MTAT.07.003 Cryptology II
Spring 2010 / Exercise Session IV / Example Solution

Exercise. Let $f : \mathcal{M} \times \mathcal{K} \to \mathcal{M}$ be a (t, ε)-pseudorandom permutation and let CTR-$ be a symmetric encryption scheme defined as follows:

- A secret key is a randomly chosen $k \xleftarrow{u} \mathcal{K}$.
- To encrypt a message $m_1, \ldots, m_n$, choose a random nonce $s_0 \xleftarrow{u} \mathcal{M}$ and output a ciphertext vector $s_0, m_1 + f(s_0 + 1, k), \ldots, m_n + f(s_0 + n, k)$.
- To decrypt $s_0, c_1, \ldots, c_n$, output $c_1 - f(s_0 + 1, k), \ldots, c_n - f(s_0 + n, k)$.

Prove that CTR-$ is IND-CPA secure cryptosystem.

Solution. Recall that a symmetric cryptosystem $\mathcal{C} = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is (t, ε)-IND-CPA secure, if for any t-time adversary $\mathcal{A}$:

$$\mathsf{Adv}^{\mathsf{ind\text{-}cpa}}_{\mathcal{C}}(\mathcal{A}) = |\Pr[\mathcal{G}^{\mathcal{A}}_0 = 1] - \Pr[\mathcal{Q}^{\mathcal{A}}_1 = 1]| \leq \varepsilon$$

where the oracle $\mathcal{O}_1(\cdot)$ in the indistinguishability games

$\mathcal{Q}^{\mathcal{A}}_0$
$\lceil \mathsf{sk} \leftarrow \mathsf{Gen}$
$(m_0, m_1) \leftarrow \mathcal{A}^{\mathcal{O}_1(\cdot)}$
$\lfloor \textbf{return } \mathcal{A}^{\mathcal{O}_1(\cdot)}(\mathsf{Enc}_{\mathsf{sk}}(m_0))$

$\mathcal{Q}^{\mathcal{A}}_1$
$\lceil \mathsf{sk} \leftarrow \mathsf{Gen}$
$(m_0, m_1) \leftarrow \mathcal{A}^{\mathcal{O}_1(\cdot)}$
$\lfloor \textbf{return } \mathcal{A}^{\mathcal{O}_1(\cdot)}(\mathsf{Enc}_{\mathsf{sk}}(m_1))$

serves encryption calls. As the first step towards a solution, lets substitute the definition of CTR-$ encryption scheme into the games $\mathcal{Q}_0$ and $\mathcal{Q}_1$:

$\mathcal{Q}^{\mathcal{A}}_0$
$\lceil k \xleftarrow{u} \mathcal{K}$
$(\boldsymbol{m}_0, \boldsymbol{m}_0) \leftarrow \mathcal{A}^{\mathcal{O}_1(\cdot)}$
$s_0 \xleftarrow{u} \mathcal{M}$
For $i \in \{1, \ldots, n\}$ do
$\lfloor c_i \leftarrow m^i_0 + f(s_0 + i, k)$
$\lfloor \textbf{return } \mathcal{A}^{\mathcal{O}_1(\cdot)}(s_0, c_1, \ldots, c_n)$

$\mathcal{Q}^{\mathcal{A}}_1$
$\lceil k \xleftarrow{u} \mathcal{K}$
$(\boldsymbol{m}_0, \boldsymbol{m}_1) \leftarrow \mathcal{A}^{\mathcal{O}_1(\cdot)}$
$s_0 \xleftarrow{u} \mathcal{M}$
For $i \in \{1, \ldots, n\}$ do
$\lfloor c_i \leftarrow m^i_0 + f(s_0 + i, k)$
$\lfloor \textbf{return } \mathcal{A}^{\mathcal{O}_1(\cdot)}(s_0, c_1, \ldots, c_n)$

As the second step, we can replace all invocations of a pseudorandom function $f_k(x) = f(x, k)$ with a truly random function. Note that the function f is invoked not only in the explicit encryption call but also in the oracle $\mathcal{O}_1(\cdot)$. Let $\mathcal{O}_f(\cdot)$ denote the new encryption oracle with the following construction

$\mathcal{O}_f(m_1, \ldots, m_n)$
$\lceil \overline{s}_0 \xleftarrow{u} \mathcal{M}$
For $i \in \{1, \ldots, n\}$ do
$\lfloor \overline{c}_i \leftarrow m^i_0 + f(s_0 + i)$
$\lfloor \textbf{return } (\overline{s}_0, \overline{c}_1, \ldots, \overline{c}_n)$

and let $\mathcal{G}_0$ and $\mathcal{G}_1$ denote the resulting games

$$
\begin{array}{ll}
\mathcal{G}_0^{\mathcal{A}} & \mathcal{G}_1^{\mathcal{A}} \\
\left[\begin{array}{l}
f \xleftarrow{u} \mathcal{F}_{\text{all}} \\
(\boldsymbol{m}_0, \boldsymbol{m}_0) \leftarrow \mathcal{A}^{\mathcal{O}_f(\cdot)} \\
s_0 \xleftarrow{u} \mathcal{M} \\
\quad \text{For } i \in \{1, \ldots, n\} \text{ do} \\
\quad \left[\, c_i \leftarrow m_0^i + f(s_0 + i) \right. \\
\textbf{return } \mathcal{A}^{\mathcal{O}_f(\cdot)}(s_0, c_1, \ldots, c_n)
\end{array}\right. &
\left[\begin{array}{l}
f \xleftarrow{u} \mathcal{F}_{\text{all}} \\
(\boldsymbol{m}_0, \boldsymbol{m}_1) \leftarrow \mathcal{A}^{\mathcal{O}_f(\cdot)} \\
s_0 \xleftarrow{u} \mathcal{M} \\
\quad \text{For } i \in \{1, \ldots, n\} \text{ do} \\
\quad \left[\, c_i \leftarrow m_0^i + f(s_0 + i) \right. \\
\textbf{return } \mathcal{A}^{\mathcal{O}_f(\cdot)}(s_0, c_1, \ldots, c_n)
\end{array}\right.
\end{array}
$$

As the function f is (t, ε)-pseudorandom function, we can easily prove that

$$\left|\Pr\left[\mathcal{Q}_0^{\mathcal{A}} = 1\right] - \Pr\left[\mathcal{G}_0^{\mathcal{A}} = 1\right]\right| \leq \varepsilon \tag{1}$$

$$\left|\Pr\left[\mathcal{Q}_1^{\mathcal{A}} = 1\right] - \Pr\left[\mathcal{G}_1^{\mathcal{A}} = 1\right]\right| \leq \varepsilon \tag{2}$$

for all adversaries $\mathcal{A}$ with sufficiently small running time. Indeed, let $\mathcal{Q}_2$ and $\mathcal{Q}_3$ be the indistinguishability games for pseudorandom functions:

$$
\begin{array}{ll}
\mathcal{Q}_2^{\mathcal{B}} & \mathcal{Q}_3^{\mathcal{B}} \\
\left[\begin{array}{l}
f \leftarrow \mathcal{F}_{\text{all}} \\
\textbf{return } \mathcal{B}^{f(\cdot)}
\end{array}\right. &
\left[\begin{array}{l}
f \leftarrow \mathcal{F} \\
\textbf{return } \mathcal{B}^{f(\cdot)}
\end{array}\right.
\end{array}
$$

Then we can define adversaries $\mathcal{B}_0$ and $\mathcal{B}_1$ by in-lining the common parts of game pairs $\mathcal{Q}_0$ and $\mathcal{G}_0$ and $\mathcal{Q}_1$ and $\mathcal{G}_1$ into the adversary construction

$$
\begin{array}{ll}
\mathcal{B}_0^{f(\cdot)} & \mathcal{B}_1^{f(\cdot)} \\
\left[\begin{array}{l}
(\boldsymbol{m}_0, \boldsymbol{m}_0) \leftarrow \mathcal{A}^{\mathcal{O}_f(\cdot)} \\
s_0 \xleftarrow{u} \mathcal{M} \\
\quad \text{For } i \in \{1, \ldots, n\} \text{ do} \\
\quad \left[\, c_i \leftarrow m_0^i + f(s_0 + i) \right. \\
\textbf{return } \mathcal{A}^{\mathcal{O}_f(\cdot)}(s_0, c_1, \ldots, c_n)
\end{array}\right. &
\left[\begin{array}{l}
(\boldsymbol{m}_0, \boldsymbol{m}_1) \leftarrow \mathcal{A}^{\mathcal{O}_f(\cdot)} \\
s_0 \xleftarrow{u} \mathcal{M} \\
\quad \text{For } i \in \{1, \ldots, n\} \text{ do} \\
\quad \left[\, c_i \leftarrow m_1^i + f(s_0 + i) \right. \\
\textbf{return } \mathcal{A}^{\mathcal{O}_f(\cdot)}(s_0, c_1, \ldots, c_n)
\end{array}\right.
\end{array}
$$

As a result, it is straightforward to prove that $\mathcal{Q}_2^{\mathcal{B}_i} \equiv \mathcal{Q}_i^{\mathcal{A}}$ and $\mathcal{Q}_3^{\mathcal{B}_i} \equiv \mathcal{G}_i^{\mathcal{A}}$. Hence, the equations (1) and (2) must hold as soon as the running times of $\mathcal{B}_0$ and $\mathcal{B}_1$ are less or equal to t. Unfortunately, these running times depend on the number of encryption queries. Let q be the number of oracle calls to $\mathcal{O}_f$ made by $\mathcal{A}$ and t_f be the time penalty of a single call to the oracle $f(\cdot)$. Then the running time of $\mathcal{B}_i$ is $t_{\mathcal{A}} + nq \cdot t_f + \mathrm{O}(n)$ and consequently the running time of $\mathcal{A}$ must be below $t - nq \cdot t_f - \mathrm{O}(n)$ for the equations (1) and (2) to hold.

For the further analysis, there are two possible options. The most straightforward way is to substitute also the encryption oracle $\mathcal{O}_f(\cdot)$ into the games according to its specification. However, this would produce rather lengthy games

that are difficult to analyse by hand. A more elegant solution is based on *horizon splitting*. Let Coll denote the event that a range $[s_0+1,\ldots,s_0+n]$ for the challenge encryption overlaps with range $[\overline{s}_0+1,\ldots,\overline{s}_0+n]$ for another encryption generated by the oracle $\mathcal{O}_f(\cdot)$ during the game. Then obviously

$$\begin{aligned}\Pr\left[\mathcal{G}_0^{\mathcal{A}}=1\right] &= \Pr\left[\mathcal{G}_0^{\mathcal{A}}=1\wedge\mathsf{Coll}\right]+\Pr\left[\mathcal{G}_0^{\mathcal{A}}=1\wedge\neg\mathsf{Coll}\right] \quad ,\\ \Pr\left[\mathcal{G}_1^{\mathcal{A}}=1\right] &= \Pr\left[\mathcal{G}_1^{\mathcal{A}}=1\wedge\mathsf{Coll}\right]+\Pr\left[\mathcal{G}_1^{\mathcal{A}}=1\wedge\neg\mathsf{Coll}\right] \quad ,\end{aligned}$$

and by triangle inequality

$$\begin{aligned}\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{G}_0,\mathcal{G}_1}(\mathcal{A}) \quad\leq\quad & \left|\Pr\left[\mathcal{G}_0^{\mathcal{A}}=1\wedge\mathsf{Coll}\right]-\Pr\left[\mathcal{G}_1^{\mathcal{A}}=1\wedge\mathsf{Coll}\right]\right| \\ & +\left|\Pr\left[\mathcal{G}_0^{\mathcal{A}}=1\wedge\neg\mathsf{Coll}\right]-\Pr\left[\mathcal{G}_1^{\mathcal{A}}=1\wedge\neg\mathsf{Coll}\right]\right| \quad .\end{aligned}$$

Hence, we can estimate the advantage $\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{G}_0,\mathcal{G}_1}(\mathcal{A})$ in terms of four new games

$\mathcal{G}_2^{\mathcal{A}}$

$f \leftarrow_{\!u} \mathcal{F}_{\mathrm{all}}$
$(\boldsymbol{m}_0,\boldsymbol{m}_0) \leftarrow \mathcal{A}^{\mathcal{O}_f(\cdot)}$
$s_0 \leftarrow_{\!u} \mathcal{M}$
For $i \in \{1,\ldots,n\}$ do
  $c_i \leftarrow m_0^i + f(s_0+i)$
if $\neg\mathsf{Coll}$ then **return** $\bot$
return $\mathcal{A}^{\mathcal{O}_f(\cdot)}(s_0,c_1,\ldots,c_n)$

$\mathcal{G}_3^{\mathcal{A}}$

$f \leftarrow_{\!u} \mathcal{F}_{\mathrm{all}}$
$(\boldsymbol{m}_0,\boldsymbol{m}_1) \leftarrow \mathcal{A}^{\mathcal{O}_f(\cdot)}$
$s_0 \leftarrow_{\!u} \mathcal{M}$
For $i \in \{1,\ldots,n\}$ do
  $c_i \leftarrow m_0^i + f(s_0+i)$
if $\neg\mathsf{Coll}$ then **return** $\bot$
return $\mathcal{A}^{\mathcal{O}_1(\cdot)}(s_0,c_1,\ldots,c_n)$

$\mathcal{G}_4^{\mathcal{A}}$

$f \leftarrow_{\!u} \mathcal{F}_{\mathrm{all}}$
$(\boldsymbol{m}_0,\boldsymbol{m}_0) \leftarrow \mathcal{A}^{\mathcal{O}_f(\cdot)}$
$s_0 \leftarrow_{\!u} \mathcal{M}$
For $i \in \{1,\ldots,n\}$ do
  $c_i \leftarrow m_0^i + f(s_0+i)$
if $\neg\mathsf{Coll}$ then **return** $\bot$
return $\mathcal{A}^{\mathcal{O}_f(\cdot)}(s_0,c_1,\ldots,c_n)$

$\mathcal{G}_5^{\mathcal{A}}$

$f \leftarrow_{\!u} \mathcal{F}_{\mathrm{all}}$
$(\boldsymbol{m}_0,\boldsymbol{m}_1) \leftarrow \mathcal{A}^{\mathcal{O}_f(\cdot)}$
$s_0 \leftarrow_{\!u} \mathcal{M}$
For $i \in \{1,\ldots,n\}$ do
  $c_i \leftarrow m_0^i + f(s_0+i)$
if $\neg\mathsf{Coll}$ then **return** $\bot$
return $\mathcal{A}^{\mathcal{O}_f(\cdot)}(s_0,c_1,\ldots,c_n)$

Lets analyse the games $\mathcal{G}_4$ and $\mathcal{G}_5$ first. Since both games end with $\bot$ when a collision occurs, the invocations of $f(s_0+i)$ can be replaced with uniform sampling $y_i \leftarrow_{\!u} \mathcal{M}$. The cases where this would be detectable are guaranteed to end with $\bot$ anyway. As $m_i + y_i$ for $y_i \leftarrow_{\!u} \mathcal{M}$ has a uniform distribution, we can further simplify the games without changing their semantics. Let $\mathcal{G}_6$ and

$\mathcal{G}_7$ denote the resulting games

$\mathcal{G}_6^{\mathcal{A}}$

$f \leftarrow_u \mathcal{F}_{\text{all}}$
$(\boldsymbol{m}_0, \boldsymbol{m}_0) \leftarrow \mathcal{A}^{\mathcal{O}_f(\cdot)}$
$s_0 \leftarrow_u \mathcal{M}$
For $i \in \{1, \ldots, n\}$ do
  $c_i \leftarrow_u \mathcal{M}$
if $\neg$Coll then **return** $\bot$
return $\mathcal{A}^{\mathcal{O}_f(\cdot)}(s_0, c_1, \ldots, c_n)$

$\mathcal{G}_7^{\mathcal{A}}$

$f \leftarrow_u \mathcal{F}_{\text{all}}$
$(\boldsymbol{m}_0, \boldsymbol{m}_1) \leftarrow \mathcal{A}^{\mathcal{O}_f(\cdot)}$
$s_0 \leftarrow_u \mathcal{M}$
For $i \in \{1, \ldots, n\}$ do
  $c_i \leftarrow_u \mathcal{M}$
if $\neg$Coll then **return** $\bot$
return $\mathcal{A}^{\mathcal{O}_f(\cdot)}(s_0, c_1, \ldots, c_n)$

Then by the reasoning given above $\mathcal{G}_4^{\mathcal{A}} \equiv \mathcal{G}_6^{\mathcal{A}} = \mathcal{G}_7^{\mathcal{A}} \equiv \mathcal{G}_5^{\mathcal{A}}$ and consequently

$$\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{G}_0,\mathcal{G}_1}(\mathcal{A}) \leq \left|\Pr\left[\mathcal{G}_0^{\mathcal{A}} = 1 \wedge \mathsf{Coll}\right] - \Pr\left[\mathcal{G}_1^{\mathcal{A}} = 1 \wedge \mathsf{Coll}\right]\right| = \mathsf{Adv}^{\mathsf{ind}}_{\mathcal{G}_2,\mathcal{G}_3}(\mathcal{A}) \ .$$

Note that if ranges $[s_0+1, \ldots, s_0+n]$ and $[\overline{s}_0+1, \ldots, \overline{s}_0+n]$ overlap at a position $s_0 + i$, the value $f(s_0 + i)$ becomes public and $\mathcal{A}$ can easily detect whether $\boldsymbol{m}_0$ or $\boldsymbol{m}_1$ was encrypted as long as $m_0^i \neq m_1^i$. Hence, we use only a trivial bound

$$\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{G}_2,\mathcal{G}_3}(\mathcal{A}) \leq \max\left\{\Pr\left[\mathsf{Coll} \text{ in } \mathcal{G}_0\right], \Pr\left[\mathsf{Coll} \text{ in } \mathcal{G}_1\right]\right\} \ .$$

Fortunately, the collision probability depends only on the number of oracle calls. For a fixed range $[s_0 + 1, \ldots, s_0 + n]$ and uniformly chosen $\overline{s}_0$, the probability of an overlap is $\frac{2n-1}{|\mathcal{M}|}$. As $\mathcal{A}$ makes q encryption calls, the union bound yields

$$\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{G}_2,\mathcal{G}_3}(\mathcal{A}) \leq \frac{q(2n-1)}{|\mathcal{M}|} \ .$$

To summarise, we have established that the IND-CPA advantage of an adversary $\mathcal{A}$ with a running time $t - nq \cdot t_f - \mathrm{O}(n)$ can be bounded:

$$\mathsf{Adv}^{\mathsf{ind\text{-}cpa}}_{\text{CTR-\$}}(\mathcal{A}) \leq \frac{q(2n-1)}{|\mathcal{M}|} + 2\varepsilon$$

where q is the upper bound on the encryption queries. This result is precise but contains unknown value q. Since $q \leq t_{\mathcal{A}}$ for obvious reasons, the inequality $t_{\mathcal{A}} \leq t - nt_{\mathcal{A}} \cdot t_f - \mathrm{O}(n)$ about running time implies

$$t_{\mathcal{A}} \leq \frac{t - \mathrm{O}(n)}{1 + nt_f}$$

and consequently

$$\mathsf{Adv}^{\mathsf{ind\text{-}cpa}}_{\text{CTR-\$}}(\mathcal{A}) \leq \frac{t - \mathrm{O}(n)}{1 + nt_f} \cdot \frac{(2n-1)}{|\mathcal{M}|} + 2\varepsilon \ .$$

Remark. The horizon splitting technique is rather powerful and allows us to analyse many settings by splitting the proof into different branches. It can be easily generalised for estimating the success against a single game, as well.