HOMEWORK-2

Şeyma ÇALIŞKAN

Intern at GSTL

1. Introduction

In this project, the method of **linear cryptanalysis** is applied to three different **Substitution-Permutation Network (SPN)** block ciphers. The goal is to statistically evaluate the linear approximations between plaintext and ciphertext bits using biased expressions, and recover parts of the secret key by analyzing the encryption behavior over multiple plaintext–ciphertext pairs. Comparative results across all three SPNs are presented to assess the relative security of their designs.

For the first block cipher, the **master key was set to 0x0081**.
As a first step, the **Linear Approximation Table (LAT)** for the cipher's S-box was constructed in order to identify input-output bit relationships with high linear bias. Figure 1 illustrates the LAT of the 4-bit S-box used in this cipher.
Based on the table, linear expressions with the largest absolute bias values were selected to construct effective approximation trails.

```
PS C:\Users\seyma\OneDrive\Desktop> python tablefor1.py
    |   0   1   2   3   4   5   6   7   8   9   a   b   c   d   e   f
 ---------------------------------------------------------------------
  0 |  +8   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
  1 |   0  +2   0  -2  -2   0  +2   0  +4  -2   0  -2  -2   0  -2  -4
  2 |   0  -2   0  +2  +2  -4  -2  -4  +2   0  -2   0   0  +2   0  -2
  3 |   0   0   0   0   0  +4   0  -4  +2  +2  +2  +2  +2  -2  +2  -2
  4 |   0  +2   0  -6   0  -2   0  -2  -2   0  -2   0  +2   0  +2   0
  5 |   0   0  -4   0  +2  -2  +2  +2  +2  +2  +2  -2  +4   0   0   0
  6 |   0   0  +4   0  +2  +2  +2  -2   0   0   0  -4  +2  +2  -2  +2
  7 |   0  -2   0  -2  +4  +2  -4  +2   0  -2   0  -2   0  -2   0  -2
  8 |   0  -6   0  -2  -2   0  +2   0  +2   0  -2   0   0  -2   0  +2
  9 |   0   0   0   0   0   0   0   0  +2  -2  +2  -2  -2  +2  +6  +2
  a |   0   0   0   0  +4   0  +4   0   0  -4   0  +4   0   0   0   0
  b |   0  -2   0  +2  -2   0  +2   0  -4  -2   0  -2  +2   0  +2  -4
  c |   0   0  +4   0  +2  -2  +2  +2   0  +4   0   0  -2  -2  +2  -2
  d |   0  +2   0  +2   0  -2   0  -2   0  -2   0  -2   0  -6   0  +2
  e |   0  -2   0  -2   0  -2   0  -2  -2   0  +6   0  -2   0  -2   0
  f |   0   0  +4   0  -2  -2  -2  +2  +2  -2  +2  +2  +4   0   0   0
```

Figure 1 LAT

Following the construction of the LAT, several **linear approximations** were selected based on their bias values to form **linear trails** across multiple rounds of the cipher. These approximations were then used to statistically recover bits of the **final round key** by analyzing the distribution of the approximated expressions over a large set of plaintext–ciphertext pairs.
All linear trails and their corresponding key recovery equations are illustrated in **Figure 2**.

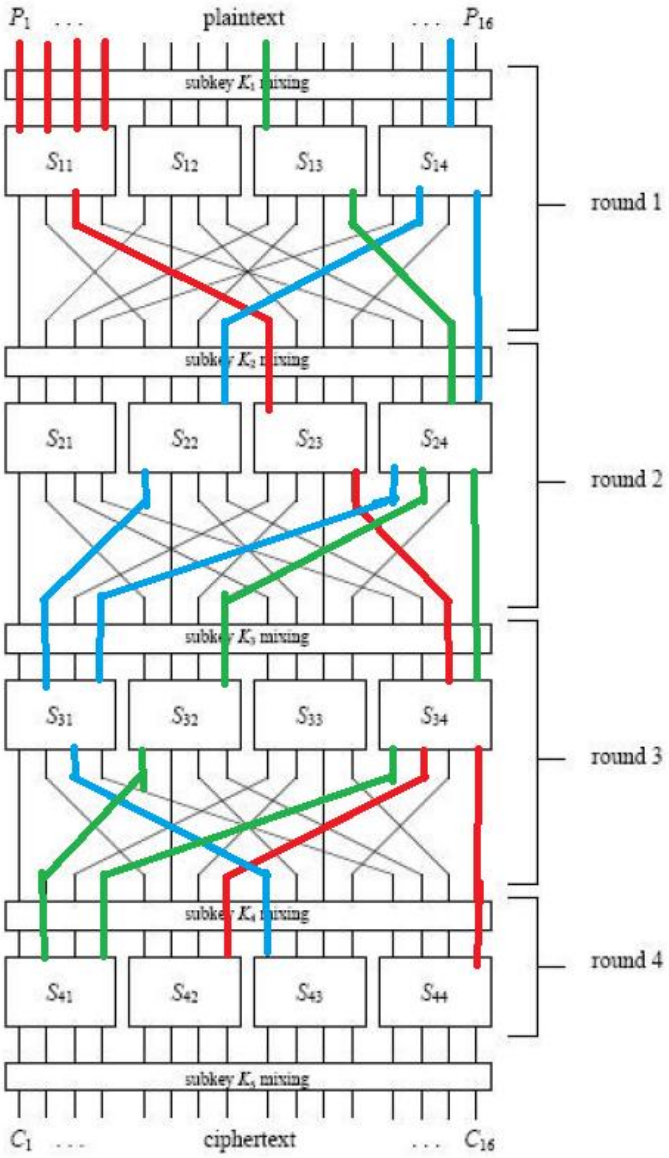

Figure 2

a. The First Approximation

The **first approximation**, represented by **red lines** in Figure 2, consists of three biased linear expressions derived from the S-box LAT of the first cipher:

s_{11}: $Y_3 = X_1 \oplus X_2 \oplus X_3 \oplus X_4$
⇨ This corresponds to $V_{1,3} = U_{1,1} \oplus U_{1,2} \oplus U_{1,3} \oplus U_{1,4}$ with ¼ bias.

---

s_{23}: $Y_1 = X_4$
⇨ This corresponds to $V_{2,12} = U_{2,9}$ with -3/8 bias.

---

s_{34}: $Y_2 \oplus Y_4 = X_3$
⇨ This corresponds to $V_{3,16} \oplus V_{3,14} = U_{3,15}$ with -1/4 bias.

By application of Piling-Up Lemma, the bias of the cipher (ε) is found as $2^2 * (-3/8) * (-1/4) * ¼ = 3/32$
From this approximation, we construct the following linear relation for the entire cipher:
$P_1 \oplus P_2 \oplus P_3 \oplus P_4 \oplus U_{4,16} \oplus U_{4,8} = 0$
which holds with a probability of: ½ - 3/32 = 13/32

The number of known plaintexts required for statistically distinguishing the key is approximately:

$N \approx \varepsilon^{-2} = (32/3)^2 \approx 114$

To generate the required data, a script (provided in the Appendices) is used to randomly create 114 plaintexts and compute their corresponding ciphertexts using the encryption function with the known master key.

These plaintext–ciphertext pairs are then used in the **linear cryptanalysis script**, which iterates over all possible **8-bit partial key values**, applies the **inverse S-box**, and evaluates whether the linear assumption holds.

For each correct evaluation, the score for the corresponding key guess is incremented in a scoreboard. In the end, the key guess with the **highest score** is considered the most probable correct key bits. The output of the script is shown in **Figure 3**.

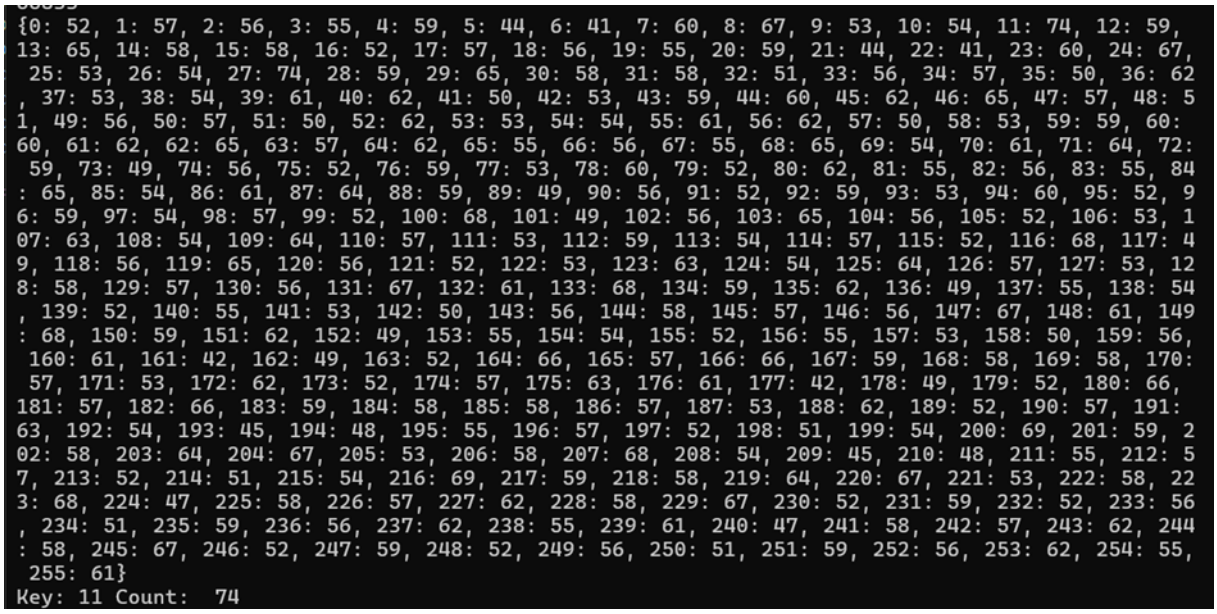

Figure 3 Red

b. The Second Approximation

The **second approximation**, represented by **blue lines** in Figure , follows a similar process to the first approximation. In this approach, four linear expressions from different S-box positions are selected based on their individual bias values:

s_{14}: $Y_2 \oplus Y_4 = X_3$
⇨ This corresponds to $V_{1,14} \oplus V_{1,16} = U_{1,15}$ with -¼ bias.

---

s_{22}: $Y_1 = X_4$
⇨ This corresponds to $V_{2,5} = U_{2,8}$ with ¼ bias.

---

s_{24}: $Y_1 = X_4$
⇨ This corresponds to $V_{2,13} \oplus V_{2,16} = U_{2,15}$ with 1/4 bias.

---

s_{31}: $Y_2 = X_2 \oplus X_4$
⇨ This corresponds to $V_{3,3} = U_{3,2} \oplus U_{3,4}$ with -1/4 bias.

Applying the **Piling-Up Lemma**, the overall trail bias ε is calculated as:
2^3 * -1/4 * -1/4 * ¼ * ¼ = 1/32

This leads to the following final round linear assumption: $P_{15} \oplus U_{4,9} = 0$

which is true with the probability: ½ - 1/32 = 15/32

To recover the relevant key bits statistically, the number of required known plaintexts is calculated as: $N \approx \varepsilon^{-2} = (32)^2 \approx 1024$

```
{0: 479, 1: 545, 2: 494, 3: 573, 4: 492, 5: 551, 6: 485, 7: 521, 8: 473, 9: 532, 10: 503, 11: 539, 12: 479, 1
3: 545, 14: 451, 15: 530}
Key: 3 Count:  573
PS C:\Users\seyma\OneDrive\Desktop>
```

Figure 4 Blue

c. The Third Approximation

The **third approximation**, illustrated with **green lines** in Figure 2, follows the same methodology as the previous two approximations. This time, four new linear expressions are selected from the S-box layers with varying bias values:

s_{13}: $Y_4 = X_1$
⇨ Corresponds to $V_{1,12} = U_{1,9}$ with -3/8 bias.

---

s_{24}: $Y_2 \oplus Y_4 = X_3$
⇨ Corresponds to $V_{2,14} \oplus V_{2,16} = U_{2,15}$ with -¼ bias.

---

s_{32}: $Y_1 = X_4$
⇨ Corresponds to $V_{3,5} = U_{3,8}$ with 1/4 bias.

---

s_{34}: $Y_1 = X_4$

⇨ Corresponds to $V_{3,13} = U_{3,16}$ with 1/4 bias.

Using the **Piling-Up Lemma**, the combined bias of this trail is calculated as:

$2^3 * 3/8 * 1/4 * ¼ * ¼ = 3/64$

This leads to the following linear expression at the final round: $P_9 \oplus U_{4,2} \oplus U_{4,4} = 0$
Which holds with probability: ½ - 1/32 = 15/32

Accordingly, the number of known plaintext–ciphertext pairs required for a successful key recovery is estimated as: $N \approx \varepsilon^{-2} = (64/3)^2 \approx 456$
Once again, these pairs are processed by the cryptanalysis script, which iterates over all partial key possibilities and uses the inverse S-box to test the linear expression. The scoreboard is updated based on how often each key guess satisfies the assumption, and the most likely key bits are determined.
The result of this attack is shown in **Figure 5**

```
PS C:\Users\seyma\OneDrive\Desktop> python cryyesil.py
{0: 232, 1: 248, 2: 242, 3: 216, 4: 244, 5: 239, 6: 221, 7: 210, 8: 213,
 9: 218, 10: 222, 11: 233, 12: 237, 13: 221, 14: 213, 15: 239}
Key: 1 Count:  248
```

Figure 5 Green

In the second block cipher, the **master key** is selected as 0x88C0.
The same cryptanalytic procedure and Python scripts used in the first cipher are applied here as well. The goal remains to identify biased linear approximations from the S-box, track their propagation across SPN rounds, and recover partial key bits from the final round.

To begin the analysis, the **Linear Approximation Table (LAT)** of the S-box used in this cipher is generated, and is presented in **Figure 7**. The entries of the table are analyzed to select input-output mask pairs with the highest absolute bias values, which serve as candidates for constructing effective linear trails.

	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
0	+8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	-2	0	+2	+2	0	-2	0	-4	+2	0	+2	+2	0	+2	+4
2	0	0	0	0	0	+4	0	-4	+2	+2	+2	+2	+2	-2	+2	-2
3	0	+2	0	-2	-2	+4	+2	+4	-2	0	+2	0	0	-2	0	+2
4	0	+2	0	-2	0	-2	+4	-2	0	+2	0	-2	+4	+2	0	+2
5	0	0	+4	+4	-2	+2	+2	-2	0	0	0	0	-2	+2	-2	+2
6	0	-2	0	+2	+4	+2	0	+2	+2	0	+2	-4	+2	0	-2	0
7	0	0	-4	+4	-2	-2	+2	+2	+2	+2	+2	+2	0	0	0	0
8	0	-4	+2	-2	0	0	+2	+2	+4	0	-2	+2	0	0	+2	+2
9	0	+2	+2	0	+2	0	0	+2	0	+2	-2	+4	+2	0	-4	-2
a	0	0	+2	+2	0	0	+2	+2	-2	-2	0	0	+2	+2	+4	-4
b	0	+2	+2	0	-2	0	-4	+2	+2	+4	0	-2	0	+2	+2	0
c	0	+2	+2	0	+4	-2	+2	0	0	+2	+2	0	-4	-2	+2	0
d	0	0	-2	-2	+2	+2	0	0	0	0	+2	+2	-2	+6	0	0
e	0	+2	+2	0	0	-2	-2	0	+2	-4	+4	+2	+2	0	0	+2
f	0	+4	-2	+2	+2	+2	0	0	+2	-2	-4	0	0	0	+2	+2

Figure 6 LAT

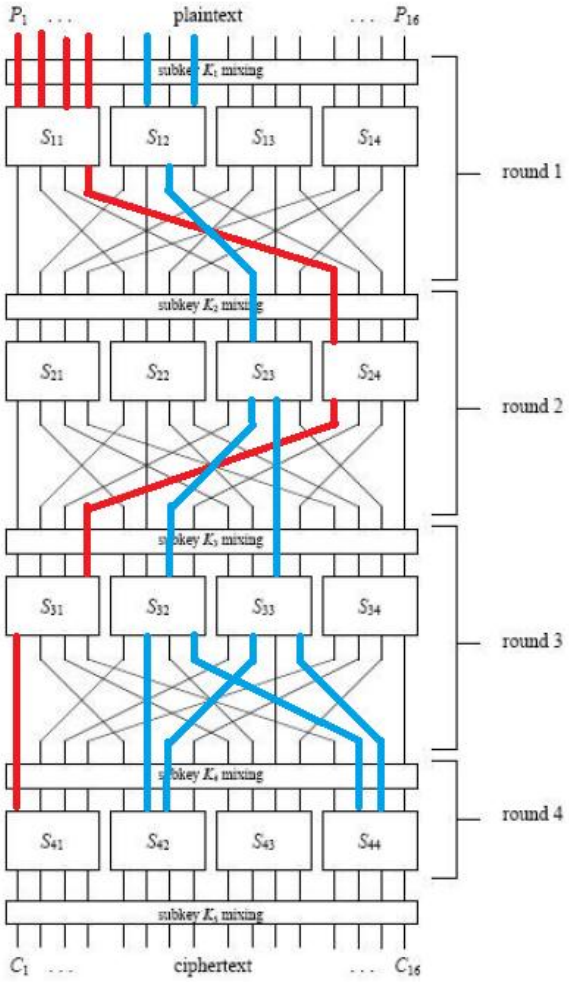

Figure 7 first and second

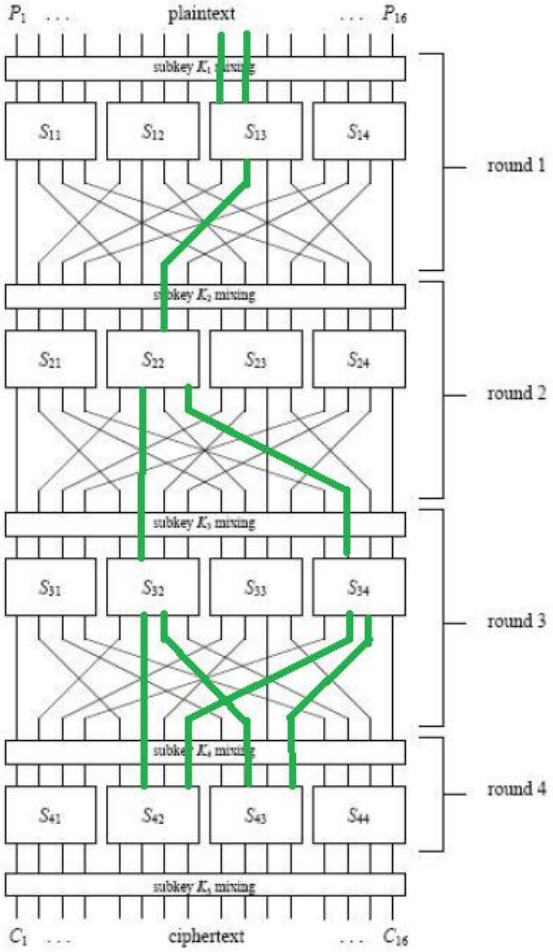

Figure 8 third

a. The First Approximation

The **first approximation** used in the cryptanalysis of the second block cipher is represented by **red lines** in **Figure 7**. This trail includes three linear expressions across different S-box layers, each selected based on their respective bias values obtained from the S-box's LAT.

s_{11}: $Y_3 = X_1 \oplus X_2 \oplus X_3 \oplus X_4$
$V_{1,3} = U_{1,1} \oplus U_{1,2} \oplus U_{1,3} \oplus U_{1,4}$ with ¼ bias.

---

s_{24}: $Y_1 = X_1$
$V_{2,12} = U_{2,12}$ with ¼ bias.

---

s_{31}: $Y_1 = X_4$
$V_{3,1} = U_{3,4}$ with -1/4 bias.

Using the **Piling-Up Lemma**, the cumulative bias ε of this linear trail is calculated as:
$\varepsilon = 2^2 * 1/4 * ¼ * -¼ = -1/16$

This leads to the following linear relation involving plaintext and final round internal bits: $P_1 \oplus P_2 \oplus P_3 \oplus P_4 \oplus U_{4,1} = 0$
which holds with the following probability: ½ + 1/16 = 9/16

$N \approx \varepsilon^{-2} = (16/1)^2 \approx 256$

According to the inverse-square relationship, the number of known plaintext–ciphertext pairs required to recover the relevant key bits with this approximation is:

```
{0: 135, 1: 140, 2: 131, 3: 121, 4: 141, 5: 144, 6: 111, 7: 137, 8: 123, 9: 121, 10: 113,
 11: 134, 12: 103, 13: 137, 14: 119, 15: 138}
Key: 5 Count:  144
```

Figure 9

As with previous approximations, these plaintext–ciphertext pairs are processed through the cryptanalysis script, which evaluates each partial key guess using inverse S-box computations and a scoreboard-based evaluation strategy.

b. The Second Approximation

The **second approximation**, illustrated with **blue lines** in **Figure 7**, includes four biased linear expressions selected from different S-box stages. Each expression was chosen from the LAT based on its non-zero bias and ease of propagation through the SPN structure.

s_{12}: $Y_3 = X_2 \oplus X_4$
$V_{1,7} = U_{1,6} \oplus U_{1,8}$ with ¼ bias.

---

s_{23}: $Y_2 \oplus Y_3 = X_2$
V2,10 ⊕ V2,11 = U2,10 with ¼ bias.

---

s_{32}: $Y_2 \oplus Y_4 = X_2$
$V_{2,6} \oplus V_{2,8} = U_{2,7}$ with ¼ bias.

---

s_{33}: $Y_2 \oplus Y_4 = X_2$
$V_{2,10} \oplus V_{2,12} = U_{2,11}$ with ¼ bias.

Applying the **Piling-Up Lemma**, the combined bias of this trail is computed as:
$\varepsilon = 2^3 * ¼ * ¼ * ¼ * ¼ = 1/8$

This leads to the following linear relation at the output of the cipher: $P_6 \oplus P_8 \oplus U_{4,6} \oplus U_{4,7} \oplus U_{4,14} \oplus U_{4,15} = 0$
which holds with probability: ½ - 1/8 = 3/8

To ensure statistically significant key recovery, the number of required known plaintext–ciphertext pairs is given by:

$N \approx \varepsilon^{-2} = (8/1)^2 \approx 64$

The cryptanalysis script utilizes these pairs to test all possible partial key guesses, incrementing the scoreboard whenever the linear assumption holds. The script output confirming the highest-scoring key guess is shown in **Figure 10**.

```
{0: 36, 1: 37, 2: 35, 3: 30, 4: 20, 5: 32, 6: 27, 7: 27, 8: 38, 9: 33, 10: 39, 11: 40, 12:
 27, 13: 33, 14: 26, 15: 32, 16: 36, 17: 37, 18: 35, 19: 30, 20: 20, 21: 32, 22: 27, 23: 2
7, 24: 38, 25: 33, 26: 39, 27: 40, 28: 27, 29: 33, 30: 26, 31: 32, 32: 29, 33: 32, 34: 30,
 35: 31, 36: 31, 37: 35, 38: 32, 39: 30, 40: 33, 41: 28, 42: 26, 43: 35, 44: 36, 45: 32, 4
6: 33, 47: 39, 48: 29, 49: 32, 50: 30, 51: 31, 52: 31, 53: 35, 54: 32, 55: 30, 56: 33, 57:
 28, 58: 26, 59: 35, 60: 36, 61: 32, 62: 33, 63: 39, 64: 26, 65: 35, 66: 33, 67: 30, 68: 3
4, 69: 34, 70: 33, 71: 35, 72: 30, 73: 29, 74: 25, 75: 32, 76: 39, 77: 29, 78: 28, 79: 40,
 80: 26, 81: 35, 82: 33, 83: 30, 84: 34, 85: 34, 86: 33, 87: 35, 88: 30, 89: 29, 90: 25, 9
1: 32, 92: 39, 93: 29, 94: 28, 95: 40, 96: 36, 97: 31, 98: 33, 99: 32, 100: 28, 101: 28, 1
02: 25, 103: 39, 104: 36, 105: 33, 106: 37, 107: 34, 108: 31, 109: 33, 110: 22, 111: 34, 1
12: 36, 113: 31, 114: 33, 115: 32, 116: 28, 117: 28, 118: 25, 119: 39, 120: 36, 121: 33, 1
22: 37, 123: 34, 124: 31, 125: 33, 126: 22, 127: 34, 128: 32, 129: 35, 130: 31, 131: 30, 1
32: 26, 133: 32, 134: 33, 135: 29, 136: 34, 137: 35, 138: 37, 139: 38, 140: 29, 141: 33, 1
42: 30, 143: 28, 144: 32, 145: 35, 146: 31, 147: 30, 148: 26, 149: 32, 150: 33, 151: 29, 1
52: 34, 153: 35, 154: 37, 155: 38, 156: 29, 157: 33, 158: 30, 159: 28, 160: 29, 161: 32, 1
62: 34, 163: 29, 164: 37, 165: 35, 166: 38, 167: 32, 168: 29, 169: 28, 170: 28, 171: 27, 1
72: 36, 173: 30, 174: 35, 175: 33, 176: 29, 177: 32, 178: 34, 179: 29, 180: 37, 181: 35, 1
82: 38, 183: 32, 184: 29, 185: 28, 186: 28, 187: 27, 188: 36, 189: 30, 190: 35, 191: 33, 1
92: 26, 193: 27, 194: 33, 195: 24, 196: 40, 197: 36, 198: 37, 199: 33, 200: 34, 201: 29, 2
02: 31, 203: 28, 204: 37, 205: 27, 206: 34, 207: 36, 208: 26, 209: 27, 210: 33, 211: 24, 2
12: 40, 213: 36, 214: 37, 215: 33, 216: 34, 217: 29, 218: 31, 219: 28, 220: 37, 221: 27, 2
22: 34, 223: 36, 224: 36, 225: 31, 226: 31, 227: 28, 228: 34, 229: 34, 230: 35, 231: 31, 2
32: 32, 233: 29, 234: 37, 235: 32, 236: 29, 237: 31, 238: 34, 239: 28, 240: 36, 241: 31, 2
42: 31, 243: 28, 244: 34, 245: 34, 246: 35, 247: 31, 248: 32, 249: 29, 250: 37, 251: 32, 2
52: 29, 253: 31, 254: 34, 255: 28}
Key: 0xb Count:  40
```

Figure 10 Blue

c. The Third Approximation

The **third approximation**, represented by **red lines** in **Figure 9**, involves four biased linear expressions identified from the LAT of the second cipher's S-box. The selected expressions allow for consistent tracking of bitwise relations across rounds and contribute to statistical key recovery in the final round.

s_{13}: $Y_2 = X_1 \oplus X_2$
$V_{1,10} = U_{1,9} \oplus U_{1,10}$ with ¼ bias.

---

s_{22}: $Y_2 \oplus Y_4 = X_2$
$V_{2,6} \oplus V_{2,8} = U_{2,7}$ with ¼ bias.

---

s_{32}: $Y_2 \oplus Y_3 = X_2$
$V_{3,7} \oplus V_{3,6} = U_{3,6}$ with ¼ bias.

---

s_{34}: $Y2 \oplus Y3 = X2$
$V_{3,15} \oplus V_{3,14} = U_{3,14}$ with ¼ bias.

Applying the **Piling-Up Lemma**, the overall bias of this trail is computed as:

$\varepsilon = 2^3 * ¼ * ¼ * ¼ * ¼ = 1/8$

From this trail, the following linear expression is assumed:

$P_9 \oplus P_{10} \oplus U_{4,6} \oplus U_{4,8} \oplus U_{4,10} \oplus U_{4,12} = 0$

which holds with probability: ½ - 1/8 = 3/8

Based on this bias, the number of known plaintext–ciphertext pairs required for reliable key recovery is:

$N \approx \varepsilon^{-2} = (8/1)^2 \approx 64$

As before, the cryptanalysis script evaluates this approximation using the generated data. It iterates over all partial key values, applies the inverse S-box operations, and updates a scoreboard based on how frequently the approximation holds. The script output identifying the most likely key guess is provided in **Figure 11**.

```
{0: 38, 1: 36, 2: 32, 3: 31, 4: 26, 5: 34, 6: 32, 7: 27, 8: 37, 9: 34, 10: 29, 11: 31, 12:
 27, 13: 36, 14: 35, 15: 27, 16: 38, 17: 36, 18: 32, 19: 31, 20: 26, 21: 34, 22: 32, 23: 2
7, 24: 37, 25: 34, 26: 29, 27: 31, 28: 27, 29: 36, 30: 35, 31: 27, 32: 32, 33: 32, 34: 32,
 35: 29, 36: 30, 37: 32, 38: 34, 39: 35, 40: 29, 41: 26, 42: 33, 43: 27, 44: 33, 45: 38, 4
6: 33, 47: 37, 48: 32, 49: 32, 50: 32, 51: 29, 52: 30, 53: 32, 54: 34, 55: 35, 56: 29, 57:
 26, 58: 33, 59: 27, 60: 33, 61: 38, 62: 33, 63: 37, 64: 28, 65: 28, 66: 30, 67: 35, 68: 3
8, 69: 28, 70: 32, 71: 37, 72: 31, 73: 26, 74: 37, 75: 33, 76: 35, 77: 30, 78: 25, 79: 39,
 80: 28, 81: 28, 82: 30, 83: 35, 84: 38, 85: 28, 86: 32, 87: 37, 88: 31, 89: 26, 90: 37, 9
1: 33, 92: 35, 93: 30, 94: 25, 95: 39, 96: 33, 97: 25, 98: 31, 99: 32, 100: 33, 101: 37, 1
02: 31, 103: 34, 104: 32, 105: 39, 106: 32, 107: 38, 108: 34, 109: 23, 110: 30, 111: 28, 1
12: 33, 113: 25, 114: 31, 115: 32, 116: 33, 117: 37, 118: 31, 119: 34, 120: 32, 121: 39, 1
22: 32, 123: 38, 124: 34, 125: 23, 126: 30, 127: 28, 128: 27, 129: 27, 130: 41, 131: 32, 1
32: 27, 133: 35, 134: 33, 135: 34, 136: 32, 137: 33, 138: 36, 139: 34, 140: 22, 141: 29, 1
42: 38, 143: 32, 144: 27, 145: 27, 146: 41, 147: 32, 148: 27, 149: 35, 150: 33, 151: 34, 1
52: 32, 153: 33, 154: 36, 155: 34, 156: 22, 157: 29, 158: 38, 159: 32, 160: 29, 161: 35, 1
62: 33, 163: 32, 164: 31, 165: 25, 166: 35, 167: 36, 168: 30, 169: 25, 170: 34, 171: 32, 1
72: 30, 173: 35, 174: 34, 175: 36, 176: 29, 177: 35, 178: 33, 179: 32, 180: 31, 181: 25, 1
82: 35, 183: 36, 184: 30, 185: 25, 186: 34, 187: 32, 188: 30, 189: 35, 190: 34, 191: 36, 1
92: 35, 193: 37, 194: 25, 195: 30, 196: 37, 197: 31, 198: 31, 199: 30, 200: 28, 201: 35, 2
02: 26, 203: 30, 204: 44, 205: 33, 206: 30, 207: 30, 208: 35, 209: 37, 210: 25, 211: 30, 2
12: 37, 213: 31, 214: 31, 215: 30, 216: 28, 217: 35, 218: 26, 219: 30, 220: 44, 221: 33, 2
22: 30, 223: 30, 224: 34, 225: 36, 226: 32, 227: 35, 228: 34, 229: 34, 230: 28, 231: 23, 2
32: 37, 233: 38, 234: 29, 235: 31, 236: 31, 237: 32, 238: 31, 239: 27, 240: 34, 241: 36, 2
42: 32, 243: 35, 244: 34, 245: 34, 246: 28, 247: 23, 248: 37, 249: 38, 250: 29, 251: 31, 2
52: 31, 253: 32, 254: 31, 255: 27}
Key: 0xcc Count:  44
```

Figure 11