A MODEL OF ERROR-DETECTING CODES BASED ON QUASIGROUPS OF ORDER 4

Nataša Ilievska
Institute of Informatics, Faculty of Natural Sciences and Mathematics
Skopje, Republic of Macedonia

Verica Bakeva
Institute of Informatics, Faculty of Natural Sciences and Mathematics
Skopje, Republic of Macedonia

ABSTRACT

Error-detecting codes are used to detect errors when messages are transmitted trough a noisy communication channel. We propose error-detecting codes using quasigroup operation * on the set $A=\{0,1,2,3\}$. In order to detect errors, we extend an input message $a_1a_2\ldots a_n$ to a block $a_1a_2\ldots a_nd_1d_2\ldots d_n$, where $d_i=a_i*a_{(i \bmod n)+1}$, $i=1,2,\ldots,n$. We calculate an approximate formula which gives the probability that there will be errors which will not be detected. Also, we find the block length such that this probability of undetected errors is smaller than some previous given value.

(**Keywords:** error-detecting codes, quasigroup, noisy channel, probability of undetected errors)

I. INTRODUCTION

A binary symmetric channel is a channel, in which inputs and outputs are 0 and 1. There are noises in the channel, because of which 0 can be transmitted as 1 and vice versa with probability $p<1/2$ (Figure 1). Therefore, the output message may not be the same as the input one. So, we need a mechanism to discover whether the correct message is received. For this reason, we concatenate some characters on the input message, defined by the code, which will help us to discover if the message is correctly transmitted or not. It is clear that these characters increase the redundancy of the message.

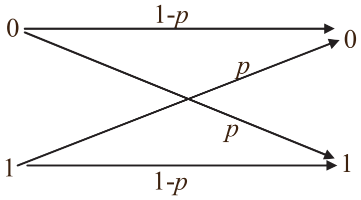

Figure 1: Binary symmetric channel

In Section II we give a proposal for designing a error-detecting code using quasigroups of order 4. In Section III, the probability of undetected errors is determined and we give a mechanism for separating the input message in blocks such that the length of a block is as small as possible and the probability of undetected errors is smaller than a given value.

II. CODE DESIGNING

Let consider an alphabet $A=\{0,1,2,3\}$ and an arbitrary quasigroup operation * on A. We define a code in the following way. Let the input message be $a_1a_2\ldots a_n$, $a_i\in A$, $i\in\{1,2,\ldots,n\}$. We define $d_i=a_i*a_{(i \bmod n)+1}$, $i\in\{1,2,\ldots,n\}$. Now, we transmit the extended message $a_1a_2\ldots a_nd_1d_2\ldots d_n$ through the noisy channel. The rate of this code is ½. Because there are noises in the channel, some of the characters may not be correctly transmitted. Let a_i be transmitted as a_i', d_i as d_i', $i\in\{1,2,\ldots,n\}$. If the character transmission is correct than a_i' will have the same value as a_i. Otherwise, a_i' will not be the same as a_i. So, the output message is $a_1'a_2'\ldots a_n'd_1'd_2'\ldots d_n'$. To check if there are any mistakes, the receiver of the message checks if $d_i'=a_i'*a_{(i \bmod n)+1}'$ for each $i=1,2,\ldots,n$. If any of these equalities are not satisfied, i.e. $d_i'\neq a_i'*a_{(i \bmod n)+1}'$ for any i, the receiver concludes that some errors occur during the block transmission and he calls the sender to send that block once again. But, some equality $d_i'=a_i'*a_{(i \bmod n)+1}'$ can be satisfied although any characters in that equality are incorrectly transmitted. In that case, incorrect transmission (error in transmission) will not be detected.

III. PROBABILITY OF UNDETECTED ERRORS

Let calculate the probability that there will be an error which will not be detected. Clearly, it is good this probability to be as small as possible. The probability of undetected errors is a function of p, so we are interested in the maximum of this function.

In general case, the formula which defines the probability that there will be an error which will not be detected is complicate, with a lot of parameters. Therefore, we have filtered the quasigroups such that the formula is independent from the distribution of the input message. After filtering 576 quasigroups of order 4, only 160 remain. Practically, only the class of fractal linear and fractal nonlinear quasigroups remains.

Theorem 1: Let $f(n,p)$ be the probability of undetected errors in a transmitted block with length n through the binary symmetric channel where p is the probability of incorrect transmission of a bit. If a fractal quasigroup is used for designing the code and the number of errors in a block is smaller than 5, then the probability of undetected errors is given by the following formulas:

$f(2,p) = 2v_0v_1+r_2$

$f(3,p) = 3v_0^3v_1 +3v_0v_2+r_3$

$f(4,p) = 4v_0^5v_1+4v_0^3v_2+2v_0^2v_1^2+4v_0v_3+r_4$

©2008 Institute of Informatics.

$$f(n,p)=nv_1v_0^{2n\text{-}3}+nv_2v_0^{2n\text{-}5}+\frac{n(n-3)}{2}v_1^2v_0^{2n\text{-}6}+nv_3v_0^{2n\text{-}7}$$
$$+\,n(n\text{-}4)v_2v_1v_0^{2n\text{-}8}+\frac{n(n-4)(n-5)}{6}v_1^3v_0^{2n\text{-}9}+nv_4v_0^{2n\text{-}9}$$
$$+\,n(n\text{-}5)v_3v_1v_0^{2n\text{-}10}+\frac{n(n-5)}{2}v_2^2v_0^{2n\text{-}10}$$
$$+\frac{n(n-5)(n-6)}{2}v_2v_1^2v_0^{2n\text{-}11}+\frac{n(n-5)(n-6)(n-7)}{24}v_1^4v_0^{2n\text{-}12},$$

for $n \geq 5$.

In the formulas, we use the following notations:

v_k – probability of undetected errors when exactly k consecutive characters of the initial message $a_1a_2\ldots a_n$ are incorrectly transmitted, i.e. the characters $a_i, a_{i+1},\ldots, a_{i+k\text{-}1}$ are incorrectly transmitted, but $a_{i\text{-}1}$ and a_{i+k} are correctly transmitted, k=1,2,3,4;

v_0 – probability of correct transmission of a character;

r_k – probability of undetected errors in a block with length k if all k characters are incorrectly transmitted, k=2,3,4.

Proof: Since each character of the alphabet $A=\{0,1,2,3\}$ can be presented by two bits, it is clear that

$$v_0=(1\text{-}p)^2 \quad (1)$$

Let denote the following random events:

A: errors occur in not more than 4 characters of the initial message $a_1a_2\ldots a_n$ and errors are not detected;

A_k: errors occur in exactly k characters of the initial message and errors are not detected, k=1,2,3,4.

Then we have that

$$A = A_1 + A_2 + A_3 + A_4 \quad (2)$$

Let calculate $P(A_1)$. The initial message has n characters and each of them can be incorrectly transmitted. There are n choices for choosing a character a_i which is incorrectly transmitted. The error will not be detected if the characters $d_{i\text{-}1}$ and d_i are incorrectly transmitted, too, but the equalities $d'_{i-1}=a'_{i-1}*a'_{((i-1)\bmod n)+1}$ and $d'_i=a'_i*a'_{(i \bmod n)+1}$ are satisfied. The rest $2n$-3 characters are correctly transmitted, so

$$P(A_1)=C_n^1\,v_1v_0^{2n\text{-}3}=nv_1v_0^{2n\text{-}3} \quad (3)$$

For calculating the probability $P(A_2)$, we denote the random events:

B_1: two consecutive characters of the initial message $a_1a_2\ldots a_n$ are incorrectly transmitted and the errors are not detected;

B_2: two nonconsecutive characters of the initial message are incorrectly transmitted and the errors are not detected.

Then

$$A_2 = B_1 + B_2 \quad (4)$$

There are n choices for two consecutive characters a_i and a_{i+1} which are incorrectly transmitted. The error will not be detected if some of characters $d_{i\text{-}1}$, d_i and d_{i+1} are incorrectly transmitted, too, but the equalities $d'_{i-1}=a'_{i-1}*a'_{((i-1)\bmod n)+1}$, $d'_i=a'_i*a'_{(i \bmod n)+1}$ and $d'_{i+1}=a'_{i+1}*a'_{((i+1)\bmod n)+1}$ are satisfied. The rest $2n$-5 characters are correctly transmitted, so we have that

$$P(B_1)=C_n^1\,v_2v_0^{2n\text{-}5}=nv_2v_0^{2n\text{-}5} \quad (5)$$

To calculate the number of two nonconsecutive characters we decrease the number of all pairs of two characters (which is C_n^2) for the number of two consecutive characters (C_n^1). Let say, a_s and a_r are two nonconsecutive incorrectly transmitted characters. Since, the random events

S: the error of a_s is not detected;

R: the error of a_r is not detected;

are independent, we conclude that the errors in both characters are not detected with the probability v_1^2. The characters $d_{s\text{-}1}$, d_s, $d_{r\text{-}1}$ and d_r are transmitted such that these errors are not detected, and the rest $2n$-6 characters are correctly transmitted, so we have that

$$P(B_2)=(C_n^2-C_n^1)v_1^2v_0^{2n\text{-}6}=\frac{n(n-3)}{2}v_1^2v_0^{2n\text{-}6} \quad (6)$$

Using (4), (5) and (6), we obtain that

$$P(A_2)=nv_2v_0^{2n\text{-}5}+\frac{n(n-3)}{2}v_1^2v_0^{2n\text{-}6} \quad (7)$$

Further on, let calculate the probability $P(A_3)$. For that goal we introduce the following random events:

C_1: three consecutive characters of the initial message $a_1a_2\ldots a_n$ are incorrectly transmitted and the errors are not detected;

C_2: two consecutive characters and one character which is not consecutive with them (of the initial message $a_1a_2\ldots a_n$) are incorrectly transmitted and the errors are not detected;

C_3: three characters (such that each two of them are not consecutive) of the initial message are incorrectly transmitted and the errors are not detected;

Then, we conclude that

$$A_3 = C_1 + C_2 + C_3 \quad (8)$$

Now, on the same way as previous, we find that

$$P(C_1)=C_n^1\,v_3v_0^{2n\text{-}7}=nv_3v_0^{2n\text{-}7} \quad (9)$$

$$P(C_2)=C_n^1\cdot C_{n-4}^1\,v_2v_1v_0^{2n\text{-}8}=n(n\text{-}4)\,v_2v_1v_0^{2n\text{-}8} \quad (10)$$

$$P(C_3)=(C_n^3-C_n^1-C_n^1\cdot C_{n-4}^1)\,v_1^3v_0^{2n\text{-}9}$$
$$=\frac{n(n-4)(n-5)}{6}v_1^3v_0^{2n\text{-}9} \quad (11)$$

From (8), (9), (10) and (11), we find the probability of the event A_3.

$$P(A_3)=nv_3v_0^{2n\text{-}7}+n(n\text{-}4)v_2v_1v_0^{2n\text{-}8}$$
$$+\frac{n(n-4)(n-5)}{6}v_1^3v_0^{2n\text{-}9} \quad (12)$$

Finally, let calculate the probability $P(A_4)$. Similarly as previous we denote the random events:

D_1: four consecutive characters of the initial message $a_1a_2\ldots a_n$ are incorrectly transmitted and the errors are not detected;

D_2: three consecutive characters and one nonconsecutive character with them (of the initial message $a_1a_2\ldots a_n$) are incorrectly transmitted and the errors are not detected;

D_3: four characters of the initial message are incorrectly transmitted such that they form two pairs of two consecutive characters and the pairs are nonconsecutive and the errors are not detected;

D_4: four characters of the initial message are incorrectly transmitted such that two of them are consecutive and other two are nonconsecutive with previous two and nonconsecutive between each other and the errors are not detected;

D_5: four nonconsecutive characters of the initial message are incorrectly transmitted such that each two of them are not consecutive.

Now, we have that

$$A_4 = D_1 + D_2 + D_3 + D_4 + D_5 \tag{13}$$

$$P(D_1) = C_n^1\, v_4 v_0^{2n-9} = n v_4 v_0^{2n-9} \tag{14}$$

$$P(D_2) = C_n^1 \cdot C_{n-5}^1\, v_3 v_1 v_0^{2n-10} = n(n-5) v_3 v_1 v_0^{2n-10} \tag{15}$$

$$P(D_3) = \frac{C_n^1 \cdot (C_{n-4}^1 - 1)}{2} v_2^2 v_0^{2n-10} = \frac{n(n-5)}{2} v_2^2 v_0^{2n-10} = \frac{n(n-5)}{2} v_2^2 v_0^{2n-10} \tag{16}$$

$$P(D_4) = C_n^1 \cdot (C_{n-4}^2 - C_{n-4}^1 + 1)\, v_2 v_1^2 v_0^{2n-11} = \frac{n(n-5)(n-6)}{2} v_2 v_1^2 v_0^{2n-11} \tag{17}$$

$$P(D_5) = [C_n^4 - C_n^1 - C_n^1 \cdot C_{n-5}^1 - \frac{C_n^1 (C_{n-4}^1 - 1)}{2} - C_n^1 \cdot (C_{n-4}^2 - C_{n-4}^1 + 1)] v_1^4 v_0^{2n-12} \tag{18}$$

$$= \frac{n(n-5)(n-6)(n-7)}{24} v_1^4 v_0^{2n-12}$$

Using (13), (14), (15), (16), (17) and (18), we obtain that

$$P(A_4) = n v_4 v_0^{2n-9} + n(n-5) v_3 v_1 v_0^{2n-10} + \frac{n(n-5)}{2} v_2^2 v_0^{2n-10} + \frac{n(n-5)(n-6)}{2} v_2 v_1^2 v_0^{2n-11} + \frac{n(n-5)(n-6)(n-7)}{24} v_1^4 v_0^{2n-12} \tag{19}$$

Now, using the probabilities obtained in (2),(3), (7), (12) and (19), we find that

$$f(n,p) = n v_1 v_0^{2n-3} + n v_2 v_0^{2n-5} + \frac{n(n-3)}{2} v_1^2 v_0^{2n-6} + n v_3 v_0^{2n-7} + n(n-4) v_2 v_1 v_0^{2n-8} + \frac{n(n-4)(n-5)}{6} v_1^3 v_0^{2n-9} + n v_4 v_0^{2n-9} + n(n-5) v_3 v_1 v_0^{2n-10} + \frac{n(n-5)}{2} v_2^2 v_0^{2n-10} + \frac{n(n-5)(n-6)}{2} v_2 v_1^2 v_0^{2n-11} + \frac{n(n-5)(n-6)(n-7)}{24} v_1^4 v_0^{2n-12}, \quad \text{for } n \geq 5 \tag{20}$$

For $n<5$, we calculate the formulas by the same way.

The probability that more than 4 characters of the initial message will be incorrectly transmitted and the errors will not be detected can be presented as a sum $\sum_{k=7}^{s} a_k p^k$. We will denote this sum as $O(p^7)$. Therefore, for $n \geq 5$, the probability that there will be an error which will not be detected can be presented by $f(n,p) + O(p^7)$. Since $O(p^7)$ is very close to 0 when p takes small values, we will use the approximate formula $f(n,p)$.

At the end we have to determine the probabilities v_k. Let S_k denote the random event that errors will not be detected if exactly k consecutive characters of the initial message $a_1a_2\ldots a_n$ are incorrectly transmitted, i.e. the characters a_i, $a_{i+1},\ldots, a_{i+k-1}$ are incorrectly transmitted, but a_{i-1} and a_{i+k} are correctly transmitted, k=1,2,3,4. Then

$$v_k = P(S_k), \quad k=1,2,3,4. \tag{21}$$

At first, we calculate v_1. Let denote

H_j: $a_i = j,\ j = 0,1,2,3$ (the real value of a character before passing through the channel is j);

B_j: $a_i' = j,\ j = 0,1,2,3$ (the transmitted value of incorrectly transmitted character is j);

D_k: $d_k' = a_k' * a_{(k \bmod n)+1}'$.

Then,

$$v_1 = \sum_{j=0}^{3} P(S_1 \mid H_j) P(H_j) \tag{22}$$

$$P(S_1 \mid H_j) = \sum_{\substack{k=0 \\ k \neq j}}^{3} P(B_k \mid H_j) P(D_{i-1}) P(D_i) \tag{23}$$

Applying in (22) that $v_1 = P(S_1)$ and $P(S_1|H_i) = P(S_1|H_j)$, for all i, j = 0,1,2,3, we obtain that $P(S_1) = P(S_1|H_0)$, i.e.

$$v_1 = P(S_1|H_0) \tag{24}$$

Now, let calculate v_2. Denote

H_{st}: $a_i = s,\ a_{i+1} = t, \quad s,t = 0,1,2,3$ (the real value of a_i is s, and the real value of a_{i+1} is t),

B_{mn}: $a_i' = m,\ a_{i+1}' = n, \quad m,n = 0,1,2,3$ (the transmitted value of a_i is m, and the new value of a_{i+1} is n)

Then, we have

$$v_2 = \sum_{s=0}^{3} \sum_{t=0}^{3} P(S_2 \mid H_{st}) P(H_{st}) \tag{25}$$

$$P(S_2 \mid H_{st}) = \sum_{\substack{m=0 \\ m \neq s}}^{3} \sum_{\substack{n=0 \\ n \neq t}}^{3} P(B_{mn} \mid H_{st}) P(D_{i-1}) P(D_i) P(D_{i+1}) \tag{26}$$

Since $P(S_2|H_{st}) = P(S_2|H_{kl})$ for all $s,t,k,l \in \{0,1,2,3\}$, we can take

$$v_2 = P(S_2|H_{00}) \tag{27}$$

Concluding on the same way, we find that $v_3 = P(S_3|H_{000})$ and $v_4 = P(S_4|H_{0000})$.

Now, using the Theorem 1 and formulas for the probabilities v_k, functions $f(n,p)$ can be determined for all 160 fractal quasigroups. These 160 quasigroups do not define 160 different functions for the probability of undetected errors, but only 7. These functions are given in Figure 2.

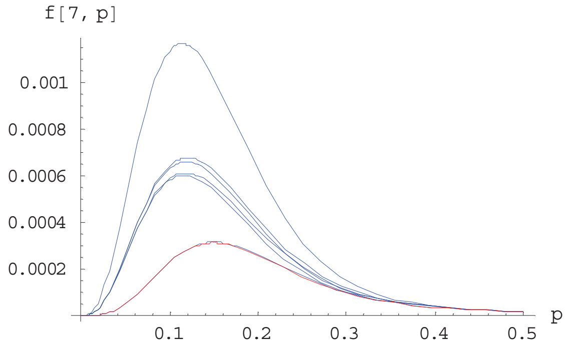

Figure 2: Seven different functions of probability of undetected errors

The quasigroups which give the smallest probability of undetected errors are the best for code design. If we use these quasigroups for coding, the probability that an undetected error occurs will be determined by the following formulas:

$$g(2,p) = 8(1-p)^4p^4 + p^8$$

$$g(3,p) = 15(1-p)^8p^4 + 33(1-p)^6p^6 + 14(1-p)^4p^8 + p^{12}$$

$$g(4,p) = 16(1-p)^{12}p^4 + 48(1-p)^{10}p^6 + 126(1-p)^8p^8 + 48(1-p)^6p^{10} + 16(1-p)^4p^{12} + p^{16}$$

$$\begin{aligned} g(n,p) = {} & 3np^4(1-p)^{2(2n-2)} + 9/2n(n-3)p^8(1-p)^{2(2n-4)} \\ & + 9/2n(n-4)(n-5)p^{12}(1-p)^{2(2n-6)} \\ & + 27/8n(n-5)(n-6)(n-7)p^{16}(1-p)^{2(2n-8)} \\ & + np^6(1-p)^{2(2n-6)}(3p^2-4p+2)(9p^4-20p^3+18p^2-8p+2) \\ & + 3n(n-5)p^{10}(1-p)^{2(2n-8)}(3p^2-4p+2)(9p^4-20p^3+18p^2-8p+2) \\ & + np^4(1-p)^{2(2n-4)}(9p^4-16p^3+12p^2-4p+1) \\ & + 3n(n-4)p^8(1-p)^{2(2n-6)}(9p^4-16p^3+12p^2-4p+1) \\ & + 9/2n(n-5)(n-6)p^{12}(1-p)^{2(2n-8)}(9p^4-16p^3+12p^2-4p+1) \\ & + 1/2n(n-5)p^8(1-p)^{2(2n-8)}(9p^4-16p^3+12p^2-4p+1)^2 \\ & + np^8(1-p)^{2(2n-8)}(81p^8-432p^7+1060p^6-1548p^5+1475p^4 \\ & -944p^3+400p^2-104p+13) + O(p^7), \end{aligned}$$

for $n \geq 5$ (25)

Let denote

$$\begin{aligned} f(n,p) = {} & 3np^4(1-p)^{2(2n-2)} + 9/2n(n-3)p^8(1-p)^{2(2n-4)} \\ & + 9/2n(n-4)(n-5)p^{12}(1-p)^{2(2n-6)} \\ & + 27/8n(n-5)(n-6)(n-7)p^{16}(1-p)^{2(2n-8)} \\ & + np^6(1-p)^{2(2n-6)}(3p^2-4p+2)(9p^4-20p^3+18p^2-8p+2) \\ & + 3n(n-5)p^{10}(1-p)^{2(2n-8)}(3p^2-4p+2)(9p^4-20p^3+18p^2-8p+2) \\ & + np^4(1-p)^{2(2n-4)}(9p^4-16p^3+12p^2-4p+1) \\ & + 3n(n-4)p^8(1-p)^{2(2n-6)}(9p^4-16p^3+12p^2-4p+1) \\ & + 9/2n(n-5)(n-6)p^{12}(1-p)^{2(2n-8)}(9p^4-16p^3+12p^2-4p+1) \\ & + 1/2n(n-5)p^8(1-p)^{2(2n-8)}(9p^4-16p^3+12p^2-4p+1)^2 \\ & + np^8(1-p)^{2(2n-8)}(81p^8-432p^7+1060p^6-1548p^5+1475p^4 \\ & -944p^3+400p^2-104p+13), \end{aligned}$$

for $n \geq 5$, (26)

and $f(n,p)=g(n,p)$ for n=2,3,4.

Now, for n≥5, $g(n,p) = f(n,p) + O(p^7)$, so further on we will work with the function $f(n,p)$. If we draw at the same graphic the functions $f(n,p)$ for n=6, n=7, n=8 and n=9, we will have:

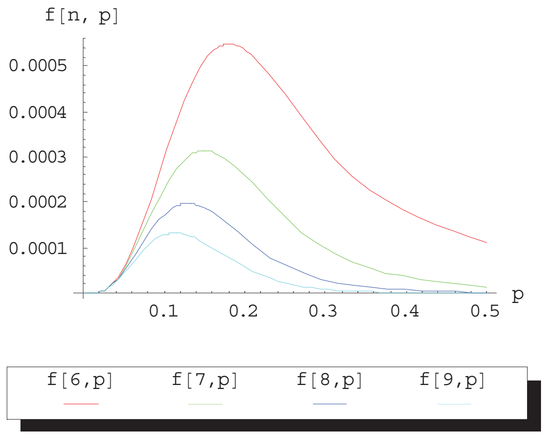

Figure 3

From the Figure 3 we can see that when the message length increases, the probability that there will be an error which will not be detected decreases. Respectively, the maximums of the functions decrease.

Now, we are creating a table with the maximums of the functions:

n	max of $f(n,p)$	n	max of $f(n,p)$
2	0.0585938	21	8.77182×10^{-6}
3	0.0153809	22	7.5905×10^{-6}
4	0.00389099	23	6.61231×10^{-6}
5	0.00112535	24	5.79537×10^{-6}
6	0.00054822	25	5.10775×10^{-6}
7	0.000313395	26	4.52483×10^{-6}
8	0.000197063	27	4.02739×10^{-6}
9	0.000132311	28	3.60032×10^{-6}
10	0.0000932542	29	3.23159×10^{-6}
11	0.0000682458	30	2.91157×10^{-6}
12	0.0000514707	31	2.63246×10^{-6}
13	0.0000397896	32	2.38794×10^{-6}
14	0.0000314013	33	2.1728×10^{-6}
15	0.0000252198	34	1.98275×10^{-6}
16	0.0000205631	35	1.81425×10^{-6}
17	0.0000169878	36	1.66432×10^{-6}
18	0.0000141968	37	1.53047×10^{-6}
19	0.000011986	38	1.4106×10^{-6}
20	0.000010212	39	1.30293×10^{-6}

Table 1

Now, we can make the probability of undetected errors arbitrary small. Namely, if we want the probability of undetected errors to be smaller then some previous given value ε, we will read from the table for which values of n the maximum of the function $f(n,p)$ is smaller then ε. Because of the nature of the functions $f(n,p)$, there will be $n_0 \in N$, such that the maximums of the functions $f(n,p)$ will be smaller than ε, for all $n \geq n_0$ and the maximums of the functions $f(n,p)$ will be greater than ε, for all $n < n_0$. We choose $n=n_0$. Since the maximum of $f(n,p)$ is smaller of ε, than $f(n,p)<\varepsilon$ for all

$p \in (0,1/2)$. Now, we separate the message in blocks with length n and we are coding every block individually.

From all values of n for which the condition $f(n,p)<\varepsilon$ is satisfied, we choose the smallest one since in this case we have fastest transmission. Namely, if the receiver detects errors in the received block, it asks for repeated transmission, it is better block length to be as small as possible.

REFERENCES

[1] S. Markovski, V. Bakeva, "Error-detecting codes with cyclically defined redundancy", *Procedings of 3-th Congress of Mathematicians of Macedonia, 2005*.

[2] S. Markovski, V. Bakeva, "Quasigroup Based Stream Error Correcting Codes", Proceedings of the 2nd CIIT, Molika, R.Macedonia, (2001).

[3] S.A. Vanstone, P.S. van Ooschot, "An Introduction to Error Correcting Codes with Applications", Kluwer Academic Publishers, Boston, 1989.

[4] R. Hill, "A First Course of Coding Theory", Clarendon Press, Oxford, 1986.