Linear Cryptanalysis and Modified DES with Embedded Parity Check in the S-boxes

Robert Tsenkov
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences

Joint work with
Yuri Borissov and Peter Boyvalenkov

BalkanCryptSec 2015 – Koper, Slovenia, September 3-4, 2015

Motivation

LINEAR CRYPTANALYSIS exploits LINEARITY.

PARITY means LINEARITY.

S-BOX "means" NONLINEARITY.

S-BOX + PARITY $\rightarrow$?

Contents

- Notations and Definitions.
- A brief overview of Matsui's Work.
- Our Experiment.
- Properties of LATs of S-boxes with Embedded Parity Check.
- The Decreasing Effectiveness for Small Number of Rounds.
- Construction of Best Characteristics.
- Best Probabilities and Best Characteristics.
- Comparison of Best Multi-round Approximate Expressions.
- Conclusions.

Notations and Definitions (1)

Let us consider a family of Feistel type cryptographic algorithms with n rounds and denote:

- P, C – the plaintext and the ciphertext, both $2b$ bits long;
- K, K_j – the secret key and the corresponding j-th round subkey;
- X_j, F_j – the j-th round transformation data input/output;
- $f_j(X_j, K_j)$ – the j-th round transformation with input/output size b;
- S_k – the k-th S-box in the cipher;

Notations and Definitions (2)

- $S(j)$ – the chosen active S-box in the j-th round, when there is at most one such S-box; $S(j) = \phi$ means no active S-box is chosen;
- $I_X(j)$, $I_K(j)$, $I_F(j)$ – sets of indices of all bits of X_j, K_j and F_j respectively, that take part in the j-th round linear approximation;
- $I_X(S_k, j)$, $I_X(S(j))$, $I_K(S_k, j)$, $I_K(S(j))$, $I_F(S_k, j)$, $I_F(S(j))$ – sets of indices of all bits of X_j, K_j and F_j respectively, that are part of the input and the output of $S_k/S(j)$ respectively in the j-th round;
- $A[i]$ – the i-th bit of the vector A;
- $A[i, j, \ldots, k] = A[i] \oplus A[j] \oplus \cdots \oplus A[k]$;
- $A[B]$ – the XOR-sum of all bits of A with indices in the set B.

Notations and Definitions (3)

- An *1-round linear characteristic* for the round j of a Feistel cipher is a pair $(I_X(j), I_F(j))$ of sets of bit indices from the input and the output of this round, respectively. An *n-round linear characteristic* for rounds $1, \ldots, n$, $n \geq 3$, is an n-tuple $((I_X(1), I_F(1)), \ldots, (I_X(n), I_F(n)))$ of 1-round linear characteristics with the property

$$I_F(j+1) = (I_F(j-1) \cup I_X(j)) \backslash (I_F(j-1) \cap I_X(j)) \qquad (1)$$

  for all $2 \leq j \leq n-1$ (i.e. if $I_F(j+1)$ is the symmetric difference of $I_F(j-1)$ and $I_X(j)$).
- Every characteristic is associated with a given pair of chosen nonempty subsets of indices of input/output bits and the corresponding probability of coincidence of their respective sums.

Notations and Definitions (4)

- For given S-box S_k with a-bit input and c-bit output and given numbers α and β, such that $0 \leq \alpha \leq 2^a - 1$ and $0 \leq \beta \leq 2^c - 1$, define

$$NS_k(\alpha, \beta) = \#\{x | 0 \leq x \leq 2^a - 1, \oplus_s(x[s] \circ \alpha[s]) = \oplus_t(S_k(x)[t] \circ \beta[t])\},$$

where the symbol $\circ$ denotes bitwise AND operator.
The table, where the vertical and the horizontal axes indicate α and β respectively, and each entry contains the value

$$NS_k^*(\alpha, \beta) = NS_k(\alpha, \beta) - 2^{a-1}$$

is referred to as *linear approximation table* (*LAT*) for S_k.

Notations and Definitions (5)

- For the purpose of comparing different characteristics we consider approximations of type

$$P[i_1, \ldots., i_p] \oplus C[j_1, \ldots, j_q] \oplus f_1(P, K_1)[u_1, \ldots, u_s] \oplus \\ \oplus f_n(C, K_n)[v_1, \ldots, v_t] = K[k_1, \ldots, k_r], \quad (2)$$

  that can be drawn from an $(n-2)$-round characteristic combined with the structural dependencies in the first and final round of the cipher, respectively.

- All key bits and text (plaintext and ciphertext) bits, that affect the left side of the equation (2) are referred to as *effective bits*.

Notations and Definitions (6)

- If a linear approximation holds with probability $p \neq 1/2$ for randomly given plaintext P and the corresponding ciphertext C, the absolute value of the *bias* $p - 1/2$ represents the *effectiveness* of that approximation.
- A linear characteristic is called *best characteristic* when the effectiveness of corresponding linear approximation is maximal. Respectively, its probability will be called *best probability*.
- **Example 1.** There is an unique best 1-round characteristic for DES, namely corresponding to the global minimum $NS_5^*(16, 15) = -20$ with effectiveness 0.31.

A brief overview of Matsui's Work (1)

Matsui (1993,1994)

- Analyzed LATs.
- Found best characteristics for 3 to 20 rounds.
- Studied some approaches for mounting attacks against different rounds of DES cipher.
- **Proposition 1.** (Matsui, 1993; as lemma)
  (i) $NS_k(\alpha, \beta)$ is even.
  (ii) If $\alpha = 1, 32$ or 33, then $NS_k(\alpha, \beta) = 32$ for all S_k and β.
- The effectiveness of an 1-round approximation is deduced directly from the LATs and for multi-round approximation the so-called *Piling-up Lemma* is applied.

A brief overview of Matsui's Work (2)

The first experimental cryptanalytic attack on DES (Matsui, 1994):

- Based on approximations of type (2) derived from best 14-round characteristics. Two such characteristics, symmetric to each other, with effectiveness of $e = 1.19 \times 2^{-21}$.
- Each of the exploited linear approximations - two active S-boxes; 13 different effective key bits, found by the maximum likelihood method. The remaining 30 unknown key - found by an exhaustive search.
- The complexity of the attack practically depends only on the effectiveness e of the approximations used and the output bits $u_1, \ldots, u_s, v_1, \ldots, v_t$.
- The number of plaintext/ciphertext pairs used is proportional to e^{-2}.
- The result: DES is breakable with complexity 2^{43} at success rate of 85% if 2^{43} known plaintexts are available.

Some references

- M. Matsui, Linear cryptanalysis method of DES cipher, *Advances in Cryptology*, EUROCRYPT'93, in Lect. Notes Comp. Sci. **765**, Springer, 1993, 386-397.
- M. Matsui, Linear cryptanalysis of DES cipher (I), version 1.03, (see, for example http://www.cs.bilkent.edu.tr/~selcuk/teaching/cs519/Matsui-LC.pdf).
- M. Matsui, The first experimental cryptanalysis of the Data Encryption Standard, *Advances in Cryptology*, CRYPTO'94, in Lect. Notes Comp. Sci. **839**, Springer, 1994, 1-11.
- M. Matsui, On correlation between the order of S-boxes and the strength of DES, *Advances in Cryptology*, EUROCRYPT'94, in Lect. Notes Comp. Sci. **950**, Springer, 1995, 366-375.
- L. R. Knudsen, Practically Secure Feistel Ciphers, FSE'93, in Lect. Notes Comp. Sci. **809**, Springer, 1994, 211-221.

Our Experiment (1)

Main steps:

- Embedding parity check bit in all S-boxes of the original DES cipher and analyzing the newly obtained LATs.
- Finding best characteristics for 3 to 20 rounds of the modified cipher when the parity bit position is the same for all S-boxes.
- Comparing the results obtained to that for the original cipher.
- Studying in details the 16-round linear approximations based on the best 14-round characteristics found for modified DES.

Without loss of generality we assume embedding of odd parity.

Our Experiment (2)

- π_k: Parity bit mask in S-box S_k.
- π_k values: 0, 1, 2, 4 and 8 or their 4-bit representations. $\pi_k = 0$ means no parity is embedded.
- "case π": When all S-boxes have the same parity bit mask π.
- $S_{k(\pi)}$, $NS_k(\pi; \alpha, \beta)$ and $NS_k^*(\pi; \alpha, \beta)$: S-box, obtained from S_k by embedding a parity bit with mask π and the corresponding values in the LATs.

- **Example 2.** Result from embedding the parity check with mask 0100 on the output of S_7 (in hexadecimal format):

index	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
S_7	4	d	b	0	2	b	e	7	f	4	0	9	8	1	d	a
$S_{7(0100)}$	4	d	b	4	2	b	e	7	b	4	4	d	8	1	d	e

LATs with Embedded Parity Check: Non-zero Masks (1)

- $\alpha \neq 0, \beta \neq 0$
- **Proposition 2.** Let S_k be an S-box of DES. Let $\pi \neq 0$ be an odd parity bit mask on the output of S_k and $\&$ denotes tuple-wise AND operator. Then:

  (i) $NS_k^*(\pi; \alpha, 15) = 0$ for all α;

  (ii) $NS_k^*(\pi; \alpha, \beta) = NS_k^*(\alpha, \beta)$ for all α and β such that $\beta \& \pi = 0$;

  (iii) $NS_k^*(\pi; \alpha, \beta) = -NS_k^*(\alpha, 15 - \beta)$ for all α and $\beta < 15$ such that $\beta \& \pi \neq 0$.
- *Remark 1.* Embedding even parity will give symmetric LATs in the part of non-zero masks.

LATs with Embedded Parity Check: Non-zero Masks (2)

- **Example 3.** A part of LAT $NS^*_{7(0100)}$ with non-zero input and output masks.

β/α	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
02	0	2	-6	0	0	-2	-2	2	2	0	0	6	-2	0	0
03	0	-2	6	4	4	-2	-2	2	2	-4	-4	-6	2	0	0
04	0	2	2	-2	-10	-4	-4	4	4	10	2	-2	-2	0	0
05	0	2	10	2	-6	8	0	0	-8	6	-2	-10	-2	0	0
06	4	0	-4	-2	2	6	2	-2	-6	-2	2	4	0	-4	0

The columns 01, 02, 03, 08, 09, 10 and 11 stay the same as in the (original) NS^*_7, because the parity bit does not participate there.

LATs with Embedded Parity Check: Zero Masks

- $\alpha = 0$ or $\beta = 0$
- **Proposition 3.** For any S-box S_k of DES and an odd parity bit mask $\pi \neq 0$ applied to its output, we have:

  (i) $NS_k^*(\pi; 0, \beta) = 0$ for all $\beta : 15 > \beta > 0$;

  (ii) $NS_k^*(\pi; 0, 15) = -32$ while $NS_k^*(\pi; 0, 0) = 32$;

  (iii) $NS_k^*(\pi; \alpha, 0) = 0$ for all $\alpha \neq 0$.
- By contrast to the original DES for every modified S-box there exists an 1-round linear characteristic with zero input mask and non-zero (namely, 15) output mask having non-zero bias, but it has probability 1. Using that characteristic does not contribute anything more to the multi-round approximation compared to the trivial zero-to-zero characteristic except an additive constant equal to the parity sum.

Decreasing Effectiveness for Small Number of Rounds (1)

- **Proposition 4.** (i) The number of non-trivial 1-round characteristics needed to create a multi-round characteristic for the considered Feistel ciphers is (at least) two for every three rounds;

  (ii) One can construct best 3-round characteristic making use two times of best 1-round non-trivial characteristic with non-zero input mask.

- Construction: The characteristic "$A - A$" for any A being best 1-round characteristic with non-zero input mask, becomes best 3-round characteristic with effectiveness $e_3 = 2(e_1)^2$ where e_1 is the effectiveness of A.

- The first claim of Proposition 4 we call *modified Knudsen observation* (Knudsen, 1994).

Decreasing Effectiveness for Small Number of Rounds (2)

- Taking in account that by embedding parity check we always eliminate the best value in the original LATs, we get the inequality

$$\max_{(k,\alpha,\beta)} |NS_k^*(\pi;\alpha,\beta)| < |NS_5^*(16,15)| = \max_{(k,\alpha,\beta)} |NS_k^*(\alpha,\beta)|.$$

- **Theorem 1.** Any non-zero parity mask applied to the S-boxes of DES leads to a reduction of the highest effectiveness of the 1-round and 3-round linear characteristics obtained within the framework of linear cryptanalysis with "at most one active S-box per round".

- *Remark 2.* An additional fact of interest concerning DES is that we can eliminate the second maximal value of the LATs in case of applying the parity mask $\pi = 0100$.

Construction of Best Characteristics: Basic Search Algorithm

- *Basic Search Algorithm (BSA)*: iterative; for any number of rounds $n \geq 3$; two main phases: 1) *Initialization* and 2) *Round chaining*.
- Initialization.
  Input: Empty sequence.
  Output: Completely constructed $I_F(1)$, $I_X(1)$ and $I_F(2)$ and partially constructed $I_X(2)$.
- Round chaining (for $j = 2, 3, \ldots, n-1$).
  Input: Completely constructed $I_F(j-1)$ and $I_F(j)$ and partially constructed $I_X(j)$.
  Output: Completely constructed $I_X(j)$ and $I_F(j+1)$, partially constructed $I_X(j+1)$ if $j+1 < n$ and completely constructed $I_X(j+1)$ if $j+1 = n$.

Construction of Best Characteristics: Construction Logic

- Construct $I_F(j+1)$ in accordance with relation (1):

$$I_F(j+1) = (I_F(j-1) \cup I_X(j)) \backslash (I_F(j-1) \cap I_X(j))$$

- Apply "modified Knudsen observation".
- **Proposition 5.** Any multi-round characteristic, based on at most one active S-box per round, has the following properties (for all relevant indices):
  (i) If $S(j) = \phi$ then $S(j-2) \neq \phi$, $S(j-1) \neq \phi$, $S(j+1) \neq \phi$ and $S(j+2) \neq \phi$.
  (ii) If $S(j) = \phi$ then $S(j+1) = S(j-1)$ and $I_F(j+1) = I_F(j-1)$.
  (iii) If $S(j-1) \neq S(j+1)$ then $I_X(j) = I_F(j-1) + I_F(j+1)$.

Construction of Best Characteristics: Case of S-boxes with Embedded Parity Check (1)

- Need to control the presence of parity sum, in our notations

$$I_F(j+1) = I_F(S(j+1))$$

  and interpret it as "structurally equivalent" to zero.

- "Structural" effects of embedding parity check:
  (i) Change of a half of biases $\rightarrow$ change of the set of usable 1-round characteristics; change of the 1-round probability contributions.
  (ii) Possibility to consider only a (proper) half of the non-zero output masks.
  (iii) Growth of the internal possibilities for chaining consecutive 1-round approximations.

Construction of Best Characteristics: Case of S-boxes with Embedded Parity Check (2)

- *Remark 3.* In our investigation, including the search algorithm, we restrict the search for best characteristics only to the first found optimal configurations for $I_X(1)$ and $I_X(n)$ when $I_F(1)$ and $I_F(n)$ are already fixed. There may exist another optimal choices for $I_X(1)$ and $I_X(n)$, which we do not examine, because the active S-boxes remain the same. The reason for doing this is that our goal is mainly to compare the best probabilities and the best characteristics structure.

Best Probabilities

- Probability biases corresponding to best characteristics (part):

$n\downarrow$	0000	0001	0010	0100	1000
3	$+0.781\cdot 2^{-2}$	$+0.632\cdot 2^{-2}$	$+0.632\cdot 2^{-2}$	$+0.5\cdot 2^{-2}$	$+0.632\cdot 2^{-2}$
4	$-0.976\cdot 2^{-4}$	$-0.562\cdot 2^{-4}$	$-0.820\cdot 2^{-5}$	$-0.984\cdot 2^{-5}$	$-0.957\cdot 2^{-5}$
14	$-0.596\cdot 2^{-20}$	$-0.617\cdot 2^{-20}$	$-0.830\cdot 2^{-30}$	$-0.75\cdot 2^{-22}$	$-0.527\cdot 2^{-27}$
15	$+0.596\cdot 2^{-21}$	$+0.926\cdot 2^{-22}$	$+0.562\cdot 2^{-31}$	$+0.656\cdot 2^{-23}$	$-0.527\cdot 2^{-29}$
16	$-0.745\cdot 2^{-23}$	$+0.617\cdot 2^{-23}$	$-0.830\cdot 2^{-34}$	$-0.984\cdot 2^{-25}$	$+0.527\cdot 2^{-31}$
17	$-0.582\cdot 2^{-25}$	$+0.772\cdot 2^{-25}$	$+0.968\cdot 2^{-37}$	$+0.861\cdot 2^{-26}$	$-0.988\cdot 2^{-34}$
18	$-0.931\cdot 2^{-27}$	$-0.579\cdot 2^{-26}$	$-0.562\cdot 2^{-38}$	$-0.656\cdot 2^{-28}$	$+0.791\cdot 2^{-36}$
19	$+0.582\cdot 2^{-27}$	$+0.869\cdot 2^{-28}$	$+0.968\cdot 2^{-41}$	$+0.820\cdot 2^{-30}$	$+0.988\cdot 2^{-38}$
20	$-0.727\cdot 2^{-29}$	$+0.579\cdot 2^{-29}$	$+0.889\cdot 2^{-44}$	$-0.75\cdot 2^{-32}$	$+0.988\cdot 2^{-40}$

- Despite of decreasing the effectiveness of best 1-round approximations, the effectiveness of best multi-round characteristics can grow ($\pi = 0001$ and $n = 14$, 17 and 18) or diminish (the remaining cases), depending on the parity bit position.

Best Characteristics

- Number and type of the best characteristics:

$n \rightarrow$	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0000	1	2	1	2	2	2	2	2	1	2	1	2	2	2	2	2	1	2
0001	1	2	1	22	2	2	11	22	11	2	1	22	2	2	11	22	11	2
0010	1	2	11	2	2	22	11	2	1	2	211	22	11	22	2	2	1	22
0100	11	2	1	2	2	2	2	2	1	2	2	2	2	2	1	2	2	2
1000	1	2	1	2	22	2	2	2	1	2	1	2	22	2	2	2	1	2

  Notations: '1' – for one symmetric characteristic, '2' – for a pair different and symmetric to each other characteristics.

- The number and the structure of the best characteristics strongly depend on the presence and the place of the parity bit. In almost all cases in our experiment the number of best characteristics in the modified cipher is at least equal to that in the original one. However, there exists also an exception in case 0100, $n = 17$.

Comparison of Best Multi-round Approximate Expressions

- Some details concerning best 16-round and 19-round linear approximate expressions of type (2):

	16-round approximations			19-round approximations		
π	effectiveness	appr	S-boxes	effectiveness	appr	S-boxes
0000	$0.596 \cdot 2^{-20}$	$appr_1, appr_2$	2	$0.582 \cdot 2^{-25}$	$appr_{15}, appr_{16}$	10
0001	$0.617 \cdot 2^{-20}$	$appr_3, appr_4$	7	$0.772 \cdot 2^{-25}$	$appr_{17}$	2
		$appr_5, appr_6$	3		$appr_{18}$	2
0010	$0.830 \cdot 2^{-30}$	$appr_7, appr_8$	9	$0.968 \cdot 2^{-37}$	$appr_{19}, appr_{20}$	9
		$appr_9, appr_{10}$	8			
0100	$0.75 \cdot 2^{-22}$	$appr_{11}, appr_{12}$	2	$0.861 \cdot 2^{-26}$	$appr_{21}$	12
1000	$0.527 \cdot 2^{-27}$	$appr_{13}, appr_{14}$	3	$0.988 \cdot 2^{-34}$	$appr_{22}, appr_{23}$	10

- There is no clear rule for dependence of complexity of a potential attack from the presence of parity. When parity is embedded there are cases of lower or higher effectiveness combined with more or less active S-boxes comparing with the original cipher.

Conclusions (1)

- In this work, from the linear cryptanalysis perspective, we have examined the effect of inserting a bit of additional linearity in the round's output of DES by embedding parity check in outputs of its S-boxes. Similar to Matsui (1993,1994), that research is focused on best characteristics obtained on the base of "at most one active S-box per each round", since our primary goal is to compare the results to those for the original cipher.
- We prove that such embedding reduces the effectiveness of optimal 1-round and 3-round characteristics. However, as shown by experiments based on our ad-hoc search algorithm, this is not true for greater number of rounds. So, in general, a modification of this type does not necessarily mean a reduction or growth in the effectiveness of interest.

Conclusions (2)

- Also, the number of yielded best characteristics varies depending on the choice of the parity position. So does the number of active S-boxes in the respective to them linear approximations with highest probability that in turn implies differences in the number of resultant effective key and text bits.
- Therefore, we could conclude that successful attacks based on this approach have varying magnitude of complexity and at the same time they are not inevitably more efficient than the corresponding primary attacks towards the original cipher.

Work in Progress: Two-Round Iterative Characteristics

Matsui's results:

- One-round "zero input mask case" for DES: using $NS_7(3, 15)$ and $NS_8(48, 13)$; bias $= -0.047$, effectiveness $= 0.047$.
- Lower effectiveness of the multi-round characteristics comparing to the case "at most one active S-box per each round".

New results (when parity check is embedded):

- One-round "zero input mask case", using two or more active S-boxes: effectiveness 0.035 (0010), 0.023 $(0001, 0100, 1000)$ – lower, compared to the same case for DES.
- One-round *"parity keeping input mask case"* (new type of characteristics!): upper bound $0,03515625$ for the effectiveness – lower compared to one-round "zero input mask case" for DES.

Work in Progress and Future Work

- Usage of one-round characteristics, based on approximations of more than one S-box.
- Embedding parity check in different positions for the different S-boxes or not in all S-boxes.
- Embedding a different kind of linear dependency in the S-boxes outputs.

- **Acknowledgments.** The authors would like to thank the anonymous reviewers for helpful comments which substantially improved the manuscript.

THANK YOU FOR YOUR ATTENTION !