UDC 004.627
DOI: 10.15587/1729-4061.2021.234492

DEVELOPMENT OF BINARY INFORMATION COMPRESSION METHODS BASED ON THE BINOMIAL NUMERICAL FUNCTION

Igor Kulyk
Corresponding author
PhD, Associate Professor*
E-mail: i.kulyk@ekt.sumdu.edu.ua
Olga Berezhna
PhD, Associate Professor*
Anatoliy Novhorodtsev
PhD, Associate Professor*
Maryna Shevchenko
Postgraduate Student*
*Department of Electronics and Computer Technology
Sumy State University
Rymskoho-Korsakova str., 2, Sumy, Ukraine, 40007

The application of data compression methods is an effective means of improving the performance of information systems. At the same time, interest is aroused to the methods of compression without information loss which are distinguished by their versatility, low needs of costs during implementation, and the possibility of self-control.

In this regard, the application of binomial numbering systems is promising. The numerical function of the binomial numbering system is used for compression. It makes it possible to put sequences in one-to-one compliance with their numbers. In this case, the transition from binary combinations to binomial numbers is used as an intermediate stage.

During the study, theorems were formulated that indicate properties of compressing and restoring the mappings as well as the ways of their implementation. Models of compression processes were obtained on the basis of a numerical function, both for the case of compressible equilibrium combinations and the case when sequences of a general form are to be compressed. The compression models include coding steps based on binary binomials.

The study results show the effectiveness of applying the compression based on the binomial numerical function. A 1.02 times increase in speed of information transmission through a communication channel was observed in the worst case and 18.29 times in the best case depending on the number of ones in 128-bit equilibrium combinations. The proposed methods are advantageous due to their high compression ratio (from 1.01 to 16 times for general 128-bit sequences) and versatility: combinations are compressed in which the number of ones is 75 % of their total variation range. The developed methods ensure control of errors during conversions. They are undemanding to computation resources and feature low implementation costs

Keywords: binomial numbering systems, binomial numerical function, binomial numbers, compression of binary information

Received date 14.03.2021
Accepted date 18.05.2021
Published date 29.06.2021

How to Cite: *Kulyk, I., Berezhna, O., Novhorodtsev, A., Shevchenko, M. (2021). Development of binary information compression methods based on the binomial numerical function. Eastern-European Journal of Enterprise Technologies, 3 (4 (111)), 6–13. doi: https://doi.org/10.15587/1729-4061.2021.234492*

1. Introduction

Existing and newly developed information systems are of distributed nature and must transfer and store huge amounts of information and perform frequent and complex queries to databases [1–3]. Insufficient bandwidth of communication channels, low speed of information storage devices, and requirements to ensuring specified levels of fail-safety and error-free transmission are the factors that limit performance gain in such systems [3–5].

One of the most effective measures taken to speed up information processing and reduce the cost of computer systems and the components that form the basis of present-day control systems and telecommunication consists in the application of methods and algorithms of data compression [1–3].

Development of compression methods and algorithms that would have the following properties and characteristics is of particular scientific and practical interest [1–3]:

– high compression ratio and simplicity of the compressed data model;

– high performance when compressing and recovering information;

– be undemanding to computing resources of terminal devices;

– adaptability to the type of processed data and their characteristics;

– the ability to control errors when compressing and recovering data.

In this regard, the development and implementation of compression methods and algorithms with the above properties is an urgent problem of scientific and practical value.

2. Literature review and problem statement

Almost any data arrays and sequences have their own structural features. It is pointed out in [6, 7] that taking into account these features can significantly improve the efficiency of data compression methods based on a combinatorial approach to their representation. Although the numbering of equilibrium combinations is carried out in [6] based on binomial coefficients, binomial numbers are not considered at all in the structure of processed combinations. This leads to more cumbersome calculations when finding numbers and does not provide an extension of the functionality of the proposed methods of compressive coding. The problem of compression is studied in [7] for information sequences of limited application which narrows the scope of the proposed methods. Similar to [6], binomial numbers are not used in [7] for compression and restoration.

Implementation of the combinatorial approach to compressive coding involves the application of structural positional numbering systems that differ from the conventional binary notation by their complexity and heterogeneity. In structural numbers generated by structural positional numbering systems, there are complex relationships between the weight coefficients of numbers and their positions in the numbers themselves. On the one hand, such complex connections between weights of numbers and their positions hamper arithmetic operations on them, however, on the other hand, they expand their functionality for self-control, encryption, and data compression. When designing various information systems and solving specialized information problems, such structural numbering systems as Fibonacci, factorial, and binomial numbering systems are already widely used [8–10].

To solve information problems including data coding, it is proposed in [8] to use Fibonacci numbering systems. But in view of their simpler structure, the Fibonacci numbers have significantly lower noise immunity and do not ensure effective error detection during information transformations. The application of a factorial numbering system is considered in [9]. This study is limited to a range of problems with permutations that do not include information compression problems.

Binomial numbering systems, in particular those dealing with a binary alphabet [10, 11], are of considerable interest and broad prospects. Being more complex in their structure, binomial numbering systems have great capabilities to detect and correct errors in information sequences, generate a wide range of combinatorial objects. Besides, algorithms of obtaining binomial numbers are quite simple and there are many classes of code combinations in the structure of which binomial numbers are found [6, 7, 11].

Description of binomial numbering systems and the binomial numbers generated by these systems is given in [10]. However, their application is limited only to organizing binomial counting. Issues of information compression using binomial numbering systems are not touched upon at all.

Only conceptual approaches to the application of binary binomial numbering systems in information control systems are given in [11] including those for compression of binary data. A significant compression effect is observed when using a binomial numeric function to obtain binary numbers of compressible information sequences. In this case, binary binomial numbers are used as intermediate combinations in the transition to numbers. This greatly simplifies the application of the binomial numerical function and expands possibilities of controlling the transformation correctness.

Binomial compression of combinations is considered in [12] but just based on binary binomial numbers. Such compression coding is fast but does not make it possible to obtain high compression ratios. Obviously, the studies do not fully reveal in full the potential of binomial numbering systems and the possibility of implementing higher compression based on a binomial numerical function.

An ideological approach to the application of binomial compression based on a numerical function is revealed in [13]. However, there is no theoretical justification and mathematical models of compression in this study that would represent a ground for further development of binomial algorithms, systems, and compression devices.

It should be pointed out that a system of numerical coding the binary combinations with constant weight is given in [14]. But like [4], binomial numbers are not used at all when numbering which in general does not simplify the implementation of compression or reduce the hardware, software, and time costs.

Thus, the following conclusions can be drawn based on analysis of [4–12]:

– application of binary binomial numeration systems determines the versatility of compressive coding due to taking into account the only and widespread limitation on the number of ones in binary sequences;

– determination of binomial numbers in the data structure should be distinguished as a separate stage in compression which will reduce costs of numbering sequences and control errors due to the noise immunity inherent in binomial numbers;

– in order to increase the degree of binomial compression, it is necessary to use a numerical function of the binary binomial numbering system which assigns corresponding numbers to binary sequences;

– for the further development of binomial compression, it is necessary to construct mathematical models and algorithms for the compression of binary data based on a binomial numerical function.

3. The study objective and tasks

The study objective implies the elimination of information redundancy of binary sequences with cutting time and hardware and software costs. This will improve the performance of information-and-control and telecommunication systems handling both binary combinations with a given limitation on the number of units and information sequences of general form.

To achieve the study objective, the following tasks were formulated:

– construct a mathematical model and an algorithm for the considered binomial compression of equilibrium combinations;

– construct a mathematical model and an algorithm for the considered binomial compression of binary sequences of general form.

4. Materials and methods used in the study

4. 1. Justification of compression expediency

The proposed binomial compression methods are based on binary (n,k)-binomial numbering systems and binary binomial numbers generated by these systems (n and k above are parameters of the used set of numbers).

Binary (n,k)-binomial numbering system is characterized by:

1) a binomial numerical function that allows one to determine the decimal quantitative equivalent F_j of a binary binomial number $X_j=x_1x_2...x_i...x_r$ [11, 12]:

$$F_j = \text{dec}\, X_j = \sum_{i=1}^{r} x_i C_{n-i}^{k-q_i}, \qquad (1)$$

where $X_j \in X[n,k]$, $x_i \in \{0,1\}$, $r<n$, $j=1,2,...,C_n^k$; q_i is the sum of unit digits x_i from the first to the $(i-1)$-th position, inclusive:

$$q_i=\sum_{t=1}^{i-1} x_t,\ q_i \le k;$$

2) systems of code-forming constraints that allow one to generate binary (n,k)-binomial numbers $X_j=x_1x_2...x_i...x_r$:

$$\begin{cases} l=n-k, \\ x_r=0 \end{cases}$$

and

$$\begin{cases} q=k, \\ x_r=1, \end{cases} \qquad (2)$$

where q and l are the numbers of ones and zeros in the binary binomial number X_j.

Binomial numerical function (1), being a numbering function, allows one to associate binary (n,k)-binomial numbers X_j with their decimal equivalents F_j or binary numbers $D_j=\mathrm{Bin}F_j$ $D_j \in D[n,k]=\left\{\mathrm{Bin}(\overline{0,C_n^k-1})\right\}$.

Binary binomial numbers $X_j \in X[n,k]$ form the basis of widespread equilibrium combinations $Y_j \in Y[n,k]$ where $Y[n,k]$ is the set of binary equilibrium n-digit combinations with a constant number of k ones. Bijective mappings of the form f_b:$Y[n,k] \to X[n,k]$ and $f_b^{-1}: X[n,k] \to Y[n,k]$ define, respectively, binomial compression and restoration based on binary binomial numbers where $0 \le k \le n$ [12].

Justification of expediency of developing a method for compression of binary equilibrium combinations $Y_j \in Y[n,k]$ based on the binomial numerical function (1) includes the following statements:

1) length of numbers D_j satisfies the inequalities $|D_j|<n$ and $|D_j|<L_{cp}$ where L_{cp} is the average length of binary (n,k)-binomial numbers [10] which provides a compression ratio of more than one;

2) functionality of compliances $H \subseteq D[n,k] \times X[n,k]$, $(D_j,X_j) \in H$ and $S \subseteq X[n,k] \times D[n,k]$, $(X_j,D_j) \in S$ between sets $D[n,k]$ of binary numbers D_j and $X[n,k]$ of binary binomial numbers X_j based on the theorems of unambiguity of binomial numbers [10, 12]. This leads to bijective binomial mappings of the form ψ^{-1}:$D[n,k] \to X[n,k]$ and ψ:$X[n,k] \to D[n,k]$ and ensures one-to-oneness of encoding and decoding, respectively;

3) uniformity of binary numbers D_j enabling significant reduction of hardware and software costs in the construction of devices and software modules for compression and restoration of equilibrium combinations Y_j based on the binomial numerical function (1).

4. 2. The method of compression of equilibrium combinations based on the binomial numerical function

Realization of binomial mapping of the form ψ: $X[n,k] \to D[n,k]$ in the presence of a bijective mapping f_b: $Y[n,k] \to X[n,k]$ for initial binary n-digit sequences $Y_j \in Y[n,k]$ is represented by compression:

$$f_e\text{:}Y[n,k] \to D[n,k]. \qquad (3)$$

of equilibrium combinations $Y_j \in Y[n,k]$ based on binary numbers $D_j \in D[n,k]$ which are calculated using the function (1). This type of compression will also be called binomial numerical compression. Thus, we are dealing with a complex function $f_e=\psi \circ f_b$:

$$D_j=f_e(Y_j)=\psi(f_b(Y_j)). \qquad (4)$$

In turn, the realization of the binomial mapping ψ^{-1}: $D[n,k] \to X[n,k]$ in the presence of a bijective mapping $f_b^{-1}: X[n,k] \to Y[n,k]$ for binary n-digit sequences $Y_j \in Y[n,k]$ means restoration:

$$f_e^{-1}: D[n,k] \to Y[n,k], \qquad (5)$$

of initial equilibrium combinations $Y_j \in Y[n,k]$ based on binary numbers $D_j \in D[n,k]$ which are calculated using the function (1). This type of restoration will also be called binomial numbering restoration. Thus, we are dealing with a complex inverse function $f_e^{-1}=f_b^{-1} \circ \psi^{-1}$: $Y_j=f_e^{-1}(D_j)=f_b^{-1}(\psi^{-1}(D_j))$.

At the heart of the methods of binomial numbering compression f_e (3) and restoration f_e^{-1} (5), it is proposed to apply binomial mappings of the form ψ:$X[n,k] \to D[n,k]$ and $\psi^{-1}: D[n,k] \to X[n,k]$ which are determined by the numerical function (1) and systems of code-forming constraints (2). In other words, when compressing and recovering based on the binomial numerical function (1), binary binomial numbers X_j are used when passing from equilibrium combinations Y_j to binary numbers D_j and vice versa. This significantly distinguishes the proposed compression method from the numbering systems considered in [6, 14].

The following Theorem 1 gives properties of compressive mapping f_e and the method of its practical implementation («/» is the decatenation symbol).

Theorem 1. Any binary sequence $Y_j=y_1y_2...y_i...y_n$, $Y_j \in Y[n,k]$, $j=\overline{1,C_n^k}$ composed of n bits y_i, the sum of values of which is equal to k, can be associated with a unique positive integer $D_j \in D[n,k]$ using mapping f_e in two steps:

1) transition from the sequence $Y_j=y_1y_2...y_i...y_n$ to the binary (n,k)-binomial number $X_j=x_1x_2...x_i...x_r$, $X_j \in X[n,k]$, $r<n$ using the function $X_j=f_b(Y_j)$ in the form:

$$X_j=x_1x_2...x_i...x_r=\begin{bmatrix} y_1y_2...y_i...y_{n-1}0/00...0, \\ y_1y_2...y_i...y_{n-1}1/11...1; \end{bmatrix}$$

2) calculation of the number $D_j=\mathrm{Bin}(\mathrm{dec}X_j)$ of the binary (n,k)-binomial number X_j in accordance with the numerical function (1).

Proof of Theorem 1 for the first stage of transformations is made proceeding from compliance of X_j to the systems of code-forming constraints (2) and the uniqueness of X_j taking into account the fact that the decatenation operation does not change values of the digits x_i. At the second stage of transformations to prove Theorem 1, the sequential division of $\mathrm{dec}X_j$ by $C_{n-i}^{k-q_i}$ is used to obtain the quotient $0 \le \delta_i \le 1$ and substantiate uniqueness of decomposition (1).

Reflecting the form of a complex function $f_e=\psi \circ f_b$, the model of f_e compression process includes the first two stages related to f_b compression based on binary binomial numbers [12].

Thus, using Theorem 1, modeling of the f_e compression process of binary equilibrium combinations $Y_j=y_1y_2...y_i...y_n$ based on the binomial numerical function (1) consists of the following stages.

Stage 1. The value of the last byte y_n is determined in the n-digit equilibrium combination $Y_j=y_1y_2...y_i...y_n$ having k of ones.

Stage 2. If $y_n=0$, then

$$X_j=Y_j/00...0=y_1y_2...y_i...y_{n-1}0/00...0=x_1x_2...x_i...x_{r-1}1,$$

that is, all zero bits are discarded from the combination $Y_j=y_1y_2...y_i...y_{n-1}0$ starting with $y_n=0$ until the first binary one $y_r=1$appears which will represent a value of the last bit $x_r=y_r=1$ of the sought (n,k)-binomial number $X_j=x_1x_2...x_i...x_{r-1}1$. Otherwise

$$X_j=Y_j/11...1=y_1y_2...y_i...y_{n-1}1/11...1=x_1x_2...x_i...x_{r-1}0,$$

that is, all unit digits are discarded from the combination $Y_j=y_1y_2...y_i...y_{n-1}1$ starting from $y_n=1$ until first binary zero appears $y_r=0$ which will represent a value of the last bit $x_r=y_r=0$ the sought (n,k)-binomial number $X_j=x_1x_2...x_i...x_{r-1}0$. In this case, in both cases, values of the remaining bits remain unchanged: $x_1=y_1$, $x_2=y_2$,..., $x_{r-1}=y_{r-1}$.

Stage 3. Quantitative equivalent $\mathrm{dec}X_j$ of the binary (n,k)-binomial number $X_j=x_1x_2...x_i...x_r$ is calculated in accordance with the numerical function (1).

Stage 4. The quantitative equivalent of $\mathrm{dec}X_j$ is converted to its binary form $D_j=\mathrm{Bin}(\mathrm{dec}X_j)$ to obtain thereby the sought number $D_j\in D[n,k]$, a compressed image of the original n-digit sequence $Y_j=y_1y_2...y_i...y_n$.

The following Theorem 2 gives properties of f_e^{-1} mapping of restoration and the way of its practical implementation («++» is the symbol of the concatenation operation).

Theorem 2. Any binary number $D_j\in D[n,k]$ can be put in one-to-one compliance with a unique binary sequence $Y_j=y_1y_2...y_i...y_n$, $Y_j\in Y[n,k]$, $j=\overline{1,C_n^k}$, composed of n bits y_i having the sum of values equal to k using the f_e^{-1} mapping in two stages:

1) calculating the values of the bits $x_i\in\{0,1\}$ of nonuniform (n,k)-binomial number $X_j=x_1x_2...x_i...x_r$, $X_j\in X[n,k]$, $r<n$ in accordance with the recurrent relation: $\mathrm{sign}(\mathrm{sign}(F(i)-\rho_i)+1)$ where $F(1)=F_j$, $F(i+1)=F_j-x_i\,\rho_i$, $\rho_i=C_{n-i}^{k-q_i}$, $F_j=\mathrm{dec}D_j$;

2) transition from a binary (n,k)-binomial number $X_j=x_1x_2...x_i...x_r$ to a binary sequence $Y_j=y_1y_2...y_i...y_n$ using a function of the form:

$$Y_j=y_1y_2...y_i...y_n=\begin{bmatrix} x_1x_2...x_i...x_{r-1}0++11...1, \\ x_1x_2...x_i...x_{r-1}1++00...0. \end{bmatrix}$$

Theorem 2 is proved for the first stage of transformations proceeding from existence and uniqueness of decomposition (1). For the second stage, it is substantiated by the fact that the sequences obtained by concatenation of binomial numbers $X_j=x_1x_2...x_i...x_r$ and series 11...1 or 00...0 are equilibrium Y_j combinations. Moreover, Y_j are unique for the corresponding X_j since the concatenation operation does not change values of the x_i digits.

The form of the complex function $f_e^{-1}=f_b^{-1}\circ\psi^{-1}$ reflects the fact that the steps related to the f_b^{-1} restoration based on binary binomial numbers [12] will be included in the model of the f_e^{-1} restoration process based on the binomial numerical function.

The f_e mapping under consideration is bijective since the compliances $D_j=f_e(Y_j)$ and $Y_j=f_e^{-1}(D_j)$ are functional (Theorems 1, 2).

When applying the binomial numeric function (1), there is a more general case when k can take any values from the range $0\le k\le n$ and the compressed array is a set $A=\bigcup_{k=0}^{n}Y[n,k]=\{0,1\}^n$ of binary n-digit sequences A_j $j=\overline{1,2^n}$, for which there is no limitation on the number k of ones.

Binary numbers D_j are uniform in length only for a constant value of k. Hence, to unambiguously restore $A_j\in\{0,1\}^n$, from the D_j numbers, it will be necessary to additionally use the k value expressed in binary form $\mathrm{Bin}\,k$. This means that the $\mathrm{Bin}\,k$ should be appended to the resulting binary numbers D_j. This approach is justified if $0<k<n$. For the case when $(k=0)\vee(k=n)$, to unambiguously restore the sequence of zeros and ones, it is sufficient to transmit and store only the binary word $\mathrm{Bin}\,k$. Therefore, additional coding is introduced in the f_k form:

1) if $(k=0)\vee(k=n)$, then f_k:$Y[n,k]\rightarrow Z_o$, $Z_o=\{Z_j/Z_j=\mathrm{Bin}\,k\}$;

2) if $0<k<n$, then f_k:$D[n,k]\rightarrow Z_e$, $Z_e=\{Z_j/Z_j=(\mathrm{Bin}\,k,D_j)\}$.

Also, in the general case for $0\le k\le n$ when compressing A_j, it is necessary to use the function f_w:$A\rightarrow M$, $M=\{(k,Y_j)/0<k<n,Y_j\in Y[n,k]\}$ which sets the original sequence A_j in compliance with the sample $(k,Y_j)\in M$ where $Y_j=A_j$. Further, if the obtained value of k satisfies the inequality $0<k<n$, then to compress the equilibrium combination Y_j complying with A_j, f_e coding based on a binomial numerical function is used. In this case, $\mathrm{Bin}\,k$ is added to the compressed combinations for unambiguous restoration, i. e., f_k is additionally encoded. If the k value satisfies the system of equalities $(k=0)\vee(k=n)$, then the encoded resulting combination will consist only of $\mathrm{Bin}\,k$, i. e., only the f_k encoding method is used.

Thus, the general case of compression when $0\le k\le n$, is represented by a mapping in the form f_{eg}:$A\rightarrow Z$ which is set by the corresponding function $Z_j=f_{eg}(A_j)$ where $A_j=a_1a_2...a_i...a_n$, $A_j\in A=\{0,1\}^n$, $Z_j=(\mathrm{Bin}\,k,D_j)$ or $Z_j=\mathrm{Bin}\,k$, $Z_j\in Z$, $j=\overline{1,2^n}$. Theorem 3 below gives properties of f_{eg} mapping and the way of its realization.

Theorem 3. Any binary sequence $A_j=a_1a_2...a_i...a_n$, $A_j\in A=\{0,1\}^n$, $j=\overline{1,2^n}$, can be put into a one-to-one compliance with a binary combination $Z_j\in Z$ of the following form:

1) if $0<k<n$, then $Z_j=\mathrm{Bin}\,k+D$, where $k=\sum_{i=1}^{n}a_i$, $D_j=f_e(Y_j)$, $D_j\in D[n,k]$, $Y_j\in Y[n,k]\subset A$;

2) otherwise, if $(k=0)\vee(k=n)$, then $Z_j=\mathrm{Bin}\,k$.

Proof of Theorem 3 is based on the fact that for the case of $0<k<n$ there cannot be two or more numbers of k ones and the numbers D_j (Theorems 1, 2) for the compressible sequence A_j and for the case of $(k=0)\vee(k=n)$, the validity of the theorem is confirmed by the uniqueness of A_j in which all digits are either zeros or ones.

The mapping f_{eg}:$A\rightarrow Z$ is bijective since each element of A_j has a unique image and each element of Z_j is the only preimage for all $A_j\in A$ and $Z_j\in Z$.

Thus, the mapping f_{eg}:$A\rightarrow Z$ which is given by a complex function in the form:

$$f_{eg}=\begin{cases} f_k\circ f_w,\ (k=0)\vee(k=n), \\ f_k\circ f_e\circ f_w,\ 0<k<n, \end{cases} \tag{6}$$

is called a generalized method of compression based on a binomial numerical function (1) or a generalized binomial-numbering compression.

Obviously, the previously considered compression f_e is a composite component of the generalized compression f_{eg} based on the binomial numerical function (1). Taking into account that $f_e=\psi\circ f_b$, binary (n,k)-binomial numbers are intermediate code objects in the process of information transformations during f_{eg} compression. Such relationships between f_b, f_e and f_{eg} simplify practical implementation when they are used together, expand their functionality, and make it possible to use an adaptive approach to binary compression of information using binary binomial numbering systems.

In turn, the inverse mapping $f_{eg}^{-1}: Z \to A$, which is given by the inverse complex function:

$$f_{eg}^{-1} = \begin{cases} f_w^{-1} \circ f_k^{-1}, \ (k=0) \vee (k=n), \\ f_w^{-1} \circ f_e^{-1} \circ f_k^{-1}, \ 0<k<n, \end{cases} \tag{7}$$

is a recovery taking into account the existing value k of the initial binary sequences A_j.

In the case when $0<k<n$, A_j is recovered on the basis of Bin k and binary numbers $D_j \in D[n,k]$ using binomial numbers $X_j \in X[n,k]$ when passing from the number D_j to the equilibrium combination Y_j. In the case of $k=0$ or $k=n$, A_j is recovered based on Bin k by generating n zeros or ones, respectively. This type of recovery will be called generalized recovery based on a binomial numerical function.

Mappings f_e and f_e^{-1} are constituent elements of the generalized compression and recovery methods f_{eg} and f_{eg}^{-1}, respectively. Hence, the methods of constructing complex functions (6) and (7) will be in many respects similar to the methods of implementing the methods of compression and restoration of equilibrium combinations (Theorems 1 and 2) when $0<k<n$. Methods of implementing the functions of f_{eg} compression (6) and f_{eg}^{-1} recovery (7) in subdomains of definition $(k=0) \vee (k=n)$ are determined by a fairly simple operation of calculating k ones when compressing f_{eg} and forming zero or one A_j in f_{eg}^{-1} recover.

Models of the processes of generalized f_{eg} compression and f_{eg}^{-1} recovery based on the binomial numerical function (1) follow from Theorem 3 substantiating the one-to-oneness of mappings f_{eg} and f_{eg}^{-1}, as well as Theorems 1 and 2.

Modeling of the process of f_{eg} compression of binary sequences $A_j=a_1a_2...a_i...a_n$, $A_j \in A=\{0,1\}^n$, $j=\overline{1,2^n}$, consists of stages which are considered as compound stages and ones related to the previous model of f_e compression.

Stage 1. The number s of bits is calculated for the binary Bin k representation of the number k of ones, $0 \leq k \leq n$, and the initial n-digit sequence $A_j=a_1a_2...a_i...a_n$: $s=\lceil \log_2(n+1) \rceil$.

Stage 2. The number $k=\sum_{i=1}^{n} a_i$ of binary ones in the original n-digit sequence $A_j=a_1a_2...a_i...a_n$ is calculated, thereby realizing the function $f_w(A_j)=(k,Y_j)$ and determining the class $Y[n,k]$ of equilibrium combinations to which $Y_j \in Y[n,k]$, $Y_j=A_j$ belongs.

Stage 3. The number k of ones is converted to its binary form Bin k consisting of s bits.

Stage 4. If the number k satisfies the system of equalities $(k=0) \vee (k=n)$, then $(k=0) \vee (k=n)$ will be the resulting combination. Otherwise, the available n value and calculated k-value are parameters of the binary (n,k)-binomial numbering system and the transition to the subsequent stages is made to implement coding $f_k(f_e(Y_j))=Z_j$, $Z_j \in Z_e$.

Stage 5. The number m of bits is calculated for binary representation of the number $D_j \in D[n,k]$ corresponding to $Y_j \in Y[n,k]$: $m=\lceil \log_2 C_n^k \rceil$.

Next, *Stages* 6–9 of the model of the f_{eg} compression process correspond, respectively, to *Stages* 1–4 of the model of the f_e compression process.

Stage 10. Concatenation of the binary value Bin k and the number D_j, i. e., encoding of the form $f_k(D_j)=Z_j$ is carried out for the case $0<k<n$: $Z_j=$Bin $k+D_j$, thereby obtaining the resulting combination $Z_j \in Z_o$.

Modeling the process of f_{eg}^{-1} restoring the sequences $A_j=a_1a_2...a_i...a_n$, $A_j \in A=\{0,1\}^n$, $j=\overline{1,2^n}$, from combinations-images Z_j, $Z_j \in Z_o \cup Z_e$, is carried out based on Theorems 2 and 3.

5. The results obtained in the study of compression methods

5. 1. Estimation of compression of equilibrium combinations based on the binomial numerical function

The results of compression f_e:$Y[8,2] \to D[8,2]$ for some $Y_j \in Y[8,2]$ using the corresponding intermediate $X_j \in X[8,2]$ are shown in Table 1. For a complete set $Y[8,2]$ when average length $L_{8,2}$ of binomial numbers $X_j \in X[8,2]$ is 5.71 bits (according to [10]) and the number of binary digits to represent the number is $D_j \in D[8,2]$ is $L_n=\lceil \log_2 C_8^2 \rceil=5$.

Thus, according to Table 1, the compression ratio of 8-bit equilibrium combinations $Y_j \in Y[8,2]$ when mapping f_e will be $K_e=n/L_n=8/5=1.6$. It is also obvious that if the degree of binomial compression should be increased, it is necessary to switch from f_b compression based on binomial numbers to $f_e=\psi \circ f_b$ compression based on the binomial numerical function (1).

In general terms, the ratio K_e of f_e compression is a ratio $K_e=n/\lceil \log_2 C_n^k \rceil$, where the denominator is the number of digits to represent the binary number D_j. Table 2 demonstrates a variation of the coefficient K_e depending on value k of ones in the 128-bit equilibrium combinations $Y_j \in Y[128,k]$ having a length of $n=128$.

Table 1

Compliance of some $Y_j \in Y[8,2]$, $X_j \in X[8,2]$ and their numbers with $D_j \in D[8,2]$

No.	Equilibrium code $Y[8,2]$								Set $X[8,2]$							Binary number				
0	0	0	0	0	0	0	1	1	0	0	0	0	0	0		0	0	0	0	0
7	0	0	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	1	1	1
10	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	1	0
16	0	1	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0	0	0
19	0	1	0	1	0	0	0	0	0	1	0	1				1	0	0	1	1
24	1	0	0	0	1	0	0	0	1	0	0	0	1			1	1	0	0	0
27	1	1	0	0	0	0	0	0	1	1						1	1	0	1	1
	$n=8$								$L_{8,2} \approx 5{,}71$							$L_n=5$				

Table 2

K_e values depending on k of ones at $n=128$

n	K_e
1	18.29
8	3.12
15	2.00
22	1.56
29	1.33
36	1.20
43	1.13
49	1.07
56	1.04
63	1.02
64	1.02
71	1.04
78	1.07
85	1.13
92	1.20
99	1.33
106	1.56
113	2.00
120	3.12
127	18.29

As seen from Table 2, compression of equilibrium combinations $Y_j \in Y[128,k]$ based on the binomial numerical function (1) is observed for any fixed value $1 \le k \le 127$ of binary ones (because of the uniqueness of corresponding combinations, the values $k=0$ and $k=128$ are not considered). In this case, a minimum value of the ratio K_e of compression f_e is 1.02, and the maximum is 18.29.

5. 2. Estimation of compression of binary sequences of general form based on the binomial numerical function

When applying the generalized binomial f_{eg} compression based on the binomial numerical function (1), one should additionally calculate the number k which is now already a variable and concatenate Bin k and the resulting number D_j.

On the one hand, the use of Bin k slightly reduces the resulting compression ratio but, on the other hand, it significantly expands the scope of f_{eg} compression since it can be used for ordinary information arrays. It should also be noted that the Bin k/n ratio tends to zero at $n \to \mu$ which means a decrease in Bin k influence on the compression ratio with an increase in length n of the compressed sequences A_j.

Some results of the considered mapping $f_{eg}: A \to Z$ and, therefore, $f_{eg}^{-1}: Z \to A$ for $n=24$ where $A_j \in A=\{0,1\}^{24}$, $j=\overline{1,2^{24}}$, are given in Table 2. If $0<k<23$, then a complex encoding function $f_k \circ f_e \circ f_w$, is used. Otherwise, i. e. when $(k=0) \vee (k=24)$, a complex encoding function $f_k \circ f_w$ is applied. The ratio of lengths of the binary representation of A_j and Z_j in this Table varies from 0.89 to 4.8 and their average value is 2.26 for the entire Table. For each sequence A_j, Table 3 shows both corresponding binary $(24,k)$-binomial number X_j used in the transition to the number and the resulting number D_j itself.

Table 3

Compliance of f_{eg} between some binary A_j and Z_j for $n=24$

Binary sequence A_j	Binary combination Z_j: Bin k	Binary combination Z_j: Digit X_j or blank string / Number D_j or blank string	Type of f_{eg}
000000000000000000001001	00010	00000000000000000000100 000000011	$f_k \circ f_e \circ f_w$
000000001000011000000000	00011	000000001000011 00111111101	$f_k \circ f_e \circ f_w$
011111011001010000101100	01100	01111101100101000010111 0101000110100101110100	$f_k \circ f_e \circ f_w$
111111111111111111111110	10111	11111111111111111111111 10111	$f_k \circ f_e \circ f_w$
111111111111111111111111	11000		$f_k \circ f_w$

In a general form, ratio K_{eg} of compression f_{eg} is the ratio $K_e = n / \left(\lceil \log_2(n+1) \rceil + \lceil \log_2 C_n^k \rceil \right)$, where the denominator is the sum of the number of digits to represent the binary Bin k and the number D_j. Table 4 shows the variation of the coefficient K_{eg} depending on the value k of ones in the 128-bit binary sequences $A_j \in A=\{0,1\}^{128}$, $j=\overline{1,2^{128}}$. The number k of ones can have any variable value in the range of $0 \le k \le 128$.

As seen from Table 4, compression of binary sequences of general form A_j based on the binomial numerical function (1) is observed at values of $0 \le k \le 48$ and $80 \le k \le 128$ binary ones. In these ranges of k variation, a minimum value of ratio K_{eg} of f_{eg} compression is 1.01, and the maximum one is 16.

Table 4

Values of K_{eg} depending on k of ones at $n=128$

k	K_{eg}
0	16
8	2.61
16	1.71
24	1.36
32	1.17
40	1.08
48	1.01
56	0.98
64	0.96
72	0.98
80	1.01
88	1.08
96	1.17
104	1.36
112	1.71
120	2.61
128	16

In the case of $48<k<80$, the f_{eg} compression ratio will have the value of $K_{eg}=0.98$ or $K_{eg}=0.96$ (for $k=64$), i. e., less than one which is the consequence of including Bin k in the resulting combination for unambiguous recovery of A_j.

6. Discussion of results obtained in the study of compression methods based on the binomial numerical function

The mathematical model (4) of the process of compression f_e of equilibrium combinations Y_j contains the function f_b of obtaining binary binomial numbers X_j from Y_j. This makes it possible to significantly reduce the number of computations when determining the binary number D_j according to (1), since length r of the numbers X_j is on average significantly less than the length n of combinations Y_j.

Therefore, there is no need to use $(n-r)$ weight coefficients $C_{n-i}^{k-q_i}$, as it is done in [6, 7]. As Table 2 shows, information redundancy in equilibrium combinations can be eliminated from 1.02 to 18.29 times. Accordingly, the speed of information transmitted through the communication channel can be increased the same number of times.

The formulated Theorems 1 and 2 determine the one-to-oneness of the compression coding of f_e and form the basis for constructing a model of the process of compression f_e. The demonstrated stages of f_e modeling are characterized by simple operations of decatenation and summation of binomial coefficients $C_{n-i}^{k-q_i}$ according to expression (1) which ensures lower hardware and software costs in practical implementation.

The mathematical model (6) of the process of f_{eg} compression of binary sequences of a general form A_j contains f_e compression coding as a composite function, and therefore also uses binary binomial numbers X_j. Accordingly, this leads to a reduction in computational costs when searching for the D_j number, as it does for f_e mapping. Inclusion of function f_k of the concatenation of binary values Bin k and the D_j number in model (6) leads to a decrease in compression ratio though this decrease is insignificant. As Table 4 shows, elimination of information redundancy in binary sequences of a general form is from 1.01 to 16 times for $n=128$. Compared to f_e compression, the use of Bin k leads to a decrease in compression ratio by approximately 6–16%. However, for $n\to\mu$, the influence of Bin k on the compression ratio f_{eg} becomes more and more insignificant since the number of $\lceil \log_2(n+1) \rceil$ bits for the Bin k representation grows much slower than that for n.

The formulated Theorem 3 determines the unambiguity of f_{eg} coding and is the basis for constructing a model of the process of the corresponding compression. The f_{eg} modeling steps shown include the model of f_e compression which makes it easier to synthesize them together. Additional operations of calculating the number k of ones and concatenating the Bin k to the compressed image D_j feature simplicity and minimal hardware and software costs in practical implementation.

As indicated in [14], the compression coefficient for the methods based on the binomial numerical function at $n\to\mu$ asymptotically tends to the limiting value, i. e., reciprocal of the entropy $H(p)$ where p is the probability of appearance of one.

Not only the compression ratio but also the time of encoding and decoding of binary sequences have a significant impact on compression efficiency from the point of view of performance improvement [1–3].

Inclusion of the function f_w in model (6) for calculating k of ones increases the time of f_{eg} compression by a maximum of $\lceil n/2 \rceil$ computation cycles [12]. But this is significantly less than the time required to get the actual number D_j.

The methods of f_e and f_{eg} compression are based on the numerical function (1) and calculation of binomial coefficients $C_{n-i}^{k-q_i}$, that play the role of weights in determining the numbers D_j of compressed binary sequences. The degree of binomial numbering compression is in direct proportional dependence on the length n and the number k of ones of a binary combination. However, an increase in n and k leads to a significant increase in computational costs when determining weights $C_{n-i}^{k-q_i}$ of the numerical function (1) which, in turn, leads to a noticeable increase in costs for hardware, software, and/or timetable [15, 16].

For the practical implementation of f_e and f_{eg} compression methods, it is proposed to use the tabular method of calculating binary numbers D_j in which all weight coefficients $C_{n-i}^{k-q_i}$ pre-calculated for parameters n and k will be used as table elements. The timetable here will be determined only by adding q values of $C_{n-i}^{k-q_i}$. Then the hardware costs will consist of the number of memory cells required to store all the weight coefficients, and the time spent from the time of adding the weight coefficients and the timetable consists of calculating the number k of ones which is limited by the value $k\leq\lceil n/2 \rceil$.

It should also be noted that knowledge of k in the case of compression f_e of equilibrium combinations and the calculated value of k in the case of generalized binomial compression f_{eg} make it possible to control errors. If, other values of k are obtained that differ from the original ones when transmitting compressed images through communication channels and their subsequent restoration, this indicates erroneousness of the resulting sequences. The obtained feature of compression methods based on the binomial numerical function (1) is an additional way to ensure the specified validity of data transmission and the level of fault tolerance of the information systems.

Thus, in contrast to [11–13], mathematical models of f_e and f_{eg} compression were obtained in this study based on the binomial numerical function (1) and stages of modeling the corresponding processes were also presented.

In contrast to [4, 5, 14], the advantage of this study consists in the abstraction of binary binomial numbers in the structure of compressible binary sequences. This makes it possible to reduce the time spent on calculating the number D_j according to the binomial numerical function (1) and introduce control of errors that may arise during the transfer of compressed images and their subsequent recovery.

As a disadvantage of the above study, the absence of models of processes of recovery of f_e^{-1} and f_{eg}^{-1} and initial combinations based on the binomial numerical function (1) can be pointed out. This is explained just by the limited volume of this paper.

Further development of the proposed compressive coding methods is seen in developing algorithms and devices for practical binomial compression as well as estimating the compression degree and time depending on probabilistic characteristics of binary information sources.

The exemplary area of application of the developed compression methods based on the binomial numeric function includes the following:

– data transmission systems using equilibrium or quasi-equilibrium codes [17];

– database management systems storing information sequences, the sum of the elements in which is limited or equal to a set limit [6, 13];

– systems for automation of scientific experiments in which their statistical characteristics are unknown at the preliminary stage of data processing;

– information and control systems using binomial sets of numbers for solving specialized problems [12].

7. Conclusions

1. A mathematical model of the method of compressing equilibrium combinations based on a binomial numerical function has been constructed. This model is based on Theorems 1 and 2 revealing the procedure of the method implementation and makes it possible to implement sufficiently effective models of the processes of binomial numbering compression and restoration of equilibrium combinations. In this case, binary binomial numbers are used as intermediate numbers in the transition to binary numbers.

2. A mathematical model of the generalized compression method based on a binomial numerical function for binary sequences with a variable number of ones has been constructed. Theorem 3 revealing the procedure of implementing this method and enabling the implementation of sufficiently effective models of processes of binomial generalized compression and restoration of information sequences of a general form is the basis of the constructed mathematical model. In this case, the numerical compression models of equilibrium combinations are an integral part of the generalized compression models based on the binomial numerical function.

References

1. Sayood, K. (2018). Introduction to Data Compression. Morgan Kaufmann, 790. doi: https://doi.org/10.1016/C2015-0-06248-7
2. Sayood, K. (2003). Lossless Compression Handbook. Academic Press, 488. doi: https://doi.org/10.1016/B978-0-12-620861-0.X5000-1
3. Smirnov, M. A. (2004). Obzor primeneniya metodov bezuscherbnogo szhatiya dannyh v SUBD. Saint Petersburg. Available at: http://compression.ru/download/articles/db/smirnov_2003_database_compression_review.pdf
4. Krasnobayev, V., Kuznetsov, A., Koshman, S., Rassomakhin, S., Zamula, A., Kavun, S. (2018). Effective Data Processing in Coding, Digital Signals and Cryptography. ASC Academic Publishing, 352. Available at: https://www.amazon.com/effective-processing-digital-signals-cryptography-ebook/dp/B07CSJHHN1
5. Yuan, Y., Zhang, Y., Liu, Z., Guan, X. (2017). Lossless coding scheme for data acquisition under limited communication bandwidth. Digital Signal Processing, 69, 204–211. doi: https://doi.org/10.1016/j.dsp.2017.06.028
6. Amel'kin, V. A. (1986). Metody numeratsionnogo kodirovaniya. Novosibirsk: Nauka, 155.
7. Amel'kin, V. A. (2008). Perechislitel'nye zadachi seriynyh posledovatel'nostey. Novosibirsk: IVMiMG SO RAN, 317.
8. Stakhov, A. P. (2014). A History, the Main Mathematical Results and Applications for the Mathematics of Harmony. Applied Mathematics, 05 (03), 363–386. doi: https://doi.org/10.4236/am.2014.53039
9. Polyakov, V. I., Skorubsky, V. I., Ekalo, Yu. V. (2015). Application of Factorial System to Combinatorial Problems Solving. Izvestiâ vysših učebnyh zavedenij. Priborostroenie, 58 (6), 436–442. doi: https://doi.org/10.17586/0021-3454-2015-58-6-436-442
10. Borisenko, A. A. (2004). Binomial'niy schet. Teoriya i praktika. Sumy: ITD «Universitetskaya kniga», 170. Available at: https://essuir.sumdu.edu.ua/handle/123456789/55161
11. Kulyk, I., Shevchenko, M. (2020). Development of information-management systems on basis of binary binomial number systems. Information Processing Systems, 2 (161), 78–85. doi: https://doi.org/10.30748/soi.2020.161.09
12. Kulyk, I., Berezhna, O., Shevchenko, M. (2018). Development of data compressing coding methods on basis of binary binomial numbers. Technology Audit and Production Reserves, 2 (2 (46)), 12–18. doi: https://doi.org/10.15587/2312-8372.2019.169897
13. Kulyk, Y. A., Novhorodtsev, A. Y., Skordyna, E. M. (2016), Binomial model of vector representation for database with columnar structure. Information Processing Systems, 4 (141), 50–55. Available at: http://www.hups.mil.gov.ua/periodic-app/article/16556
14. Schalkwijk, J. (1972). An algorithm for source coding. IEEE Transactions on Information Theory, 18 (3), 395–399. doi: https://doi.org/10.1109/tit.1972.1054832
15. Knuth, D. (1997). The Art of Computer Programming, Vol. 1: Fundamental Algorithms. Addison-Wesley Professional, 672. Available at: https://www.amazon.com/Art-Computer-Programming-Vol-Fundamental/dp/0201896834
16. Anderson, J. A., Anderson, J., Bell, J. (2000). Discrete mathematics with combinatorics. Prentice Hall College Div, 799. Available at: https://www.amazon.com/Discrete-Mathematics-Combina-torics-James-Anderson/dp/0130869988
17. Kohzuki, K., Tokiwa, K., Tanaka, H. (1997). A class of single error correcting constant weight codes. Electronics and Communications in Japan (Part III: Fundamental Electronic Science), 80 (7), 55–64. doi: https://doi.org/10.1002/(sici)1520-6440(199707)80:7<55::aid-ecjc7>3.0.co;2-0