is more than 2^ℓ samples.

Any attempt to look at the content of these pages before the signal will be severly punished.

Please be patient.