CS5760: Cryptanalysis of DES and DES-like Iterated Cryptosystems

Gautam Singh

Indian Institute of Technology Hyderabad

February 3, 2025

Differential Cryptanalysis

1. Chosen plaintext attack.
2. Exploit XOR between plaintext pairs to find key bits.

Differential Cryptanalysis

1. Chosen plaintext attack.
2. Exploit XOR between plaintext pairs to find key bits.
3. Per DES round, XOR of respective inputs is:
   - *Linear* in expansion E to get S_E.
   - *Invariant* in key mixing with subkey S_K to get $S_I = S_E \oplus S_K$.
   - *Linear* in permutation P on S_O after S boxes.
   - *Invariant* in XOR operation connecting rounds.

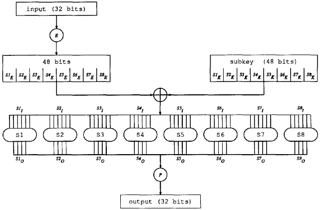

Figure 1: F function of DES.

Differential Cryptanalysis

1. Chosen plaintext attack.
2. Exploit XOR between plaintext pairs to find key bits.
3. Per DES round, XOR of respective inputs is:
   - *Linear* in expansion E to get S_E.
   - *Invariant* in key mixing with subkey S_K to get $S_I = S_E \oplus S_K$.
   - *Linear* in permutation P on S_O after S boxes.
   - *Invariant* in XOR operation connecting rounds.
4. S boxes are *nonlinear*. Probability analysis performed between input and output XOR.

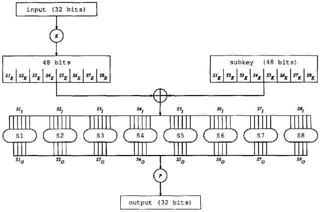

Figure 1: F function of DES.

Probability Analysis of S Boxes

1. Suppose $Si'_I = Si_I \oplus Si^*_I$ is the input XOR to the i^{th} S box, and Si'_O is the output XOR $(1 \leq i \leq 8)$.

Probability Analysis of S Boxes

1. Suppose $Si'_I = Si_I \oplus Si^*_I$ is the input XOR to the i^{th} S box, and Si'_O is the output XOR ($1 \leq i \leq 8$).
2. We create a *pairs XOR distribution table* for each S box.
   - Each entry (Si'_I, Si'_O) equals the number of 6-bit key blocks Si_K for which $Si'_I \rightarrow Si'_O$.
   - 64-by-16 joint probability mass function.

Probability Analysis of S Boxes

1. Suppose $Si_I' = Si_I \oplus Si_I^*$ is the input XOR to the i^{th} S box, and Si_O' is the output XOR $(1 \leq i \leq 8)$.
2. We create a *pairs XOR distribution table* for each S box.
   - Each entry (Si_I', Si_O') equals the number of 6-bit key blocks Si_K for which $Si_I' \to Si_O'$.
   - 64-by-16 joint probability mass function.
3. This joint PMF can reduce the number of possible (sub)keys. Used to drive choice for the plaintext XOR.
   - $\approx 80\%$ entries are non-zero/possible for each S box (some have lesser percentages).
   - Given Si_I' and Si_O', we can narrow down Si_K to a few possibilities.

Probability Analysis of S Boxes

1. Suppose $Si'_I = Si_I \oplus Si^*_I$ is the input XOR to the i^{th} S box, and Si'_O is the output XOR $(1 \le i \le 8)$.
2. We create a *pairs XOR distribution table* for each S box.
   - Each entry (Si'_I, Si'_O) equals the number of 6-bit key blocks Si_K for which $Si'_I \to Si'_O$.
   - 64-by-16 joint probability mass function.
3. This joint PMF can reduce the number of possible (sub)keys. Used to drive choice for the plaintext XOR.
   - $\approx 80\%$ entries are non-zero/possible for each S box (some have lesser percentages).
   - Given Si'_I and Si'_O, we can narrow down Si_K to a few possibilities.
4. i^{th} S box contributes probability p_i for $Si'_I \to Si'_O$.
   - For $X \to Y$ over a round, $P = \prod_i p_i$.
   - Over n rounds, $P = \prod_{i=1}^{n} P_i$.

Probability Analysis of S Boxes

1. Suppose $Si'_I = Si_I \oplus Si^*_I$ is the input XOR to the i^{th} S box, and Si'_O is the output XOR $(1 \leq i \leq 8)$.
2. We create a *pairs XOR distribution table* for each S box.
   - Each entry (Si'_I, Si'_O) equals the number of 6-bit key blocks Si_K for which $Si'_I \rightarrow Si'_O$.
   - 64-by-16 joint probability mass function.
3. This joint PMF can reduce the number of possible (sub)keys. Used to drive choice for the plaintext XOR.
   - $\approx 80\%$ entries are non-zero/possible for each S box (some have lesser percentages).
   - Given Si'_I and Si'_O, we can narrow down Si_K to a few possibilities.
4. i^{th} S box contributes probability p_i for $Si'_I \rightarrow Si'_O$.
   - For $X \rightarrow Y$ over a round, $P = \prod_i p_i$.
   - Over n rounds, $P = \prod_{i=1}^{n} P_i$.

Desirable for cryptanalysis: high P with large n.

Characteristic

Formalizes notion of high-probability plaintext XORs.

Definition 1 (Characteristic)

An *n-round chracteristic* is a tuple $\Omega = (\Omega_P, \Omega_\Lambda, \Omega_T)$ where $\Omega_P = (L', R')$ and $\Omega_T = (l', r')$ are m bit numbers, $\Omega_\Lambda = (\Lambda_1, \ldots, \Lambda_n)$, $\Lambda_i = (\lambda_I^i, \lambda_O^i)$ and $\lambda_I^i, \lambda_O^i, L', R', l', r'$ are $\frac{m}{2}$ bit numbers and m is the block size of the cryptosystem satisfying

$$\lambda_I^1 = R' \tag{1}$$

$$\lambda_I^2 = L' \oplus \lambda_O^1 \tag{2}$$

$$\lambda_I^n = r' \tag{3}$$

$$\lambda_I^{n-1} = l' \oplus \lambda_O^n \tag{4}$$

$$\forall\ 1 < i < n,\ \lambda_O^i = \lambda_I^{i-1} \oplus \lambda_I^{i+1} \tag{5}$$

Characteristic

Definition 2 (Right Pair)

A *right pair with respect to an n-round characteristic* $\Omega = (\Omega_P, \Omega_\Lambda, \Omega_T)$ *and an independent key* K is a pair for which $P' = \Omega_P$ and for each round i of the first n rounds of the encryption of the pair using K the input XOR of the i^{th} round equals λ_I^i and the output XOR of the F function equals λ_O^i. Pairs that do not satisfy these conditions are called *wrong pairs*.

Characteristic

Definition 2 (Right Pair)

A *right pair with respect to an n-round characteristic* $\Omega = (\Omega_P, \Omega_\Lambda, \Omega_T)$ *and an independent key* K is a pair for which $P' = \Omega_P$ and for each round i of the first n rounds of the encryption of the pair using K the input XOR of the i^{th} round equals λ_I^i and the output XOR of the F function equals λ_O^i. Pairs that do not satisfy these conditions are called *wrong pairs*.

Definition 3 (Probability of a Round of a Characteristic)

Round i of an n-round characteristic Ω has probability p_i^Ω if $\lambda_I^i \to \lambda_O^i$ with probability p_i^Ω by the F function.

Probability of a Characteristic

Definition 4 (Probability of a Characteristic)

An n-round characteristic Ω has probability p^{Ω} given by

$$p^{\Omega} = \prod_{i=1}^{n} p_i^{\Omega} \tag{6}$$

Probability of a Characteristic

Definition 4 (Probability of a Characteristic)

An n-round characteristic Ω has probability p^{Ω} given by

$$p^{\Omega} = \prod_{i=1}^{n} p_i^{\Omega} \tag{6}$$

Theorem 5 (Probability of a Characteristic and Right Pairs)

The formally defined probability of a characteristic $\Omega = (\Omega_P, \Omega_\Lambda, \Omega_T)$ is the probability that any fixed plaintext pair satisfying $P' = \Omega_P$ is a right pair when random independent keys are used.

Example of a Characteristic

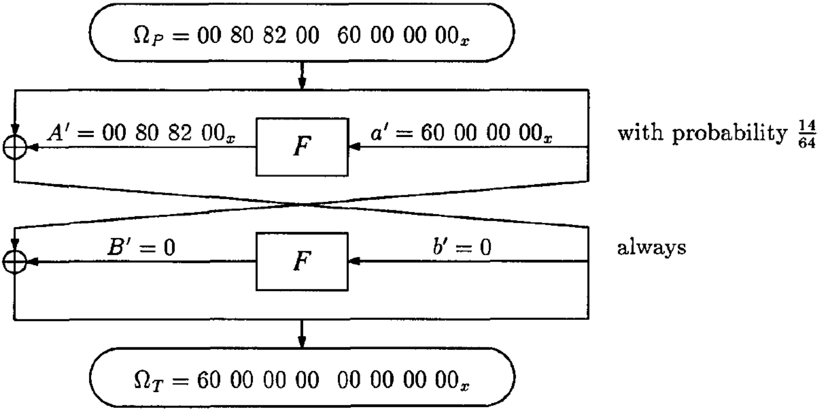

Figure 2: Example of a two-round characteristic with probability $\frac{14}{64}$.

Signal to Noise Ratio

1. Right pairs will always suggest the right key value. But right pairs occur with probability p^{Ω}.

Signal to Noise Ratio

1. Right pairs will always suggest the right key value. But right pairs occur with probability p^{Ω}.
2. On the other hand, wrong pairs suggest a randomly chosen key (not necessarily the right key in the worst case).

Signal to Noise Ratio

1. Right pairs will always suggest the right key value. But right pairs occur with probability p^{Ω}.
2. On the other hand, wrong pairs suggest a randomly chosen key (not necessarily the right key in the worst case).
3. Suitable counting approach on the key values will "spike" at the right key and have smaller but approximately equal counts at other keys.

Signal to Noise Ratio

1. Right pairs will always suggest the right key value. But right pairs occur with probability p^{Ω}.
2. On the other hand, wrong pairs suggest a randomly chosen key (not necessarily the right key in the worst case).
3. Suitable counting approach on the key values will "spike" at the right key and have smaller but approximately equal counts at other keys.
4. The key with the largest count is likely the actual key.

Signal to Noise Ratio

1. Right pairs will always suggest the right key value. But right pairs occur with probability p^{Ω}.
2. On the other hand, wrong pairs suggest a randomly chosen key (not necessarily the right key in the worst case).
3. Suitable counting approach on the key values will "spike" at the right key and have smaller but approximately equal counts at other keys.
4. The key with the largest count is likely the actual key.

Definition 6 (Signal-to-Noise Ratio)

The ratio between the number of right pairs and the average count of incorrect subkeys in a counting scheme is called the *signal to noise ratio of the counting scheme* and is denoted by S/N.

Computing the SNR

Consider the variables shown in Table 1.

Variable	Definition
p	Probability of the characteristic
m	Number of created pairs
α	Average count per analyzed pair
β	Fraction of analyzed pairs
k	Number of key bits counted on

Table 1: Table of variables to compute the SNR.

Computing the SNR

Consider the variables shown in Table 1.

Variable	Definition
p	Probability of the characteristic
m	Number of created pairs
α	Average count per analyzed pair
β	Fraction of analyzed pairs
k	Number of key bits counted on

Table 1: Table of variables to compute the SNR.

Then,

$$S/N = \frac{m \cdot p}{\frac{m \cdot \beta \cdot \alpha}{2^k}} = \frac{2^k \cdot p}{\alpha \cdot \beta} \tag{7}$$

Structures

1. Many attacks on DES use more than one characteristic.

Structures

1. Many attacks on DES use more than one characteristic.
2. Requirement to minimize the amount of plaintexts generated.

Structures

1. Many attacks on DES use more than one characteristic.
2. Requirement to minimize the amount of plaintexts generated.

Definition 7 (Quartet and Octet)

A *quartet* is a structure of four ciphertexts that simultaneously contains two ciphertext pairs of one characteristic and two ciphertext pairs of a second characteristic. An *octet* is a structure of eight ciphertexts that simultaneously contains four ciphertext pairs of each of three characteristics.

3. As an example, $(P, P \oplus \Omega_P^1, P \oplus \Omega_P^2, P \oplus \Omega_P^1 \oplus \Omega_P^2)$ is a quartet.

Structures

1. Many attacks on DES use more than one characteristic.
2. Requirement to minimize the amount of plaintexts generated.

Definition 7 (Quartet and Octet)

A *quartet* is a structure of four ciphertexts that simultaneously contains two ciphertext pairs of one characteristic and two ciphertext pairs of a second characteristic. An *octet* is a structure of eight ciphertexts that simultaneously contains four ciphertext pairs of each of three characteristics.

3. As an example, $(P, P \oplus \Omega_P^1, P \oplus \Omega_P^2, P \oplus \Omega_P^1 \oplus \Omega_P^2)$ is a quartet.
4. Quartets save $\frac{1}{2}$ of the data and octets save $\frac{2}{3}$ of the data.

DES Reduced to Four Rounds

1. Use two one-round characteristics, as shown in Figure 3.

DES Reduced to Four Rounds

1. Use two one-round characteristics, as shown in Figure 3.
2. Both characteristics have probability 1.

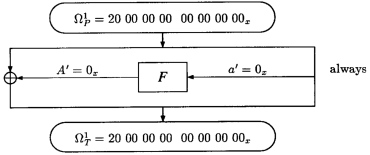

Figure 3: Characteristics used for cryptanalysis of DES reduced to four rounds.

DES Reduced to Four Rounds

1. Use two one-round characteristics, as shown in Figure 3.
2. Both characteristics have probability 1.
3. Example of a *3R-attack*. There are *three* extra rounds after the characteristic is applied.

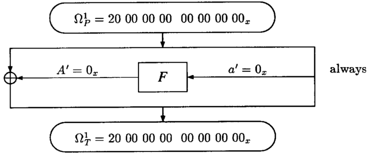

Figure 3: Characteristics used for cryptanalysis of DES reduced to four rounds.

DES Reduced to Four Rounds

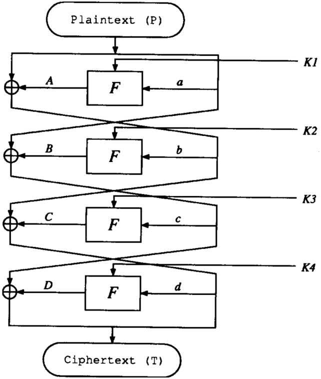

Figure 4: DES reduced to four rounds.

DES Reduced to Four Rounds

1. Using Ω^1, we have

$$c' = D' \oplus l' = a' \oplus B' \implies D' = B' \oplus l' \quad (8)$$

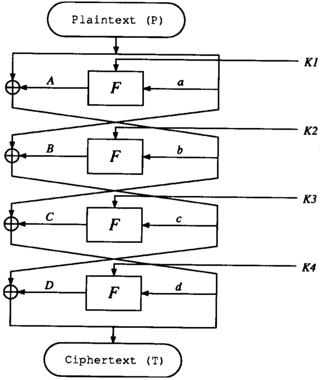

Figure 4: DES reduced to four rounds.

DES Reduced to Four Rounds

1. Using Ω^1, we have

$$c' = D' \oplus l' = a' \oplus B' \implies D' = B' \oplus l' \quad (8)$$

2. We have $a' = 0_x \implies A' = 0_x$ and $b' = A' \oplus L' = L'$.

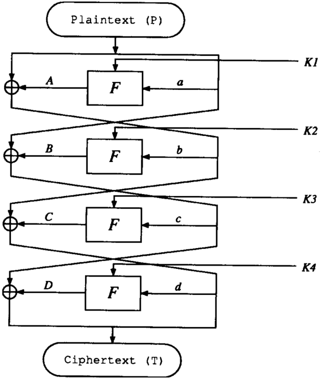

Figure 4: DES reduced to four rounds.

DES Reduced to Four Rounds

1. Using Ω^1, we have

$$c' = D' \oplus l' = a' \oplus B' \implies D' = B' \oplus l' \quad (8)$$

2. We have $a' = 0_x \implies A' = 0_x$ and $b' = A' \oplus L' = L'$.
   - In the second round S2, ..., S8 receive zero XOR input.

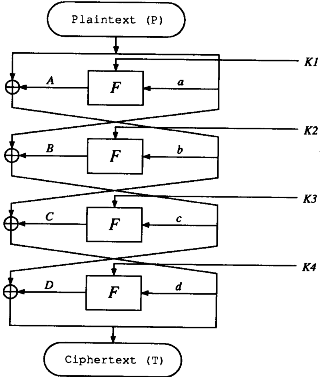

Figure 4: DES reduced to four rounds.

DES Reduced to Four Rounds

1. Using Ω^1, we have

$$c' = D' \oplus l' = a' \oplus B' \implies D' = B' \oplus l' \quad (8)$$

2. We have $a' = 0_x \implies A' = 0_x$ and $b' = A' \oplus L' = L'$.
   - In the second round S2, . . . , S8 receive zero XOR input.
   - 28 bits of B' are zero and hence we can find *28 bits of D'*.

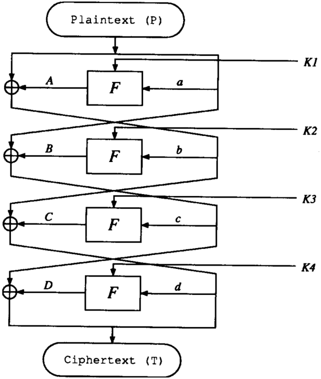

Figure 4: DES reduced to four rounds.

DES Reduced to Four Rounds

1. Using Ω^1, we have

$$c' = D' \oplus l' = a' \oplus B' \implies D' = B' \oplus l' \quad (8)$$

2. We have $a' = 0_x \implies A' = 0_x$ and $b' = A' \oplus L' = L'$.
   - In the second round S2, ..., S8 receive zero XOR input.
   - 28 bits of B' are zero and hence we can find *28 bits of* D'.
   - We already know $d' = r'$. So, we employ a counting approach to get $K4$.

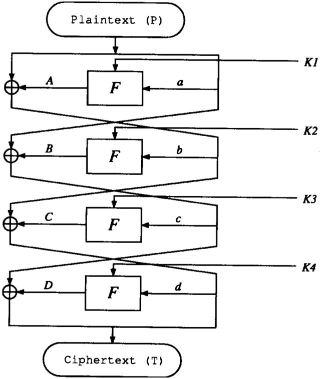

Figure 4: DES reduced to four rounds.

DES Reduced to Four Rounds

1. To get Si_{Kd} for $2 \leq i \leq 8$, we verify (9).

$$S(S_{Ed} \oplus S_{Kd}) \oplus S(S^*_{Ed} \oplus S_{Kd}) = S'_{Od} \tag{9}$$

2. Only *one* plaintext pair is needed since characteristic probability is 1.
3. We recover $7 \times 6 = 42$ key bits of $K4$, which correspond to 42 bits of the master key.
4. Exhaustively search the other 14 key bits to get the entire master key.
5. We have used the key schedule to our advantage here? *What if all the keys were independent?*

DES Reduced to Four Rounds: Independent Subkeys

1. We now use Ω^2 to get the remaining 6 subkey bits of $K4$, as the input to S1 in the second round is now zero.

DES Reduced to Four Rounds: Independent Subkeys

1. We now use Ω^2 to get the remaining 6 subkey bits of $K4$, as the input to S1 in the second round is now zero.
2. We have $C' = b' \oplus d'$. Peeling off/decrypting one round will give us c' completely.

DES Reduced to Four Rounds: Independent Subkeys

1. We now use Ω^2 to get the remaining 6 subkey bits of $K4$, as the input to S1 in the second round is now zero.
2. We have $C' = b' \oplus d'$. Peeling off/decrypting one round will give us c' completely.
    - Since c' and C' are both completely known, $K3$ can be completely found using a similar counting argument.

DES Reduced to Four Rounds: Independent Subkeys

1. We now use Ω^2 to get the remaining 6 subkey bits of $K4$, as the input to S1 in the second round is now zero.
2. We have $C' = b' \oplus d'$. Peeling off/decrypting one round will give us c' completely.
   - Since c' and C' are both completely known, $K3$ can be completely found using a similar counting argument.
3. Since $a' = A' = 0_x$, all keys are equally likely. Other characteristics Ω^3 and Ω^4 are chosen such that
   - $S'_{Ea} \neq 0_x$ for all S boxes for both characteristics.
   - For every S box, the S'_{Ea} values differ between the characteristics.
   - Similar counting methods used to get $K1$ and $K2$.

DES Reduced to Four Rounds: Independent Subkeys

1. We now use Ω^2 to get the remaining 6 subkey bits of $K4$, as the input to S1 in the second round is now zero.
2. We have $C' = b' \oplus d'$. Peeling off/decrypting one round will give us c' completely.
   - Since c' and C' are both completely known, $K3$ can be completely found using a similar counting argument.
3. Since $a' = A' = 0_x$, all keys are equally likely. Other characteristics Ω^3 and Ω^4 are chosen such that
   - $S'_{Ea} \neq 0_x$ for all S boxes for both characteristics.
   - For every S box, the S'_{Ea} values differ between the characteristics.
   - Similar counting methods used to get $K1$ and $K2$.
4. 16 chosen plaintexts are needed for this attack.
   - 8 pairs of Ω^1 and Ω^2 each.
   - 4 pairs of Ω^3 and Ω^4 each.

To reduce the data needed, two octets are used.

DES Reduced to Six Rounds

1. Two three-round characteristics used, each with probability $\frac{1}{16}$.

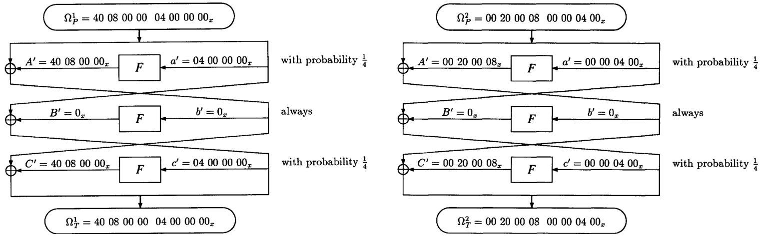

Figure 5: Characteristics used for cryptanalysis of DES reduced to 6 rounds.

DES Reduced to Six Rounds

1. Two three-round characteristics used, each with probability $\frac{1}{16}$.
2. We have,

$$e' = c' \oplus D' = F' \oplus l' \implies F' = c' \oplus D' \oplus l' \tag{10}$$

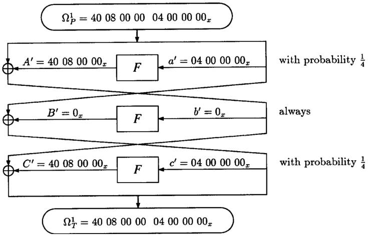

Figure 5: Characteristics used for cryptanalysis of DES reduced to 6 rounds.

DES Reduced to Six Rounds

1. In the fourth round,
   - with Ω^1, S2, S5, ..., S8 have zero input XORs.
   - with Ω^2, S1, S2, S4, S5 and S6 have zero input XORs.

DES Reduced to Six Rounds

1. In the fourth round,
   - with Ω^1, S2, S5, . . . , S8 have zero input XORs.
   - with Ω^2, S1, S2, S4, S5 and S6 have zero input XORs.
2. Combining both characteristics, 42 key bits of $K6$ can be found.

DES Reduced to Six Rounds

1. In the fourth round,
   - with Ω^1, S2, S5, ..., S8 have zero input XORs.
   - with Ω^2, S1, S2, S4, S5 and S6 have zero input XORs.
2. Combining both characteristics, 42 key bits of $K6$ can be found.
3. Counting on more bits gives high S/N at the cost of exponentially more memory.

DES Reduced to Six Rounds

1. In the fourth round,
    - with Ω^1, S2, S5, . . . , S8 have zero input XORs.
    - with Ω^2, S1, S2, S4, S5 and S6 have zero input XORs.
2. Combining both characteristics, 42 key bits of $K6$ can be found.
3. Counting on more bits gives high S/N at the cost of exponentially more memory.
4. Due to higher S/N, fewer plaintext pairs are analyzed. *This is exploited to get a faster counting algorithm*.

The Clique Method

1. Used to reduce memory when few plaintexts are used to count on more subkey bits.

The Clique Method

1. Used to reduce memory when few plaintexts are used to count on more subkey bits.
2. Create a graph where
   - Each plaintext pair is a vertex.
   - There is an edge between two vertices if corresponding pairs suggest the same key value for an S box.

The Clique Method

1. Used to reduce memory when few plaintexts are used to count on more subkey bits.
2. Create a graph where
   - Each plaintext pair is a vertex.
   - There is an edge between two vertices if corresponding pairs suggest the same key value for an S box.
3. The edges are labelled with five 64-bit masks (one mask per S box, one bit per suggested key value in the mask).
   - A pair suggests a key value if it passes the check in (9).

The Clique Method

1. Used to reduce memory when few plaintexts are used to count on more subkey bits.
2. Create a graph where
   - Each plaintext pair is a vertex.
   - There is an edge between two vertices if corresponding pairs suggest the same key value for an S box.
3. The edges are labelled with five 64-bit masks (one mask per S box, one bit per suggested key value in the mask).
   - A pair suggests a key value if it passes the check in (9).
4. Goal is to find the largest clique such that the bitwise AND of all masks in the subgraph induced by that clique is nonzero.

The Clique Method

1. Used to reduce memory when few plaintexts are used to count on more subkey bits.
2. Create a graph where
   - Each plaintext pair is a vertex.
   - There is an edge between two vertices if corresponding pairs suggest the same key value for an S box.
3. The edges are labelled with five 64-bit masks (one mask per S box, one bit per suggested key value in the mask).
   - A pair suggests a key value if it passes the check in (9).
4. Goal is to find the largest clique such that the bitwise AND of all masks in the subgraph induced by that clique is nonzero.
5. Apply this method for both Ω^1 and Ω^2, ensuring that the suggested keys at S2, S5 and S6 match. Otherwise, use more data.

Completing the Cryptanalysis

1. 42 key bits have been found, thus exhaustive search can be performed on the remaining 14 bits.
2. To speed up the search, we can find the remaining 6 key bits of $K6$ using Figure 6. Count using checks on S2, S3 and S8 of the fifth round.
   - Remaining 8 bits can be exhaustively searched.
   - Wrong pairs should be discarded by checking if they satisfy the characteristic and expected value of F'.
   - This will leave us with $\frac{1}{16}$ of the pairs, which boosts S/N greatly.

Into S box number	e bits S_{Ee}	Key bits S_{Ke}
S1	+ + + + + +	**3** + . . + +
S2	+ + **3** + + +	+ **3** + **3** **3** **3**
S3	+ + + + + +	+ + + + + +
S4	+ + + + **3** +	+ + . . + +
S5	**3** + + + + +	+ + + . + +
S6	+ + + + **3** +	+ . + . + +
S7	**3** + + + + +	+ + + . + +
S8	+ + **3** + + +	+ + + + + +

Figure 6: Dependence of $K5$ on bits of $K6$. '3' indicates dependence on $S3_{Kf}$, '.' indicates bits unused in K6 and '+' indicates dependence on known key bits of K6.

Data Requirements

1. The first phase has

$$S/N = \frac{2^{30} \cdot \frac{1}{16}}{4^5} = 2^{16}. \tag{11}$$

Only 7-8 pairs are needed for each characteristic. Since each characteristic has probability $\frac{1}{16}$, we require about 120 pairs of plaintexts.

2. The second phase has

$$S/N = \frac{2^6 \cdot 1}{4} = 16. \tag{12}$$

Though S/N is lesser, we can use the 7-8 right pairs from the first part.

3. We can reduce the data required by using quartets. In total, about 240 ciphertexts are needed.