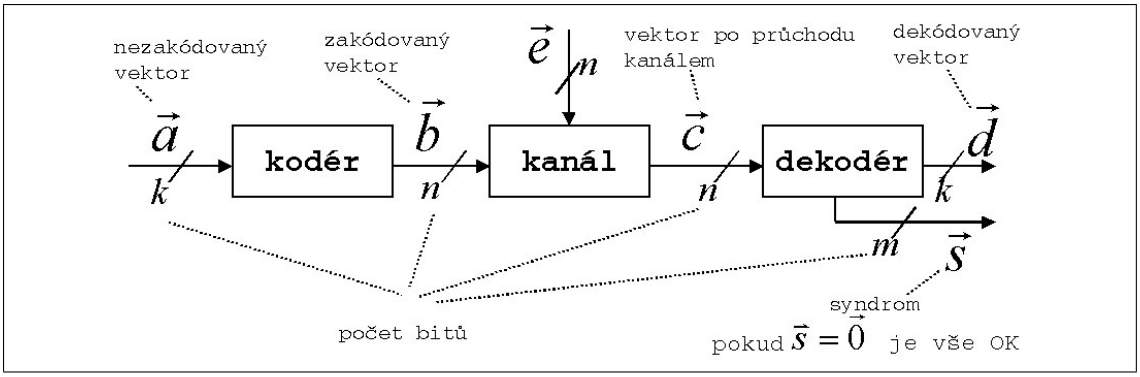
Kapitola 6

Bezpečnostní kódy

6.1 Základní principy a pojmy



Obrázek 6.1: bezpečnostní kód

Bezpečnostní kód - kód, ve kterém je záměrně zavedena tzv. informační redundance při kódování dat tak, aby bylo možné detekovat a případně také opravit nejpravděpodobnější typy chyb.

Bezpečnostní kód tedy slouží pro detekování a opravu chyb.

Typy chyb v informaci

Rozdělení podle pravděpodobnosti chybných změn $0 \rightarrow 1$ a $1 \rightarrow 0$ na chyby symetrické a asymetrické, podle počtu chybných bitů na chyby nezávislé a závislé, podle počtu chybných bitů na chyby jednonásobné, dvojnásobné atp.

Důležitým hlediskem je také to, zda k chybě došlo při přenosu nebo při zpracování informace.

Symetrická chyba - chyba při níž se s přibližně stejnou pravděpodobností změní $0 \rightarrow 1$ nebo $1 \rightarrow 0$

Asymetrická chyba - vlivem poruchy se může změnit jen $0 \rightarrow 1$ a pravděpodobnost změny $1 \rightarrow 0$ je nulová nebo naopak.

Kanál s výmazem - zná se poloha chybného bitu, ale jeho hodnota není známa.

Kanál bez paměti - nezávislost výskytu chyby v určitém bitu na výskytu chyb v ostatních bitech, tj. chyba je nezávislá na ostatních bitech, ve kterých došlo k chybě.

Kanál s pamětí - v zakódované informaci vznikají shluky chyb. *Shluk* je posloupnost bitů přijatého slova, kde krajní bity posloupnosti jsou chybné a v mezilehlých bitech mohlo, ale nemuselo dojít k chybě.

Dvojkový kód - $b_i \in \{0, 1\}$, tj. kódová slova jsou tvořena 0 a 1
Blokový kód (n,k):

- k - informační obsah
- n - délka
- $m = n - k$ - redundance (nadbytečnost), tj. bity navíc
- $\frac{k}{n}$ - norma kódu, přenosová rychlost, proměnná (relativní) míra informace - udává se v %
- $\frac{m}{n}$ - proměnná relativní redundance - udává se v %

Systematický kód - slovo je tvořeno informační částí a kontrolní částí, kde prvních k bitů je stejných jako nezakódovaná informace.

	$\vec{b}$	
$\vec{a}$	$\mathcal{K}_1$	$\mathcal{K}_2$
00	000	00000
01	011	01011
10	101	10101
11	110	11110

Kódy $\mathcal{K}_1$ a $\mathcal{K}_2$ jsou tedy systematické.

Hammingova váha - číslo udávající počet jedniček vektoru $w(\vec{V})$

Hammingova vzdálenost - počet odlišných bitů dvou vektorů. Vektory musí být stejně dlouhé $d(\vec{U}, \vec{V})$. Platí $d(\vec{U}, \vec{V}) = w(\vec{V})$.

Jestliže dojde při přenosu nebo zpracování zakódované informace k chybě, změní se kódový vektor $\vec{V}$ na vektor $\vec{U} = \vec{U} \oplus \vec{E}$, kde $\vec{E}$ je nenulový chybový vektor. Hemmingova váha vektoru $\vec{E}$ je váha chyby. Chyba s vahou $= 1$ jednonásobná, s vahou $= 2$ dvojnásobná atp.

Minimální kódová vzdálenost kódu $\mathcal{K}$ je rovna minimu kódových vzdáleností.

Tj. $kvzd = min\ d(\vec{U}, \vec{V})$.

Detekce chyb: $dch < kvzd$, kde dch je počet detekovaných chyb.
Oprava (korekce): $dch + och < kvzd \Rightarrow och < kvzd - dch$, kde och je počet opravitelných chyb.

6.2 Lineární kódy

Kódová slova lineárního kódu tvoří lineární prostor. Popisujeme je pomocí matice.

$$\vec{b} = \vec{a} \cdot \mathbb{G}$$

Kde $\vec{b}$ je zakódovaný vektor, $\vec{a}$ je nezakódovaný vektor a $\mathbb{G}$ je generovací matice. V této matici jsou řádky lineárně nezávislé.

Kontrolní matice $\mathbb{H}$ $[(n-k)*n]$ obsahuje všechny možné nenulové $(n-k)$ bitové sloupce a každý právě jednou. Kde k je informační délka a n je celková délka. Tj. $\mathbb{H}$ má $(n-k)$ řádek a n sloupců. Počet lineárně nezávislých sloupců matice $\mathbb{H}$ udává počet detekovaných chyb. Platí následující:

$$\mathbb{G} \cdot \mathbb{H}^T = \mathbb{O}$$

Řádky matice $\mathbb{G}$ a $\mathbb{H}$ tvoří bázi lineárního prostoru. *Lineární prostor* je množina všech slov dvojkového blokového kódu a jsou zde definované operace (XOR, AND, součet vektorů, násobení vektoru skalárem, skalární součin). Všechny lineární kombinace vektorů tvoří právě *lineární prostor.*

Syndrom závisí pouze na chybě:

$$\vec{s} = \vec{e} \cdot \mathbb{H}^T$$

Pokud je $\vec{s} = \vec{o}$, pak buď nedošlo k chybě ($\vec{e} = \vec{o}$ nebo má $\vec{e}$ tvar kódového slova.

Pro vektor chyb platí:

$$\vec{e} = \vec{b} \oplus \vec{c}$$

6.3 Cyklické kódy

Cyklické kódy jsou lineární kódy, které mají tu vlastnost, že je-li $\vec{v} = (v_{n-1}, v_{n-2}, \ldots, v_0)$ kódovým vektorem, je kódovým vektorem také vektor $\vec{v} = (v_{n-2}, v_{n-3}, \ldots, v_0, v_{n-1})$. Každý cyklický kód (n, k) obsahuje kódové slovo $g(x)$, které jako mnohočlen má stupeň $n-k$. Tento mnohočlen se nazývá *generující.* Je to jediný kódový mnohočlen stupně $n-k$, ostatní mají stupeň vyšší. *Generující mnohočlen* dělí bezezbytku dvojčlen $X^n - 1$, tj. platí:

$$h(x) = \frac{X^n - 1}{g(x)}$$

Kde $h(x)$ je kontrolní mnohočlen. Kód generovaný mnohočlenem je lineární.

$$G(x) = x^m + g_{m-1} \cdot x^{m-1} + \cdots + g_1 \cdot x + 1$$

m je redundance. Platí následující:

$$\text{detekce shluků chyb délky } l \leq m = \deg G(x)$$

V cyklickém kódu, ve kterém se vyskytují trojčleny, je minimální kódová vzdálenost $d_m in = 3 \Rightarrow$ mohou být opraveny jednonásobné chyby.

Nejběžnější cyklické kódy jsou:

$G(x)$	
$x+1$	parita
$x^3 + x + 1$	Hammingův kód (7,4)
	další Hammingovy kódy
	kódy BCH
	kódy RM

6.4 Srovnání lineárních a cyklických kódů

V následující tabulce jsem se pokusil srovnat jednotlivé vlastnosti lineárních a cyklických kódů. Pokud budete sledovat obrázek 6.1, tak jednotlivým kódům porozumíte. Alespoň já jsem si to ujasnil.

$B(x) \sim \vec{b}$	vyslané slovo
$C(x) \sim \vec{c}$	přijaté slovo
$E(x) \sim \vec{e}$	chybový mnohočlen

Lineární k.	Cyklický k.
$\vec{b} = \vec{a} \cdot \mathbb{H}$ $\mathbb{G} \cdot \mathbb{H}^T = \mathbb{O}$ $\vec{s} = \vec{c} \cdot \mathbb{H}^T$ $\vec{s} = \vec{e} \cdot \mathbb{H}^T$	$B(x) = G(x) \cdot A(x)$ $G(x) = x^m + g_{m-1} \cdot x^{m-1} + \cdots + g_1 \cdot x + 1$ $S(x) = C(x)\%G(x)$ $S(x) = E(x)\%G(x)$
Systematický k.	
$\mathbb{G} = [\mathbb{I}_k, \mathbb{F}]$ $\mathbb{H} = [\mathbb{F}^T, \mathbb{I}_m]$, $\mathbb{I}$ je jednotková matice	$B(x) = A(x) \cdot x^m + A(x) \cdot x^m \% G(x)$
detekce chyb = počet LN sloupců $\mathbb{H}$	detekce shluků chyb délky $l \leq m = \deg G(x)$

6.5 Jednoduché bezpečnostní kódy

1. *opakovací kód (n,1)* - opakuje jeden bit.
   – velká redundance ⇒ malý informační obsah
2. *koktavý kód (jk,k)* - j-krát se opakuje totéž (k bitů).
   – velká redundance ⇒ malý informační obsah
3. *parita (k+1,k)* - $\vec{b} = (a_1, \ldots, a_k, p)$
   p - parita $\begin{cases} \text{sudá:} p = a_1 \oplus \cdots \oplus a_k \\ \text{lichá: } p = a_1 \oplus \cdots \oplus a_k \oplus 1 \end{cases}$
   – minimální redundance
   – detekce pouze jedné chyby

6.6 Hammingovy kódy

Všechny Hammingovy kódy jsou SEC (Single Error Correcting), tj. opravují jednu chybu. Kódová vzdálenost je $d_{min} = 3$. Jsou zadávány kontrolní maticí, ve které jsou **vystřídány všechny možné nenulové** $(n-k)$**-bitové sloupce, a to každý právě jednou**. Tj:

$$\mathbb{H} = [\mathbb{P}, \mathbb{I}_{n-k}]$$
$$\mathbb{G} = [\mathbb{I}_k, \mathbb{P}^T],$$

kde matice $\mathbb{P}$ je binární matice typu $[(n-k) \times k]$. Například pro Hammingův kód $(7, 4)$ mohou být tyto matice:

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

První část matice (4×4) je matice $\mathbb{P}$. Druhá část matice (3×3) je jednotková matice $\mathbb{I}$. Z matice $\mathbb{H}$ tedy okamžitě vyplývá matice $\mathbb{G}$:

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

V této matici je naopak první část (4×4) jednotková matice a druhá část (3×4) je matice $\mathbb{P}^T$.
$\mathbb{H} \cdot \vec{v}^T = O$, $\vec{v} = (v_0, \ldots, v_6) \Rightarrow$

$$\begin{array}{lllllllllllll} v_0 & \oplus & v_1 & \oplus & v_2 & & & \oplus & v_4 & & & = & 0 \\ v_0 & \oplus & v_1 & & & \oplus & v_3 & & & \oplus & v_5 & = & 0 \\ v_0 & & & \oplus & v_2 & \oplus & v_3 & & & & v_6 & = & 0 \end{array}$$

Bity v_0, v_1, v_2, v_3 jsou informační. Kontrolní bity jsou tedy v_4, v_5, v_6 a získáme je z předchozí soustavy $\Rightarrow$

$$\begin{aligned} v_4 &= v_0 \oplus v_1 \oplus v_2 \\ v_5 &= v_0 \oplus v_1 \oplus v_3 \\ v_6 &= v_0 \oplus v_2 \oplus v_3 \end{aligned}$$

$\vec{s} = \vec{c} \cdot \mathbb{H}^T$
Pokud došlo k chybě, pak je syndrom roven číslu udávající chybný bit v chybovém vektoru $\Rightarrow$ oprava spočívá v negaci bitu.

6.7 Rozšířený Hammingův kód

Rozšířený Hammingův kód je kód SEC-DEC (korekce jedné chyby, detekce dvou chyb). Kódová slova jsou doplněna na sudou paritu. V kontrolní matici $\mathbb{H}$ se to projeví doplněním všech řádků o nulu na konci a přidáním řádků samých jedniček.

Pro konstrukci generátoru syndromu je výhodné v kontrolní matici minimalizovat počet jedniček $\Rightarrow$ matice se konstruuje tak, aby její sloupce měly lichou paritu a minimální počet jedniček a aby její řádky měly pokud možno stejnou Hammingovu váhu. Syndromy jednonásobných chyb pak mají lichý počet jedniček a syndormy dvojnásobných chyb mají sudý počet jedniček:
$\vec{s} = \times \times \times 1 \Rightarrow$ 1 chyba
$\vec{s} = \times \times \times 0 \Rightarrow$ 2 chyby kromě $\vec{s} = 0000$

6.8 Systematické cyklické kódy

$B(x) = A(x) \cdot x^m + A(x) \cdot x^m \% G(x)$, kde $m = \deg G(x)$