# Chuẩn mã dữ liệu (DES)

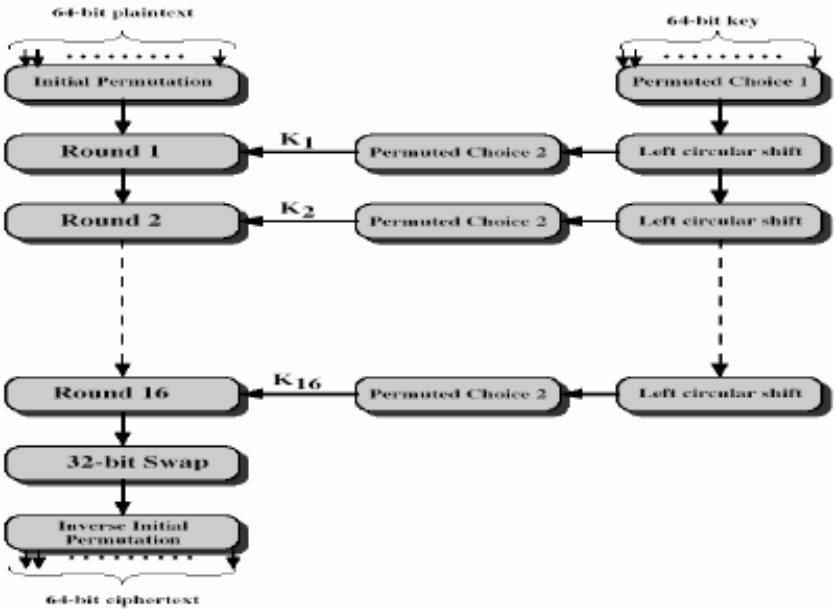
Bởi:

TS. Trần Văn Dũng

## Lịch sử DES

Cuối những năm 1960, IBM phát triển mã Lucifer, được lãnh đạo bởi Fiestel. Ban đầu Lucifer sử dụng khối dữ liệu 64 bít và khoá 128 bít. Sau đó tiếp tục phát triển như mã thương mại. Năm 1973 NBS yêu cầu đề xuất chuẩn mã Quốc gia. IBM đề nghị bản sửa đổi Lucifer, sau này gọi là DES. Đã có các tranh luận về thiết kế của DES. Vì chuẩn của DES được công khai, mọi người đóng góp ý kiến về tốc độ, độ dài khoá và mức độ an toàn, khả năng thám mã. Người ta đề xuất chọn khoá 56 bít thay vì 128 để tăng tốc độ xử lý và đưa ra các tiêu chuẩn thiết kế một chuẩn mã dữ liệu. Các suy luận và phân tích chứng tỏ rằng thiết kế như vậy là phù hợp. Do đó DES được sử dụng rộng rãi, đặc biệt trong lĩnh vực tài chính.

## Sơ đồ mã DES




- **Hoán vị ban đầu IP:** đây là bước đầu tiên của tính toán dữ liệu, hoán vị IP đảo thứ tự các bít đầu vào: các bít chẵn sang nửa trái và các bít lẻ sang nửa phải. Hoán vị trên dễ dàng thực hiện trên phần cứng. Mỗi số trong hệ 16 biểu diễn

bởi 4 bit, 16 số được thể hiện bởi 64 bit. Mỗi bit có một vị trí xác định qua hoán vị ban đầu (xem bảng phụ lục cuối tài liệu).

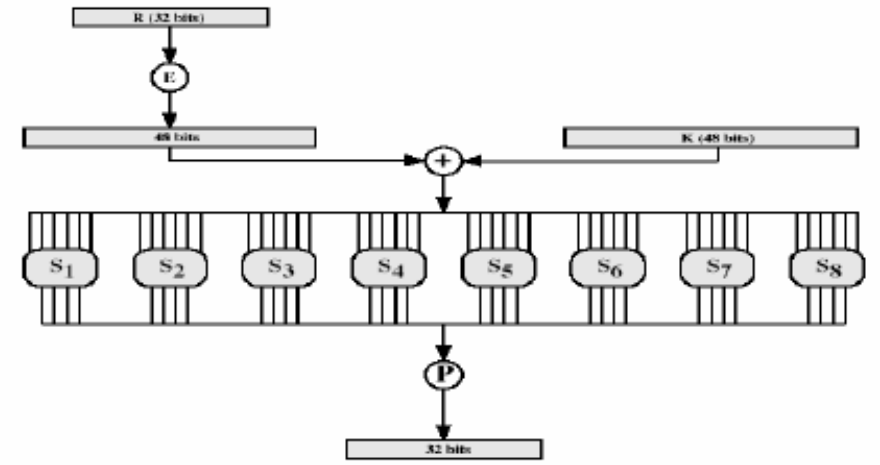
IP(675a6967 5e5a6b5a) = (ffb2194d 004df6fb)

- **Cấu tạo một vòng của DES**

Sử dụng hai nửa 32 bít trái và 32 bít phải. Như đối với mọi mã Fiestel, nửa phải của vòng trước được chuyển qua nửa trái của bước sau và lấy đầu ra của hàm vòng trên nửa phải và khoá con cộng cơ số 2 với nửa trái. Có thể biểu diễn bằng công thức như sau:

$L_i = R_{i-1}$
$R_i = L_{i-1} \text{ xor } F(R_{i-1}, K_i)$

Ở đây F lấy 32 bít nửa phải R, mở rộng thành 48 bít nhờ hoán vị E, rồi cộng vào với khoá con 48 bít. Sau đó chia thành 8 cụm 6 bít và cho qua 8 S-box để nhận được kết quả 32 bít. Đảo lần cuối sử dụng hoán vị 32 bít P nhận được 32 bít đầu ra, rồi cộng với nửa trái để chuyển thành nửa phải của bước sau.




- **Các hộp thế S (xem phụ lục cuối tài liệu)**

Có 8 hộp S khác nhau ánh xạ 6 bít vào 4 bít. Các hộp S box thực hiện các phép thế, chúng được cấu tạo không có qui luật và cố định. Mỗi S box là hộp 4 x 16 bít, mỗi hàng là một hoán vị của 16 phần tử. Giả sử ta có 6 bít đầu vào. Ta lấy hai bít ngoài 1-6 ghép lại được số nhị phân xác định chọn hàng từ 0 đến 3 trong S box. Bốn bít từ 2 đến 5 là một số nhị phân xác định cột từ 0 đến 15 trong S box. Lấy phần tử tương ứng trên hàng và cột mới được xác định, đây là một số từ 0 đến 15, chuyển sang số nhị phân ta được 4 bít đầu ra. Như vậy 48 bít chia thành có 8 cụm 6 bít, qua 8 S box được chuyển thành 8 cụm 4 bít, tổng cộng là 32 bít Việc chọn hàng trong các S box phụ thuộc cả dữ liệu và khoá - đặc trưng này được gọi là khoá tự xác định

S(18 09 12 3d 11 17 38 39) = 5fd25e03

- **Sinh khoá con của DES**

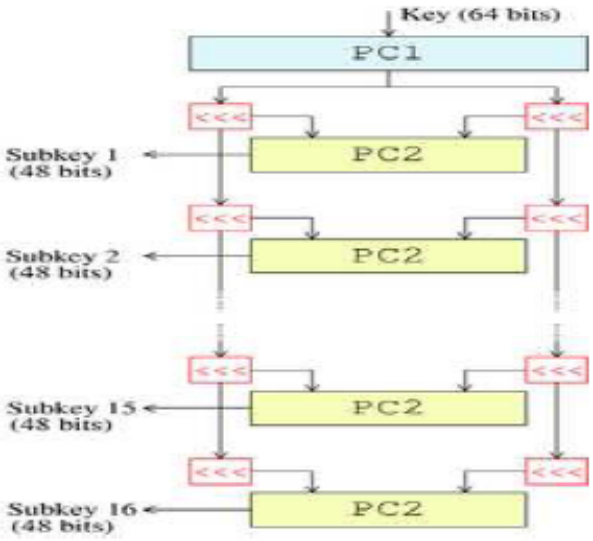
- Tạo 16 khoá con sử dụng cho 16 vòng của DES. 56 bit khoá đầu vào được sử dụng như bảng 8 x 8, trong đó cột thứ 8 không sử dụng.

- Hoán vị ban đầu của khoá PC1 và tách 56 bít thành hai nửa 28 bít.

- 16 giai đoạn bao gồm

  - Ở mỗi vòng nửa trái và nửa phải được dịch trái vòng quanh tương ứng 1 và 2 bit. Hai nửa này được dùng tiếp cho vòng sau.
  - Đồng thời hai nửa cũng cho qua hoán vị PC2 và chọn mỗi nửa 24 bít gộp lại thành 48 bít để sinh khoá con..

- Ứng dụng thực tế trên cả phần cứng và phần mềm đều hiệu quả




Các thông số cụ thể về hoán vị ban đầu, các hộp Box và thuật toán sinh khoá của DES được cho cuối tài liệu trong phần phụ lục.

- **Giải mã DES**

Giải mã làm ngược lại quá trình mã hoá. Với thiết kế Fiestel thực hiện mã hoá tiếp với các khoá con từ SK16 ngược lại về SK1. Nhận thấy rằng hoán vị ban đầu IP sẽ trả lại tác dụng của hoán vị cuối FP. Vòng đầu với SK16 sẽ trả lại tác dụng của vòng mã thứ 16. Vòng thứ 16 với SK1 sẽ trả lại tác dụng của vòng mã đầu tiên. Hoán vị cuối FP trả lại tác dụng hoán vị ban đầu IP. Như vậy đã khôi phục lại được dữ liệu ban đầu.

## Tính chất của DES

**Tác dụng đồng loạt**. Khi ta thay đổi 1 bit trong khoá sẽ gây ra tác động đồng loạt làm thay đổi nhiều bit trên bản mã. Đây là tính chất mong muốn của khoá trong thuật toán mã hoá. Nếu thay đổi 1 bít đầu vào hoặc khoá sẽ kéo theo thay đổi một nửa số bít đầu ra. Do đó không thể đoán khoá được. Co thể nói rằng DES thể hiện tác động đồng loạt mạnh.

- **Sức mạnh của DES – kích thước khoá.**

Độ dài của khoá trong DES là 56 bít có $2^{56} = 7.2 \times 10^{16}$ giá trị khác nhau. Đây là con số rất lớn nên tìm kiếm duyệt rất khó khăn. Các thành tựu gần đây chỉ ra rằng thời gian cần thiết để giải một trang mã DES mà không biết khoá là: sau một vài tháng trên Internet trong năm 1997; một vài ngày trên thiết bị phần cứng tăng cường trong năm 1998; sau 22 giờ nếu kết hợp các biện pháp trong năm 1999. Như vậy vẫn có thể đoán được bản rõ sau một khoảng thời nhất định, nếu có nguồn lực máy tính mạnh. Chính vì vậy bây giờ người ta đã xét một vài biến thể của DES nhằm nâng cao sức mạnh cho DES.

- **Sức mạnh của DES – tấn công thời gian.**

Đây là dạng tấn công vào cài đặt thực tế của mã. Ở đây sử dụng hiểu biết về quá trình cài đặt thuật toán mà suy ra thông tin về một sô khoá con hoặc mọi khoá con. Đặc biệt sử dụng kết luận là các tính toán chiếm khoảng thời gian khác nhau phụ thuộc vào giá trị đầu vào của nó. Do đó kẻ thám mã theo dõi thời gian thực hiện mà phán đoán về khoá. Có thể kẻ thám mã sáng tạo ra các loại card thông minh phán đoán khoá, mà còn phải bàn bạc thêm về chúng.

- **Sức mạnh của DES – tấn công thám mã.**

Có một số phân tích thám mã trên DES, từ đó đề xuất xây dựng một số cấu trúc sâu về mã DES. Rồi bằng cách thu thập thông tin về mã, có thể đoán biết được tất cả hoặc một số khoá con đang dùng. Nếu cần thiết sẽ tìm duyệt những khoá còn lại. Nói chung, đó là những tấn công dựa trên phương pháp thống kê bao gồm: thám mã sai phân, thám mã tuyến tính và tấn công khoá liên kết.

- **Thám mã sai phân**

Một trong những thành tựu công khai gần đây trong thám mã là phương pháp thám mã sai phân. Nó được biết đến bởi NSA trong những năm 70, chẳng hạn trong thiết kế DES. Murphy, Birham và Shamir công bố phương pháp sai phân năm 1990. Đây là phương pháp mạnh để phân tích mã khối. Nó sử dụng phân tích hầu hết các mã khối hiện tại với mức độ thành công khác nhau. Nhưng DES có thể kháng cự lại các tấn công đó. Thám mã sai phân là tấn công thống kê chống lại các mã Fiestel. Mã Fiestel dùng các cấu trúc

mã chưa được sử dụng trước kia như thiết kế S-P mạng có đầu ra từ hàm f chịu tác động bởi cả đầu vào và khoá. Do đó không thể tìm lại được giá trị bản rõ mà không biết khoá.

Thám mã sai phân so sánh hai cặp mã có liên quan với nhau

o Với sự khác biệt đã biết ở đầu vào

o Khảo sát sự khác biệt ở đầu ra

o Khi với cùng khoá con được dùng

o Trong công thức sau với hai đầu vào khác nhau, vế trái là sự khác biệt mã ở cùng vòng thứ i được biểu diễn qua sự khác biệt mã ở vòng trước đó i-1 và sự khác biệt của hàm f trong ngoặc vuông.

$$\begin{aligned}\Delta m_{i+1} &= m_{i+1} \oplus m'_{i+1} \\ &= \left[m_{i-1} \oplus \mathrm{f}(m_i, K_i)\right] \oplus \left[m'_{i-1} \oplus \mathrm{f}(m'_i, K_i)\right] \\ &= \Delta m_{i-1} \oplus \left[\mathrm{f}(m_i, K_i) \oplus \mathrm{f}(m'_i, K_i)\right]\end{aligned}$$

Sự khác biệt ở đầu vào cho sự khác biệt ở đầu ra với một xác suất cho trước.

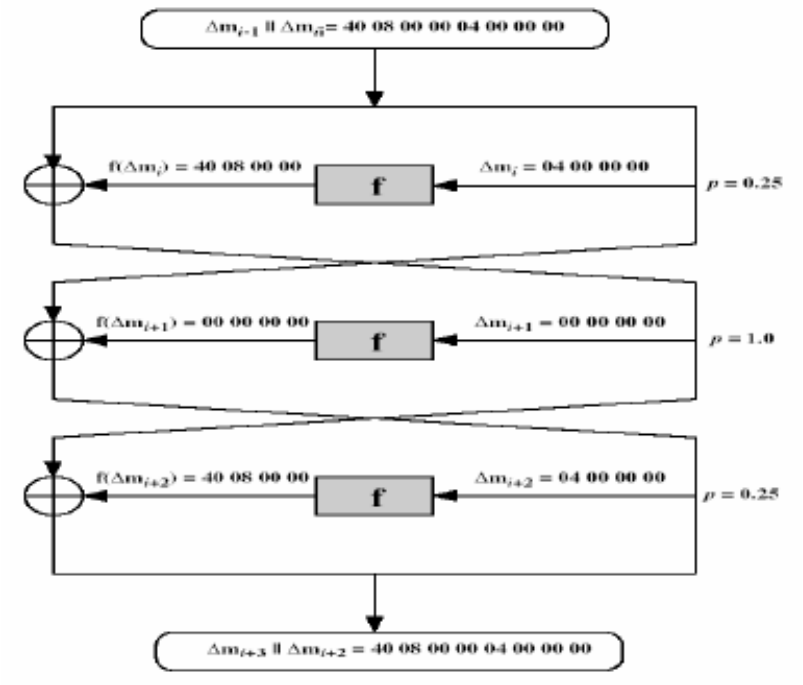
- Nếu tìm được một thể hiện đầu vào - đầu ra với xác suất cao. Thì có thể luận ra khoá con được sử dụng trong vòng đó

- Sau đó có thể lặp lại cho nhiều vòng (với xác suất giảm dần)

  - Cặp đúng cho bít khoá như nhau
  - Cặp sai cho giá trị ngẫu nhiên

- Đối với số vòng lớn, xác suất để có nhiều cặp đầu vào 64 bít thoả mãn yêu cầu là rất nhỏ.

-Birham và Shamir chỉ ra rằng làm như thế nào để các đặc trưng lặp của 13 vòng có thể bẻ được DES 16 vòng đầy đủ.




- Qui trình thám mã như sau: thực hiện mã hoá lặp lại với cặp bản rõ có XOR đầu vào biết trước cho đến khi nhận được XOR đầu ra mong muốn

- Khi đó có thể tìm được

  - nếu vòng trung gian thỏa mãn XOR yêu cầu thì có cặp đúng
  - nếu không thì có cặp sai, tỷ lệ sai tương đối cho tấn công đã biết trước dựa vào thống kê.

- Sau đó có thể tạo ra các khoá cho các vòng theo suy luận sau

  - **Thám mã tuyến tính**

Đây là một phát hiện mới khác. Nó cũng dùng phương pháp thống kê. Ở đây cần lặp qua các vòng với xác suất giảm, nó được phát triển bởi Matsui và một số người khác vào đầu những năm 90. Cơ sở của phương pháp dựa trên tìm xấp xỉ tuyến tính. Và có nhận định rằng có thể tấn công DES với 247 bản rõ đã biết. Như vậy thám mã tuyến tính vẫn không khả thi trong thực tế.

- Tìm xấp xỉ tuyến tính với xác suất

p != ½ P[i1,i2,...,ia] (+) C[j1,j2,...,jb] = K[k1,k2,...,kc]

trong đó ia, jb, kc là các vị trí bit trong bản rõ, mã, khoá.

- Điều kiện trên cho phương trình tuyến tính của các bít khoá. Để nhận được 1 bít khoá sử dụng thuật toán lân cận tuyến tính

- Sử dụng một số lớn các phương trình thử nghiệm. Hiệu quả cho bởi |p– 1/2|

Trong quá trình tìm hiểu DES người ta đã hệ thống lại các tiêu chuẩn thiết kế DES. Như báo cáo bởi Copperscmith trong [COPP94]:

o Có 7 tiêu chuẩn đối với S box được cung cấp để đảm bảo

- tính phi tuyến tính
- chống tham mã sai phân
- Rối loạn tốt

o Có 3 tiêu chuẩn cho hoán vị P để tăng độ khuếch tán

- **Các nguyên lý mã khối**

Các nguyên lý cơ bản của mã khối giống như Fiestel đề xuất trong những năm 70:

o Có một số vòng: càng nhiều càng tốt; tấn công tốt nhất phải tìm tổng thể

o Trong mỗi vòng có hàm cung cấp độ rối loạn là phi tuyến, tác động đồng loạt

o Qui trình sinh khoá con phức tạp, khoá tác động đồng loạt đến bản mã.

## Các kiểu thao tác của DES

Mã khối mã các block có kích thước cố định. Chẳng hạn DES mã các block 64 bít với khoá 56 bít Cần phải có cách áp dụng vào thực tế vì các thông tin cần mã có kích thước tùy ý. Trwosc kia có 4 kiểu thao tác được định nghĩa cho DES theo chuẩn **ANSI: ANSI X3.106-1983** Modes of Use. Bây giờ mở rộng them có 5 cách cho DES và chuẩn mã nâng cao (AES – Advanced Encryption Standards). Trong đó có kiểu áp dụng cho khối và có kiểu áp dụng cho mã dòng.

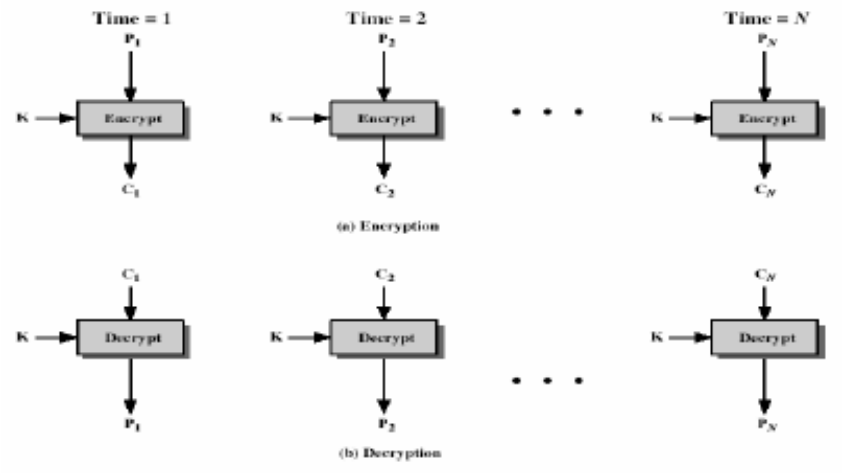
### 1. Sách mật mã điện tử (Electronic Codebook Book - ECB)

o Mẫu tin được chia thành các khối độc lập, sau đó mã từng khối

o Mỗi khối là giá trị cần thay thế như dùng sách mã, do đó có tên như vậy

o Mỗi khối được mã độc lập với các mã khác Ci = DESK1 (Pi)

o Khi dùng: truyền an toàn từng giá trị riêng lẻ




**o Ưu và nhược của ECB**

- Lặp trên bản mã được chỉ rõ lặp trên bản tin

+ Nếu dóng đúng khối

+ Đặc biệt với hình ảnh

- Hoặc với bản tin mà thay đổi rất ít sẽ trở thành đối tượng để thám mã

+ Nhược điểm là các khối được mã độc lập

+ Được sử dụng chủ yếu khi gửi một ít dữ liệu

**2. Dây chuyền mã khối (Cipher Block Chaining - CBC)**

o Các mẫu tin được chia thành các khối

o Nhưng chúng được liên kết với nhau trong quá trình mã hoá

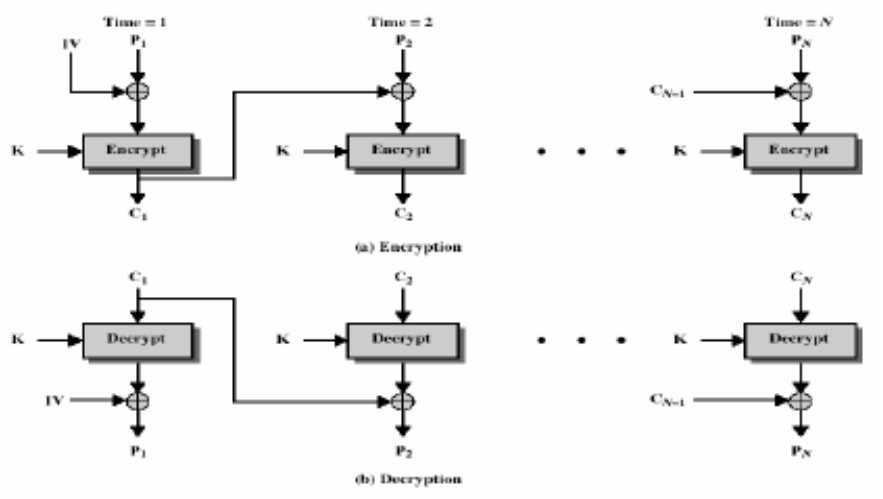
o Các block được sắp thành dãy, vì vậy có tên như vậy

o Sử dụng véctơ ban đầu IV để bắt đầu quá trình

Ci = DESK1(Pi XOR Ci-1)

$C_{-1}$ = IV

o Dùng khi: mã dữ liệu lớn, xác thực




**o Ưu và nhược của CBC**

- Mỗi khối mã phụ thuộc vào tất cả các khối bản rõ

- Sự thay đổi của bản tin ở đâu đó sẽ kéo theo sự thay đổi của mọi khối mã

- Cần giá trị véc tơ ban đầu IV được biết trước bởi người gửi và người nhận

+ Tuy nhiên nếu IV được gửi công khai, kẻ tấn công có thể thay đổi bít đầu tiên và thay đổi cả IV để bù trừ

+ Vậy IV cần phải có giá trị cố định trước hoặc mã hoá trong chế độ ECB và gửi trước phần còn lại của mẩu tin

- Ở cuối bản tin, để kiểm soát các block ngắn còn lại

+ Có thể bổ sung các giá trị không phải dữ liệu như NULL

+ Hoặc dùng bộ đệm cuối với số byte đếm kích thước của nó.

[ b1 b2 b3 0 0 0 0 5] <- 3 data bytes, vậy có 5 bytes dành cho đệm và đếm.

**3. Mã phản hồi ngược (Cipher FeedBack - CFB)**

o Bản tin coi như dòng các bít

o Bổ sung vào đầu ra của mã khối

o Kết quả phản hồi trở lại cho giai đoạn tiếp theo, vì vậy có tên như vậy.

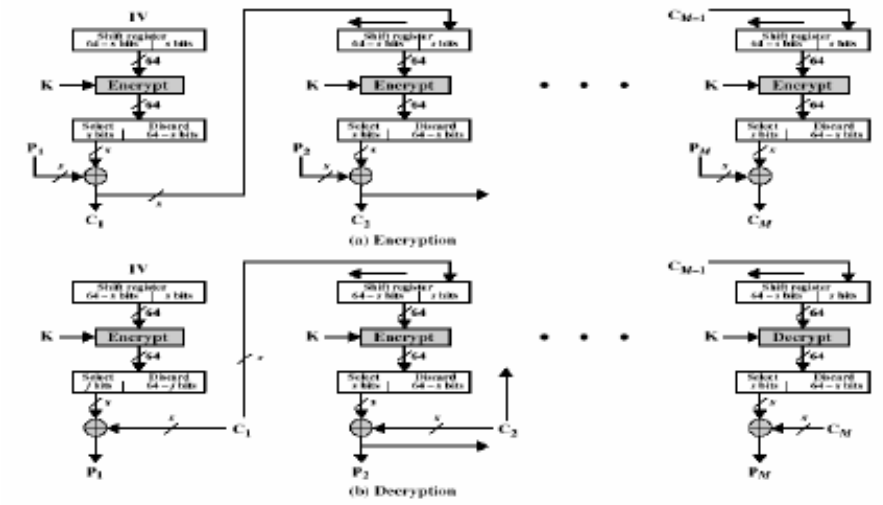
o Nói chung cho phép số bít phản hồi là 1, 8, 64, hoặc tuỳ ý: ký hiệu tương ứng là CFB1, CFB8, CFB64,…

o Thường hiệu quả sử dụng cả 64 bít

Ci = Pi XOR DESK1(Ci-1)

C-1 = IV

o Được dùng cho mã dữ liệu dòng, xác thực




**Ưu và nhược điểm của mã phản hồi ngược**

o Được dùng khi dữ liệu đến theo byte/bit

o Chế độ dòng thường gặp nhất

o Hạn chế là cần ngăn chuồng khi mã khối sau mỗi n bit

o Nhận xét là mã khối được dùng ở chế độ mã ở cả hai đầu

o Lỗi sẽ lan ra một vài block sau lỗi

**4. Phản hồi ngược đầu ra (Output FeedBack - OFB)**

o Mẩu tin xem như dòng bit

o Đầu ra của mã được bổ sung cho mẩu tin

o Đầu ra do đó là phản hồi, do đó có tên như vậy

o Phản hồi ngược là độc lập đối với bản tin

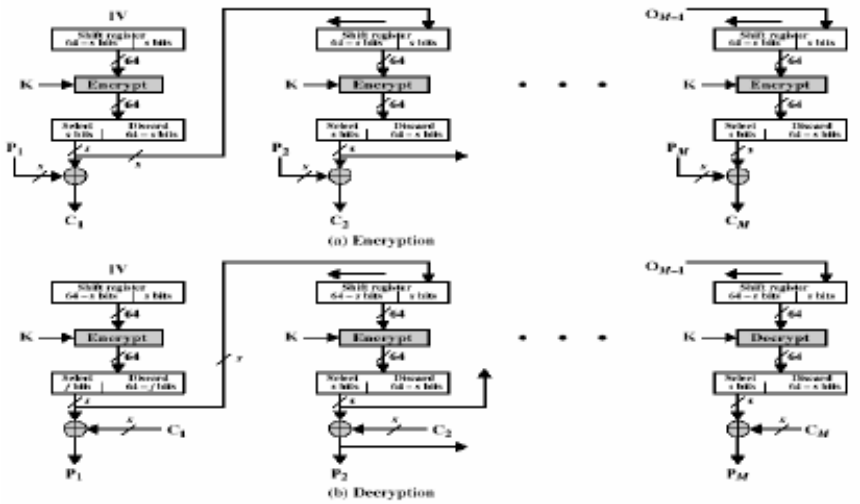
o Có thể được tính trước

Ci = Pi XOR Oi

Oi = DESK1(Oi-1)

O-1 = IV

o Được dùng cho mã dòng trên các kênh âm thanh




**Ưu điểm và nhược điểm của OFB**

o Được dùng khi lỗi phản hồi ngược lại hoặc ở nơi cần mã trước khi mẩu tin sẵn sàng

o Rất giống CFB

o Nhưng phản hồi là từ đầu ra của mã và độc lập với mẩu tin

o Là biến thể của mã Vernam, suy ra không sử dụng lại với cùng một dãy (Key + IV)

o Người gửi và người nhận phải đồng bộ, có phương pháp khôi phục nào đó là cần thiết để đảm bảo việc đó.

o Nguyên bản chỉ rõ m bit phản hồi ngược theo các chuẩn

o Các nghiên cứu tiếp theo chỉ ra rằng chỉ có OFB64 là dùng được

**5. Bộ đếm CTR (Counter)**

o Là chế độ mới, tuy đã được đề xuất từ lâu

o Giống như OFB, nhưng mã giá trị đếm thay vì giá trị phản hồi tuỳ ý.

o Cần phải có khoá khác và giá trị đếm cho mỗi khối bản rõ (không bao giờ dùng lại)

Ci = Pi XOR Oi

Oi = DESK1(i)

o Được dùng mã trên mạng với tốc độ cao

o Ưu và nhược điểm của CTR

- Hiệu quả

- Do có thể mã song song

- Chuẩn bị trước nếu cần

- Tốt cho các kết nối với tốc độ rất cao

- Truy cập ngẫu nhiên đến các khối dữ liệu mã
- Tính an toàn có thể chứng minh được
- Nhưng phải tin tưởng không bao giờ dùng lại khoá/đếm, nếu không có thể bẻ.