25 Years of Linear Cryptanalysis

-Early History and Path Search Algorithm-

Asiacrypt 2018, December 3 2018

Mitsuru Matsui

Back to 1990…

FEAL (Fast data Encipherment ALgorithm)

- Designed by Miyaguchi and Shimizu (NTT).
- 64-bit block cipher family with the Feistel structure.
  - 4 rounds (1987)
  - 8 rounds (1988)
  - N rounds(1990) N=32 recommended
- Key size is 64 bits (later extended to 128 bits).
- Optimized for 8-bit microprocessors (no lookup tables).
- First commercially successful cipher in Japan.
- Inspired many new ideas, including linear cryptanalysis.

FEAL-NX Algorithm [Miyaguchi 90]

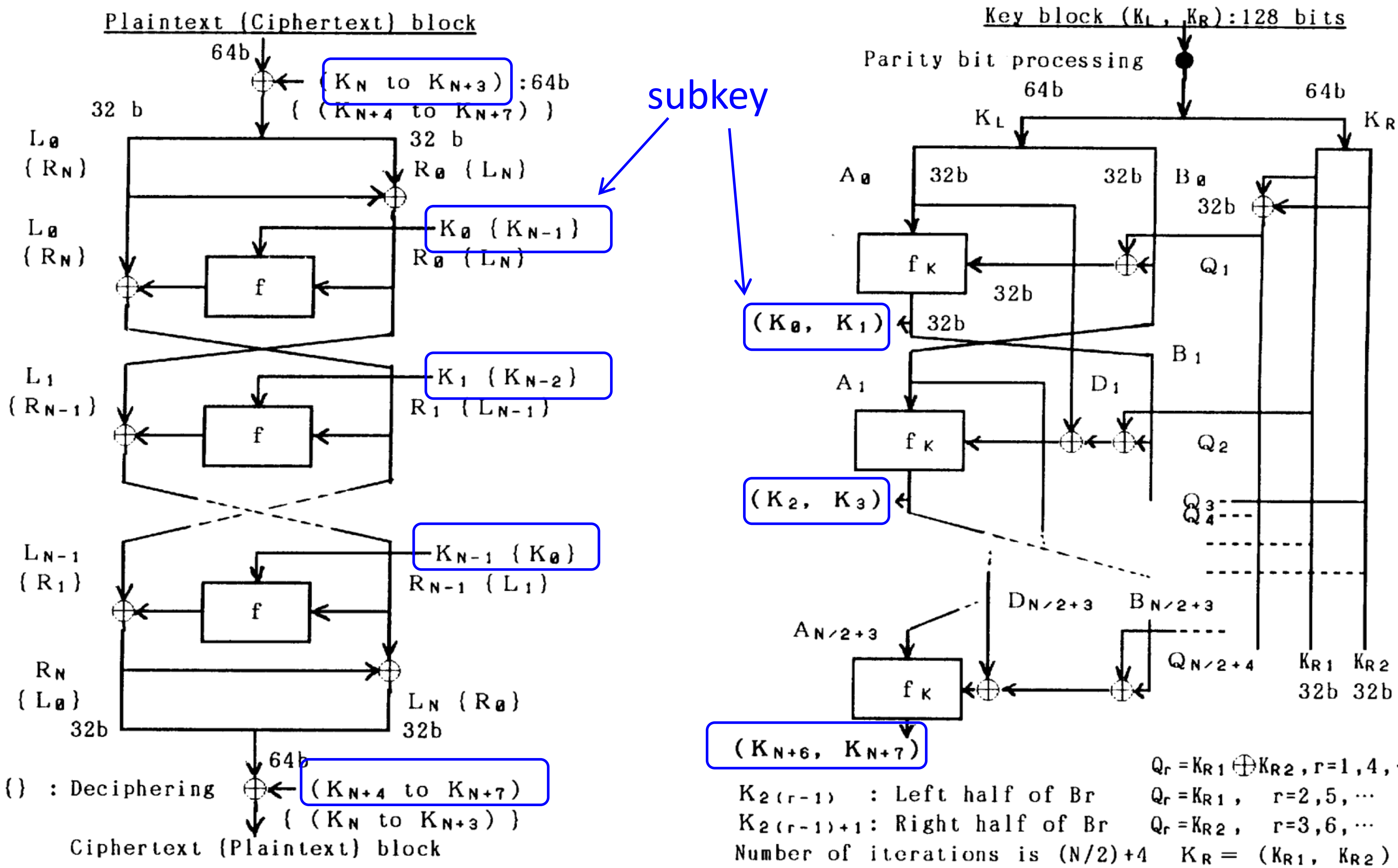
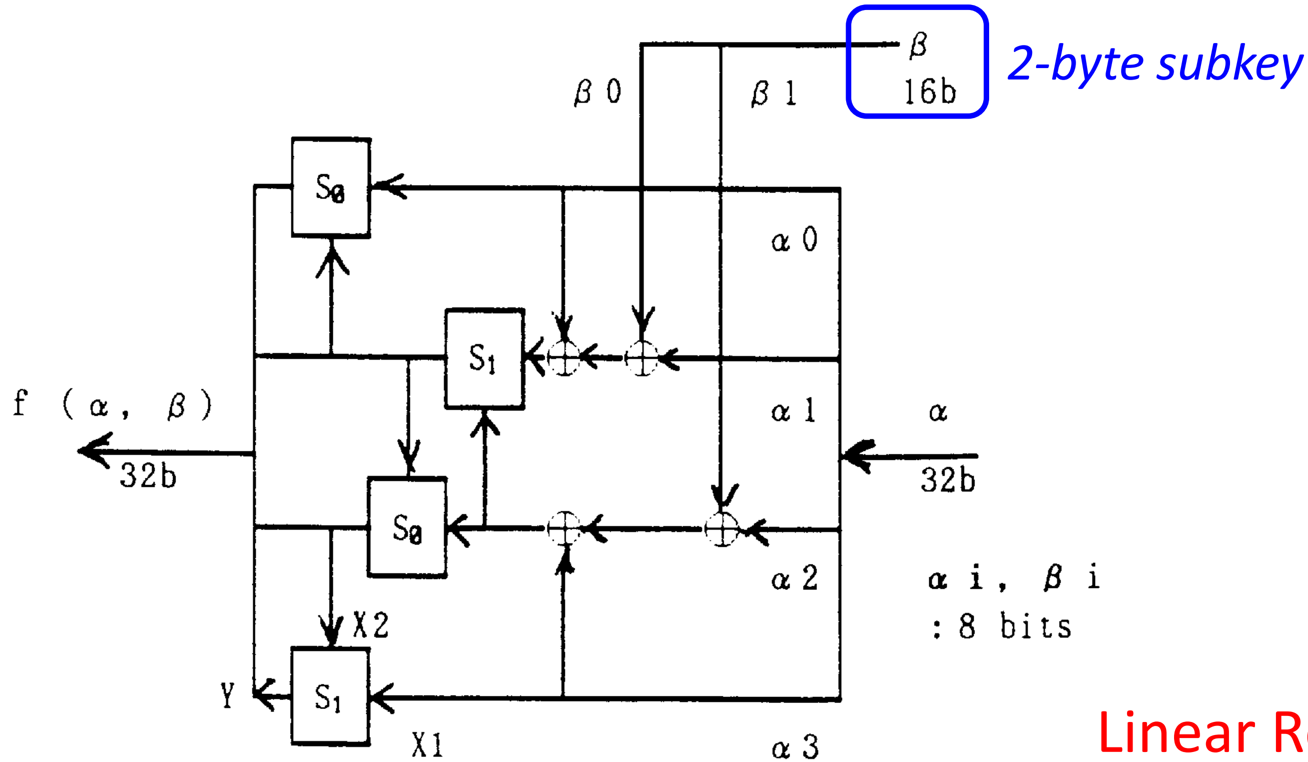


Fig. 1 Data Randomization

Fig. 2 Key Schedule of FEAL (FEAL-NX)

The Round Function of FEAL



Y = S0 (X1,X2) = Rot2((X1+X2) mod256)
Y = S1 (X1,X2) = Rot2((X1+X2+1) mod256)
Y: output(8 bits), X1/X2(8 bits): inputs,
Rot2(Y): a 2-bit left rotation on 8-bit data Y

Fig. 3 f-function

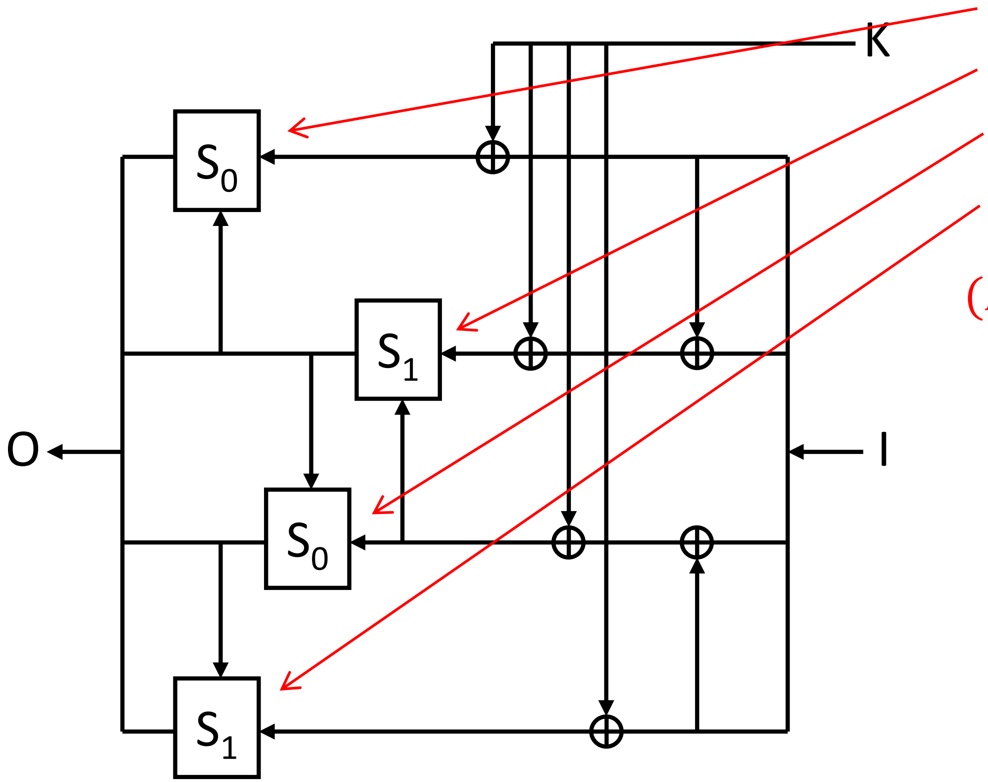
Linear Relations

$Y[2] = X1[0] \oplus X2[0]$

$Y[2] = X1[0] \oplus X2[0] \oplus 1$

$(Notation: Y[i] = i\text{-th bit of } Y)$

Linear Relations of the Round Function



modified round function f
(with whitening key)

Linear Relations

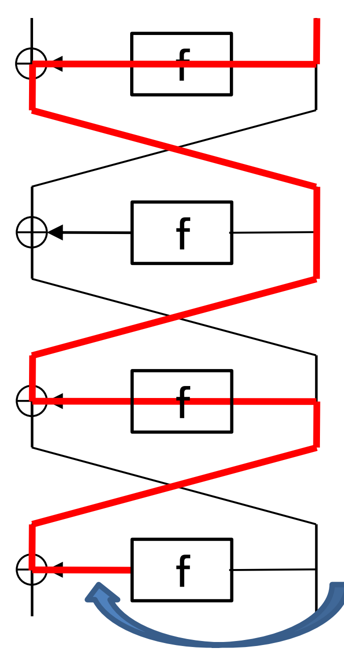
$O[16,26] \oplus I[24] = K[24]$

$O[18] \oplus I[0,8,16,24] = K[8,16] \oplus 1$

$O[10,16] \oplus I[0,8] = K[8]$

$O[2,8] \oplus I[0] = K[0] \oplus 1$

$(Notation: A[i,j,k] = A[i] \oplus A[j] \oplus A[k])$



3-round linear relations with p=1

at least 3 subkey bytes affect output

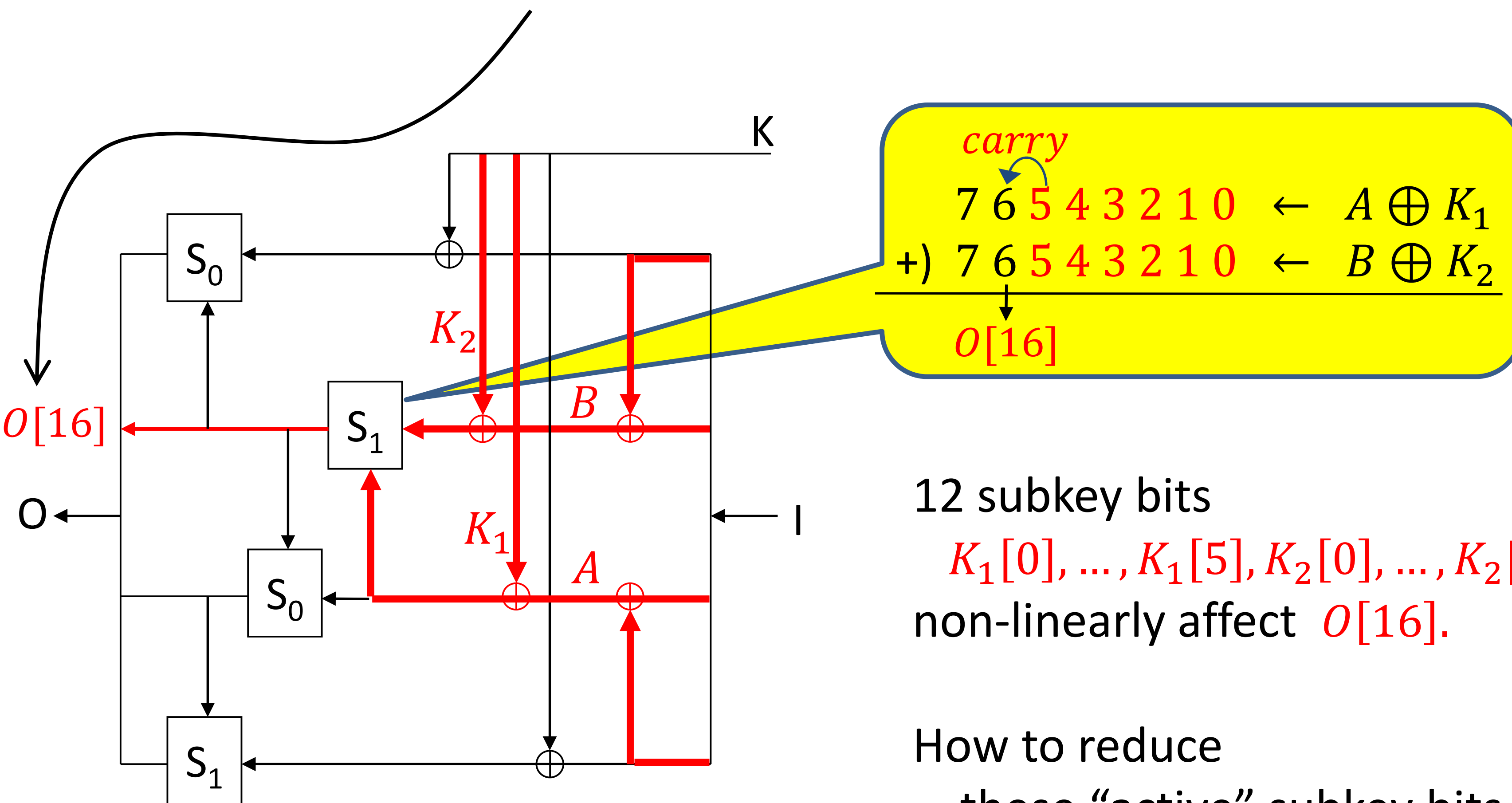
History of Cryptanalysis of FEAL

- 4-round version
  - 100-10000 chosen plaintexts [Boer 88]
  - 20 chosen plaintexts [Murphy 90]
  - 8 chosen plaintexts [Biham, Shamir 91] differential
  - 200 known plaintexts [Tardy-Corfdir, Gilbert 91]
  - 5 known plaintexts [Matsui, Yamagishi 92] pre-linear

- 8-round version
  - 10000 chosen plaintexts [Tardy-Corfdir, Gilbert 90] differential
  - 2000 chosen plaintexts [Biham, Shamir 91] differential
  - 2^{15}-2^{28} known plaintexts [Matsui, Yamagishi 92] pre-linear
  - 2^{24} known plaintexts [Biham 94] linear
  - 2^{10}-2^{11} known plaintexts [Sakikoyama et al 2016] multiple linear

A Better Relation and Pre-linear Attack

$O[10,18,26] \oplus I[16] = K[16,24] \oplus 1$ [Tardy-Corfdir, Gilbert 91]



Fourth (modified) round function

12 subkey bits

$K_1[0], \ldots, K_1[5], K_2[0], \ldots, K_2[5]$

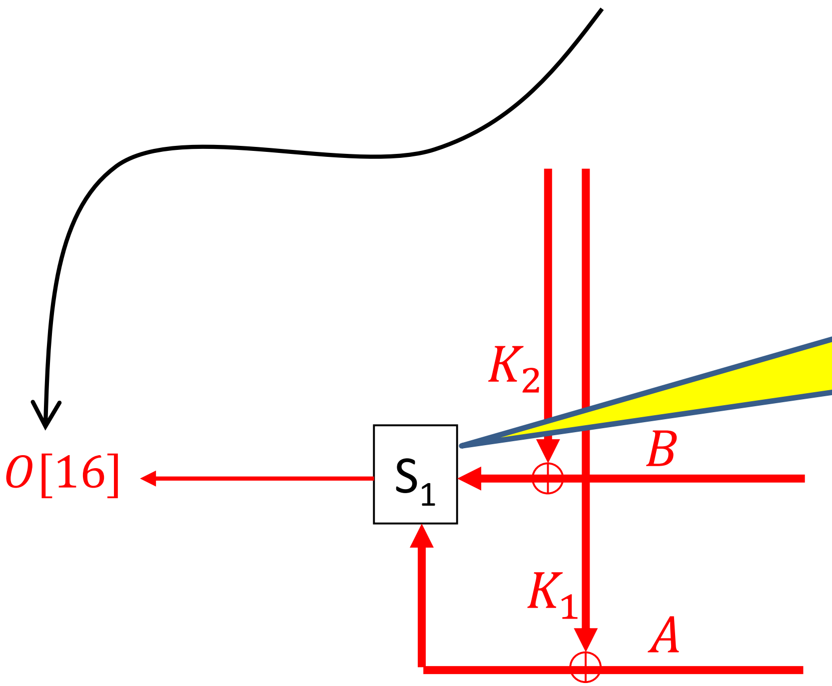
non-linearly affect $O[16]$.

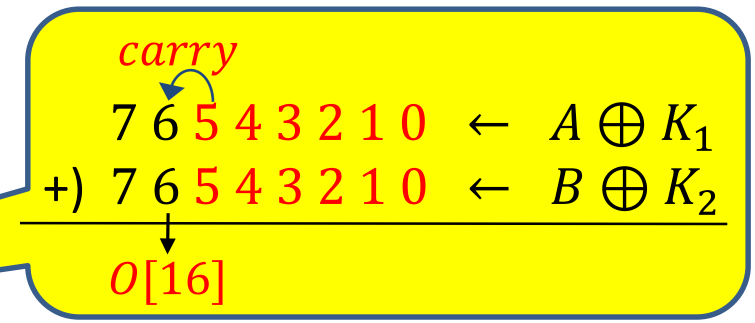
How to reduce

these "active" subkey bits ?

A Better Relation and Pre-linear Attack

$O[10, 18, 26] \oplus I[16] = K[16,24] \oplus 1$ [Tardy-Corfdir, Gilbert 91]





Observation [Matsui, Yamagishi 92]

Divide inputs into two groups
$A[5] = B[5]$ and $A[5] = \sim B[5]$.
Then for one of the two groups,
$A[6] \oplus A[5] \oplus B[6] \oplus O[16]$ is const.
This reveals
$K_1[5] = K_2[5]$ or $K_1[5] = \sim K_2[5]$.

The Cut-off Method

Later generalized as "Partitioning Cryptanalysis" [Harpes, Massey 97].

Divide plaintext-ciphertext pairs into two groups.

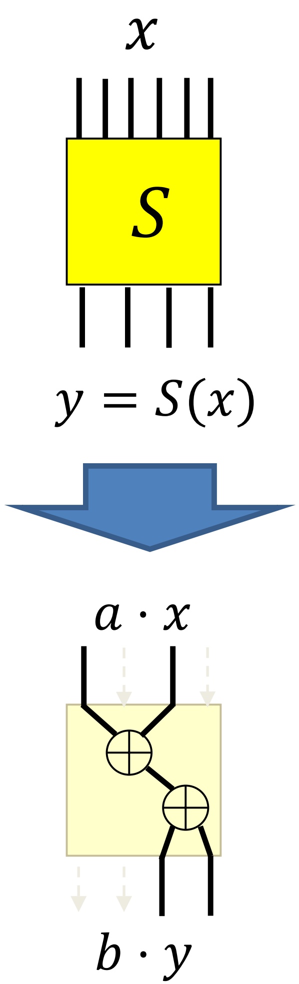
- For one group, a relation always holds or fails (p=1 or 0).
- For the other, the relation does not always hold (p≈1/2).
- "Which is which" depends on key.

A few pairs are enough to distinguish.

- Also reducing the number of relevant text/key bits.
- Better than (probabilistic) linear approximation.

Successful for FEAL, but applying this method to DES is difficult…

Moving to Linear Approximation



Linear Approximation Probability

$$p(a, b) = \{x \mid a \cdot x = b \cdot y\}/2^n$$

a, b: linear characteristic or masking value

$0 \leq p(a, b) \leq 1$

Linear Correlation

$$\lambda p_S(a, b) = 2p(a, b) - 1$$

$|\lambda p_S(a, b)|$ is larger = more linear.

$0 \leq |\lambda p_S(a, b)| \leq 1$

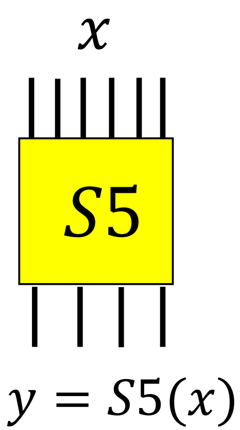
Differential Characteristic Probability

$$\delta p_S(a, b) = \{x \mid S(x) \oplus S(x \oplus a) = b\}/2^n$$

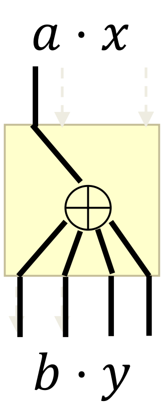
a, b: differential characteristic

$0 \leq \delta p_S(a, b) \leq 1$

An Example: S5 of DES



$X[3] = Y[2]$	$\lambda p_{S5}(8,4) = 0$	no use
$X[1] = Y[0]$	$\lambda p_{S5}(2,1) = 1/8$	makes sense
$X[4] = Y[0,1,2,3]$	$\lambda p_{S5}(16,15) = -5/8$	very biased

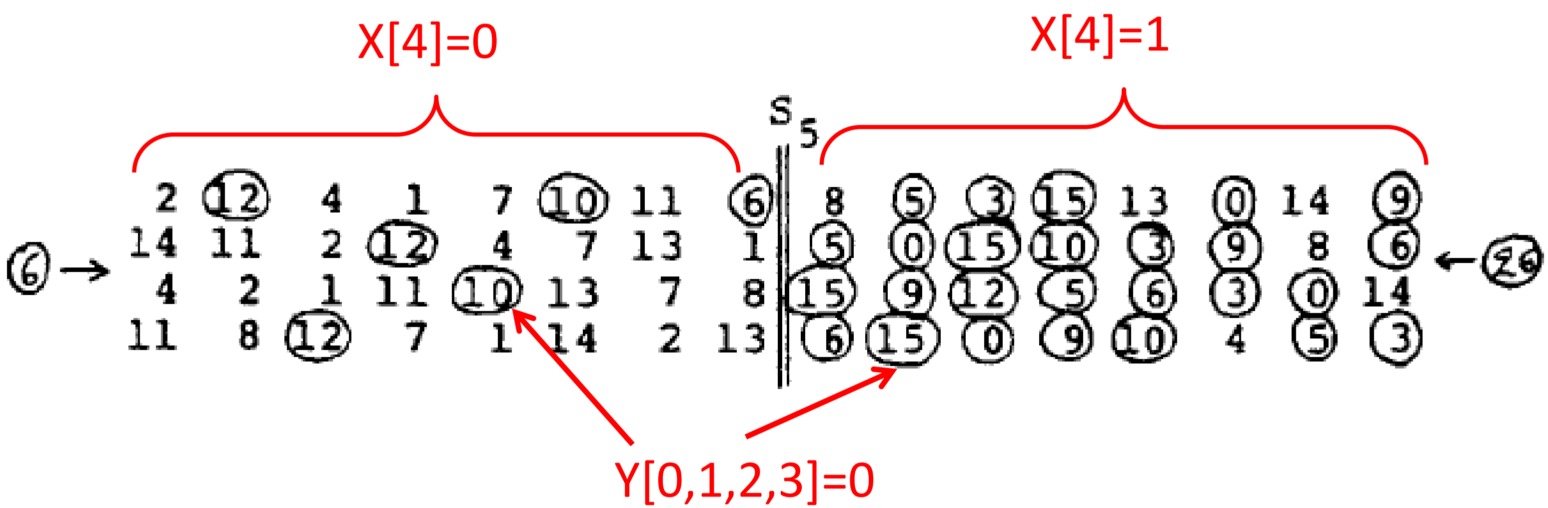


This relation was fully used for cryptanalyzing the full 16-round DES.

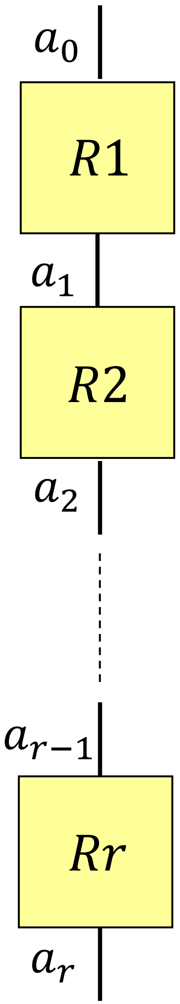
Adi Shamir (and independently Matt Franklin) had found this correlation already in 1985.

“On the Security of DES” [Shamir Crypto 85]

Fig. 1 describes our main observation. It circles all the WXYZ entries in which W⊕X⊕Y⊕Z = 0 (0,3,5,6,9,10,12,15). There is a clear correlation between this function and input bit B (which determines the left/right half of each S-box). Furthermore, the minorities in each half are located in such a way that there are exceptionally simple boolean polynomials (XOR's of AND's) which describe the 64 values of W⊕X⊕Y⊕Z in each S-box with very small number of errors. A detailed description of these observations, along with possible lines of attack based on them, will appear in the full paper.



Best Path Search Algorithm



- An algorithm deriving linear (resp. differential) path $a_0, a_1, \dots, a_r$ that maximize

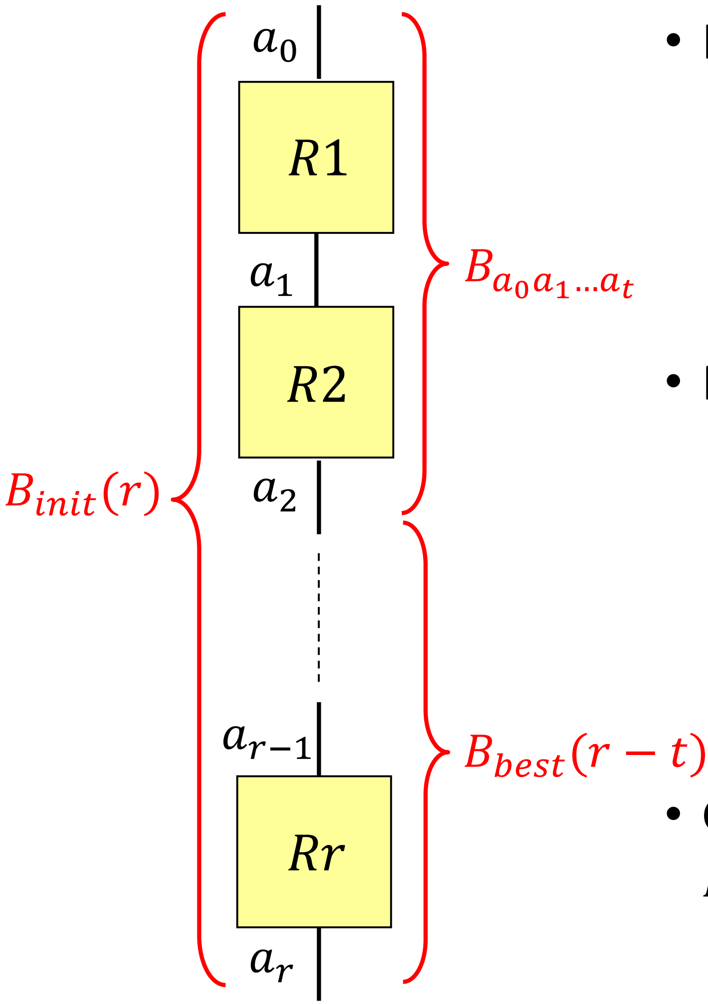
$$\prod_{i=1\dots r} |\lambda p_{R_i}(a_{i-1}, a_i)| \text{ (resp. } \delta p\text{).}$$

- Expected to lead to the most efficient attack.

- The best linear (resp. differential) path of the full DES was completely determined.

- Still being used as a tool for evaluating block ciphers (e.g. "multiple linear cryptanalysis").

The Algorithm: Sketch



- Induction on the number of rounds:
  - Computes $B_{best}(r)$ (and the corresponding $a_0, a_1, \ldots, a_r$) assuming $B_{best}(i)$ (i=1, … , r-1) is known.
- Needs an initial value $B_{init}(r)(< B_{best}(r))$
  - Works for any small value, but significantly affects performance.
  - Updated when a better path is found.
  - When finished, $B_{init}(r)$ must be $B_{best}(r)$.
- Given $a_0, a_1, \ldots, a_{t-1}$, choose a_t such that $B_{init}(r) < B_{a_0 a_1 \ldots a_t} * B_{best}(r-t)$ holds.
  - If not holds, no hope of finding a better path.

$B_{a_0 a_1 \ldots a_t} = \prod_{i=1 \ldots t} \lambda p_{R_i}(a_{i-1}, a_i)$

$B_{best}(r) = \max_{a_0 a_1 \ldots a_r} \prod_{i=1 \ldots t} \lambda p_{R_i}(a_{i-1}, a_i)$

The Branch-and-Bound Method

```
Path Search Algorithm
   Define B_init(n) for n=1,2,..,r
   Search1();

Search1()        /* 1st round characteristic */
   for each characteristic a_1
        choose a_0 so as to maximize |λp_R1(a_0,a_1)|
        Search2(2);

Search2(i)       /* i-th round characteristic i>1 */
   for each characteristics a_i  [in an decreasing order of |λp_Ri(a_i-1,a_i)|]
        if(a_i is not 'connectable' to a_i-1) continue;
        if(B_a0,a1,..,ai * B_best(r-i) < B_init(r)) return;
                                         /* no hope of this path */
        if(i == r)                       /* reached bottom */
             B_init(r) = B_a0,a1,..,ai;  /* found a better path ! */
        else
            Search2(i+1);                /* go to next round */
```

Time Complexity

- In practice, we need to design a double recursive algorithm to handle multiple active S-boxes in a single round.

- Generally exponentially heavy.
- Difficult to estimate computing time in advance.
- Found the best linear path of DES in a few minutes (in 1993), which is equivalent to an exhaustive path search.

- Speed depends on the quality of the initial value B_{init}.
- Limiting the number of active S-boxes per round often improves performance significantly.
  - Step1: Limited search using arbitrary small initial values.
  - Step2: Full search using the output of Step1 as initial values.

First Experiment of Breaking the Full DES

- July to September, 1993 (50 days)
- 12 computers (PA-RISC)
- 2^{43} known plaintext-ciphertext pairs (generated by M-sequence).
- 26 key bits from 2 equations derived from best 14-round linear path.
- Remaining 30 key bits were found by an exhaustive search.
- The biggest threat was the “thunder attack”.

One of the 12 computers (PA-RISC 99MHz)

DES variants with permuted S-boxes

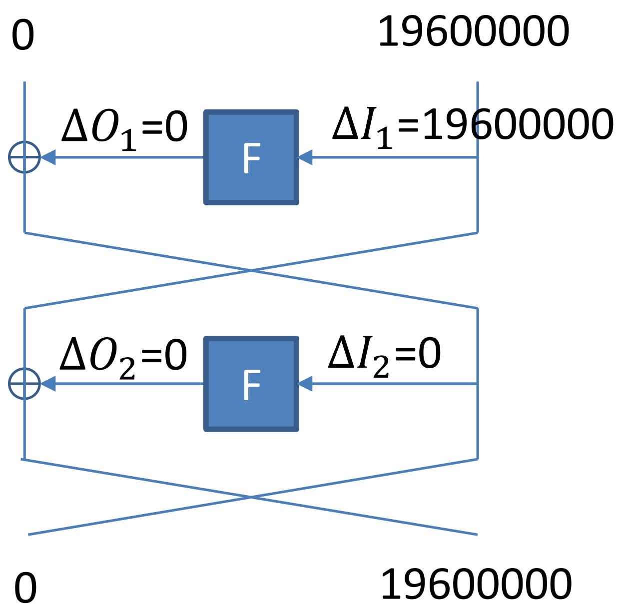
- Best linear/differential path search for all (8!=40320) permutations.
- For permuted DES variants, partial results were known [Matsui94]

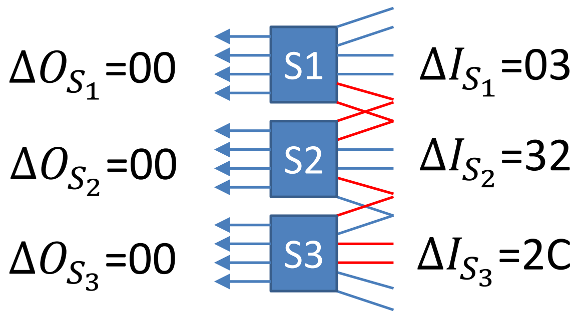
  The only following paths were searched:
  - (differential) Biham-Shamir's two-round iterative characteristics.
  - (linear) one active S-box per round (type I)

    or two-round iterative characteristics (type II).
- Complete search for all permutation patterns [Unpublished]
  - (differential) confirmed that known best is actually best.
  - (linear) better paths newly found for some permutations.

Biham-Shamir's 2-round Iterative Characteristic

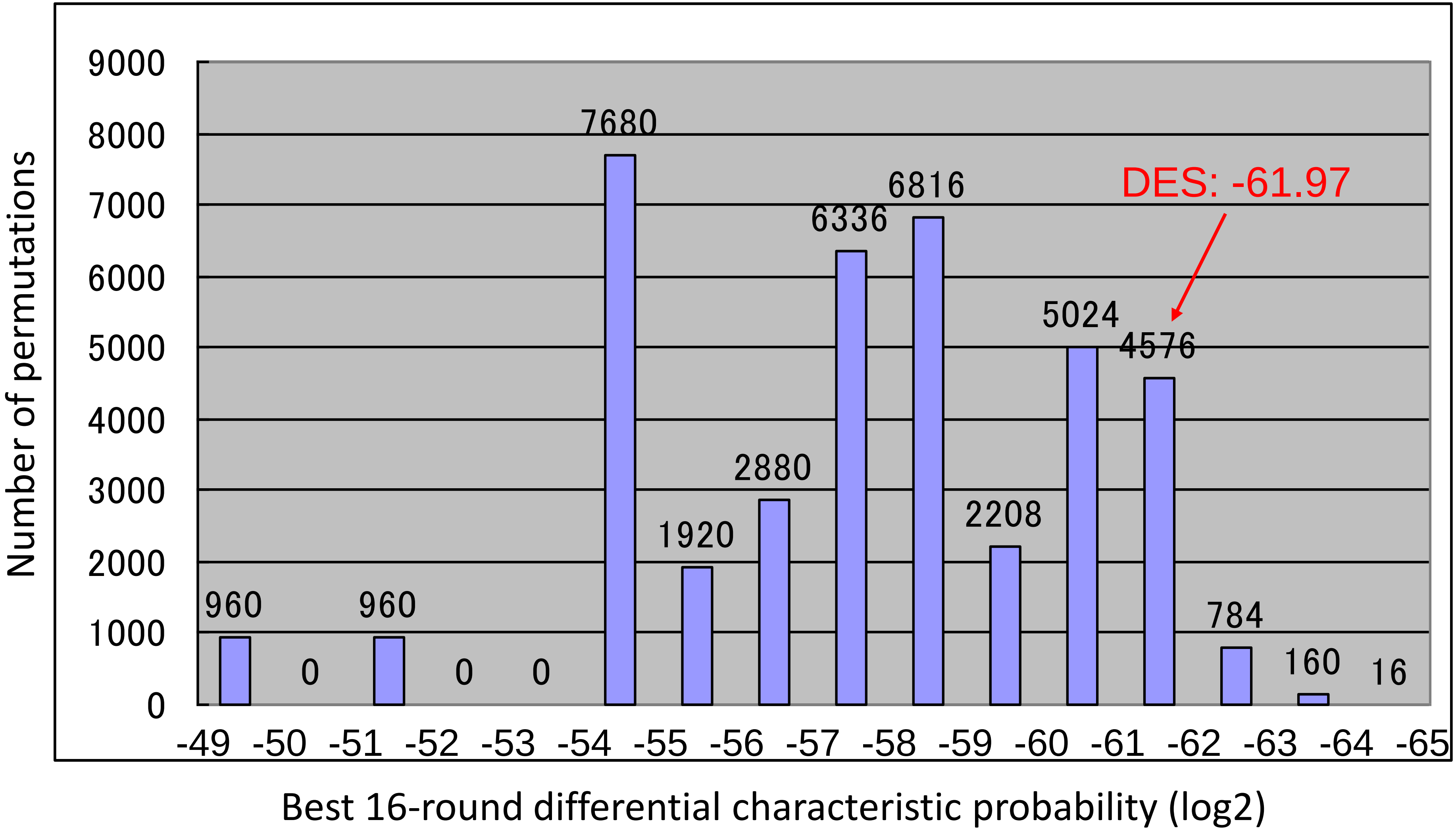
Three adjacent active S-boxes in every other round



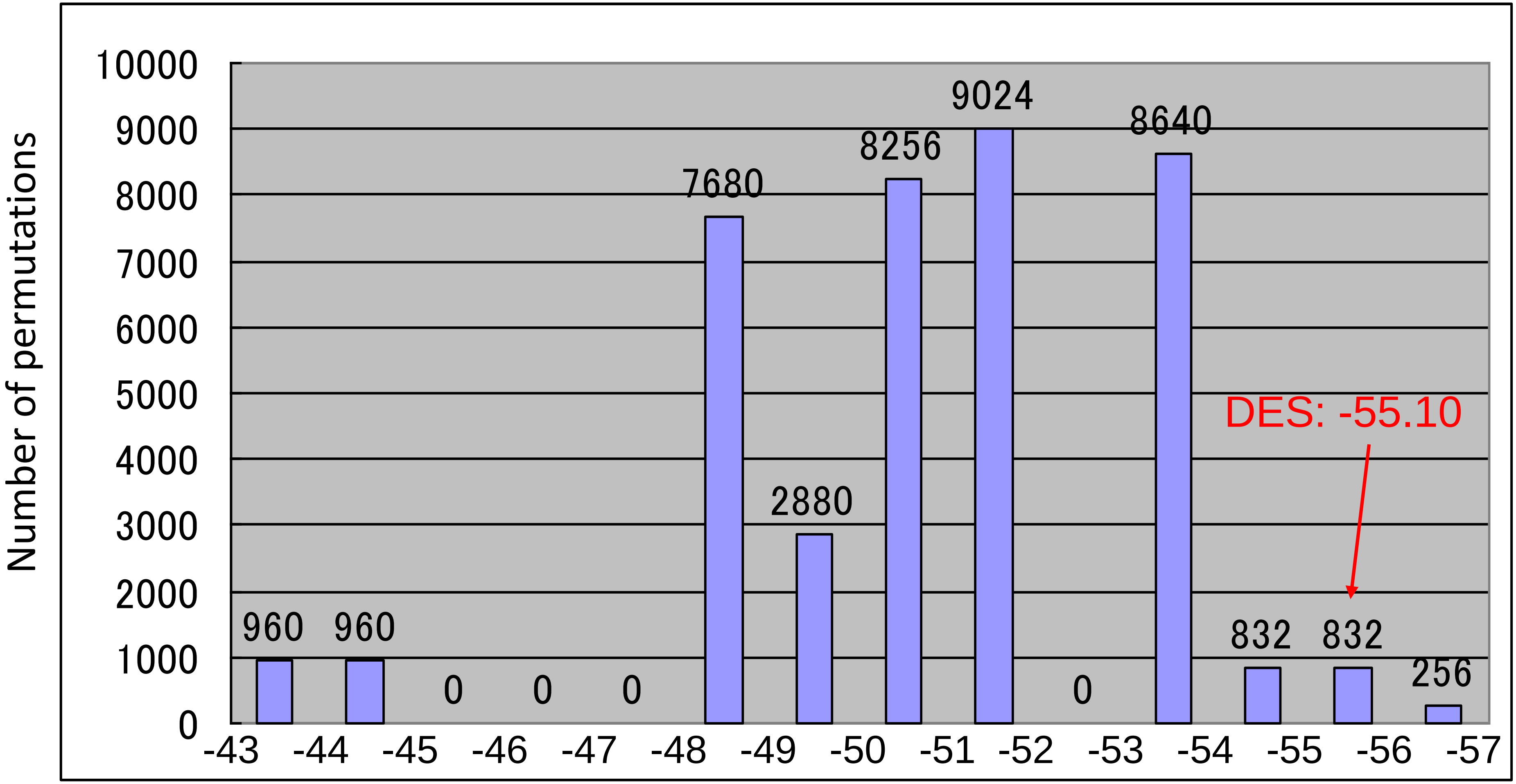


For all permuted DES variants, this type of characteristic achieves best path (part of design criteria of DES).

Distribution of Best 16-round Differential Char Probability of Permuted DES Variants

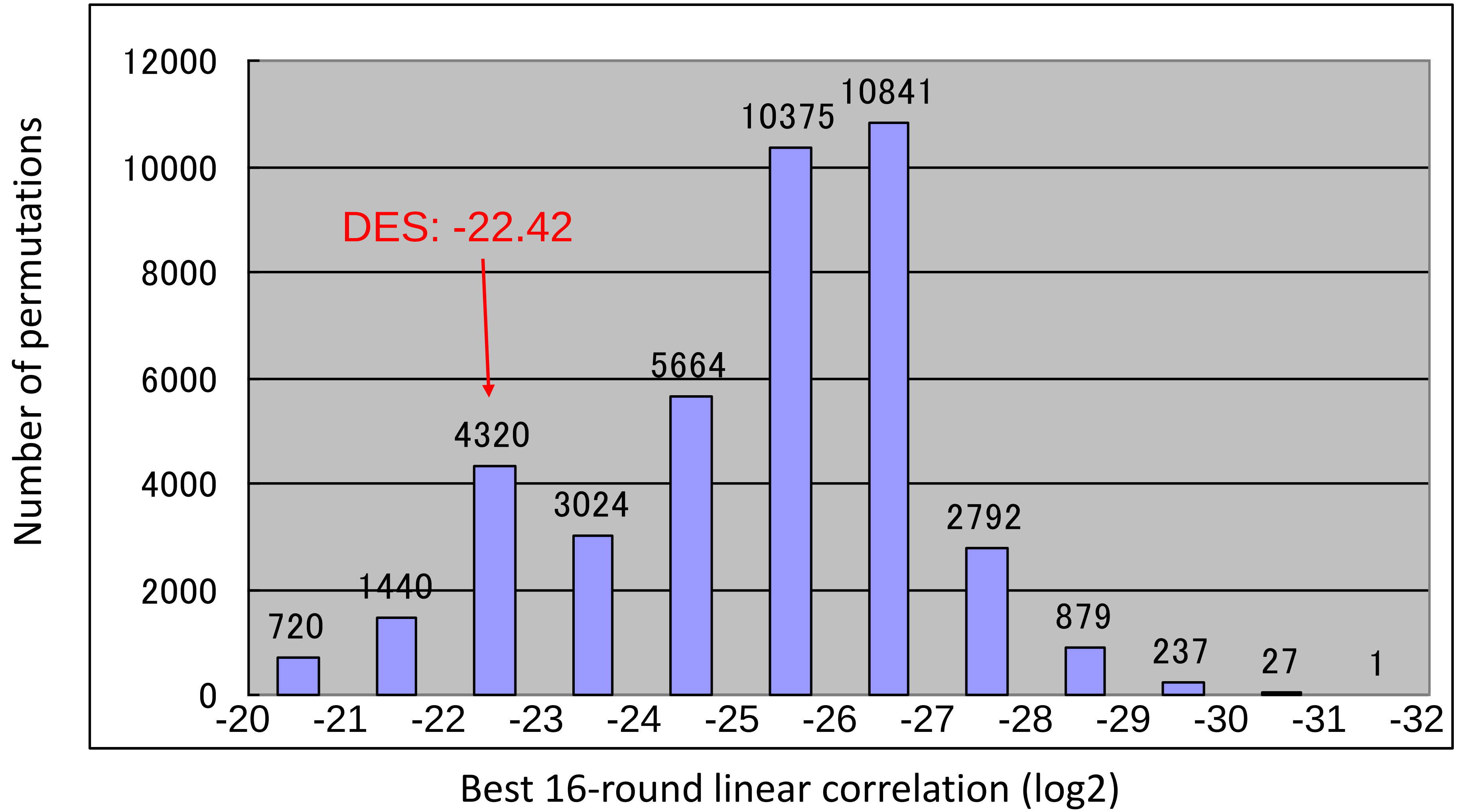


Distribution of Best 15-round Differential Char Probability of Permuted DES Variants

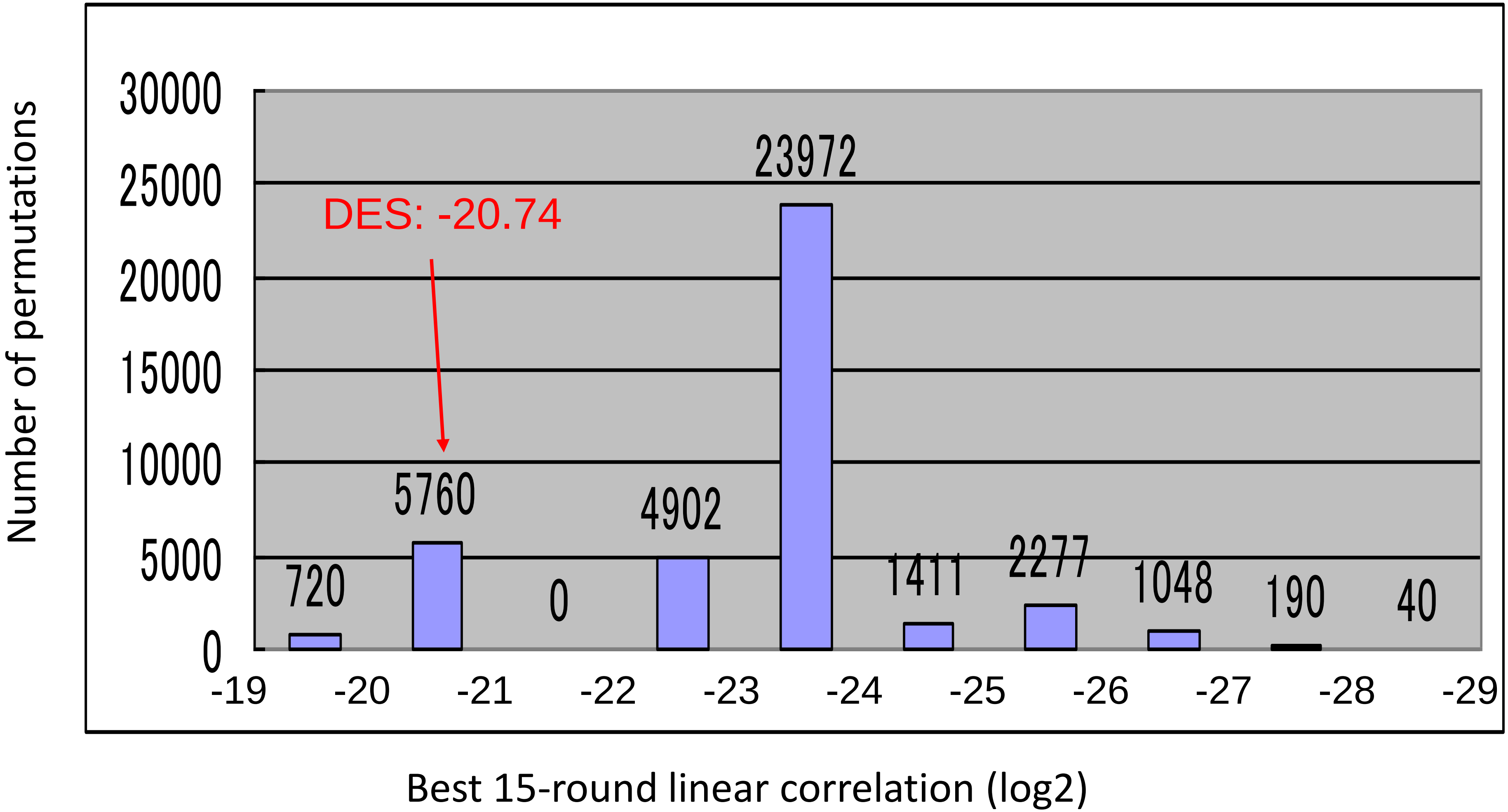


Best 15-round differential characteristic probability (log2)

Distribution of Best 16-round Linear Correlation of Permuted DES Variants



Distribution of Best 15-round Linear Correlation of Permuted DES Variants



Some Observations

- Some interesting linear paths:

  Ex.1: Non iterative 16-round best path

  7 – 7 8 7 – 7 8 7 – 7 32 7 – 7 5 $2^{-28.68}$ (3 perms)

  Ex.2: 11-round best path with 3 active S-boxes in a round

  1 5 – 5 1 456 1 5 – 5 1 $2^{-18.98}$ (29 perms)

- The permutation choice of the original DES is very weak.

  85% of the permutations are stronger. (=correlation is smaller)

- The designers of DES were unlucky.

  many permutations that strengthen DES against linear attack without weakening it against differential attack.

How to Find Better S-boxes for DES

"DES S-box Generator" [De Meyer, Vaudenay 2016]

- Quickly finds DES S-boxes satisfying Coppersmith's criteria with additional criteria for strong S-boxes.
- Found very strong S-boxes against differential cryptanalysis even stronger than any permuted DES variant.

	Differential characteristic prob.			Linear correlation		
	DES	Best of Permuted DES	[DV2016]	DES	Best of Permuted DES	[DV2016]
15-round	$2^{-55.10}$	$2^{-56.00}$	$2^{-61.81}$	$2^{-20.75}$	$2^{-28.00}$	$2^{-23.63}$
16-round	$2^{-61.97}$	$2^{-64.00}$	$2^{-69.32}$	$2^{-22.42}$	$2^{-31.32}$	$2^{-25.40}$

Coppersmith's Design Criteria [1994]

- Published after differential cryptanalysis was public.
- S-1 to S-8 for S-boxes, P-1 to P-3 for permutation P.
- Not mentioned how to generate such S-boxes.

(S-1) 6 input bits and 4 output bits.

(S-2) not too close to linear functions.

(S-3) fixing leftmost and rightmost input bits makes 4-bit permutation.

Coppersmith's Design Criteria [1994]

(S-4) If w(ΔI)=1, then w(ΔO)≥2. w: weight

(S-5) If ΔI=001100, then w(ΔO)≥2.

(S-6) If ΔI=11xy00, then ΔO≠0.

(S-7) For any j, ΔI≠0 and ΔO, #{x | S_j(x+ΔI)+S_j(x)=ΔO}≤16.

(S-8) Arrange S-boxes so as to minimize $\max_{j=1,\ldots,8} q_{0,j}q_{1,j+1}q_{2,j+2}$.

$q_{0,j}=\max_{c,d}$ prob{ΔO_{S_j}=0 | ΔI_{S_j}=00cd11}

$q_{1,j}=\max_{g,h}$ prob{ΔO_{S_j}=0 | ΔI_{S_j}=11gh10}

$q_{2,j}=\max_{k,m}$ prob{ΔO_{S_j}=0 | ΔI_{S_j}=10km00}

Coppersmith's Design Criteria Aimed…

- For any differential characteristic path, the average number of active S-boxes per round is at least 1.6 in 13 and 16 rounds, except the following case.

- The best differential characteristic path should be given by a repetition of (S-8) in every other round.
  - average number of active S-boxes per round is 1.5.
  - (2n+1)-round best probability = (prob. of (S-8))^{n}.
  - Equivalent to Biham-Shamir's 2-round iterative characteristic.

A Couple of Comments on (S-8)

- It looks that (S-8) says the order of S-boxes should be arranged so as to minimize $\max_{j=1,\ldots,8} q_{0,j}q_{1,j+1}q_{2,j+2}$. However, original DES does not hold this condition. Any other criteria about the order of S-boxes?

- For all known DES S-box variants satisfying Coppersmith's criteria, the characteristic of (S-8) gave the best differential path. Is this always correct?
→This presentation shows the first(?) counterexample.

Exploring Permuted Variants of [DV2016]

Permuting S-boxes (8!=40320 patterns)

Step 1: Compute (S-8) for all 40320 patterns and select good ones.
Step 2: Compute their best linear correlation and select good ones.
Step 3: Compute their best diff characteristic prob. to make sure they are actually good. (i.e. (S-8) actually gives best diff characteristic prob.)

	Differential characteristic prob.			Linear correlation		
	DES	[DV2016]	New Variant	DES	[DV2016]	New Variant
15-round	$2^{-55.10}$	$2^{-61.81}$	$2^{-61.81}$	$2^{-20.75}$	$2^{-23.63}$	$2^{-27.15}$
16-round	$2^{-61.97}$	$2^{-69.32}$	$2^{-69.48}$	$2^{-22.42}$	$2^{-25.40}$	$2^{-29.05}$

Step 3 was time consuming. All the Step 3 candidates passed the test.

Exploring Permuted Variants of [DV2016]

Allowing duplication of S-boxes (8^8=16777216 patterns)

Step 1: Compute (S-8) for all 16777216 patterns and select good ones.
Step 2: Compute their best linear correlation and select good ones.
Step 3: Compute their best diff characteristic prob. to make sure they are actually good. (i.e. (S-8) actually gives best diff characteristic prob.)

	Differential characteristic prob.			Linear correlation		
	DES	[DV2016]	New Variant	DES	[DV2016]	New Variant
15-round	$2^{-55.10}$	$2^{-61.81}$	$2^{-70.00}$	$2^{-20.75}$	$2^{-23.63}$	$2^{-29.13}$
16-round	$2^{-61.97}$	$2^{-69.32}$	$2^{-78.41}$	$2^{-22.42}$	$2^{-25.40}$	$2^{-31.13}$

For some permutation variants, non 2-round iterative characteristic gives the best path.

New type of best differential path

- Permutation Pattern 66666666 (eight same S-boxes!)

best 16-round path based on (S-8):

– 567 – 567 – 567 – 567 – 567 – 567 – 567 – 567 – 567

characteristic prob. = $2^{-81.58}$. (active S-boxes per round = 1.5)

actual best 16-round path derived by the search algorithm:

57 – 57 234 57 – 57 234 57 – 57 234 57 – 57 234

characteristic prob. = $2^{-73.56}$. (active S-boxes per round = 2)

Q: Does anything bad happen if using an S-box eight times?

Conclusions

- I would like to thank DES cipher, FEAL cipher, differential cryptanalysis, and all works that led to linear cryptanalysis.

- DES is still an attractive algorithm.
  - We can learn a lot.
  - We can attack a lot.
  - We can improve a lot.

- Let's thank FEAL cipher for its contribution to the history of modern block ciphers.