

Samoopravné kódy

Alexandr Kazda

Univerzita Karlova

4. května 2020

Theorem

Nechť $p \in (0, 1/2)$. Mějme kanál s chybovostí p a kapacitou $C(p) = 1 - H(p)$. Nechť $\kappa < C(p)$. Potom pro každé $\varepsilon > 0$ existuje binární kód s hustotou k/n aspoň κ a spolehlivostí $> 1 - \varepsilon$.

- Mám kanál o kapacitě o něco větší než κ
- Vyberu si $\varepsilon > 0$, vypadne mi (n, k, d) kód, že $k/n \geq \kappa$ a spolehlivost je $1 - \varepsilon$
- **Není zaručeno** d velké, ale vysoká spolehlivost

- Zvolím $\alpha > 0$, aby $p + \alpha < 1/2$ (pomocný parametr)
- Zvolím n velké (mj. chceme $e^{-n\alpha^2/2} < \varepsilon/2$)
- Zvolím náhodně kód $C \subset \{0, 1\}^n$ velikosti $2^{\lceil n\kappa \rceil}$
- Náhodně vyberu i , pošlu c_i skrz kanál, dekóduji \tilde{c} na nejbližšího souseda (to je fce D)
- Cíl $P[D(\tilde{c}) = c_i] \geq 1 - \varepsilon$

- Předpokládáme, že kanál udělal $\leq n(p + \alpha)$ chyb
- Pak \tilde{c} a odeslané slovo c_i jsou $n(p + \alpha)$ -blízko
- Chceme: Pst, že \tilde{c} neleží $n(p + \alpha)$ -blízko nějakého $c_j \in C$ ($j \neq i$) je aspoň $1 - \varepsilon/2$
- Máme

$$P[D(\tilde{c}) = c_i | X] \geq 1 - P[\exists j \neq i, d(c_j, \tilde{c}) \leq n(p + \alpha) | c_i + e = \tilde{c}, X]$$

- Fixuji $i, j \neq i$. Pak $\tilde{c} = c_i + e$ a c_j jsou nezávislé rovnoměrně rozdělené veličiny
- Jaká je pst, že c_j leží v kouli se středem \tilde{c} a poloměrem $n(p + \alpha)$?
- Představím si, že nejdřív zvolím \tilde{c} , pak c_j . Máme

$$P[d(c_j, \tilde{c}) \leq n(p + \alpha) | c_i + e = \tilde{c}, X] \leq \frac{V(n, \lfloor n(p + \alpha) \rfloor)}{2^n}$$

- Pravou stranu odhadneme (entropie je na $(0, 1/2)$ rostoucí)

$$\frac{V(n, \lfloor n(p + \alpha) \rfloor)}{2^n} < \frac{2^{nH(\lfloor n(p + \alpha) \rfloor / n)}}{2^n} \leq \frac{2^{nH(n(p + \alpha) / n)}}{2^n} = \frac{2^{nH(p + \alpha)}}{2^n}$$

- Máme

$$P[d(c_j, \tilde{c}) \leq n(p + \alpha) | c_i + e = \tilde{c}, X] < \frac{2^{nH(p+\alpha)}}{2^n} = 2^{-n(1-H(p+\alpha))}$$

- Chceme znát

$$P[\exists j \neq i, d(c_j, \tilde{c}) \leq n(p + \alpha) | c_i + e = \tilde{c}, X]$$

- Kolik je kandidátských $j \neq i$? Přesně $2^k - 1$

$$P[d(c_j, \tilde{c}) \leq n(p + \alpha) | c_i + e = \tilde{c}, X] < 2^{-n(1-H(p+\alpha))}$$

- Uděláme hrubý odhad: Pokud máme $2^k - 1$ stejně pravděpodobných jevů Y_j , tak

$$P\left[\bigcup_j Y_j\right] \leq \sum_j P[Y_j] = (2^k - 1)P[Y_1] \leq 2^k P[Y_1]$$

- První nerovnost je tzv. Booleova nerovnost
- Tedy

$$P[\exists j \neq i, d(c_j, \tilde{c}) \leq n(p + \alpha) | c_i + e = \tilde{c}, X] < 2^{k-n(1-H(p+\alpha))}$$

- Máme

$$P[D(\tilde{c}) = c_i | X] > 1 - 2^{k-n(1-H(p+\alpha))} = 1 - 2^{n(k/n - (1-H(p+\alpha)))}$$

- Značme $\zeta = k/n - (1 - H(p + \alpha))$
- Máme $\kappa < 1 - H(p)$, pravá strana je spojitá fce p
- Zvolme $\alpha > 0$ dost malé, aby $\kappa < 1 - H(p + \alpha)$
- Přitom jsme volili $k = \lceil n\kappa \rceil \leq n\kappa + 1$ pro $\kappa < 1 - H(p + \alpha)$
- Pro dost velké n bude $k < n(1 - H(p + \alpha))$
- Tedy $\zeta = k/n - (1 - H(p + \alpha)) < 0$
- Máme

$$P[D(\tilde{c}) = c_i | X] > 1 - 2^{n\zeta}$$

přitom $\zeta < 0$ nezávisí na n (pro n velké)

- Volme n dost velké, aby $2^{n\zeta} < \varepsilon/2$ a vyhráváme

- Jak víme, že $k/n \geq \kappa$? Volili jsme $|C| = 2^{\lceil n\kappa \rceil}$, tedy $k = \log_2 |C| \geq n\kappa$
- V jakém pořadí volíme n, α ?
- Nejdřív α , potom n dost velké
- Námitka: Co když náhodou $c_i = c_j$ pro $i \neq j$ (narozeninový paradox)?
- To se stane jen vzácně a spolehlivost to ovlivní málo (kódových slov je $2^{\kappa n}$, tak pár špatných nevadí)
- Pokud je nějaký kód spolehlivý a má $c_i = c_j$ stejná, zvolme c_j znova náhodně z $\{0, 1\}^n \setminus C$; spolehlivost nezhoršíme

Theorem

Nechť $p \in (0, 1/2)$. Mějme kanál s chybovostí p a kapacitou $C(p) = 1 - H(p)$. Nechť $\kappa < C(p)$. Potom pro každé $\varepsilon > 0$ existuje binární kód s hustotou k/n aspoň κ a spolehlivostí $> 1 - \varepsilon$.

- Z důkazu je vidět, že spolehlivý kód existuje pro všechna dost velká n , ale nevíme, jak ten kód vyrobit
- Všimněte si, že mluví o kódech velké délky (asymptotický výsledek)
- Pro $p = 0,01$, $\kappa = 0,9$, $\varepsilon = 0,1$ je největší α aby $\kappa < 1 - H(p + \alpha)$ jen asi 0,0029
- Pak potřebuji n asi 1,26 milionu (blok 158 kB) aby $e^{-n\alpha^2/2} \leq \varepsilon/2$
- Co dělat pro malá n ? Zkoušet kódy, co někdo navrhl

Bonus: Spolehlivý kód pro $p = 0,01$ a 99 bitů

- Uvažme Reed-Mullerův kód s parametry $m = 7$, $r = 4$
- Délka $n = 2^7 = 128$
- $k = 1 + \binom{7}{1} + \binom{7}{2} + \binom{7}{3} + \binom{7}{4} = 99$
- Minimální vzdálenost $2^{m-r} = 2^3 = 8$
- Dolní odhad spolehlivosti:

$$(1-p)^{128} + 128(1-p)^{127}p + \binom{128}{2}(1-p)^{126}p^2 + \binom{128}{3}(1-p)^{125}p^3 \doteq 0,9$$

- Na přenos zdrojového slova délky 99 potřebují cca 130 bitů; o cca 45 méně než přes H_3 , o 170 méně než pro opakovací kód