CSL759: Cryptography and Computer Security

Ragesh Jaiswal

CSE, IIT Delhi

CPA Security

- Until now, we have seen encryption schemes that are secure in some limited sense:
 - One-time encryption
 - Ciphertext-only adversary.
- We would now like to transition to stronger notions of security for symmetric encryption schemes that allows multiple encryptions and where the adversary can obtain encryptions of its choice (CPA security).
- Pseudorandom function (PRF) and Pseudorandom Permutation (PRP) are Cryptographic primitives that help us to design such schemes that are "CPA-secure".

Pseudorandom Function (PRF)

- We consider functions of the form $F: \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$.
- These are called *keyed* functions since we have a *collection* of 2^k functions, one for each value of the key $K \in \{0,1\}^k$. This function is denoted by $F_K: \{0,1\}^n \to \{0,1\}^n$ and is defined as $F_K(x) = F(K, x)$.
- This collection of functions is also known as a *function family*.
- We will use such function families as a primitive in designing symmetric encryption schemes that are CPA-secure.
- Th useful security notion for this primitive is how similar this family is to the family of random functions from {0,1}ⁿ to {0,1}ⁿ.

- Th useful security notion for this primitive is how similar this family is to the family of random functions from {0,1}ⁿ to {0,1}ⁿ.
- For this, we define the following two *Experiments* and then compare the bahavior of adversaries in these two experiments.

• Real_{A,F}

- Randomly pick $K \leftarrow \{0,1\}^k$.
- When A queries with an input $x \in \{0,1\}^n$, return $F_K(x)$.
- Finally *A* outputs a bit *b*.
- Output *b*.

• Random_A

- Pick a random function f from $\{0,1\}^n$ to $\{0,1\}^n$.
- When A queries with an input $x \in \{0,1\}^n$, return f(x).
- Finally *A* outputs a bit *b*.
- Output *b*.

- Th useful security notion for this primitive is how similar this family is to the family of random functions from {0,1}ⁿ to {0,1}ⁿ.
- For this, we define the following two *Experiments* and then compare the bahavior of adversaries in these two experiments.
- Why did we not have to define these "experiments" while discussing the security of PRGs?

• Real_{A,F}

- Randomly pick $K \leftarrow \{0,1\}^k$.
- When A queries with an input $x \in \{0,1\}^n$, return $F_K(x)$.
- Finally *A* outputs a bit *b*.
- Output *b*.

• Random_A

- Pick a random function f from $\{0,1\}^n$ to $\{0,1\}^n$.
- When A queries with an input $x \in \{0,1\}^n$, return f(x).
- Finally *A* outputs a bit *b*.
- Output *b*.

- Th useful security notion for this primitive is how similar the family is to the family of random functions from {0,1}ⁿ to {0,1}ⁿ.
- For this, we define the following two *Experiments* and then compare the bahavior of adversaries in these two experiments.

• Real_{A,F}

- Randomly pick $K \leftarrow \{0,1\}^k$.
- When A queries with an input $x \in \{0,1\}^n$, return $F_K(x)$.
- Finally *A* outputs a bit *b*.
- Output *b*.

• Random_A

- When A queries with an input $x \in \{0,1\}^n$, return a random value from $\{0,1\}^n$.
- Finally *A* outputs a bit *b*.
- Output *b*.

The adversary is not allowed to repeat a query.

- Th useful security notion for this primitive is how similar the family is to the family of random functions from {0,1}ⁿ to {0,1}ⁿ.
- The PRF advantage of an adversary A is defined as follows: $Adv_{PRF}(A,F) = |\Pr[Real_{A,F} = 1] - \Pr[Random_A = 1]|$
- Real_{A,F}
 - Randomly pick $K \leftarrow \{0,1\}^k$.
 - When A queries with an input $x \in \{0,1\}^n$, return $F_K(x)$.
 - Finally *A* outputs a bit *b*.
 - Output *b*.

• Random_A

- When A queries with an input $x \in \{0,1\}^n$, return a random value from $\{0,1\}^n$.
- Finally *A* outputs a bit *b*.
- Output *b*.

The adversary is not allowed to repeat a query.

- The PRF advantage of an adversary A is defined as follows: $Adv_{PRF}(A,F) = |\Pr[Real_{A,F} = 1] - \Pr[Random_A = 1]|$
- A function F: {0,1}^k × {0,1}ⁿ → {0,1}ⁿ is called (t, q, ε)-secure PRF if for every adversary A that runs in time ≤ t and asks ≤ q queries, Adv_{PRF}(A, F) ≤ ε.
- Real_{A,F}
 - Randomly pick $K \leftarrow \{0,1\}^k$.
 - When A queries with an input $x \in \{0,1\}^n$, return $F_K(x)$.
 - Finally *A* outputs a bit *b*.
 - Output *b*.

• Random_A

- When A queries with an input $x \in \{0,1\}^n$, return a random value from $\{0,1\}^n$.
- Finally *A* outputs a bit *b*.
- Output *b*.

The adversary is not allowed to repeat a query.

- The PRF advantage of an adversary A is defined as follows: $Adv_{PRF}(A,F) = |\Pr[Real_{A,F} = 1] - \Pr[Random_A = 1]|$
- A function F: {0,1}^k × {0,1}ⁿ → {0,1}ⁿ is called (t, q, ε)-secure PRF if for every adversary A that runs in time ≤ t and asks ≤ q queries, Adv_{PRF}(A, F) ≤ ε.
- We can define asymptotic security for *length-preserving functions*, *F*: {0,1}* × {0,1}* → {0,1}*, where the length of the key, input, and output are the same.
 - Such a function is called a secure pseudorandom function (or just PRF) if for every adversary A that runs in polynomial time, and makes polynomial number of queries, there is a negligible function negl such that $Adv_{PRF}(A,F) \leq negl(k)$.

- Borrowing ideas from one-time, ciphertext-only attack scenario, we can try to use message-indistinguishability as our notion of security.
- What is the main issue with this idea?

- Borrowing ideas from one-time, ciphertext-only attack scenario, we can try to use message-indistinguishability as our notion of security.
- What is the main issue with this idea?
 - In CPA, the adversary is allowed multiple encryptions of messages of its choice.

- Borrowing ideas from one-time, ciphertext-only attack scenario, we can try to use message-indistinguishability as our notion of security.
- What is the main issue with this idea?
 - In CPA, the adversary is allowed multiple encryptions of messages of its choice.
- How do we define security then?

- Borrowing ideas from one-time, ciphertext-only attack scenario, we can try to use message-indistinguishability as our notion of security.
- A symmetric encryption scheme SE = (E, D) is said to be IND-CPA insecure if an efficient adversary is able to figure out which world it is in.

• Left_{SE,A}

- Randomly pick key $K \leftarrow \{0,1\}^k$.
- When A queries message pair (M_0^i, M_1^i) return $E(M_0^i)$ to A.
- Finally *A* outputs *b*.
- Output *b*.

• Right_{SE,A}

- Randomly pick key $K \leftarrow \{0,1\}^k$.
- When A queries message pair (M_0^i, M_1^i) return $E_K(M_1^i)$ to A.
- Finally *A* outputs *b*.
- Output *b*.

- Borrowing ideas from one-time, ciphertext-only attack scenario, we can try to use message-indistinguishability as our notion of security.
- A symmetric encryption scheme SE = (E, D) is said to be IND-CPA insecure if an efficient adversary is able to figure out which world it is in.

• Left_{SE,A}

- Randomly pick key $K \leftarrow \{0,1\}^k$.
- When A queries message pair (M_0^i, M_1^i) return $E_K(M_0^i)$ to A.
- Finally *A* outputs *b*.
- Output *b*.

• Right_{SE,A}

- Randomly pick key $K \leftarrow \{0,1\}^k$.
- When A queries message pair (M_0^i, M_1^i) return $E_K(M_1^i)$ to A.
- Finally *A* outputs *b*.
- Output *b*.
- The IND-CPA advantage of an adversary A is defined as follows: $Adv_{ind-cpa}(A, SE) = \left| \Pr[Left_{SE,A} = 1] - \Pr[Right_{SE,A} = 1] \right|$

- $Left_{SE,A}$
 - Randomly pick key $K \leftarrow \{0,1\}^k$.
 - When A queries message pair (M_0^i, M_1^i) return $E_K(M_0^i)$ to A.
 - Finally *A* outputs *b*.
 - Output *b*.

- Right_{SE,A}
 - Randomly pick key $K \leftarrow \{0,1\}^k$.
 - When A queries message pair (M_0^i, M_1^i) return $E_K(M_1^i)$ to A.
 - Finally *A* outputs *b*.
 - Output *b*.
- The IND-CPA advantage of an adversary *A* is defined as follows:

 $Adv_{ind-cpa}(A,SE) = \left| \Pr[Left_{SE,A} = 1] - \Pr[Right_{SE,A} = 1] \right|$

• A symmetric encryption scheme SE = (E, D) is called (t, q, ϵ) ind-cpa secure if for every adversary A that runs in time $\leq t$ and asks $\leq q$ quesries, $Adv_{ind-cpa}(A, SE) \leq \epsilon$.

• The IND-CPA advantage of an adversary *A* is defined as follows:

 $Adv_{ind-cpa}(A,SE) = \left| \Pr[Left_{SE,A} = 1] - \Pr[Right_{SE,A} = 1] \right|$

- A symmetric encryption scheme SE = (E, D) is called (t, q, ϵ) ind-cpa secure if for every adversary A that runs in time $\leq t$ and asks $\leq q$ queries, $Adv_{ind-cpa}(A, SE) \leq \epsilon$.
- A symmetric encryption scheme SE = (E, D) is said to be indcpa secure if for every adversary A that runs in polynomial time and makes polynomial number of queries, there exist a negligible function negl such that $Adv_{ind-cpa}(A, SE) \leq negl(k)$.

- IND-CPA allows adversaries to make multiple queries.
- How much advantage do adversaries who is allowed to ask q > 1 queries, have over adversaries who can only make 1 "left/right" query?

- IND-CPA allows adversaries to make multiple queries.
- How much advantage do adversaries who is allowed to ask q > 1 queries, have over adversaries who can only make 1 query?

• Left'_{SE,A}

- Randomly pick key $K \leftarrow \{0,1\}^k$.
- When A queries challenge message pair (M_0, M_1) return $E_K(M_0)$ to A.
- When A queries a message M^j , then return $E_K(M^j)$ to A
- Finally A outputs b.

• Output *b*.

• Right'_{SE,A}

- Randomly pick key $K \leftarrow \{0,1\}^k$.
- When A queries challenge message pair (M_0, M_1) return $E_K(M_1)$ to A.
- When A queries a message M^j , then return $E_K(M^j)$ to A
- Finally *A* outputs *b*.
- Output *b*.
- The FTG-CPA advantage of an adversary A is defined as follows: $Adv_{ftg-cpa}(A, SE) = |\Pr[Left'_{SE,A} = 1] - \Pr[Right'_{SE,A} = 1]|$
- A symmetric encryption scheme SE = (E, D) is called (t, q, ϵ) -ftg-cpa secure if for every adversary A that runs in time $\leq t$ and asks $\leq q$ quesries, $Adv_{ftg-cpa}(A, SE) \leq \epsilon$.

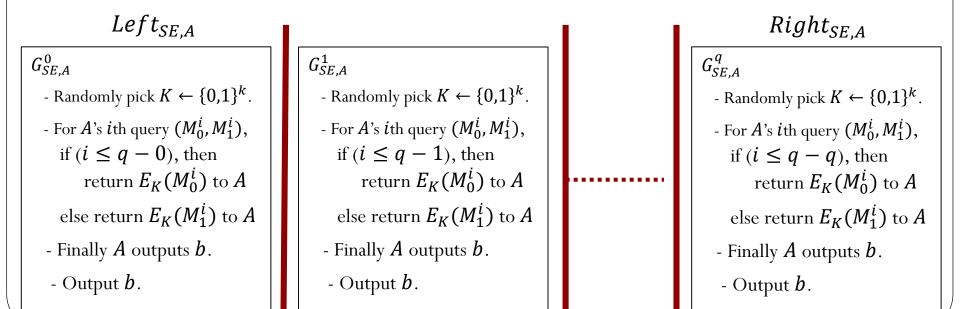
- IND-CPA allows adversaries to make multiple queries.
- How much advantage do adversaries who is allowed to ask q > 1 queries, have over adversaries who can only make 1 query?
- <u>Theorem</u>: If a symmetric encryption scheme SE = (E, D) is (t, q, ϵ) -ftg-cpa secure, then SE is also $(t, q, \epsilon \cdot q)$ -ind-cpa secure.
 - We prove the following: Let *A* be any ind-cpa adversary that runs in time *t* and makes *q* queries, then there exists an ftg-cpa adversary that runs in time *t* and makes *q* queries such that

 $Adv_{ind-cpa}(A,SE) \leq q \cdot Adv_{ftg-cpa}(B,SE).$

• <u>Theorem</u>: Let *A* be any ind-cpa adversary that runs in time *t* and makes *q* queries, then there exists an ftg-cpa adversary that runs in time *t* and makes *q* queries such that

$$Adv_{ind-cpa}(A,SE) \le q \cdot Adv_{ftg-cpa}(B,SE).$$

• To prove this, we define hybrid experiments.



<u>Theorem</u>: Let A be any ind-cpa adversary that runs in time t and makes q queries, then there exists an ftg-cpa adversary that runs in time t and makes q queries such that

 $Adv_{ind-cpa}(A,SE) \leq q \cdot Adv_{ftg-cpa}(B,SE).$

•
$$Adv_{ind-cpa}(A, SE) = |\Pr[Left_{SE,A} = 1] - \Pr[Right_{SE,A} = 1]|$$

 $= |\Pr[G_{SE,A}^0 = 1] - \Pr[G_{SE,A}^q = 1]|$
• Let $P_0 = \Pr[G_{SE,A}^0 = 1], P_1 = \Pr[G_{SE,A}^1 = 1], \dots, P_q = \Pr[G_{SE,A}^q = 1]$
Left_{SE,A}
• Randomly pick $K \leftarrow \{0,1\}^k$.
• For A's ith query (M_0^i, M_1^i) ,
if $(i \le q - 0)$, then
return $E_K(M_0^i)$ to A
else return $E_K(M_1^i)$ to A
 $|\operatorname{For} K_K(M_1^i)$ to A
 $|\operatorname{For} K_K(M_1^i)$ to A

- Finally *A* outputs *b*.

- Output *b*.

- Finally *A* outputs *b*.

- Output **b**.

- Finally *A* outputs *b*.

- Output **b**.

 $G_{SE,A}^0$

• <u>Theorem</u>: Let A be any ind-cpa adversary that runs in time t and makes q queries, then there exists an ftg-cpa adversary that runs in time t and makes q queries such that

 $Adv_{ind-cpa}(A,SE) \leq q \cdot Adv_{ftg-cpa}(B,SE).$

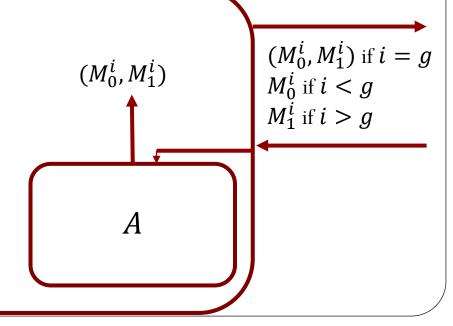
•
$$Adv_{ind-cpa}(A, SE) = |\Pr[Left_{SE,A} = 1] - \Pr[Right_{SE,A} = 1]|$$

= $|\Pr[G_{SE,A}^0 = 1] - \Pr[G_{SE,A}^q = 1]|$

• Let
$$P_0 = \Pr[G_{SE,A}^0 = 1], P_1 = \Pr[G_{SE,A}^1 = 1], \dots, P_q = \Pr[G_{SE,A}^q = 1]$$

В

- Pick $g \leftarrow [q]$ randomly
- When A makes its i^{th} query (M_0^i, M_1^i) :
 - If (i < g) make a query with M_0^i and return the value to A
 - If (i > g) make a query with M_1^i and return the value to A
 - If (i = g) make a query (M_0^i, M_1^i) and return the value to A
- Output *A*'s result



• <u>Theorem</u>: Let A be any ind-cpa adversary that runs in time t and makes q queries, then there exists an ftg-cpa adversary that runs in time t and makes q queries such that

 $Adv_{ind-cpa}(A,SE) \leq q \cdot Adv_{ftg-cpa}(B,SE).$

•
$$Adv_{ind-cpa}(A, SE) = |\Pr[Left_{SE,A} = 1] - \Pr[Right_{SE,A} = 1]|$$

= $|\Pr[G_{SE,A}^0 = 1] - \Pr[G_{SE,A}^q = 1]|$

- Let $P_0 = \Pr[G_{SE,A}^0 = 1], P_1 = \Pr[G_{SE,A}^1 = 1], \dots, P_q = \Pr[G_{SE,A}^q = 1]$
- $\Pr[Left'_{SE,B} = 1] = ?$

•
$$\Pr[Right'_{SE,B} = 1] = ?$$

• <u>Theorem</u>: Let A be any ind-cpa adversary that runs in time t and makes q queries, then there exists an ftg-cpa adversary that runs in time t and makes q queries such that

 $Adv_{ind-cpa}(A,SE) \leq q \cdot Adv_{ftg-cpa}(B,SE).$

•
$$Adv_{ind-cpa}(A, SE) = |\Pr[Left_{SE,A} = 1] - \Pr[Right_{SE,A} = 1]|$$

= $|\Pr[G_{SE,A}^0 = 1] - \Pr[G_{SE,A}^q = 1]|$

• Let
$$P_0 = \Pr[G_{SE,A}^0 = 1], P_1 = \Pr[G_{SE,A}^1 = 1], \dots, P_q = \Pr[G_{SE,A}^q = 1]$$

•
$$\Pr[Left'_{SE,B} = 1] = \frac{1}{q} \cdot (P_0 + P_1 + \dots + P_{q-1})$$

•
$$\Pr[Right'_{SE,B} = 1] = \frac{1}{q} \cdot (P_1 + P_2 + ... + P_q)$$

•
$$Adv_{ftg-cpa}(B,SE) = |\Pr[Left'_{SE,B} = 1] - \Pr[Right'_{SE,B} = 1]|$$

 $= \frac{1}{q} \cdot |((P_0 - P_q))|$
 $= \frac{1}{q} \cdot Adv_{ind-cpa}(A,SE)$

- Alternate definition of FTG-CPA security.
- GuessLR_{SE,A}
 - Randomly pick a key $K \leftarrow \{0,1\}^n$.
 - Pick a random bit $b \leftarrow \{0,1\}$
 - When A makes a encryption query M^i , return the value $E_K(M^i)$.
 - When A makes the challenge query (M_0, M_1) , return the value $E_K(M_b)$.
 - Finally, A outputs a bit b'
 - If (b = b') output 1 else output 0
- <u>Theorem</u>: $\Pr[GuessLR_{SE,A} = 1] = \frac{1}{2} \pm \frac{1}{2} \cdot Adv_{ftg-cpa}(A, SE)$

- Alternate definition of FTG-CPA security.
- GuessLR_{SE,A}
 - Randomly pick a key $K \leftarrow \{0,1\}^n$.
 - Pick a random bit $b \leftarrow \{0,1\}$
 - When A makes a encryption query M^i , return the value $E_K(M^i)$.
 - When A makes the challenge query (M_0, M_1) , return the value $E_K(M_b)$.
 - Finally, A outputs a bit b'
 - If (b = b') output 1 else output 0
- <u>Theorem</u>: $\Pr[GuessLR_{SE,A} = 1] = \frac{1}{2} \pm \frac{1}{2} \cdot Adv_{ftg-cpa}(A, SE)$
- So, summing up all the discussion until now, for CPA-security of an encryption scheme, we just need to analyse the performance of an adversary in the experiment *GuessLR*_{SE,A}.

- Suppose we have a *secure* pseudorandom permutation family $F: \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$.
 - We saw a few examples(AES, 3DES etc.) in the last lecture.
- Consider the following encryption scheme SE = (E, D)that encrypts messages of length n.
 - $E_K(M) = F_K(M)$ and $D_K(C) = F_K^{-1}(C)$
- Is *SE* ind-cpa secure?
- Is *SE* ftg-cpa secure?
- Is *SE* "GuessLR" secure?

- Suppose we have a *secure* pseudorandom permutation family $F: \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$.
 - We saw a few examples(AES, 3DES etc.) in the last lecture.
- Consider the following encryption scheme SE = (E, D)that encrypts messages of length n.
 - $E_K(M) = F_K(M)$ and $D_K(C) = F_K^{-1}(C)$
- Is *SE* "GuessLR" secure?
 - No
 - Adversary *A*
 - Query the message 0^n and get back $C = E_K(0^n)$.
 - Make the challenge query $(0^n, 1^n)$ and get back C'.
 - If (C = C'), then output 0 else output 1
 - $\Pr[GuessLR_{SE,A} = 1] = ?$

- Suppose we have a *secure* pseudorandom permutation family $F: \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$.
 - We saw a few examples(AES, 3DES etc.) in the last lecture.
- Consider the following encryption scheme SE = (E, D) that encrypts messages of length n.
 - $E_K(M) = F_K(M)$ and $D_K(C) = F_K^{-1}(C)$
- Is *SE* "GuessLR" secure?
 - No
 - Adversary *A*
 - Query the message 0^n and get back $C = E_K(0^n)$.
 - Make the challenge query $(0^n, 1^n)$ and get back C'.
 - If (C = C'), then output 0 else output 1
 - $\Pr[GuessLR_{SE,A} = 1] = 1$

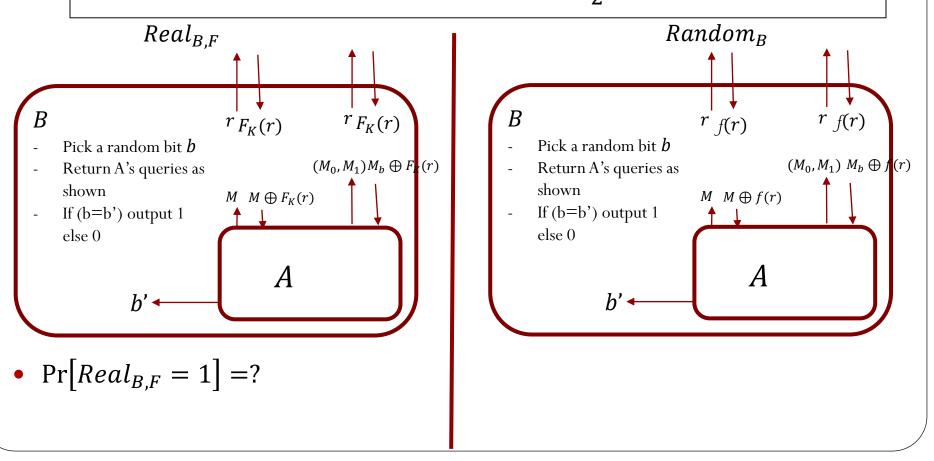
- Suppose we have a *secure* pseudorandom permutation family $F: \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$.
 - We saw a few examples(AES, 3DES etc.) in the last lecture.
- Consider the following encryption scheme SE = (E, D)that encrypts messages of length n.

• $E_K(M) = F_K(M)$ and $D_K(C) = F_K^{-1}(C)$

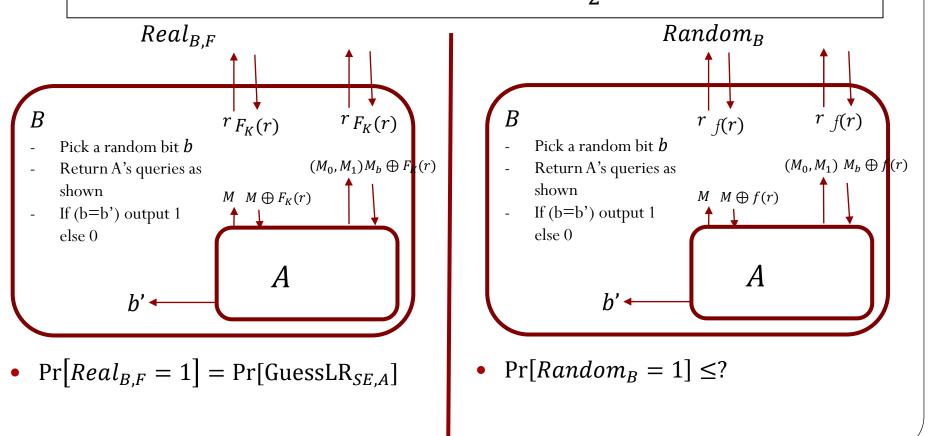
- In fact, any deterministic encryption scheme cannot be IND-CPA secure!
 - For *SE* to be IND-CPA secure, everytime you encrypt a message *M*, you should get a different ciphertext!

- Suppose we have a *secure* pseudorandom permutation family $F: \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$.
 - We saw a few examples(AES, 3DES etc.) in the last lecture.
- Consider the following encryption scheme SE = (E, D) that encrypts messages of length n.
 - $E_K(M)$
 - Pick a random $r \leftarrow \{0,1\}^n$
 - Output $C = \langle r, F_K(r) \bigoplus M \rangle$
 - $D_K(C)$
 - Parse *C* as < *r*, *s* >
 - Output $M = F_K(r) \bigoplus s$
- <u>Theorem</u>: If *F* is $\left(2t, q, \frac{\epsilon}{2} \frac{q}{2^n}\right)$ -secure PRF, then *SE* is (t, q, ϵ) -ftg-cpa secure symmetric encryption scheme.

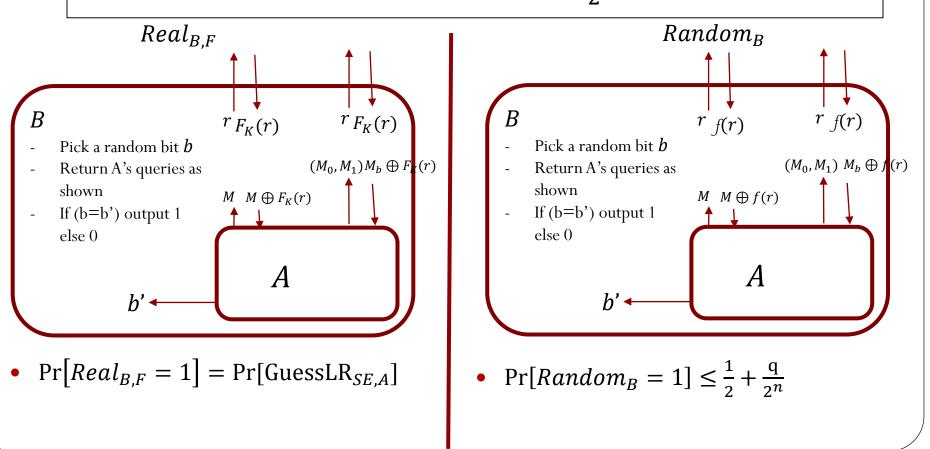
• <u>Theorem</u>: Consider an adversary A that runs in time t, makes qqueries such that $\Pr[GuessLR_{SE,A} = 1] > \frac{1}{2} + \epsilon$, then there is an adversary B that runs in time at most 2t, makes (q + 1)queries such that $Adv_{PRF}(B,F) > \epsilon - \frac{q}{2^n}$.



• <u>Theorem</u>: Consider an adversary A that runs in time t, makes qqueries such that $\Pr[GuessLR_{SE,A} = 1] > \frac{1}{2} + \epsilon$, then there is an adversary B that runs in time at most 2t, makes (q + 1)queries such that $Adv_{PRF}(B,F) > \epsilon - \frac{q}{2^n}$.



• <u>Theorem</u>: Consider an adversary A that runs in time t, makes qqueries such that $\Pr[GuessLR_{SE,A} = 1] > \frac{1}{2} + \epsilon$, then there is an adversary B that runs in time at most 2t, makes (q + 1)queries such that $Adv_{PRF}(B,F) > \epsilon - \frac{q}{2^n}$.



End