Combined Attacks — from Boomerangs to Sandwiches and Differential-Linear

Orr Dunkelman

Department of Computer Science, University of Haifa

June 5th, 2014

Outline

Differential Cryptanalysis

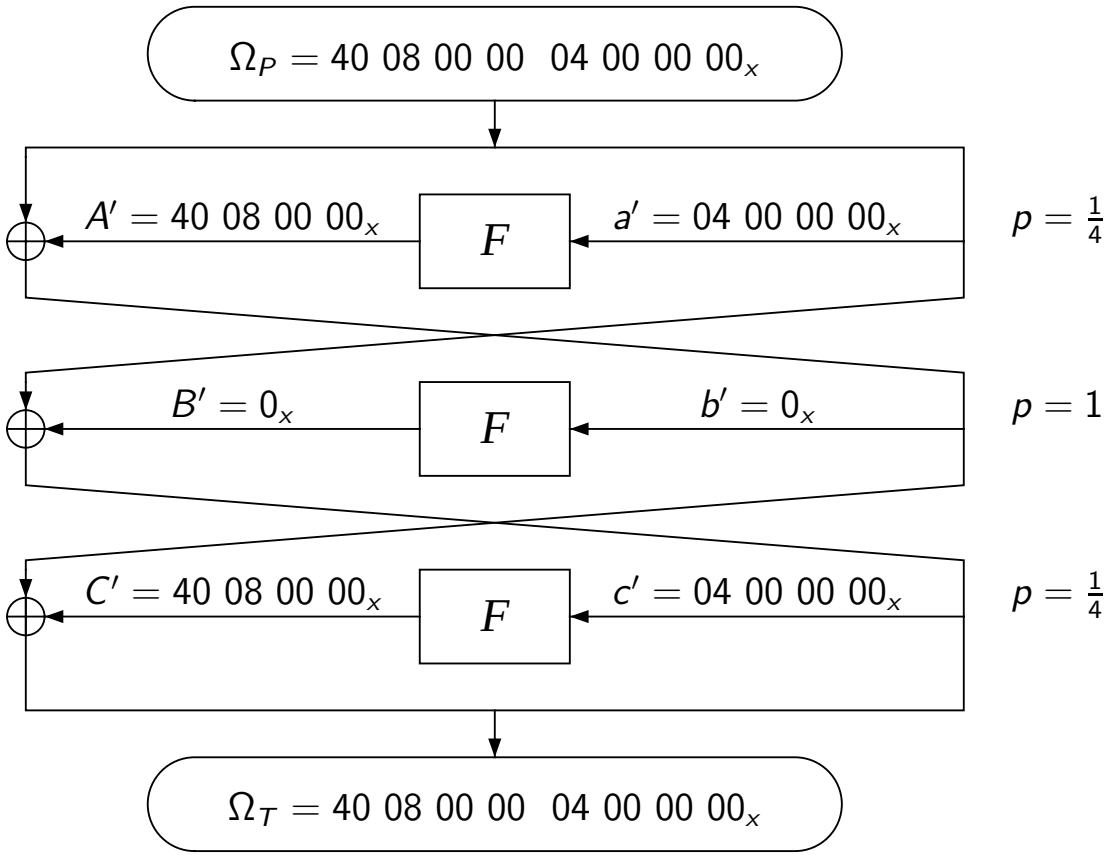
- Considers the development of differences through the encryption process.
- The core of the attack: a differential characteristic (a prediction of the development of differences through the encryption process).
- Given a differential characteristic with probability p, the adversary asks for $O(1/p)$ pairs of plaintexts $(P, P^* = P \oplus \Omega_P)$.
- The attack tries to locate "right pairs", i.e., a pair whose corresponding ciphertexts satisfy $C^* = C \oplus \Omega_C$.
- Information about the key can be learnt from the right pair.

Differential Cryptanalysis (cont.)

- To attack more rounds of the cipher than in the differential characteristic:
  - Guess subkey material in the additional rounds,
  - Partially encrypt/decrypt the plaintext/ciphertext pairs,
  - Count how many "right pairs" exist,
  - The counter for the right subkey is expected to be the highest.
- In such attacks, we care less about "which pair is a right pair", and more about how many such pairs exist.
- Hence, for this sort of attacks, we are only interested in the input and output differences.
- This set of (Ω_P, Ω_C) and the associated probability is called a differential. Its probability is the sum of the probabilities of all differential characteristics that share Ω_P and Ω_C.

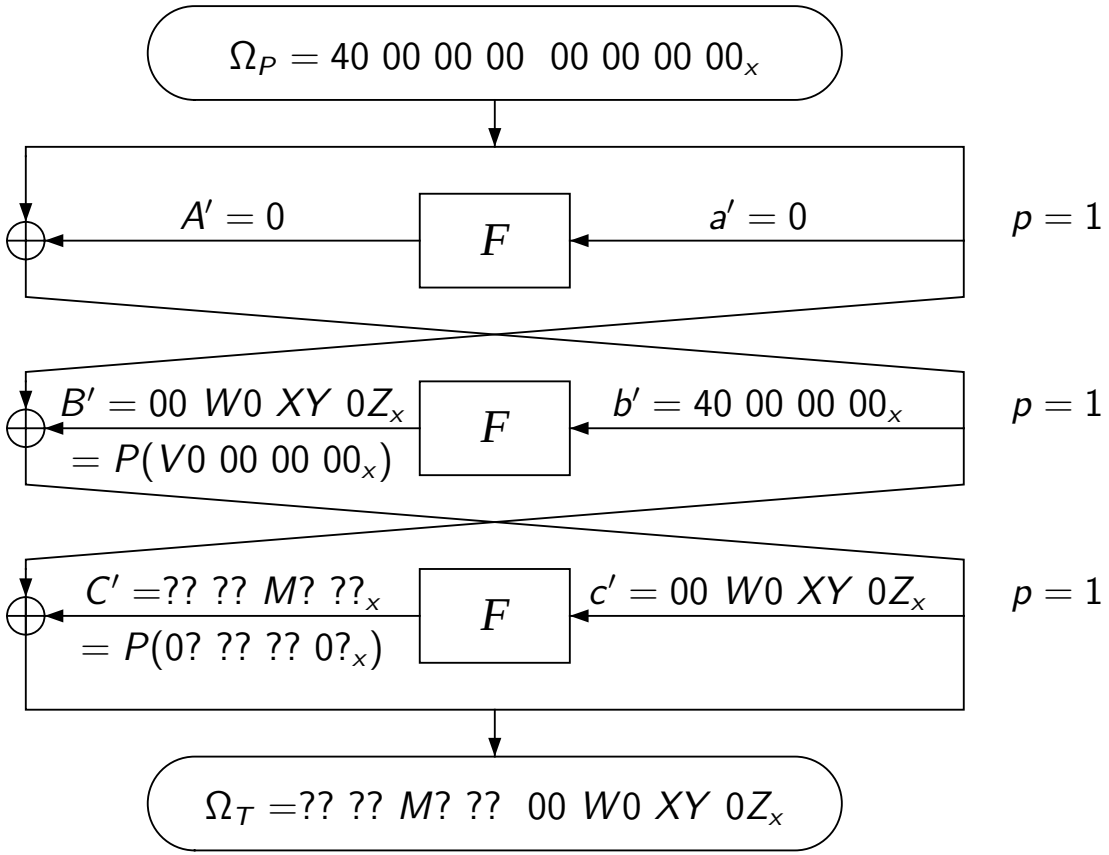
Differential Characteristic of DES

A three-round differential characteristic of DES with probability 1/16:



Differential Characteristic of DES (cont.)

A 3-round truncated differential characteristic of DES:



Linear Cryptanalysis

- Tries to approximate the cipher (or a reduced-round variant of it) as a linear equation:

$$\lambda_P \cdot P \oplus \lambda_C \cdot C = \lambda_K \cdot K$$

  with probability $1/2 + \epsilon$.
- Collect $N = O(\epsilon^{-2})$ known plaintext/ciphertext pairs. The majority are expected to satisfy $\lambda_P \cdot P \oplus \lambda_C \cdot C = \lambda_K \cdot K$ (when $\epsilon > 0$).
- To attack more rounds than in the linear approximation:
  - Guess subkey material in the additional rounds,
  - Partially encrypt/decrypt the plaintext/ciphertext pairs,
  - Count how many times $\lambda_P \cdot P \oplus \lambda_C \cdot C = 0$,
  - The counter for the right subkey is expected to be more biased.

Linear Cryptanalysis (cont.)

- The attack is actually a random process.
- Consider the following scenario:
  - There are 2^s possible subkeys.
  - We want the right subkey to be among the 2^a most biased ones.
- Let $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$.
- A linear attack with $N = c/\epsilon^2$ known plaintexts has a success probability of

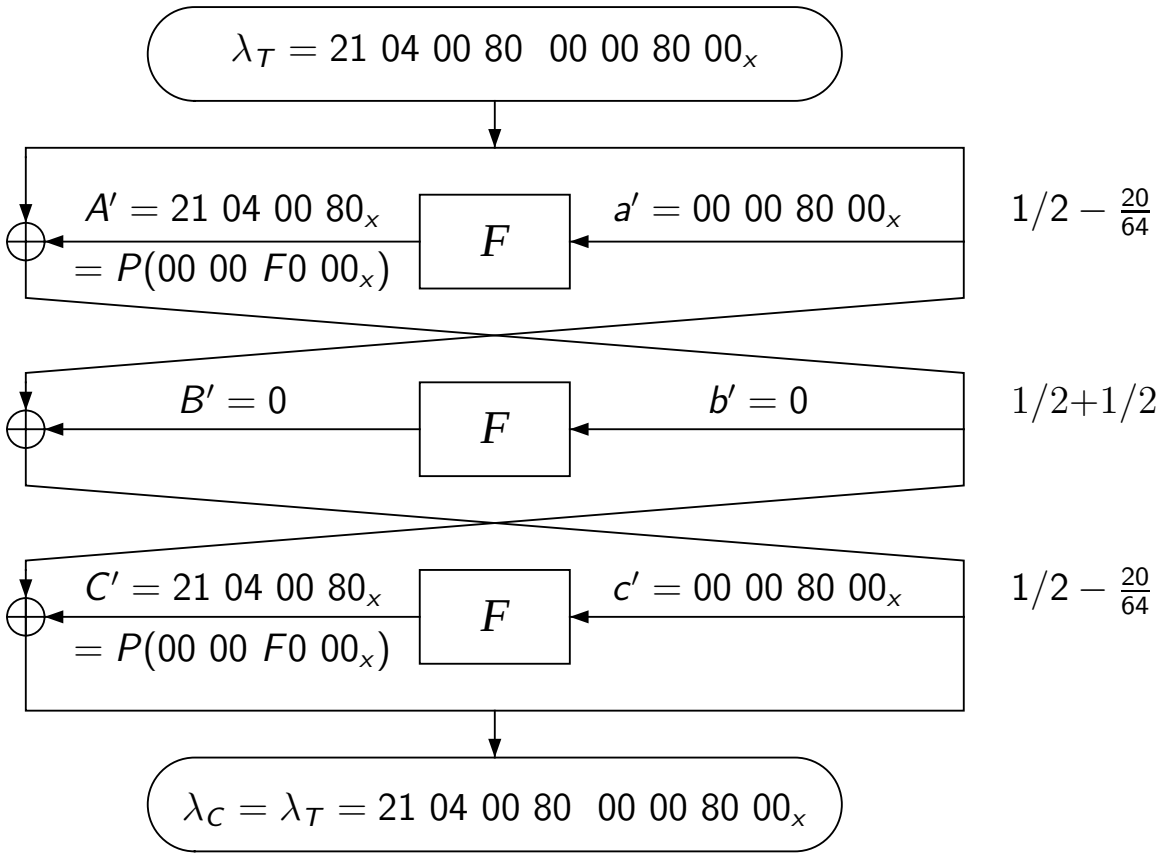
$$P_s = \Phi\left(2c - \Phi^{-1}\left(1 - 2^{-a-1}\right)\right).$$

To achieve a success probability of P_s, set

$$N = \left(\frac{\Phi^{-1}(P_s) + \Phi^{-1}\left(1 - 2^{-a-1}\right)}{2}\right)^2 \cdot \epsilon^{-2}.$$

Linear Approximation of DES

A three-round linear approximation of DES with bias $1/2 + 2 \cdot (\frac{20}{64})^2 = 1/2 + \frac{25}{128}$:

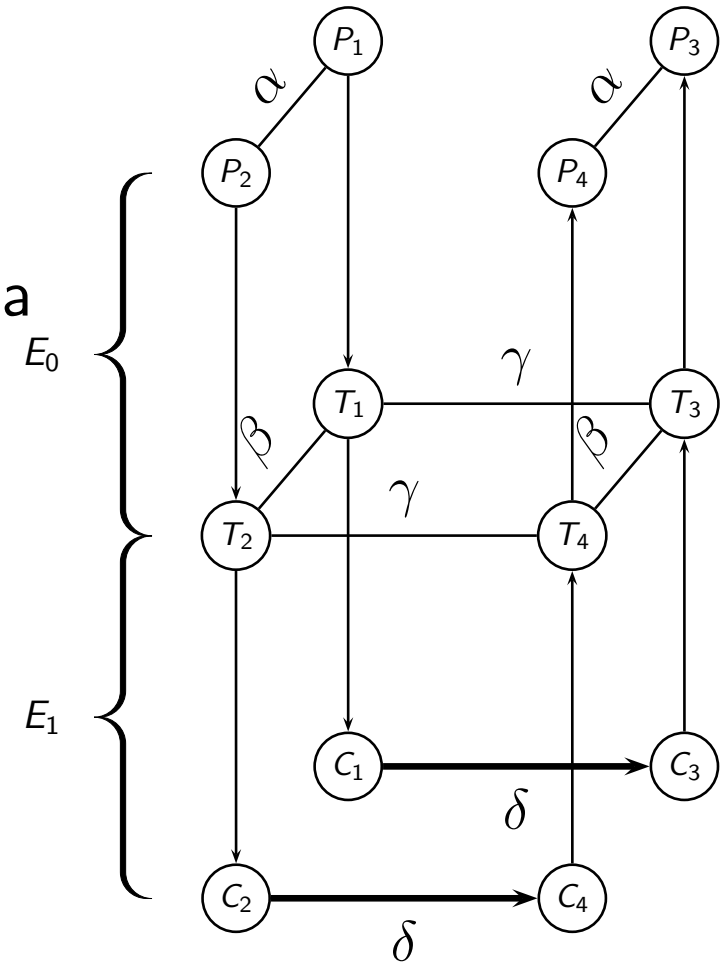


Some General Comments

- Finding good differential characteristics/linear approximation is a hard task.
- Some automatic tools exist (Matsui's method), but it is better to study the algorithm.
- Sometimes, a better attack is obtained when using differentials (approximations) of lower probability (bias).
- Many optimizations for both attacks exist. Consider differential cryptanalysis:
  - Structures of plaintexts,
  - Discarding wrong pairs (early abort),
  - Using multiple differentials,

The Boomerang Attack

- Introduced by [W99].
- Targets ciphers with good short differentials, but bad long ones.
- The core idea: Treat the cipher as a cascade of two sub-ciphers. Where in the first sub-cipher a differential $\alpha \xrightarrow{E_0} \beta$ exists, and a differential $\gamma \xrightarrow{E_1} \delta$ exists for the second.
- The process starts with a pair of plaintexts: $P_1, P_2 = P_1 \oplus \alpha$.
- After the first sub-cipher, $T_1 \oplus T_2 = \beta$.



The Boomerang Attack — Some Details

- If the probability of the first differential is p, and of the second differential is q, the total probability of the boomerang quartet is

$$\Pr[\alpha \to \beta]^2 \cdot \Pr[\gamma \to \delta]^2 = (pq)^2.$$

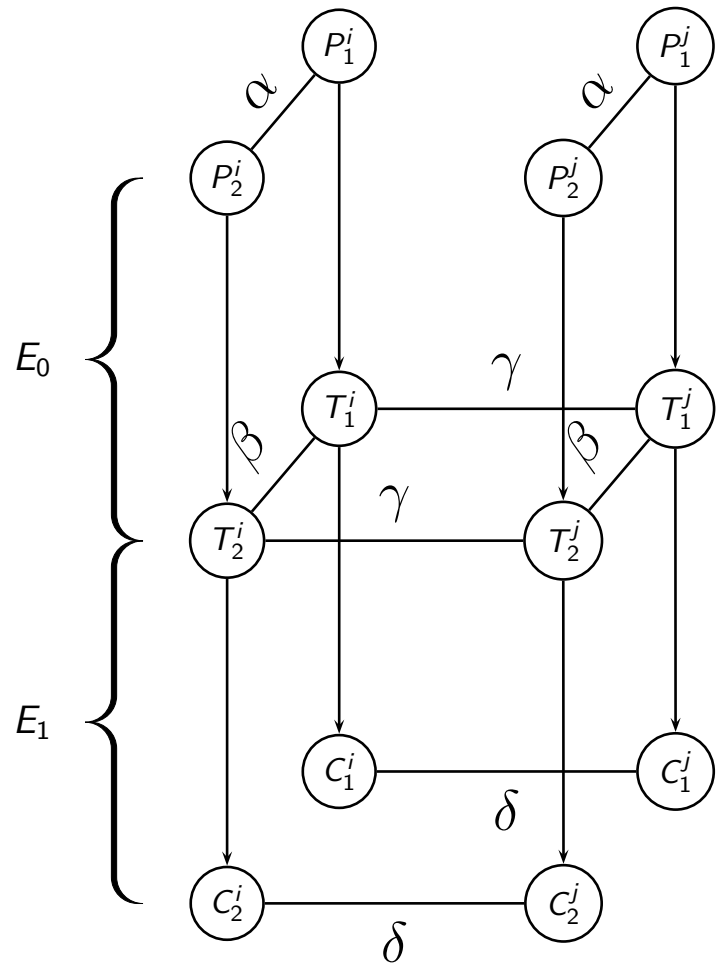
- Note that we use three out of the four differentials in the backward direction.
- For regular differentials, the probability is the same.
- However, for truncated differentials, the probability is not necessarily the same.

The Boomerang Attack — Some More Details

- A right boomerang quartet discloses information about the key.
- At the same time, the attack is an adaptive chosen plaintext and ciphertext attack.
- This prevents us from using many of the cryptanalytic techniques that were proposed over the years.
- To overcome this, we need to transform the attack into a chosen plaintext attack.

The Amplified Boomerang Attack

- Introduced by [KKS00].
- Similar idea to the boomerang attack, but in a chosen plaintext scenario.
- Again, assume the existence of two differentials: $\alpha \xrightarrow{E_0} \beta$ for the first sub-cipher and $\gamma \xrightarrow{E_1} \delta$ for the second.
- Take many pairs of plaintext with difference α: $P_1^i, P_2^i = P_1^i \oplus \alpha$.
- After the first sub-cipher, for some of them $T_1^i \oplus T_2^i = \beta$.



The Amplified Boomerang Attack — Some Details

- If the probability of the first differential is p, and of the second differential is q, the total probability of the amplified boomerang quartet is

$$\Pr[\alpha \to \beta]^2 \cdot \Pr[\gamma \to \delta]^2 \cdot \mathbf{2}^{-n} = (pq)^2 \cdot 2^{-n}.$$

- **In other words, the probability is less than 2^{-n}!**

The Amplified Boomerang Attack — Some Details (cont.)

- If we take N pair with input difference α, we obtain about $N^2/2$ quartets.
- Hence, we expect

$$N^2/2 \cdot (pq)^2 \cdot 2^{-n}$$

right amplified boomerang quartets.

- Start with $N = O(2^{n/2}/pq)$ pairs.
- As long as $(pq) > 2^{-n/2}$, we can have enough data to run the attack.
- Which is the same condition as for the boomerang attack...

The Rectangle Attack — Three Improvements

1. If the quartet $((P_1^i, P_2^i), (P_1^j, P_2^j))$ is not a right quartet, then maybe $((P_1^i, P_2^i), (P_2^j, P_1^j))$ is a right one?
2. If $T_1^i \oplus T_2^i = \beta'$, but so does $T_1^j \oplus T_2^j = \beta'$, we can still get a right quartet.
3. If $T_1^i \oplus T_1^j = \gamma'$, but so does $T_2^i \oplus T_2^j = \gamma'$, we can still get a right quartet.

Expected number of right quartets starting with N pairs:

$$N^2 \cdot 2^{-n+1} \cdot (pq)^2$$

$$N^2 \cdot 2^{-n} \cdot (pq)^2$$

$$N^2 \cdot 2^{-n} \cdot \left(\sum_{\beta'} \Pr[\alpha \xrightarrow{E_0} \beta']^2 \right) q^2$$

$$N^2 \cdot 2^{-n} \cdot \left(\sum \Pr[\alpha \xrightarrow{E_0} \beta']^2 \right) \cdot \left(\sum \Pr[\gamma' \xrightarrow{E_1} \delta]^2 \right)$$

A Technical Problem...

- In the boomerang attack the quartet is fully known.
- In the amplified boomerang attack, one needs to find the quartets among all possible ones.
- This task is hard, as the number of candidate quartets is at least 2^n.

Underlying Assumptions for Differential Attacks

Formally, define

$$G_K\left(\alpha \xrightarrow{E} \beta\right) = \left\{P \middle| E_K(P) \oplus E_K(P \oplus \alpha) = \beta\right\}.$$

and

$$G_K^{-1}\left(\alpha \xrightarrow{E} \beta\right) = \left\{C \middle| E_K^{-1}(C) \oplus E_K^{-1}(C \oplus \beta) = \alpha\right\}.$$

These two sets contain all the right pairs (i.e., X is in the set if it is a part of a right pair).

Independence Assumptions for Differential Attacks

1. The probability of the differential characteristic in round i is independent of other rounds.

(formally: the event $X \in G_K^{-1}(\alpha \xrightarrow{E_0} \beta)$ is independent of the event $X \in G_K(\beta \xrightarrow{E_1} \gamma)$ for all K's and β)

2. Partial encryption/decryption under the wrong key makes the cipher closer to a random permutation.

Independent Subkeys

- A cipher whose subkeys are all chosen at random (independently of each other) can be modeled as a Markov chain.
- For such a cipher, the previous conditions are satisfied (under reasonable use of the keys) as the independent subkeys assure that the inputs to each round are truly random and independent.

Independent Subkeys — Where we Cheated

- The above assumes that the keys are chosen *during* the differential attack, and for each new pair of plaintexts, they are chosen again at random.
- This is of course wrong, as the key is fixed **a priori**, and the only source of "randomness" in the experiment is the plaintext pair.
- Hence, we need to assume *Stochastic Equivalence*, i.e.,

$$\Pr[\Delta C = \beta | \Delta P = \alpha] =$$

$$\Pr[\Delta C = \beta | \Delta P = \alpha \wedge K = (k_1, k_2, \ldots)]$$

for almost all keys K.

Underlying Assumptions for the Boomerang Attack

For $E = E_1 \circ E_0$, and any set of differences α, γ' and $\delta,$, we require that T is (part of) a right pair with respect to $\gamma' \xrightarrow{E_1} \delta$ independently of the following three events:

1. T is (part of) a right pair with respect to $\alpha \xrightarrow{E_0} \beta'$ for all β'.
2. $T \oplus \beta'$ is (part of) a right pair with respect to $\gamma'' \xrightarrow{E_1} \delta$ for all β', γ''.
3. $T \oplus \gamma_1$ is (part of) a right pair with respect to $\alpha \xrightarrow{E_0} \beta''$ for all β''.

When Independence Fails — Part I

- The independence may fail if
  - There is one β whose most significant bit is 0 for which $\Pr\left[\alpha \xrightarrow{E_0} \beta\right] = 1/2$.
  - For all other β': $\Pr\left[\alpha \xrightarrow{E_0} \beta'\right]$ is either 0 or 2^{-n+1}.
  - All the pairs (T, T^*) which satisfy the differential $\alpha \xrightarrow{E_0} \beta$ are such that the most significant bit of both T and T^* is set to 0.
  - There is one γ whose most significant bit is 1 for which $\Pr\left[\gamma \xrightarrow{E_1} \delta\right] = 1/2$.
  - For all other γ': $\Pr\left[\gamma' \xrightarrow{E_1} \delta\right]$ is either 0 or 2^{-n+1}.

When Independence Fails — Part II

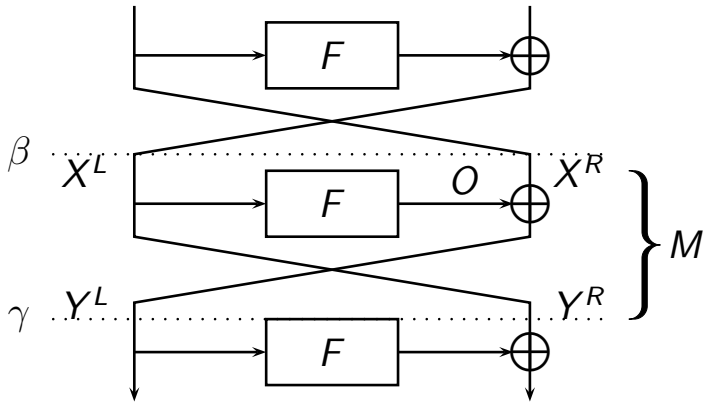
- Consider the case where the last round of the first differential characteristic relies on the transformation $x \xrightarrow{S} y$ for some S-box S.
- If the difference distribution table of S satisfies that $DDT_S(x, y) = 2$, and if the difference in γ is such that the two pairs (T_a, T_c) and (T_b, T_d) have a non-zero difference in the bits of x, then the transition is impossible.

Is it Serious?

- It is possible to construct not-so-artificial examples of boomerangs that fail one of the above two examples [M09].
- On the other hand, the failure is with respect to a pair of intermediate differences β', γ'.
- When truly taking all possible differences (in the boomerang attack or in the rectangle attack), this problem tends to "shrink".
- Sometimes, the dependence can be used for the benefit of the adversary:
  - Boomerang switch [BK09],
  - Sandwich attach [DKS10]

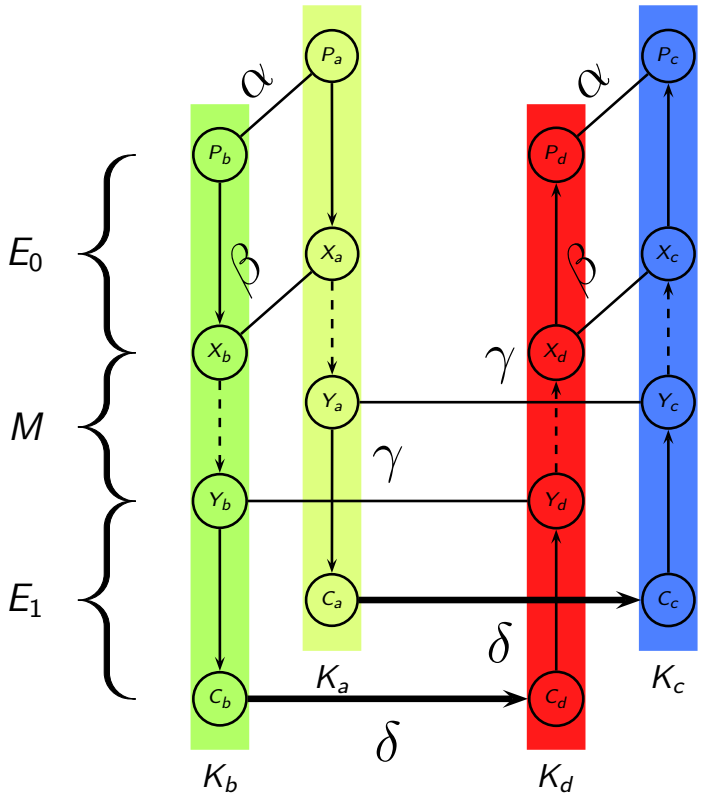
For more details: Kim et al.
`http://eprint.iacr.org/2010/019`

The Bright Side of Dependence



- Assume that $\gamma^R = 0$.
- In other words, $X_a^R = Y_a^R = Y_c^R = X_c^R$ and $X_b^R = Y_b^R = Y_d^R = X_d^R$.
- Hence, if $X_a^R \to O_a$ and $X_b^R \to O_b$, then $X_c^R \to O_a$ and $X_d^R \to O_b$ as well.
- Which ensures that the last round of the differential characteristic $\alpha \to \beta$ is satisfied for the second pair!

The Sandwich



The probability of a quartet to be a right one is:

$$\Pr[P_c \oplus P_d = \alpha] = \Pr[X_a \oplus X_b = \beta] \cdot \Pr[Y_a \oplus Y_c = \gamma] \cdot \Pr[Y_b \oplus Y_d = \gamma] \cdot \Pr[X_c \oplus X_d = \beta \,|\, \text{Previous conditions hold}] \cdot$$

The Transition M

- As noted before, M may prove that the transition happens with a lower or higher probability than expected.
- In Feistels, $\gamma^R = 0$ is indeed quite useful (as well as $\gamma^R = \beta^R$).
- For SPNs similar cases can be constructed, as demonstrated by Biryukov and Khrovatovich in the *boomerang switch*.
- This transition has various interpretations, but it is actually a (constructive) use of the **dependence**.

Differential-Linear Cryptanalysis

- Introduced by Langford and Hellman in 1994.
- The idea is to combine two statistical properties: a differential characteristic and a linear approximation.

Differential-Linear Cryptanalysis (cont.)

- Consider 6-round DES.
- Take two plaintexts $(P_1, P_2 = P_1 \oplus \Omega_P)$ for $\Omega_P = 40\ 00\ 00\ 00\ \ 00\ 00\ 00\ 00_x$.
- After three rounds, the intermediate encryption values (T_1, T_2) have no difference in more than 30 bits.
- Interestingly, five of these bits are masked by $\lambda_T = 21\ 04\ 00\ 80\ \ 00\ 00\ 80\ 00_x$.

Differential-Linear Cryptanalysis (cont.)

- In other words,

$$\lambda_T \cdot T_1 = \lambda_T \cdot T_2.$$

- We know that $\lambda_T \cdot T_1 \oplus \lambda_C \cdot C_1 = \lambda_K \cdot K$ and that $\lambda_T \cdot T_2 \oplus \lambda_C \cdot C_2 = \lambda_K \cdot K$ (each with probability of $1/2 + \frac{25}{128}$).
- Hence, $\lambda_C \cdot C_1 = \lambda_C \cdot C_2$ with probability of $1/2 + 0.0763$ (about $1/2 + 1/13.1$).
- For a random permutation, this probability is expected to be $1/2$, and about $1/(1/13.1)^2 \approx 172$ pairs with input difference Ω_P are needed.

A Differential-Linear Attack on 8-Round DES

- The attack starts with *structures* of plaintexts.
- In each structure, after the first round, there are 16 pairs of plaintexts with input difference $\Omega_P = 40\ 00\ 00\ 00\ \ 00\ 00\ 00\ 00_x$.
- After obtaining their ciphertexts:
  1. For each guess of the 6-bit subkey of $S1$ in round 1, find the pairs with input difference $\Omega_P = 40\ 00\ 00\ 00\ \ 00\ 00\ 00\ 00_x$ to the second round.
  2. For each guess of the 6-bit subkey of $S5$ in round 8, partially decrypt the pair, and check whether $\lambda_C \cdot C_1 = \lambda_C \cdot C_2$.
  3. The subkey for which $\lambda_C \cdot C_1 = \lambda_C \cdot C_2$ happens the most is likely to be the correct one.

Several Extensions

- One can deal with (truncated) differentials with probability lower than 1.
- If the differential has probability p, and the linear approximation has bias ϵ, the total bias of the differential-linear is $2p\epsilon^2$.
- If you can evaluate $\Pr[\Omega_T \cdot \lambda_T = 0]$ for many differentials — even better ([L12]).
- The sign of the bias, depends on $\Omega_T \cdot \lambda_T$.
- Even if $\Omega_T \cdot \lambda_T$ is unknown, as long as it has some more probable value, the relation $\lambda_C \cdot C_1 = \lambda_C \cdot C_2$ will be biased.

Research Directions in Cryptanalysis

- Attack various ciphers,
- Develop new attacks,
- Better mathematical foundation to some attacks,
- Better understanding of security,

Questions?

Thank you for your Attention!