Lecture Notes: Differential Cryptanalysis (DC)

Course: INFO-F-537 Cryptanalysis (Topic 2).
Document compiled by Luca Volonterio

Based on: *Gilbert-DC*, *SecretKeyPrimitive*, & *Notes-DC*.

January 5, 2026

Abstract

This document covers **Topic 2** for the oral exam. Differential Cryptanalysis is a statistical **Chosen-Plaintext Attack (CPA)** that targets the non-linear elements (S-boxes) of a cipher. This guide explains the Difference Distribution Table (DDT), the concept of Characteristics vs. Differentials, statistical aspects of distinguishers, refinements such as the Differential Effect and Truncated Differentials, and the Wide Trail Strategy used to defend AES.

Contents

1 Introduction: Definitions and Scope

1.1 What is Differential Cryptanalysis?

Differential Cryptanalysis (DC) is a Shortcut Attack (White Box) introduced by Biham and Shamir. Unlike generic attacks, it analyzes the internal structure of the cipher.

- The Core Idea: Instead of analyzing plain values (x), we analyze how a difference in the input (Δ_{in}) propagates to a difference in the output (Δ_{out}).
- The Difference Metric: In most ciphers (like DES or AES), the difference is defined by XOR ($\oplus$):

$$\Delta X = X \oplus X'$$
$$\Delta Y = Y \oplus Y'$$

- Why XOR? Because in linear operations (like MixColumns or AddRoundKey), the difference propagation is deterministic and probability 1. Non-linearity (S-boxes) is where the probability drops.

Exam Tip: Crucial Distinction: In Generic Attacks (Topic 1), we looked for collisions ($C_i = C_j$). In Differential Cryptanalysis, we look for fixed differences ($\Delta C = \text{constant}$).

—

2 The Mathematics of Differences

(Based on INFOF537-Notes-DC and Gilbert)

2.1 Propagation through Layers

1. Linear Layer (L): If $y = L(x)$, then:

$$\Delta y = L(x) \oplus L(x') = L(x \oplus x') = L(\Delta x)$$

   The difference propagates deterministically with Probability 1.

2. Key Addition ($AddRoundKey$):

$$y = x \oplus k \implies \Delta y = (x \oplus k) \oplus (x' \oplus k) = x \oplus x' = \Delta x$$

   The key cancels out! Differential Cryptanalysis is independent of the key in the intermediate rounds. This is why it is so powerful.

3. Non-Linear Layer (S-Box): This is the probabilistic part. For a given input difference Δ_{in}, not all output differences Δ_{out} are equally likely.

2.2 Difference Distribution Table (DDT)

To analyze an S-box ($S : \{0,1\}^n \to \{0,1\}^m$), we build a table called the DDT (or NDP - Number of Difference Pairs).

Definition: For every possible input difference α and output difference β, we count how many inputs x satisfy:

$$S(x) \oplus S(x \oplus \alpha) = \beta$$

The probability of a differential transition is:

$$DP_S(\alpha \to \beta) = \frac{\#\{x \in \{0,1\}^n : S(x) \oplus S(x \oplus \alpha) = \beta\}}{2^n} \quad (1)$$

Properties (from Notes):

- The sum of counts in a row is always 2^n.
- The first entry (0 to 0) always has count 2^n (Prob 1).
- Impossible transitions have count 0.

2.3 Statistical View and Markov Assumption

In practice, an iterated block cipher is seen as a random process over differences, and differential trails are analyzed under a Markov-style assumption.

- A **differential characteristic** over r rounds is assumed to have probability equal to the product of the probabilities of its round transitions, if round keys are independent and the transition behavior does not depend on the actual intermediate value (*Markov cipher* hypothesis).
- One distinguishes between the **expected differential probability** (EDP or EDCP), which is the average over all keys, and the probability for a fixed key; in many attacks one uses an additional "stochastic equivalence" hypothesis that these two are close enough for the distinguisher to behave predictably.

Exam Tip: Exam Tip: It is good to be able to state in words that "for a Markov cipher with independent round keys, the probability of a trail is the product of the one-round probabilities, and the attack uses the *expected* differential probability over keys".

—

3 Differentials vs. Characteristics

(Source: Gilbert - DC.pdf)

This distinction is often a specific exam question.

3.1 Characteristic (or Trail)

A Characteristic (or Differential Trail) specifies the exact difference pattern at every single round of the cipher.

$$Q = (\Delta_0 \xrightarrow{r_1} \Delta_1 \xrightarrow{r_2} \Delta_2 \dots \xrightarrow{r_n} \Delta_n)$$

Probability of a Trail: Assuming the rounds are independent (Markov Cipher assumption), the probability of the trail is the product of the probabilities of each round's active S-boxes.

$$P(Q) = \prod_{i=1}^{n} P(\Delta_{i-1} \to \Delta_i)$$

3.2 Differential

A Differential specifies only the Input difference and the Output difference, treating the internal steps as a black box.

$$(\Delta_{in} \to \Delta_{out})$$

The probability of a differential is the sum of the probabilities of all possible trails/characteristics that lead from Δ_{in} to Δ_{out}.

$$P(\Delta_{in} \to \Delta_{out}) = \sum_{Q} P(Q)$$

Exam Tip: Teacher's Insight: Why does this matter? Often, finding one strong trail is enough to break a cipher. However, if many weak trails cluster together (Differential Effect), the total probability might be much higher than predicted by a single trail.

—

4 Refinements: Differential Effect and Truncation

(Source: Gilbert - DC.pdf)

4.1 Differential Effect

In practice, there are often many different trails between the same $(\Delta_{in}, \Delta_{out})$, not just the "best" one.

- The **differential effect** refers to the fact that the differential probability $P(\Delta_{in} \to \Delta_{out})$ can be significantly larger than the probability of the single best characteristic, because many low-probability trails contribute to the same input–output difference.
- For some ciphers (e.g. DES-like structures) these clusters of trails must be taken into account when estimating the true data complexity of a differential attack, otherwise the designer or the attacker may under- or over-estimate security.

4.2 Truncated Differentials

Instead of fixing the full bit pattern of the output difference, one can work with partially specified differences.

- A **truncated difference** specifies which bit (or byte) positions are zero, non-zero, or "don't care", while allowing any actual non-zero value in the active positions.
- A **truncated differential (characteristic)** tracks only this coarse pattern through the rounds; for example, on DES there exist 4-round truncated differentials of probability 1 that guarantee that some pattern of active S-boxes appears, even if the exact output difference bits are not fixed.

Exam Tip: Exam Tip: Be ready to define in one sentence what a truncated differential is: "we predict only the *support* (positions of non-zero differences), not the exact bit values".

—

5 The Attack Logic: Signal-to-Noise

(Source: Gilbert - DC.pdf)

How do we actually find the key?

1. Distinguishing Phase: We send pairs with difference Δ_{in}. If the cipher were random, the output difference Δ_{out} would appear with probability $1/2^n$. If our trail (or differential) holds, it appears with probability $p \gg 1/2^n$.
2. Key Recovery Phase (Last Round Attack):
   - Choose a trail that covers $R-1$ rounds with high probability.
   - Request encryption of pairs (P, P') with $\Delta P = \Delta_{start}$.
   - Get ciphertexts (C, C').
   - Guess a portion of the last round key k_R.
   - Partially decrypt C and C' one round back to check if the difference matches the trail's expected output.
   - Signal-to-Noise Ratio (S/N): A counter is incremented for the key guess. The correct key will result in a high count (Signal). Wrong keys will look random (Noise).

Exam Tip: In many classical DC attacks the number of required pairs M is roughly proportional to $1/p_S$, where p_S is the signal probability of the exploited differential, assuming a good signal-to-noise ratio.

—

6 Defense: The Wide Trail Strategy (AES)

(Source: INFOF537-SecretKeyPrimitive.pdf and Gilbert - DC.pdf)

Designers use this strategy to prove resistance against DC (and LC).

6.1 The Goal

We cannot eliminate differentials (mathematically impossible). We can only make their probability so low that an attack requires more data than the entire codebook (e.g., $> 2^{128}$ pairs).

To lower the probability of a trail:

$$P(\text{trail}) \approx \prod p_{sbox}$$

We need to minimize the max probability of S-boxes and maximize the number of active S-boxes.

6.2 Branch Number & Diffusion

In AES (Rijndael), the MixColumns layer ensures high diffusion.

Theorem: The sum of active input bytes (a) and active output bytes (b) of the MixColumns operation is at least the Branch Number ($\mathcal{B}$).

$$a + b \geq 5 \tag{2}$$

(For AES MixColumns).

Implication:

- If you have 1 active byte coming into MixColumns (difference $\neq 0$), you will have at least 4 active bytes coming out.
- This forces a "multiplication" of active S-boxes in the next round, and over 4 rounds of AES any differential trail must activate at least 25 S-boxes in total.
- Since for the AES S-box $P_{sbox}^{\max} \leq 2^{-6}$, a 4-round trail then has probability at most $(2^{-6})^{25} = 2^{-150}$, which is already below 2^{-128} and thus below exhaustive search on a 128-bit block.

Exam Tip: Exam Tip: A nice one-liner is: “Because of the wide trail (branch number 5), any 4-round differential in AES has at least 25 active S-boxes, so its probability is upper-bounded by 2^{-150}.”

—

7 Example: The χ (Chi) Function

(Source: INFOF537-Notes-DC.pdf)

The notes provide an example of analyzing a small non-linear function, χ (used in Keccak), to calculate differential probabilities.

Function: $y_i = x_i + (x_{i+1} + 1)x_{i+2}$ (indices mod 3). This is a 3-bit S-box.

Calculating a specific entry: If we want to check the transition $\Delta_{in} = (1, 0, 0) \rightarrow \Delta_{out} = (1, 0, 0)$:

1. We write the algebraic expression for the difference.
2. $\Delta y = \chi(x) \oplus \chi(x \oplus \Delta_{in})$.
3. We check for how many inputs x this equals $(1, 0, 0)$.
4. If it holds for 4 inputs out of 8, the probability is $4/8 = 1/2$.
5. The Weight of this transition is $-\log_2(1/2) = 1$.

—

8 Self-Assessment Questions

Level 1: Concepts

Q: Why does the key addition layer not affect the probability of a differential trail?
Answer: Because in Differential Cryptanalysis we look at XOR differences. $\Delta y = (x \oplus k) \oplus (x' \oplus k) = x \oplus x'$. The key k cancels out with itself.

Q: What is the difference between a Characteristic and a Differential?
Answer: A Characteristic (Trail) defines the difference at every intermediate round. A Differential only defines the input and output difference (summing up all internal trails).

Level 2: Design & Strategy

Q: Explain the "Wide Trail Strategy" in simple terms.
Answer: It is a design method (used in AES) to guarantee security against DC. By using a diffusion layer with a high Branch Number (like MixColumns), the cipher forces any difference trail to activate many S-boxes. Since each active S-box reduces the probability, having many active S-boxes makes the attack probability negligible.

Q: What does the Branch Number constraint $a + b \geq 5$ mean for AES?
Answer: It means that for the MixColumns operation, the number of non-zero input bytes plus the number of non-zero output bytes is at least 5. If you input 1 difference byte, you get 4 out. If you input 2, you get at least 3 out. This ensures rapid diffusion.

Level 3: Analysis

Q: How do we determine the validity of a key guess in the last round of a Differential Attack?
Answer: We use the Signal-to-Noise ratio. We decrypt the last round using the guessed key for many pairs. If the resulting difference matches the expected trail, we increment a counter. The correct key will produce a statistically significant peak (Signal) compared to random keys (Noise).

Q: What is the Differential Effect, and why can it make an attack stronger than what a single trail suggests?
Answer: The differential effect is the phenomenon where many different trails share the same input and output difference, so the differential probability is the sum of all their probabilities. This can make the effective probability of $(\Delta_{in} \to \Delta_{out})$ much higher than the probability of the single best trail, giving a stronger distinguisher or key-recovery attack.

Q: What is a truncated differential, and when is it useful?

Answer: A truncated differential only fixes which bit or byte positions have non-zero difference, without fixing the exact non-zero values. It is useful when it is easier to control or predict the *pattern* of activity (which S-boxes or bytes are active) than the exact difference, for instance in some DES or AES-like constructions.