Differential Cryptanalysis

Vincent Rijmen

WCAP 2005

Institute for Applied Information Processing and Communications (IAIK) - Krypto Group

Faculty of Computer Science
Graz University of Technology

Thanks to Stefan Mangard

Outline

- Basic
  - Differential Cryptanalysis of 3 rounds of DES
  - Differential Cryptanalysis of 6 rounds of DES
- Advanced
  - Probability
  - Differentials and characteristics
  - Markov ciphers
  - Decorrelation theory
  - Key-alternating ciphers
- Applied
  - 2 rounds of AES

DC – the setting

- First published by Biham and Shamir in 1991
- Goal: Determine the secret key
- DC is a chosen plaintext attack
- "Exploits mapping properties of differences within DES"

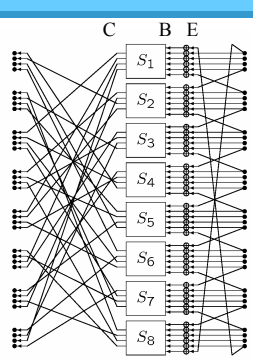
Notation

- B_j, B_j^* … Inputs to Sbox j
- C_j, C_j^* … Outputs of Sbox j
- The differences are denoted by B_j' and C_j'

$$B_j' = B_j \oplus B_j^*$$

$$C_j' = C_j \oplus C_j^*$$

The DES "F" function



The set of possible inputs to an Sbox

- Important observation: Not every input difference can be mapped to every output difference
- The set of possible inputs for given input- and output differences for Sbox j:

$$IN(B_j', C_j') = \{B_j \in (Z_2)^6 : S(B_j) \oplus S_j(B_j \oplus B_j') = C_j'\}$$

The set of possible inputs to an Sbox

S_1

14	4	13	1	2	15	11	8	3	10	6	12	5	9	0	7
0	15	7	4	14	2	13	1	10	6	12	11	9	5	3	8
4	1	14	8	13	6	2	11	15	12	9	7	3	10	5	0
15	12	8	2	4	9	1	7	5	11	3	14	10	0	6	13

39 19

Example:
IN(110100,0100) = {010011 (19),100111 (39)}

$$IN(B_j^{'},C_j^{'}) = \{B_j \in (Z_2)^6 : S(B_j) \oplus S_j(B_j \oplus B_j^{'}) = C_j^{'}\}$$

The key XOR

$$B^{'} = B \oplus B^{*} = (E \oplus K) \oplus (E^{*} \oplus K) = E \oplus E^{*} = E^{'}$$

- The input difference of the Sboxes of a round does not depend on the round key
- Important observation:

$$E_j \oplus K_j \in IN_j(E_j^{'},C_j^{'})$$

The set of all keys that are possible

- The set of possible input values

$$IN(B_j^{'},C_j^{'}) = \{B_j \in (Z_2)^6 : S(B_j) \oplus S_j(B_j \oplus B_j^{'}) = C_j^{'}\}$$

- The set of possible keys:

$$test_j(E_j,E_j^{*},C_j^{'}) = \{E_j \oplus B_j : B_j \in IN_j(E_j^{'},C_j^{'})\}$$

$$K_j = B_j \oplus E_j$$

Summary

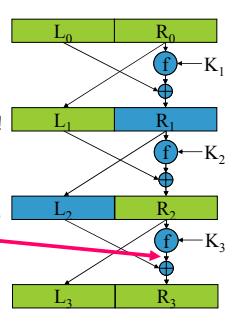
$$test_j(E_j,E_j^{*},C_j^{'}) = \{E_j \oplus B_j : B_j \in IN_j(E_j^{'},C_j^{'})\}$$

- Given E, E* und C', we can start to narrow down the key space

Attack on 3 rounds of DES

E and E* of the last round are known!

The important question is: What is C´?



Attack on 3 rounds of DES

$$R_3 = L_0 \oplus f(R_0,K_1) \oplus f(R_2,K_3)$$

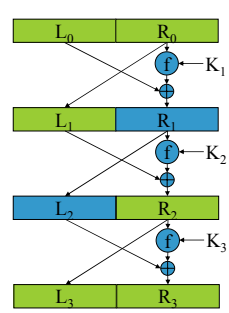
The differences can be expressed as follows, if $R_0^{'} = 0$:

$$R_3^{'} = L_0^{'} \oplus f(R_2,K_3) \oplus f(R_2^{*},K_3)$$

C'

The set of possible values for K_3 can be determined:

$$test_j(E_j,E_j^{*},C_j^{'})$$

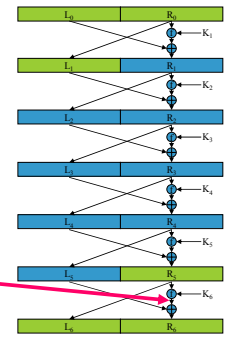


Attack on 3 rounds of DES

- Last round key can be determined by using several pairs of plaintexts with $R'_0 = 0$
- The remaining 8 (= 56 – 48) key bits can be determined easily by a brute-force search

Attack on 6 rounds of DES

- C' cannot be calculated as easily as before
- Hence, a probabilistic approach is pursued

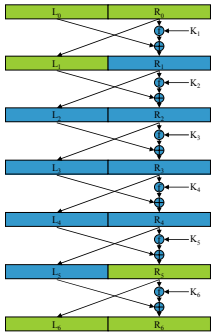


Attack on 6 rounds of DES

- Definition of a characteristic:

  L'_0, R'_0
  L'_1, R'_1, p_1
  L'_2, R'_2, p_2
  …
  L'_n, R'_n, p_n

- p_i is the probability that L'_{i-1}, R'_{i-1} is mapped to L'_i, R'_i

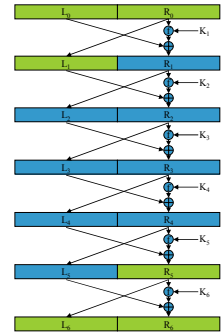


Attack on 6 rounds of DES

- 1-round characteristic:

  L'_0 = anything $\quad R'_0 = 00000000$
  $L'_1 = 0000000 \quad R'_1 = L'_0$

  $p_1 = 1$

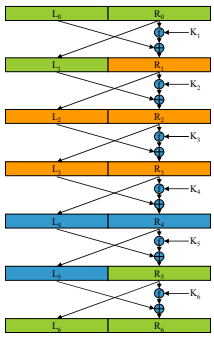


Attack on 6 rounds of DES

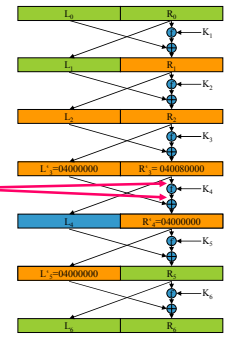
- 3-round characteristic:

  $L'_0 = 4008000 \quad R'_0 = 04000000$
  $L'_1 = 0400000 \quad R'_1 = 00000000$
  $L'_2 = 0000000 \quad R'_2 = 04000000$
  $L'_3 = 0400000 \quad R'_3 = 04008000$

  $p_1 = 0.25$
  $p_2 = 1$
  $p_3 = 0.25$



Attack on 6 rounds of DES



$$C'_1C'_2C'_3C'_4C'_5C'_6C'_7C'_8 = P^{-1}(R'_6 \oplus 04000000)$$

The keys J_2, J_5, J_6, J_7 and J_8 can be determined!

Attack on 6 rounds of DES

$$C_1'C_2'C_3'C_4'C_5'C_6'C_7'C_8' = P^{-1}(R_6' \oplus 04000000)$$

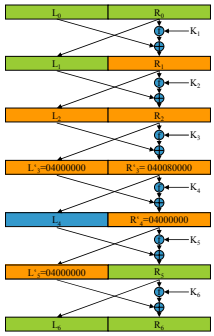
$E_{1\text{-}8}$ and $E'_{1\text{-}8}$ are known from the ciphertext

$$test_j(E_j, E_j^*, C_j') = \{E_j \oplus B_j : B_j \in IN_j(E_j', C_j')\}$$

Apply test function for j=2,5,6,7,8

Only works for 1/16 of the plaintext pairs!

Empty sets of $test_j$ filter 2/3 of the wrong plaintext pairs

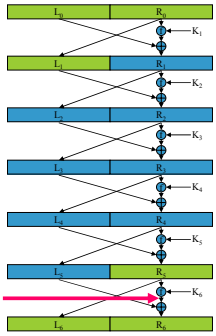


Determining the secret key

- We get wrong key suggestions, because not all bad pairs are filtered
- Tabulating all suggested key strings $J_2J_5J_6J_7J_8$ would require a lot of memory
- Use several pairs of plaintexts to reduce the set of possible values to one element for J_2,J_5,J_6,J_7, and J_8 before combining the parts to a key string

Summary of the attack

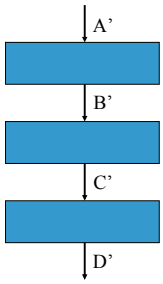
- It is necessary to determine the output differences of the Sboxes in the last round
- A "good" characteristic needs to be found in order to get there



Security against differential attacks

- Make prediction of differences difficult
- Ensure that there are no good characteristics
  - Compute bounds for existing ciphers
  - Design ciphers with low bounds on the probability
  - Design ciphers with easily computable bounds

Predicting a difference



p_1

p_2

p_3

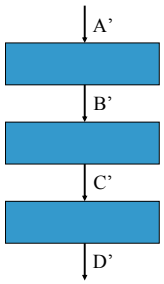
$Pr(A' \rightarrow B') = p_1p_2p_3$???

Technical problems

Computing the probability

1. Characteristics and differentials
2. Independence of rounds
3. Key-averaging

Characteristics and differentials



$Pr(A' \to B') =$

$Pr(A' \to B' \to C' \to D')$

$+ Pr(A' \to B_1' \to C_1' \to D')$

$+ \ldots$

$= \Sigma_{B'} \Sigma_{C'} Pr(A' \to B' \to C' \to D')$

(A',D'): differential

(A',B',C',D'): characteristic (trail)

Characteristic and differential probabilities

- Pr(A',D') ≥ Pr(A',B',C',D')
- Computing Pr(A',D') is much more difficult than computing Pr(A',B',C',D') (folk lore)
  - Both are difficult
- In a `weak' cipher, one characteristic dominates the probability: Pr(A',D') ≈ Pr(A',B',C',D')
  - In `strong' ciphers: open problem

Independence of rounds

- Computing Pr(A',B',C',D')
- Pr(A',B') * Pr(B',C') * Pr(C',D') ??
- Pr(A',B') * Pr(B',C' | A') * Pr(C',D' | A', B')
- Theory of Markov ciphers [Lai,Massey,Murphy]
  - First deal with the key-averaging problem

Key averaging

- What `probability' do we want?
  - Over all possible plaintext pairs with the given input difference
  - And what about the key?
    1. For the 1 key we are attacking
    2. Over all keys (→ averaging)
    3. Over all possible independent round keys (→ averaging): known as Expected Differential Probability (EDP)
       - Note that EDP is not the expected value of (1)
       - EDP of a characteristic is easy to compute in Markov ciphers

Markov cipher

- Definition: cipher such that over one round:

  EDP(A',B' | X) = EDP(A',B')
- With:
  - X: input value
- Fundamental Theorem:

  EDP(A',B',C',D') = EDP(A',B') * EDP(B',C') * EDP(C',D')

Hypothesis of stochastic equivalence

- [Lai,Massey,Murphy]
- For almost all keys k,

  Pr[k](A',B',C',D') ≈ EDP(A',B',C',D')
- We now denote Pr[k] with DP (differential probability)

Problems with the hypothesis of s.e.

- For the flagship cipher IDEA, the hypothesis is wrong [Dae94]
- In the limit, it never holds
  - Example: 8-round AES:
    - all characteristics have EDP < 2^{-300}
    - 2^{127} pairs with given input difference → DP is multiple of 2^{-127}
- Computing EDP of a differential remains a problem

Exhaustive approach

- Keliher, Meijer, Tavares; Hong, Lee, Lim, Sung, Cheon, Park
- Compute/bound EDP of a differential by going through all characteristics
  - Exhaustive analysis
  - Some structure based speedups for up to 4 rounds
  - Some programming speedups

Provable security in the sense of Knudsen/Nyberg

- Developed for Feistel ciphers by Knudsen, Nyberg
- Extended by Matsui and others
- Prove upper bounds on the EDP of a differential through the cipher (minus ε rounds)
- Examples: Misty, KASUMI

Decorrelation theory

- Developed by Vaudenay
- Borrows techniques from universal hash function design
- Example: $F(X,K) = K_1 \times X + K_2$
  - $DP(A' \rightarrow B') = 1$ if $B' = A' \times K_1$
  - $= 0$ otherwise
  - $EDP(A' \rightarrow B') = (\#\text{ of keys } K_1)^{-1}$
- Hypothesis of stochastic equivalence ?

Decorrelation theory relevance

- AES submission DFC was a decorrelated cipher `pur sang'
  - Criticised by Knudsen & Rijmen
- Decorrelation new style: decorrelation modules, to plug in into other ciphers

Daemen-Rijmen approach

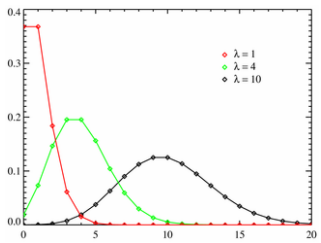
- ePrint report 2005/212 (submitted to the Journal of Cryptology)
- Stronger restrictions than Markov ciphers
- Look at distribution (mean, variance) of DP

Poisson distribution

$Ps(i;\lambda) = \lambda^i e^{-\lambda} / i!$

Mean: λ

Variance: λ

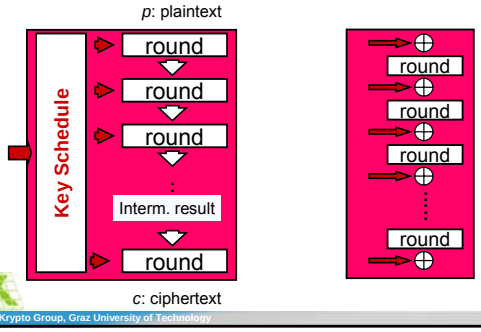


Poisson distribution

- Arises as limit for the binomial distribution B(n,p)
  - When n goes to infinity,
  - While np (=λ) stays constant
- Very good approximation for B(n,p) if np^2 is small
- Discrete distribution for events with low probability
  - For instance: pairs following a characteristic

Key-alternating cipher

Iterative cipher | Key-alternating cipher



Theorem 1

- In a key-alternating cipher, the DP of a characteristic (A',B',C',D') is a stochastic variable with:

  $Pr(DP(A',B',C',D') = i*2^{1-n}) = Ps(i;EDP(A',B',C',D')*2^{n-1})$

  where probability is over all keys and over all key schedules
- Proof sketch:
  - With uniform, independent round keys:
    number of pairs following Q = EDP(Q)*total number of pairs
  - Key schedule defines sampling in the population of pairs
  - Sampling in a large set → approximated by Poisson distribution

Theorem 2

- In a key-alternating cipher, the DP of a differential (A',D') is a stochastic variable with:

  $Pr(DP(A',D') = i*2^{1-n}) = Ps(i;EDP(A',D')*2^{n-1})$

  where probability is over all keys and over all key schedules
- Proof sketch:
  - Sum of Poisson distributions is again a Poisson distribution

Advantages of the approach (for key-alternating ciphers)

- Hypothesis of s.e. replaced by statistical distribution
- No problem if EDP smaller than the smallest possible DP
- Remaining problem: computing/bounding EDP of a differential
  - Proposal: do this also by studying statistical distributions
  - Some claims in the paper

Conclusions

- Differential cryptanalysis
  - Practice is easy
  - Theory is difficult
- For some simple cipher structures, a kind of theory is emerging
  - Theory is good!