Computer Science and Engineering, UCSD — Fall 09
CSE 207: Modern Cryptography — **Instructor:** Mihir Bellare
Problem Set 3 Solutions — October 19, 2009

Problem Set 3 Solutions

Problem 1. [100 points] Let $E\colon \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^n$ be a block cipher and let algorithm $\mathcal{K}$ return $K \xleftarrow{\$} \{0,1\}^k$. Assume messages to be encrypted have length $\ell < n$. Let $\mathcal{E}$ be the following encryption algorithm:

```
algorithm $\mathcal{E}_K(M)$
    if $|M| \neq \ell$ then return $\perp$    // Only encrypts $\ell$-bit messages
    $R \xleftarrow{\$} \{0,1\}^{n-\ell}$
    $C \leftarrow E_K(R \,\|\, M)$
    return $C$
```

Above, "$x \,\|\, y$" denotes the concatenation of strings x and y.

1. **[10 points]** Specify a decryption algorithm $\mathcal{D}$ such that $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ is a symmetric encryption scheme providing correct decryption.

We use the fact that E is a block cipher and thus given the key one can easily compute its inverse E^{-1}. Given a n-bit string C, the decryption algorithm is then as follows:

```
algorithm $\mathcal{D}_K(C)$
    $X \leftarrow E_K^{-1}(C)$
    $M \leftarrow X[n-\ell+1..n]$
    return $M$
```

Above $X[a..b]$ means bits a through b of string X.

2. **[40 points]** Give the best attack you can on this scheme. Given an even number q, your attack should take the form of an ind-cpa adversary A that makes q oracle queries and has running time around that for $O(q)$ applications of E. Specify $\mathbf{Adv}^{\text{ind-cpa}}_{\mathcal{SE}}(A)$ as a function of q, n, ℓ. Letting $n = 128$, make a table showing, for values $\ell = 1, 16, 32, 64, 96$, the smallest value of q for which the advantage is at least 1/4. (The better the attack, the more points you get.) For the analysis, you may find Lemma A.1 below useful.

Based on attacks in class, one might propose the following adversary, where q is an integer parameter:

```
adversary $A$
for $i = 1, \ldots, q$ do $C_i \xleftarrow{\$} \mathbf{LR}(\langle i \rangle, 0^\ell)$
if $\exists i_1 < i_2$ such that $C_{i_1} = C_{i_2}$ then return 1
else return 0
```

But $\langle i \rangle$ must be an allowed message, which here is an ℓ-bit string, and i ranges from 1 to q. So the adversary is only valid if $q < 2^\ell$. But our ℓ may be very small, such as $\ell = 1$, and then we are not making enough queries for a successful attack. We need to make about $2^{n-\ell}$ queries, so this adversary only works if $\ell \geq n - \ell$, meaning $\ell \geq n/2$, which for $n = 128$ rules out several of the values of ℓ we were asked to consider.

Instead, letting $q = 2r$ be an even integer parameter, our adversary works as follows:

```
adversary $A$
for $i = 1, \ldots, r$ do $C_{0,i} \xleftarrow{\$} \mathbf{LR}(0^\ell, 0^\ell)$ ; $C_{1,i} \xleftarrow{\$} \mathbf{LR}(1^\ell, 0^\ell)$
if $\exists i_1, i_2$ such that $C_{0,i_1} = C_{1,i_2}$ then return 1
else return 0
```

For the analysis, let $R_{0,i}, R_{1,i}$ denote the random choices made by the encryption algorithm in the computations of $C_{0,i}, C_{1,i}$, respectively.

In game $\text{Left}_{\mathcal{SE}}$ we have $C_{0,i} = E_K(R_{0,i} \,\|\, 0^\ell)$ and $C_{1,i} = E_K(R_{1,i} \,\|\, 1^\ell)$ for all $i = 1, \ldots, r$. The fact that E_K is a permutation implies that for all i_1, i_2 we will have $C_{0,i_1} \neq C_{1,i_2}$. This means that A always returns 0 in game $\text{Left}_{\mathcal{SE}}$. Thus

$$\Pr\left[\text{Left}_{\mathcal{SE}}^A \Rightarrow 1\right] \;=\; 0\,.$$

In game $\text{Right}_{\mathcal{SE}}$ we have $C_{0,i} = E_K(R_{0,i} \,\|\, 0^\ell)$ and $C_{1,i} = E_K(R_{1,i} \,\|\, 0^\ell)$ for all $i = 1, \ldots, r$. The fact that E_K is a permutation implies that for any i_1, i_2 we will have $C_{0,i_1} = C_{1,i_2}$ iff $R_{0,i_1} = R_{1,i_2}$. Thus the probability that A returns 1 is the probability that there exist i_1, i_2 such that $R_{0,i_1} = R_{0,i_2}$. This probability is $D(2^{n-\ell}, r)$ where the function D is defined in the appendix below. By Lemma A.1 we have

$$\Pr\left[\text{Right}_{\mathcal{SE}}^A \Rightarrow 1\right] \;=\; D(2^{n-\ell}, r) \;\geq\; \frac{C(2^{n-\ell}, 2r)}{2} \;=\; \frac{C(2^{n-\ell}, q)}{2}\,.$$

The advantage of A is thus

$$\begin{aligned}\mathbf{Adv}_{\mathcal{SE}}^{\text{ind-cpa}}(A) \;&=\; \Pr\left[\text{Right}_{\mathcal{SE}}^A \Rightarrow 1\right] - \Pr\left[\text{Left}_{\mathcal{SE}}^A \Rightarrow 1\right] \\ &\geq\; \frac{C(2^{n-\ell}, q)}{2}\,.\end{aligned}$$

Theorem A.1 of the Appendix on the birthday problem tells us that

$$C(2^{n-\ell}, q) \;\geq\; 1 - e^{-q(q-1)/2^{n-\ell+1}}\,.$$

The table of Fig. 1 shows the smallest values of q for which the right hand side of the above inequality exceeds 1/4, for $n = 128$ and the requested values of ℓ.

3. [**40 points**] Give a reduction of the IND-CPA security of $\mathcal{SE}$ to the PRF security of E. This means you must state a theorem that upper bounds the ind-cpa advantage of a given ind-cpa adversary A as a function of the prf-advantage of a constructed prf-adversary B and (possibly) n, ℓ and the number q of LR-queries made by A. This is analogous to results we have seen in class for CTRC and CBC\$ encryption. Prove your theorem using a game sequence. The better the theorem (meaning the quantitative relationship) the more points you get.

The attack above provides the intuition for what one should expect to be able to prove. The attack relies on collisions in the R values across different queries. We expect to prove that in

ℓ	q
1	2^{64}
16	2^{57}
32	2^{49}
64	2^{33}
96	2^{17}

Figure 1: Number of queries for our attack to get a 0.25 advantage, for different message lengths.

the absence of such collisions, the adversary would have no advantage above that of breaking E as a PRF, so its advantage is bounded by the advantage of another adversary against E plus the probability of a collision in the R values. The following makes this precise and proves it.

Theorem. Let $E : \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^n$ be a block cipher. Let $\ell < n$. Let $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be the corresponding symmetric encryption scheme as defined above. Let A be an ind-cpa adversary against $\mathcal{SE}$ that has running time t and makes at most q **LR**-queries. Then there is a prf-adversary B against E such that

$$\mathbf{Adv}^{\text{ind-cpa}}_{\mathcal{SE}}(A) \leq 2 \cdot \mathbf{Adv}^{\text{prf}}_{E}(B) + \frac{q^2}{2^{n-\ell+1}} \, .$$

Furthermore, B makes at most q **Fn**-queries and has running time t.

The proof resembles those in class. Consider the games G_0 and G_1 from Fig. 2. The "If" statement treats a table entry $T[R \parallel M_b]$ as a boolean having value 1 if the entry is defined (different from $\perp$) and 0 otherwise.

Claim 1: There is a prf-adversary B with resource usage as indicated in the theorem statement, such that

$$\Pr\left[G_0^A \Rightarrow \mathsf{true}\right] - \Pr\left[G_1^A \Rightarrow \mathsf{true}\right] \leq \mathbf{Adv}^{\text{prf}}_{E}(B)$$

Proof: Adversary B works as follows:

```
Adversary $B$
$b \xleftarrow{\$} \{0,1\}$
$b' \xleftarrow{\$} A^{\text{SIM}}$
If $(b = b')$ then return 1
Else return 0
```

```
Subroutine SIM$(M_0, M_1)$
$R \xleftarrow{\$} \{0,1\}^{n-\ell}$
If not $T[R \parallel M_b]$ then
    $T[R \parallel M_b] \xleftarrow{\$} \mathbf{Fn}(R \parallel M_b)$
return $T[R \parallel M_b]$
```

It is easy to check that:

$$\Pr\left[\text{Real}_E^B \Rightarrow 1\right] \quad = \quad \Pr\left[G_0^A \Rightarrow \mathsf{true}\right]$$

```
Game $G_0$

procedure Initialize
$K \xleftarrow{\$} \{0,1\}^k$ ; $b \xleftarrow{\$} \{0,1\}$

procedure LR($M_0, M_1$)
$R \xleftarrow{\$} \{0,1\}^{n-\ell}$
If not $T[R \,\|\, M_b]$ then
    $T[R \,\|\, M_b] \leftarrow E_K(R \,\|\, M_b)$
return $T[R \,\|\, M_b]$

procedure Finalize($b'$)
return $(b = b')$
```

```
Game $G_1$

procedure Initialize
$K \xleftarrow{\$} \{0,1\}^k$ ; $b \xleftarrow{\$} \{0,1\}$

procedure LR($M_0, M_1$)
$R \xleftarrow{\$} \{0,1\}^{n-\ell}$
If not $T[R \,\|\, M_b]$ then
    $T[R \,\|\, M_b] \xleftarrow{\$} \{0,1\}^n$
return $T[R \,\|\, M_b]$

procedure Finalize($b'$)
return $(b = b')$
```

Figure 2: Games G_0 and G_1 used in the proof of the theorem.

$$\Pr\left[\mathrm{Rand}^B_{\{0,1\}^n} \Rightarrow 1\right] \quad = \quad \Pr\left[G_1^A \Rightarrow \mathsf{true}\right]$$

which suffices to prove Claim 1. ▌

Given Claim 1, we have:

$$\Pr\left[G_0^A \Rightarrow \mathsf{true}\right] = \Pr\left[G_1^A \Rightarrow \mathsf{true}\right] + \underbrace{(\Pr[G_0^A \Rightarrow \mathsf{true}] - \Pr[G_1^A \Rightarrow \mathsf{true}])}_{\leq\ \mathbf{Adv}_E^{\mathrm{prf}}(B)}$$

which means,

$$\begin{aligned} \mathbf{Adv}^{\text{ind-cpa}}_{\mathcal{SE}}(A) &= 2 \cdot \Pr\left[G_0^A \Rightarrow \mathsf{true}\right] - 1 \\ &\leq 2 \cdot (\Pr[G_1^A \Rightarrow \mathsf{true}] + \mathbf{Adv}_E^{\mathrm{prf}}(B)) - 1 \\ &= 2 \cdot \mathbf{Adv}_E^{\mathrm{prf}}(B) + 2 \cdot \Pr[G_1^A \Rightarrow \mathsf{true}] - 1 \end{aligned}$$

Next, we have:

Claim 2: $\Pr\left[G_1^A \Rightarrow \mathsf{true}\right] \leq \frac{1}{2} + \frac{q^2}{2^{n-\ell+1}}$

Proof: For the proof of this claim, we use the games G_2, G_3 shown in Fig. 3. It is straightforward to check that

$$\begin{aligned} \Pr[G_1^A \Rightarrow \mathsf{true}] &= \Pr[G_2^A \Rightarrow \mathsf{true}] \\ &= \Pr[G_3^A \Rightarrow \mathsf{true}] + \Big(\Pr[G_2^A \Rightarrow \mathsf{true}] - \Pr[G_3^A \Rightarrow \mathsf{true}]\Big) \end{aligned}$$

Our aim will be to show the following:

```
Game $\boxed{G_2}$, $G_3$

procedure Initialize
$b \xleftarrow{\$} \{0,1\}$

procedure LR$(M_0, M_1)$
$R \xleftarrow{\$} \{0,1\}^{n-\ell}$ ; $C \xleftarrow{\$} \{0,1\}^n$
If $T[R \| M_b]$ then $\mathsf{bad} \leftarrow \mathsf{true}$ ; $\boxed{C \leftarrow T[R \| M_b]}$
$T[R \| M_b] \leftarrow C$
return $C$

procedure Finalize$(b')$
return $(b = b')$
```

```
Game $G_4$

procedure Initialize
$b \xleftarrow{\$} \{0,1\}$ ; $S \leftarrow \emptyset$

procedure LR$(M_0, M_1)$
$R \xleftarrow{\$} \{0,1\}^{n-\ell}$ ; $C \xleftarrow{\$} \{0,1\}^n$
If $R \in S$ then $\mathsf{bad} \leftarrow \mathsf{true}$
$S \leftarrow S \cup \{R\}$
return $C$

procedure Finalize$(b')$
return $(b = b')$
```

Figure 3: Games G_2, G_3 and G_4. Game G_2 includes the boxed code and G_3 does not.

- $\Pr[G_3^A \Rightarrow \mathsf{true}] = \frac{1}{2}$
- $\Pr[G_2^A \Rightarrow \mathsf{true}] - \Pr[G_3^A \Rightarrow \mathsf{true}] \leq \frac{q^2}{2^{n-\ell+1}}$

From this, Claim 2 will follow immediately.

The first part is true from the very definition of G_3: in this game, the ciphertext C given in reply to any LR-query of A is always a uniformly random string in $\{0,1\}^n$, sampled independently of b. Since b itself is chosen uniformly at random from $\{0,1\}$, A cannot guess its value with probability better than 1/2.

To prove the second part, we will invoke the fundamental lemma of game playing. Notice that G_2 and G_3 are identical-until-bad, which means

$$\Pr\left[G_2^A \Rightarrow \mathsf{true}\right] - \Pr\left[G_3^A \Rightarrow \mathsf{true}\right] \leq \Pr\left[G_3^A \text{ sets } \mathsf{bad}\right] .$$

It thus suffices to argue that $\Pr[G_3^A \text{ sets } \mathsf{bad}]$ is at most $\frac{q^2}{2^{n-\ell+1}}$. For this, we consider game G_4 shown in figure 3. But

$$\begin{aligned}\Pr\left[G_3^A \text{ sets } \mathsf{bad}\right] &\leq \Pr\left[G_4^A \text{ sets } \mathsf{bad}\right] \\ &\leq \sum_{j=1}^{q} \frac{j-1}{2^{n-\ell}} \\ &= \frac{q(q-1)}{2^{n-\ell+1}} \leq \frac{q^2}{2^{n-\ell+1}} .\end{aligned}$$

This completes the proof of Claim 2, and thus of the theorem as well. ▌

4. **[10 points]** As a result of the above, do you consider the scheme to be secure or insecure? Discuss this for $E = \mathsf{AES}$ and $\ell = 1, 16, 32, 64, 96$.

Consider $\ell = 64$. Here the attack provided above breaks the scheme using 2^{33} queries, and computation of the same order. We are talking about encrypting $2^{33} = 8,589,934,592$ or about 8 billion messages. This is not a large number in practice. Imagine a situation where packets are being encrypted on a fast (1Ggit/sec, say) network. It will take very little time before the number of messages encrypted is this number. But the attack says that you cannot encrypt this many message securely.

Of course there are other issues and arguments too. If we are encrypting email, you might say 8 billion messages is a large number and thus the attack is not practical. True, but that does not mean the scheme is secure. Remember as cryptographers we do not want to constrain the use of the scheme to certain applications. This is too complicated: no matter what, once a scheme is out there, people use it for everything. It better work even in extreme situations. Along the same lines you might argue that the particular attack is not important because it leaks very little information. Again, not a good viewpoint. We have seen many reasons to view IND-CPA as the right notion of security, so the attack is relevant.

For $\ell = 96$ the attack succeeds even faster. So for $\ell = 64, 96$ I would consider the scheme insecure.

For $\ell = 32$, the attack takes 2^{49} queries. That's borderline. For $\ell = 16$ the attack takes 2^{57} queries. This is getting prohibitive. Maybe not in time, but it is in queries, and for $\ell = 1$ the number is even higher.

But we are not yet ready to say even $\ell = 1$ is secure because there may be better attacks. That's where the theorem comes in. The assumed PRF security of AES means that we can assume

$$\mathbf{Adv}_{E}^{\mathrm{prf}}(B) \leq \frac{q^2}{2^{n+1}}$$

in the theorem. Since $\ell < n$ this means that

$$\mathbf{Adv}_{\mathcal{SE}}^{\mathrm{ind\text{-}cpa}}(A) \leq \frac{q^2}{2^{n-\ell}} \; .$$

So the theorem is effectively saying the attack we found is the best possible. With that, we have some confidence in saying the scheme is secure for $\ell = 1, 16$.

A Generalized birthday lemma

Let N, r be positive integers and let S be a set of size N. Suppose we pick $y_1, \ldots, y_r$ at random from S and also pick $z_1, \ldots, z_r$ at random from S. Let $D(N, r)$ be the probability that there exist i, j such that $y_i = z_j$.

Lemma A.1 Let N, r be positive integers. Then

$$D(N, r) \; \geq \; \frac{C(N, 2r)}{2} \; .$$
