Лекция 11. Линейный криптоанализ блочных шифров

Цель лекции: рассмотреть один из универсальных методов криптоанализа.

План лекции:
Введение.
1 Линейный криптоанализ блочных шифров
Заключение
Контрольные вопросы

Ключевые слова: [выбрать самостоятельно].

Содержание лекции:

Введение

1 Линейный криптоанализ блочных шифров

Линейный криптоанализ [1] является атакой при известном открытом и шифрованном текстах на итеративные шифры типа DES, в которых «криптографически слабая» функция повторяется r раз. Предполагается, что открытый текст выбирается случайно и равновероятно, а подключи в каждом раунде независимы друг от друга.

Замечание 1. По-видимому, не имеет значения каким образом выбирается открытый текст.

Часто линейный криптоанализ является наиболее эффективным методом, тогда сложность линейного криптоанализа итеративных блочных шифров может являться мерой их слабости.

Линейный криптоанализ Мацуи r-раундового DES требует значительного числа открытых текстов. Это следует из следующей таблицы.

Количество раундов	Количество известных открытых текстов для нахождения ключа
8	2^{21}
12	2^{33}
16	2^{47} $\xrightarrow{Crypto'94}$ 2^{43} $\rightarrow Eurocrypt'93$

Сложность атаки связана с количеством необходимых известных открытых текстов, т.к. для любой пары (открытый текст, шифртекст) требуется небольшое количество вычислений для реализации алгоритма.

Например, атака на r=8 DES занимает 40 сек. на рабочей станции, а атака на r=12 DES занимает 50 час, что примерно в 4500 раз больше.

Эта атака может быть проведена, если имеется только шифртекст. Мацуи были получены следующие результаты:

- ✓ *если известно, что открытый текст на английском языке, представленный в ASCII кодах, то при r = 8 DES вскрывается при 2^{29} известных шифртекстах;*
- ✓ *если открытый текст является случайным ASCII кодом, то при r = 8 DES вскрывается при 2^{37} известных шифртекстах.*

Рассмотрим основную идею линейного криптоанализа. Пусть S – {0,1} – случайная величина с P{S = 1} = p. $\Delta(S) = |1 - 2p|$.

Лемма 1. Если $S_{(1)}, \ldots, S_{(n)}$ – независимые бинарные случайные величины, тогда

[1] Matsui M. Linear Cryptanalysis Method for DES Cipher//Pre- Proceedings of Eurocrypt'93, 1993, с. 112-133.

$$\Delta\left(S_{(1)} \oplus \ldots \oplus S_{(n)}\right) = \prod_{i=1}^{n} \Delta\left(S_{(i)}\right).$$

Доказательство. В силу независимости $S_{(1)}$ и $S_{(2)}$

$$P\left(S_{(1)} \oplus S_{(2)} = 1\right) = P\left(S_{(1)} = 1\right)P\left(S_{(2)} = 0\right) + P\left(S_{(1)} = 0\right)P\left(S_{(2)} = 1\right) = p_1 + p_2 - 2p_1p_2.$$

Тогда

$$\Delta\left(S_{(1)} \oplus S_{(2)}\right) = 1 - 2p\left(S_{(1)} \oplus S_{(2)} = 1\right) = 1 - 2p_1 - 2p_2 + 4p_1p_2 =$$

$$= (1 - 2p_1)(1 - 2p_2) = \Delta\left(S_{(1)}\right)\Delta\left(S_{(2)}\right).$$

Утверждение леммы следует из-за ассоциативности операции $\oplus$:

$$S_{(1)} \oplus \left(S_{(2)} \oplus S_{(3)}\right) = \left(S_{(1)} \oplus S_{(2)}\right) \oplus S_{(3)}.$$

Заметим, что $0 \le \Delta(S) \le 1$. При этом $\Delta(S) = 0$ при $p = 1/2$, а $\Delta(S) = 1$ при $P(S = 1) = 1$ или $P(S = 0) = 1$.

Рассмотрим схему произвольного итеративного блочного шифра в i-ом раунде.

$$\begin{array}{c} \vec{X}(i) \\ \downarrow \\ E \\ \downarrow \\ \vec{Y}(i) \end{array} \leftarrow \vec{K}(i) \quad i = 1, \ldots, r.$$

$$\vec{Y}(i) = E\left(\vec{X}(i); \vec{K}(i)\right). \qquad (1)$$

Здесь E – функция шифрования, $\vec{X}(i)$ – блок открытого текста в i-ом раунде, $\vec{Y}(i)$ – блок шифртекста, $\vec{K}(i)$ – подключ, используемый в i-ом раунде. $\vec{Y}(i), \vec{X}(i) \in V_n, \vec{K}(i) \in V_m$, n - размер блока, m - размер подключа.

Обозначим через $\left(\vec{X}, \vec{\alpha}\right) = X_1\alpha_1 \oplus \ldots \oplus X_n\alpha_n = X_{i_1} \oplus \ldots \oplus X_{i_k} = X[i_1, \ldots, i_k]$ – скалярное произведение двоичных векторов $\vec{X}$ и $\vec{\alpha}$, где $\left(\alpha_{i_1}, \ldots, \alpha_{i_k}\right)$ единичные координаты вектора $\vec{\alpha}$.

Определение 1. Линейным статистическим аналогом нелинейной функции (1) будем называть случайную величину

$$S(i) = \left(\vec{Y}(i), \vec{\alpha}(i)\right) \oplus \left(\vec{X}(i), \vec{\beta}(i)\right) \oplus \left(\vec{K}(i), \vec{\gamma}(i)\right), \qquad (2)$$

если вероятность $P(S(i) = 1) = p \ne ½$ для случайно выбранного открытого текста $\vec{X}(i)$.

Отклонение $\Delta(S(i)) = |1 - 2p|$ определяет эффективность линейного статистического аналога (2).

Замечание 2. Массе [2] называет S(i) тройной суммой и предлагает обобщение линейного криптоанализа, рассматривая тройную сумму для i-го раунда как случайную величину S(i) вида

[2] Massey J. Cryptography: Fundamentals and Applications. Advanced Technology Seminars, 1994.

$$S(i) = f_i\left(\vec{X}(i)\right) \oplus g_i\left(\vec{Y}(i)\right) \oplus h_i\left(\vec{K}(i)\right),$$

где f_i, g_i, h_i – равновероятные булевы функции. В определении Мацуи f_i, g_i, h_i – линейные функции.

Определение 2. Последовательные тройные суммы S(i + 1) и S(i) называются связанными, если $f_{i+1} = g_i$.

Из определения следует, что

$$S(i) \oplus S(i+1) = f_i\left(\vec{X}(i)\right) \oplus g_i\left(\vec{Y}(i)\right) \oplus h_i\left(\vec{K}(i)\right) \oplus f_{i+1}\left(\vec{X}(i+1)\right) \oplus g_{i+1}\left(\vec{Y}(i+1)\right) \oplus h_{i+1}\left(\vec{K}(i+1)\right),$$

$$\vec{Y}(i) = \vec{X}(i+1).$$

Определение 3. Если S(1), …, S(n) – связанные тройные суммы, то тогда

$$S_{1\ldots n} = S(1) \oplus \ldots \oplus S(n) = f_1\left(\vec{X}(1)\right) \oplus g_n\left(\vec{Y}(n)\right) \oplus \sum_{i=1}^{n} h_i\left(\vec{K}(i)\right) \quad (3)$$

и $S_{1\ldots n}$ называется n-раундовой тройной суммой.

В случае линейных функций формула (3) дает линейное соотношение

$$S_{1\ldots n} = \left(\vec{X}(1), \vec{\alpha}\right) \oplus \left(\vec{Y}(n), \vec{\beta}\right) \oplus \sum_{i=1}^{n} \left(\vec{K}(i), \vec{\gamma}(i)\right), \quad (4)$$

которое выполняется с вероятностью, определяемой $\Delta(S(1) \oplus \ldots \oplus S(n))$ и леммой 1.

Определение 4. Эффективным линейным статистическим аналогом называется линейный статистический аналог (4) из заданного множества с наибольшим Δ.

Для применения линейного криптоанализа необходимо решить следующие задачи:

1. *нахождения эффективного линейного статистического аналога и вычисление его вероятности;*
2. *определения ключа (или некоторых бит ключа) с использованием эффективного линейного статистического аналога.*

Для нахождения эффективного линейного статистического аналога рассмотрим сначала задачу нахождения статистических аналогов для S-боксов алгоритма DES.

Нелинейная функция, реализующая a-ый S-бокс алгоритма DES, может быть записана в виде

$$\vec{Y} = F_a\left(\vec{X} \oplus \vec{K}\right), \quad a = \overline{1,8}, \quad \vec{Y} \in V_4, \quad \vec{X}, \vec{K} \in V_6. \quad (5)$$

Пусть $1 \le i < 64$, $1 \le j < 16$, а $\vec{k}$ – двоичное разложение числа $k \in N$. Линейным статистическим аналогом каждого из уравнений (5) будет уравнение вида

$$\left(\vec{Y}, \vec{j}\right) = \left(\vec{X}', \vec{\imath}\right), \quad \vec{X}' = \vec{X} \oplus \vec{K}. \quad (6)$$

Обозначим через $S_a(i,j), a = \overline{1,8}, i = \overline{1,63}, j = \overline{1,15}$, число ненулевых $\vec{X} \in V_6$ для a-го S-бокса DES таких, что для i и j выполняется равенство (6). Когда $S_a(i, j) \neq 32$ можно сказать, что есть статистическая связь между входными и выходными битами a-го S-бокса.

Пусть

$$S_a^*(i^*,j^*): |S_a^*(i^*,j^*) - 32| = \max_{\substack{1 \le i \le 63 \\ 1 \le j \le 15}} |S_a(i,j) - 32|. \quad (7)$$

Тогда уравнение

$$(\vec{X},\vec{i}^*) \oplus (\vec{Y},\vec{j}^*) = (\vec{K},\vec{i}^*) \quad (8)$$

является эффективным линейным статистическим аналогом a-го S-бокса в классе всех линейных статистических аналогов вида (6) с вероятностью

$$p_a = \frac{S_a^*(i^*,j^*)}{64}, \qquad a = \overline{1,8}. \quad (9)$$

Все значения для $S_5(i, j)$ приведены в таблице 1. Значение $S_5(i, j)$, имеющее наибольшее отклонение от 1/2, помечено * в таблице 1. Следовательно, $S_5^*(i^*,j^*) = 12$, где $\vec{j}^*$ = (0, 1, 0, 0, 0, 0), $\vec{i}^*$ = (1, 1, 1, 1).

Отсюда эффективным линейным статистическим аналогом 5-го S-бокса является уравнение

$$\vec{Y}[1,2,3,4] \oplus \vec{X}[2] = \vec{K}[2], \quad (10)$$

и это уравнение выполняется с вероятностью $p_5 = 3/16$.

Таблица 1. Значения $S_5(i, j)$.

i	Значения j														
S(i, j)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32
2	36	30	34	30	34	28	32	36	32	34	30	34	30	32	28
3	32	30	38	30	30	36	28	32	32	30	38	30	30	36	28
4	34	30	32	32	34	30	32	32	34	34	36	28	30	30	32
5	34	34	28	32	42	26	28	32	34	22	32	36	30	34	36
6	30	28	26	30	28	34	32	32	30	32	30	26	24	34	32
7	34	32	34	30	40	38	32	28	38	32	26	30	32	26	28
8	32	34	38	32	32	30	26	30	34	36	20	34	38	28	36
9	28	38	30	32	28	26	26	38	30	32	28	34	26	24	28
10	36	32	32	30	26	34	34	34	34	30	34	36	28	28	32
11	36	36	36	38	34	30	30	30	30	30	34	32	24	28	32
12	34	32	30	32	34	36	42	30	36	30	24	30	36	26	28
13	38	32	34	32	30	36	22	30	32	30	36	30	40	26	32
14	30	30	32	30	36	32	34	30	32	36	34	28	38	30	28
15	30	30	40	38	36	32	34	34	36	40	30	40	26	34	32
16	34	30	32	32	30	26	24	32	30	30	28	32	34	42	12*
17	34	30	32	36	34	30	28	36	34	34	32	24	26	34	36
18	30	32	30	34	28	30	24	36	38	36	38	30	36	26	32
19	26	32	34	30	36	34	32	36	26	36	34	26	36	30	32

20	36	28	32	32	32	32	32	28	28	36	36	32	36	28	32
21	36	32	28	28	36	24	24	32	32	28	36	40	36	32	36
22	32	38	38	34	30	36	32	36	32	38	34	34	34	32	32
23	36	26	30	38	30	28	36	36	28	26	34	30	34	32	36
24	38	32	34	36	22	28	34	34	32	30	32	34	36	30	28
25	34	36	26	32	30	36	30	38	40	38	36	42	32	34	28
26	34	34	24	30	36	32	34	30	32	36	34	32	30	30	32
27	34	38	28	26	32	32	34	38	40	32	30	28	26	30	32
28	32	30	34	36	32	30	34	30	38	28	32	34	30	32	32
29	36	30	38	24	32	30	34	42	30	24	24	34	34	32	36
30	28	24	32	30	30	30	34	30	34	30	38	36	36	36	32
31	28	40	24	34	26	26	30	30	34	30	30	24	32	32	28
32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32
33	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32
34	28	30	34	30	34	28	40	28	32	26	38	34	30	16	20
35	32	30	30	38	30	28	36	32	32	30	30	30	38	36	28
36	30	38	36	32	38	30	36	36	26	30	36	32	46	34	32
37	38	34	32	32	38	34	32	28	26	34	24	32	30	38	28
38	34	36	30	30	32	34	28	36	30	28	30	38	32	30	32
39	22	32	30	38	36	38	28	32	38	20	34	34	32	38	28
40	36	30	30	32	36	26	34	34	26	36	32	38	30	28	32
41	32	34	38	32	32	38	34	34	30	24	32	30	26	32	32
42	32	28	24	38	38	38	26	38	34	30	30	24	36	28	36
43	40	32	36	38	30	26	38	34	38	30	38	28	32	36	36
44	34	36	26	32	26	32	38	30	28	34	28	30	36	38	32
45	30	28	30	32	30	24	34	30	32	26	24	30	32	30	36
46	38	34	28	38	36	36	30	22	24	32	30	36	30	34	32
47	38	26	28	38	28	36	30	34	36	36	26	32	34	30	28
48	34	30	32	28	26	30	28	36	34	34	32	32	34	34	36
49	34	30	32	32	30	34	32	32	30	30	28	32	34	34	36
50	38	32	30	30	40	34	36	32	42	32	34	30	36	34	32
51	26	32	42	34	32	30	28	32	38	32	22	34	36	30	32
52	32	20	36	28	32	36	24	28	32	28	32	28	28	32	32
53	24	32	32	40	28	36	32	32	28	28	32	36	36	28	36
54	36	30	26	30	30	40	32	36	28	30	30	38	34	28	32
55	24	26	26	26	38	32	36	44	32	34	30	34	34	36	28
56	34	36	26	32	30	36	30	26	36	26	32	38	36	30	32
57	30	40	34	28	38	28	26	30	28	34	36	30	32	34	32
58	38	22	32	34	36	32	30	38	28	32	34	36	30	30	28
59	30	26	28	22	32	24	30	22	36	36	30	32	34	30	36
60	24	26	30	32	28	34	34	26	34	36	32	42	30	36	36
61	36	34	34	36	36	30	34	30	42	32	32	34	34	36	32
62	28	36	28	34	34	30	34	34	30	30	30	36	28	32	36
63	28	28	28	46	38	26	30	34	30	38	30	32	32	28	32

Уравнения для эффективных линейных статистических аналогов всех S-боксов приведены в следующей таблице 2. Здесь $\vec{X}$ = (x_1, ..., x_6), $\vec{Y}$ = (y_1, ..., y_4), $\vec{K}$ = (k_1, ..., k_6) – входные, выходные и ключевые векторы соответственно.

Таблица 2. Уравнения для эффективных линейных статистических аналогов S-боксов

№ S-бокса	Эффективное линейное уравнение	p	Δ = |1 – 2p|
1	$x_2 \oplus y_1 \oplus y_2 \oplus y_3 \oplus y_4 = k_2$	7/32	9/16
2	$x_1 \oplus x_5 \oplus y_1 \oplus y_3 \oplus y_4 = k_1 \oplus k_5$	1/4	1/2
3	$x_1 \oplus x_5 \oplus y_1 \oplus y_2 \oplus y_3 \oplus y_4 = k_1 \oplus k_5$	1/4	1/2
4	$x_1 \oplus x_5 \oplus y_1 \oplus y_2 \oplus y_3 \oplus y_4 = k_1 \oplus k_5 \oplus$	1/4	1/2
	$x_1 \oplus x_3 \oplus y_1 \oplus y_2 \oplus y_3 \oplus y_4 = k_1 \oplus k_3 \oplus$	1/4	1/2
	$x_1 \oplus x_3 \oplus x_5 \oplus x_6 \oplus y_2 \oplus y_3 = k_1 \oplus k_3 \oplus k_5 \oplus k_6$	1/4	1/2
	$x_1 \oplus x_3 \oplus x_5 \oplus x_6 \oplus y_1 \oplus y_4 = k_1 \oplus k_3 \oplus k_5 \oplus k_6$	1/4	1/2
5	$x_2 \oplus y_1 \oplus y_2 \oplus y_3 \oplus y_4 = k_2$	3/16	5/8
6	$x_2 \oplus y_2 \oplus y_3 \oplus y_4 = k_2$	9/32	7/16
	$x_1 \oplus x_5 \oplus y_1 \oplus y_3 \oplus y_4 = k_1 \oplus k_5$		
7	$x_1 \oplus x_2 \oplus x_3 \oplus x_5 \oplus x_6 \oplus y_2 = k_1 \oplus k_2 \oplus k_3 \oplus k_5 \oplus k_6$	7/32	9/16
8	$x_2 \oplus y_1 \oplus y_2 \oplus y_3 \oplus y_4 = k_2$	1/4	1/2
	$x_1 \oplus x_5 \oplus y_1 \oplus y_2 \oplus y_3 = k_1 \oplus k_5$		

Исходя из результатов таблицы 2, можно сделать вывод, что эффективным линейным статистическим аналогом одного раунда DES является уравнение (10). С учётом функции расширения и перестановки алгоритма DES (см. таблицу 2 лекции 1) имеем эффективный линейный статистический аналог i-го раунда DES

$$\vec{Y}(i)[2,8,14,24] \oplus \vec{X}(i)[17] = \vec{K}(i)[26], \qquad (11)$$

с $\Delta = 5/8$.

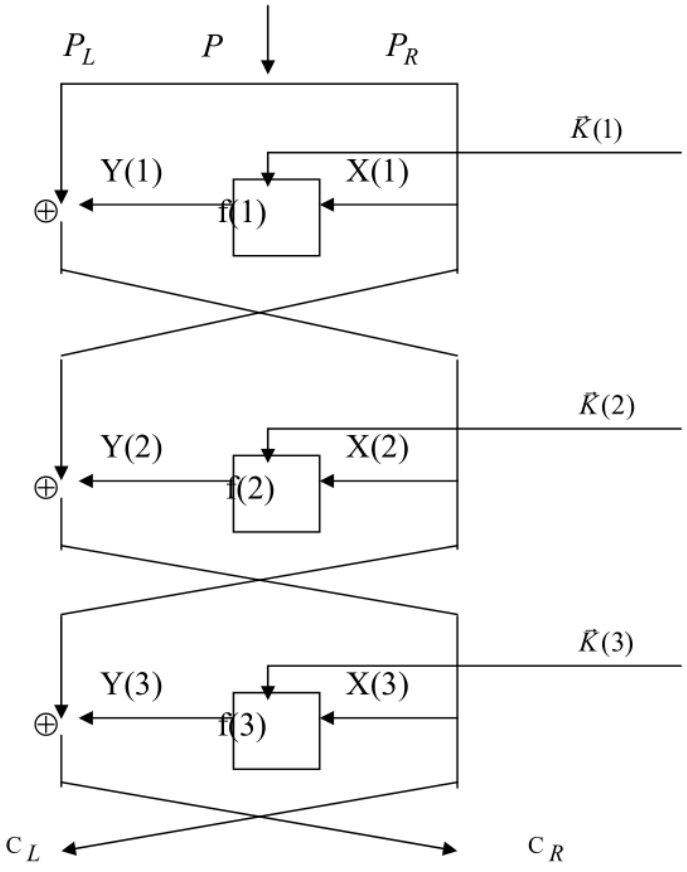

Рис. 1. Схема 3-х раундов DES

Мацуи не описывал алгоритма получения эффективных линейных аналогов для r-раундового DES, но из финальных результатов можно предположить, что основная идея состоит в следующем: использовать связанные тройные суммы с наибольшим Δ.

Поясним это на примере. Для этого рассмотрим схему 3-раундового DES.

Эффективными линейными статистическими аналогами 1-ой и 3-ей итераций являются уравнения

$$\vec{Y}(1)[2,8,14,24] \oplus \vec{X}(1)[17] = \vec{K}(1)[26],$$
$$\vec{Y}(3)[2,8,14,24] \oplus \vec{X}(3)[17] = \vec{K}(3)[26],$$

каждое из которых выполняется с вероятностью 3/16.

Учитывая , что $\vec{X}(1) = P_R$, $\vec{X}(3) = C_L$, $\vec{Y}(1) = P_L \oplus \vec{X}(2)$, $\vec{Y}(3) = C_R \oplus \vec{X}(2)$, после сложения этих уравнений получим

$$(P_R \oplus C_L)[17] \oplus (P_L \oplus C_R)[2,8,14,24] = \left(\vec{K}(1) \oplus \vec{K}(3)\right)[26]. \qquad (12)$$

По лемме 1 Δ = (5/8)×(5/8) = 25/64 и это уравнение выполняется с вероятностью

$$p = (1 - \Delta)/2 = 39/128.$$

Найти биты ключа $\vec{K}(1) \oplus \vec{K}(3)$ можно, решая уравнение (12) с использованием следующего алгоритма.

Алгоритм. Пусть N – число всех открытых текстов и T – число открытых текстов, для которых левая часть (11) равна 0.

Если T > N/2, то

$$\vec{K}(1) \oplus \vec{K}(3) = \begin{cases} 0, & p > 1/2 \\ 1, & p < 1/2 \end{cases}.$$

Если T < N/2, то

$$\vec{K}(1) \oplus \vec{K}(3) = \begin{cases} 1, & p > 1/2 \\ 0, & p < 1/2 \end{cases}.$$

Успех алгоритма возрастает с ростом N и Δ = |1 – 2p|. Вероятность успеха определяется следующей леммой.

Лемма 2. Пусть N – число открытых текстов и p - вероятность выполнения линейного уравнения (2). Предположим, что Δ = |1 – 2p| → 0. Тогда вероятность успеха алгоритма имеет вид

$$P \to \int_{-\sqrt{N}\Delta}^{\infty} \varphi(x)dx,$$

где

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \qquad (13)$$

Числовые вычисления (13) сведены в таблице 3.

Таблица 3. Вероятностей успеха алгоритма

N	$1/16\Delta^2$	$1/8\Delta^2$	$1/4\Delta^2$	$1/2\Delta^2$
$P_{усп}$	0,841	0,921	0,977	0,998

Заключение

Контрольные вопросы

Смотри руководство по организации самостоятельной работы магистрантов.