

Article

The Lossless Adaptive Binomial Data Compression Method

Oleksiy Borysenko ¹, Svitlana Matsenko ^{2,*}, Toms Salgals ^{2,3}, Sandis Spolitis ^{2,3} and Vjaceslavs Bobrovs ³¹ Department of Electronics and Computer Technology, Sumy State University, 40007 Sumy, Ukraine² Communication Technologies Research Center, Riga Technical University, 1048 Riga, Latvia³ Institute of Telecommunications, Riga Technical University, 1048 Riga, Latvia* Correspondence: svitlana.matsenko@rtu.lv

Abstract: In this paper, we propose a new method for the binomial adaptive compression of binary sequences of finite length without loss of information. The advantage of the proposed binomial adaptive compression method compared with the binomial compression method previously developed by the authors is an increase in the compression rate. This speed is accompanied in the method by the appearance of a new quality—noise immunity of compression. The novelty of the proposed method, which makes it possible to achieve these positive results, is manifested in the adaptation of the compression ratio of compressible sequences to the required time, which is carried out by dividing the initial set of binary sequences into compressible and incompressible sequences. The method is based on the theorem proved by the authors on the decomposition of a stationary Bernoulli source of information into the combinatorial and probabilistic source. The last of them is the source of the number of units. It acquires an entropy close to zero and practically does not affect the compression ratio at considerable lengths of binary sequences. Therefore, for the proposed compression method, a combinatorial source generating equiprobable sequences is paramount since it does not require a set of statistical data and is implemented by numerical coding methods. As one of these methods, we choose a technique that uses binomial numbers based on the developed binomial number system. The corresponding compression procedure consists of three steps. The first is the transformation of the compressible sequence into an equilibrium combination, the second is its transformation into a binomial number, and the third is the transformation of a binomial number into a binary number. The restoration of the compressed sequence occurs in reverse order. In terms of the degree of compression and universalization, the method is similar to statistical methods of compression. The proposed method is convenient for hardware implementation using noise-immune binomial circuits. It also enables a potential opportunity to build effective systems for protecting information from unauthorized access.

Keywords: data compression; binomial code; lossless data compression; binomial numbers

Citation: Borysenko, O.; Matsenko, S.; Salgals, T.; Spolitis, S.; Bobrovs, V. The Lossless Adaptive Binomial Data Compression Method. *Appl. Sci.* **2022**, *12*, 9676. <https://doi.org/10.3390/app12199676>

Academic Editors: Aleksandr Cariow and Oleg Finko

Received: 26 July 2022

Accepted: 23 September 2022

Published: 26 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With an increase in the volume of transmitted information and speed and reliability, the task arises of an effective method for reducing the volume of transmitted data, mainly using lossless data compression methods. Various lossless compression algorithms are widely used in areas such as medicine [1], industrial [2], the Internet of Things (IoT) [3], databases [4], cloud computing [5], especially communication networks [6], and space data system [7].

Finding an effective data compression algorithm for any application remains relevant [8]. Modern lossless data compression methods combine entropy coding and general-purpose compression techniques. For example, the main algorithms of universal compressions, such as LZ77/78 (Lempel–Ziv coding 1977/1978), LZW (Lempel–Ziv–Welch), and LZMA (Lempel–Ziv–Markov chain algorithm), are based on a dictionary, zip (LZ77 + Huffman), gzip (LZ77 + Huffman), 7-zip (LZMA + arithmetic coding) and bzip2

(block sorting + Move-to-Front + Huffman) [9,10]. Also used is LZSS [10,11], which optimizes matching operations in the look-up table to assign fewer bits for compressed data than LZ77 [12–14].

One of the main compression methods is the class of statistical compression methods since their implementation requires knowledge only of the probability distribution of compressed messages. They are mainly intended for communication systems and are quite versatile for them. Their main disadvantage is the need for statistical data. In parallel with statistical methods, other compression methods are being developed that, as a rule, are of a special nature and do not require direct statistical tests. One such method is the numbering method, which usually compresses mathematical objects such as permutations or combinations. Their use is applicable in information security and error protection systems. In [15], for the need to improve health information systems for health monitoring, lossless electroencephalogram (EEG) data compression is applicable [16], and proposed a lossy EEG compression model based on fractals for network traffic. manifest the advantage of this method. In [17], a simpler alternative to arithmetic coding is presented. The advantage of this method is its simplicity and the possibility of using it for simultaneous data encryption or as a block of cryptosystems. The disadvantage of this method is that it requires the creation and storage of encoding tables and decoding. In [18], a mathematical model of the method of compression of equilibrium combinations based on binary binomial numbers and a mathematical model of generalized binomial compression of binary sequences are shown, but this method is not adaptive. Meanwhile, [19] presents a mathematical model for compressing binary sequences by binomial numbers.

This paper proposes a universal adaptive method for compressing binary sequences of unlimited length n without losing information. The code length of the original binary sequence n will be greater than the length of the number m , which determines the compression effect. Its value is determined by the length difference between the original code length n of the binary sequence and the compressed length m . The more significant this difference, the greater the compression effect. It will take the largest value in the case when the original binary sequence consists of either only zeros ($k = 0$) or only ones ($k = n$), and the smallest value when $k = n/2$. In the first case, the compression effect will be determined only by the number of bits required to transmit the value k , which is always less than n . In this case, the compressed binary sequence number is not transmitted.

Therefore, this case will be the most favorable for compression. In practice, it occurs, for example, in television binary images, where there are empty or completely dark lines. In the second case, when $k = n/2$, the code length of the number m approximately coincides with the code length of n . Accordingly, when transmitting the values m and n , instead of compressing the binary sequence, its lengthening will be observed due to redundant coding since the transmission of the length of the compressed number m is also added to the transmission of the length of the binary number for the number of units k . In all other cases of compression of binary sequences most often encountered in practice, the inequality $n > k > 0$ takes place.

This inequality allows the compression to be adapted to its required speed. A restriction is introduced on the value of k , the excess of which prohibits the compression of binary sequences. Changing this value reduces or increases the information transfer rate due to the compression ratio. The possibility of such regulation is the effect of the adaptability of the proposed compression method. When restrictions generally prohibit the compression of a number of binary sequences, it is possible to increase their noise immunity by transmitting only the k value to the receiver during compression.

The compression effect of the proposed method depends on the probabilities of the distribution of the number of units in the compressed binary code sequences. The greater the likelihood of sequences with a number of ones different from $k = n/2$, the greater this effect will be. However, it cannot exceed the bounds given by Shannon's equation for the entropy of binary sequences. Therefore, this method is comparable in terms of compression efficiency with existing lossless information compression methods. Its main

advantage over them is the ability to change the compression efficiency depending on the requirements for its performance. Another essential property of this binomial compression method is it increases the noise immunity of a part of transmitted binary messages. In addition, using the binomial number system in this compression method makes it possible to effectively implement it in hardware while increasing data compressions speed, reliability, and noise immunity.

A distinctive feature of the proposed method for compressing binary sequences of finite length n is their numbering using the binomial number system (BNS). In addition, the method allows the compression ratio K to be adapted to its required speed and error detection. Adaptation manifests itself in the fact that for each compressed sequence, the compression coefficient K is preliminarily determined, and, depending on its value and the requirements for speed and error detection, a decision is made whether or not to perform the compression operation. As a result, only those sequences are compressed for which the compression coefficients K are more significant than the specified value. The remaining possible sequences are transmitted or stored in their original form. Using the binomial number system and binomial numbers in the method makes it possible to implement it in hardware. The availability of keys for each compressed sequence can be used in systems to protect information from unauthorized access.

The rationale for developing methods for compressing information based on binary binomial numbers generated by binomial number systems is:

- (1) The non-uniformity of binomial numbers whose length r is less than the length n of the original compressible code combinations, which provides a compression coefficient more significant than one;
- (2) The functionality of correspondences between sets of binary binomial numbers and combination codes and provides one-to-one coding and decoding;
- (3) The prefix of binary binomial numbers, which allows, without additional hardware (and/or software) and time costs for separators, compression coding, and restoration of binary sequences;
- (4) Prevalence of code combinations based on binomial numbers (for example, equilibrium and quasi-equilibrium codes, combinations with restrictions on mutual arrangement of zeros and units, etc.) for data presentation in information-control systems.

The method of compressing data based on binary binomial numbers has properties shown in Figure 1:

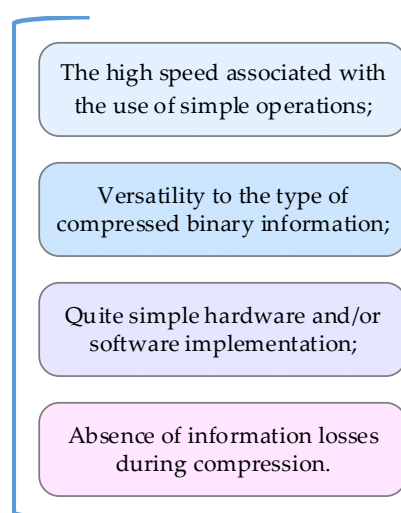


Figure 1. Properties of the method of compressing data based on binary binomial numbers.

In addition, an essential advantage of the method is that compressed sequences are endowed with numerical characteristics that are expressed by their corresponding

binomial numbers. Thus, information is compressed without loss, adapting to speed and error detection.

The paper is structured as follows. Section 2 describes the general analysis of the binomial adaptive compression method. Section 2.1 describes the binomial decomposition of information sources, while Section 2.2 describes the analysis of the proposed compression method. Section 3 describes the lossless adaptive data compression method based on the binomial number system (BNS). Section 3.1 describes the data compression method with BNS, and Section 3.2 describes the methods for converting binomial code to binary. Section 3.3 describes the methods for converting binary sequences to binomial code and vice versa. Theoretical aspects of the lossless adaptive binomial data compression method are described in Section 4. Finally, we conclude this paper in Section 5.

2. General Analysis of the Binomial Adaptive Compression Method

2.1. The Binomial Decomposition of Information Sources

The lossless binomial adaptive compression of binary sequences of length n proposed by the authors is based on the binomial information compression they developed based on the theorem on the decomposition of Bernoulli sources of information [20]. The theorem proves that the Bernoulli source of probabilistic binary sequences A^* decomposes into two sources, one of which is a combinatorial source A with conditional entropy

$H(A/B) = \sum_{k=1}^n \hat{P}_k \log_2 C_n^k$, generating equally probable binary sequences of different code

lengths, and another probabilistic source of information B with entropy

$H(B) = -\sum_{k=1}^n \hat{P}_k \log_2 \hat{P}_k$, generating quantities of k number of units 0, 1, ..., n in these

sequences.

Probabilities \hat{P}_k determine the probabilities of occurrence of these quantities. It is proved that the sum of the entropies (entropy of the union) of the sources A and B are

equal to the entropy $H(A^*) = -\sum_{j=1}^{2^n} p_j \log_2 p_j$ original source of binary sequences A^* , where

p_j —source generation probability A^* of j th binary code combination.

The Bernoulli information source entropy equation is decomposed into two entropies. They provide the basis for new compression methods, one of which is discussed in this article. The entropy decomposition equations show that the Shannon–Fano and Huffman statistical methods of optimal coding can be effectively replaced by numerical coding methods that do not require preliminary statistical tests for their implementation. In this case, the compression efficiency in practice is not less than in statistical methods of optimal coding. With that compression occurring with zero errors in statistics, it is even higher. This idea underlies the compression method [20] previously presented by the authors, which was carried out using binomial numbers [21–23]. However, compression occurs for different binary sequences with different efficiencies. The difference here depends on the probability distribution of generating compressible sequences.

To eliminate the time spent on compression, it was proposed to abandon the compression of inefficiently compressible sequences. As a result, practically without reducing the compression efficiency, it was possible to reduce the compression time several times while increasing its noise immunity. However, this process requires the authors' dynamic control system called adaptive coding. The solution to the problem of improving efficiency proposed by the authors formed the basis of this paper. The binomial compression method itself, without adaptation, has long been studied by the authors and has shown its practical performance. This adaptive compression method development has been proposed for the first time for publication. The authors also developed a program that confirms the practical performance of the adaptive compression method. In any case, the efficiency of the proposed adaptive compression method is higher than without adaptation.

2.2. Analysis of the Proposed Compression Method

The considered compression method belongs to the class of numbering compression methods and does not require preliminary statistical tests, which is its advantage. Compared with other numbering compression methods, it is versatile and dynamic, expressed in adaptation to the source of information, reducing the compression time. As a result, the compression ratio increases, and the cost of the equipment implementing the method becomes cheaper. In addition, the method allows for detecting errors, thereby improving the noise immunity of compression. The method also includes the ability to protect information from unauthorized access.

The compression methods proposed above in the introduction, which have a high compression ratio and speed, are primarily specialized. Universal methods, especially those requiring statistical studies, are less fast. At the same time, both methods, during compression, reduce the noise immunity of the compressed information by eliminating redundancy. It is possible to eliminate the shortcomings of existing compression methods by solving the task of the optimal adaptive control of the values of their compression ratios, noise immunity, and compression time. The proposed adaptive method is based on the method developed for the universal compression of binary sequences based on binomial numbers. The universality of the proposed method of adaptive binomial compression is explained by the fact that it compresses binary sequences. Moving from binary sequences to processing symbolic, graphic, and more complex information is relatively easy. In addition, many messages and images themselves are represented as binary sequences. These include television binary images, graphics, and symbols.

However, the binomial compression method compressed all binary sequences without exception, significantly increasing the compression time since many of the binary sequences gave almost zero compression ratio or even negative. In this paper, we propose to solve this task using adaptive selective compression of binary sequences, in which inefficient sequences are not compressed. Accordingly, the task of this paper is to develop a universal binomial information compression method that does not require preliminary statistical tests, adaptively optimally regulating the value of the compression ratio, time, and noise immunity. This paper is research; therefore, it does not provide information on the practical application of results. The reliability of the obtained scientific results was verified experimentally on computer models and in separate works on the theoretical study of the method. It is easy to check it for specific examples in manual mode without additional calculations by the method proposed in the paper.

Practical applications of the proposed adaptive binomial compression method are most effective in mobile communication systems operating under interference. This method also can compress texts, graphics, and images. In this case, the longer the compressed binary block, the more efficient the compression. This method can be especially effective in hardware implementation. In terms of speed, compression ratio, and noise immunity, the method, under certain conditions, can significantly surpass the known compression methods. The peculiarity of the method is its adaptability, which means obtaining maximum efficiency under the conditions of changing probabilities of generated binary sequences. It can give a gain in performance and a loss in the compression ratio and vice versa, giving the most significant overall effect. Therefore, its comparison with conventional compression methods is not correct. To date, adaptive binomial methods in compression systems were first proposed by the authors.

The disadvantage of the method is the complexity of implementation and lack of universality. Changing the probabilities requires changing the adaptive properties of the method, which complicates its structure. Many compression problems do not obey the binomial probability distribution and have forbidden sequences and other restrictions. The inability to take them into account in the method under consideration reduces the degree of compression and the versatility of its application. Additionally, converting binomial numbers to binary numbers and vice versa is a time-consuming procedure that requires a reduction in conversion time.

3. The Lossless Adaptive Data Compression Method Based on the Binomial Number System (BNS)

3.1. Description of the Data Compression Method with BNS

According to the proposed method, compressible binary sequences, by determining the number k of units contained in them, are first converted into equilibrium binomial code of the binary BNS with code length $m < n$. Then, by removing ones to the LSB to the first zero or zeros to the first LSB unit are converted by a binomial numerical (numbering) function into numbers of one of the natural number systems (natural numbers), as a rule, of length $l < n$ (see Figure 2).

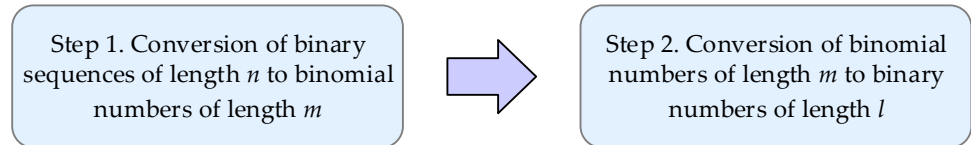


Figure 2. The adaptive binary sequence compression method.

In practice, the binary number system numbers are usually used as such natural numbers. Conversely, restoring a compressed number is transferred from it using a binomial numbering function to a binomial number and then by adding zeros or ones to the least significant digits to the original binary sequence of length n .

3.2. Methods for Converting Binomial Code to Binary

Converting binomial numbers to binary numbers is essential to the described compression method. It is implemented using the binary binomial number system [23,24].

Definition 1. A binary k —BNS is a system that includes a numeric function:

$$F = x_{r-1}C_{n-1}^{k-q_{r-1}} + \dots + x_i C_{n-r+i}^{k-q_i} + \dots + x_1 C_{n-r+1}^{k-q_1} + x_0 C_{n-r}^{k-q_0}, \quad (1)$$

set of binomial numbers $F = x_{r-1} \dots x_i + \dots + x_1 x_0$, binary alphabet $x_i = 0, 1$, and systems of conditions, generating binomial code [23,24]:

$$q \leq r \leq n-1, \quad (2)$$

$$x_0 = 1, q = k, \quad (3)$$

and

$$n-k = r-q, \quad (4)$$

$$x_0 = 0, 0 \leq q \leq k-1, \quad (5)$$

where r —the number of binary digits in binomial numbers (code length); $r \in 1, 2, \dots$,

i —is a position index $i = 0, 1, \dots, r-1$;

q —is a number of 1's in a binomial code;

n, k —is an integer parameter of the BNS $-1, 2, \dots$;

q_i —is a sum of 1's bits of the x_i from $(r-1)$ -th to the $(i+1)$ -th bits,

$$q_i = \sum_{j=i+1}^r x_j \quad (6)$$

$i = 0, 1, \dots, r-1; x_r = 0$.

Step 2 (see Figure 2) is an essential part of this method, converting the binomial number using the binomial numeric function to the number of the natural positional number system. An example of a conversion of binomial codes with parameters $n = 6$ and $k = 4$ to

a binary number is shown in Table 1. The algorithm for obtaining it using the binomial numerical function is demonstrated in [24].

Table 1. Binomial code with $n = 6, k = 4$.

№	Binomial Code	Binomial Numeric Function	№	Binomial Code	Binomial Numeric Function
0	00	$0C_5^4 + 0C_4^4$	8	10111	$1C_5^4 + 0C_4^3 + 1C_3^3 + 1C_2^2 + 1C_1^1$
1	010	$0C_5^4 + 1C_4^4 + 0C_3^3$	9	1100	$1C_5^4 + 1C_4^3 + 0C_3^2 + 0C_2^2$
2	0110	$0C_5^4 + 1C_4^4 + 1C_3^3 + 0C_2^2$	10	11010	$1C_5^4 + 0C_4^3 + 0C_3^2 + 1C_2^2 + 0C_1^1$
3	01110	$0C_5^4 + 1C_4^4 + 1C_3^3 + 1C_2^2 + 0C_1^1$	11	11011	$1C_5^4 + 1C_4^3 + 0C_3^2 + 1C_2^2 + 1C_1^1$
4	01111	$0C_5^4 + 1C_4^4 + 1C_3^3 + 1C_2^2 + 1C_1^1$	12	11100	$1C_5^4 + 1C_4^3 + 1C_3^2 + 0C_2^1 + 0C_1^1$
5	100	$1C_5^4 + 0C_4^3 + 0C_3^3$	13	11101	$1C_5^4 + 1C_4^3 + 1C_3^2 + 0C_2^1 + 1C_1^1$
6	1010	$1C_5^4 + 0C_4^3 + 1C_3^3 + 0C_2^2$	14	1111	$1C_5^4 + 1C_4^3 + 1C_3^2 + 1C_2^1$
7	10110	$1C_5^4 + 0C_4^3 + 1C_3^3 + 1C_2^2 + 0C_1^1$			

Since the binomial coefficient determines the number of binomial numbers, $C_n^k = n! / k!(n-k)!$ it will be equal $C_6^4 = 15$ in the example. A feature of binomial numbers is that they have a variable code length m , varying from k or $n - k$ and up to $n - 1$ bit, and have either k units or $n - k$ zeros in their composition. Since in the example $k > n - k$, the minimum length of binomial numbers in it is $n - k = 2$ bits, and the maximum $n - 1 = 6 - 1 = 5$ bits. Accordingly, in the end, they contain either 0, which will be $(n - k)$ in a row or 1, which will be k in a row.

3.3. Methods for Converting Binary Sequences to Binomial Code and Vice Versa

The binomial numbers of the binary sequences required in step 1 of the method (see Figure 1) are obtained after their preliminary conversion to equilibrium combinations. In a compressible binary sequence of code length n , the number $k = 0, 1, 2, \dots, n$ units are calculated. Thus, the binary sequence is transformed into a k -equilibrium code combination with code length n and k units. Then, by removing the units from the LSB to the first zero or zeros to the first unit, it is converted into a binomial length of $m < n$ (see Figure 3). As a result of this conversion to a binomial number, the binary sequence is compressed by the $n - m$.

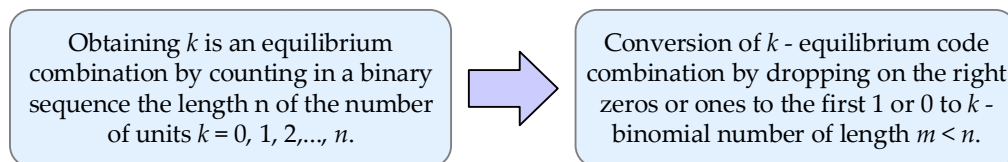


Figure 3. Obtaining binomial numbers.

For an example of the implementation of step 1, the binary sequence 001111 of code length $n = 6$ is shown. After determining the number of units $k = 4$, it is converted into a combination of 0011110 of equilibrium code, as shown in Table 2.

Table 2. Conversion of binary code to binomial code.

№	Binomial Code	Equilibrium Combinations	№	Binomial Code	Equilibrium Combinations
0	00	001111	8	10111	101110
1	010	010111	9	1100	110011
2	0110	011011	10	11010	110101
3	01110	011101	11	11011	110110
4	01111	011110	12	11100	111001
5	100	100111	13	11101	111010
6	1010	101011	14	1111	111100
7	10110	101101			

Deletion in 4 LSB of the code combinations containing units converts it to binomial code 00 with code length $m = n - k = 6 - 4 = 2$. Compression result $n - m = 4$. For an arbitrary binary sequence with code length $n = 1024$ and number of units $k = 224$, the number of zeros $n - k = 1024 - 224 = 800$. If it is 64 ones at its end, then after discarding them, the code length of the binomial code decreases to $m = 1024 - 64 = 960$ bits, which includes 800 zeros and 160 ones. The result of compression is 64 bits. The examples show that the result of the compression of the binary sequence in the first step largely depends on the number of zeros or ones in their continuous sequence located at LSB. Therefore, such compression is effective only in exceptional cases, such as binary television TV lines [23].

In most other cases, another compression step is needed, carried out by the numbering function of the BNS, which converts the binomial c of code length m into the corresponding binary number of the code length l . However, the need to store or transmit the number of units k for its restoration increases the code length l of the compressed sequence per $q = \log_2 n$ bit, thereby reducing the compression effect to $n - (l + q) = n - l - q$. In this case, $(l + q)$ may be greater than n , and there will be no compression. However, for long binary sequences and small values of k and $n > (l + q)$, which determines the compression effect.

In Table 3, the binomial number 00 with code length $m = 2$ is replaced by the binary number 0000 of length $l = 4$. In this case, it is evident that $l > m$. However, the overall compression effect for all binomial numbers in Table 3 will be higher since the total number of bits for all binary numbers of 56 will be less than the number of all bits of binomial numbers 64 by 8 bits. As the length n increases, this difference will only increase.

Table 3. Converting binomial code to binary code.

№	Binomial Code	Binary Code	№	Binomial Code	Binary Code
0	00	0000	8	10111	1000
1	010	0001	9	1100	1001
2	0110	0010	10	11010	1010
3	01110	0011	11	11011	1011
4	01111	0100	12	11100	1100
5	100	0101	13	11101	1101
6	1010	0110	14	1111	1110
7	10110	0111			

Accordingly, the proposed binomial compression method can be represented by the following three key steps:

1. In a compressible binary sequence of code length n , the number of units k is counted. Thus, there is a transition from the binary sequence to the code combination k – equilibrium code. Accordingly, the number of zeros in it is $n - k$.
2. In the k -equilibrium code combination, the LSB zeros to the first unit or the LSB units to the first zero are removed. As a result, the corresponding k -binomial number is obtained, compressing the equilibrium code combination.

3. The obtained k is a binomial number using the binomial numbering function converted into the number of the binary number system. End procedure.

Meanwhile, the binary sequence is restored in reverse order in two following key steps:

1. By the known value k and the binary number l , the transition to the corresponding binomial number m occurs.
2. The transition to the original binary sequence from the binomial code m by adding to LSB ending in 1, $n - k$ zeros or ending in 0, k ones to n .

4. Theoretical Aspects of the Lossless Adaptive Binomial Data Compression Method

Following the above-explained compression method, any binary sequence for which the number of units k is known is converted into a combination of an equilibrium code with $C_n^k = \frac{n!}{k!(n-k)!}$ combinations. Since k can vary from 0 to n , the number of possible equilibrium codes will equal $n + 1$, and each of them will have the same value of the number of units in their combinations $k = 0, 1, 2, \dots, n$. Accordingly, the number of equilibrium combinations is equal to $C_n^0 = n, \dots, C_n^k = \frac{n!}{k!(n-k)!}, \dots, C_n^n = 1$. Together they form a binary code with a total number of combinations of code length n equal to:

$$C_n^0 + C_n^1 + \dots + C_n^k + \dots + C_n^n = 2^n \quad (7)$$

If we assume that compressible sequences are equally probable, then the equilibrium combinations corresponding to them will be the same in length. Then the average amount of information transmitted by one equilibrium combination is

$$I = \log_2(C_n^0 + C_n^1 + \dots + C_n^k + \dots + C_n^n) = \log_2 2^n = n \quad (8)$$

It is impossible to compress information without losing it in such a case. For binary messages to be compressed, different probabilities must characterize them $p_j, 0 \leq p_j \leq 1, j = 1, 2, \dots, 2^n$. Then, the amount of data transmitted by one binary sequence from the information source is determined by the Shannon equation:

$$I(A^*) = - \sum_{j=1}^{2^n} p_j \log_2 p_j \quad (9)$$

The difference $2^n - I$ determines the information redundancy of this source and, accordingly, the maximum possible compression ratio of the information transmitted:

$$K = n / I \quad (10)$$

Obviously, in the case of compression, only a part of the possible sequences from the number 2^n the compression ratio K decreases. Still, this decrease will be insignificant if only sequences with a value of k close to $n/2$ are excluded from the compression procedure. However, the overall compression speed will increase significantly. Thus, it becomes possible to adjust it adaptively by changing the compression ratio. Since k must be determined for all sequences, its presence makes it possible to increase the noise immunity of uncompressed sequences and, accordingly, to detect errors in them. This means that changing the compression ratio makes it possible to optimally adjust the compression rate and the noise immunity of the compressed sequences.

This compression method requires prior knowledge of the statistics of transmitted messages, unlike, for example, the Shannon-Fano or Huffman statistical compression method, since it is automatically determined during the compression process, and with the highest possible accuracy, by analyzing the value of k for each compressible sequence. The same k also determines the compression ratio (10), taken for each binary sequence

$$K = \frac{n}{[\log_2 C_n^k]} \quad (11)$$

Expression $\log_2 C_n^k$ is the amount of information in bits contained in the compressed sequence number with k ones. This information should be supplemented with information $q = \log_2 n$ about the value of k . Then the amount of data after compression equal to

$$I = \log_2 n + [\log_2 C_n^k] \quad (12)$$

Accordingly, the compression ratio of the compressed sequence:

$$K = \frac{n}{([\log_2 n] + [\log_2 C_n^k])} \quad (13)$$

For $k = 0$ and $k = n$ —compression ratio

$$K = \frac{n}{[\log_2 n]}, \quad (14)$$

reaches its maximum.

The most significant value of the compression ratio $k = 1$ is observed. So, for example, with the compression $n = 32, k = 1$ ratio $K = \frac{32}{(5+5)} = 3.2$ and with $n = 1024, k = 1$,

$K = \frac{1024}{(10+10)} = 512$. At the same time, the compression $k = \frac{n}{2}$ coefficient K reaches its

minimum since the binomial coefficient C_n^k takes the highest value in this case. Therefore, in such a case, the compression of the equilibrium combination should be eliminated by transmitting it in its original form, which increases its average transmission compression speed. At the same time, q value information can be used k to improve the noise immunity of the transmitted equilibrium combination. In this case, comparing k with the number of units in the transmitted equilibrium combination determines its correctness.

Table 4 shows examples of compression ratios K for various k . The coefficient k is 1, 2, 16, 32, and 64.

Table 4. The compression ratios K for various k .

	n	32	64	128	256	512	1024
$k = 1$	k	1	1	1	1	1	1
	K	3.20	5.33	9.14	16.00	28.44	51.20
$k = 2$	k	2	2	2	2	2	2
	K	2.29	3.76	6.40	11.13	19.69	35.31
$k = 4$	k	4	4	4	4	4	4
	K	1.58	2.53	4.21	7.23	12.67	22.55
$k = 8$	k	8	8	8	8	8	8
	K	1.12	1.68	2.70	4.52	7.80	13.71
$k = 16$	k	16	16	16	16	16	16
	K	0.93	1.16	1.74	2.81	4.72	8.15

The presence of information about the number of units k , transmitted separately from the compressed binary sequence, allows it to be used as a key, without which it is impossible to restore the original sequence. Therefore, it can be used to protect a compressed sequence. The key k can, in principle, be obtained by iterating its values, changing them from 0 to n . However, this overkill still requires the expenditure of computing power and the corresponding program. Therefore, such a search can be used to protect confidential information that is not of particular importance or is rapidly losing its value.

However, embedding this compression method into an information protection system from unauthorized access can significantly increase its resistance since the general enumeration of the protection system keys becomes more complicated. For each possible key from k of the compression system, the possible keys of the security system are sorted out, and only after that, in the absence of decryption, does the transition to a new possible key from k occur.

In the worst case, the result is manifested in the need to enumerate the product of all possible keys k by the number of possible keys of the protected system. The use of binomial adaptive compression makes it possible to protect particular sequences from opening, and not all, as with conventional compression, which makes it possible to find the optimal ratio of protected and unprotected sequences from the point of view of the compression rate criterion.

5. Conclusions

This paper proposes a lossless adaptive binomial data compression method that is efficient for hardware implementation and based on binomial numbers that allow for compressing binary messages of code length n while controlling their correctness and, if necessary, changing the compression ratio depending on the requirements for it in terms of speed and noise immunity. The presence in the compression method of a separate source of information that generates keys k makes it possible to use it for the transmission or storage of confidential information and, if integrated into the information protection system, to increase its efficiency. The compression coefficients for different lengths of sequences and the number of k units contained in them are studied. For these adaptations, a two-sided restriction on the value of k is introduced. The excess on one or the other side prohibits the compression of binary sequences. The value of k can be either large or small by the symmetry properties of binomial coefficients. Changing these values allows, by changing the compression ratio, to reduce or increase the speed of information compression. Using noise-immune binomial circuits makes the proposed method convenient for hardware implementation and protects information from unauthorized access. The practical application of the proposed adaptive binomial compression method is most effective in mobile communication systems operating under interference conditions and compressing texts, graphics, and images. In this case, the longer the length of the compressed binary block, the more efficient the compression. The method, under certain conditions, can surpass the known compression methods in terms of speed, compression ratio, and noise immunity. The peculiarity of the method is its adaptability, which means the possibility of finding the maximum efficiency under the conditions of changing probabilities of generated binary sequences. It can give a gain in performance and, at the same time loss in the compression ratio and vice versa. The disadvantage of the method is its Bernoulli sources of information and the presence of increased complexity associated with the need to introduce a tracking system for compressible sequences. It allows for the exclusion of inefficient sequences from the compression procedure.

Author Contributions: Conceptualization, S.M. and O.B.; methodology, S.M. and O.B.; software, S.M.; validation, S.M., O.B., and V.B.; formal analysis, S.M.; investigation, S.M.; resources, S.M. and O.B.; data curation, S.M. and O.B.; writing—original draft preparation, S.M., O.B. and V.B.; writing—review and editing, S.M., O.B., T.S., and S.S.; visualization, S.M.; supervision, V.B.; project administration, V.B.; funding acquisition, S.M. and V.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the European Regional Development Fund within Activity 1.1.1.2 “Postdoctoral Research Aid” of the Specific Aid Objective 1.1.1 “To increase the research and innovative capacity of scientific institutions of Latvia and the ability to attract external financing, investing in human resources and infrastructure” of Operational Programme “Growth and Employment” (No. 1.1.1.2/VIAA/3/19/421).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Reddy, B.V.; Reddy, P.B.; Kumar, P.S.; Reddy, A.S. Lossless Compression of Medical Images for Better Diagnosis. In Proceedings of the 2016 IEEE 6th International Conference on Advanced Computing (IACC), Bhimavaram, India, 27–28 February 2016; pp. 404–408.
- Gómez-Brandón, A.; Paramá, J.R.; Villalobos, K.; Illarramendi, A.; Brisaboa, N.R. Lossless compression of industrial time series with direct access. *Comput. Ind.* **2021**, *132*, 103503.
- Gia, T.N.; Qingqing, L.; Queralta, J.P.; Tenhunen, H.; Zou, Z.; Westerlund, T. Lossless Compression Techniques in Edge Computing for Mission-Critical Applications in the IoT. In Proceedings of the 2019 Twelfth International Conference on Mobile Computing and Ubiquitous Network (ICMU), Kathmandu, Nepal, 4–6 November 2019; pp. 1–2.
- Huang, W.; Wang, W.; Xu, H. A Lossless Data Compression Algorithm for Real-time Database. In Proceedings of the 2006 6th World Congress on Intelligent Control and Automation, Dalian, China, 21–23 June 2006; Volume 2, pp. 6645–6648.
- Nivedha, B.; Priyadharshini, M.; Thendral, E.; Deenadayalan, T. Lossless Image Compression in Cloud Computing. In Proceedings of the 2017 International Conference on Technical Advancements in Computers and Communications (ICTACC), Melmaurvathur, India, 10–11 April 2017; pp. 112–115.
- Routray, S.K.; Javali, A.; Sharmila, K.P.; Semunigus, W.; Pappa, M.; Ghosh, A.D. Lossless Compression Techniques for Low Bandwidth Networks. In Proceedings of the 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 3–5 December 2020; pp. 823–828.
- CCSDS. CCSDS 121.0-B-3; Lossless Data Compression. National Aeronautics and Space Administration, Recommendation for Space Data System Standards. CCSDS Secretariat: Washington, DC, USA, 2020.
- Li, M.; Vitányi, P.M. An Introduction to Kolmogorov Complexity and Its Applications; Springer: New York, NY, USA, 2008.
- Willems, F.; Shtarkov, Y.; Tjalkens, T. The context-tree weighting method: Basic properties. *IEEE Trans. Inf. Theory* **1995**, *41*, 653–664.
- Cleary, J.; Witten, I. Data Compression Using Adaptive Coding and Partial String Matching. *IEEE Trans. Commun.* **1984**, *32*, 396–402.
- Storer, J.A.; Szymanski, T.G. Data Compression via Textual Substitution. *J. ACM* **1982**, *29*, 928–951.
- Oord, A.V.; Kalchbrenner, N.; Kavukcuoglu, K. Pixel Recurrent Neural Networks. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; Volume 48, pp. 1747–1756.
- Mentzer, F.; Agustsson, E.; Tschannen, M.; Timofte, R.; Van Gool, L. Practical Full Resolution Learned Lossless Image Compression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern, Salt Lake City, UT, USA, 18–23 June 2018.
- Cover, T.M.; Thomas, J.A. In *Elements of Information Theory*, *Proceedings of the Computer Vision and Pattern Recognition (CVPR)*, Long Beach, SC, USA, 15–20 June 2019; John Wiley & Sons: Hoboken, NJ, USA, 2012, pp. 10621–10630.
- Kadhum Idrees, A.; Kadhum Idrees, S.; Couturier, R.; Ali-Yahiya, T. An Edge-Fog Computing-Enabled Lossless EEG Data Compression with Epileptic Seizure Detection in IoMT Networks. *IEEE Internet Things J.* **2022**, *9*, 13327–13337.
- Idrees, S.K.; Idrees, A.K. New fog computing enabled lossless EEG data compression scheme in IoT networks. *J. Ambient Intell. Human. Comput.* **2022**, *13*, 3257–3270. <https://doi.org/10.1007/s12652-021-03161-5>.
- Duda, J. Asymmetric numeral systems: Entropy coding combining the speed of Huffman coding with a compression rate of arithmetic coding. *arXiv* **2013**. arXiv: 1311.2540.
- Kulyk, I.; Shevchenko, M. Development of information-management systems on the basis of binary binomial number systems. *Inf. Technol. Control Syst.* **2020**, *2*, 78–85.
- Borysenko, O.; Kulyk, I.; Kostel, S.; Skordina, O. Binary Image Compression Based on Binomial Numbers. *Bull. PG Univ. Ploiesti Ser. Math. Inform. Phys. Ploiesti* **2010**, *62*, 1–12.
- Borysenko, O. On the decomposition of Bernoulli's sources of information. *Bull. Sumy State Univ.* **1995**, *1*, 57–59.
- Borysenko, O.; Matsenko, S.; Novhorodtsev, A.; Kobiakov, O.; Spolitis, S.; Bobrovs, V. Estimating the Indivisible Error Detecting Codes Based on an Average Probability Method. *East. Eur. J. Enterp. Technol.* **2020**, *6/9*, 25–33. <https://doi.org/10.15587/1729-4061.2020.218076>.
- Matsenko, S.; Borysenko, O.; Horiachev, O.; Kobiakov, O. In *Noise-Immune Codes Based on Permutations*, *Proceedings of the 9th IEEE International Conference on Dependable Systems, Services and Technologies (DESSERT 2018)*, Kiev, Ukraine, 24–27 May 2018; IEEE: Kyiv, Ukraine, 2018, pp. 609–612. ISBN 978-1-5386-5904-5; ISBN 978-1-5386-5903-8. <https://doi.org/10.1109/dependable.2018.8409204>.
- Borysenko, O. *Introduction to the Theory of Binomial Counting*, Monograph; Sumy ITD University Book: Sumy, Ukraine, 2004; p. 88.
- Borysenko, O.; Matsenko, S.; Bobrovs, V. Binomial Number System. *Appl. Sci.* **2021**, *11*, 1110. ISSN 2076-3417. <https://doi.org/10.3390/app112311110>.