# Advanced Cryptology

## Linear Cryptanalysis

prof. Ing. Róbert Lórencz, CSc.

České vysoké učení technické v Praze, Fakulta informačních technologií
Katedra počítačových systémů


Příprava studijních programů Informatika pro novou fakultu ČVUT je spolufinancována Evropským sociálním fondem a rozpočtem Hlavního města Prahy v rámci Operačního programu Praha — adaptabilita (OPPA) projektem CZ.2.17/3.1.00/31952 – „Příprava a zavedení nových studijních programů Informatika na ČVUT v Praze“.
Praha & EU: Investujeme do vaší budoucnosti

Tato přednáška byla rovněž podpořena z prostředků projektu č. 347/2013/B1a Fondu rozvoje vysokých škol Ministerstva školství, mládeže a tělovýchovy

# Lecture overview

- Keys
- Cryptanalysis
- Types of an attack
- Linear Cryptanalysis (LC)
- Basic properties

# Keys

## Weak key

Key, which special mathematical properties enables easy cipher cracking.

## Weak key in DES

Lets choose key $k = (0101010101010101)_{16}$. All round keys (sub)keys generated from $k$ are the same. Because DES in an Feistel type algorithm , interaction cancels the subkeys and $E_k(\text{PT})=\text{PT}$!

## Half-weak key in DES

Keys $k_1 \neq k_2$, are half-weak, if::

$$E_{k_1}(E_{k_2}(PT)) = PT \ .$$

# Cryptanalysis

## Cryptanalysis

Science about breaking the ciphers without knowledge of $k$. Block ciphers are most often analyzed using linear and differential cryptanalysis.

## Linear cryptanalysis

Searches for linear dependencies (approximations) to each cipher action.

## Differential cryptanalysis

Searches for dependencies (differences) between inputs and outputs of ciphers.

# Kinds of an attack

- Brute force attack — we try all possible keys, and that is why it is often not applicable. In case of DES it is necessary to try as many as $2^{56} = 64P \approx 7.2 \cdot 10^{16}$ keys!
- Attack with known ciphertext — we know only CT. This attack can succeed, if the $k$ or the PT is predictable. In opposite case it is very complex process.
- Attack with known plaintext — we got samples of CT, and corresponding PT, from which we are trying to find $k$, or another informations about encryption system.
- Attack with knowledge of chosen PT — we got samples CT for arbitrary, and also chosen, PT. Goal of this attack is gaining of informations about weak points in encryption process.
- Attack with knowledge of chosen CT — we got CT and PT without knowledge of $k$, which we are trying to find out.

# Linear cryptanalysis (LC)

## Basic properties

- LC used for cryptanalysis of block ciphers.
- LC uses high probability of occurrence of linear description of PT and CT bits and subkey bits for given round.
- LC looks for dependencies between inputs and outputs of S-boxes.
- LC is an attack with knowledge of PT and corresponding CT, but PT we cannot chose.
- Basic idea is to approximate operations of parts of cipher with expression, which is linear. Operations between bits are bit operations exclusive-OR "$\oplus$"modulo 2. Generally we can express expression in form of:

$$X_1 \oplus X_2 \oplus \ldots \oplus X_u \oplus Y_1 \oplus Y_2 \ldots \oplus Y_v = 0, \qquad (1)$$

# Linear cryptanalysis (LC)

## Basic properties

- where $X_i$ is $i$th bit of an input $X = [X_1, X_2 \ldots X_u]$ and $Y_j$ is $j$th bit of an output $Y = [Y_1, Y_2 \ldots Y_v]$
- Whole equation express the sum of exclusive additions modulo 2 of input and output bits.

## Goal of LC

Find out such expressions, which are in form (1) and has high or low probability of occurrence.

## Example:

- Lets have 2 random bits $a$ and $b$.
- Probability, that $a \oplus b = 0$ is $1/2$.
- Unless it is not true, that they are random, it is possible to find out bias in probability $1/2$.
- LC is using this fact.

# Linear cryptanalysis (LC)

*Linear probability bias (LPB)*

- If we had a probability $p$ that, randomly chosen expression is true (1), then the bias *LPB* is calculated as $p - 1/2$.
- Size is then $|p - 1/2|$.
- With increasing size of *LPB*, the ability to analyze the cipher increases.
- $p = 1$ imply, that linear expression (1) is perfect representation of cipher behavior and that cipher has a catastrophic weaknesses.
- If $p = 0$ then expression (1) represents an affine dependencies in cipher and also indicates catastrophic weaknesses in cipher.
- For modulo 2 operation is affine function simple complement to linear function.
- As linear, also affine approximation of cipher behavior indicate for $p > 1/2$ or $p < 1/2$ easy usage of LC. For both examples we will be using linear approximation - expression (1)

# Linear cryptanalysis (LC)

Question: How to create expression, which is 'strongly linear' and how to use it?

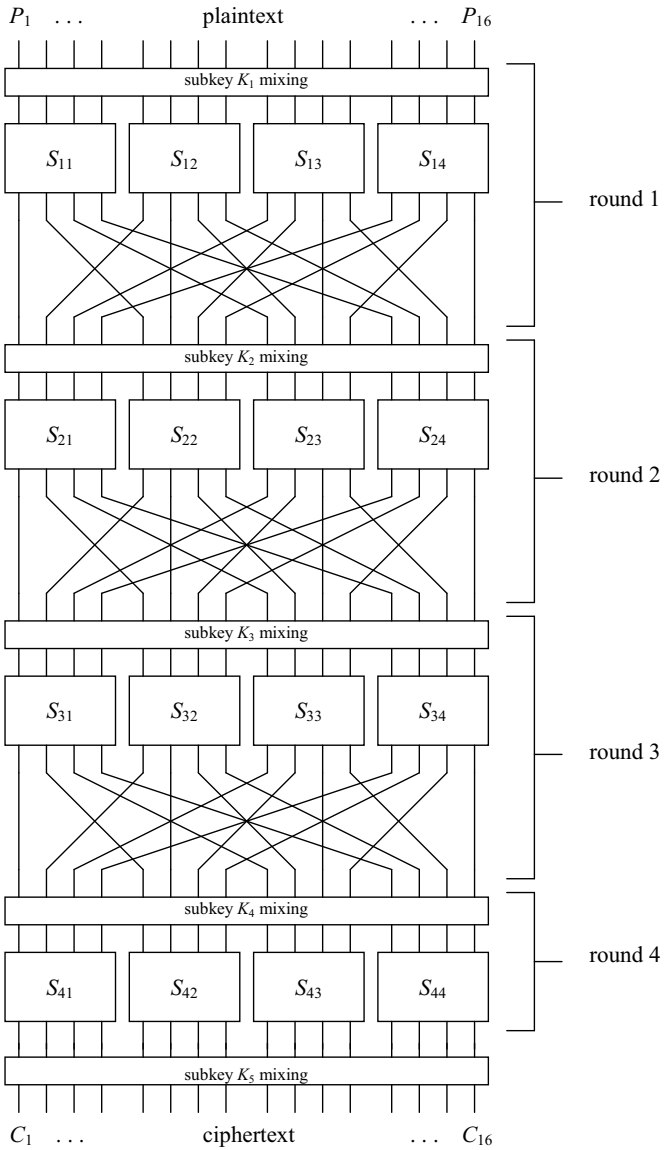
- Lets assume properties of nonlinear cipher parts: S-boxes.
- If linear properties of S-box are 'discoverable', then we can create linear approximations between input and output bits of S-Box
- Further it is possible to chain linear approximations of S-boxes in a way, that the intermediate bits can cancel themselves (bits pervading between S-boxes).
- Then linear expression describing cipher behavior includes only bits of PT and bits of the last round and has a high *LPB*.
- Bits of subkeys of individual round we moved to right side of linear expression with that, that in a sum they can have a value of a "0"or a "1". This causes the change in sign on *LPB*. During searching suitable linear approximation we are interested only in size of a *LPB*, thus its absolute value.

# SPN

## Substitution and permutation net (SPN)

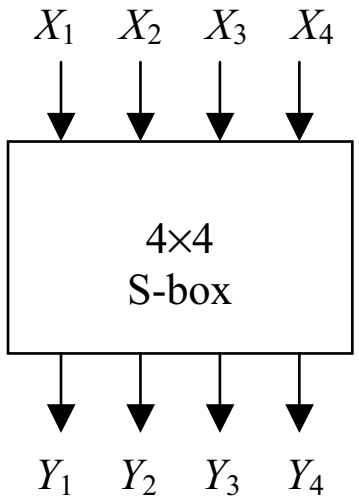
- Lets have a basic Substitution and permutation net on picture (next slide).
- SPN is a cipher containing substitution blocks, transposition nets and operations for generating subkeys.
- SPN has a 16bit word, thus input and output block has a size of a length 16bits.
- SPN is a simple block cipher, on which it is possible to show basic principle of a LC. These principles can be generalized for more complex block ciphers as are: DES, AES etc.
- SPN divides 16bit block into four 4bits subblocks. Each subblock enters a S-box, where is substitution executed of 4bits to 4bits.
- Very important property of S-box is nonlinear mapping of input to output, i.e. outputting bits cannot be represented as some linear operation of input bits.

## Substitution and permutation net (SPN)



## S-box

| Input | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Output | E | 4 | D | 1 | 2 | F | B | 8 | 3 | A | 6 | C | 5 | 9 | 0 | 7 |




## Permutation

| Input | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Output | 1 | 5 | 9 | 13 | 2 | 6 | 10 | 14 | 3 | 7 | 11 | 15 | 4 | 8 | 12 | 16 |

# SPN

## Substitution and permutation network (SPN)

- SPN uses same S-Boxes for each subword and in each round (difference against DES).
- Substitution is derived from S-box of a DES
- Permutation in each round is same. Executed simple transposition of bits or permutation of bit positions. Permutation is given by permutation table.
- Permutation in last round has no justification and thus it is not executed.
- Subkeys are added with operation XOR in each round to pervading 16 bit word and also at the end of a 4th round from reason of securing the last substitution.
- For decryption is used SPN in revers order. That means, that S-box has an inverse substitution and by this ensures the bijective view of an S-box.

## Piling-Up Lemma

- For constructing the relation (1) it is necessary to state some basic principles.
- Let assume 2 random variables $X_1$ and $X_2$.
- Let $X_1 \oplus X_2 = 0$ (equivalent $X_1 = X_2$) is a linear expression.
- $X_1 \oplus X_2 = 1$ (equivalent $X_1 \neq X_2$) is an affine expression.
- Let assume, that probability of separation for $X_1$ is.

$$\Pr(X_1 = 0) = p_1 \quad \text{and} \quad \Pr(X_1 = 1) = 1 - p_1$$

- Further assume, that probability of separation for $X_2$ is.

$$\Pr(X_2 = 0) = p_2 \quad \text{and} \quad \Pr(X_2 = 1) = 1 - p_2$$

## Piling-Up Lemma

- If the $X_1$ and $X_2$ are mutually independent, then
  $\Pr(X_1 = 0, X_2 = 0) = p_1 p_2$,
  $\Pr(X_1 = 0, X_2 = 1) = p_1(1 - p_2)$
  $\Pr(X_1 = 1, X_2 = 0) = (1 - p_1)p_2$,
  $\Pr(X_1 = 1, X_2 = 1) = (1 - p_1)(1 - p_2)$.
- And we can write:

$$\begin{aligned} \Pr(X_1 \oplus X_2 = 0) &= \Pr(X_1 = X_2) \\ &= \Pr(X_1 = 0, X_2 = 0) + \Pr(X_1 = 1, X_2 = 1) \\ &= p_1 p_2 + (1 - p_1)(1 - p_2). \end{aligned}$$

- If we note $p_1 = 1/2 + \varepsilon_1$ and $p_2 = 1/2 + \varepsilon_2$, where $\varepsilon_1$ and $\varepsilon_2$ as a probabilistic bias and is true that $-1/2 \leq \varepsilon_1, \varepsilon_2 \leq 1/2$, we can write: $\Pr(X_1 \oplus X_2 = 0) = 1/2 + 2\varepsilon_1\varepsilon_2$

## Piling-Up Lemma

- and LPB $\varepsilon_{1,2}$ of expression $X_1 \oplus X_2 = 0$ is $\varepsilon_{1,2} = 2\varepsilon_1\varepsilon_2$.
- This conclusion can be extended to more than two random variables For variables from $X_1$ to $X_n$, having probabilities $p_1 = 1/2 + \varepsilon_1$ and $p_2 = 1/2 + \varepsilon_n$ and probabilities of expression $X_1 \oplus \ldots \oplus X_n = 0$ holds so called Piling-Up Lemma.

### Piling-Up Lemma

For $n$ independent and random binary variables $X_1, X_2, \ldots, X_n$ holds

$$\Pr(X_1 \oplus X_2 \oplus \ldots \oplus X_n = 0) = 1/2 + 2^{n-1} \prod_{i=1}^{n} \varepsilon_i$$

or

$$\varepsilon_{1,2,\ldots,n} = 2^{n-1} \prod_{i=1}^{n} \varepsilon_i,$$

where$\varepsilon_{1,2,\ldots,n}$ represents *LPB* expression $X_1 \oplus X_2 \oplus \ldots \oplus X_n = 0$.

## Piling-Up Lemma

- If $p_i = 0$ or $p_i = 1$ for all $i$, then $\Pr(X_1 \oplus \ldots \oplus X_n = 0) = 0$ or 1.
- If only one $p_i = 1/2$, then $\Pr(X_1 \oplus \ldots \oplus X_n = 0) = 1/2$.

## Example:

- In developing the linear approximation of a cipher, the $X_i$ values will actually represent linear approximations of the S-boxes.
- Consider 4 independent random binary variables $X_1, X_2, X_3$ a $X_4$.
- Let $\Pr(X_1 \oplus X_2 = 0) = 1/2 + \varepsilon_{1,2}$ and $\Pr(X_2 \oplus X_3 = 0) = 1/2 + \varepsilon_{2,3}$.
- and consider the sum $X_1 \oplus X_3$ to be derived by adding $X_1 \oplus X_2$ and $X_2 \oplus X_3$ together.
- Hence

$$\Pr(X_1 \oplus X_3 = 0) = \Pr([X_1 \oplus X_2] \oplus [X_2 \oplus X_3] = 0)$$

.

## Piling-Up Lemma

- So we are combining linear expressions to form a new linear expression.
- Since we may consider random variables $X_1 \oplus X_2$ and $X_2 \oplus X_3$ to be independent, we can use Piling-Up Lemma to be determine:

$$\Pr(X_1 \oplus X_3 = 0) = 1/2 + 2\varepsilon_{1,2}\varepsilon_{2,3}$$

- and, consequently, $\varepsilon_{1,3} = 2\varepsilon_{1,2}\varepsilon_{2,3}$.
- As we shall see, the expressions $X_1 \oplus X_2 = 0$ and $X_2 \oplus X_3 = 0$ are analogous to linear approximations of S-boxes and $X_1 \oplus X_3 = 0$ is analogous to a cipher approximation where the intermediate bit $X_2$ is eliminated.
- The real analysis will be more complex involving many S-box approximations.

## Analyzing the Cipher Components

- Consider the S-box representation with input $X = [X_1, X_2, X_3, X_4]$ and a corresponding output $Y = [Y_1, Y_2, Y_3, Y_4]$.
- All linear approximations can be examined to determine their usefulness by computing the probability bias for each. Hence, we are examining all expressions of the form of equation (1) where $X$ and $Y$ are the S-box input and outputs, respectively.

## Example:

- for the S-box used in our cipher, consider the linear expression $X_2 \oplus X_3 \oplus Y_1 \oplus Y_3 \oplus Y_4 = 0$ or equivalently $X_2 \oplus X_3 = Y_1 \oplus Y_3 \oplus Y_4$.
- Applying all 16 possible input values for $X$ and examining the corresponding output values $Y$, it may be observed that for exactly 12 out the 16 cases, the expression above holds true.
- Hence, the probability bias is $12/16 - 1/2 = 1/4$ (see Table).

## Linear Approximations of S-box

| $X_1$ | $X_2$ | $X_3$ | $X_4$ | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ | $X_2 \oplus X_3$ | $Y_1 \oplus Y_3 \oplus Y_4$ | $X_1 \oplus X_4$ | $Y_2$ | $X_3 \oplus X_4$ | $Y_1 \oplus Y_4$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |
| 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 |
| 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 |
| 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 |

## Analysis cipher components

- Similarly, for equation $X_1 \oplus X_4 = Y_2$ is $LPB = 0$, see Table: $8/16 - 1/2 = 0$.
- For equation $X_3 \oplus X_4 = Y_1 \oplus Y_4$ is $LPB = 2/16 - 1/2 = -3/8$. In the this case, the best approximation is an affine approximation as indicated by the minus sign.
- However, the success of the attack is based on magnitude of the bias and, as we shall see, affine approximations can be used equivalently to linear approximations.
- A complete enumeration of all linear approximations of the S-box in our cipher is given in the linear approximation table (next Table).
- Each element in the table represents the number of matches between the linear equation represented in hexadecimal as "Input Sum"and the sum of the output bits represented in hexadecimal as "Output Sum"$-8$.

## Linear Approximation Table

| Input Sum | Output Sum 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 | 0 | −2 | −2 | 0 | 0 | −2 | +6 | +2 | +2 | 0 | 0 | +2 | +2 | 0 | 0 |
| 2 | 0 | 0 | −2 | −2 | 0 | 0 | −2 | −2 | 0 | 0 | +2 | +2 | 0 | 0 | −6 | +2 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | +2 | −6 | −2 | −2 | +2 | +2 | −2 | −2 |
| 4 | 0 | +2 | 0 | −2 | −2 | −4 | −2 | 0 | 0 | −2 | 0 | +2 | +2 | −4 | +2 | 0 |
| 5 | 0 | −2 | −2 | 0 | −2 | 0 | +4 | +2 | −2 | 0 | −4 | +2 | 0 | −2 | −2 | 0 |
| 6 | 0 | +2 | −2 | +4 | +2 | 0 | 0 | +2 | 0 | −2 | +2 | +4 | −2 | 0 | 0 | −2 |
| 7 | 0 | −2 | 0 | +2 | +2 | −4 | +2 | 0 | −2 | 0 | +2 | 0 | +4 | +2 | 0 | +2 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | −2 | +2 | +2 | −2 | +2 | −2 | −2 | −6 |
| 9 | 0 | 0 | −2 | −2 | 0 | 0 | −2 | −2 | −4 | 0 | −2 | +2 | 0 | +4 | +2 | −2 |
| A | 0 | +4 | −2 | +2 | −4 | 0 | +2 | −2 | +2 | +2 | 0 | 0 | +2 | +2 | 0 | 0 |
| B | 0 | +4 | 0 | −4 | +4 | 0 | +4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C | 0 | −2 | +4 | −2 | −2 | 0 | +2 | 0 | +2 | 0 | +2 | +4 | 0 | +2 | 0 | −2 |
| D | 0 | +2 | +2 | 0 | −2 | +4 | 0 | +2 | −4 | −2 | +2 | 0 | +2 | 0 | 0 | +2 |
| E | 0 | +2 | +2 | 0 | −2 | −4 | 0 | +2 | −2 | 0 | 0 | −2 | −4 | +2 | −2 | 0 |
| F | 0 | −2 | −4 | −2 | −2 | 0 | +2 | 0 | 0 | −2 | +4 | −2 | −2 | 0 | +2 | 0 |

## Analysis cipher components

- Dividing an element value by 16 gives the probability bias for the particular linear combination of input and output bits.
- The hexadecimal value representing a sum, when viewed as a binary value indicates the variables involved in the sum.
- For a linear combination of input variables represented as $a_1 \cdot X_1 \oplus a_2 \cdot X_2 \oplus a_3 \cdot X_3 \oplus a_4 \cdot X_4$, where $a_i \in 0, 1$ and "$\cdot$"represents binary AND, the hexadecimal value represents the binary value $a_1 a_2 a_3 a_4$ ($a_1$ is MSB).
- Similarly, for a linear combination of output bits $b_1 \cdot Y_1 \oplus b_2 \cdot Y_2 \oplus b_3 \cdot Y_3 \oplus b_4 \cdot Y_4$, where $b_i \in 0, 1$, the hexadecimal value represents the binary vector $b_1 b_2 b_3 b_4$.
- *LPB* of linear equation $X_3 \oplus X_4 = Y_1 \oplus Y_4$ (hex input 3 and hex output 9) is $-6/16 = -3/8$ and the probability that the linear equation holds true is given by $1/2 - 3/8 = 1/8$.

## Basic properties of the linear approximation

- The first column is all zeros except for the topmost value.
- The top row of the table is all zeros, except for the leftmost value.
- the sum of any row or any column must be either +8 or -8.

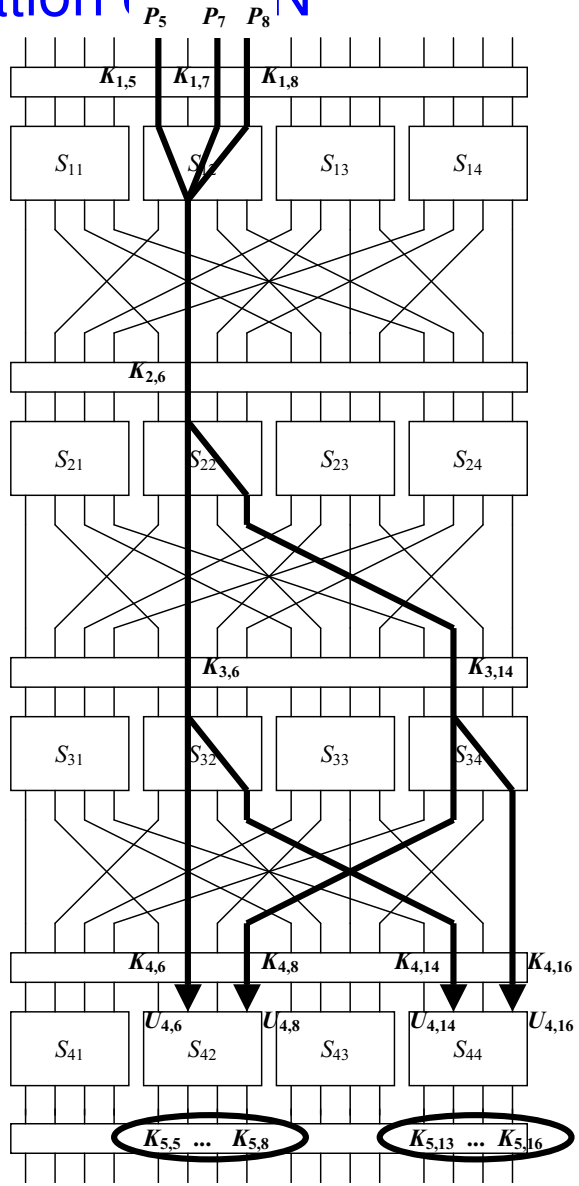
## Constructing Linear Approximations for SPN

- Once the linear approximation information has been compiled for the S-boxes in an SPN, we have the data to proceed with determining linear approximations of the overall cipher of the form of equation (1).
- This can be achieved by concatenating appropriate linear approximations of S-boxes.
- By constructing a linear approximation involving plaintext bits and data bits from the output of the second last round of S-boxes, it is possible to attack the cipher by recovering a subset of the subkey bits that follow the last round.
- We illustrate with an example.

## Example of linear approximation of SPN

- Consider an approximation involving $S_{1,2}, S_{2,2}, S_{3,2}, S_{3,4}$ as illustrated in following Figure.
- Note that this actually develops an expression for the first 3 rounds of the cipher and not the full 4 rounds.
- We shall see how this is useful in deriving the subkey bits after the last round in the next slides.
- We use the following approximations of the S-box:
  - $S_{12} : X_1 \oplus X_3 \oplus X_4 = Y_2$ with probability $12/16$ and $LPB = +1/4$
  - $S_{22} : X_2 = Y_2 \oplus Y_4$ with probability $4/16$ and $LPB = -1/4$
  - $S_{32} : X_2 = Y_2 \oplus Y_4$ with probability $4/16$ and $LPB = -1/4$
  - $S_{34} : X_2 = Y_2 \oplus Y_4$ with probability $4/16$ and $LPB = -1/4$
- $P = [P_1, P_2, \ldots, P16]$ are 16 bits of PT. $U_i(V_i)$ represent the 16-bit block of bits at the input (output) of the round $i$ S-boxes and $U_{i,j}$ ($V_{i,j}$) represent the $j$-th bit block $U_i(V_i)$ (where bits are numbered from 1 to 16 from left to right in the figure of the cipher).

## Sample Linear Approximation of SPN



## Sample Linear Approximation of SPN

- Let $K_i$ represent the subkey block of bits exclusive-ORed at the input to round $i$, with the exception that K5 is the key exclusive-ORed at the output of round 4.
- $U_1 = P \oplus K_1$. Using the linear approximation of the 1st round, we then have:

$$V_{1,6} = U_{1,5} \oplus U_{1,7} \oplus U_{1,8} = (P_5 \oplus K_{1,5}) \oplus (P_7 \oplus K_{1,7}) \oplus (P_8 \oplus K_{1,8}) \quad (2)$$

  with probability $3/4$.
- For the approximation in the 2nd round, we have:

$$V_{2,6} \oplus V_{2,8} = U_{2,6} = V_{1,6} \oplus K_{2,6}$$

  with probability $1/4$.

# LK II

- Substituting for $V_{1,6}$ from equation (2) with probability $3/4$ obtain the equation:

$$V_{2,6} \oplus V_{2,8} \oplus P_5 \oplus P_7 \oplus P_8 \oplus K_{1,5} \oplus K_{1,7} \oplus K_{1,8} \oplus K_{2,6} = 0, \quad (3)$$

  for which holds probability $1/2 + 2(3/4 - 1/2)(1/4 - 1/2) = 3/8$, follows from Piling Up Lemma. $LPB = -1/8$.
- Note that we are using the assumption that the approximations of S-boxes are independent which, although not strictly correct, works well in practice for most ciphers.
- For round 3, we note that:

$$V_{3,6} \oplus V_{3,8} = U_{3,6} \text{ a } V_{3,14} \oplus V_{3,16} = U_{3,14}$$

  both with probability $1/4$.

# LK III

- For $U_{3,6} = V_{2,6} \oplus K_{3,6}$ a $U_{3,14} = V_{2,8} \oplus K_{3,14}$ from the previous two equations we get:

$$V_{3,6} \oplus V_{3,8} \oplus V_{3,14} \oplus V_{3,16} \oplus V_{2,6} \oplus K_{3,6} \oplus V_{2,8} \oplus K_{3,14} = 0 \quad (4)$$

  with probability $1/2 + +2(1/4 - 1/2)^2 = 5/8$ and $LPB = 1/8$
- Combining (3) and (4) we get:

$$V_{3,6} \oplus V_{3,8} \oplus V_{3,14} \oplus V_{3,16} \oplus P_5 \oplus P_7 \oplus P_8 \oplus K_{1,5} \oplus$$

$$\oplus K_{1,7} \oplus K_{1,8} \oplus K_{2,6} \oplus K_{3,6} \oplus K_{3,14} = 0$$

- It holds: $U_{4,6} = V_{3,6} \oplus K_{4,6}$, $U_{4,8} = V_{3,14} \oplus K_{4,8}$, $U_{4,14} = V_{3,8} \oplus K_{4,14}$ a $U_{3,16} = V_{3,6} \oplus K_{4,16}$.

# LK IV

- Based on previous, we can write:

$$U_{4,6} \oplus U_{4,8} \oplus U_{4,14} \oplus U_{4,16} \oplus P_5 \oplus P_7 \oplus P_8 \oplus \sum_K = 0 \quad \text{,kde}$$

$$\sum_K = K_{1,5} \oplus K_{1,7} \oplus K_{1,8} \oplus K_{2,6} \oplus K_{3,6} \oplus K_{3,14} \oplus K_{4,6} \oplus K_{4,8} \oplus K_{4,14} \oplus K_{4,16}$$

- $\sum_K$ has a fixed value and either 0 or 1 depending on the cipher key.
- Using Piling-Up Lemma for previous expression we get the probability: $1/2 + 2^3(3/4 - 1/2)(1/4 - 1/2)^3 = 15/32$ a $LPB = -1/32$.
- Then $\sum_K$ is fixed, then:

$$U_{4,6} \oplus U_{4,8} \oplus U_{4,14} \oplus U_{4,6} \oplus P_5 \oplus P_7 \oplus P_8 = 0 \qquad (5)$$

with probability $15/32$ for $\sum_K = 0$ or $(1 - 15/32) = 17/32$ for $\sum_K = 1$.

- Now we have a linear approximation of the 3 rounds of the SPN with $LPB = 1/32$.

# LC - Extracting Key Bits I

- Once an R-1 round linear approximation is discovered for a cipher of R rounds with a suitably large enough linear probability bias, it is conceivable to attack the cipher by recovering bits of the last subkey.
- In the case of our example cipher, it is possible to extract bits from subkey $K_5$ given a 3 round linear approximation. The linear expression (5) contains inputs to S-boxes $S_{42}$ and $S_{44}$ in the last round. For each pair of PT and its competent CT we will test all 256 values of the selected subkey $[K_{5,5} \ldots K_{5,8}, K_{5.13} \ldots K_{5.16}]$.
- For each partial subkey value, which is backward substituted by S-boxes from $S_{42}$ and $S_{44}$ to values $[U_{4,5} \ldots U_{4,8}, U_{4,13} \ldots U_{4,16}]$ putt to (5) with the values of bits corresponding PT $P_5, P_7$ and $P_8$ is evaluated the linear expression.
- In case that expression (5) is true for a given case is incremented appropriate counter value of the subkey.

# LC - Extracting Key Bits II

- The counter value will be the largest in absolute value minus the number of samples of PT/CT, for the expected value of the subkey.
- The deviation will be positive or negative depending on the values of the subkey bits included in the $\sum_K$. If $\sum_K = 0$, the probability of a linear approximation (5) is less than $1/2$, and if $\sum_K = 1$ will probably be more than $1/2$.
- The following table presents part of the results of experiments performed with 10,000 pairs of PT/CT (complete table has 256 entries). Hledaná část podklíče $[K_{5,5} \ldots K_{5,8}, K_{5,13} \ldots K_{5,16}] = [0010\ 0100]$.
- Calculation $|bias| = |LPB|$:

$$|bias| = |count - 5000|/10000$$

- As the table shows, the largest $|bias|$ is the value of the parts of the subkey $[2, 4]$.

## Experimental Results for Linear Attack

| *partial subkey* $[K_{5,5}...K_{5,8}, K_{5,13}...K_{5,16}]$ | \| bias \| | *partial subkey* $[K_{5,5}...K_{5,8}, K_{5,13}...K_{5,16}]$ | \| bias \| |
|---|---|---|---|
| 1 C | 0.0031 | 2 A | 0.0044 |
| 1 D | 0.0078 | 2 B | 0.0186 |
| 1 E | 0.0071 | 2 C | 0.0094 |
| 1 F | 0.0170 | 2 D | 0.0053 |
| 2 0 | 0.0025 | 2 E | 0.0062 |
| 2 1 | 0.0220 | 2 F | 0.0133 |
| 2 2 | 0.0211 | 3 0 | 0.0027 |
| 2 3 | 0.0064 | 3 1 | 0.0050 |
| **2 4** | **0.0336** | 3 2 | 0.0075 |
| 2 5 | 0.0106 | 3 3 | 0.0162 |
| 2 6 | 0.0096 | 3 4 | 0.0218 |
| 2 7 | 0.0074 | 3 5 | 0.0052 |
| 2 8 | 0.0224 | 3 6 | 0.0056 |
| 2 9 | 0.0054 | 3 7 | 0.0048 |

# LC - Evaluation

- The experimentally determined $|bias|$ has value 0.0336 and is very close to the theoretical vypočítané hodnotě $1/32 = 0,03125$.
- Deviations of the experimental and theoretical values ??is also caused by a smaller number of tested pairs PT/CT, and incomplete mutual independence of the samples.