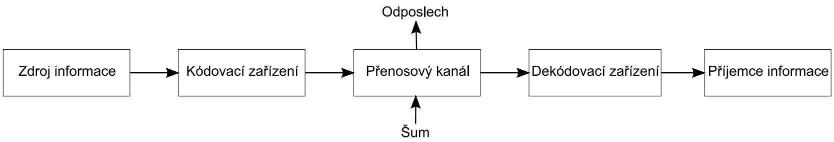
# Teorie kódování

# Kódování bez šumu

# Kódování




## Důvody pro kódování:

- přenosové vlastnosti kanálu
- ekonomičnost (rychlost) přenosu kanálem
- bezpečnostní kódy
- šifrování

# Obecné pojmy

| | |
|---|---|
| **Abeceda** | $\mathbb{T} = \{t_1, t_2, \dots, t_m\}$ |
| **Znak** | $t \in \mathbb{T}$ |
| **Slovo** | $\tau \in \mathbb{T}^n, \tau = t_{i_1} \dots t_{i_n}$ |
| **Délka slova** | $\lvert\tau\rvert = n$ |
| **Velikost abecedy** | $\lvert\mathbb{T}\rvert = m$ |

| | |
|---|---|
| **Prázdné slovo** | $\varepsilon$ |
| **Množina všech slov** | $\mathbb{T}^*$ |
| **Množina všech neprázdných slov** | $\mathbb{T}^+ = \mathbb{T}^* \setminus \{\varepsilon\}$ |
| **Množina všech slov délky $n$** | $\mathbb{T}^n$ |

# Kódování

**Kódování** = předpis, který každému znaku zdrojové abecedy $\mathbb{Z}$ přiřazuje právě jedno kódové slovo z množiny $\mathbb{A}^+$.

$$\mathcal{K} : \mathbb{Z} \to \mathbb{A}^+$$

| | |
|---|---|
| $\mathbb{Z} = \{z_1, \ldots, z_r\}$ | zdrojová abeceda |
| $z$ | zdrojový znak |
| $\zeta = z_{i_1} \ldots z_{i_k}$ | zdrojové slovo |
| $\mathbb{A} = \{a_1, \ldots, a_s\}$ | kódová abeceda |
| $a$ | kódový znak |
| $\alpha = a_{i_1} \ldots a_{i_l}$ | kódové slovo |

# Kód

**Kód** = množina všech kódových slov.

$$K = \mathcal{K}(\mathbb{Z})$$

$$\begin{array}{ccc} z_1 & \rightarrow & \alpha_1 \\ z_2 & \rightarrow & \alpha_2 \\ & \vdots & \\ z_r & \rightarrow & \alpha_r \end{array}$$

**Prostý kód** = kód, ve kterém jsou dvěma různým zdrojovým znakům přiřazena dvě různá kódová slova.

**Kódování zdrojových zpráv** – $\mathcal{K}^+ : \mathbb{Z}^+ \rightarrow \mathbb{A}^+$

$$\zeta = z_{i_1} \ldots z_{i_k} \Rightarrow \mathcal{K}^+(\zeta) = \mathcal{K}(z_{i_1}) \ldots \mathcal{K}(z_{i_k})$$

# Typy kódování

**Jednoznačně dekódovatelné kódování**

Řekneme, že kódování $\mathcal{K}$ je jednoznačně dekódovatelné $\Leftrightarrow \mathcal{K}^+$ je prosté.

**Prefixové kódování**

Řekneme, že kódování $\mathcal{K}$ je prefixové, není-li žádné kódové slovo prefixem (předponou) jiného kódového slova.

**Blokové kódování**

Řekneme, že kódování $\mathcal{K}$ je blokové (s délkou $n$), jestliže mají všechna kódová slova stejnou délku $n$.

$$K \subseteq \mathbb{A}^n$$

**Poznámka:** Každé blokové kódování je prefixové, a tedy jednoznačně dekódovatelné.

# Existence prefixových kódů

**Věta:** Při kódování pomocí $s$ znaků můžeme sestavit prefixový kód s délkami slov $n_1, n_2, \dots, n_r \Leftrightarrow$ platí Kraftova nerovnost

$$s^{-n_1} + s^{-n_2} + \cdots + s^{-n_r} \leq 1$$

**Poznámka:** Kraftova nerovnost zajišťuje pouze existenci prefixového kódování.

**Mc Millanova věta:** Pro každé jednoznačně dekódovatelné kódování platí Kraftova nerovnost.

# Nejkratší kód

$z_1, \ldots, z_r$ – znaky zdrojové abecedy
$p_1, \ldots, p_r$ – pravděpodobnost výskytu znaku $z_i$ ve zprávě

$$p_1 + p_2 + \cdots + p_r = \sum_{i=1}^{r} p_i = 1, \qquad p_i \in \langle 0, 1 \rangle$$

$n_1, \ldots, n_r$ – délky kódových slov
$\overline{n}$ = průměrná délka kódového slova

$$\overline{n} = n_1\, p_1 + n_2\, p_2 + \cdots + n_r\, p_r = \sum_{i=1}^{r} n_i\, p_i$$

**Nejkratší kód**

Nejkratším $s$-znakovým kódováním zdrojové abecedy $z_1, \ldots, z_r$ s pravděpodobnostmi výskytu $p_1, \ldots, p_r$ se rozumí prefixové kódování této abecedy pomocí $s$ znaků, které má nejmenší možnou průměrnou délku slova $\overline{n}$.

# Huffmanova metoda

**Huffmanova metoda** – konstrukce nejkratšího kódu

**A) konstrukce binárního nejkratšího kódu**

1. zdrojové znaky uspořádáme sestupně podle $p_i$
2. sečteme poslední dvě pravděpodobnosti a výsledek zařadíme podle velikosti mezi ostatní pravděpodobnosti = redukce
3. opakujeme 2) dokud nedojdeme k součtu 1
4. posledním dvěma znakům přiřazujeme 0 a 1
5. zpětným postupem přiřazujeme sčítancům kódové znaky 0 a 1

# Huffmanova metoda

## B) konstrukce obecného nejkratšího kódu

$\mathbb{Z} = \{z_1, \ldots, z_r\}$ – zdrojová abeceda
$\mathbb{A} = \{a_1, \ldots, a_s\}$– kódová abeceda

$r \leq s \Rightarrow z_1 \to a_1, \ldots, z_r \to a_r$
$r > s$: výjimečná první redukce:

- sečteme posledních $q$ nejmenších pravděpodobností, $q \in \{2, 3, \ldots, s\}$
- $q$ získáme tak, že v seznamu zdrojových znaků oddělujeme skupiny po $s - 1$ znacích, až zbyde skupina s počtem $q$ znaků

další redukce: po $s$ znacích