

PTS

CODAREA SURSELOR PENTRU CANALE FĂRĂ PERTURBAȚII

Mihai Ivanovici

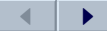
Universitatea Transilvania din Brașov



Titlul

Codarea
Coduri unic decodabile
Lungimea medie a u...
Capacitatea,...

Page 1 of 26



Full Screen

Search

Close

PTS 2008

1 Codarea

În general, alfabetul sursei diferă de alfabetul canalului

Primul scop al codării: trecerea de la alfabetul sursei la cel al canalului

Al doilea scop al codării: transinformația să fie maximă (eficiența maximă)

Adaptarea statistică a sursei la canal: sursa secundară (de la ieșirea din blocul de codare) să genereze simbolurile cu acele probabilități care asigură valoarea maximă a transinformației



Titlul

Codarea
Coduri unic decodabile
Lungimea medie a u...
Capacitatea,...

Page 2 of 26



Full Screen

Search

Close

PTS 2008

În cazul canalelor fără perturbații, capacitatea canalului este atinsă atunci când sursa secundară este una de entropie maximă:

$$C = \max[H(X)] = \log D$$

unde D este numărul de simboluri din alfabetul canalului (respectiv al codului)

Obiectivul codării surselor este de a *transforma o sursă dată, cu un set de probabilități determinat, într-o sursă de entropie maximă și deci anularea redundanței sursei*

Vom considera în continuare că sursele au debit controlabil și că nu au memorie



Titlul

Codarea

Coduri unic decodabile

Lungimea medie a u...

Capacitatea,...

Page 3 of 26



Full Screen

Search

Close

PTS 2008

2 Coduri unic decodabile

Fie o sursă discretă, fără memorie, care generează simboluri din alfabetul:

$$[S] = [s_1 s_2 \dots s_N]$$

având probabilitățile:

$$[P] = [p(s_1) p(s_2) \dots p(s_N)]$$

Fie $[X]$ alfabetul finit al codului / al canalului:

$$[X] = [x_1 x_2 \dots x_D]$$



Titlul

Codarea

Coduri unic decodabile

Lungimea medie a u...

Capacitatea,...

Page 4 of 26



Full Screen

Search

Close

PTS 2008

Cu aceste simboluri se formează un număr N de cuvinte de cod:

$$[C] = [c_1 c_2 \dots c_N]$$

Cuvintele de cod sunt succesiuni finite de litere din alfabetul $[X]$

Codarea reprezintă stabilirea unei corespondențe biunivoce între simbolurile $s_k \in S$ și cuvintele $c_k \in C$

Totalitatea cuvintelor c_k formează un cod

- **Cuvinte cu sens** - cuvinte cărora le corespund simboluri din alfabetul sursei



Titlul

Codarea

Coduri unic decodabile

Lungimea medie a u...

Capacitatea,...

Page 5 of 26



Full Screen

Search

Close

PTS 2008

- **Cuvinte fără sens** - cu alfabeul $[X]$ pot fi formate cuvinte cărora să nu le corespundă simboluri din alfabetul sursei

Cod unic decodabil: atunci când fiecărei succesiuni de cuvinte de cod sa-i corespundă o singură succesiune de litere/cuvinte ale sursei

Un cod unic decodabil fără semne de demarcare între cuvinte se numește *separabil*



Titlul

Codarea

Coduri unic decodabile

Lungimea medie a u...

Capacitatea,...

Page 6 of 26



Full Screen

Search

Close

PTS 2008

Mesaje s_k	Codul A	Codul B	Codul C	Codul D
s_1	00	0	0	0
s_2	01	10	01	10
s_3	10	110	011	110
s_4	11	1110	0111	111

- Codul A - cuvintele de cod au aceeași lungime; este cel mai simplu cod unic decodabil
- Codul B - simbolul "0" indică sfârșitul fiecărui cuvânt
- Codul C - simbolul "0" indică începutul fiecărui cuvânt



Titlul

Codarea

Coduri unic decodabile

Lungimea medie a u...

Capacitatea,...

Page 7 of 26



Full Screen

Search

Close

PTS 2008

Coduri instantanee

Pentru codul B: pe măsură ce se recepționează succesiunea de litere din alfabetul codului, se pot determina cuvintele codului fără referință la literele următoare:

0	10	110	0	1110	0
s_1	s_2	s_3	s_1	s_4	s_1

Cod instantaneu - prin adăugarea unor litere la un cuvânt de cod nu se poate obține un nou cuvânt

Codul C nu este instantaneu - pentru decodare este necesar ca după recepționarea unui cuvânt din vocabularul codului să se recepționeze și câteva litere care urmează pt. a putea decide care cuvânt a fost transmis



Titlul

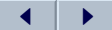
Codarea

Coduri unic decodabile

Lungimea medie a u...

Capacitatea,...

Page 8 of 26



Full Screen

Search

Close

PTS 2008

Definiție: Fie $c_i = [x_{i_1}x_{i_2}\dots x_{i_m}]$ un cuvânt din vocabularul unui cod. Șirul de litere $x_{i_1}x_{i_2}\dots x_{i_k}$, unde $k < m$, se numește **prefixul** cuvântului c_i

Condiția necesară și suficientă ca un cod să fie instantaneu este ca nici un cânt al codului să nu fie un prefix al unui alt cuvânt de cod

Un cod instantaneu are proprietatea de prefix (este *ireductibil*)



Titlul

Codarea

Coduri unic decodabile

Lungimea medie a u...

Capacitatea,...

Page 9 of 26



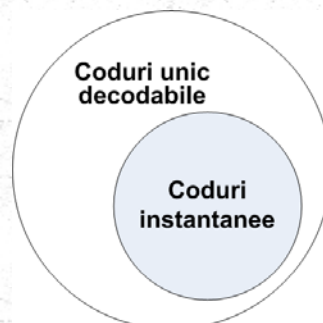
Full Screen

Search

Close

PTS 2008

Un cod instantaneu este unic decodabil (reciproca nu este întotdeauna valabilă)



Titlul

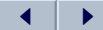
Codarea

Coduri unic decodabile

Lungimea medie a u...

Capacitatea,...

Page 10 of 26



Full Screen

Search

Close

PTS 2008

Algoritm - cod binar instantaneu

Pasul 1°. Mulțimea $[S]$ a mesajelor sursei se împarte în două submulțimi S_0 și S_1 astfel:

$$\underbrace{S_1, S_2, \dots, S_k}_{S_0}, \underbrace{S_{k+1}, \dots, S_N}_{S_1}$$

Se atribuie litera "0" tuturor mesajelor din S_0 și litera "1" tuturor mesajelor din S_1

Pasul 2°. Se continuă aceeași operație pentru S_0 și S_1 , divizându-le în S_{00} și S_{01} , respectiv S_{10} și S_{11} . Se atribuie fiecărei mulțimi litera 0 sau 1, a.î. mesajele din S_{00} vor avea la începutul cuvântului 00 iar cele din S_{01} vor avea 01, ș.a.m.d.

Pasul 3°. Operația de mai sus se continuă până când în



Titlul

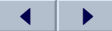
Codarea

Coduri unic decodabile

Lungimea medie a u...

Capacitatea, ...

Page 11 of 26



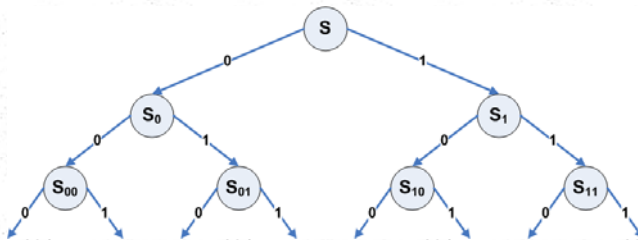
Full Screen

Search

Close

PTS 2008

fiecare submulțime rămâne un singur element s_j căruia îi va corespunde un cuvânt de cod



Codul astfel generat este *ireductibil* - nici un cuvânt al codului nu este prefixul altui cuvânt de cod



Titlul

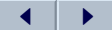
Codarea

Coduri unic decodabile

Lungimea medie a u...

Capacitatea, ...

Page 12 of 26



Full Screen

Search

Close

PTS 2008

3 Lungimea medie a unui cuvânt de cod

Prin codare urmărim mărirea eficienței transmisiunii informației. Mărirea eficienței se referă în general la minimizarea unei anumite *funcții de cost*

Cea mai simplă funcție de cost - fiecărui cuvânt c_i i se asociază un anumit coeficient de cost t_i . De exemplu: durata cuvântului c_i , dacă prețul exploataării unui sistem de transmisiune este aproximativ liniar crescător cu timpul

Costul mediu pe mesaj:

$$\bar{C} = \sum_{i=1}^N t_i p(c_i) = \sum_{i=1}^N t_i p(s_i)$$



Titlul
Codarea
Coduri unic decodabile
Lungimea medie a u...
Capacitatea, ...

Page 13 of 26



Full Screen

Search

Close

PTS 2008

Cea mai eficientă transmisiune - cea care minimizează costul mediu \bar{C}

Dacă toate literele x_i din alfabetul $[X]$ al codului au aceeași durată τ (considerată egală cu unitatea $\tau = 1$), atunci:

$$t_i = l_i \tau = l_i$$

adică durata unui cuvânt este egală cu numărul de litere care formează acel cuvânt

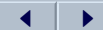
Costul mediu devine:

$$\bar{C} = \sum_{i=1}^N p(s_i) l_i = \bar{l}$$



Titlul
Codarea
Coduri unic decodabile
Lungimea medie a u...
Capacitatea, ...

Page 14 of 26



Full Screen

Search

Close

PTS 2008

Costul mediu este egal cu lungimea medie a unui cuvânt

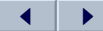
Mărirea eficienței transmisiunii se obține atribuind în mod convenabil fiecărui mesaj s_i dat de sursă un cuvânt de cod c_i în care numărul literelor să fie astfel ales încât lungimea medie \bar{l} a unui cuvânt de să fie cât mai mică

Eficiența transmisiunii poate fi definită numai dacă se cunoaște marginea inferioară a lui \bar{l}



Titlul
Codarea
Coduri unic decodabile
Lungimea medie a u...
Capacitatea, ...

Page 15 of 26



Full Screen

Search

Close

PTS 2008

Limita inferioară a lungimii medii a unui cuvânt de cod

Fie o sursă caracterizată de mulțimea mesajelor:

$$[S] = [s_1 s_2 \dots s_N]$$

având probabilitățile:

$$[P] = [p(s_1) p(s_2) \dots p(s_N)]$$

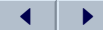
Fie cuvintele de cod:

$$[C] = [c_1 c_2 \dots c_N]$$



Titlul
Codarea
Coduri unic decodabile
Lungimea medie a u...
Capacitatea, ...

Page 16 of 26



Full Screen

Search

Close

PTS 2008

care apar cu aceleași probabilități ca mesajele sursei:

$$[P_C] = [P] = [p_1 p_2 \dots p_N]$$

Lungimile cuvintelor de cod sunt:

$$[L] = [l_1 l_2 \dots l_N]$$

unde l_i este numărul de litere din alfabetul codului care compun cuvântul c_i

Alfabetul codului este:

$$[X] = [x_1 x_2 \dots x_D]$$

Entropia sursei este:



Titlul
Codarea
Coduri unic decodabile
Lungimea medie a u...
Capacitatea, ...

Page 17 of 26



Full Screen

Search

Close

PTS 2008

$$H(S) = H(C) = - \sum_{i=1}^N p(s_i) \log p(s_i)$$

unde $H(C)$ este entropia cuvintelor codului $[C]$

Entropia alfabetului codului $[X]$ este:

$$H(X) = - \sum_{i=1}^D p(x_i) \log p(x_i)$$

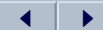
Informația medie pe un cuvânt de cod este dată de produsul dintre numărul mediu de litere \bar{l} și informația medie pe literă $H(X)$:

$$H(S) = H(C) = \bar{l}H(X)$$



Titlul
Codarea
Coduri unic decodabile
Lungimea medie a u...
Capacitatea, ...

Page 18 of 26



Full Screen

Search

Close

PTS 2008

Valoarea maximă a entropiei $H(X)$ se obține atunci când probabilitățile $p(x_i)$ sunt egale:

$$p(x_1) = p(x_2) = \dots = p(x_D) = \frac{1}{D}$$

deci în cazul general:

$$H(X) \leq \log D$$

Rezultă că:

$$H(S) = H(C) = \bar{l}H(X) \leq \bar{l}\log D$$

de unde:



Titlul
Codarea
Coduri unic decodabile
Lungimea medie a u...
Capacitatea, ...

Page 19 of 26



Full Screen

Search

Close

PTS 2008

$$\bar{l} \geq \frac{H(S)}{\log D} = \bar{l}_{min}$$

Lungimea medie \bar{l} a unui cuvânt de cod are o margine inferioară egală cu entropia sursei raportată la valoarea maximă a entropiei alfabetului codului

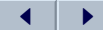
Informația medie pe o literă din alfabetul codului $\frac{H(S)}{\bar{l}}$ nu poate fi mai mare decât valoarea maximă a entropiei alfabetului codului $\log D$

$$\frac{H(S)}{\bar{l}} \leq \log D$$



Titlul
Codarea
Coduri unic decodabile
Lungimea medie a u...
Capacitatea, ...

Page 20 of 26



Full Screen

Search

Close

PTS 2008

4 Capacitatea, eficiența și redundanța codului

Capacitatea codului este valoarea maximă a entropiei alfabetului codului

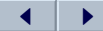
$$C = \max[H(X)] = \log D$$

Eficiența codului este raportul dintre lungimea medie minimă și lungimea medie a unui cuvânt de cod



Titlul
Codarea
Coduri unic decodabile
Lungimea medie a u...
Capacitatea,...

Page 21 of 26



Full Screen

Search

Close

PTS 2008

$$\eta = \frac{\bar{l}_{min}}{\bar{l}}$$

Dar marginea inferioară a lungimii unui cuvânt de cod este:

$$\bar{l}_{min} = \frac{H(S)}{\log D} = \frac{H(S)}{C}$$

iar lungimea medie a unui cuvânt de cod este:

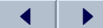
$$\bar{l} = \frac{H(S)}{H(X)}$$

Rezultă că:



Titlul
Codarea
Coduri unic decodabile
Lungimea medie a u...
Capacitatea,...

Page 22 of 26



Full Screen

Search

Close

PTS 2008

$$\eta = \frac{H(S)}{\bar{l} \log D} = \frac{H(X)}{\log D}$$

Redundanța codului este mărimea complementară eficienței

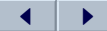
$$\rho = 1 - \eta = \frac{\bar{l} \log D - H(S)}{\bar{l} \log D}$$

$$\rho = \frac{\log D - H(X)}{\log D}$$



Titlul
Codarea
Coduri unic decodabile
Lungimea medie a u...
Capacitatea, ...

Page 23 of 26



Full Screen

Search

Close

PTS 2008

Exemplu

Se consideră o sursă cu alfabetul: $[S] = [s_1 s_2 s_3 s_4]$

și probabilitățile: $[P] = [\frac{1}{2} \frac{1}{4} \frac{1}{8} \frac{1}{8}]$

$$\begin{aligned} \text{Entropia sursei este: } H(S) &= - \sum_{i=1}^4 p(s_i) \log p(s_i) = \\ &= -\frac{1}{2} \log \frac{1}{2} - \frac{1}{4} \log \frac{1}{4} - \frac{1}{8} \log \frac{1}{8} - \frac{1}{8} \log \frac{1}{8} = \frac{7}{4} \text{ bit/simbol} \end{aligned}$$

Se consideră alfabetul codului format din două litere:

$[X] = [01]$

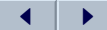
Se presupune că se face următoarea codare:

s_1	s_2	s_3	s_4
00	01	10	11



Titlul
Codarea
Coduri unic decodabile
Lungimea medie a u...
Capacitatea, ...

Page 24 of 26

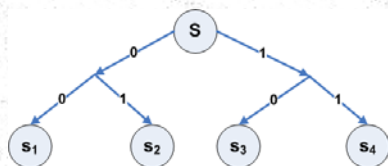


Full Screen

Search

Close

PTS 2008



Lungimea medie a cuvintelor este: $\bar{l} = \sum_{i=1}^4 p_i l_i = 2$

Eficiența este: $\eta = \frac{7/4}{2 \log 2} = \frac{7}{8} = 0.875$

Redundanța este: $\rho = 1 - \frac{7}{8} = \frac{1}{8} = 0.125$

Dacă se face următoarea codare:

s_1	s_2	s_3	s_4
0	10	110	111



Titlul
Codarea
Coduri unic decodabile
Lungimea medie a u...
Capacitatea,...

Page 25 of 26

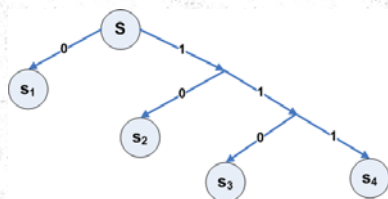


Full Screen

Search

Close

PTS 2008



În acest caz, lungimea medie va fi: $\bar{l} = \sum_{i=1}^4 p_i l_i = 1.75$

Eficiența este: $\eta = \frac{7/4}{1.75 \log 2} = 1$

Redundanța este: $\rho = 1 - 1 = 0$



Titlul
Codarea
Coduri unic decodabile
Lungimea medie a u...
Capacitatea,...

Page 26 of 26



Full Screen

Search

Close

PTS 2008