Chương 1. Mã hoá và giải mã

Encoding and decoding

[**Chap 1.** Jiri Adamek, ***Foundations of Coding***]

ntnhut@hcmus.edu.vn

Vấn đề chính của lý thuyết thông tin

Truyền đi (transmit)

Nhiễu (noisy)

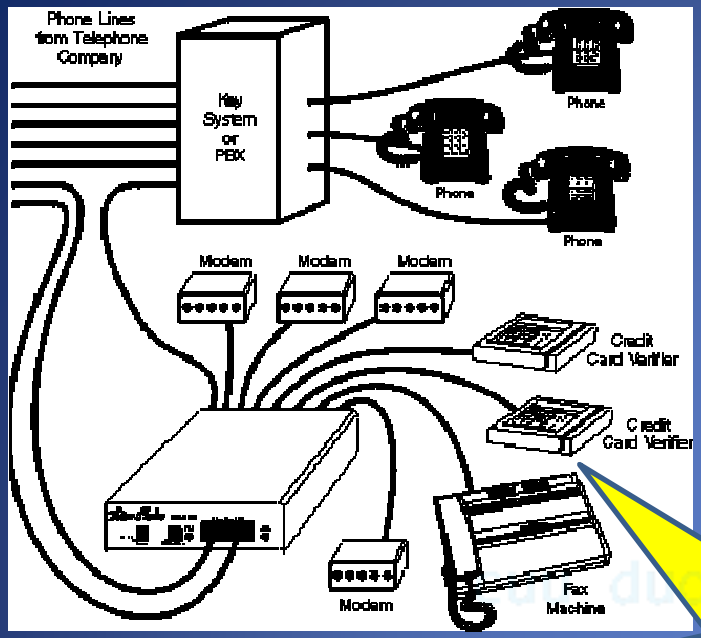
a 5 MB image

Nén (compress)

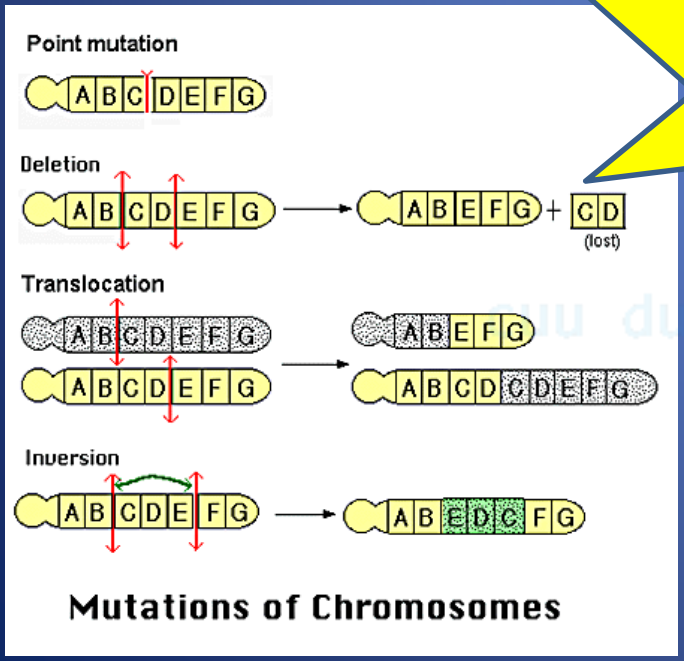
an 100 KB image

Mất chi tiết (loss of details)

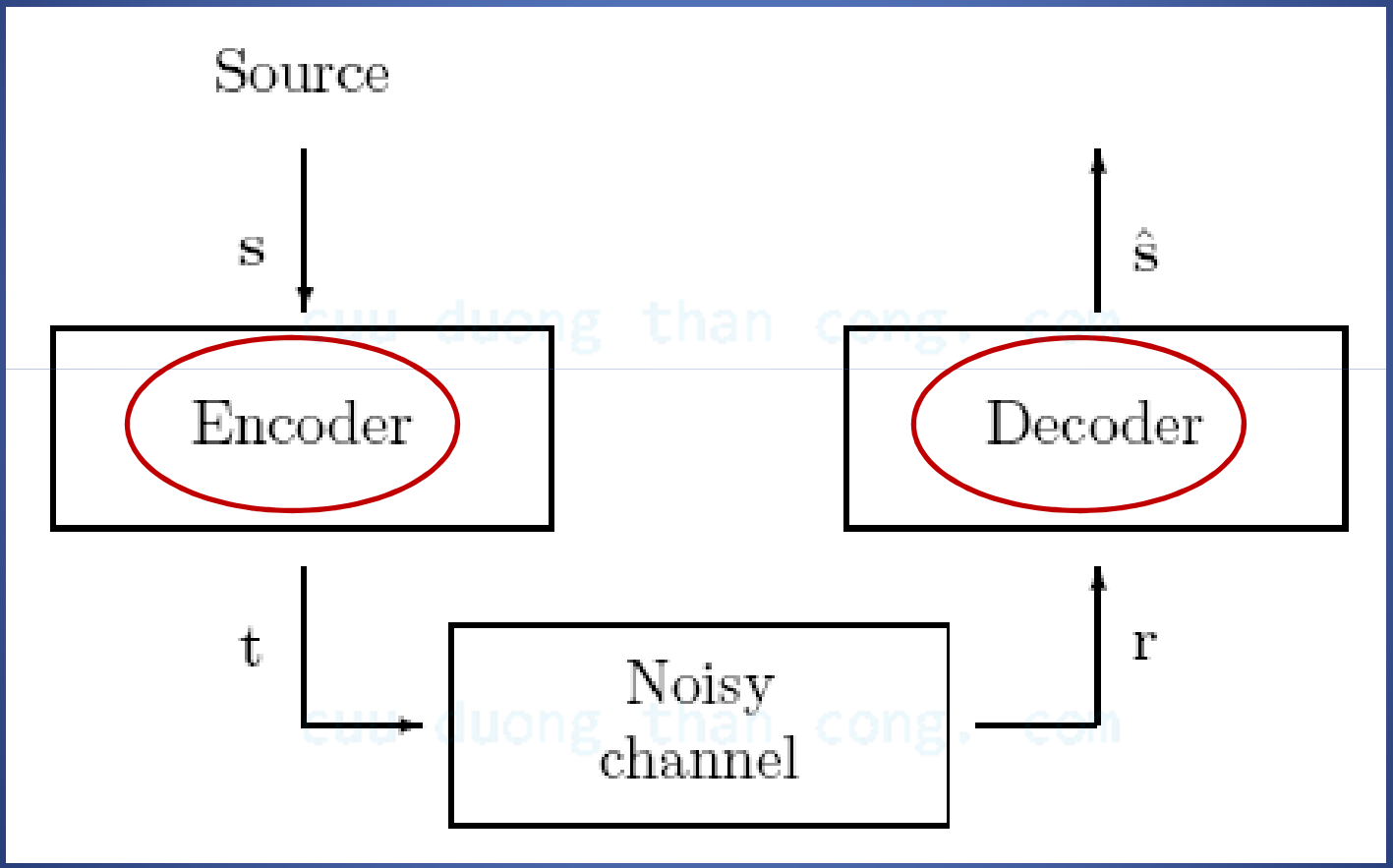
ntnhut@hcmus.edu.vn



Noisy channels



Cách khắc phục (solution)



Các khái niệm cơ sở

- ***Ký tự (letter)***: 'A', 'B', …
- ***Bảng ký tự (alphabet)*** = {các ký tự}
  - Source alphabet: bảng ký tự nguồn.
  - Code alphabet: bảng ký tự mã.
- ***Từ (word)*** = "ANHLAAI"
- Binary → bảng ký tự đang xét chỉ có 2 ký tự
  - Thường dùng '0', '1'

Mã hoá (encoding)

- $A = \{A_1, \ldots, A_p\}$: source alphabet
- $B = \{B_1, \ldots, B_q\}$: code alphabet
- ***Phép mã hoá*** ***là một đơn ánh*** **K** ***đi từ tập*** **A** ***đến tập các từ trong*** **B,**

$$\mathbf{K : A \rightarrow B^*}$$

- $K(A_i) =$ "$B_{i1} \ldots B_{ik}$" : ***từ mã (code word)***
- ***Mã (code)*** $= \{K(A_1), \ldots, K(A_p)\}$

Ví dụ

Source Symbol	Code Word
1	11000
2	10100
3	01100
4	10010
5	01010
6	00110
7	10001
8	01001
9	00101
0	00011

Mã "2-out-of-5"

- Chuỗi "173" được mã hoá thành "110001000101100"
- Giải mã thử chuỗi "100100100101010"

Giải mã (decoding)

Giải mã duy nhất

Giải mã không duy nhất

Giải mã không duy nhất

- Mã 2-out-of-5 có giải mã duy nhất không?

Ví dụ giải mã không duy nhất

- Xét mã sau

$$a \longmapsto 00 \qquad b \longmapsto 10 \qquad c \longmapsto 101 \qquad d \longmapsto 110 \qquad e \longmapsto 1001$$

- Giải mã thử "10110"

Mã khối và mã tức thời

- ***Mã khối (block code):***
  - Các từ mã khác nhau từng đôi một có cùng độ dài
  - Dễ giải mã
  - Dùng cho **mã tự sửa lỗi (error-correcting code)**.
- ***Mã tức thời (instantaneous code):***
  - Mỗi từ mã không là ***tiền tố (prefix)*** của các từ mã khác.
  - Độ dài mỗi từ mã có thể khác nhau.
  - Dùng cho **mã nén (compressing code)**.

Ví dụ mã tức thời – mã Morse

A	·—	N	—·
B	—···	O	———
C	—·—·	P	·——·
D	—··	Q	——·—
E	·	R	·—·
F	··—·	S	···
G	——·	T	—
H	····	U	··—
I	··	V	···—
J	·———	W	·——
K	—·—	X	—··—
L	·—··	Y	—·——
M	——	Z	——··

Figure 2: Morse code

Ví dụ mã khối

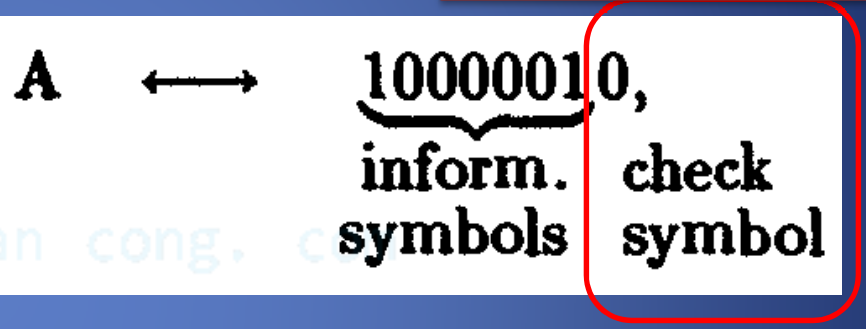
- Mã octal

0	000	4	100
1	001	5	101
2	010	6	110
3	011	7	111

Mã ASCII
(American Standard Code for Information Interchange)

- **$2^7 = 128$ ký tự nguồn:**
  - {A, B, …, Z}
  - {a, b, …, z}
  - {0, 1, …, 9}
  - {@, %, (, <, …}
- **Từ mã nhị phân độ dài 8:**
  - 7 bit mang thông tin
  - 1 bit kiểm tra chẵn lẻ (parity check).

Parity check



- 0: even number of 1's.
- 1: odd number of 1's.

Source Symbol	Code	Source Symbol	Code	Source Symbol	Code	Source Symbol	Code
@	$1(00)_8$	‘	$1(40)_8$	NUL	$0(00)_8$	SP	$0(40)_8$
A	$1(01)_8$	a	$1(41)_8$	SOH	$0(01)_8$	!	$0(41)_8$
B	$1(02)_8$	b	$1(42)_8$	STX	$0(02)_8$	"	$0(42)_8$
C	$1(03)_8$	c	$1(43)_8$	ETX	$0(03)_8$	#	$0(43)_8$
[	$1(33)_8$	{	$1(73)_8$	ESC	$0(33)_8$	;	$0(73)_8$
\	$1(34)_8$	|	$1(74)_8$	FS	$0(34)_8$	<	$0(74)_8$
]	$1(35)_8$	}	$1(75)_8$	GS	$0(35)_8$	=	$0(75)_8$
^	$1(36)_8$	~	$1(76)_8$	RS	$0(36)_8$	>	$0(76)_8$
_	$1(37)_8$	DEL	$1(77)_8$	US	$0(37)_8$	?	$0(77)_8$

Figure 4: ASCII code (7 information bits)

- Ghi chú: $(01)_8 = 000001.$

Mã ISBN (international standard book number)

- Bảng ký tự mã = {0, 1, …, 9, X}
- **Từ mã độ dài 10** (kiểu mới có độ dài 13)
- **Nội dung:**
  - Mã quốc gia/mã ngôn ngữ
  - Nhà xuất bản
  - Mã ấn phẩm
  - Check number

$$\sum_{i=1}^{10} i a_{11-i} = 10a_1 + 9a_2 + 8a_3 + \cdots + 2a_9 + a_{10}$$

chia hết cho 11

Short break

Xây dựng mã tức thời

Bài toán: xây dựng bộ mã tức thời cho các ký tự nguồn.

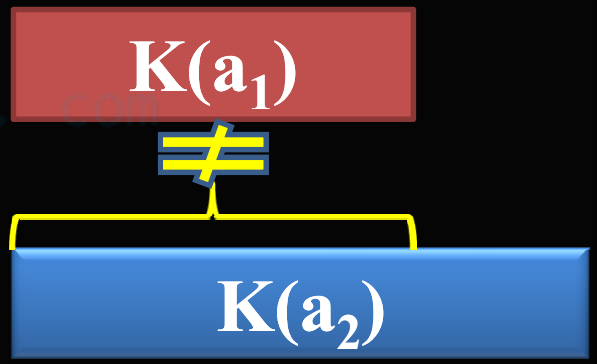
- ***Yêu cầu***: độ dài các từ mã bằng các giá trị d_1, d_2, …, d_n nguyên dương cho trước.

Ví dụ: mã hoá nhị phân các ký tự {0, 1, 2, 3} bằng mã tức thời với độ dài các từ mã là 1, 2, 3, 3.

$$0 \longmapsto 0 \qquad 1 \longmapsto 10 \qquad 2 \longmapsto 110 \qquad 3 \longmapsto 111.$$

Cách xây dựng mã tức thời

- Giả sử $d_1 \leq d_2 \leq \ldots \leq d_n$.
- Chọn một từ mã $K(a_1)$ nào đó có độ dài d_1.
- Chọn một từ mã $K(a_2)$ có độ dài d_2 thoả không chứa $K(a_1)$ là tiền tố.



- Tiếp tục quá trình trên cho $d_3, \ldots, d_n$.

Khả thi không?

- Tổng số từ có độ dài d_2 chứa $K(a_1)$ làm tiền tố:

$$2^{d2-d1}.$$

- Mà $2^{d2} \geq 2^{d2-d1} + 1$

→ Tồn tại ít nhất một từ mã $K(a_2)$ thoả yêu cầu.

- **$K(a_3)$?**
  - không chứa $K(a_1)$ hay $K(a_2)$ làm tiền tố.
  - Tồn tại $K(a_3)$ nếu

$$2^{d3} \geq 2^{d3-d1} + 2^{d3-d2} + 1.$$

  - Chia hai vế cho 2^{d3}:

$$1 \geq 2^{-d1} + 2^{-d2} + 2^{-d3}.$$

→Tổng quát?

Bất đẳng thức Kraft

Định lý: *Với bảng ký tự nguồn có* ***n*** *ký tự, bảng ký tự mã có* ***k*** *ký tự. Để xây dựng được mã tức thời với các độ dài mã* $d_1, d_2, \ldots, d_n$ *cho trước, bất đẳng thức Kraft sau phải thoả:*

$$k^{-d1} + k^{-d2} + \ldots + k^{-dn} \leq 1.$$

Ví dụ: Mã hoá nhị phân {0, 1, 2, 3} với các độ dài mã 1, 2, 3, 3 là khả thi vì:

$$2^{-1} + 2^{-2} + 2^{-3} + 2^{-3} \leq 1.$$

Chẳng hạn $0 \longmapsto 0 \quad 1 \longmapsto 10 \quad 2 \longmapsto 110 \quad 3 \longmapsto 111.$

- BĐT Kraft kiểm tra sự tồn tại hay không của một mã tức thời thoả các điều kiện độ dài từ mã cho trước.

- ẩ gược lại, cho trước một mã giải mã duy nhất bất kỳ, các độ dài từ mã có thoả BĐT Kraft không?

Định lý McMillan

Định lý: *Mọi mã giải mã duy nhất đều thoả bất đẳng thức Kraft*:

$$k^{-d1} + k^{-d2} + \ldots + k^{-dn} \leq 1.$$

Ví dụ:

- ***Mã khối***
  - *Mã 2-out-of-5*: n = 10, k = 2, $d_i = 5\ \forall i$.

    $$10(2^{-5}) \leq 1\ !$$
- ***Mã tức thời***
  - $0 \longmapsto 0 \quad 1 \longmapsto 10 \quad 2 \longmapsto 110 \quad 3 \longmapsto 111.$

Tóm tắt

- LT thông tin nghiên cứu cách:
  - Mã hoá thông tin có thể tự sửa lỗi nhiễu
  - ả én dữ liệu
- Khái niệm cơ bản:
  - Ký tự, bảng ký tự
  - Từ mã
  - Giải mã duy nhất
  - Mã khối
  - Mã tức thời
  - BĐT Kraft

Homework

- [1] Jiri Adamek, ***Foundations of Coding***
- Đọc lại chương 1 [1] và làm các bài tập cuối chương.
- Đọc trước chương 2 [1] Mã Huffman.
- Đăng ký thành viên trên web môn học
- Đăng ký nhóm tối đa 3SV/nhóm.

Bài tập 1

- ả ếu dùng mã khối thì độ dài mã ngắn nhất có thể dùng là bao nhiêu để mã hoá bảng ký tự nguồn {A, B, …, Z} bằng bảng ký tự mã {•, —, ‘ ’} giống mã Morse.

Bài tập 2

1	...	01
2	...	011
3	...	10
4	...	1000
5	...	1100
6	...	0111

A	...	1010
B	...	001
C	...	101
D	...	0001
E	...	1101
F	...	1011

a) Xét tính giải mã duy nhất của các mã trong hai hình bên? Chỉ ra một đoạn mã làm phản ví dụ cho trường hợp giải mã không duy nhất.

b) Chúng có là mã tức thời không? Giải thích.

c) Thay thế chúng bằng các mã tức thời khác có cùng các độ dài mã.

Bài tập 3

- Dùng bất đẳng thức Kraft để xét tính giải mã duy nhất của các mã sau

A	...	001	A	...	00
B	...	1001	B	...	10
C	...	0010	C	...	011
D	...	1110	D	...	101
E	...	1010	E	...	111
F	...	01110	F	...	110
G	...	0101	G	...	010

Bài tập 4

- Xây dựng mã nhị phân tức thời cho bảng ký tự nguồn sau với độ dài mã tương ứng

Symbol	A	B	C	D	E	F	G	H	I	J	K	L
Length	2	4	7	7	3	4	7	7	3	4	7	7

Bài tập 5

- Để mã hoá bảng ký tự nguồn sau với độ dài mã tương ứng, cần ít nhất bao nhiêu ký tự mã.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	2	2	2	1	2	2	2	1	2	2	2	2	2	1	2