MH4311 Cryptography

Lecture 8 Block Cipher

Part 5: Attacks on Block Cipher

Wu Hongjun

Lecture Outline

- **Classical ciphers**
- **Symmetric key encryption**
  - **One-time pad & information theory**
  - **Block cipher**
    - **Introduction**
    - **DES, Double DES, Triple DES**
    - **AES**
    - **Modes of Operation**
    - **Attacks: double DES, differential cryptanalysis, linear cryptanalysis**
  - **Stream cipher**
- **Hash function and Message Authentication Code**
- **Public key encryption**
- **Digital signature**
- **Key establishment and management**
- **Introduction to other cryptographic topics**

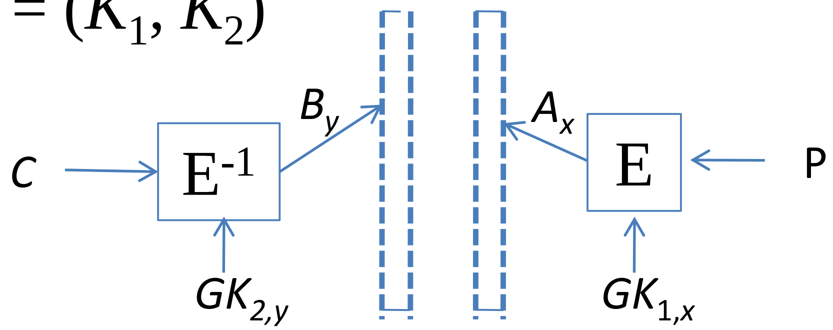
Recommended Reading

- CTP Section 3.3, 3.4
- Wikipedia
  - Meet-in-the-middle attack
    http://en.wikipedia.org/wiki/Meet-in-the-middle_attack
  - Differential cryptanalysis
    http://en.wikipedia.org/wiki/Differential_cryptanalysis
  - Linear cryptanalysis
    http://en.wikipedia.org/wiki/Linear_cryptanalysis

Meet-in-the-middle attack on double DES

- Double DES: $C = E_{K_2}(E_{K_1}(P))$
- Attack (Given a plaintext P and ciphertext C)
  - Re-write the above equation as $E^{-1}_{K_2}(C) = E_{K_1}(P)$
  - Guess all the possible values of K_1, encrypt P and obtain a table T_1 (2^{56} elements, each element is ($GK_{1,x}$, A_x))
  - Guess all the possible values of K_2, decrypt C and obtain a table T_2 (2^{56} elements, each element is ($GK_{2,y}$, B_y))
  - Now compare those two tables: if $A_i = B_j$, then **maybe** $(GK_{1,i}\ , GK_{2,j}\ ) = (K_1, K_2)$

.



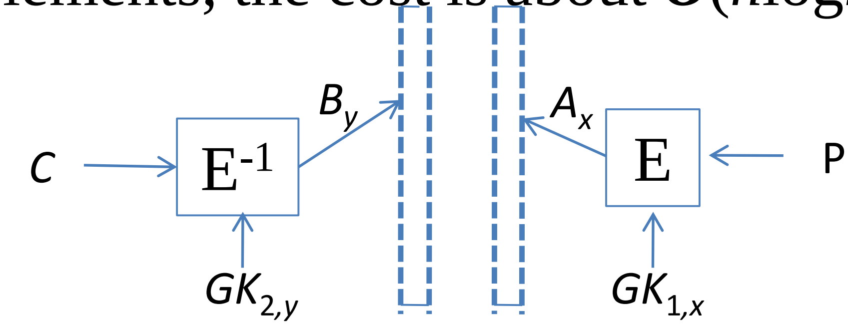
Meet-in-the-middle attack on double DES

Two tables:

T_1 (2^{56} elements, each element is (A_x, $GK_{1,x}$))

T_2 (2^{56} elements, each element is (B_y, $GK_{2,y}$))

- Q1. How to compare these two tables and find out the identical elements (i.e., to identify $A_x = B_y$) efficiently?
  - Solution: Sort those two tables first, then compare.
    - Sorting n elements, the cost is about $\boldsymbol{O}(n\log n)$



Meet-in-the-middle attack on double DES

Two tables

T_1 (2^{56} elements, each element is $(A_x, GK_{1,x})$)

T_2 (2^{56} elements, each element is $(B_y, GK_{2,y})$)

- Q2. What is the probability that $A_x = B_y$?

- Q3. How many pairs of (A_x , B_y) satisfying

  $A_x = B_y$?

Meet-in-the-middle attack on double DES

Two tables

T_1 (2^{56} elements, each element is (A_x, $GK_{1,x}$))

T_2 (2^{56} elements, each element is (B_y, $GK_{2,y}$))

- Q4. Given $A_i = B_j$, what is the probability that

$$(GK_{1,i}\ , GK_{2,\ j}\ ) = (K_1, K_2)\ \ ?$$

  - There are about $2^{112\text{-}64} = 2^{48}$ cases that $A_y = B_x$, so we now have 2^{48} possible keys, one of them is correct.
  - Try all these 2^{48} possible keys with another plaintext-ciphertext pair (P', C') to find the secret key

Cryptanalysis of Block Cipher

Two main approaches:

- Algebraic approaches
  - Solve algebraic equations
  - ………
- Statistical approaches (powerful)
  - *Differential cryptanalysis
  - *Linear cryptanalysis
  - ……..

Solve Algebraic Equations

- Algebraic equations
  - Equations over a field

- Example: two variables, two equations over GF(p)

$$x^2 + xy + y \equiv 16 \pmod{p}$$

$$x^2 + y^2 + x + y \equiv 1 \pmod{p}$$

$$x^2 + xy + y \equiv 16 \pmod{p}$$
$$x^2 + y^2 + x + y \equiv 1 \pmod{p}$$

- **How to solve the above equations?**
  - **Brute force: try all the possible values of *x* and *y***
    - **Example: *p* = 17, only need to try 17^2 possible values**
    - **Impractical if there are many variables, or large *p***
  - **Linearization**
    - **if the algebraic equations are over-defined**
      (i.e., the number of equations > the number of variables)

Linearization of Overdefined Algebraic Equations

$$x^2 + xy + y \equiv 16 \pmod{p} \quad (1)$$
$$x^2 + y^2 + x + y \equiv 1 \pmod{p} \quad (2)$$
$$2x^2 + 3xy + y \equiv 0 \pmod{p} \quad (3)$$
$$x^2 + y^2 + 4x + y \equiv 16 \pmod{p} \quad (4)$$
$$x^2 + 2xy + 5y \equiv 11 \pmod{p} \quad (5)$$

let $z_1 = x^2$, $z_2 = xy$, $z_3 = y^2$, $z_4 = x$, $z_5 = y$

$$z_1 + z_2 + z_5 \equiv 16 \pmod{p} \quad (6)$$
$$z_1 + z_3 + z_4 + z_5 \equiv 1 \pmod{p} \quad (7)$$
$$2\,z_1 + 3\,z_2 + z_5 \equiv 0 \pmod{p} \quad (8)$$
$$z_1 + z_3 + 4\,z_4 + z_5 \equiv 16 \pmod{p} \quad (9)$$
$$z_1 + 2\,z_2 + 5\,z_5 \equiv 11 \pmod{p} \quad (10)$$

Algebraic Equations Over GF(2)

Solve Algebraic Equations

- **Two basic operations over GF(2)**
  - **Addition (XOR, $\oplus$)**

    (in C programming language "^")

    $0+0 = 0$

    $0+1 = 1$

    $1+0 = 1$

    $1+1 = 0$
  - **Multiplication (AND) (in C language, "&")**

    $0 \cdot 0 = 0$

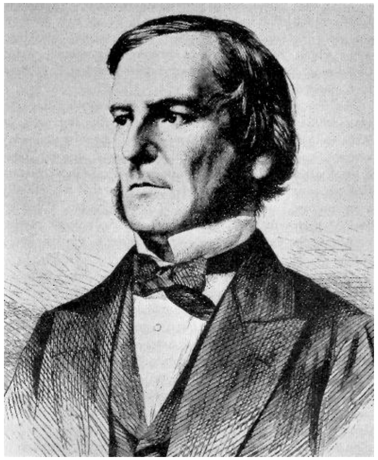
    $0 \cdot 1 = 0$

    $1 \cdot 0 = 0$

    $1 \cdot 1 = 1$

Solve Algebraic Equations

- **Foundation of digital computer: any digital computation can be carried out as computations over GF(2)**
  - **George Boole (1815-1864)**
  - **addition, multiplication, ….**



Example: two-bit A and B, three-bit D.

Let $A = A_1A_0$, $B = B_1B_0$, $D = D_2D_1D_0$,

$D = A + B$ is computed as:

$$D_0 = A_0 \oplus B_0 ,$$

$$D_1 = A_1 \oplus B_1 \oplus (A_0 \,\&\, B_0)$$

$$D_2 = (A_1 \,\&\, B_1) \oplus ((A_0 \,\&\, B_0) \oplus (A_1 \oplus B_1))$$

http://en.wikipedia.org/wiki/Adder_(electronics)

Solve Algebraic Equations

- **How to implement computations over GF(2)**
  - **Claude Shannon, 1937**
    - **Electrical relays can be used to construct logic gates to perform computations over GF(2)**
    - **“possibly the most important and famous master’s thesis in the century”**
  - **Today transistors are used to build the logic gates to perform operations over GF(2) => electronic computers**

Solve Algebraic Equations

- **Any cipher can be expressed as algebraic equations involving plaintext, ciphertext and the key**
  - **these algebraic equations are normally overdefined (since an attacker may obtain many plaintext-ciphertext pairs for a secret key)**

Solve Algebraic Equations

- **A natural approach to attack a cipher is to solve those overdefined equations**
  - **Linearization technique**
    - **How to defend: increase the number of monomials**
      - **High algebraic degree**
      - **Randomness: a lot of random monomials in the equations**
  - **Other methods**
    - **Some methods were proposed to attack AES by solving algebraic equations efficiently over GF(2) or GF(2^8), but these methods are not recognized (and not verified)**
    - **Cube attack (new)**
    - **…**

Cryptanalyis using Statistical Approach*

Statistical Approach

- **Basic idea: to find the largest statistical correlation between plaintext & ciphertext, then recover key**

- **Two powerful techniques**
  - **Differential cryptanalysis**
    - **NSA discovered this attack attack in 19??, kept secret**
    - **IBM knew this attack around 1974—1976, kept secret**
    - **Eli Biham, 1990**
  - **Linear cryptanalysis**
    - **Mitsuru Matsui, 1993**

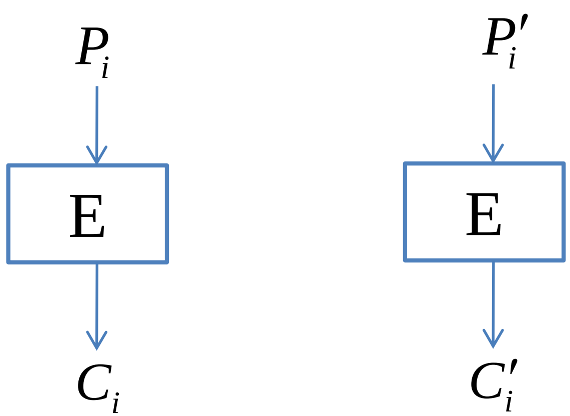
Statistical Approach

- **Differential cryptanalysis**
  - **Basic idea: if the input difference and output difference are statistically strongly correlated, differential attack can be applied!**
  - **Differential cryptanalysis is a type of chosen-plaintext attack**
    - **Chosen plaintext attack: the attacker is able to choose some plaintexts and obtain their ciphertexts**

Statistical Approach

- **Differential cryptanalysis**
  - **For a particular input difference (we consider the XOR difference between plaintexts), if the output differences are not random, then differential cryptanalysis can be launched.**

$$\Delta P = P_i \oplus P_i'$$



ΔP is a fixed value, P_i is random chosen

For a strong cipher,

ΔC_i should appear with probability about 2^{-n} for any ΔP,

otherwise, insecure.

$$\Delta C_i = C_i \oplus C_i'$$

Statistical Approach

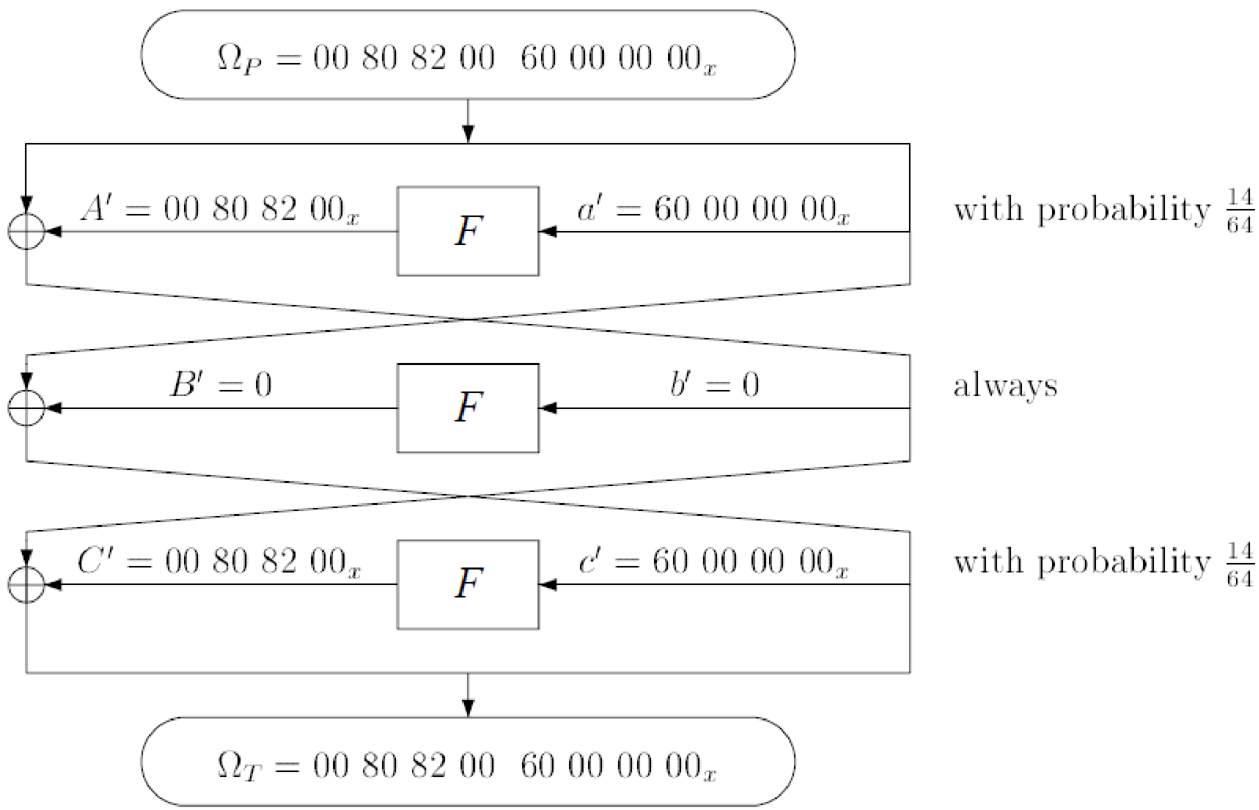
- **Differential cryptanalysis**
  - Basic steps of differential attack:
    - Suppose that $\Delta P \Rightarrow \Delta C$ with probability $p > 2^{-n}$
    - And suppose that within a cipher, the difference is propagated as follows to achieve the highest prob.:

$$\Delta P \Rightarrow \Delta_1 \Rightarrow \Delta_2 \Rightarrow \Delta_3 \cdots\cdots \Rightarrow \Delta_{r-1} \Rightarrow \Delta C$$

    - After observing 1/p ciphertext pairs,
      - we are expected to find one ciphertext pairs $C_i \oplus C_i' = \Delta C$;
      - then we know that likely for the two plaintexts $P_i \oplus P_i' = \Delta P$, the difference propagates as $\Delta P \Rightarrow \Delta_1 \cdots\cdots \Rightarrow \Delta C$
        - » We are able to attack the first round separately!

Statistical Approach

- **Differential cryptanalysis**
  - Example: differential propagation for 3-round DES



Statistical Approach

- **Differential cryptanalysis**
  - **DES**
    - **Designed to resist differential attack**
    - **2^{47} chosen plaintexts are required in the attack**
  - **AES**
    - **Strong against differential attack**

Statistical Approach

- **How to resist the differential attack**
  - **Strong Sbox**
    - **Reduce the maximal diff. prob. of each Sbox !**
    - **AES: for the Sbox, the maximal diff. prob. is 2^{-6}**
  - **Enforce the difference to propagate through many Sboxes**
    - **Diffusion should be properly designed**
      - **Example: AES**
        - » **ShiftRows + MixColumns**
        - » **An input difference passes through at least 25 Sboxes in 4 rounds**

Statistical Approach

- Linear cryptanalysis
  - Basic idea: for plaintext and ciphertext, if some input bits and output bits are statistically correlated, linear cryptanalysis may be applied!
  - We have the following linear approximation equation that involves some plaintext bits, ciphertext bits and key bits:

$$k_a + k_b + \ldots = p_i + p_j + \ldots + c_{i'} + c_{j'} + \ldots \quad \text{with prob. } p = 0.5 + x$$

(x is a small value)

    - After collecting enough plaintext-ciphertext pairs ($1/x^2$), we can obtain the equation with high confidence:

$$k_a + k_b + \ldots = 0$$

or $$k_a + k_b + \ldots = 1$$

Another way to solve overdefined nonlinear equations!

(the attack in the textbook is a bit different from the above attack)

Statistical Approach

- **Linear cryptanalysis**
  - **Example:**

DES: 5-round linear approximation with p = 0.519

$$P_H[15] \oplus P_L[7,18,24,27,28,29,30,31] \oplus C_H[15] \oplus C_L[7,18,24,27,28,29,30,31]$$
$$= K_1[42,43,45,46] \oplus K_2[22] \oplus K_4[22] \oplus K_5[42,43,45,46]. \qquad (11)$$

Statistical Approach

- **Linear cryptanalysis**
  - **DES**
    - **2^{43} known plaintexts (not that strong)**
  - **AES**
    - **Strong against linear cryptanalysis**

Statistical Approach

- **Linear cryptanalysis**
  - **Strong Sbox**
    - **Let the prob. of linear approximation be close to 0.5!**
    - **For AES Sbox, the prob. of linear approximation is within 0.5 ± 2^{-3}**
  - **Enforce the linear approximation to pass through many Sboxes**
    - **Diffusion should be properly designed**
      - **AES: at least 25 Sboxes are involved in 4-round linear approximation**

Summary

- **Meet-in-the-middle attack on double DES**

- **Attacks on block cipher**
  - **Solving algebraic equations**
  - **Statistical approach***
    - ***Differential cryptanalysis**
    - ***Linear cryptanalysis**
    - **……..**

Important for block cipher design: Sbox (confusion), diffusion