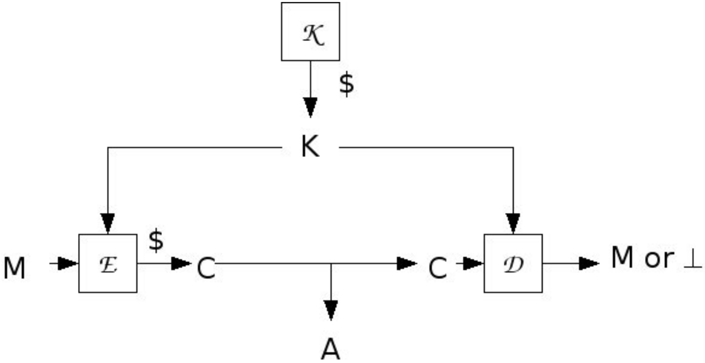
SYMMETRIC ENCRYPTION

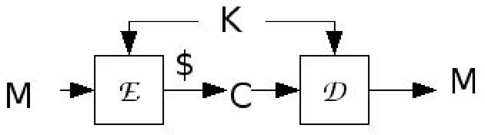
Syntax

A symmetric encryption scheme $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ consists of three algorithms:



$\mathcal{K}$ and $\mathcal{E}$ may be randomized, but $\mathcal{D}$ must be deterministic.

Correct decryption requirement



More formally: For all keys K that may be output by $\mathcal{K}$, and for all M in the *message space*, we have

$$Pr[\mathcal{D}_K(\mathcal{E}_K(M)) = M] = 1 ,$$

where the probability is over the coins of $\mathcal{E}$.

A scheme will usually specify an associated message space.

Modes of operation

$E : \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^\ell$ a family of functions

Usually a block cipher, in which case $\ell = n$.

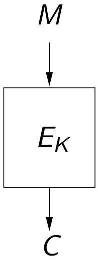
Notation: $x[i]$ is the i-th block of a string x, so that $x = x[1] \ldots x[m]$. Length of blocks varies.

Always:

Alg $\mathcal{K}$
$K \stackrel{\$}{\leftarrow} \{0,1\}^k$
return K

Modes of operation

Block cipher provides parties sharing K with



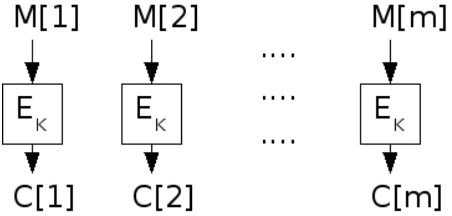
which enables them to encrypt a 1-block message.

How do we encrypt a long message using a primitive that only applies to n-bit blocks?

ECB: Electronic Codebook Mode

$\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ where:

Alg $\mathcal{E}_K(M)$	**Alg** $\mathcal{D}_K(C)$
for $i = 1, \ldots, m$ do	for $i = 1, \ldots, m$ do
$\quad C[i] \leftarrow E_K(M[i])$	$\quad M[i] \leftarrow E_K^{-1}(C[i])$
return C	return M

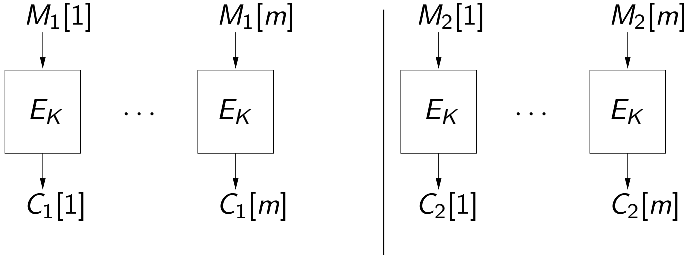


Correct decryption relies on E being a block cipher, so that E_K is invertible

Security of ECB

Weakness: $M_1 = M_2 \Rightarrow C_1 = C_2$

Why is the above true? Because E_K is deterministic:



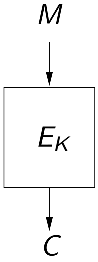
Why does this matter?

Security of ECB

Suppose we know that there are only two possible messages, $Y = 1^n$ and $N = 0^n$, for example representing

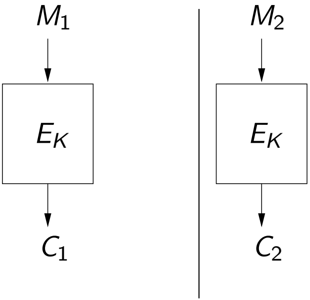
- FIRE or DON'T FIRE a missile
- BUY or SELL a stock
- Vote YES or NO

Then ECB algorithm will be $\mathcal{E}_K(M) = E_K(M)$.



Security of ECB

Votes $M_1, M_2 \in \{Y, N\}$ are ECB encrypted and adversary sees ciphertexts $C_1 = E_K(M_1)$ and $C_2 = E_K(M_2)$



Adversary may have cast the first vote and thus knows M_1; say $M_1 = Y$. Then adversary can figure out M_2:

- If $C_2 = C_1$ then M_2 must be Y
- Else M_2 must be N

Is this avoidable?

Let $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be ANY encryption scheme.

Suppose $M_1, M_2 \in \{Y, N\}$ and

- Sender sends ciphertexts $C_1 \leftarrow \mathcal{E}_K(M_1)$ and $C_2 \leftarrow \mathcal{E}_K(M_2)$
- Adversary A knows that $M_1 = Y$

Adversary says: If $C_2 = C_1$ then M_2 must be Y else it must be N.

Does this attack work?

Is this avoidable?

Let $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be ANY encryption scheme.

Suppose $M_1, M_2 \in \{Y, N\}$ and

- Sender sends ciphertexts $C_1 \leftarrow \mathcal{E}_K(M_1)$ and $C_2 \leftarrow \mathcal{E}_K(M_2)$
- Adversary A knows that $M_1 = Y$

Adversary says: If $C_2 = C_1$ then M_2 must be Y else it must be N.

Does this attack work?

Yes, if $\mathcal{E}$ is deterministic.

Randomized encryption

For encryption to be secure it must be randomized

That is, algorithm $\mathcal{E}_K$ flips coins.

If the same message is encrypted twice, we are likely to get back different answers. That is, if $M_1 = M_2$ and we let

$$C_1 \stackrel{\$}{\leftarrow} \mathcal{E}_K(M_1) \text{ and } C_2 \stackrel{\$}{\leftarrow} \mathcal{E}_K(M_2)$$

then

$$Pr[C_1 = C_2]$$

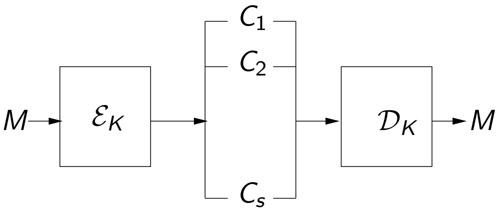
will (should) be small, where the probability is over the coins of $\mathcal{E}$.

Randomized encryption

There are many possible ciphertexts corresponding to each message.

If so, how can we decrypt?

We will see examples soon.



Randomized encryption

A fundamental departure from classical and conventional notions of encryption.

Clasically, encryption (e.g., substitution cipher) is a code, associating to each message a unique ciphertext.

Now, we are saying no such code is secure, and we look to encryption mechanisms which associate to each message a number of different possible ciphertexts.

CBC$: Cipher Block Chaining with random IV mode

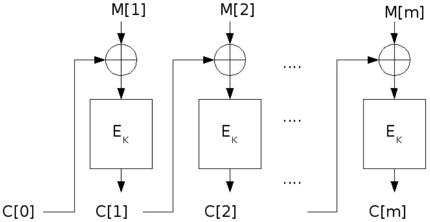
$\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ where:

Alg $\mathcal{E}_K(M)$

$C[0] \xleftarrow{\$} \{0,1\}^n$
for $i = 1, \ldots, m$ do
    $C[i] \leftarrow E_K(M[i] \oplus C[i-1])$
return C

Alg $\mathcal{D}_K(C)$

for $i = 1, \ldots, m$ do
    $M[i] \leftarrow E_K^{-1}(C[i]) \oplus C[i-1]$
return M



Correct decryption relies on E being a block cipher.

CTR\$ mode

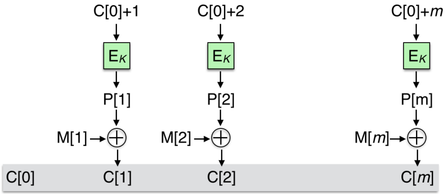
Let E: $\{0,1\}^k \times \{0,1\}^n \to \{0,1\}^\ell$ be a family of functions. If $X \in \{0,1\}^n$ and $i \in \mathbf{N}$ then $X + i$ denotes the n-bit string formed by converting X to an integer, adding i modulo 2^n, and converting the result back to an n-bit string. Below the message is a sequence of ℓ-bit blocks:

Alg $\mathcal{E}_K(M)$

$C[0] \xleftarrow{\$} \{0,1\}^n$
for $i = 1, \ldots, m$ do
  $P[i] \leftarrow E_K(C[0] + i)$
  $C[i] \leftarrow P[i] \oplus M[i]$
return C

Alg $\mathcal{D}_K(C)$

for $i = 1, \ldots, m$ do
  $P[i] \leftarrow E_K(C[0] + i)$
  $M[i] \leftarrow P[i] \oplus C[i]$
return M



CTR$ mode

Alg $\mathcal{E}_K(M)$

$C[0] \xleftarrow{\$} \{0,1\}^n$
for $i = 1, \ldots, m$ do
    $P[i] \leftarrow E_K(C[0] + i)$
    $C[i] \leftarrow P[i] \oplus M[i]$
return C

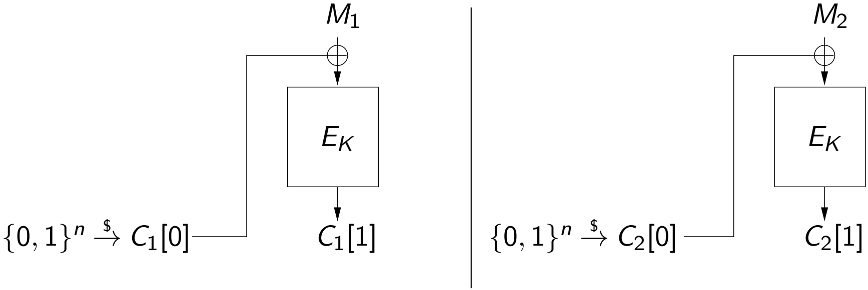
Alg $\mathcal{D}_K(C)$

for $i = 1, \ldots, m$ do
    $P[i] \leftarrow E_K(C[0] + i)$
    $M[i] \leftarrow P[i] \oplus C[i]$
return M

- $\mathcal{D}$ does not use E_K^{-1}! This is why CTR$ can use a family of functions E that is not required to be a blockcipher.
- Encryption and Decryption are parallelizable.

Voting with CBC$

Suppose we encrypt $M_1, M_2 \in \{Y, N\}$ with CBC$.



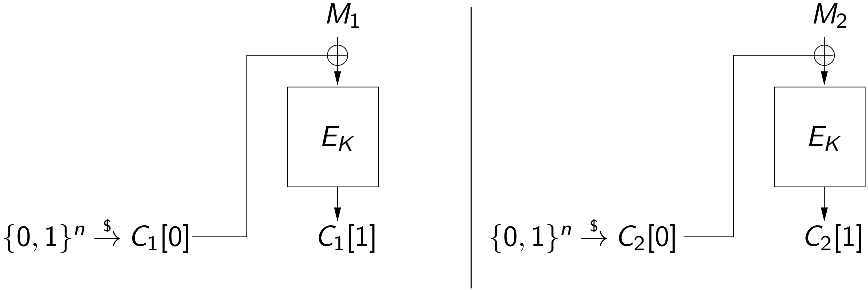
Adversary A sees $C_1 = C_1[0]C_1[1]$ and $C_2 = C_2[0]C_2[1]$.

Suppose A knows that $M_1 = Y$.

Can A determine whether $M_2 = Y$ or $M_2 = N$?

Voting with CBC$

Suppose we encrypt $M_1, M_2 \in \{Y, N\}$ with CBC$.



Adversary A sees $C_1 = C_1[0]C_1[1]$ and $C_2 = C_2[0]C_2[1]$.

Suppose A knows that $M_1 = Y$.

Can A determine whether $M_2 = Y$ or $M_2 = N$?

NO!

Assessing security

So CBC$ is better than ECB. But is it secure?

CBC$ is widely used so knowing whether it is secure is important

To answer this we first need to decide and formalize what we mean by secure.

Security requirements

Suppose sender computes

$$C_1 \stackrel{\$}{\leftarrow} \mathcal{E}_K(M_1)\,;\; \cdots\,;\; C_q \stackrel{\$}{\leftarrow} \mathcal{E}_K(M_q)$$

Adversary A has $C_1, \ldots, C_q$

What if A	
Retrieves K	Bad!
Retrieves M_1	Bad!

But also we want to hide all partial information about the data stream, such as

- Does $M_1 = M_2$?
- What is first bit of M_1?
- What is XOR of first bits of M_1, M_2?

Something we won't hide: the length of the message

What we seek

We want a single "master" property MP of an encryption scheme such that

- MP can be easily specified
- We can evaluate whether a scheme meets it
- MP implies ALL the security conditions we want: it guarantees that a ciphertext reveals NO partial information about the plaintext.

Intuition for definition of IND-CPA

The master property MP is called IND-CPA (indistinguishability under chosen plaintext attack).

Consider encrypting one of two possible message streams, either

$$M_0^1, ..., M_0^q$$

or

$$M_1^1, ..., M_1^q \; ,$$

where $|M_0^i| = |M_1^i|$ for all $1 \leq i \leq q$. Adversary, given ciphertexts $C^1, \ldots,$ C^q and both data streams, has to figure out which of the two streams was encrypted.

We will even let the adversary pick the messages: It picks (M_0^1, M_1^1) and gets back C^1, then picks (M_0^2, M_1^2) and gets back C^2, and so on.

Games for ind-cpa-advantage of an adversary A

Let $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be an encryption scheme

Game $\mathrm{Left}_{\mathcal{SE}}$	Game $\mathrm{Right}_{\mathcal{SE}}$
procedure Initialize	**procedure Initialize**
$K \xleftarrow{\$} \mathcal{K}$	$K \xleftarrow{\$} \mathcal{K}$
procedure LR(M_0, M_1)	**procedure LR**(M_0, M_1)
Return $C \xleftarrow{\$} \mathcal{E}_K(M_0)$	Return $C \xleftarrow{\$} \mathcal{E}_K(M_1)$

Associated to $\mathcal{SE}, A$ are the probabilities

$$\Pr\left[\mathrm{Left}_{\mathcal{SE}}^{A} \Rightarrow 1\right] \qquad \Big| \qquad \Pr\left[\mathrm{Right}_{\mathcal{SE}}^{A} \Rightarrow 1\right]$$

that A outputs 1 in each world. The (ind-cpa) advantage of A is

$$\mathbf{Adv}_{\mathcal{SE}}^{\mathrm{ind\text{-}cpa}}(A) = \Pr\left[\mathrm{Right}_{\mathcal{SE}}^{A} \Rightarrow 1\right] - \Pr\left[\mathrm{Left}_{\mathcal{SE}}^{A} \Rightarrow 1\right]$$

Message length restriction

It is required that $|M_0| = |M_1|$ in any query M_0, M_1 that A makes to **LR**. An adversary A violating this condition is considered invalid.

This reflects that encryption is not aiming to hide the length of messages.

The measure of success

$\mathbf{Adv}^{\mathrm{ind\text{-}cpa}}_{\mathcal{SE}}(A) \approx 1$ means A is doing well and $\mathcal{SE}$ is not ind-cpa-secure.

$\mathbf{Adv}^{\mathrm{ind\text{-}cpa}}_{\mathcal{SE}}(A) \approx 0$ (or ≤ 0) means A is doing poorly and $\mathcal{SE}$ resists the attack A is mounting.

Adversary resources are its running time t and the number q of its oracle queries, the latter representing the number of messages encrypted.

Security: $\mathcal{SE}$ is IND-CPA-secure if $\mathbf{Adv}^{\mathrm{ind\text{-}cpa}}_{\mathcal{SE}}(A)$ is "small" for ALL A that use "practical" amounts of resources.

Insecurity: $\mathcal{SE}$ is not IND-CPA-secure if we can specify an explicit A that uses "few" resources yet achieves "high" ind-cpa-advantage.

ECB is not IND-CPA-secure

Let $E : \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^n$ be a block cipher. Recall that ECB mode defines symmetric encryption scheme $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ with

$$\mathcal{E}_K(M) = E_K(M[1])E_K(M[2]) \cdots E_K(M[m])$$

Can we design A so that

$$\mathbf{Adv}^{\text{ind-cpa}}_{\mathcal{SE}}(A) = \Pr\left[\text{Right}^A_{\mathcal{SE}} \Rightarrow 1\right] - \Pr\left[\text{Left}^A_{\mathcal{SE}} \Rightarrow 1\right]$$

is close to 1?

ECB is not IND-CPA-secure

Let $E : \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^n$ be a block cipher. Recall that ECB mode defines symmetric encryption scheme $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ with

$$\mathcal{E}_K(M) = E_K(M[1])E_K(M[2]) \cdots E_K(M[m])$$

Can we design A so that

$$\mathbf{Adv}^{\text{ind-cpa}}_{\mathcal{SE}}(A) = \Pr\left[\text{Right}^A_{\mathcal{SE}} \Rightarrow 1\right] - \Pr\left[\text{Left}^A_{\mathcal{SE}} \Rightarrow 1\right]$$

is close to 1?

Exploitable weakness of $\mathcal{SE}$: $M_1 = M_2$ implies $\mathcal{E}_K(M_1) = \mathcal{E}_K(M_2)$.

ECB is not IND-CPA-secure

Let $\mathcal{E}_K(M) = E_K(M[1]) \cdots E_K(M[m])$.

adversary A

$C_1 \leftarrow \mathbf{LR}(0^n, 0^n)\,;\; C_2 \leftarrow \mathbf{LR}(1^n, 0^n)$

if $C_1 = C_2$ then return 1 else return 0

Right game analysis

$\mathcal{E}$ is defined by $\mathcal{E}_K(M) = E_K(M[1]) \cdots E_K(M[m])$.

adversary A
$C_1 \leftarrow \mathbf{LR}(0^n, 0^n)\,;\ C_2 \leftarrow \mathbf{LR}(1^n, 0^n)$
if $C_1 = C_2$ then return 1 else return 0

Game $\text{Right}_{\mathcal{SE}}$
procedure Initialize
$K \stackrel{\$}{\leftarrow} \mathcal{K}$
procedure LR(M_0, M_1)
Return $\mathcal{E}_K(M_1)$

Then

$$\Pr\left[\text{Right}_{\mathcal{SE}}^A \Rightarrow 1\right] =$$

Right game analysis

$\mathcal{E}$ is defined by $\mathcal{E}_K(M) = E_K(M[1]) \cdots E_K(M[m])$.

adversary A

$C_1 \leftarrow \mathbf{LR}(0^n, 0^n)\,;\ C_2 \leftarrow \mathbf{LR}(1^n, 0^n)$

if $C_1 = C_2$ then return 1 else return 0

Game $\mathrm{Right}_{\mathcal{SE}}$
procedure Initialize
$K \stackrel{\$}{\leftarrow} \mathcal{K}$
procedure LR(M_0, M_1)
Return $\mathcal{E}_K(M_1)$

Then

$$\Pr\left[\mathrm{Right}_{\mathcal{SE}}^{A} \Rightarrow 1\right] = 1$$

because $C_1 = E_K(0^n)$ and $C_2 = E_K(0^n)$.

Left game analysis

$\mathcal{E}$ is defined by $\mathcal{E}_K(M) = E_K(M[1]) \cdots E_K(M[m])$.

adversary A
$C_1 \leftarrow \mathbf{LR}(0^n, 0^n)\,;\ C_2 \leftarrow \mathbf{LR}(1^n, 0^n)$
if $C_1 = C_2$ then return 1 else return 0

Game $\mathrm{Left}_{\mathcal{SE}}$
procedure Initialize
$K \xleftarrow{\$} \mathcal{K}$
procedure LR(M_0, M_1)
Return $\mathcal{E}_K(M_0)$

Then

$$\Pr\left[\mathrm{Left}_{\mathcal{SE}}^{A} \Rightarrow 1\right] =$$

Left game analysis

$\mathcal{E}$ is defined by $\mathcal{E}_K(M) = E_K(M[1]) \cdots E_K(M[m])$.

adversary A
$C_1 \leftarrow \mathbf{LR}(0^n, 0^n)$; $C_2 \leftarrow \mathbf{LR}(1^n, 0^n)$
if $C_1 = C_2$ then return 1 else return 0

Game $\mathrm{Left}_{\mathcal{SE}}$
procedure Initialize
$K \xleftarrow{\$} \mathcal{K}$
procedure LR(M_0, M_1)
Return $\mathcal{E}_K(M_0)$

Then

$$\Pr\left[\mathrm{Left}_{\mathcal{SE}}^A \Rightarrow 1\right] = 0$$

because $C_1 = E_K(0^n) \neq E_K(1^n) = C_2$.

ECB is not IND-CPA secure

adversary A
$C_1 \leftarrow \mathbf{LR}(0^n, 0^n)\,;\; C_2 \leftarrow \mathbf{LR}(1^n, 0^n)$
if $C_1 = C_2$ then return 1 else return 0

$$\mathbf{Adv}^{\text{ind-cpa}}_{\mathcal{SE}}(A) = \overbrace{\Pr\left[\text{Right}^{A}_{\mathcal{SE}} = 1\right]}^{1} - \overbrace{\Pr\left[\text{Left}^{A}_{\mathcal{SE}} = 1\right]}^{0}$$
$$= 1$$

And A is very efficient, making only two queries.

Thus ECB is **not** IND-CPA secure.

Exercise

Let $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be any symmetric encryption scheme for which $\mathcal{E}$ is deterministic. Show that $\mathcal{SE}$ is **not** IND-CPA secure by giving pseudo-code for an efficient adversary A that achieves

$$\mathbf{Adv}^{\text{ind-cpa}}_{\mathcal{SE}}(A) = 1 \ .$$

Assume the message space is $\{0,1\}^*$, meaning any string is a legitimate message.

Exercise

Let $\mathcal{K}$ be the key-generation algorithm that returns a random 128-bit string as the key K. Define

Alg $\mathcal{E}_K(M)$

$R \xleftarrow{\$} \{0,1\}^{128}$; $M[1]\ldots M[m] \leftarrow M$; $C[0] \leftarrow R$
for $i = 1, \ldots, m$ do
  $W[i] \leftarrow (R + i) \bmod 2^{128}$; $C[i] \leftarrow \mathrm{AES}_K(M[i] \oplus W[i])$
$C \leftarrow C[0]C[1]\ldots C[n]$; Return C

Above $W[i] \leftarrow (R + i) \bmod 2^{128}$ means we regard R as an integer, add i to it, take the result modulo 2^{128}, view this as a 128-bit string, and assign it to $W[i]$. A message must be a string whose length is a positive multiple of 128.

Exercise

1. Specify a decryption algorithm $\mathcal{D}$ such that $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ is a symmetric encryption scheme satisfying the correct decryption condition.
2. Present in pseudocode a practical adversary A making one **LR** query and achieving $\mathbf{Adv}^{\text{ind-cpa}}_{\mathcal{SE}}(A) = 1$.

There should be a succinct analysis justifying the claimed advantage.

Why is IND-CPA the "master" property?

We claim that if encryption scheme $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ is IND-CPA secure then the ciphertext hides ALL partial information about the plaintext.

For example, from $C_1 \stackrel{\$}{\leftarrow} \mathcal{E}_K(M_1)$ and $C_2 \stackrel{\$}{\leftarrow} \mathcal{E}_K(M_2)$ the adversary cannot

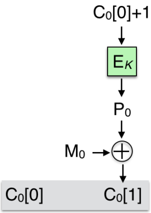
- get M_1
- get 1st bit of M_1
- get XOR of the 1st bits of M_1, M_2
- etc.

Exercise: State and prove such claims formally.

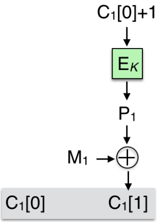
Birthday attack on CTR$

Let $E : \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^\ell$ be a family of functions and $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ the corresponding CTR$ symmetric encryption scheme. Suppose 1-block messages M_0, M_1 are encrypted:

$C_0[0]C_0[1] \xleftarrow{\$} \mathcal{E}(K, M_0)$



$C_1[0]C_1[1] \xleftarrow{\$} \mathcal{E}(K, M_1)$



Let us say we are **lucky** If $C_0[0] = C_1[0]$. If so:

$$C_0[1] = C_1[1] \text{ if and only if } M_0 = M_1$$

So if we are lucky we can detect message equality and violate IND-CPA.

Birthday attack on CTR$

Let $1 \le q < 2^n$ be a parameter and let $\langle i \rangle$ be integer i encoded as an ℓ-bit string.

adversary A

for $i = 1, ..., q$ do
    $C^i[0]C^i[1] \xleftarrow{\$} \mathbf{LR}(\langle i \rangle, \langle 0 \rangle)$
$S \leftarrow \{(j,t)\colon C^j[0] = C^t[0] \text{ and } j < t\}$
If $S \neq \emptyset$, then
    $(j,t) \xleftarrow{\$} S$
    If $C^j[1] = C^t[1]$ then return 1
return 0

Birthday attack on CTR\$: Right game analysis

adversary A

for $i = 1, \ldots, q$ do

$\quad C^i[0]C^i[1] \xleftarrow{\$} \mathbf{LR}(\langle i \rangle, \langle 0 \rangle)$

$S \leftarrow \{(j,t) \colon C^j[0] = C^t[0] \text{ and } j < t\}$

If $S \neq \emptyset$, then

$\quad (j,t) \xleftarrow{\$} S$

$\quad$ If $C^j[1] = C^t[1]$ then return 1

return 0

Game $\mathrm{Right}_{\mathcal{SE}}$

procedure Initialize

$K \xleftarrow{\$} \mathcal{K}$

procedure LR(M_0, M_1)

$C[0] \xleftarrow{\$} \{0,1\}^n$

$P \leftarrow E(K, C[0]+1)$

$C[1] \leftarrow P \oplus M_1$

Return $C[0]C[1]$

If $C^j[0] = C^t[0]$ (lucky) then

$$C^j[1] = \langle 0 \rangle \oplus E_K(C^j[0]+1) = \langle 0 \rangle \oplus E_K(C^t[0]+1) = C^t[1]$$

so

$$\Pr\left[\mathrm{Right}_{\mathcal{SE}}^{A} \Rightarrow 1\right] = \Pr\left[S \neq \emptyset\right] = C(2^n, q)$$

Birthday attack on CTR\$: Left game analysis

adversary A

for $i = 1, \ldots, q$ do
  $C^i[0]C^i[1] \xleftarrow{\$} \mathbf{LR}(\langle i \rangle, \langle 0 \rangle)$
$S \leftarrow \{(j,t) \colon C^j[0] = C^t[0] \text{ and } j < t\}$
If $S \neq \emptyset$, then
  $(j,t) \xleftarrow{\$} S$
  If $C^j[1] = C^t[1]$ then return 1
return 0

Game $\text{Left}_{\mathcal{SE}}$

procedure Initialize
$K \xleftarrow{\$} \mathcal{K}$

procedure LR(M_0, M_1)
$C[0] \xleftarrow{\$} \{0,1\}^n$
$P \leftarrow E(K, C[0]+1)$
$C[1] \leftarrow P \oplus M_0$
Return $C[0]C[1]$

If $C^j[0] = C^t[0]$ (lucky) then

$$C^j[1] = \langle j \rangle \oplus E_K(C^j[0]+1) \neq \langle t \rangle \oplus E_K(C^t[0]+1) = C^t[1]$$

so

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

Birthday attack on CTR\$

$$\begin{aligned}\mathbf{Adv}^{\text{ind-cpa}}_{\mathcal{SE}}(A) &= \Pr\left[\text{Right}^{A}_{\mathcal{SE}}\Rightarrow 1\right] - \Pr\left[\text{Left}^{A}_{\mathcal{SE}}\Rightarrow 1\right] \\ &= C(2^n, q) - 0 \;\geq\; 0.3 \cdot \frac{q(q-1)}{2^n}\end{aligned}$$

Conclusion: CTR\$ can be broken (in the IND-CPA sense) in about $2^{n/2}$ queries, where n is the block length of the underlying block cipher, regardless of the cryptanalytic strength of the block cipher.

Exercise

The above attack on CTR\$ uses 1-block messages. Letting $\mathcal{SE}$ be the same scheme, give an adversary A that makes q **LR**-queries, each consisting of two m-block messages, and achieves

$$\mathbf{Adv}^{\text{ind-cpa}}_{\mathcal{SE}}(A) = \Omega\left(\frac{mq^2}{2^n}\right)$$

The running time of A should be about $\mathcal{O}(mq(n+\ell) \cdot \log(mq(n+\ell)))$.

Security of CTR$

So far: A q-query adversary can break CTR$ with advantage $\approx \frac{q^2}{2^{n+1}}$

Question: Is there any better attack?

Security of CTR$

So far: A q-query adversary can break CTR$ with advantage $\approx \frac{q^2}{2^{n+1}}$

Question: Is there any better attack?

Answer: NO!

We can prove that the best q-query attack short of breaking the block cipher has advantage at most

$$\frac{2(q-1)\sigma}{2^n}$$

where σ is the total number of blocks across all messages encrypted.

Example: If q 1-block messages are encrypted then $\sigma = q$ so the adversary advantage is not more than $2q^2/2^n$.

For $E = \text{AES}$ this means up to about 2^{64} blocks may be securely encrypted, which is good.

Security of CTR$

Theorem: [BDJR97] Let $E : \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^\ell$ be a family of functions and $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ the corresponding CTR$ symmetric encryption scheme. Let A be an ind-cpa adversary against $\mathcal{SE}$ that has running time t and makes at most q **LR** queries, the messages across them totaling at most σ blocks. Then there is a prf-adversary B against E such that

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

Furthermore, B makes at most σ oracle queries and has running time $t + \Theta(\sigma \cdot (n + \ell))$.

Intuition and proof

We first give some intuition for the proof and then the proof itself.

We assume for simplicity that both messages in each **LR** query of A are m blocks long. Thus $\sigma = mq$.

Note a block is ℓ bits, so each message in a query is $m\ell$ bits.

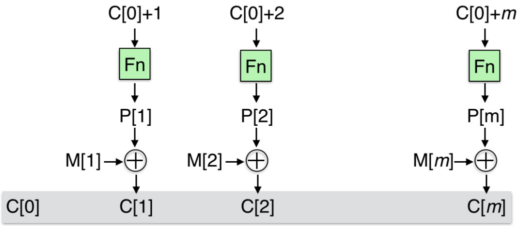
Extending the proof to the general case is an exercise.

We let $C_i = C_i[0]C_i[1]\ldots C_i[m]$ denote the response of the **LR** oracle to A's i-th query.

Intuition for IND-CPA security of CTR$

Consider the CTR$ scheme with E_K replaced by a random function **Fn** with range $\{0,1\}^\ell$.

$\underline{\textbf{Alg } \mathcal{E}_{\textbf{Fn}}(M)}$
$C[0] \xleftarrow{\$} \{0,1\}^n$
for $i = 1, \ldots, m$ do
  $P[i] \leftarrow \textbf{Fn}(C[0] + i)$
  $C[i] \leftarrow P[i] \oplus M[i]$
return C



Analyzing this is a thought experiment, but we can ask whether it is IND-CPA secure.

If so, the assumption that E is a PRF says CTR$ with E is IND-CPA secure.

CTR$ with a random function

Let E be the event that the points

$$C_1[0]+1, \ldots, C_1[0]+m, \ldots, C_q[0]+1, \ldots, C_q[0]+m \ ,$$

on which **Fn** is evaluated across the q encryptions, are all distinct.

Case 1: E happens. Then the encryption is a one-time-pad: ciphertexts are random, independent strings, regardless of which message is encrypted. So A has zero advantage.

Case 2: E doesn't happen. Then A may have high advantage but it does not matter because $\Pr[E]$ doesn't happen is small. (It is the small additive term in the theorem.)

The Union bound

Let $\mathrm{E}_1, \ldots, \mathrm{E}_n$ be events in a probability space. Then

$$\Pr[\bigvee_{1 \le i \le n} \mathrm{E}_i] \le \sum_{i=1}^{n} \Pr[\mathrm{E}_i] \, .$$

Example: In the case $n = 2$ we have

$$\begin{aligned} \Pr[\mathrm{E}_1 \vee \mathrm{E}_2] &= \Pr[\mathrm{E}_1] + \Pr[\mathrm{E}_2] - \Pr[\mathrm{E}_1 \wedge \mathrm{E}_2] \\ &\le \Pr[\mathrm{E}_1] + \Pr[\mathrm{E}_2] \, . \end{aligned}$$

Interval intersection probability

Let $N, q, m \geq 1$ be integers and let $\mathbf{Z}_N = \{0, 1, \ldots, N-1\}$. Let $+$ be addition modulo N. Consider the game

For $i = 1, \ldots, q$ do
  $c_i \stackrel{\$}{\leftarrow} \mathbf{Z}_N$; $I_i \leftarrow \{c_i + 1, \ldots, c_i + m\}$

For $1 \leq i < j \leq q$ define the events

$$\mathrm{B}_{i,j} \;:\; S_i \cap S_j \neq \emptyset \quad \text{and} \quad \mathrm{B} \;:\; \bigvee_{1 \leq i < j \leq q} \mathrm{B}_{i,j} \; .$$

Then let

$$\mathrm{IIP}(N, q, m) = \Pr[\mathrm{B}] \; .$$

Problem: Upper bound $\mathrm{IIP}(N, q, m)$ as a function of N, q, m.

Interval intersection probability

Claim: $\mathrm{IIP}(N,q,m) \leq \dfrac{q(q-1)}{2}\dfrac{(2m-1)}{N}$

Why? For any $1 \leq i < j \leq q$ we have

$$\begin{aligned}\Pr[\mathrm{B}_{i,j}] &= \Pr[S_i \cap S_j \neq \emptyset] \\ &= \Pr[c_j \in \{c_i - m + 1, \ldots, c_i + m - 1\}] = \frac{2m-1}{N}\;.\end{aligned}$$

Then by the Union bound

$$\begin{aligned}\mathrm{IIP}(N,q,m) &= \Pr[\textstyle\bigvee_{1\leq i<j\leq q}] \leq \displaystyle\sum_{1\leq i<j\leq q} \Pr[\mathrm{B}_{i,j}] \\ &= \binom{q}{2}\frac{2m-1}{N}\;.\end{aligned}$$

Alternative formulation of advantage

Let $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be a symmetric encryption scheme and A an adversary.

Game $\mathrm{Guess}_{\mathcal{SE}}$	**procedure LR**(M_0, M_1)
	return $C \xleftarrow{\$} \mathcal{E}_K(M_b)$
procedure Initialize	**procedure Finalize**(b')
$K \xleftarrow{\$} \mathcal{K}$; $b \xleftarrow{\$} \{0,1\}$	return $(b = b')$

Proposition: $\mathbf{Adv}^{\mathrm{ind\text{-}cpa}}_{\mathcal{SE}}(A) = 2 \cdot \Pr\left[\mathrm{Guess}^{A}_{\mathcal{SE}} \Rightarrow \mathsf{true}\right] - 1.$

Proof: Observe

$$\Pr\left[b' = 1 \mid b = 1\right] = \Pr\left[\mathrm{Right}^{A}_{\mathcal{SE}} \Rightarrow 1\right]$$

$$\Pr\left[b' = 1 \mid b = 0\right] = \Pr\left[\mathrm{Left}^{A}_{\mathcal{SE}} \Rightarrow 1\right]$$

Proof of Proposition, continued

$$\Pr\left[\text{Guess}_{\mathcal{SE}}^{A} \Rightarrow \mathsf{true}\right]$$

$$
\begin{aligned}
&= \Pr\left[b = b'\right] \\
&= \Pr\left[b = b' \mid b = 1\right] \cdot \Pr\left[b = 1\right] + \Pr\left[b = b' \mid b = 0\right] \cdot \Pr\left[b = 0\right] \\
&= \Pr\left[b = b' \mid b = 1\right] \cdot \frac{1}{2} + \Pr\left[b = b' \mid b = 0\right] \cdot \frac{1}{2} \\
&= \Pr\left[b' = 1 \mid b = 1\right] \cdot \frac{1}{2} + \Pr\left[b' = 0 \mid b = 0\right] \cdot \frac{1}{2} \\
&= \Pr\left[b' = 1 \mid b = 1\right] \cdot \frac{1}{2} + \left(1 - \Pr\left[b' = 1 \mid b = 0\right]\right) \cdot \frac{1}{2} \\
&= \frac{1}{2} + \frac{1}{2} \cdot \left(\Pr\left[b' = 1 \mid b = 1\right] - \Pr\left[b' = 1 \mid b = 0\right]\right) \\
&= \frac{1}{2} + \frac{1}{2} \cdot \left(\Pr\left[\text{Right}_{\mathcal{SE}}^{A} \Rightarrow 1\right] - \Pr\left[\text{Left}_{\mathcal{SE}}^{A} \Rightarrow 1\right]\right) \\
&= \frac{1}{2} + \frac{1}{2} \cdot \mathbf{Adv}_{\mathcal{SE}}^{\text{ind-cpa}}(A) \,.
\end{aligned}
$$

A game-playing proof

The proof uses a sequence of games and invokes the fundamental lemma of game playing [BR96].

The games have the following **Initialize** and **Finalize** procedures:

Initialize // G_0	**Initialize** // G_1, G_2, G_3	**Finalize** // All games
$b \stackrel{\$}{\leftarrow} \{0,1\}$; $S \leftarrow \emptyset$	$b \stackrel{\$}{\leftarrow} \{0,1\}$; $S \leftarrow \emptyset$	Return $(b = b')$
$K \stackrel{\$}{\leftarrow} \{0,1\}^k$		

For brevity we omit writing these procedures explicitly in the games, but you should remember they are there.

Also for brevity if G is a game and A is an adversary then we let

$$\Pr[G^A] = \Pr[G^A \Rightarrow \text{true}]$$

Games G_0, G_1 for the proof

Game G_0

procedure LR(M_0, M_1)
$C[0] \xleftarrow{\$} \{0,1\}^n$
for $i = 1, \ldots, m$ do
  $P \leftarrow C[0] + i$
  if $P \notin S$ then $\mathsf{T}[P] \leftarrow E_K(P)$
  $C[i] \leftarrow T[P] \oplus M_b[i]$
  $S \leftarrow S \cup \{P\}$
return C

Game G_1

procedure LR(M_0, M_1)
$C[0] \xleftarrow{\$} \{0,1\}^n$
for $i = 1, \ldots, m$ do
  $P \leftarrow C[0] + i$
  if $P \notin S$ then $\mathsf{T}[P] \xleftarrow{\$} \{0,1\}^\ell$
  $C[i] \leftarrow \mathsf{T}[P] \oplus M_b[i]$
  $S \leftarrow S \cup \{P\}$
return C

Then

$$\mathbf{Adv}^{\text{ind-cpa}}_{\mathcal{SE}}(A) = 2 \cdot \Pr\left[G_0^A\right] - 1$$

Analysis

Clearly $\Pr[G_0^A] = \Pr[G_1^A] + \left(\Pr[G_0^A] - \Pr[G_1^A]\right)$.

Claim 1: We can design prf-adversary B so that

$$\Pr[G_0^A] - \Pr[G_1^A] \leq \mathbf{Adv}_E^{\mathrm{prf}}(B)$$

Claim 2: $\Pr[G_1^A] \leq \dfrac{1}{2} + \dfrac{(q-1)\sigma}{2^n}$

Given these, we have

$$\begin{aligned}\mathbf{Adv}_{\mathcal{SE}}^{\mathrm{ind\text{-}cpa}}(A) &\leq 2 \cdot \left(\frac{1}{2} + \frac{(q-1)\sigma}{2^n}\right) - 1 + 2 \cdot \mathbf{Adv}_E^{\mathrm{prf}}(B) \\ &= \frac{2(q-1)\sigma}{2^n} + 2 \cdot \mathbf{Adv}_E^{\mathrm{prf}}(B)\end{aligned}$$

which proves the theorem. It remains to prove the claims.

Proof of Claim 1

adversary B

$b \xleftarrow{\$} \{0,1\}$; $S \leftarrow \emptyset$
$b' \xleftarrow{\$} A^{\mathrm{LRSim}}$
if $(b = b')$ then return 1
else return 0

subroutine $\mathrm{LRSim}(M_0, M_1)$
$C[0] \xleftarrow{\$} \{0,1\}^n$
for $i = 1, ..., m$ do
  $P \leftarrow C[0] + i$
  if $P \notin S$ then $T[P] \leftarrow \mathbf{Fn}(P)$
  $C[i] \leftarrow T[P] \oplus M_b[i]$
  $S \leftarrow S \cup \{P\}$
return C

$$\Pr\left[\mathrm{Real}_E^B \Rightarrow 1\right] = \Pr\left[G_0^A\right]$$

$$\Pr\left[\mathrm{Rand}_{\{0,1\}^n}^B \Rightarrow 1\right] = \Pr\left[G_1^A\right]$$

Subtracting, we get Claim 1.

Where we are

It remains to prove:

Claim 2: $\Pr[G_1^A] \leq \frac{1}{2} + \frac{(q-1)\sigma}{2^n}$

Introducing the "bad" flag

Game G_1

procedure $\mathbf{LR}(M_0, M_1)$
$C[0] \xleftarrow{\$} \{0,1\}^n$
for $i = 1, \ldots, m$ do
  $P \leftarrow C[0] + i$
  If $P \notin S$ then
    $\mathrm{T}[P] \xleftarrow{\$} \{0,1\}^\ell$
  $C[i] \leftarrow \mathrm{T}[P] \oplus M_b[i]$
  $S \leftarrow S \cup \{P\}$
return C

Game $\boxed{G_2}$, G_3

procedure $\mathbf{LR}(M_0, M_1)$
$C[0] \xleftarrow{\$} \{0,1\}^n$
for $i = 1, \ldots, m$ do
  $P \leftarrow C[0] + i$
  $C[i] \xleftarrow{\$} \{0,1\}^\ell$
  If $P \in S$ then
    bad $\leftarrow$ true ; $\boxed{C[i] \leftarrow \mathrm{T}[P] \oplus M_b[i]}$
  $\mathrm{T}[P] \leftarrow C[i] \oplus M_b[i]$
  $S \leftarrow S \cup \{P\}$
return C

$$\Pr[G_1^A] = \Pr[G_2^A] = \Pr[G_3^A] + \left(\Pr[G_2^A] - \Pr[G_3^A]\right)$$

Analysis of G_3

Game G_3

procedure LR(M_0, M_1)
$C[0] \xleftarrow{\$} \{0,1\}^n$
for $i = 1, \ldots, m$ do
  $P \leftarrow C[0] + i$; $C[i] \xleftarrow{\$} \{0,1\}^\ell$
  If $P \in S$ then $\mathsf{bad} \leftarrow \mathsf{true}$
  $\mathrm{T}[P] \leftarrow C[i] \oplus M_b[i]$; $S \leftarrow S \cup \{P\}$
return C

Ciphertext C in G_3 is always random, independently of b, so

$$\Pr\left[G_3^A\right] = \frac{1}{2}.$$

Fundamental Lemma of game playing

Games G, H are identical-until-bad if their code differs only in statements following the setting of bad to true.

Lemma: [BR96] If G, H are identical-until-bad, then for any adversary A and any y

$$\left|\Pr\left[G^A \Rightarrow y\right] - \Pr\left[H^A \Rightarrow y\right]\right| \leq \Pr\left[H^A \,\mathrm{sets}\, \mathsf{bad}\right]$$

Using the fundamental lemma

Game $\boxed{G_2}$, G_3

procedure LR(M_0, M_1)
$C[0] \xleftarrow{\$} \{0,1\}^n$
for $i = 1, ..., m$ do
  $P \leftarrow C[0] + i$; $C[i] \xleftarrow{\$} \{0,1\}^\ell$
  If $P \in S$ then
    bad $\leftarrow$ true ; $\boxed{C[i] \leftarrow \mathsf{T}[P] \oplus M_b[i]}$
  $\mathsf{T}[P] \leftarrow C[i] \oplus M_b[i]$; $S \leftarrow S \cup \{P\}$
return C

G_2 and G_3 are identical-until-bad, so Fundamental Lemma implies

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

Bounding the probability that G_3 sets bad

Game G_3

procedure LR(M_0, M_1)
$C[0] \xleftarrow{\$} \{0,1\}^n$
for $i = 1, \ldots, m$ do
  $P \leftarrow C[0] + i$; $C[i] \xleftarrow{\$} \{0,1\}^\ell$
  If $P \in S$ then $\mathsf{bad} \leftarrow \mathsf{true}$
  $\mathrm{T}[P] \leftarrow C[i] \oplus M_b[i]$; $S \leftarrow S \cup \{P\}$
return C

$$\begin{aligned}\Pr\left[G_3^A \text{ sets bad}\right] &\leq \mathrm{IIP}(2^n, q, m) \leq \frac{q(q-1)}{2}\,\frac{2m-1}{2^n} \\ &\leq \frac{mq(q-1)}{2^n} \\ &\leq \frac{(q-1)\sigma}{2^n}\,.\end{aligned}$$

Security of CTR$

Theorem: [BDJR97] Let $E : \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^\ell$ be a family of functions and $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ the corresponding CTR$ symmetric encryption scheme. Let A be an ind-cpa adversary against $\mathcal{SE}$ that has running time t and makes at most q **LR** queries, the messages across them totaling at most σ blocks. Then there is a prf-adversary B against E such that

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

Furthermore, B makes at most σ oracle queries and has running time $t + \Theta(\sigma \cdot (n + \ell))$.

Security of CBC$

Let $E : \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$ be a block cipher and $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ the corresponding CBC$ symmetric encryption scheme.

Exercise: Give an adversary A that makes q **LR**-queries, each consisting of two 1-block messages, and achieves

$$\mathbf{Adv}^{\text{ind-cpa}}_{\mathcal{SE}}(A) = \Omega\left(\frac{q^2}{2^n}\right)$$

The running time of A should be about $\mathcal{O}(qn \cdot \log(qn))$.

Can you generalize this to m-block messages? What is the best advantage you can get?

Theorem: [BDJR97] Let $E : \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^n$ be a block cipher and $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ the corresponding CBC$ symmetric encryption scheme. Let A be an ind-cpa adversary against $\mathcal{SE}$ that has running time t and makes at most q **LR** queries, the messages across them totaling at most σ blocks. Then there is a prf-adversary B against E such that

$$\mathbf{Adv}^{\mathrm{ind\text{-}cpa}}_{\mathcal{SE}}(A) \leq 2 \cdot \mathbf{Adv}^{\mathrm{prf}}_{E}(B) + \frac{\sigma^2}{2^n}$$

Furthermore, B makes at most σ oracle queries and has running time $t + \Theta(\sigma \cdot n)$.

Exercise: Prove the above theorem.

Exercise

You are hired at a top company with an extravagant salary. Your boss asks you how secure is CBC$ based on AES. Give a clear and full answer which includes an explanation of security metrics, their relative merits, attacks and proofs. This should include an interpretation of the theorem we just saw. Your description should cover both the value and the limitations of this theorem and give a realistic picture of security aimed at someone with little understanding of cryptography.

Have a friend play the role of boss and make this a conversation. Your friend should be critical and curious about what you say.

Exercise

In CTR\$ and CBC\$ based on a family of functions $E\colon \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^\ell$, the message must have length a positive multiple of ℓ. Specify —by giving clear pseudocode for all three algorithms— a symmetric encryption scheme $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ based on E in which the message M can be a string of arbitrary length. Encryption and decryption should use at most $\lceil |M|/\ell \rceil$ applications of E or E^{-1} and the length of the ciphertext should be $|M| + n$. The scheme should achieve IND-CPA with the same kinds of bounds as in the theorems about CTR\$ and CBC\$. You may assume E is a block cipher if needed.