Coding for information systems security and viability

© Bohdan Zhurakovskyi 1[0000-0003-3990-5205], © Serhii Toliupa 2[0000-0002-1919-9174],
© Serhiy Otrokh 1[0000-0001-9008-0902], © Valeriy Kuzminykh 1[0000-0002-8258-0816],
© Hanna Dudarieva 3[0000-0002-9887-021X] and © Vladislav Zhurakovskyi 4[0000-0002-6510-9969]

[1] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
[2] Taras Shevchenko National University of Kyiv, Ukraine
[3] State University of Telecommunications, Kyiv, Ukraine
[4]National Aviation University, Kyiv, Ukraine

zhurakovskiybyu@tk.kpi.ua, tolupa@i.ua,
2411197@ukr.net, vakuz0202@gmail.com,
Annett.13.86@gmail.com, vbzh116@gmail.com

Abstract. In this article, models of discrete channels are analyzed. It is proposed to use the L.P. Purtov for building channels of information systems. Redundant binary and non-binary codes and their error-correcting capability are analyzed. Formulas are derived for finding the probability of an uncorrected error for error-correcting codes, for codes that find errors, and for codes that detect and correct errors. The coefficient of increasing the information array is calculated for different probabilities of the detection of one element of the codeword in the channel and different error rates. Calculation of the redundancy of the error burst correction codes is performed. The use of Fire codes as anti-noise codes in the channels of the information system is proposed.

Keywords: error-correcting codes, error bursts, discrete channel model, code redundancy.

1 Introduction

The issue of transmitting discrete information in information systems and networks in the case when it is impossible or unprofitable from an economic point of view to use a reverse communication channel to increase the reliability of messages is of interest to a wide range of specialists in the field of security, wire and radio communications. A large number of articles and monographs are devoted to this topic. However, most of the publications are devoted to the coverage of information channels in which binary correction codes are used as error-correcting codes. This is mainly due to the simplicity of operations in the binary number system, which are performed by various computer devices in the processing, accumulation and storage of information in information systems and networks.

The use of combined types of modulation in information channels allows to significantly increase the data transmission rate in comparison with binary systems. And the use of error-correcting codes with $q > 2$ alphabet, which allow you to identify and correct errors, in such systems is quite a logical step. Such codes are called Non-binary (multi-positional, multi-base or q-ary) [3, 4, 5, 7, 8, 15, 16]. Despite the fact that some redundant nonbinary codes, such as the generalized Hamming code, the Reed-Solomon code, and others, have been widely used in information channels [6, 9, 10, 11, 12, 13, 14], the development issues of the nonbinary coding theory remain quite relevant.

2 Statement of research problem

2.1 Discrete channel mathematical model

The description of the channel consists of determining the restrictions on the signals $S(t)$, the transmission of which it provides, and the nature of the transformation $S(t)\to S^*(t)$, which it performs. For discrete channels, it is enough to indicate the allowed signals (symbols) or simply their number m and the allowed moments of signal change. For m = 2 (such channels are called binary), the allowed times of signal change (significant moments) are completely determined by the modulation rate. The transformation $S(t) \to S^*(t)$ in the ideal case reduces to the equality $S(t)=S^*(t)$. The difference arises from the non-improvement of the communication line. The transformation $S(t)\to S^*(t)$ has two components - deterministic and random. Deterministic changes in the signal in a discrete channel are delayed (time offset). The resolving device is a threshold device, it emits signals of a fixed form, so no signal changes are observed. Random signal changes are caused by interference.

When converting symbols in a discrete channel, the received sequence of message elements B_i differs from the transmitted one B_i. In the general case, the probabilistic laws of transformation $B_i \to B_i^*$ are very complex due to the complex nature of the interference. The influence of noise can be represented as an elemental summation of a sequence with a sequence of errors. Then $B_i^* = B_i + E_i$.

In binary channels, two categories of output elements can be distinguished: unaffected and affected. In the simple case, the beginning of the affected elements can be obtained by comparing the received sequences with the transmitted ones. This comparison, as a rule, is performed not on an infinite, but on a finite segment of sequences (code combination, block, group of blocks).

Errors can be distinguished by type (transition 1 → 0 or 0 → 1) and multiplicity t. If $p(1 \to 0) = p(0 \to 1)$, the errors are considered symmetric. The multiplicity (number of errors) is the number of errors t, which accounted for a given number of unit elements: errors can be single, double, etc.

From the point of view of the influence of interference on the fidelity of the transmission, the channels are divided into symmetrical and asymmetrical, channels with memory and without memory, stationary and non-stationary. In a symmetric channel, the transmission probability does not depend on the statistics of the transmitted sequence, but is completely determined by the error statistics. A stationary channel without memory is called one in which the probability of an error in receiving an element does not depend on the number of an element in the transmitted sequence and does not depend on the values (zero or one) of the elements preceding or following it.

The probabilistic model of a discrete channel is called the final description of the channel, which allows, with known probabilities of the input signal, to find expensive probabilities of the characteristics of a discrete channel without resorting to experimental data or additional assumptions. Since statistics in a symmetric channel are completely determined by the error statistics and do not depend on statistics, an error stream is sufficient to describe it.

For independent errors, it is enough to know the only parameter p_{int} to find the distributions of the random variable. Based on Bernoulli's theorem, the probability of occurrence in an n-element combination is equal to t errors $P(t,n)$ is found by the binomial distribution

$$P(t,n) = C_n^t p_{int}^t (1-p_{int})^{n-1}, \quad (1)$$

for $0 < t < n$ it is seen that the probability of receiving an undistorted combination $(t=0)$ the above equation becomes:

$$P(0,n) = (1-p_{int})^n. \quad (2)$$

In the case of the probability of receiving a false combination, that is, a combination that contains at least one error, the above equation becomes:

$$P(\geq 1,n) = 1 - P(0,n) = (1-p_{err})^n, \quad (3)$$

In the case of the probability of occurrence of m or more errors in a combination of length n, the above equation becomes:

$$P(\geq m,n) = \sum_{t=m}^{t=n} C_n^t \, p_{int} (1-p_{int})^{n-1}. \quad (4)$$

A simple error model gives only an approximate state in real channels. Further investigation of the stream of errors in real channels showed that errors in communication channels are grouped. The probability of an error in a group (pack) of errors increases sharply and is significantly greater than p_{int}.

To describe the probabilistic laws of a sequence of binary elements in complex mathematical models of a channel, it is convenient to use the statistics of elementary channel states, in which the channel is described by several possible states. The number of channel states and conditional error probabilities ε in each of the states can be different. In the very first approximation, a channel can have two states - "good" and "bad". State statistics have two implementations. Let us denote the probabilities of finding a channel in a "good" state $P(D=0)=P_0$, and in a "bad" state $P(D=1)=P_1$. Obviously, $P_0+P_1=1$. Let there be possible independent errors with conditional probabilities ε_0 and ε_1 in both channel states. Then the error probabilities are calculated by the formula:

$$p_{int} = P_0\varepsilon_0 + P_1\varepsilon_1. \quad (5)$$

For $\varepsilon_0 = 0$ and $\varepsilon_1 = 1$, the statistics coincide with statistics.

The mathematical model of the channel should provide the engineering ability to calculate the basic characteristics [1], the knowledge of which may be required when evaluating various methods of increasing the reliability of the transmission [2] of discrete information. These characteristics primarily include: the probability of an error (incorrect reception of a single element) p_{int}; distribution of intervals between errors, as well as distribution of lengths of series of errors, bursts of errors; distribution of probabilities $P_n(t)$ - occurrence of t errors in a block of information of length *n*.

2.2 Discrete Channel Models with Error Bursts

As a general mathematical model of an information system channel, the model of L.P. Purtov is used most of all. Models with Markov chains require a more accurate description of the initial parameters and a large amount of computation, so they are used to describe specific channels. L.P. Purtov's model describes the probability of a multiplicity error equal to or greater than *t* in a block of length *n* elements, depending on the error probability in the *p* element and the error grouping coefficient α:

$$P(N_k) = \left(\frac{N_k}{t}\right)^{1-\propto}, \tag{6}$$

where N_k is the length of the codeword.

The application of the model is limited by the values of the error rate $< \frac{N_k}{3}$, which are typical for most channels.

The error grouping factor varies within 0.5 ... 0.7 - for cable and 0.2 ... 0.4 - for radio channels. Limit values: $\propto = 0$ - independent errors and $\propto = 1$ - all errors are collected in one batch.

Let us investigate the dependences $P(t)$ of the error probability of multiplicity t in the code word using the above model. Let the code combination have $N_k = 128$ bits with the probability of an independent error $p = 10^{-6}$ and the error grouping factors $\propto =$ 0.5; 0.6 and 0.7. Figure 1 shows graphs of these dependencies.

Figure 1 shows that, firstly, with an increase in the error grouping coefficient, all other things being equal, the probability of an error of the multiplicity t decreases, and this pattern remains in the entire field of application of the L.P. Purtov.

In other cases, the probability of an error of multiplicity t is always greater than the probability of an error of multiplicity $t + 1$. And finally, with an increase in the error grouping coefficient, the difference between the probabilities of errors of multiplicity t and $t + 1$ changes.

Analyzing the above figure, we can conclude that in real transmission channels of control information, errors tend to be grouped. In this case, there is a tendency, for a given number of bits of information ($N_k = 128$), the greatest probability of occurrence of a burst of errors is when the length of this burst is $t \rightarrow 5$. So, suppose that the multiplicity of the error will be 4. To correct this error, it will be necessary apply noise immunity code, will have a large correcting ability, with little redundancy.

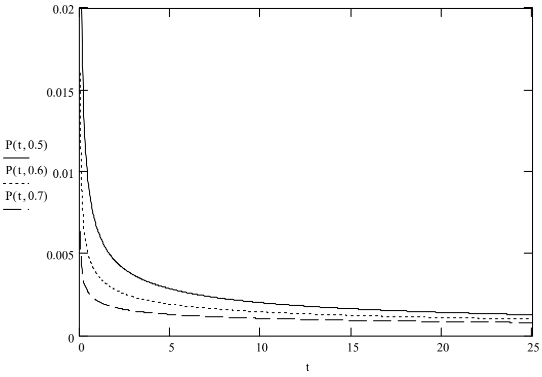

Fig. 1. Dependence *P* (*t*) in channels with grouping errors.

2.3 Discrete Channel Models with Error Bursts

There are such coding methods that increase the efficiency of information transmission: optimal coding, error detection coding and error correction coding [3].

Error detection and correction codes increase the noise immunity of information transmission.

To select the type of correction code that is used for secondary coding, consider the error rate and the nature of their distribution in time.

A careful distribution of the statistical characteristics of the sequence of errors in real communication channels showed that the errors are dependent and have a tendency to grouping (batching), that is, there is a correlation between them.

Most of the time, information on communication channels passes without interference, and at some points in time there are bursts of errors, within which the error probability turns out to be much higher than the average error probability, which is found for a sufficiently long transmission time interval.

With the group nature of the distribution of errors, one parameter - the probability of error - does not fully characterize the channel; additional parameters are needed that reflect the degree of grouping of errors in various types of channels.

The reasons for the group nature of errors in channels of various types are as follows: tropospheric radio links - fading and interference; radio relay lines - short-term interruptions and impulse noise; over cable communication lines - overload of group paths of high-frequency equipment, short-term interruptions and various kinds of station and operational interference.

For dedicated telephone cable channels, the error rate ranges from (2 ... 5) * 10^{-4} to 10^{-5}. In switching telephone channels $p = 10^{-3} \dots 10^{-4}$, depending on the length of the channel, the number and load of automatic telephone exchanges involved in the connection of subscribers. For radio relay and tropospheric channels $p = 10^{-3} \dots 10^{-4}$ [3].

In all real channels, the errors are very unevenly distributed over time. This is especially typical for cable channels, in which the bulk of errors are concentrated at certain hours of the day, responding mainly to an increase in the load of channel-forming equipment and the3 time of changing the duty of the service personnel.

The assessment of the reliability of information exchange, when designing information transmission systems, is determined by the permissible probability of interference in messages P_{int}, At the same time, the probability of interference of the binary symbol of the transmitted message Pbs is indicated., In the event of independent errors in the communication channels and (1- P_{bs}) - the probability of no interference. Then for a binary sequence, that has n characters [4], equation becomes:

— for the probability of a correctly received sequence:

$$P_{rec} = (1 - P_{bs})^n, \tag{7}$$

— for the probability of error in the received sequence:

$$P_{int} = 1 - (1 - P_{bs})^n. \tag{8}$$

When using redundant codes, it becomes possible to find and correct one or another error rate depending on the code distance:

$$P_{int}^i = C_n^i (1 - P_{bs})^{n-1}, \tag{9}$$

where i = 1,2,3, etc. is frequency of errors, and C_n^i determined from the formula:

$$C_n^i = \frac{n!}{i!(n-i)!} \tag{10}$$

For codes that correct errors of multiplicity to t, the probability of correction P_{cor} is:

$$P_{cor} = \sum_{i=1}^{t} P_{int}^i \text{ or } P_{cor} = \sum_{i=1}^{t} C_n^i \, P_{bs}(1 - P_{bs})^{n-1}. \tag{11}$$

Then four situations can be formulated:

1) The code word was received without errors (the probability of this event is P_{rec}).

2) The code word was received with an error (probability P_{int}).

$$P_{int} + P_{rec} = 1 \tag{12}$$

3) The code word was received with an error, which is corrected with the probability of P_{cor}.

4) The code word was received in error and cannot be corrected by this code. The probability of this event is P_{ncor}, then

$$P_{int} = P_{cor} + P_{ncor} \tag{13}$$

From formula (13) it follows that the probability of uncorrected errors:

$$P_{ncor} = P_{int} - P_{cor} \tag{14}$$

Let us substitute formulas (8) and (11) into this expression and obtain:

$$P_{ncor} = 1 - (1 - P_{bs})^n \text{-} \sum_{i=1}^{t} C_n^i P_{bs}(1 - P_{bs})^{n-1} \tag{15}$$

This expression allows you to calculate the probability of an uncorrectable error in transmitting information using a code that corrects t-fold errors.

In practice, communication channels are characterized by the dependence of the probability of interference of the next transmitted symbol on the distortion of the previous one, as well as seasonal and daily changes in the weather, the presence of industrial interference that changes the intensity at the time of the day and days of the week, mutual interference, etc. All this leads to an abrupt nature errors in information transmission channels.

The calculation of error probabilities for batch distribution is carried out according to the formula:

$$P_{int} = \frac{p_e}{q} \cdot \sum_{b=1}^{b_{max}} \left(1 + \frac{n-1}{b}\right) \cdot \frac{b \cdot p_b}{\sum_{b=1}^{b_{max}} b \cdot p_e} \tag{16}$$

where b is the length of the error packet; p_b - conditional probability of occurrence of a packet of errors of length b; p_e is the probability of distortion of a binary symbol; q is the density of errors in a packet, which is equal to the ratio of the number of errors in a packet to the length of this packet b.

Equation (16) determines the transmission error probability:

Probability $P_{ncor}$and $P_{cor}$for codes that determine bursts of errors, are calculated by the formulas:

$$P_{ncor} = \frac{1}{2^r} \cdot \frac{p_e}{q} \cdot \sum_{b=l_k}^{b_{max}} \left[1 + \frac{n-(2l_k+1)}{b}\right] \cdot \frac{b \cdot p_b}{\sum_{b=1}^{b_{max}} b \cdot p_e}; \tag{17}$$

$$P_{cor} = \frac{p_e}{q}\left[\sum_1^{b_{max}}\left(1 + \frac{n+1}{b}\right)\frac{bp_b}{\sum_1^{b_{max}} bp_e} - \frac{1}{2^r}\sum_{b=l_k}^{b_{max}}\left[1 + \frac{n-(2l_k+1)}{b}\right]\frac{bp_b}{\sum_1^{b_{max}} bp_e}\right]\frac{bp_b}{\sum_{b=1}^{b_{max}} bp_e}, \tag{18}$$

where l_k is the length of the error packet found.

Approximate formulas for determining P_{ncor} look like:

— for error-correcting codes:

$$P_{ncor} = \left(\frac{n}{t+1}\right)^{1-s}, \tag{19}$$

where t is the multiplicity of the error being corrected; s - error grouping indicator;

— for codes that find errors:

$$P_{ncor} = \frac{p_e}{2^r}\left(\frac{n}{d}\right)^{1-s}, \tag{20}$$

— for codes that detect and correct errors:

$$P_{ncor} = \frac{\sum_{i=0}^{t} C_n^i}{2^r}\left(\frac{n}{d-1}\right)^{1-s} p_e. \tag{21}$$

These formulas give good results if the number of errors e in combination with n symbols satisfying the requirements $e < 0.3n$ [6].

When transmitting information with a simple non-redundant code, the reliability of reception depends on the type of channel and the type of interference in it. In most cases, the reliability that is found is insufficient. It must be increased so that the probability of erroneous reception of the message by the consumer is less than the probability of errors in the message without special measures.

The use of redundant code is one of the ways to improve the reliability.

All redundant codes can be used for:

— identification of errors;
— error correction;
— identification and correction of errors.

In order to increase the reliability with the help of codes designed to detect errors, it is necessary to introduce a feedback channel. Then the code word received on the forward channel is analyzed to determine if it belongs to the allowed combination. The allowed combination comes to the consumer after the check bits are rejected. If an error is detected, a request signal is sent over the reverse channel, through which the transmitter repeats the transmission of information. Therefore, the transmitting device must store information about the sent signals for a period of time sufficient for the receiving device to analyze the combination and receive a possible error request.

Error correction is usually applied when there are independent errors or short bursts of errors in the communication channel. If the error weight is the same as the codeword length, then the error burst corrections lead to unjustified losses of equipment for encoders and decoders.

Error-correcting codes can correct errors, the weight of which numerically does not exceed 20-25% of the length of the code combination. Most probable errors with a weight close to 50% of the length of the code word. Therefore, if it is necessary to correct, those methods are appropriate that allow you to determine the verification pulses from the information for a time exceeding the probable length of the burst of errors [8].

Thus, the choice of a method for increasing the reliability of information transmission depends on many factors: the reliability of reception, the permissible transmission rate, and dependence on errors in the communication channel are necessary.

The degree of information protection from errors by the appropriate coding method depends mainly on the minimum coding distance d_{min} of the given code.

There are three types of code distance: Hamming, Lee and matrix. The first is most widely used in coding theory. The Hamming code distance is inextricably linked with the concept of the weight w of the code combination - the number of its elements that are not equal to zero. Hamming code distance d between two combinations of the same length n is defined as the number of similarly named bits (positions) that have unequal elements. So, for binary codes, since in binary arithmetic adding identical elements gives 0, and unequal ones - 1, the Hamming distance between two code combinations can be determined by adding them bitwise modulo 2 and then counting the number of nonzero elements, that is, determining the weight w of such a sum.

The total number of code combinations of length n is equal to 2n, and the number of those of them that are distant from the given in view d is the number of messages from n to d:

$$C_n^d = \frac{n!}{[d!(n-d)!]}. \tag{22}$$

To identify all elements with multiplicity, the code distance must be $d \geq \nu_d + 1$, and error correction with multiplicity $\nu_{cor} - d \geq 2\nu_{cor} + 1$. To correct and identify all errors, the condition must be satisfied:

$$d \geq \nu_{cor} + \nu_i + 1.. \tag{23}$$

Due to the fact that, on the whole, each element (bit) of a combination of a Non-binary (multi-position) code can have, in contrast to a binary, and more than one position *(m* $\geq$1) from the alphabet q, the code distance is determined by the expression:

$$d = \sum_{i=1}^{m} d_i \tag{24}$$

where m is the number of positions in each bit (single new interval, corresponding to the duration of one element) of the code combination.

In the Hamming metric, the code distance, as for a binary code, is determined by the number of bits of the same name with different positions (symbols):

$$d_i(x_k, x_l) = \begin{cases} 0, x_k = x_l; \\ 1, x_k \neq x_l. \end{cases} \tag{25}$$

In the Lie metric

$$d_i(x_k, x_l) = min\{|x_k - x_l|, q - |x_k - x_l|\}=min\{d_{jmod}, q - d_{jmod}\}, \qquad (26)$$

where $d_{jmod} = |x_k - x_l|$.

In modular metric $d_i(x_k, x_l) = |x_k - x_l|$, that is, you should perform subtraction modulo q.

The most rational should be considered information transmission systems in which redundant codes are used only for detecting errors, because in real channels there are often bursts of errors with a length of several tens or hundreds of symbols and to correct them one would need a code with a code combination length that is measured in thousands and tens of thousands of discharges, which is technically difficult to implement.

If the nature of the distribution of errors in a channel is known, this does not mean that such errors are present only in this channel, and not in the modem, which is used for research and can only change when the transmission rate, signal power, etc. The nature of the errors is highly dependent on the modem being used. Therefore, it is always necessary to consider the choice of modem and code as a single task and find the optimal solution.

After determining the characteristics of the communication channel, further selection of the code is characterized by the error probability. By this parameter, codes are selected in which the probability of detecting an error P_{ncor} less $P_{int.dop.}$.

To ensure a given information transfer rate, you need to choose a code with the minimum required number of check bits, which provide P_{ncor}. In this case, one should not forget that the detecting properties are determined not only by the number of check bits, but also by the type of check ratios, and for cyclic codes - a generating polynomial. Compared to other codes, the best code that not only detects but also corrects independent (single) errors and has recommendations of international organizations and is relatively simple to implement is Hamming code, both binary and generic. As for both packets and independent errors, the BCH, Fire and Reed-Solomon codes are the best for the same parameters. These codes can be used for the transmission the ring codes shift indexes vectors that are described in [17, 18]. But ring codes may have limited use for the transmission of service information.

Despite all the advantages of error-correcting coding, one of its disadvantages is an increase in the information array. Table 1 shows the coefficient of increasing the information array for the above defined optimal codes for various error probabilities of one element in the communication line (channel).

According to the OSI hierarchy, the control information transmission channel consists of physical and data link layers.

The physical layer of the channel ensures that the parameters of the signal are matched to such characteristics of the physical channel as the amplitude-frequency characteristic and the distribution of the noise power density, and must provide the specified constant component of the signal delay and the probability of an error in the binary bit of the transmitted digital stream, which, in turn, defines the requirements for the average bit rate. Building a physical layer subsystem is a multi-criteria optimization problem, the criteria of which are:

Table 1. Increase in the information array with an error rate equal to 2

	Codes			
	BCH	Compound	Reed-Solomon	Fire
Distortion probability of one element in a line	10^{-6}			
Coefficient of increasing the information array (at k = 8)	2.13	2.66	2.08	1.29
Distortion probability of one element in a line	10^{-5}			
Coefficient of increasing the information array (at k = 8)	2.28	2.33	2,17	1,59
Distortion probability of one element in a line	10^{-4}			
Coefficient of increasing the information array (at k = 8)	2.37	2.38	2.23	2.22
Distortion probability of one element in a line	10^{-3}			
The coefficient of increasing the information array (at k = 8)	2.46	2.56	2.29	2.35
Distortion probability of one element in a line	10^{-2}			
Coefficient of increasing the information array (at k = 8)	2.56	2.52	2.37	1.39

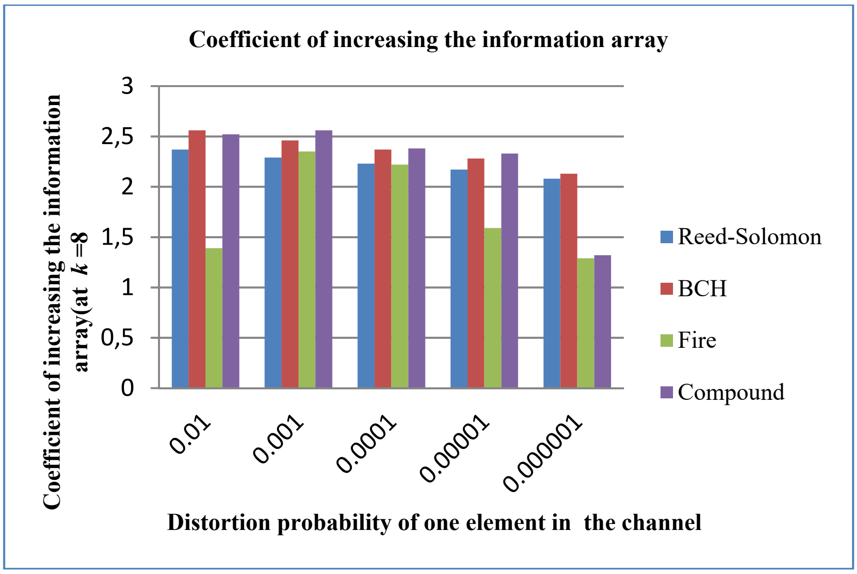

Fig. 2. Coefficient of increasing the information array

- constant value of the digital stream delay in the channel;
- the value of the error probability in a binary digit of a digital stream;
- average bit rate of digital stream; the cost of the physical layer subsystem.

The link layer provides the combination of control messages into a single stream for transmission to the physical layer, control of the error probability of the received information, control of the message delay in the channel depending on the probability of drinking, which is provided by the physical layer subsystem and the length of a specific control message transmitted by the upper layers of the network control system. The construction of the link level subsystem is also a multi-criteria optimization problem, the criteria of which are:

- variance of the control message delay value in the channel;
- the value of the error probability in the control message;
- average transmission rate of the message flow [2];
- cost of the link layer subsystem [1].

The random nature of the message delay in the channel and the error probability in the message are determined by the non-stationarity of the interference characteristics in most physical channels and, as a consequence, by the variable value of the bit error probability itself and the message processing time at the channel layer. There are possibilities to reduce the variance of the message error probability and the variance of the message delay in the channel:

- the use of correction codes with a large correction capacity (for example, convolutional ones, which leads to the introduction of significant redundancy, will not be effectively used when reducing the probability of an error in a bit and will significantly reduce the channel capacity;
- block coding when transmitting very large blocks, the length of which is so large that the probability of errors in a block differs very little from the average. It is widely known block codes with optimal redundancy for a given correction ability, however, coding in very large blocks leads to an unacceptable value for control information of a constant value of the digital stream delay in the channel and an increased complexity of correction code decoders.

The following restrictions are most typical for such a control information transmission channel:

- control information arrives at the channel input in the form of messages (blocks) with a length that is a multiple of 8 bits (due to the specifics of the development of a modern element base);
- the block nature of the control information determines the use of block codes (more often Bose-Chowdhury-Hawkingham (BCH) codes or their non-binary subclass - Reed-Solomon (RS) codes);
- the physical layer provides the error probability for the binary symbol y, but errors are grouped into packets.

Among the anti-jamming codes, we will choose a code, using which, the indicators of quality and survivability of the information system will increase. When choosing a code, it is necessary that the total number of characters is approximately the same. Then, comparing the redundancy of the codes with the redundancy of the BCH code, namely, it is used as a noise immunity code when transmitting information in information systems, where bursts of errors prevail.

For BCH codes to correct 4 errors, it is necessary that the code distance adheres to the condition $d_{min} = v_{cor} + v_{def} + 1$, therefore $d_{min} = 4 + 4 + 1 = 9$ the length n of the BCH code combination can be determined as follows:

$$n = 2^k - 1 \quad or \qquad n = \frac{2^{k-1}}{g}, \tag{27}$$

where $h > 0$ is integer; g is an odd positive number, when divided by which n becomes an odd number. Thus, the length n can only have an odd number of elements.

The number of verification code elements is determined by the expression:

$$r \leq \frac{h(d-1)}{2} = [log_2(n+1)]\frac{d-1}{2}, \tag{28}$$

and the number of information items - by the expression:

$$k \geq (2^h - 1) - \frac{h(d-1)}{2} \quad or \;\; k = n - r. \tag{29}$$

Thus, $n = 127 = 2^7 - 1 = 127$, $r = [log(127+1)]\frac{9-1}{2} = 28,$ $k = 127 - 28 = 99$. Code redundancy $R = 28/127 = 0{,}22$.

For the Fire code, we find h, provided that the length of the error burst is 4. $h = 2^4 - 1 = 15$. For this situation, the generating polynomial for the Fire code is $P_F(x) = (x^4 + x + 1)(x^7 + 1) = x^{11} + x^8 + x^7 + x^4 + x + 1$, then the degree of the irreducible polynomial P(x) is l = 4, $c = 7$. Referring to formulas (3.12), (3.13), (3.14), we find n, r, k.

n = HCK (7, 15) = 7·15 = 105; r = 4 + 7 = 11; k = 105 - 7 - 4 = 94. Redundancy of the code $R = 11/105 = 0{,}1$.

It follows from this that it is much easier to correct four errors that are in the same place than the same four errors, which are randomly distributed over the entire length of the combination.

Table 2. Redundancy of error burst correction codes

Codes	Code distance, d					
	5	7	9	11	13	15
Compound	0,23	0,165	0,146	0,147	0,183	0,22
Fire	0,24	0,165	0,1	0,05	0,084	0,055
BCH	0,323	0,28	0,22	0,203	0,197	0,168
Reed-Solomon	0,22	0,126	0,118	0,078	0,197	0,082

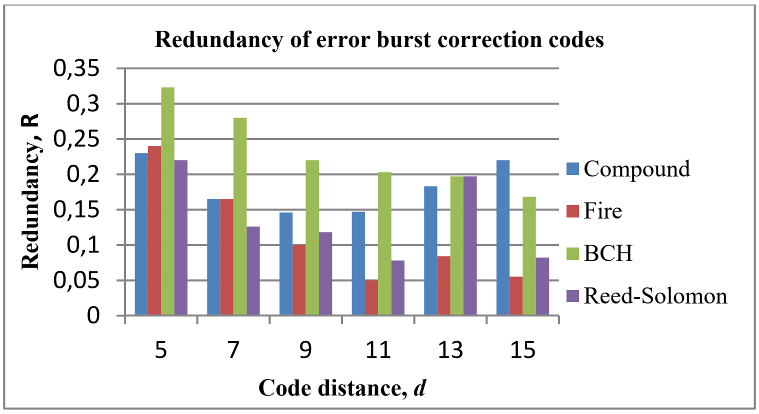

Fig. 3. Dependence of the code redundancy on the code distance

3 Conclusions

After analyzing the data obtained, we can conclude that the Fire code provides a lower probability of an uncorrected error than the BCH, compound, or Reed-Solomon codes at the same coding distance, that is, these codes require greater message redundancy to provide the same probability uncorrected error as the Fire code.

So, after considering the codes that allow you to correct single bursts of errors, you can propose the use of the Fire code instead of the BCH or Reed-Solomon code, since it has less redundancy compared to other noise immunity codes that correct bursts of errors.

References

1. B.Zhurakovskiy, N.Tsopa Assessment Technique and Selection of Interconnecting Line of Information Networks. —IEEE, 3rd International Conference on Advanced Information and Communications Technologies (AICT). —2019. — pp. 71-75.
2. B. Zhurakovskyi, Y. Boiko, V. Druzhynin, I.Zeniv, O.Eromenko. Increasing the efficiency of information transmission in communication channels. Indonesian Journal of Electrical Engineering and Computer Science. – 2020. – Vol 19, №3 - C. 1306–1315. DOI: http://doi.org/10.11591/ijeecs.v19.i3.pp1306-1315
3. Жураковський Ю. П. Полторак В. П. Теорія інформації та кодування. – К.: «Вища школа», 2001. – 256 с.
4. Березюк Н. Т. Кодирования информации. – Харьков: «Вища школа», 1978. – 252 с.
5. Richard E. Blahut Theory and Practice of Error Control Codes Hardcover – January 1, 1984
6. Berrou C, Glavieux A., "Near Optimum Error Correcting Coding and Decoding: Turbo-Codes", IEEE Trans. On Comm., Vol. 44, No. 10, pp. 1261-1271, Oct. 1996.
7. Y. Wu. "Implementation of parallel and serial concatenated convolutional codes." Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University. April, 2000. - 206 p.

8. Берликэмп Э. Алгебраическая теория кодирования. Пер.с англ.- М.: Мир, 1972. – 478 с.
9. J. Hagenauer, "The turbo principle: Tutorial introduction and state of the art," in Proc. of The Int. Symp. on Turbo Codes and Related Topics (Brest, France), Sept., 1997. - pp. 1-11.
10. P. Jung, J. Plechinger, "Performance of rate compatible punctured Turbo-codes for mobile radio applications," Electronics Lettes, 1997, vol. 33, No.25, pp. 2102-2103.
11. G. Caire and E Biglieri, "Parallel concatenated codes with unequal error protection," IEEE Transactions on Communications, vol. 46, No. 5, May 1998, pp. 565-567.
12. P. Robertson, K. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain," in Proc. IEEE Int. Conf on Commun., 1995. - pp. 1009- 1013,
13. Tilavat, V., & Shukla, Y. (2014). Simplification of procedure for decoding Reed–Solomon codes using various algorithms: an introductory survey. International Journal of Engineering Development and Research, 2(1), 279-283.
14. Sarwate, D. V., & Morrison, R. D. (1990). Decoder malfunction in BCH decoders. Information Theory, IEEE Transactions on, 36(4), 884-889.
15. Richard E. Blahut, "Algebraic Codes for Data Transmission", 2003, chapter 7.6 "Decoding in Time Domain"
16. Lin, S. J., Chung, W. H., & Han, Y. S. (2014, October). Novel polynomial basis and its application to reed-solomon erasure codes. In Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on (pp. 316-325). IEEE.
17. Otrokh S., Kuzminykh V., Hryshchenko O.. Method of forming the ring codes// CEUR WorkshopProceedings - Vol-2318, - 2018, pp. 188-198.
18. Berkman, L., Otrokh, S., Kuzminykh, V., Hryshchenko O. Method of formation shift indexes vector by minimization of polynomials// CEUR Workshop Proceedings -Vol-2577, - 2019, pp. pp. 259-269.