DEPARTMENT OF
NETWORKED SYSTEMS
AND SERVICES

Hibakontroll kódolás

VIHIBB01 – Coding and IT Security, 2020

István Vajda

CrySyS Lab, BME
vajda@crysys.hu

www.crysys.hu

MŰEGYETEM 1782

Tartalom

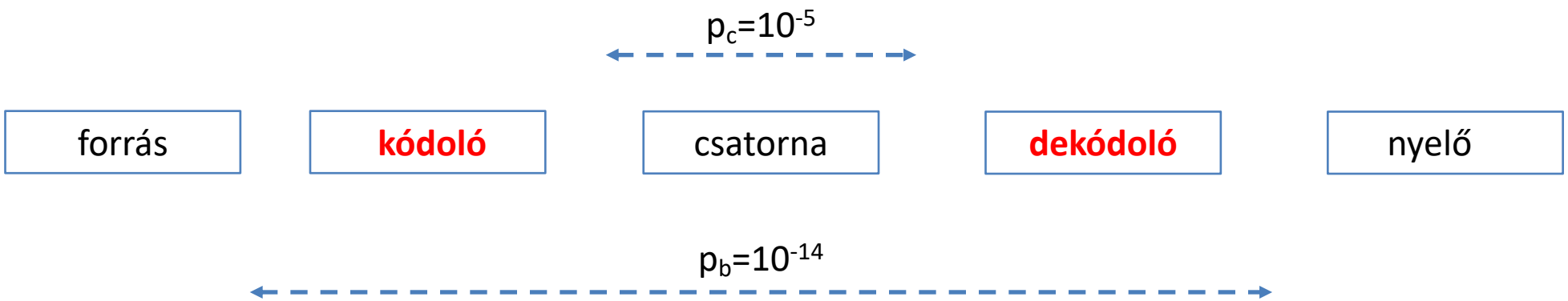
- Probléma: megbízható átvitel zajos csatornán, alapkoncepció
- Alkalmazási területek
- Alapfogalmak, definíciók
- Lineáris kódok
- Hibadetekciós kódok
- Hibajavító blokk kódok
  - Hamming kód
  - Reed-Solomon kód

Probléma: megbízható átvitel zajos csatornán

A probléma:

Csatorna bithibaarány: $p_c=10^{-5}$

A megkövetelt üzenet bithibaarány a forrás és nyelő között: $p_b=10^{-14}$



Alapötlet:

Kódolásként ismételjünk meg n-szer minden üzenetet és dekódolásként alkalmazzunk többségi döntést.

Probléma: megbízható átvitel zajos csatornán

Konstrukció: (ismétléses kód)

Legyen pl. a kódszóhossz n=5.

Kódolás: 0 üzenetbitet ***00000*** kódszóba kódoljuk

1 üzenetbitet ***11111*** kódszóba kódoljuk (itt ***0*** és ***1*** csatornabiteket jelöl)

Dekódoló: 0 az output, ha a vett szóban az ***1*** bitek darabszáma ≤2, egyébként 1 az output

Analízis:

Prob(the number of errors > 2)= $\binom{5}{3}p^3(1-p)^2+\binom{5}{4}p^4(1-p)+\binom{5}{5}p^5 \approx \binom{5}{3}p^3$=10$\cdot\ 10^{-15} = 10^{-14}$

Ezen látványos javulás a hibaarányban költséges: az üzenetbit-sebesség (R bits/sec) csak töredéke a csatornabit-sebességnek (C bits/sec). (sávszélesség-pazarlás: R/C=1/5)

Szerencsére, *hatékonyabb kódolás*sal javítani tudjuk az *R/C arányt ,* megtartva a hibaarány javulást.

Alkalmazási területek

Internet:

- Ethernet keretek a CRC-32 hibadetekciót alkalmaznak,
- IPv4 fejléc (header) hibadetekcióval védi a fejléc tartalmát,
- TCP hibadetekcióval védi a payload-ot and TCP and IP fejlécekbeli címzési információt

Adattárak:

- optikai tárak (CD, DVD),
- hard drive-ok (RAID technika)

Hibajavító memóriák:

- mission-critical alkalmazások esetén

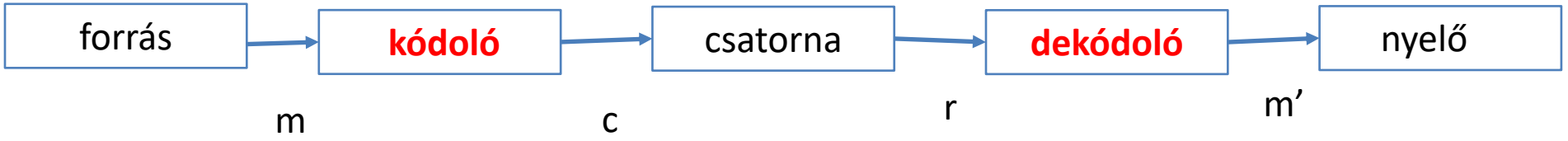
Satellite broadcasting (DVB):

- televízió és IP adat továbbítás

Deep space communication

Fogalmak, definíciók

Az átvitel blokksémája:



Forrás abc: F

- bináris eset: F={0,1} ; nem-bináris eset: F={0,1,…,q-1}

Üzenet: m, $m = (m_0, \dots, m_{k-1}), m_i \epsilon F$

Csatorna abc: Q

bináris, nem-bináris

Kódszó: c, $c = (c_0, \dots, c_{n-1}), c_i \epsilon Q$

Kód: C

A kódszavak halmaza, pl. C={00000, 11111} az ismétléses kód esetén

Kódoló: m üzenet inputra, c kódszó outputot ad

Vett szó: r

Dekódolt üzenet: m‘

Fogalmak, definíciók

Hamming távolság: dist$((a_0, \dots, a_{j-1}), (b_0, \dots, b_{j-1}))$

azon pozíciók száma, amelyekben ***a*** és ***b*** szavak különböznek, ahol ***a*** és ***b*** azonos hosszúságúak és azonos abc felettiek

pl. dist(10111, 01101)=3

Hamming súly: w$((b_0, \dots, b_{n-1}))$

a nemzérus karakterek darabszáma

pl. w(10111)=4

Minimális Hamming távolság: d_{min}

minimális Hamming távolság a kódszópárok halmaza felett

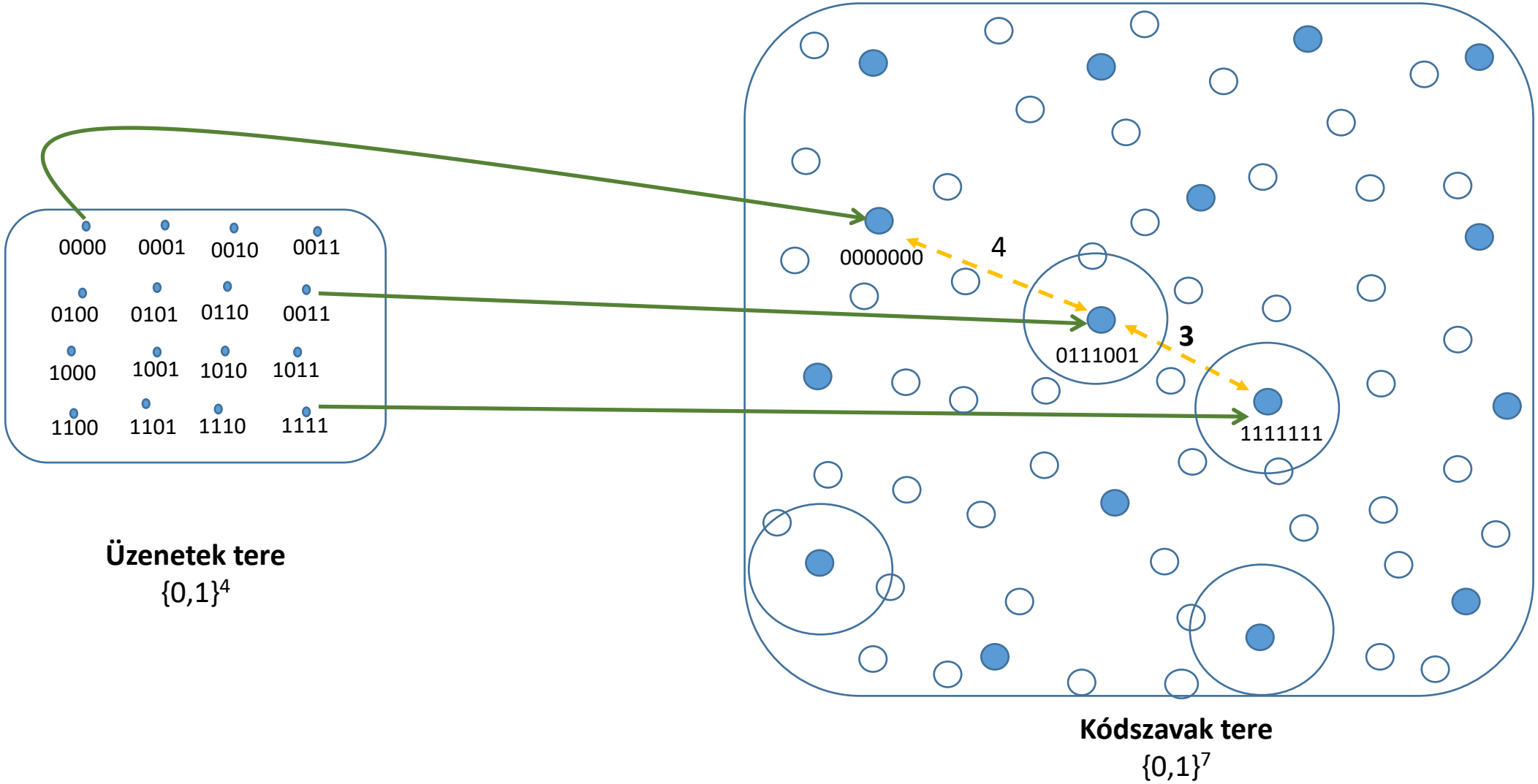
pl. d_{min} =dist(00000,11111)=5 az ismétléses kód esetén

Kódparaméterek: C(n,k,d_{min}) ; C(kódszóhossz, üzenethossz, minimális Hamming távolság)

pl. C(5,1,5) az ismétléses kód esetén

Fogalmak, definíciók

C(7,4,3) kódolás illusztráció:



Üzenetek tere
$\{0,1\}^4$

Kódszavak tere
$\{0,1\}^7$

Fogalmak, definíciók

Hibajavítás: célja a kódszó rekonstruálása a vett szóból

Hibajavítás algoritmusa: r vett szó inputra c' output az r-hez (Hamming-távolságban) legközelebbi kódszó

Hibadetekció: célja átviteli hiba detektálása

Javítás (detekció) hibavalószínűsége: P_e

annak valószínűsége hogy sikertelen a javítás (detekció)

A kód javító (detekciós) képessége $C(n,k,d_{min})$:

Tétel 1: Egy $C(n,k,d_{min})$ kód $t_{corr} = \lfloor d_{min}/2 \rfloor$ darab hibát képes sikeresen javítani a vett szóban.

Magyarázat: Tekintsük az összes lehetséges vett szó halmazát a kódszavak körül ≤ t_{corr} hiba esetén. Vegyük észre, hogy ezek a halmazok (Hamming-gömbök) nem metsződnek. Következésképp, a vett szóhoz legközelebbi szó az átküldött kódszó lesz.

Tétel 2: Egy $C(n,k,d_{min})$ kód $t_{det} = d_{min} - 1$ darab hibát képes sikeresen detektálni a vett szóban.

Magyarázat: ≤ t_{det} hiba esetén a vett szó nem eshet egybe egy, az átküldött kódszótól különböző kódszóval

Példa: C(n,1,n) ismétléses kód esetén, ahol n páratlan szám, t_{corr}=(n-1)/2 , t_{det} =n-1.

Linear codes

Lineáris kód: Code C(n,k,d_{min}) lináris ha a kódszavak tetszőleges lineáris kombinációja is kódszó.
(Más szavakkal C kód a Q^n tér lineáris altere.)

Generátor mátrix: G(k,n) mátrix k sorában k lineárisan független n hosszú kódszó áll.

Kódgenerálás: c=mG

$$(c_0, \dots, c_{n-1}) = (m_0, \dots, m_{k-1}) \begin{pmatrix} g_{0,0}, & \cdots & , g_{0,n-1} \\ & \cdots & \\ g_{k-1,0}, & \cdots & , g_{k-1,n-1} \end{pmatrix}$$

Paritásellenőrző mátrix: H(n-k,n) mátrix n-k sorában álló n hosszúságú szavak, ortogonálisak G mátrix soraira. A H mátrixot a dekódolás algoritmusa használja.

Példa: (110101) és (011100) bináris vektorok ortogonálisak, mivel skalárszorzatuk 0:

1·0+1·1+0·1+1·1+0·0+1·0=0+1+0+1+0+0=0

Tétel 3: Lineáris kódok esetén $d_{min}=w_{min}$.

Magyarázat: A zérus kódszó és a hozzá (Hamming távolságban) legközelebbi kódszó távolsága d_{min} és ez egyben w_{min}.

Hibadetekció

Az egyszerű paritásellenőrző kód: C(k+1,k,2)

$c=(m_0,m_1,\ldots m_{k-1}, par)$

ahol *par* az üzenetbitek bináris összege.

$d_{min}=2$ (magyarázat: ha egy bitet megváltoztatunk az üzenetben, az üzenet paritása is változik)

CRC (Ciklikus Redundancia Ellenőrző kód):

A bináris szavakat bináris polinomként ábrázoljuk.

Pl. $b=(b_0,b_1,\ldots,b_m) \rightarrow b(x)= b_0+b_1x+\ldots+b_mx^m$.

C(n,k) CRC kódot, az n-k fokszámú g(x) generátorpolinomja a következőképp generálja:

$$\mathbf{c(x)=m(x)\cdot x^{n-k} - ([m(x)\cdot x^{n-k}] \bmod g(x))},$$

ahol $[m(x)\cdot x^{n-k}] \bmod g(x)$ jelöli azon polinomosztás osztási maradékát, amikor $m(x)\cdot x^{n-k}$ polinomot g(x) polinommal osztjuk.

CRC generator polynomials can be found in communication standards.

Pl. CRC-32 generátor polinom kitevő-sora: (0,1,2,4,5,7,8,10,11,12,16,22,23,26,32) (i.e. $g(x)=1+x+x^2+x^4+x^5+\ldots+x^{32}$)

Pl. alkalmazások: Ethernet, SATA, MPEG2,PKZIP, Gzip,…

(7,4,3) Hamming kód

A generátor és a paritásellenőrző polinomok a következők:

$$G=\begin{pmatrix}1&0&0&0&1&1&0\\0&1&0&0&1&0&1\\0&0&1&0&0&1&1\\0&0&0&1&1&1&1\end{pmatrix} \qquad H=\begin{pmatrix}1&1&0&1&1&0&0\\1&0&1&1&0&1&0\\0&1&1&1&0&0&1\end{pmatrix}$$

Kódgenerálás: c=mG (pl. m=1100, c= ***1100011***)

0000 - ***0000000***, 0001 – ***0001111***, 0010 – ***0010011***, 0011 - ***0011100***

0100 – ***0100101***, 0101 – ***0101010***, 0110 – ***0110110***, 0111 – ***0111001***

1000 – ***1000110***, 1001 – ***1001001***, 1010 – ***1010101***, 1011 – ***1011010***

1100 – ***1100011***, 1101 – ***1101100***, 1110 – ***1110000***, 1111 – ***1111111***

A kódszavak Hamming-súlyai: 0,3,4,7 → $w_{min}=3$ → $d_{min}=3$

Szindróma dekódolás: (7,4,3) Hamming kód

Vegyük észre, hogy H mátrix oszlopai különbözők:

$$H=\begin{pmatrix}1&1&0&1&1&0&0\\1&0&1&1&0&1&0\\0&1&1&1&0&0&1\end{pmatrix}$$

Tétel 1 alapján $t_{corr}=1$.

r=c+e, ahol e 1-súlyú hibavektorban az i pozícióban áll 1 ($0\leq i\leq 6$).

Dekódolás:

1) Szindróma számítás: $s=Hr^T$, ahol r^T vektor r vektor transzponáltja
2) Szindróma táblázat használata: keressük meg i indexet s szindrómához; index i definiálja a hibavektor e' becslését
3) Dekódolt kódszó: c'=r-e'
4) Dekódolt üzenet: m' = c' első k bitje (szisztematikus kód)

s	i
001	6
010	5
011	2
100	4
101	1
110	0
111	3

Reed-Solomon kód

F=Q=q, ahol q primszám

mod q aritmetika (mod q összeadás és szorzás)

Legyen α egy primitív elem mod q, azaz ^{0} (=1), α^1, α^2, α^3, α^4, α^{q-2} (mod q) különbözők.

C(n,k), n=q-1 Reed-Solomon (RS) kód generátor mátrixa:

$$G=\begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \alpha & \alpha^2 & \dots & \alpha^{n-1} \\ & & & \dots & \\ 1 & \alpha^{(k-1)} & \alpha^{2(k-1)} & \dots & \alpha^{(n-1)(k-1)} \end{pmatrix}$$

Tétel 4: Az RS-kód minimális távolsága n-k+1.

Reed-Solomon codes

q=7 , F=Q={0,1,2,3,4,5,6}

mod 7 aritmetika

e.g. 4+6=3 (mod 7), -2=5 (mod 7), 3·4=5 (mod 7), 2^4=1 (mod 7), 5^{-1}=3 (mod 7) i.e. 5·3=1 (mod 7)

α=3 primitív elem (ellenőrzés: hatványai 1, 3, 2, 6, 4, 5)

C(6,2,5) RS-kód generátor mátrixa:

$$G=\begin{pmatrix}1 & 1 & 1 & 1 & 1 & 1\\ 1 & 3 & 2 & 6 & 4 & 5\end{pmatrix}$$

Kódszavak száma: $q^k=7^2=49$.

Hibajavító képesség: ≤ 2 (nem-bináris) hiba

Hibadetekciós képesség: ≤ 4 (nem-bináris) hiba.

Ellenőrző kérdések

- Nevezd meg a hibakorlátozó kódolás főbb alkalmazási területeit!
- Mi a hibadetekciós feladat?
- Hogy mérjük a hibajavító kód hatékonyságát?
- Mik a hibakorlátozó kód paraméterei?
- Mik az n bit kódszóhosszú bináris ismétléses kód paraméterei?
- Mi a hibajavító képessége egy 7 minimális távolságú kódnak?
- Lineáris-e a bináris ismétléses kód illetve az egyszerű paritásellenőrző kód?
- Hogyan használjuk a kód generátor mátrixát?
- Mi a szindróma értéke, ha a hibavektor egy kódszóval esik egybe?
- Mikor hívunk egy kódot nem-binárisnak?
- Mi a CRC? Nevezzen meg néhány alkalmazását!