MTAT.07.003 Cryptology II
Spring 2012 / Exercise session ?? / Example Solution

Exercise (CBC mode as a pseudorandom function). *Let $\mathcal{F}$ be (t,ℓ,ε)-secure pseudorandom function with the domain $\{0,1\}^n$. Show that a single outcome of Cipher Block Chaining mode of operation*

$$c_1 = f(c_0 \oplus m_1),\ c_2 = f(c_1 \oplus m_2),\ \ldots,\ c_\ell = f(c_{\ell-1} \oplus m_\ell)$$

is indistinguishable from a random string even $c_0, m_1, \ldots, m_\ell$ are known or chosen by the adversary. Generalise the proof for the case when the adversary can query many inputs of the same size. More precisely, treat the mode as a randomised function $\textsc{Cbc}_f(m_1,\ldots,m_\ell) = (c_0, c_1, \ldots, c_\ell)$ for $c_0 \xleftarrow{u} \{0,1\}^n$. Show that $\textsc{Cbc}_f$ is a pseudorandom function if one considers adversaries that query the same argument $(m_1,\ldots,m_\ell)$ only once.

Solution.

Indistinguishability for a single Cbc query. Let us first correctly model the problem statement. Obviously, we need two games where the adversary $\mathcal{A}$ first specifies $c_0, m_1, \ldots, m_\ell$, then obtains $c_1, \ldots, c_\ell$ and finally makes its decision. Note that in both games $\mathcal{A}$ cannot have oracle access to the function f or otherwise it could redo the computations and directly verify whether $c_i = f(c_{i-1} \oplus m_i)$ or not. Consequently, the right formalisation of the problem is

$\mathcal{G}_0$
$f \leftarrow \mathcal{F}$
$(c_0, m_1, \ldots, m_\ell) \leftarrow \mathcal{A}$
For $i \in \{1,\ldots,\ell\}$ do
$\quad c_i \leftarrow \{0,1\}^n$
return $\mathcal{A}(c_1, c_2, \ldots, c_\ell)$

$\mathcal{G}_1$
$f \leftarrow \mathcal{F}$
$(c_0, m_1, \ldots, m_\ell) \leftarrow \mathcal{A}$
For $i \in \{1,\ldots,\ell\}$ do
$\quad c_i \leftarrow f(c_{i-1} \oplus m_i)$
return $\mathcal{A}(c_1, c_2, \ldots, c_\ell)$

where the first game models random reply and the second game models behaviour of $\textsc{Cbc}_f$.

As the first step, we replace function family $\mathcal{F}$ with the family of all functions $\mathcal{F}_{\text{all}}$. Recall that a function family $\mathcal{F}$ is a (t,ℓ,ε)-secure pseudorandom function family if for any t-time adversary $\mathcal{B}$ games

$\mathcal{Q}_0$
$f \leftarrow \mathcal{F}_{\text{all}}$
return $\mathcal{B}^f$

$\mathcal{Q}_1$
$f \leftarrow \mathcal{F}$
return $\mathcal{B}^f$

are (t,ε)-indistinguishable if $\mathcal{B}$ makes at most ℓ queries to f. Now if we plug the following adversary

$\mathcal{B}^f$
$(c_0, m_1, \ldots, m_\ell) \leftarrow \mathcal{A}$
For $i \in \{1,\ldots,\ell\}$ do
$\quad c_i \leftarrow f(c_{i-1} \oplus m_i)$
return $\mathcal{A}(c_1, c_2, \ldots, c_\ell)$.

into the $\mathcal{Q}_1$ then we get the game $\mathcal{G}_1$ where $\mathcal{B}$ makes ℓ queries. If we plug $\mathcal{B}$ into the game $\mathcal{Q}_0$ then we get a game $\mathcal{G}_2$ where the function family $\mathcal{F}$ is replaced

$\mathcal{G}_2$
$f \leftarrow \mathcal{F}_{\text{all}}$
$(c_0, m_1, \ldots, m_\ell) \leftarrow \mathcal{A}$
For $i \in \{1,\ldots,\ell\}$ do
$\quad c_i \leftarrow f(c_{i-1} \oplus m_i)$
return $\mathcal{A}(c_1, c_2, \ldots, c_\ell)$.

Therefore, we obtain a bound

$$|\Pr[\mathcal{G}_1^{\mathcal{A}} = 1] - \Pr[\mathcal{G}_2^{\mathcal{A}} = 1]| \leq \varepsilon$$

provided that $\mathcal{A}$ is $(t - c\ell)$-time adversary where the constant $c > 1$ captures the overhead in the reduction.

For a conceptual clarity, we will unroll the for loop and try to use hybrid argument to bound the maximal success in the game $\mathcal{G}_2$. Let us first define the sequence of following

$$\begin{array}{lll}
\mathcal{G}_2 & \mathcal{G}_3 & \mathcal{G}_{2+\ell} \\
\left[\begin{array}{l} f \leftarrow \mathcal{F}_{\text{all}} \\ (c_0, m_1, \ldots, m_\ell) \leftarrow \mathcal{A} \\ c_1 \leftarrow f(c_0 \oplus m_1) \\ c_2 \leftarrow f(c_1 \oplus m_2) \\ \ldots \\ c_{\ell-1} \leftarrow f(c_{\ell-2} \oplus m_{\ell-1}) \\ c_\ell \leftarrow f(c_{\ell-1} \oplus m_\ell) \\ \textbf{return } \mathcal{A}(c_1, c_2, \ldots, c_\ell) \end{array}\right. & \left[\begin{array}{l} f \leftarrow \mathcal{F}_{\text{all}} \\ (c_0, m_1, \ldots, m_\ell) \leftarrow \mathcal{A} \\ c_1 \leftarrow f(c_0 \oplus m_1) \\ c_2 \leftarrow f(c_1 \oplus m_2) \\ \ldots \\ c_{\ell-1} \leftarrow f(c_{\ell-2} \oplus m_{\ell-1}) \\ c_\ell \leftarrow_u \{0,1\}^n \\ \textbf{return } \mathcal{A}(c_1, c_2, \ldots, c_\ell) \end{array}\right. & \cdots \quad \left[\begin{array}{l} f \leftarrow \mathcal{F}_{\text{all}} \\ (c_0, m_1, \ldots, m_\ell) \leftarrow \mathcal{A} \\ c_1 \leftarrow_u \{0,1\}^n \\ c_2 \leftarrow_u \{0,1\}^n \\ \ldots \\ c_{\ell-1} \leftarrow_u \{0,1\}^n \\ c_\ell \leftarrow_u \{0,1\}^n \\ \textbf{return } \mathcal{A}(c_1, c_2, \ldots, c_\ell) \end{array}\right. .
\end{array}$$

where each game replaces a function call with a random sample. Clearly, the only difference in games $\mathcal{G}_2$ and $\mathcal{G}_3$ is how c_ℓ is computed: c_ℓ is a function call $f(c_{\ell-1} \oplus m_\ell)$ in $\mathcal{G}_2$ and a uniform choice over $\{0,1\}^n$ in $\mathcal{G}_3$. As f is chosen uniformly form all functions $\mathcal{F}_{\text{all}}$, the value of a function $f(x)$ is uniformly distributed and independent form all other inputs. However, the input $c_{\ell-1} \oplus m_\ell$ might be already computed as

$$c_0 \oplus m_1, \quad c_1 \oplus m_2, \quad \ldots, \quad c_{\ell-1} \oplus m_{\ell-1} \ .$$

In this case, replacing $f(c_{\ell-1} \oplus m_\ell)$ with a uniformly chosen value alters the game and thus might influence the outcome of the game. Consequently, we need to estimate the probability of collision

$$\Pr\left[\begin{array}{l} f \leftarrow \mathcal{F}_{\text{all}}, c_1 \leftarrow f(c_0 \oplus m_1), c_2 \leftarrow f(c_1 \oplus m_2), \ldots, c_{\ell-1} \leftarrow f(c_{\ell-2} \oplus m_{\ell-1}) : \\ c_0 \oplus m_1 = c_{\ell-1} \oplus m_\ell \vee c_1 \oplus m_2 = c_{\ell-1} \oplus m_\ell \vee \ldots \vee c_{\ell-2} \oplus m_{\ell-1} = c_{\ell-1} \oplus m_\ell \end{array}\right]$$

which is technically hard as we must consider many function calls at once. Let us therefore abandon the initial attempt and define the sequence differently so that the modifications start from the beginning:

$$\begin{array}{lll}
\mathcal{G}_2 & \mathcal{G}_3 & \mathcal{G}_{2+\ell} \\
\left[\begin{array}{l} f \leftarrow \mathcal{F}_{\text{all}} \\ (c_0, m_1, \ldots, m_\ell) \leftarrow \mathcal{A} \\ c_1 \leftarrow f(c_0 \oplus m_1) \\ c_2 \leftarrow f(c_1 \oplus m_2) \\ \ldots \\ c_{\ell-1} \leftarrow f(c_{\ell-2} \oplus m_{\ell-1}) \\ c_\ell \leftarrow f(c_{\ell-1} \oplus m_\ell) \\ \textbf{return } \mathcal{A}(c_1, c_2, \ldots, c_\ell) \end{array}\right. & \left[\begin{array}{l} f \leftarrow \mathcal{F}_{\text{all}} \\ (c_0, m_1, \ldots, m_\ell) \leftarrow \mathcal{A} \\ c_1 \leftarrow_u \{0,1\}^n \\ c_2 \leftarrow f(c_1 \oplus m_2) \\ \ldots \\ c_{\ell-1} \leftarrow f(c_{\ell-2} \oplus m_{\ell-1}) \\ c_\ell \leftarrow f(c_{\ell-1} \oplus m_\ell) \\ \textbf{return } \mathcal{A}(c_1, c_2, \ldots, c_\ell) \end{array}\right. & \cdots \quad \left[\begin{array}{l} f \leftarrow \mathcal{F}_{\text{all}} \\ (c_0, m_1, \ldots, m_\ell) \leftarrow \mathcal{A} \\ c_1 \leftarrow_u \{0,1\}^n \\ c_2 \leftarrow_u \{0,1\}^n \\ \ldots \\ c_{\ell-1} \leftarrow_u \{0,1\}^n \\ c_\ell \leftarrow_u \{0,1\}^n \\ \textbf{return } \mathcal{A}(c_1, c_2, \ldots, c_\ell) \end{array}\right. .
\end{array}$$

Again, the only difference in games $\mathcal{G}_2$ and $\mathcal{G}_3$ is how c_1 is computed: c_1 is a function call $f(c_0 \oplus m_1)$ in $\mathcal{G}_2$ and a uniform choice over $\{0,1\}^n$ in $\mathcal{G}_3$. Note that the replacement is correct only if the subsequent function calls are different from $c_0 \oplus m_1$ and thus we must estimate the probability of a collision

$$\Pr\left[\begin{array}{l} f \leftarrow \mathcal{F}_{\text{all}}, c_1 \leftarrow f(c_0 \oplus m_1), c_2 \leftarrow f(c_1 \oplus m_2) \ldots, c_\ell \leftarrow f(c_{\ell-1} \oplus m_\ell) : \\ c_0 \oplus m_1 = c_1 \oplus m_2 \vee c_0 \oplus m_1 = c_2 \oplus m_3 \vee \ldots \vee c_0 \oplus m_1 = c_{\ell-1} \oplus m_\ell \end{array}\right] \ .$$

Again, this is technically quite difficult as we must consider many function calls at once.

As the final attempt, let us model function calls without errors. This leads to the following definition

$\mathcal{G}_3$

$f \leftarrow \mathcal{F}_{\text{all}}$
$(c_0, m_1, \ldots, m_\ell) \leftarrow \mathcal{A}$
$c_1 \xleftarrow{u} \{0,1\}^n$
if $c_0 \oplus m_1 = c_1 \oplus m_2$ then $c_2 \leftarrow c_1$
else $c_2 \xleftarrow{u} \{0,1\}^n$
if $c_0 \oplus m_1 = c_2 \oplus m_3$ then $c_3 \leftarrow c_1$
else if$c_1 \oplus m_2 = c_2 \oplus m_3$ then $c_3 \leftarrow c_2$
else $c_3 \xleftarrow{u} \{0,1\}^n$
$\ldots$
$\ldots$
if $c_0 \oplus m_1 = c_{\ell-1} \oplus m_1$ then $c_\ell \leftarrow c_1$
$\ldots$
else $c_\ell \xleftarrow{u} \{0,1\}^n$
return $\mathcal{A}(c_1, c_2, \ldots, c_\ell)$

where each function call is replaced with an if-else block that checks if the argument $c_{i-1} \oplus m_i$ has not been queried before and if such a query was made then provides a consistent answer. As a result $\mathcal{G}_2$ and $\mathcal{G}_3$ are identical. Now we can start to simplify $\mathcal{G}_3$ by omitting the effect of if branches. The effect of the first change

$\mathcal{G}_3$

$f \leftarrow \mathcal{F}_{\text{all}}$
$(c_0, m_1, \ldots, m_\ell) \leftarrow \mathcal{A}$
$c_1 \xleftarrow{u} \{0,1\}^n$
if $c_0 \oplus m_1 = c_1 \oplus m_2$ then $c_2 \leftarrow c_1$
else $c_2 \xleftarrow{u} \{0,1\}^n$
if $c_0 \oplus m_1 = c_2 \oplus m_3$ then $c_3 \leftarrow c_1$
else if$c_1 \oplus m_2 = c_2 \oplus m_3$ then $c_3 \leftarrow c_2$
else $c_3 \xleftarrow{u} \{0,1\}^n$
$\ldots$
$\ldots$
if $c_0 \oplus m_1 = c_{\ell-1} \oplus m_1$ then $c_\ell \leftarrow c_1$
$\ldots$
else $c_\ell \xleftarrow{u} \{0,1\}^n$
return $\mathcal{A}(c_1, c_2, \ldots, c_\ell)$

$\mathcal{G}_4$

$f \leftarrow \mathcal{F}_{\text{all}}$
$(c_0, m_1, \ldots, m_\ell) \leftarrow \mathcal{A}$
$c_1 \xleftarrow{u} \{0,1\}^n$
if $c_0 \oplus m_1 = c_1 \oplus m_2$ then $c_2 \xleftarrow{u} \{0,1\}^n$
else $c_2 \xleftarrow{u} \{0,1\}^n$
if $c_0 \oplus m_1 = c_2 \oplus m_3$ then $c_3 \leftarrow c_1$
else if$c_1 \oplus m_2 = c_2 \oplus m_3$ then $c_3 \leftarrow c_2$
else $c_3 \xleftarrow{u} \{0,1\}^n$
$\ldots$
$\ldots$
if $c_0 \oplus m_1 = c_{\ell-1} \oplus m_1$ then $c_\ell \leftarrow c_1$
$\ldots$
else $c_\ell \xleftarrow{u} \{0,1\}^n$
return $\mathcal{A}(c_1, c_2, \ldots, c_\ell)$

is easy to describe. The first if branch is selected only if a collision $c_0 \oplus m_1 = c_1 \oplus m_2$ occurs. As c_1 is uniformly chosen this collision occurs with probability 2^{-n} and we obtain a nice bound

$$|\Pr[\mathcal{G}_3^{\mathcal{A}} = 1] - \Pr[\mathcal{G}_4^{\mathcal{A}} = 1]| \leq 2^{-n} \ .$$

In the modified game $\mathcal{G}_4$ the variable c_2 is uniformly chosen and thus the probability that one of the conditions $c_0 \oplus m_1 = c_2 \oplus m_3$ and $c_1 \oplus m_2 = c_2 \oplus m_3$ is satisfied is $2 \cdot 2^{-n}$ and thus defining $c_3 \xleftarrow{u} \{0,1\}^n$ alters the outcome of the game only by $2 \cdot 2^{-n}$. By repeating the simplification procedure, we indeed end up in the game were $c_1, \ldots, c_\ell$ are uniformly sampled and this costs us

$$|\Pr[\mathcal{G}_3^{\mathcal{A}} = 1] - \Pr[\mathcal{G}_{3+\ell}^{\mathcal{A}} = 1]| \leq 2^{-n} + 2 \cdot 2^{-n} \cdots + (\ell - 1) \cdot 2^{-n} = \frac{\ell(\ell-1)}{2^{n+1}} \ .$$

Since the last game $\mathcal{G}_{3+\ell}$ is identical to $\mathcal{G}_0$, we have proven

$$|\Pr[\mathcal{G}_0^{\mathcal{A}} = 1] - \Pr[\mathcal{G}_1^{\mathcal{A}} = 1]| \leq |\Pr[\mathcal{G}_0^{\mathcal{A}} = 1] - \Pr[\mathcal{G}_2^{\mathcal{A}} = 1]| + |\Pr[\mathcal{G}_2^{\mathcal{A}} = 1] - \Pr[\mathcal{G}_1^{\mathcal{A}} = 1]| \leq \frac{\ell(\ell-1)}{2^{n+1}} + \varepsilon \ .$$

Alternatively, we could have estimated directly the probability of a collision event

$$\Pr[f \leftarrow \mathcal{F}_{\text{all}} : \exists i \neq j : \quad c_{i-1} \oplus m_i = c_{j-1} \oplus m_j] \ .$$

However, this would have been tractable only by considering all function calls simultaneously and decomposing the formula into tree of comparison events. The latter together with over-estimations would have been equivalent to the gradual nullification of collision cases described in the sequence of games $\mathcal{G}_3, \mathcal{G}_4, \ldots, \mathcal{G}_{3+\ell}$.

Generalisation to many Cbc queries. **To be completed!**