

MAS 433: Cryptography

Lecture 9
Block Cipher
(Part 5: Attacks on Block Cipher)

Wu Hongjun

Lecture Outline

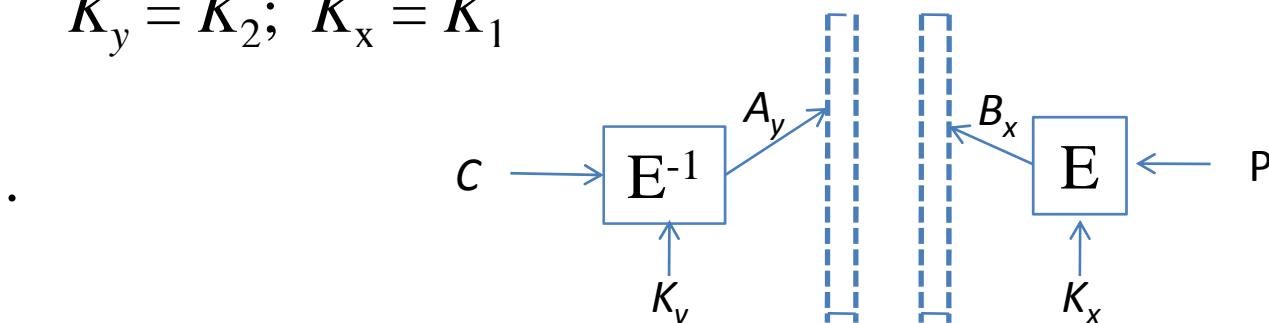
- Classical ciphers
- Symmetric key encryption
 - One-time pad & information theory
 - Block cipher
 - DES, Double DES, Triple DES
 - AES
 - Modes of Operation
 - Attacks: double DES, differential cryptanalysis, linear cryptanalysis
 - Stream cipher
- Hash function and Message Authentication Code
- Public 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

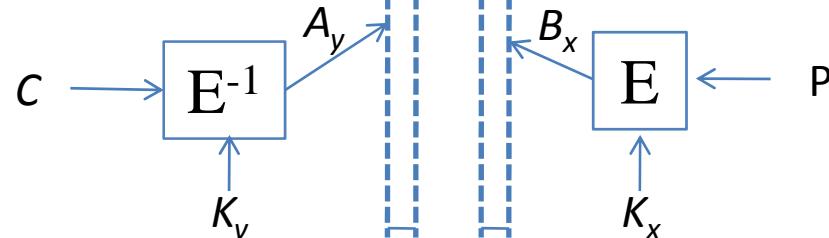
Meet-in-the-middle attack on double DES

- Double DES: $C = E_{K_2}(E_{K_1}(P))$
- Attack (Given a plaintext P and ciphertext C)
 - Re-write the above equation as $E_{K_2}^{-1}(C) = E_{K_1}(P)$
 - Guess all the possible values of K_2 , decrypt C and obtain a table T_2 (2^{56} elements, each element is (A_y, K_y))
 - Guess all the possible values of K_1 , encrypt P and obtain a table T_1 (2^{56} elements, each element is (B_x, K_x))
 - Now compare those two tables: if $A_y = B_x$, then maybe $K_y = K_2; K_x = K_1$



Meet-in-the-middle attack on double DES

- Two tables
 - T_2 (2^{56} elements, each element is (A_y, K_y))
 - T_1 (2^{56} elements, each element is (B_x, K_x))
How to compare these two tables and find out identical elements, i.e., $A_y = B_x$?
- Sort those two tables first, then compare.
 - Sorting n elements, cost is about $O(n \log n)$



Meet-in-the-middle attack on double DES

- Two tables
 - T_2 (2^{56} elements, each element is (A_y, K_y))
 - T_1 (2^{56} elements, each element is (B_x, K_x))
- If $A_y = B_x$, what is the probability that
$$K_y = K_2; K_x = K_1 ?$$
 - There are $2^{56} \times 2^{56} = 2^{112}$ (A, B) pairs
 - The chance for two random A, B to be equal is 2^{-64}
 - There are about $2^{112-64} = 2^{48}$ cases that $A_y = B_x$, so we now have 2^{48} possible keys, one of them is correct.
 - Try all these keys with another (P', C') to find the correct key

Cryptanalysis of Block Cipher

Two main approaches:

- Solve algebraic equations
- Statistical approach
 - *Differential cryptanalysis
 - *Linear cryptanalysis
 -

Solve Algebraic Equations

- Algebraic equation
 - Equation over a given field
- Example: two variables, two equations over GF(p)

$$x^2 + xy + y - 16 \mod p$$

$$x^2 + y^2 + x + y = 1 \mod p$$

$$\begin{aligned}x^2 + xy + y &= 16 \pmod{p} \\x^2 + y^2 + x + y &= 1 \pmod{p}\end{aligned}$$

- 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 for many variables/large field
 - Linearization
 - if algebraic equations are over-defined
(i.e., number of equations > number of variables)

Linearization of Overdefined Algebraic Equations

$$x^2 + xy + y = 16 \quad \text{mod } p \quad (1)$$

$$x^2 + y^2 + x + y = 1 \quad \text{mod } p \quad (2)$$

$$2x^2 + 3xy + y = 0 \quad \text{mod } p \quad (3)$$

$$x^2 + y^2 + 4x + y = 16 \quad \text{mod } p \quad (4)$$

$$x^2 + 2xy + 5y = 11 \quad \text{mod } p \quad (5)$$

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

$$z_1 + z_2 + z_5 = 16 \quad \text{mod } p \quad (1)$$

$$z_1 + z_3 + z_4 + z_5 = 1 \quad \text{mod } p \quad (2)$$

$$2z_1 + 3z_2 + z_5 = 0 \quad \text{mod } p \quad (3)$$

$$z_1 + z_3 + 4z_4 + z_5 = 16 \quad \text{mod } p \quad (4)$$

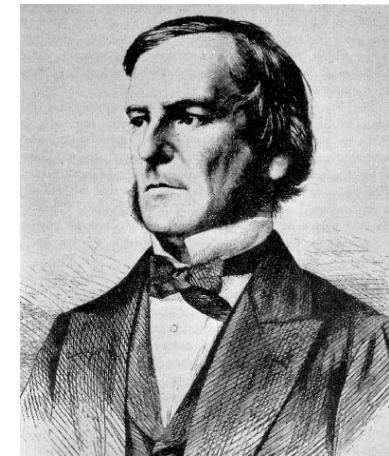
$$z_1 + 2z_2 + 5z_5 = 11 \quad \text{mod } p \quad (5)$$

Solve Algebraic Equations (contd.)

- Two basic operations over GF(2)
 - Addition (XOR) (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 (contd.)

- Basis of digital computer: any digital computation can be carried out as computations over GF(2)
 - George Boole (1815-1864)
 - addition, multiplication, table lookup
- 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)



Solve Algebraic Equations (contd.)

- Any cipher can be expressed as algebraic equations involving plaintext, ciphertext and key
 - these algebraic equations are normally overdefined (since an attacker may obtain many plaintexts, ciphertexts)
- 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(2⁸), but not recognized (and not verified)
 - Any more ?

Statistical Approach

- Basic idea: to find strong correlation between plaintext & ciphertext, then recover key
- Two basic and powerful techniques
 - Differential cryptanalysis
 - NSA, 19??, kept secret
 - IBM, around 1974--1976, kept secret
 - Eli Biham, 1990
 - Linear cryptanalysis
 - Mitsuru Matsui, 1993

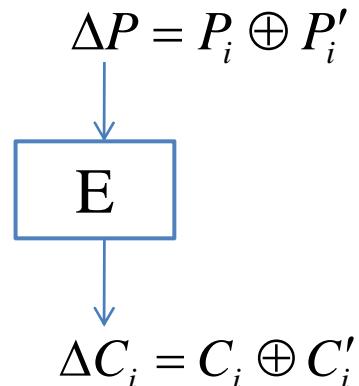


Statistical Approach (contd.)

- 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 (contd.)

- Differential cryptanalysis
 - For a particular difference between plaintexts, there are many plaintext pairs with this difference. Now if the distribution of those ciphertext differences are not random, then differential cryptanalysis can be launched.



ΔP is fixed, P_i is random chosen

For a strong cipher,

ΔC_i should appear with probability about 2^{-n}

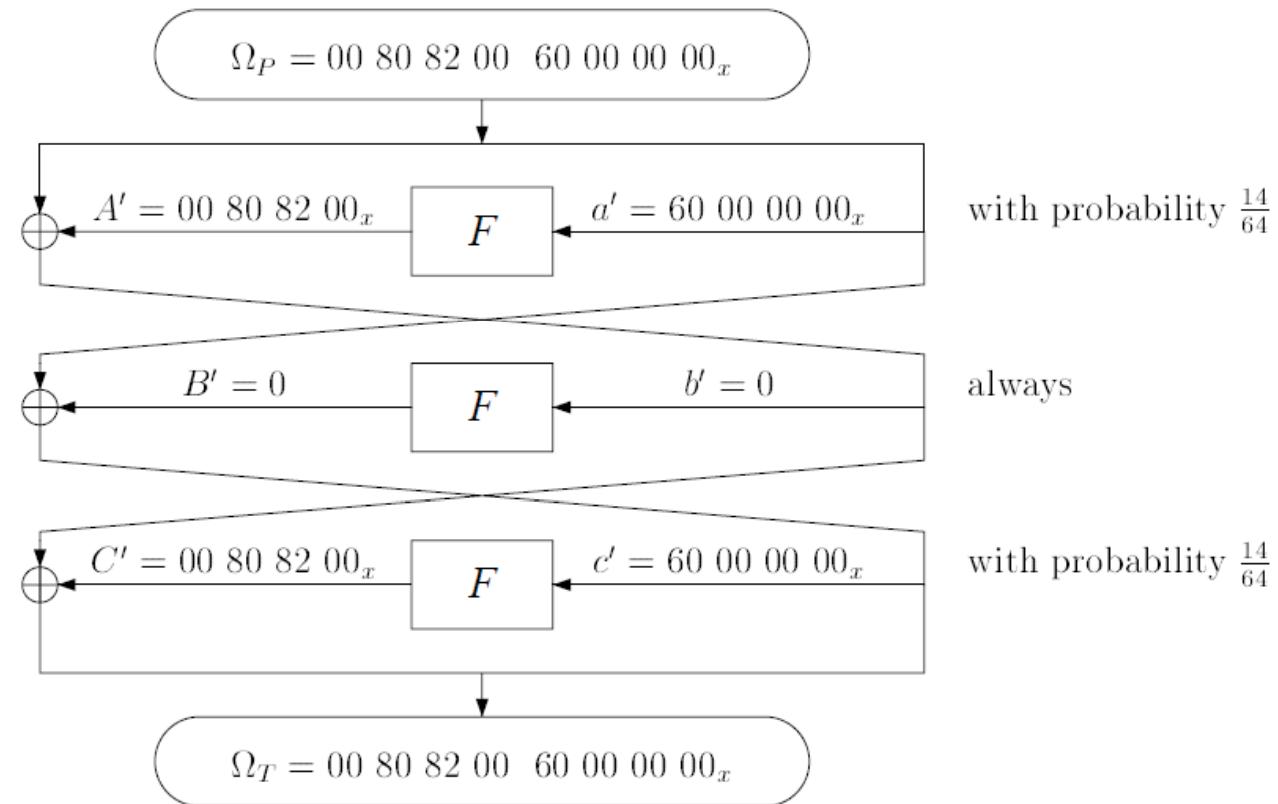
Otherwise, insecure.

Statistical Approach (contd.)

- Differential cryptanalysis (contd.)
 - 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.:
$$\Delta P \Rightarrow \Delta_1 \Rightarrow \Delta_2 \Rightarrow \Delta_3 \dots \dots \Rightarrow \Delta_{r-1} \Rightarrow \Delta C$$
 - 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 $\Delta P \Rightarrow \Delta_1 \dots \dots \Rightarrow \Delta C$
 - » We are able to attack the first round separately!

Statistical Approach (contd.)

- Differential cryptanalysis (contd.)
 - Example: differential propagation for 3-round DES



Statistical Approach (contd.)

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

Statistical Approach (contd.)

- 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 propagation to pass through many Sboxes
 - Diffusion should be properly designed
 - Example: AES
 - » ShiftRows+
 - MixColumns(maximum distance separable code)
 - » At least 25 active Sboxes are involved in 4 rounds

Statistical Approach (contd.)

- Linear cryptanalysis
 - Basic idea: for plaintext and ciphertext, if some input bits and output bits are statistically correlated, linear cryptanalysis may be applied!
 - We can have the following linear approximation equation that involves some plaintext bits, ciphertext bits and key bits:
$$k_a + k_b + \dots - p_i + p_j + \dots + c_{i'} + c_{j'} + \dots \text{ with prob. } p = 0.5 + x$$
(x is a small value)
 - After collecting enough plaintext-ciphertext ($1/x^2$), we can obtain the equation with high confidence:

$$\begin{aligned} & k_a + k_b + \dots = 0 \\ \text{or } & k_a + k_b + \dots = 1 \end{aligned}$$

Another way to solve
overdefined nonlinear equations!

Statistical Approach (contd.)

- Linear cryptanalysis (contd.)
 - Example:
DES: 5-round linear approximation with $p = 0.519$

$$\begin{aligned} & P_H[15] \oplus P_L[7, 18, 24, 27, 28, 29, 30, 31] \oplus C_H[15] \oplus C_L[7, 18, 24, 27, 28, 29, 30, 31] \\ &= K_1[42, 43, 45, 46] \oplus K_2[22] \oplus K_4[22] \oplus K_5[42, 43, 45, 46]. \end{aligned}$$

Statistical Approach (contd.)

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

Statistical Approach (contd.)

- Linear cryptanalysis (contd.)
 - 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