Usage of Linear Erasure Codes for Increasing Reliability and Efficiency of Information Delivery on the Internet

Nikolaos G. Bardis, Oleksandr P. Markovskyi and Kostiantyn V. Koliada

***Abstract*—**A new method of redundant packet formation for data transmission in global networks, as well as a technology of their usage for the retrieval of lost and damaged data packets is proposed. Technical justification of the proposed method is given. The developed technologies are illustrated via examples. It is proved that the proposed method can increase the reliability of data transmission in global networks.

***Keywords*—** erasure codes, reconstruction packet in network, linear codes.

I. Introduction

OVER the last decade, in all areas of human activity, there has been a steadily upward trend in Internet dominance in the field of information processing. In particular, there is a clearly visible gradual replacement of traditional television and a change of the way of operation of telephone communication with the technology of global networks. In addition, the process of expanding the Internet to new areas of application continues.

The technological breakthrough, caused by the emergence of cheap radio modems, has led to widespread usage of the Internet as a tool for data transmission in remote control systems and real-world objects management in real time [1]. The usage of the Internet in real-time systems puts forward qualitatively more strict requirements for the efficiency and reliability of data transmission. This means that the specificity of a particular task of applying a global network technology determines the critical time limit, to which the receiver must obtain information. Internet technology provides the transfer of information by packets by different routes depending on traffic. Accordingly, packets of an information message can be delivered at different times. One way to ensure a given efficiency of data delivery through a global network is to use redundant packets that are sent via different routes in order to reconstruct those that are late, thus achieving reception of the entire information message before the deadline. This requires the creation of an effective technology for the formation of redundant packets and the development of recommendations for determining their number.

One of the priority areas for the development of global networks is the usage expansion of peer-to-peer technologies [2], [14]. This technology provides the connection at a logical level of a certain set of physical network devices. In addition, it allows the nodes of the network function flexibly, that is, they can act as both receivers and transmitters. This provides the opportunity to reduce duplication of broadcast information transmission and, accordingly, reduce traffic on global network. Data transmission through non-controlling device systems requires reconfiguration with every exclusion of each one of them. On the other hand, such an uncontrolled shutdown of the device, which in fact plays the role of the server and through which the data transfer is carried out, can lead to loss of part of the packets.

A possible solution to this problem is to use redundant packets that are transmitted over the main network and provide the ability to reconstruct lost packets.

Thus, the scientific task of increasing the efficiency of redundancy and reconstruction of data transmitted over global networks is relevant in the context of the modern stage of development of computer and network technologies.

II. A Review and Analysis of Known Technology for Troubleshooting Network Errors

The problem of the data transmission reliability in global networks has a leading place in network technologies since the very beginning of their development. Dealing with data transmission errors in networks has a dynamic character, which is determined by the development of the technological base.

The efficiency of special tools that guarantee error-free delivery of data in networks, is determined by the solution of the compromise between the reliability of the transmission (the probability of receiving the message without errors) and the amount of resources involved (delays in the data delivery, the amount of additional information transmitted to correct the errors that occur, computational complexity of procedures for localization and error correction) [3].

In accordance with the specifics of the task and the constraints on available resources, a wide range of tools is

Nikolaos Bardis, is with Hellenic Army Academy, Department of Military Sciences, Section of Mathematics and Engineering Sciences, Vari - 16673, Greece, (email: bardis@ieee.org)

Oleksandr P. Markovskyi is with Department of Computer Engineering, National Technical University of Ukraine, (Polytechnic Institute of Kiev) Peremohy pr., Kiev 252056, KPI 2003, Ukraine, (email: markovskyy@i.ua).

Kostiantyn V. Koliada is with Department of Computer Engineering, National Technical University of Ukraine, (Polytechnic Institute of Kiev) Peremohy pr., Kiev 252056, KPI 2003, Ukraine, (email: kost@ukr.net).

used to ensure reliable data transmission in global networks.

In modern networks [1], two fundamental technologies used are, (i) retransmitting a packet when detecting transmission errors and (ii) correcting a limited number of errors using the correction codes that localize and recover distorted packet symbols.

When using the first technology, control symbols are transmitted in the package for error detection. These symbols are mostly CRC-codes (Cyclic Redundancy Codes) and are formed as a remainder from the polynomial division of a sequence of packet information symbols into a simple polynomial in the Galois fields. In the case of an error, a request for retransmission of the package is sent. When applying the second technology, with the use of specially formed redundant symbols of the package, the localization and correction of a limited number of distorted information symbols of this package is carried out. Reed-Solomon codes are used as correction codes. The main disadvantage of these correction codes is the considerable computational complexity of the procedures for the localization of distorted symbols, and this complexity grows rapidly with increasing numbers of symbols that can be corrected.

Detection and correction of errors can be carried out at each transfer of the packet between the switching nodes participating in its delivery. Such an organization eliminates the accumulation of errors, but significantly slows down the process of the packet delivery, because on each node through which it passes, operations are performed to detect erroneous symbols, fix them or retransmit.

An alternative to the described step-by-step organization is the detection of errors and their elimination at the end point of the packet delivery.

Structurally, error correction can be performed at the level of the individual symbols of each packet, or entire packets using redundant packets. In the latter case, the error handling procedure is performed at the final delivery point of the packets.

The main disadvantage of implementing error correction at the level of individual symbols is that the exponential growth of the computational complexity of procedures for localization and correction of distorted symbols with an increase in their number significantly limits the corrective ability. In modern conditions, the increase of the loading of the broadcast channels in data transmission in networks has increased the probability of transmission errors due to the influence of external electromagnetic interference. Moreover, the duration of the interference significantly exceeds the time of transmission of one symbol. This results from the fact that the dominant type of error is the group of contiguous distorted symbols, corrected with a help of the Reed-Solomon correction codes, which require significant time and computational resources.

In addition, error correction technologies at the symbol level do not allow to reconstruct packets that were lost during the transmission process.

Therefore, in recent years, more attention is paid to methods of increasing the reliability that work at the packet level. The vast majority of these methods are based on the usage of erasure codes [2], [3]. In addition to global networks, such codes are also widely used in remote distributed information storage systems [4]. Unlike the correcting codes, erasure codes do not solve the problem of detecting packets corrupted during transmission: this is accomplished by integrating CRC-codes into packets. To reconstruct damaged or lost packets, additional redundant packets are transferred over the network, which partially accumulates information contained in the main packets. The task of erasure codes is to reconstruct the lost or damaged packet while transmission based on the information contained in the redundant packets.

The vast majority of erasure codes use linear operations (LT-coding) to form redundant packets. This ensures fast and efficient computing implementation of the reconstruction process. The most well-known type of erasure codes is Raptor [10], which allows to reconstruct an arbitrary number of lost packets from n transmitted. This high data reconstruction capability is achieved through a high level of redundancy - greater than 100%.

The main difference between existing linear erasure codes is the way of forming redundant packets. In the simplest case, this is a simple duplication of the main packets [9], which allows to simplify the procedures for creating redundant packets and reconstruction of lost or damaged packets.

In most of the existing erasure codes, the formation of redundant packets is carried out in such a way as to guarantee the reconstruction of a given number of lost packets [9]. This approach is justifiable for the reconstruction of information, stored in distributed memory on remote media.

For example, if an informational sending consists of 3 packets, then in order to guarantee their reconstruction, it is necessary to create and send five redundant packets to ensure their reconstruction. But even with three redundant packets, provided that they are rationally formed, the probability of reconstruction of the main packets at the loss of three of the six is 0.93, and the probability of reconstruction of the main packet when one or two packets are lost is one. This indicates that existing erasure codes, that are guided to the guaranteed reconstruction of a certain number of lost packets, do not achieve high efficiency. So, if in assessing efficiency only the number of redundant packets that are further transmitted over the networks are considered as resources, then the efficiency of using the three packages is 34% higher compared to the usage of five redundant packets.

Thus, existing erasure coders do not provide high efficiency of data packets reconstruction, considering the capability for reconstruction of the main packet, in relation to the number of redundant packets.

Therefore, the required task is that of developing a method for the formation of redundant packets, which would provide the highest probability of reconstruction of main data packets. The research goal is hence to increase the efficiency of redundant packet usage and reconstruction of lost data packets during transmission in global networks.

III. Method of Redundant Packet Creation and Reconstruction

To achieve the goal, a method is proposed for the formation of redundant packages and the reconstruction of lost, damaged or delayed over critical time the main data packets.

The method is derived from the following model of data transmission in the global network. Transmitted data is organized into n main packets P_1, P_2,…, P_n. Each packet has built-in means of detecting transmission errors. Packets are transmitted over different routes of a heterogeneous network, which may include snippets of peer to peer networks (P2P). For peer-to-peer networks, when information delivery occurs through a chain of nodes, it is possible to disable one of them. This leads to a rupture of the virtual transmission channel and, accordingly, to the loss of some packages. The specificity of usage determines some critical time of information delivery. Packets delivered after a critical time lose relevance.

The proposed method involves the formation of k redundant packets R_1, R_2, …, R_k, which are transmitted over the network along with the main ones. It is believed that in the process of transmission, part of the main and redundant packets may be lost, damaged or delayed exceeding the critical delivery time. The problem is to reconstruct main data packets that are lost, damaged or delayed using the main and redundant intact packets received by the receiver before the critical time limit. If the data delivery time is not critical, the problem of correction the main packets that are lost or damaged resolves using the redundant packets.

Since the correction of lost main packets is critical, the method involves the usage of linear Boolean transformations. Such transformations are simple and can be executed in parallel, which ensures high speed of redundant packets formation and reconstruction of lost main packets. A significant advantage of using linear Boolean transformations is the simplicity of hardware implementation.

Thus, it is proposed to create redundant packets using linear transformations over the main packets:

$$\forall i \in \{1,2,\ldots,k\}: \quad R_i = \bigoplus_{j=1}^{n} \lambda_{i,j} \cdot P_j \tag{1}$$

where $\forall i \in \{0, \ldots, k\}, j \in \{1, 2, \ldots, n\}: \lambda_{i,j} \in \{0, 1\}$.

On the receiver's side, in the absence of receiving a certain set of Ω main packets at a critical moment of time, they are considered lost and can be corrected by performing linear transformations over the main and redundant packets, received in time.

The probability of reconstruction of lost main packets is determined by their number h, characteristics and status of network, the number of k redundant packets, the number of n main packets, as well as the way of redundant packets formation.

For effective reconstruction of lost data packets, based on the network characteristics, determined by the specific task of the requirements of the reliability for correction and the specified number of main packets, it is needed to determine the required number of redundant packets, as well as the way they formed.

Determining the required number of redundant packets is based on the requirements for the reliability of the reconstruction, network characteristics and the dependence of the probability of lost packets reconstruction from the number of main and redundant packets.

Thus, in order to achieve this goal, the following tasks need to be solved:

- Theoretically substantiate and develop the way of redundant packets formation, which provides the greatest probability of lost packets reconstruction.
- Determine the dependency of the probability of lost packets reconstruction from their number, number of main and redundant packets.

Each of n main packets is related with binary vectors V_1, V_2, …, V_n, in every j-th of which only j-th component equals to one, and all others is equals to zero: $V_1 = \{v_{11}, v_{12}, \ldots, v_{1n}\}$, $V_2 = \{v_{21}, v_{22}, \ldots, v_{2n}\}$, …, $V_n = \{v_{n1}, v_{n2}, \ldots, v_{nn}\}$, $\forall i,j \in \{1, 2, \ldots, n\}$, $i \neq j$: $v_{ij}=0$, $v_{ii} = 1$.

The formation method of each i of redundant packets P_i, $i \in \{1, 2, \ldots, k\}$ is determined by corresponding binary vector $\lambda_i = \{\lambda_{i,1}, \lambda_{i,2}, \ldots, \lambda_{i,n}\}$. Efficiency of the formation method of k redundant packets is determined by the probability of reconstruction of the lost main packets provided that during the transmission lost u packets of main and redundant number. The efficiency of the method of redundant packets formation, that is vectors λ_1, …, λ_k, is determined by the selection of the matrix Λ columns and does not depend on their order:

$$\Lambda = \begin{vmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{vmatrix} = \begin{vmatrix} \lambda_{1,1} & \lambda_{1,2} \cdots \lambda_{1,n} \\ \lambda_{2,1} & \lambda_{2,2} \cdots \lambda_{2,k} \\ & \vdots \\ \lambda_{k,1} & \lambda_{k,2} \cdots \lambda_{k,n} \end{vmatrix} = \left| s_1 \; s_2 \cdots s_n \right| \tag{2}$$

where $s_1=\{\lambda_{1,1}, \lambda_{2,1}, \ldots, \lambda_{k,1}\}$, $s_2=\{\lambda_{1,2}, \lambda_{2,2}, \ldots, \lambda_{k,2}\}$,…, $s_n=\{\lambda_{1,n}, \lambda_{2,n}, \ldots, \lambda_{k,n}\}$.

Assertion. When h main packets are lost, whose numbers form a set Ω and the loss of $\eta = u-h$ redundant packets, so that the numbers of unlost redundant packets form a set ϑ, the lost main packets can be corrected, if a matrix Θ is formed by components of the vectors $\lambda \in \vartheta$, the numbers of which belong to the set Ω, contains an orthogonal submatrix consisting of h rows and h columns.

Proof. The lost packet P_j, $j \in \Omega$ can be reconstructed, if the vector V_j, which related to the lost packet P_j can be represented as linear function of vectors V_i of unlost packets $i \notin \Omega$ and vectors λ_l of unlost redundant packets $l \in \vartheta$:

$$V_j = \bigoplus_{i \notin \Omega} a_i \cdot V_i \oplus \bigoplus_{l \in \vartheta} b_l \lambda_l, \tag{3}$$

where $a_1, \ldots, a_n \in \{0, 1\}$, $b_1, b_2, \ldots, b_n \in \{0, 1\}$.

If there is an orthogonal submatrix M, h columns of which related to the numbers $i_1, i_2, \ldots, i_h$ of lost main packets, $i_1, i_2, \ldots, i_h \in \Omega$, and h rows of which related with the numbers $j_1, j_2,$

…, j_h of subset $\Delta \subseteq \vartheta$ of unlost redundant packets, i.e. there is an orthogonal submatrix M:

$$M = \begin{vmatrix} \lambda_{j_1,i_2} & \lambda_{j_1,i_2} & \cdots \lambda_{j_1,i_h} \\ \lambda_{j_2,i_1} & \lambda_{j_2,i_2} & \cdots \lambda_{j_2,i_h} \\ & \vdots & \\ \lambda_{j_h,i_1} & \lambda_{j_h,i_2} & \cdots \lambda_{j_h,i_h} \end{vmatrix}, \quad (4)$$

then vectors $\lambda_{j_1}, \lambda_{j_2}, ..., \lambda_{j_h} \in \Delta$ can be represented as follows

$$\forall n \in \vartheta : \lambda_{j_n} = \bigoplus_{m \in \Omega} \lambda_{j_n,m} \cdot V_m \oplus \bigoplus_{l \notin \Omega} \lambda_{j_n,l} \cdot V_l, \quad (5)$$

A system (5) can be represented as a system of linear boolean equations, the unknowns of which are vectors $V_{i_1}, V_{i_2}, ..., V_{j_h}$, which are related to the lost main packets

$$\forall n \in \vartheta : \bigoplus_{m \in \Omega} \lambda_{j_n,m} \cdot V_m = \lambda_{j_n} \oplus \bigoplus_{l \notin \Omega} \lambda_{j_n,l} \cdot V_l. \quad (6)$$

In a system (6) coefficients for the unknowns $V_{i_1}, V_{i_2}, ..., V_{j_h}$ are elements of the matrix M. Accordingly, if the matrix M is orthogonal, the system (6) can be solved relatively $V_{i_1}, V_{i_2}, ..., V_{j_h}$, and each of the vectors $V_{i_1}, V_{i_2}, ..., V_{j_h}$ is expressed as a linear combination of vectors V_i of unlost packets: $i \notin \Omega$ and a subset $\Delta \subseteq \vartheta$ of vectors λ_l of unlost redundant packets $l \in \Delta$:

$$\forall q \in \Omega : V_q = \bigoplus_{i \notin \Omega} a_{q,i} \cdot V_i \oplus \bigoplus_{l \in \Delta} b_{q,l} \cdot \lambda_l, \quad (7)$$

where $\forall\, q \in \Omega$, $i \notin \Omega$: $a_{q,i} \in \{0,1\}$, $\forall\, q \in \Omega$, $l \in \Delta$: $b_{q,l} \in \{0,1\}$.

Respectively, the reconstruction of h lost main packets $P_{i_1}, P_{i_2}, ..., P_{i_h}$ can be accomplished using correctly received main and redundant packets according to the following formula:

$$\forall q \in \{i_1, i_2, .., i_h\} = \Omega : \quad P_q = \bigoplus_{i \in \Omega} a_{q,i} \cdot P_i \oplus \bigoplus_{l \in \Delta} b_{q,l} \cdot R_l, \quad (8)$$

which was necessary to prove.

Obviously, the orthogonal sub matrix M of matrix Θ for a specific localization of lost packets can only exist under the following two conditions:

- among the columns, whose numbers belong to the set Ω, there are no such that contain only zero components.
- among the columns, whose numbers belong to the set Ω, there are no such that repeat.

The probability g_u is determined by the choice of columns s_1, s_2, …, s_n of the matrix Λ and does not depend on their order. In particular, if $h=u$, i.e. all lost packets belong to the main ones, then the maximum probability g_u is achieved provided all the columns s_1, s_2, …, s_n are different and none of them consists of only zero components. Obviously, all the columns can be different only if the condition $2^k > n$ is fulfilled.

If $i<u$, i.e. u-i redundant packets are lost, from columns s_1, s_2, …, s_n exclude components that are related to lost redundant packets. In this case, columns containing less than u-i+1 ones can be transformed into a columns containing only zeros, which results in a decreasing of the probability g_i.

Therefore, columns containing a small number of ones should be considered with less priority. On the other hand, if the number of zero components of the columns s_1, s_2, …, s_n is small, with the exception of the components that are related to the lost redundant packets, the probability of their transformation in identical ones increases.

In terms of loss of ability to reconstruct lost main packets, the transformation of the column of the matrix Λ to zero when redundant packets are lost, is more critical than the transformation of the column into one, all components of which consist of ones. In the last situation the reduction of the reconstruction ability is due to the appearance of repeating columns.

If h the main packets are lost and after the transformation of the matrix Λ it contains a zero column, then if the number of the lost packet coincides with the number of the zero column, it cannot be reconstructed for all variants of localization h-1 of other lost packets of n-1. Accordingly, this generates $\rho_0 = \binom{n-1}{h-1}$ localization variants of lost main packets that cannot be reconstructed.

If at loss of h the main packets and after the transformation of the matrix Λ it contains a column consisting of ones, there arises the situation of the presence of two identical columns of the transformed matrix Λ. If the numbers of the lost main packets coincide with the numbers of the same columns of the transformed matrix Λ, they cannot be reconstructed in all variants of localization h-2 of other lost packets of n-2. Accordingly, this generated $\rho_1 = \binom{n-2}{h-2}$ localizations variants of lost main packets that cannot be reconstructed.

The ratio $\upsilon = \rho_0 / \rho_1$ of the number of localization variants of the main packets, that cannot be reconstructed, as a result of the transformation of the matrix Λ appearance of the zero and one columns, can be represented as:

$$\upsilon = \frac{\rho_0}{\rho_1} = \frac{\binom{n-1}{h-1}}{\binom{n-2}{h-2}} = \frac{n-1}{h-1} \quad (9)$$

Given that in practice $n>>h$, the value of $\upsilon>>1$, i.e. the appearance of a zero column while transforming the matrix Λ significantly has a much more negative effect on reconstruction ability than the appearance of a one column. For example, for n=15 and k=4 in the situation of loss of 4 packets, one of which is redundant and three main (h=3), the value of υ is 7. This means that the appearance as a result of

the transformation of the matrix Λ of a zero column leads to increasing the number of main packets, that cannot be corrected and this number is 7 times greater than the number of non-reconstructed main packets due to the appearance of the column containing only one components as a result of the transformation of the matrix Λ.

Based on the above, the following procedure for determining the priorities $\chi_1, \chi_2,\ldots,\chi_n$ for the columns $s_1, s_2, \ldots, s_n$ of the matrix Λ is proposed. For the column s_i, $i\in\{1,2,\ldots,n\}$ the number of its one components ε_i can be determined. Columns than do not contain one components, those for which $\varepsilon_i = 0$, are not used. For the column s_i with $\varepsilon_i > 0$ its priority χ_i is proposed to define as follows:

1. If $\varepsilon_i = 1$, i.e. the column s_i contains only single one, then it is assigned the lowest priority $\chi_i = 0$.

2. If the column s_i contains only one components, i.e. $\varepsilon_i = k$, then $\chi_i = 1$.

3. Provided $1< \varepsilon_i < k$, then if $\varepsilon_i < k/2$, then $\chi_i = 2\cdot\min\{\varepsilon_i, k-\varepsilon_i\}-1$, otherwise, i.e. if $\varepsilon_i >= k/2$, then $\chi_i = 2\cdot\min\{\varepsilon_i, k-\varepsilon_i\}$.

As an illustration, Table 1 lists the priorities of 5-component columns (k=5), depending on the numbers of ones in them.

Table 1. Dependence of the column priorities of the matrix Λ on the number of ones in them for k=5

Number of ones ε	Priority value χ
1	0
2	2
3	4
4	3
5	1

IV. Technical Implementation

As was shown above, theoretically, the condition for the reconstruction of lost packets during transmission of h main packets, whose numbers form the set Ω, consists in the presence of the orthogonal submatrix Θ in the transformed matrix Λ.

The reconstruction process consists of solving a system of linear equations, whose coefficients are formed by the components of the matrix Θ. The analysis of the transformed matrix Λ for the detection of the orthogonal submatrix Θ in it, the formation on the basis of the matrix Θ the system of linear equations with its subsequent solution requires certain computational and time resources, which are critical for the reconstruction of lost data in real time. Therefore, it is considered as the most expedient the implementation of the specified list of action for all possible localizations of lost packets during the setup. The received results are proposed to be organized in the form of special tables, which, depending on the localization of the lost packets, contain a specification of the operations for the reconstruction of lost data, or information that lost packets cannot be reconstructed. Accordingly, the foregoing procedures of transformation of the matrix Λ, the derivation of the orthogonal submatrix Θ in it, the formation and solving of the system of linear equations, can be reduced to reading the specification for the reconstruction of lost packets from tabular memory. This solution makes it possible to implement the specified list of actions in real time.

While using k redundant packets it is possible to reconstruct no more than k packets. Thus, the table should provide a specification for the reconstruction of lost packets, if their number does not exceed k.

An important technological facilitation for the use of the table of precomputations, is to develop an algorithm for its addressing according to a given L localization of lost packets. The code L consists of $n+k$ bits, each of which corresponds to the main or redundant packet and equals to one, if the corresponding packet is lost.

It is clear that the number δ_j of possible loss localizations of j packets (the number of possible L codes, containing exactly j ones) is equal to $\binom{n+k}{j}$, where $j\in\{1, 2, \ldots, k\}$. The specification of the reconstruction of the i^{th} packet, $i\in\{1, 2, \ldots, n+k\}$ is $(n+k)$-bit code $C = \{c_1, c_2, \ldots, c_{n+k}\}$, $\forall i\in\{1, 2, \ldots, n+k\}$: $c_i\in\{0, 1\}$, the one components of which indicate packets, the sum of which is equal to the lost. In other words, the lost packet P_i can be reconstructed according to the following formula:

$$P_i = \bigoplus_{l=1}^{n} c_l \cdot P_l \oplus \bigoplus_{t=1}^{k} c_{n+t} \cdot R_t \qquad (10)$$

Accordingly, the total number N_c of specifications for the reconstruction of lost main packets contained in the table is determined by the following formula:

$$N_c = \sum_{j=1}^{k} \binom{n+k}{j} \cdot j \qquad (11)$$

For example, for n = 8, k = 4, the number N_c of specifications for the reconstruction not more than 4 main packets, contained in the precomputation table is $N_c = 12\cdot1 + 66\cdot2 + 220\cdot3 + 495\cdot4 = 2784$. The length of each of the specification is $n+k$ = 12 bits, so that the total volume of the table is 33408 bits.

The procedure for specifying specifications at the setup stage can be illustrated by such example.

Example. The number of main packets n = 8, the number of redundant packets k = 4, then the matrix Λ will be as follows:

$$\Lambda = \begin{vmatrix} 11100010 \\ 10011011 \\ 01010111 \\ 00101101 \end{vmatrix} \qquad (12)$$

Accordingly, the vectors for forming of redundant packets are as follows: λ_1 = [1 1 1 0 0 0 1 0], λ_2 = [1 0 0 1 1 0 1 1], λ_3=[0 1 0 1 0 1 1 1], λ_4 = [0 0 1 0 1 1 0 1]. This means, that redundant packets are formed in such way:

$$R_1 = P_1 \oplus P_2 \oplus P_3 \oplus P_7,$$
$$R_2 = P_1 \oplus P_4 \oplus P_5 \oplus P_7 \oplus P_8, \qquad (13)$$

$$R_3 = P_2 \oplus P_4 \oplus P_6 \oplus P_7 \oplus P_8,$$
$$R_4 = P_3 \oplus P_5 \oplus P_6 \oplus P_8$$

The vector of lost packets is as follows: [0 0 1 0 1 0 0 1 | 0 0 0 1], that is 3 main P_3, P_5, P_8 and 1 redundant R_4 data packets are lost while transmission. Therefore,

$$\lambda_1 = P_1 \oplus P_2 \oplus \gamma_1 \oplus P_7,$$
$$\lambda_2 = P_1 \oplus P_4 \oplus \gamma_2 \oplus P_7 \oplus \gamma_3,$$
$$\lambda_3 = P_2 \oplus P_4 \oplus P_6 \oplus P_7 \oplus \gamma_3$$
(14)

Then after transformation of the matrix Λ the matrix Λ′ will be received, which contains the orthogonal submatrix M:

$$\Lambda' = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix}$$
(15)

Thus, the vectors γ_1, γ_2, γ_3, which correspond to the sum of the packets which are lost and which are included in the corresponding index of the redundant packet, can be written in the form:

$$\begin{cases} \gamma_1 = P_3 \\ \gamma_2 = P_5 \oplus P_8 \\ \gamma_3 = P_8 \end{cases}$$
(16)

Having solved the linear equation relative to P_i, the system will have the form:

$$\begin{cases} P_3 = \gamma_1 \\ P_5 = \gamma_2 \oplus \gamma_3 \\ P_8 = \gamma_3 \end{cases}$$
(17)

After substituting for γ_i values of packets, that are not lost during transmission, the system will turn into the form:

$$\begin{cases} P_3 = R_1 \oplus P_1 \oplus P_2 \oplus P_7 \\ P_5 = R_2 \oplus R_3 \oplus P_1 \oplus P_2 \oplus P_6 \\ P_8 = R_3 \oplus P_2 \oplus P_4 \oplus P_6 \oplus P_7 \end{cases}$$
(18)

Then the recovery of lost packets P_3, P_5, P_8 using the received data packets is as follows:

$$\begin{cases} P_3 = R_1 \oplus P_1 \oplus P_2 \oplus P_7 \\ P_5 = R_2 \oplus R_3 \oplus P_1 \oplus P_2 \oplus P_6 \\ P_8 = R_3 \oplus P_2 \oplus P_4 \oplus P_6 \oplus P_7 \end{cases}$$
(19)

So, the specification will look like this:

s_1 = [1 1 0 0 0 0 1 0 | 1 0 0 0]
s_2 = [1 1 0 0 0 1 0 0 | 0 1 1 0]
s_3 = [0 1 0 1 0 1 1 0 | 0 0 1 0]
s_4 = [0 0 0 0 0 0 0 0 | 0 0 0 0]
(20)

It is suggested the addressing of the specification table for the reconstruction of lost packets by code L be organized in the following way. For a given code L, the numbers of ones in it is $E(L)$, and the serial number $W(L)$ among the codes, containing $E(L)$ ones. The A_L address of the first of the specifications, which is related to the given code L, it is suggested to be calculated in accordance with the formula:

$$A_L = B(E(L)) + W(L) \cdot E(L),$$
(21)

where $B(E(L))$ – shifting the start of the specification, describing the reconstruction of $E(L)$ lost packets. For example, for $n = 8$ i $k = 4$; $B(1) = 0$, $B(2) = 12$, $B(3) = 12 + 66 \cdot 2 = 144$, $B(4) = 144 + 220 \cdot 3 = 804$.

To calculate the shifting of the specification address in the table memory of code L, it must be transformed into serial number $W(L)$ on the set $(n+k)$-digit codes, containing $E(L)$ ones. For example, for $n = 8$ and $k = 4$ the code L = 1111 0000 0000 can be transformed into serial number 0, code 1110 1000 0000 – into serial number 1. Accordingly, for the code L = 0000 0000 1111 the code $W(L)$ of the serial number is:

$$W(L) = \binom{12}{4} - 1 = 494$$
(22)

To calculate the serial number $W(L)$ of code L, containing $E(L)$ ones, the following algorithm is proposed.

1. The number i of the current bit of code L set to one: $i = 1$, the code r of result set to zero: $r = 0$, the variable g assigns a value to the number of ones $E(L)$ in the code L: $g = E(L)$.
2. If the current bit l_i of the code L is equal to one, then execute the decrement g: $g = g - 1$; otherwise, if $l_i = 0$, add the result to the code r of result:

$$\binom{n+k-1}{g-1} : r = r + \binom{n+k-1}{g-1}$$
(23)

3. Make the transition to the next bit of the code L: $i = i + 1$. If $i < n + k$ and $g > 0$, make a transition to re-execution of item 2.
4. The end of the algorithm: $W(L) = r$.

The operation of the proposed algorithm can be illustrated by the example of the calculation $W(L)$ for L = {0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1}. The diagram of the values of the algorithm variables in the calculation of $W(L)$ is shown in Table 2.

Table 2. The diagram of the variables of the shifting calculation algorithm W(L) for L = {0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1}

The number i of current bit of the code L	Current bit of the code L: l_i	The number g of unprocessed ones of the code L	The result r
1	0	4	165
2	0	4	285
3	1	3	285
4	0	3	313
5	1	2	313

6	0	2	319
7	0	2	324
8	1	1	324
9	0	1	325
10	0	1	326
11	0	1	327
12	1	0	327

Numeric factorial codes $\binom{n+k-1}{g-1}$ for all possible values of g and i can be calculated at the configuring stage with maintaining the results in the tabular memory, addressed by concatenation of the codes g and i. For $n = 8$ and $k = 4$, it is necessary to store $11 \cdot 4 = 44$ values of the factorials.

In this example $W(L) = r = 327$. This means that the specification that determines the procedure of reconstruction packets P_3, P_5 and P_8, is stored in memory starting from address 804 + 327 = 1131.

V. Performance Evaluation

The method efficiency is determined by the ratio of its ability to reconstruct lost data and the amount of resources spent on solving this problem. As an estimate of the volume of resources the number of k redundant packets can be considered, for which delivery data channels resources need to be taken and for the formation of which it is necessary to perform certain calculations, the volume of which is proportional to k.

Therefore, in order to make a comparative assessment of the efficiency of packet redundancy by different methods, it is expedient to determine the efficiency E of redundancy as the ratio of the reconstruction probability of lost main packets to the number of redundant packets:

$$E = \frac{1}{k} \cdot \sum_{i=1}^{n+k} q_i \cdot p_{i,k} \qquad (24)$$

where q_i – the probability that during the transfer will be lost i packets of $n+k$ transmitted; $p_{i,k}$ – the probability that when lost i packets of $n+k$ transmitted, all lost main packets can be reconstructed using k redundant packets.

Assuming that loss of packets occurs independently, the number of lost packets is subordinate to the binomial law for which the probability ρ of loss of one packet is determined. Then, formula (24) can be represented as:

$$E = \frac{1}{k} \cdot \sum_{i=1}^{n+k} \binom{n+k}{i} \cdot \rho^i \cdot (1-\rho)^{n+k-i} \cdot p_{i,k} \qquad (25)$$

As noted in the overview section, existing erasure-codes, in their majority, are aimed at guaranteeing the reconstruction of a fixed number of packets transmitted. Often, as such number is the number of main packets [14]. With this approach, the number k of redundant packets is defined as $w(n)$ [15]. For example, in order to guarantee the reconstruction of three packets for n=3, it is necessary to transfer $w(3)$=5 redundant packets, and for guaranteed reconstruction of five packets with the same number n=5 of main packets $w(5)$=7 redundant packets are used.

Accordingly, in guaranteeing the reconstruction of n packages of $n+k$ transmitted, the efficiency of E_0 can be represented as:

$$E_0 = \frac{1}{w(n)} \cdot \sum_{i=1}^{n+k} \binom{n+w(n)}{i} \cdot \rho^i \cdot (1-\rho)^{n+w(n)-i} \qquad (26)$$

Then, the gain ε in the efficiency of the redundancy of the proposed method in relation to known methods can be estimated from the ratio of the efficiency E of the redundancy of the proposed method to the efficiency of E_0 redundancy of known methods:

$$\varepsilon = \frac{E}{E_0} \qquad (27)$$

Table 3 presents the results of a comparative analysis of the redundancy efficiency for known and proposed redundancy methods for the transmission of 3 and 5 main packets. When calculating the efficiency of the proposed method for n=3 and for n=5, 3 backup packages were used: k=3.

Table 3. The results of a comparative analysis of the efficiency redundancy ***of proposed method and known erasures-codes***

n	***w(n)***	***k***	**ε = E/E₀**			
			ρ=0.001	**ρ=0.005**	**ρ=0.01**	**ρ=0.05**
3	5	3	1.24	1.25	1.26	1.31
5	7	3	1.556	1.561	1.563	1.582

The analysis of the data presented in Table 3 indicates that the developed method provides better efficiency of using redundant packets for the reconstruction of lost main packets in comparison with known methods. The difference in the efficiency of the proposed method increases with the increase in the number of main packets and the probability of loss or damage of the packet during its delivery.

VI. Conclusions

As a result of the conducted research, the method of formation of redundant packets and their usage for the reconstruction of main information packets that can be lost during transmission to the Internet is theoretically substantiated, developed and researched.

Each of the redundant packets is formed as the logical sum of certain subsets of information packets, and the choice of the specified subsets is regulated by the developed method, which ensures the greatest probability of the existence of an orthogonal system of equations, which solves the process of reconstruction of lost information packets. This provides greater efficiency in using redundant packets in comparison with known methods for reconstruction of packets in global networks.

To accelerate the reconstruction of lost packets, the method involves the usage of special pre-computational tables. In the tables, for each reconstructed packet, predefined subsets of

unlost main and redundant packets are stored, the logical sum of which is a lost packet.

The proposed method is oriented for usage in remote control computer management systems, communication with which is carried out via the Internet using modern radio modems.

REFERENCES

[1] Tanenbaum A.S. Computer networks / A.S. Tanenbaum. – Prentice Hall PTR. – 2016. – 960 p.

[2] Czap L. Secure Network Coding with Erasures and Feedback / L. Czap, C. Fragouli, V. Phabhakaran, S. Diggavi // IEEE Transaction on Information Theory. – 2015. – Vol. 61– No. 4. – P. 1667-1686.

[3] Mladenov T. Raptor Codes for P2P Streaming / T. Mladenov, U. Krieger // Parallel, Distributed and Network-Based Processing. – Feb. 2012. – P. 327 – 332.

[4] Adler N. Burst-Erasure Correcting Codes with Optimal Average Delay / A. Nitzan., Y.Cassuto // IEEE Transaction on Information Theory. – 2017. – Vol. 63. – No. 5. – P. 2848-2865.

[5] Fan X. Variable Packet-Error Coding / X.Fan, O.Kosut, A.B. Wagner // IEEE Transaction on Information Theory. – 2018. – Vol. 64. – No. 3. – P. 1530-1547.

[6] Wing Q. End-to-End Error-Correcting Codes on Networks with Wors-Case Bit Errors / Q. Wing, S. Jaggi // IEEE Transaction on Information Theory. – 2018. – Vol. 64. – No. 6. – P. 4467-4479.

[7] Leong D. Erasure coding for real-time streaming / D. Leong, T. Ho // Proceedng IEEE Int. Symposium Information Theory – ISIT-2012. – 200.

[8] Cisco Visual Networking Index: Forecast and methodology 2014-2019 [Virtual Resource]. – Access Mode: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf

[9] The statistics portal [Virtual Resource]. – Access Mode: http://www.statista.com/statistics/272835/share-of-internet-users-who-watch-online-videos

[10] IEEE 802.3 Industry Connections Ethernet Bandwidth Assessment [Virtual Resource]. – Access Mode: http://www.ieee802.org/3/ad_hoc/bwa/BWA_Report.pdf

[11] Deshpande H. Streaming live media over a peer-to-peer network / H. Deshpande, M. Bawa, and H. Garcia-Molina // In Work at CS-Stanford. – 2002.

[12] Prakash C. Data Communications and Computer Networks / C. Prakash // PHI Learning. – 2006. – P. 7-8.

[13] Hei X. Inferring Network-Wide Quality in P2P Live Streaming Systems / X. Hei, Y. Liu and K. Ross // Selected Areas in Communications. – 2007.

[14] Begen A. RFC 7198 – Duplicating RTP Streams / A. Begen, C. Perkins // IETF. – April 2014.

[15] Brinkmeier M. Methods for Improving Resilience in Communication Networks and P2P Overlays. PIK / M. Brinkmeier, M. Fischer, S. Grau, G. Schaefer, T. Strufe // Praxis der Informationsverarbeitung und Kommunikation 32. – 2009

Nikolaos G. Bardis received the diploma of Computer Engineering and the PhD degree from National Technical University of Ukraine (Polytechnic Institute of Kiev) in 1995 and 1999 respectively. He is currently an Associate Professor at the Hellenic Army Academy and Head of the Department of Mathematics and Engineering Sciences at the same institution. Collaborates as a lecturer and researcher at the University of Athens - Department of Mathematics and entered the postgraduate course of Cryptography and Security of Information Systems. His research interests include cryptography and data security, information theory, coding theory, systems engineering and applications in defence. He has published in over 50 peer-reviewed journals and conferences. He is a member of the Technical Chamber of Greece, Technical Program Committee (TPC) of the IEEE Communication Society (COMSOC), IEEE Computer Society Technical Committee on Computer Communications (TCCC), Technical Council on Software Engineering (TCSE) of the IEEE Computer Society and IEEE Information Theory Society.

Oleksandr P. Markovskyi obtained his specialty in "Computer engineering" in 1978 after graduating from Kiev Polytechnic Institute. His Ph.D. was conferred to him in 1987 at the same Institute. He is currently Professor of the Department of Computer Engineering of National Technical University of Ukraine “Kiev Polytechnic Institute". At present time, his scientific interests cover information security, security systems, coding theory, coding theory, arrangement of computational processes, associative memory. Oleksandr P. Markovskyi is the author more than 150 published works and 60 patents in the indicated fields. He has also taken part in many research programs carried out in the former Soviet Union, Japan, Greece and other counties

Kostiantyn V. Koliada obtained his specialty in "Computer Engineering" in 1985 after graduating from Kiev Polytechnic Institute. He is currently Assistant Professor of the Department of Computer Engineering of National Technical University of Ukraine “Kiev Polytechnic Institute". At present time, his scientific interests cover information security, data bases and, coding theory. Kostiantyn V. Koliada is the author more than 20 published scientific works and patents in the indicated fields.