Mária Franeková *

MATEMATICKÝ APARÁT PRE VÝPOČET PRAVDEPODOBNOSTI CHYBY DEKODÉROV BLOKOVÝCH KÓDOV

MATHEMATICAL APPARATUS FOR ERROR PROBABILITY DETERMINATION OF BLOCK CODE DECODERS

Článok sa zaoberá problematikou bezpečnej komunikácie v uzatvorených prenosových systémoch pri použití blokových systematických (n, k) kódov. Je uvedený matematický aparát pre výpočet pravdepodobnosti nedetegovanej chyby pre kódy so známou a neznámou váhovou štruktúrou. Číselné výsledky sú platné pre symetrický binárny kanál, za predpokladu nezávislých chýb v kanáli.

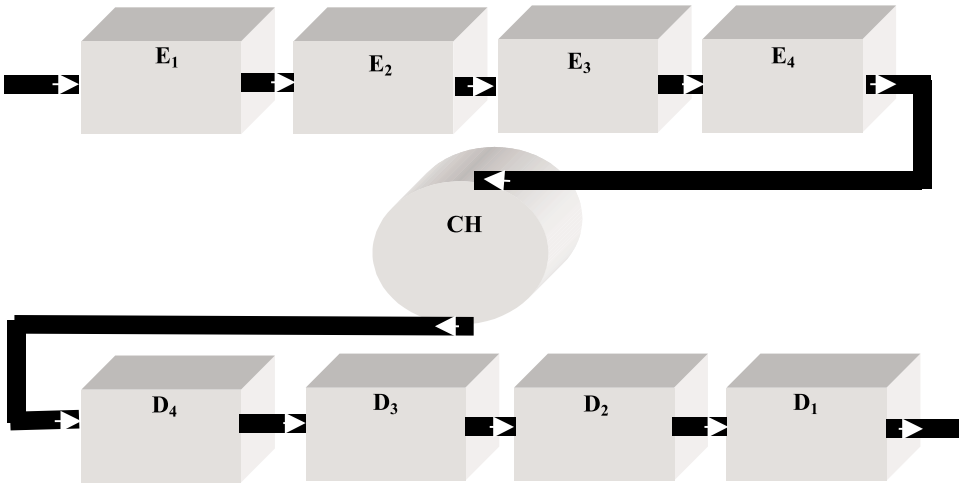
1. Úvod

Úroveň bezpečnosti pri prenose informácie cez komunikačný kanál je daná sumárnou hodnotou zloženou z integrity, dostupnosti a dôvernosti. Typické poradie kodérov a dekodérov v komunikačnom reťazci je ukázané na obr. 1.

This paper deals with safety - related communication in closed transmission systems when block systematic (n, k) safety codes are used. The mathematical apparatus for calculation of probability of undetected sequence error is described for codes, with known and unknown weight structure. The numerical results are valid for a Symmetric Binary Channel when the occurrence of independent errors is assumed.

1. Introduction

The level of security for transmission of information across the communication channel is calculated as a summary value of integrity, availability and confidentiality. A typical order of coders and decoders in the communication chain is shown in the Fig. 1.



Obr. 1. Komunikačný reťazec pre prenos informácie
Fig. 1. Communication chain for transmission of information

Poradie komponentov pre prenos informácie v komunikačnom reťazci nie je náhodné, každý blok má presne určenú lokalitu.

The orders of components in a communication chain for transmission of information are not random; every block has its fixed locality.

* **Ing. Mária Franeková, PhD.**
Department of Information and Safety Systems, Faculty of Electrical Engineering, University of Žilina, Veľký diel, SK-010 26 Žilina, Slovak Republic, Tel.: +421-89-5133 248, E-mail: frane@fel.utc.sk

Poradie dvojíc komponentov kodér/dekodér v komunikačnom reťazci je nasledovné:

E_1/D_1 – kompresný (zdrojový) kodér/dekodér
E_2/D_2 – šifrovací kodér/dekodér
E_3/D_3 – zabezpečovací (kanálový) kodér/dekodér
E_4/D_4 – modulátor/demodulátor
CH – prenosový kanál

V niektorých prípadoch nie je nutné použiť všetky uvedené komponenty, to závisí od konkrétnej aplikácie. Tiež kanál môže byť napríklad pamäť, do ktorej je informácia zapísaná alebo zakódovaná a potom je čítaná alebo dekódovaná.

Nezmenenie a nenarušenie informácie počas prenosu možno zabezpečiť pomocou zabezpečovacích kódov, ktoré by mali byť schopné ochrániť systém podľa požadovanej úrovne integrity.

V uzavretých prenosových systémoch podľa normy prEN 159-1 [1] môže byť použitý ako zabezpečovací kód iba detekčný kód. Pre splnenie požadovanej úrovne bezpečnosti je potrebné detegovať typické zhluky. Tieto môžu byť: náhodné chyby, zhluky chýb, systematické alebo kombinované chyby. Úroveň bezpečnosti je efektívna vtedy, ak pravdepodobnosť nedetegovanej chyby kódového slova je vyššia než požadovaná hodnota. Tieto požiadavky možno realizovať pomocou veľkého množstva zabezpečovacích kódov [2], [5]. Frekventovanou závislosťou, ktorá ohodnocuje úroveň bezpečnosti je pravdepodobnosť chyby kódového slova p_e v závislosti od bitovej chybovosti p_b používaného kanála. Každý odporučený alebo použitý kód v uzatvorenom prenosovom systéme je potrebné analyzovať a vypočítať maximum pravdepodobnosti nedetegovanej chyby kódového slova pre použitý prenosový kanál. Pri výpočte možno použiť štatistické hodnoty pravdepodobnosti chyby elementárneho symbolu (bitu) niektorých typických kanálov. Binárny symetrický kanál alebo kanál s gausovským aditívnym šumom sú veľmi často používané matematické modely kanálov, pri vyjadrení matematického aparátu v sledovanej oblasti [3].

V článku sú uvedené metódy výpočtu pravdepodobnosti nedetegovanej chyby kódového slova systematických blokových kódov:
1. v prípade, keď je známa váhová funkcia kódových zložiek (napr. Hammingove kódy)
2. v prípade, keď nie je jednoduché generovať váhovú funkciu kódových slov (napr. binárne BCH kódy).

2. Metódy výpočtu pravdepodobnosti nedetegovanej chyby kódového slova

Pri niektorých postupoch výpočtu pravdepodobnosti chyby [2], [4] je potrebné poznať všetky kódové slová kódu. Potom pravdepodobnosť nedetegovanej chyby kódového slova p_{e1} pre binárny symetrický kanál (BSC) možno vypočítať podľa (1):

$$P_{e1} = \sum_{i=\left\|\frac{d_{min}+1}{2}\right\|}^{n} A_i \, p_b^i \, (1-p_b)^{n-1} \qquad (1)$$

The order of couples of components in the communication chain is the following:

E_1/D_1 – compression (source) encoder/decoder
E_2/D_2 – ciphering encoder/decoder
E_3/D_3 – safety (channel) encoder/decoder
E_4/D_4 – modulator/demodulator
CH – transmission channel

In some cases it is not necessary to use all the above shown components, depending on their concrete application. Also the channel can be e.g. a memory, where information is written or coded and then is read or decoded.

The non-modification and non-corruption of information during transmission can be secured by safety codes, which should be capable of protecting the system, in the required level of integrity.

In the closed transmission system according to the norm prEN 159-1 [1] only the error detection code can be used as a safety code. For fulfilling the required safety integrity level it is necessary to detect typical errors. These errors can be random errors, burst errors, systematic or combined errors. The level of security is effective when the probability of undetected sequence error is bigger than a demanded value. These requirements can be realized with the sufficiently complex safety codes [2], [5]. A frequent function relation used for evaluation of a safety level is error probability of code word p_e depending on a bit error rate p_b of the used channel. It is necessary to analyze and determine the utmost probability of an undetected sequence error for the used transmission channel for every recommended or used code in a closed transmission system. In the process of calculation the statistical values of probability of bit error for some typical channels can be used. Binary Symmetric Channel or Additive White Gaussian Noise Channel is an often-used mathematical model of a channel for the expression of mathematical apparatus in this area [3].

In the article the following methods of the calculation of probability of undetected sequence error of systematic block codes are introduced:
1. in a case when a weight function of code words is known (e.g. Hamming codes)
2. in a case when it is not easy to generate the weight function of code words (e.g. BCH codes).

2. Methods for calculation of probability of an undetected sequence error

An approach of determination of error probability [2], [4] needs to know all sequences (or code words) of code. Then the probability of an undetected sequence error p_{e1} for Binary Symmetric Channel (BSC) can be calculated according to (1):

$$P_{e1} = \sum_{i=\left\|\frac{d_{min}+1}{2}\right\|}^{n} A_i \, p_b^i \, (1-p_b)^{n-1} \qquad (1)$$

Kde: d_{min} je minimálna Hammingova vzdialenosť kódu
A_i je celkový počet kódových slov s váhou i
p_b je bitová chybovosť kanála

Predpokladáme, že chyby v kanáli sú nezávislé a vyskytujú sa s bitovou chybovosťou p_b a ich výskyt možno aproximovať hustotou rozdelenia pravdepodobnosti podľa binomického rozloženia.

Hammingove perfektné (n, k) kódy [5] sú jedny z mála skupín kódov, kde je známa kompletná váhová funkcia $A(x)$:

$$A(x) = \sum_{i=0}^{n} A_i x^i \tag{2}$$

Váhová funkcia pre Hammingove kódy s $d_{min} = 3$ a dĺžkou kódového slova $n = 2^r - 1$ (r je počet redundantných bitov) je [2]:

Where: d_{min} is minimal Hamming distance of code
A_i is the total number of sequences in the code with weight i
p_b is a bit error rate of channel

Provided that errors in the channel are independent and occur with a bit error rate p_b and the occurrence of error might be approximated by a binomial density of probability.

Hamming perfect (n, k) codes [5] belong to a few classes of codes for which the complete weight function of code words $A(x)$ is known, where A(x) is the weight-enumerating function of a code:

$$A(x) = \sum_{i=0}^{n} A_i x^i \tag{2}$$

The weight-enumerating function for the distance-3 Hamming codes of code words length $n = 2^r - 1$ (where, r is number of redundancy bits) is [2]:

$$A(x) = \frac{1}{n+1}\left[(1+x)^n + n(1+x)^{(n-1)/2} \cdot (1-x)^{(n-1)/2}\right] \tag{3}$$

Pre kódové slová s väčšou dĺžkou n môže byť výpočet p_{e1} komplikovaný. Veľkou výhodou je, že hodnoty funkcie $A(x)$ sú symetrické pozdĺž $(n-1)/2$, čo zjednodušuje výpočet. Hammingove kódy s $d_{min} = 3$ sa dajú ľahko prekonvertovať na Hammingove kódy s $d_{min} = 4$ pridaním jedného paritného bitu. To spôsobí, že všetky kódové slová s váhou i (pre nepárnu váhu) sa stanú kódovými slovami s váhou $i + 1$.

Pozn.: *Platí len pre rozšírené Hammingove kódy s párnou paritou. Výsledky pravdepodobnosti nedetegovanej chyby kódového slova pre rozšírený Hammingov kód sú uvedené v článku [5].*

Nie pre všetky zabezpečené blokové kódy je jednoduché generovať váhovú funkciu kódových slov $A(x)$. Potom pravdepodobnosť nedetegovanej chyby kódového slova nemôže byť počítaná podľa vzťahu (1). V uzatvorených prenosových systémoch sú veľmi často používanými detekčnými kódmi binárne cyklické kódy, s dĺžkou kódového slova n a generačným polynómom $g(x)$ [5]. Rovnicu (1) možno modifikovať podľa [4], ak hodnota A_i je aproximovaná pomocou (4):

$$A_i \cong \frac{1}{2^{n-k}}\binom{n}{i} \tag{4}$$

Potom pravdepodobnosť nedetegovanej chyby kódového slova je daná vzťahom (5):

$$p_{e2} \cong \frac{1}{2^{n-k}} \sum_{i=d_{min}}^{n} \binom{n}{i} p_b^i (1-p_b)^{n-i} \tag{5}$$

Ak súčin $np_b << 1$ môže byť suma v (5) aproximovaná prvým členom sumy (6):

$$p_{e3} \cong \frac{1}{2^{n-k}}\binom{n}{d_{min}} p_b^{d_{min}} (1-p_b)^{n-d_{min}} \tag{6}$$

Je evidentné, že vo vzťahu (6) je potrebné poznať okrem parametrov (n, k) aj minimálnu Hammingovu vzdialenosť kódu d_{min} kódových slov. Ak je jej hodnota neznáma, na výpočet d_{min} možno

For sequences of a larger length n calculations of p_{e1} can be complicated a big advantage is that the values of function $A(x)$ are symmetrical along the point $(n-1)/2$, what eases the calculation The distance-3 Hamming codes can be converted in a very easy way to a minimum distance-4 code by appending one additional parity symbol to all code words. It causes that all weight-i code words (with odd weight) become weight- $i + 1$.

Note: *It is valid for expanded Hamming codes for even parity, only. The results of values of probability of undetected error for extended Hamming codes are shown in paper [5].*

It is not easy to generate the weight function of code words $A(x)$ for all safety block codes. Then the probability of undetected sequence error cannot be calculated by given (1). In the closed transmission system some types of binary symmetric cyclic (n, k) codes with code word of lengths n and with generate polynomial $g(x)$ as the detection codes are very often used. Equation (1) can be modified [3], where value of A_i is approximated by (4).

$$Ai \cong \frac{1}{2n-k}\binom{n}{i} \tag{4}$$

Then probability of an undetected sequence error is given by (5):

$$p_{e2} \cong \frac{1}{2^{n-k}} \sum_{i=d_{min}}^{n} \binom{n}{i} p_b^i (1-p_b)^{n-i} \tag{5}$$

If $np_b << 1$ in (5) then, can be sum approximated by the first member of sum (6):

$$p_{e3} \cong \frac{1}{2^{n-k}}\binom{n}{d_{min}} p_b^{d_{min}} (1-p_b)^{n-d_{min}} \tag{6}$$

It is evident that in the relationship (6) it is necessary to know besides the parameters of n, k, also the minimum Hamming distance of code words d_{min}. If this value is unknown for determina-

použiť Gilbertovu nerovnosť pre párne kódové slová (7) a pre nepárne kódové slová (8).

$$2^k \sum_{i=0}^{(d_{min}-1)/2} \binom{n}{i} \leq 2^n \tag{7}$$

$$2^k \sum_{i=0}^{(d_{min}-2)/2} \binom{n-1}{i} \leq 2^{n-1} \tag{8}$$

Pre niektoré hodnoty (n, k) binárnych cyklických kódov a BCH kódov sú výsledky Gilberovej nerovnosti tabelované [2], [4], [5].

3. Výsledky simulované počítačom

V príspevku sú uvedené príklady výpočtu hodnôt pravdepodobnosti nedetegovanej chyby kódového slova p_{e1}, p_{e2} a p_{e3} vyjadrené vzťahmi (1), (5) a (6) pre interval bitovej chybovosti kanála od $p_b = 10^{-9}$ do $p_b = 0{,}5$. Prvá metóda výpočtu, zo znalosti váhovej funkcie kódu, je demonštrovaná na príklade rozšíreného Hammingovho kódu (128,120) s $d_{min} = 4$, ktorý možno použiť v uzatvorenom prenosovom systéme pre detekciu dvoch chýb v kódových slovách. Číselné hodnoty sú uvedené v tab. 1 a grafické závislosti na obr. 2. Pre porovnanie hodnôt počítaných podľa (5) a (1) sú vypočítané aj hodnoty p_{e2} a v tab. 1 je uvedená odchýlka medzi nimi $\delta = abs\,(p_{e1} - p_{e2})$. Váhová funkcia Hammingovho (128,120) kódu je určená pomocou programu MATLAB. Výsledky pravdepodobnosti nedetegovanej chyby kódového slova podľa (1) sú vypočítané pomocou programu DERIVE.

Pravdepodobnosť nedetegovanej chyby kódového slova Hammingovho (128,120) kódu — Tab. 1

p_b	2^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
p_{e1}	3,9 e-3	2,58 e-4	7,54 e-8	8,42 e-12	8,52 e-16	8,53 e-20	8,53 e-24	8,53 e-28	8,53 e-32
p_{e2}	3,9 e-3	1,57 e-4	3,77 e-8	4,12 e-12	4,16 e-16	4,16 e-20	4,16 e-24	4,16 e-28	4,16 e-32
δ	0	1,01 e-4	3,77 e-8	4,30 e-12	4,36 e-16	4,37 e-20	4,37 e-24	4,37 e-28	4,37 e-32

Výsledky v tab. 1 a na obr. 2 ukazujú, že hodnoty pravdepodobnosti chyby počítané podľa vzťahov (1) a (5) sa odlišujú minimálne. Odchýlka δ medzi hodnotami je konštantná v mantise (hodnota 4,37) od hodnoty bitovej chybovosti $p_b = 10^{-6}$. Od hodnoty $pb = 10^{-2}$ sa odchýlka δ líši pravidelne o štyri rády v exponente.

Druhá metóda výpočtu pravdepodobnosti chyby pre blokové kódy s neznámou váhovou funkciou je ukázaná na príklade binárneho BCH kódu (255,199) s $d_{min} = 15$. Kód dokáže detegovať 14 chýb v kódovej zložke. Výsledky sú uvedené v tab. 2 pre rovnaký interval bitovej chybovosti. Hodnoty p_{e2} a p_{e3} sú určené podľa

tion of d_{min} Gilbert's unequation for even length of code words (7) and for odd length of code words (8) can be used.

$$2^k \sum_{i=0}^{(d_{min}-1)/2} \binom{n}{i} \leq 2^n \tag{7}$$

$$2^k \sum_{i=0}^{(d_{min}-2)/2} \binom{n-1}{i} \leq 2^{n-1} \tag{8}$$

For some valid combination (n, k) of binary cyclic codes and BCH codes the values of Gilbert's unequation are tabulated [2], [4], [5].

3. Computer simulation results

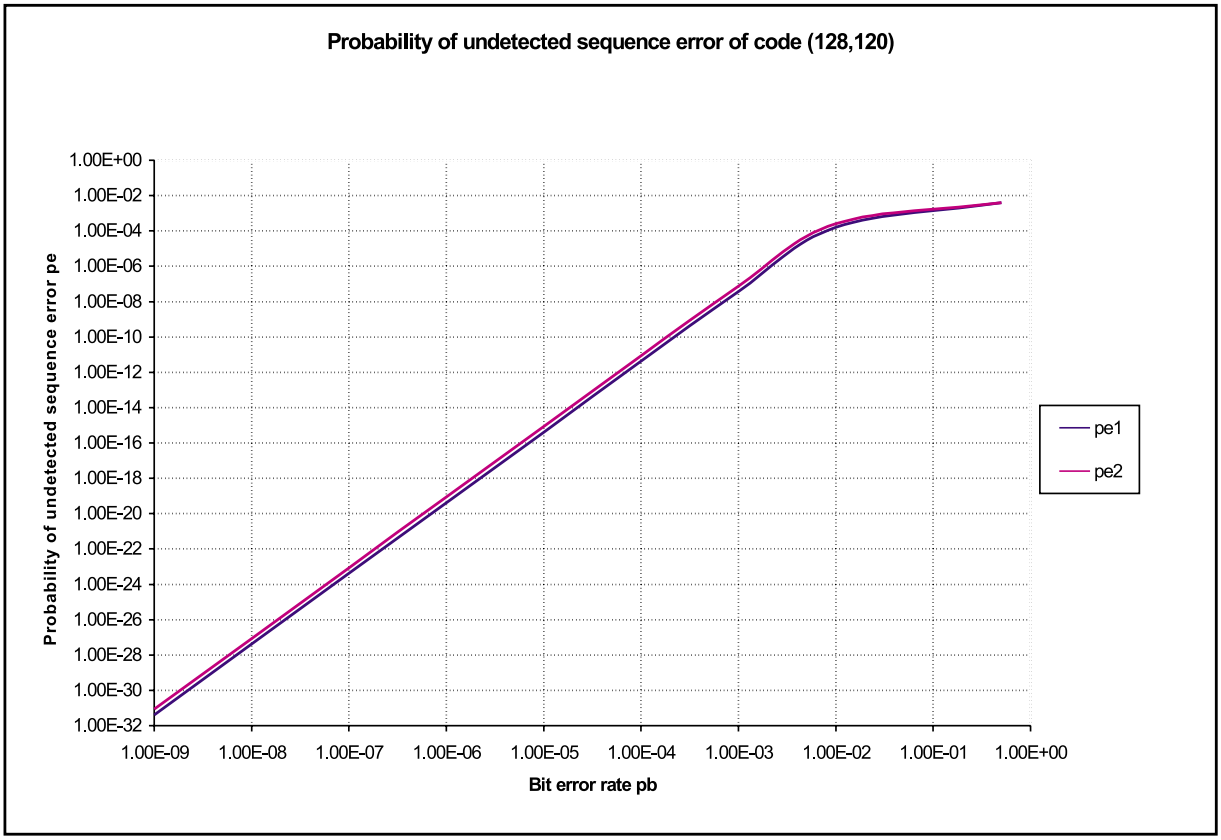
In the paper are shown examples of calculation of the values of probability of an undetected sequence error p_{e1}, p_{e2} and p_{e3} given by relations (1), (5) and (6) for interval of a bit error rate of a channel from $p_b = 10^{-9}$ to $p_b = 0.5$. The first method of calculation, based on knowledge of the weight function of code, is demonstrated by an example of extended Hamming (128,120) code with $d_{min} = 4$, which can be used in the closed transmission system for detection of two errors in every code word. The numerical results are shown in Tab. 1 and the graphical results in Fig. 2. For comparison of the values given by (5) and (1) are calculated the results of p_{e2}, too and in Tab. 1 the deviation between them $\delta = abs\,(p_{e1} - p_{e2})$ is determined. The weight function of Hamming (128,120) code according to (3) and undetected sequence error probability according to (1) are calculated via program MATLAB. The values of undetected sequence error probability given by (5), (when the weight function of code words is unknown), are calculated with the help of program DERIVE.

The undetected sequence error probability for Hamming (128, 120) code — Tab. 1

p_b	2^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
p_{e1}	3.9 e-3	2.58 e-4	7.54 e-8	8.42 e-12	8.52 e-16	8.53 e-20	8.53 e-24	8.53 e-28	8.53 e-32
p_{e2}	3.9 e-3	1.57 e-4	3.77 e-8	4.12 e-12	4.16 e-16	4.16 e-20	4.16 e-24	4.16 e-28	4.16 e-32
δ	0	1.01 e-4	3.77 e-8	4.30 e-12	4.36 e-16	4.37 e-20	4.37 e-24	4.37 e-28	4.37 e-32

The results in Tab. 1 and Fig. 2 show that the values of an undetected sequence error probability, calculated by relations (5) and (1) are minimally distinguished. The deviation between the values is constant in mantisa (value 4.37) for bit error rate from $p_b = 10^{-6}$. The deviation from value $p_b = 10^{-2}$ is periodically different by four orders in the exponent.

The second method of calculation of error probability for block codes with an unknown weight function is shown for binary BCH (255,199) code with $d_{min} = 15$. The code can detect 14 independent errors in every code word. The results are shown in Tab. 2 for the same interval of a bit error rate.



Obr. 2. Grafické výsledky pravdepodobnosti chyby nedetegovaného slova pre Hammingov (128,120) kód
Fig. 2. Graphical results of the probability of undetected sequence error for Hamming (128,120) code

vzťahov (5) a (6). Hraničná hodnota bitovej chybovosti p_b, od ktorej možno vzťah (5) aproximovať vzťahom (6) pre dĺžku kódového slova $n = 255$, je $p_b = 10^{-3}$.

Pravdepodobnosť nedetegovanej chyby pre binárny BCH (255, 199,14) kód — Tab. 2

p_b	2^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
p_{e2}	1,3 e-17	9,2 e-25	6,9 e-39	8,5 e-54	8,7 e-69	8,7 -84	8,7 e-99	8,7 e-114	8,7 -129
p_{e3}	1,5 e-17	7,8 e-25	6,8 e-39	8,5 e-54	8,7 e-69	8,7 e-84	8,7 e-99	8,7 e-114	8,7 e-129

4. Záver

Počas prenosu informácie medzi zabezpečovacími systémami musí byť garantovaná požadovaná hranica pravdepodobnosti chyby. Preto nutnou súčasťou celkovej analýzy bezpečnosti je aj analýza bezpečnosti použitého blokového kódu. V príspevku sú uvedené dve metódy výpočtu pravdepodobnosti nedetegovanej chyby kódového slova, keď poznáme všetky kódové zložky kódu a v prípade, keď je veľmi obtiažne generovať váhovú funkciu kódu. Porovnanie

The undetected sequence error probability for binary BCH (255, 199,14) code — Tab. 2

p_b	2^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
p_{e2}	1.3 e-17	9.2 e-25	6.9 e-39	8.5 e-54	8.7 e-69	8.7 -84	8.7 e-99	8.7 e-114	8.7 -129
p_{e3}	1.5 e-17	7.8 e-25	6.8 e-39	8.5 e-54	8.7 e-69	8.7 e-84	8.7 e-99	8.7 e-114	8.7 e-129

The values of p_{e2} and p_{e3} are calculated according to (5) and (6). The bounded value for a bit error rate pb from which relation (5) can be approximated by relation (6) for code word length $n = 255$ is $p_b = 10^{-3}$.

4. Conclusion

During communication between safety-related systems the required bound of undetected sequence error probability must be guaranteed. Therefore with some type of safety codes used of data transmission the analysis is a necessary part of the total analysis of security. In the paper two methods of calculation of the probability of an undetected sequence error are dealt with – when all code sequences of code are known and when the weight function

výsledkov pravdepodobnosti nedetegovanej chyby kódového slova počítaných podľa vzťahov (1), (5) a (6) je realizované na príkladoch systematických blokových kódov – Hammingov (128,120) a binárny BCH kód (255, 199, 14). Výsledky pre p_{e1} a p_{e2} a pre p_{e2} a p_{e3} uvedené v tab. 1 a tab. 2 sú rádovo ekvivalentné, čo je veľmi dôležitý poznatok. Výsledky sú platné len pre nezávislé chyby v kanáli. Z obr. 1 vyplýva, že výsledky pravdepodobnosti p_{e1} a p_{e2} limitujú pre $p_b = 0{,}5$ k hodnote $2^{-(n-k)}$, čo podľa [1] je najnižšia možná hodnota pravdepodobnosti nedetegovanej chyby pri použití blokových (n, k) kódov.

of code words is very difficult to generate or it is unknown. The comparison of results of probability of undetected sequence error calculated by relations' (1), (5) and (6) are performed on the examples of systematic block codes – Hamming code (128,120) and binary BCH code (255, 199). The results of probability p_{e1}, p_{e2} and p_{e2}, p_{e3} shown in Tab. 1 and Tab. 2 are equivalent with the order, which is very a important evidence. The results are valid only for independent errors in the channel. From Fig. 1 follows that curves of p_{e1} and p_{e2} for $p_b = 0{,}5$ are limited to the value $2-(N-K)$, which is according to [1] the lowest value of error probability valid for block (n,k) codes.

5. Literatúra – References

[1] PrEN 159-1: *Railway applications - Communication, signalling and processing systems,* Part1: Safety-related communication in closed communication system
[2] ADÁMEK, J.: *Foundation of coding,* Theory and applications of Error-Correcting Codes with an Introduction to Cryptography and information Theory,
[3] CLARK, G., CAIN, B.: *Error-Correction Coding for Digital Communications,* Plenum Press, New York, 1988
[4] HRDINA, Z., VEJRAŽKA, F.: *Digitální rádiová komunikace,* skriptum ČVUT, Praha, 1994
[5] FRANEKOVÁ, M., BUBENÍKOVÁ, E.: *The calculation of the probability of undetected sequence error for the ARQ systems*, MOSIS 2000, máj 2000, Rožnov pod Radhoštěm, pp. 203-208