Conference Applications of Mathematics 2015, in honor of the birthday anniversaries of Ivo Babuška (90), Milan Práger (85), and Emil Vitásek (85)
J. Brandts, S. Korotov, M. Křížek, K. Segeth, J. Šístek, T. Vejchodský (Eds.)
Institute of Mathematics CAS, Prague 2015

MESSAGE DOUBLING AND ERROR DETECTION IN THE BINARY SYMMETRICAL CHANNEL

Lucie Kárná[1,2], Štěpán Klapka[2]

[1] Faculty of Transportation Sciences CTU
Na Florenci 25, Praha 1, Czech Republic
karna@fd.cvut.cz

[2] AŽD Praha s.r.o., Research and Development
Žirovnická 2, Praha 10, Czech Republic
klapka.stepan@azd.cz

Abstract: The error correcting codes are a common tool to ensure safety in various safety-related systems. The usual technique, employed in the past, is to use two independent transmission systems and to send the safety relevant message two times. This article focuses on analysis of the detection properties of this strategy in the binary symmetrical channel (BSC) model.
Besides, various modifications of the mentioned technique can be used. Their impact on the detection properties can be significant, positively or negatively. This article demonstrates one of these modifications.

Keywords: error correcting code, undetected error, message repetition

MSC: 62P30, 94A40, 94B70

1. Introduction

Communication safety is a small, but important part of the safety of every electronics-based system, particularly in railway interlocking systems. A special position in this issue has the safety code, because it is the unique tool to protect messages against corruption.

The basic motivation for this paper was the cooperation on design of interlocking systems. The communication protocol, used by our partner, includes sending the safety relevant messages twice using two transmission lines. It turns up, that safety analysis of this simple approach is not quite simple.

The first part of the article describes some basic terms of coding theory. The second part introduces the concept of probability of undetected error in the binary symmetrical channel as a basic tool for evaluating detection quality of the code. The next part investigates the main approaches to message doubling. The problem of calculating the probability of an undetected error in these cases is studied.

2. Coding theory

This section defines the basic terminology for linear binary codes and the related binary symmetrical channel (BSC) model. The "code-related" terminology in this paper is based on terms used in the mathematical coding theory (see for example [2]).

2.1. Linear binary codes

A linear binary $(n, k)-$code K is any $k-$dimensional subspace of the space $(\mathbf{Z}_2)^n$. Traditionally, binary vectors from $(\mathbf{Z}_2)^n$ are called *words*; the words from the code K are the *code words*. In an $(n, k)-$code the code word length is n, the number of information bits is equal to k and the number of redundant bits is equal to $c = n-k$. Any linear $(n, k)-$code K can be described by its *generator matrix*, whose rows are exactly the words forming a basis of the subspace K.

In practice, usually the code word of an $(n, k)-$code is created by the addition of c bits (the *redundant* or *control part* of the code word) to a word of length k (the *information part* of the code word). This technique is called a *systematic encoding*, the code is a *systematic code*. A generator matrix of the systematic code has the form $G = (E|B)$, where E denotes the identity matrix of the order k and B is some $k \times c$ matrix.

2.2. Error detection

During the transfer of a message unwanted modifications can occur. Usually, it is supposed that a number of bits is preserved and these modifications are manifested by altered bit(s). The adverse situation occurs, when the modification during transfer unfortunately creates another code word, different from the sent one. The receiver has no possibility to recognize this state.

This scenario is dangerous and results in an undetected error. The probability of such an undetected error of the detection codes used in safety relevant applications (including transportation control) is a very important safety parameter.

We define the *Hamming weight* of a word as the count of non-zero bits in the word. Then we define the *minimal distance* of a linear code as the smallest non-zero Hamming weight of its code word.

The minimal distance of a linear code sets the ability of the code to detect some classes of transmission errors. A code with a minimal distance d will detect all errors with at most $d-1$ modified bits in the transmitted code word (see [2, 3]).

For a more detailed description of the code, a *weight structure* of the code is defined as a vector $A = (A_0, A_1, A_2, \dots, A_n)$, where A_i denotes the number of code words with Hamming weight equal to i. For linear codes, the weight structure is fully sufficient for the description of its ability to detect errors.

3. Probability of undetected error

The most useful approach for measuring the detection properties of a code uses its maximal value of the probability of undetected error in a binary symmetrical channel.

3.1. Description of the BSC model

The binary symmetrical channel (BSC) is a simple probabilistic model based on a bit (binary symbol) transmission. The BSC model does not describe the reality completely, but it is an appropriate tool for comparison of the detection properties of the codes.

In this model the probability of an error is supposed to be independent from one bit to the next one. The probability p_e that the bit changes its value during the transmission (*bit error rate*) is the same for both possibilities ($0 \to 1$ and $1 \to 0$). The probability that the code word with n symbols is corrupted exactly in i symbols is then equal to

$$p_e^i \, (1 - p_e)^{n-i} \, . \tag{1}$$

The probability of an undetected error in the BSC model for a linear binary code K with code words of length n and with minimal Hamming distance d is given by the following formula

$$P_{ud}(K, p_e) = \sum_{i=d}^{n} p_e^i \, (1 - p_e)^{n-i} \, A_i, \tag{2}$$

where A_i is the number of code words with exactly i nonzero symbols and p_e is the bit error rate in the BSC channel.

For every linear $(n,k)-$code the value of the function $P_{ud}(K,.)$ for $p_e = 1/2$ is equal to $(2^k - 1)/2^n$ and this is a local maximum of this function. Although the use of a transmission channel with bit error rate near to 1/2 is virtually excluded, the standard EN 50159 for safety-related communication in railway applications [1] recommends not to use a better detection estimate than this value for calculations in a safety model.

Actually, for the codes used in safety relevant applications it is necessary to know (or, at least, estimate) an upper bound of the function $P_{ud}(K,.)$ on the entire interval $[0, 1/2]$. In particular, it is recommended to use codes with a monotone function $P_{ud}(K,.)$ or, at least, this function should not exceed the value $P_{ud}(K, 1/2)$ (see [1]).

3.2. Indirect calculation using dual code

The formula (2) for the probability of an undetected error of a code is quite simple in principle. However, the coefficients A_i (the number of code words with i nonzero symbols) cannot be expressed by some elegant formula (with exception of rare family of codes). They have to be calculated by counting the weight of every individual code word. As the number of code words is equal to 2^k, these calculations are not feasible for long code words.

To get more effective calculations, it is useful to apply the MacWilliams Identity, which links the weight structure of the given code and its dual code. These computations use another representation of the weight structure by the weight enumerator $\mathbf{pw}(x, K)$. It is the following formal polynomial:

$$\mathbf{pw}(K, x) = \sum_{i=0}^{n} A_i x^i. \tag{3}$$

3.2.1. Dual code

We define for the binary words $u = u_1 u_2 \dots u_n$ and $v = v_1 v_2 \dots v_n$

$$u \cdot v = \sum_{i=1}^{n} u_i \cdot v_i. \tag{4}$$

This bilinear form is usually referred as *inner product*, despite it does not satisfy condition that from $u \cdot u = 0$ follows $u = (0, 0, \dots, 0)$. This is a consequence of the fact that in the space $\mathbf{Z}_2$ it is $1 + 1 = 0$.

A *dual code* to the linear binary $(n, k)-$code K is a linear binary $(n, n-k)-$code $K^\perp$ consisting from all words $u \in (\mathbf{Z}_2)^n$, whose inner product with every code word from the code K is equal to zero:

$$u \in K^\perp \iff u \cdot v = 0 \text{ for every } v \in K. \tag{5}$$

If the code K is a systematic code with generator matrix $G = (E|B)$, where E is the identity matrix and B is some $k \times c$ matrix, then the dual code $K^\perp$ has a generator matrix $G^\perp = (B^{\mathrm{T}}|E)$, where B^{T} is the transposed of the matrix B.

3.2.2. MacWilliams Identity

The following formula is the MacWilliams Identity for binary codes:

$$2^k \mathbf{pw}(K^\perp, x) = (1 + x)^n \mathbf{pw}\left(K, \frac{1 - x}{1 + x}\right). \tag{6}$$

The advantage of this formula is that the dual code has much fewer code words ($2^{n-k} << 2^k$, because typically, $n - k = c << k$) and then it is significantly easier to compute the weight distribution for a dual code.

4. Message doubling

A natural procedure to ensure authenticity of the message is to use two independent transmission systems and to send the safety relevant message twice. The received message is considered undamaged only if both copies are delivered and theirs contents are matching.

The situation with a missing message is trivial, so we focus only on the case when both copies arrived and their length is preserved (verification of the correct length of the message is done by other techniques). In the BSC model (independent transmission of single symbols – bits), it is equivalent to a serial transmission using a single transmission channel.

4.1. Repetition of the message

A plain repetition of the message with length equal to k is represented by the linear binary $(2k, k)-$code with binomial weight structure, where

$$A_{2j} = \binom{n}{j} \quad \text{for } j = 0, \ldots, k \tag{7}$$

$$A_{2j-1} = 0 \quad \text{for } j = 1, \ldots, k. \tag{8}$$

The minimal distance of the code is equal to 2, which is insufficient for most purposes.

More useful is a repetition of the message already protected by some linear code. Consider a binary message of length k. This message we protect by a linear binary (n, k)-code K_A with minimal distance d and with known weight structure $A = (A_0, A_1, A_2, \ldots, A_n)$. Then we send this message twice.

This procedure corresponds to the protection of the message with linear binary $(2n, k)-$code K_D. The minimal distance of this code is equal to $2d$ and its weight structure, denoted as $D = (D_0, D_1, D_2, \ldots, D_n)$, is given by the weight structure of the code K_A:

$$D_{2j} = A_j \quad \text{for } j = 0, \ldots, n \tag{9}$$

$$D_{2j-1} = 0 \quad \text{for } j = 1, \ldots, n. \tag{10}$$

The probability of undetected error in the BSC of the code K_D is then

$$P_{ud}(K_D, p_e) = \sum_{i=2d}^{2n} p_e^i \,(1-p_e)^{n-i}\, D_i = \sum_{i=d}^{n} \left(p_e^i \,(1-p_e)^{n-i}\right)^2 A_i. \tag{11}$$

Obviously, we have

$$P_{ud}(K_D, .) < P_{ud}(K_A, .). \tag{12}$$

The following graph illustrates the situation for one sample code with length $n = 32$ and with $c = 8$ control bits. (Note: it is a shortened cyclic code generated by the polynomial $x^8 + x^7 + x^2 + 1$ – for explanation see e. g. [3].) The upper curve represents the probability of an undetected error for the sample code, the lower curve represents the corresponding probability with repetition of the message. The vertical axis is in logarithmic scale.

Let us consider the lower bound of the function $P_{ud}(K_A, .)$ as a $A_d\, p_e^d (1 - p_e)^{n-d}$. The ratio between the lower bounds for the codes K_D and K_A is $p_e^d (1 - p_e)^{n-d}$, and the minimal improvement is obtained for $p_e = d/n$. Hence with increased length n the maximal value of the lower bound decreases significantly slower than the value $P_{ud}(K_D, p_e)$. From this it is evident that the minimal distance is a very important parameter, which has a dominant influence to the detection properties of doubling messages.

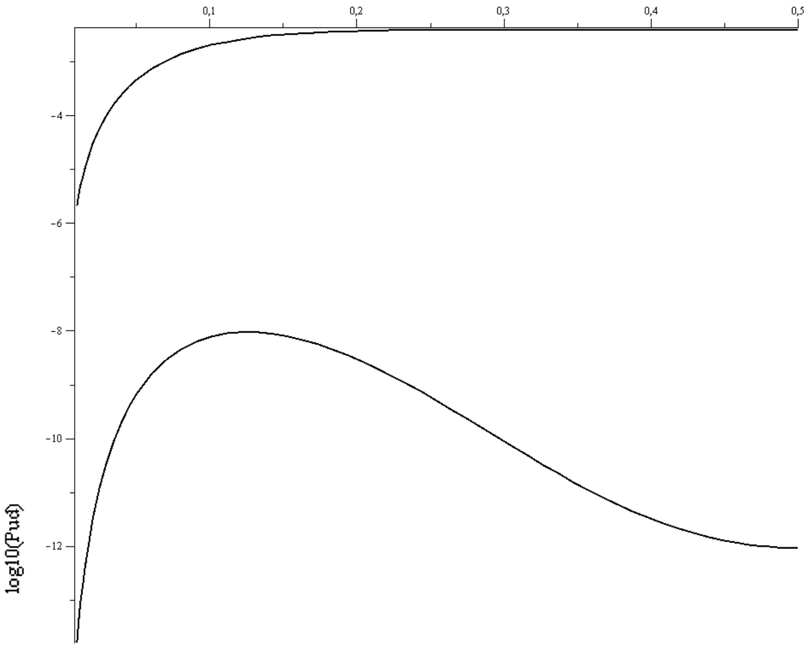

Figure 1: The probability of an undetected error for the sample code (upper curve) and for the same code combined with repetition of the message. Horizontal axis: bit error rate p_e, vertical axis: logarithm of probability of undetected error $P_{ud}(p_e)$.

4.2. Double encoding of the message

In some situations a more sophisticated approach can be useful. We protect a binary message M of length k by a linear binary $(n, k)-$code K_A with known weight structure $A = (A_0, A_1, A_2, \ldots, A_n)$; we denote this encoded message by M_A. Then we repeat this procedure with the original message M and with another linear binary $(n, k)-$code K_B with weight structure $B = (B_0, B_1, B_2, \ldots, B_n)$; denote the encoded message by M_B. Finally we send both messages M_A and M_B using two separate transmission lines.

One advantage of this approach is that the received messages are "signed" – if one of the messages M_A and M_B is wrong, we know on which transmission line (or in which encoder) the failure occurred. More important, this technique protects against the situation, when two copies of one received message are handled as two independent messages.

4.2.1. Weight structure

The two-transmission-lines configuration is in the BSC model equivalent with transmission of concatenated messages M_A and M_B. This corresponds with some linear binary $(2n, k)-$code K_{AB}. Unfortunately, the weight structure of the code K_{AB} cannot be derived from the weight structures of the codes K_A and K_B. However, the number of information bits k is equal for all three codes K_A, K_B and K_{AB} and therefore if the calculation of the weight structure of the codes K_A, K_B is manageable, then for the code K_{AB} the computation is practicable as well.

The questionable situation occurs, when the number of information bits k is too high and it is impossible to generate 2^k code words in a reasonable time. The dual codes to the K_A and K_B are $(n, n-k)-$codes, and if the number of the redundant bits $c = n - k$ is acceptably small, it is possible to compute the weight structures of these duals and then use the MacWilliams identity (6) to compute the weight structures of the codes K_A and K_B.

However, the dual code to the code K_{AB} is a $(2n, n+c)-$code and generation of the 2^{n+c} code words may be impossible, as in a typical case the number of information bits k is considerably greater than the number of control bits $c = n-k$. This problem can be solved by utilization of the special form of the code dual to K_{AB}.

Let us assume that the codes K_A and K_B are systematic codes. This is a reasonable assumption, because every linear code is equivalent with a systematic code. Then the codes K_A and K_B have generator matrices in the form $G_A = (E|A)$ and $G_B = (E|B)$, respectively. A generator matrix of the code K_{AB} is $G_{AB} = (E|A|E|B)$, and there exists an equivalent generator matrix $(E|E|A|B)$. Then a generator matrix of the dual code $K_{AB}^{\perp}$ has the following form:

$$G_{AB}^{\perp} = \left(\begin{array}{c|ccc} E & E & \mathbf{0} & \mathbf{0} \\ A^T & \mathbf{0} & E & \mathbf{0} \\ B^T & \mathbf{0} & \mathbf{0} & E \end{array} \right), \tag{13}$$

where $\mathbf{0}$ denotes a zero matrix.

The matrix

$$G^* = \left(\begin{array}{c|cc} A^T & E & \mathbf{0} \\ B^T & \mathbf{0} & E \end{array} \right), \tag{14}$$

derived from the $G_{AB}^{\perp}$, is a generator matrix of some $(k + 2c, 2c)-$code K^*. In the favourable case it is acceptable to generate 2^{2c} code words and enumerate their weights.

Computation of the weight structure of the code K_{AB} is based on more detailed information about weights of the code words of the code K^*. Rather than the weight structure we compute a matrix of weight structures. We split a code word into two parts: the information part of length $2c$ and the control part of length k. Then we construct a matrix $N = (n_{ij})$, where n_{ij} is the number of code words of the code K^* with weight of the information part equal to i and weight of the control part equal to j.

Every code word of the code $K_{AB}^{\perp}$ is the sum of two words $v + w$:

- $v = (u, u, o)$, where u is an arbitrary binary word of length k and o is a zero vector of length $2c$, and

- $w = (w_1, o, w_2)$, where (w_1, w_2) is a code word of the code K^* (w_1 consists of its first k bits, w_2 is the rest) and o is a zero vector of length k.

Consider a word w with weight of w_1 equal to i and weight of w_2 equal to j. We add to this word every possible word of the type $v = (u, u, o)$. For every position, where it is one in the word w_1 and zero in the word u, the weight of the sum $v + w$ increases by 2. Then, for given w there exist $\binom{k-i}{m}$ words with weight $i + j + 2m$. The number of these words w is $2^j n_{ij}$. Adding these contributions for all indices i and j we obtain the desired weight structure of the code $K_{AB}^{\perp}$ and finally by means of the MacWilliams Identity (6) the weight structure of the code K_{AB}.

This procedure is quite complicated, nevertheless, our computations show, that for a code with 16 control bits it is fully manageable on ordinary personal computer.

4.2.2. Upper estimate of $P_{ud}(K_{AB}, .)$

In case the enumeration of the 2^{2c} code words of the code K^* is computationally too difficult, but 2^c code words of the codes K_A and K_B is still computationally accessible, we can estimate the maximal value of $P_{ud}(K_{AB}, .)$ by the following construction.

We use the known weight structures $A = (A_0, A_1, \ldots, A_n)$ of the code K_A and $B = (B_0, B_1, \ldots, B_n)$ of the code K_B to create a new weight structure $C = (C_0, C_1, \ldots, C_n)$ of the fictive code K_f. The value of C_i we define as the maximum value of A_i, B_i. Then we consider doubling of the message with this fictive code K_f as described in Section 4.1 and enumerate the upper bound of the $P_{ud}(K_f, .)$. This is the upper bound for the function $P_{ud}(K_{AB}, .)$ as well.

5. Conclusions

Repetition of the message is a natural and undemanding method of protecting its content. In the safety relevant applications it is not a sufficient technique. Therefore, more sophisticated variations of this principle can be useful as additional defence.

Providing the probabilistic analysis of the code using some of these variants of message doubling is surprisingly complicated. Nevertheless, an effective, though not elegant, method for necessary computations was developed.

References

[1] EN 50159 Railway applications – Communication, signalling and processing systems – Safety-related communication in transmission systems. European standard, CENELEC, September 2010.

[2] Huffman, W. C. and Pless, V.: *Fundamentals of error-correcting codes.* Cambridge University Press, Cambridge, 2003.

[3] Sweeney, P.: *Error control coding. From theory to practice.* John Wiley & Sons, 2002.