Directions of Two-Symbol Encoding Application in Information and Cyber Security*

Mykola Pratsiovytyi[1,2,†], Volodymyr Astapenya[3,†], Yuliia Zhdanova[3,†], Svitlana Shevchenko[3,*,†] and Iryna Lysenko[2,†]

[1] Institute of Mathematics of NAS of Ukraine, 3 Tereschenkivska str, 01024 Kyiv, Ukraine
[2] Drahomanov Ukrainian State University, 9 Pirogova str., 01601 Kyiv, Ukraine
[3] Borys Grinchenko Kyiv Metropolitan University, 18/2 Bulvarno-Kudriavska str., 04053 Kyiv, Ukraine

Abstract

Reliability and accuracy of information transmission are critically important in today's digital society. At the same time, error-free data transmission should not only guarantee security and efficiency but also optimize the use of resources, such as time, energy, and network bandwidth. To achieve these results, mathematical methods are used, which become the foundation for the development of new algorithms and solutions in the field of information and communication technologies, in particular in the theory of information coding. This article considers the possibility of using two-symbol encoding for efficient data transmission and storage. The paper presents a two-symbol encoding system for real numbers, which is a unique partial case of a system with two bases of different signs. Probabilistic and fractal theory of binary representations is covered. It was determined that this two-symbol information encoding system can be used in cryptography, using numbers recorded in this way as components of a series of pseudo-random numbers, and in steganography for hiding information in video images or other open data streams of both static and dynamic types based on fractal theory. The approaches considered in this study can be used in the training of specialists in the field of information and cyber security.

Keywords

two-symbol number encoding system, G_2-representation of numbers, information protection, information and cybernetic systems, cryptography, steganography

1. Introduction

Encoding is the process of converting information from one form to another that conforms to a specific algorithm, standard, or protocol. Encoding information is an important component of information and cyber security. Encoding information ensures compatibility, efficiency, and security during data transmission, processing, and storage [1]. In other words, as a result of encoding, if there is no interference, the message is presented in a more compact (compressed) form, otherwise, the encoding process detects and corrects errors that arise due to interference. The latter ensures the reliability of the information. A set of message characters receives a unique numerical code that corresponds to a certain encoding standard.

"A code is a system of symbols or signals for transmitting, processing, and storing information" [2, p. 551]. In information transmission systems, codes can be classified according to various characteristics:

- Regarding the construction of a code based on logical and mathematical theories (algebraic, combinatorial, probabilistic, and others).
- Regarding resistance to distortion (interference-resistant, non-interference-resistant).
- Regarding the structure of code combinations (block, tree-like, and others).

* CPITS 2025: Workshop on Cybersecurity Providing in Information and Telecommunication Systems, February 28, 2025, Kyiv, Ukraine

* Corresponding author.
† These authors contributed equally.
prats4444@gmail.com (M. Pratsiovytyi); v.astapenia@kubg.edu.ua (V. Astapenya); s.shevchenko@kubg.edu.ua (S. Shevchenko); y.zhdanova@kubg.edu.ua (Y. Zhdanova); i.m.lysenko@udu.edu.ua (I. Lysenko)
0000-0001-6130-9413 (M. Pratsiovytyi); 0000-0003-0124-216X (V. Astapenya); 0000-0002-9736-8623 (S. Shevchenko); 0000-0002-9277-4972 (Y. Zhdanova); 0009-0000-5299-7787 (I. Lysenko)

- Regarding the length of the code combination (uniform, non-uniform).
- Regarding the code base (binary, ternary, etc.).
- Regarding the method of transmitting code elements by signals (serial, parallel, etc.).
- Regarding the form of information representation (text, numeric, graphic, audio, video, combined).
- Encryption.
- Steganography.

In information theory and programming, there are a variety of encoding methods, many of which are named after prominent scientists who made significant contributions to their development. These codes are used for efficient and reliable data representation, including compression, error protection, and information transmission:

- The Hamming code developed by Richard Hamming to detect and correct errors in data, is widely used in computer memory and communication.
- The Berger code, developed by Jerry Berger to detect errors in arithmetic operations is used in processors and other digital circuits.
- The Huffman code, developed by David Huffman for data compression, provides optimal encoding by using variable-length codewords and is used in many file formats, such as JPEG and MP3.
- The Shannon-Fano code, developed by Claude Shannon and Robert Fano for data compression, is a predecessor of the Huffman code and also uses variable-length codewords.

This list is not exhaustive and there are many other codes named after their authors. Some of them may be less well-known but are also important for the development of information technology. The main goal in the process of encoding information, scientists identify:

- Search for codes that allow for efficient, unhindered transmission of information.
- Search for codes that allow for the reliable transmission of information.

The first direction is studied by the theory of economic coding [3–7], and the second—by the theory of noise-resistant coding [8–10]. Obviously, the solution to these problems is best solved by encoded information, so this problem will be relevant in the future. The current state of coding is characterized by the implementation of artificial intelligence, in particular machine learning. Mathematical methods will play a key role in creating efficient and reliable coding systems that ensure the transmission and storage of information [11].

Today, exact sciences boldly operate with both finite and infinite sets and data arrays. At the same time, the ideas of coordination and coding are effectively used. As a result, powerful families of mathematical objects are described by a small number of basic (reference) objects and relations. Meaningful information about dependencies, and relations (correspondences) takes on a digital form, encrypted by codes: sets, matrices, and sequences of elements of the alphabet, which can be finite and infinite, constant and variable [12].

2. Representation of real numbers through a two-symbol alphabet

In mathematics and its applications today, various representations and images (encodings) of real numbers are used, which use finite and infinite, constant and variable alphabets, i.e. a real number has different forms of existence. One of the simplest is the representation and image in the s-base (binary, ternary, decimal, etc.) number system.

Two-symbol information encoding systems traditionally use the alphabet $A=\{0,1\}$. Two-symbol systems deserve special attention, in particular, due to the minimal alphabet. Next, we will focus on encodings of real numbers.

The coding of real numbers of the interval $[0;1]$ using the alphabet A is called the correspondence between the sets $[0;1]$ and $L=A\times A\times A\times\ldots$, in which each number $x\in[0;1]$ corresponds to at least one element of the set L and each element of the set L is the image of at least one number of the interval $[0;1]$. The sequence itself $(a_n)=(a_1,a_2,\ldots,a_n)\in L$, which corresponds to the number x, is called its representation (or code), and a a_n is the n^{th} digit (or symbol) of its representation.

It seems that the coding system has zero redundancy since each number can have no more than two images, and what is important is the majority of numbers in one representation.

The binary representation of numbers is the simplest example of two-symbol continuous encodings with zero redundancy [13].

2.1. Existence of a G-representation of a natural number

Definition 1. *If for a natural number a there exists a set $(a_1,a_2,\ldots,a_n)$ of zeros and ones such that*

$$a=2^n+\sum_{k=1}^{n}\left[(-1)^{1+\sigma_k}a_k 2^{n-k}\right]\equiv(1a_1\ldots a_n)_G, \tag{1}$$

where $\sigma_k=a_1+\ldots+a_{k-1}$, then we will say that the number a has a G-representation, which is symbolically written $a=(1a_1\ldots a_n)_G$ and is $(n+1)$-digit.

For example, each of the numbers within ten has a G-representation, and there are exactly two of them, which is easy to verify.

$1=2^0=(1)_G=2^1-2^0=(11)_G$,

$2=2^1=(10)_G=2^2-2^1=(110)_G$,

$3=2^2-2^1+2^0=(111)_G=2^2-2^0=(101)_G$,

$4=2^2=(100)_G=2^3-2^2=(1100)_G$,

$5=2^3-2^2+1=(1101)_G=2^3-2^2+2-1=(1111)_G$,

$6=2^3-2^2+2=(1110)_G=2^3-2+1=(1011)_G$,

$7=2^3-1=(1001)_G=2^3-2+1=(1011)_G$,

$8=2^3=(1000)_G=2^4-2^3=(11000)_G$,

$9=2^4-2^3+1=(11001)_G=2^4-2^3+2-1=(11011)_G$,

$10=2^4-2^3+2=(110110)_G=2^4-2^3+2-1=(11110)_G$.

Note. As we can see, the G-representation of the same natural number can have a different number of digits. For example, $8=(1000)_G=(11000)_G$, and in general

$$a=2^n=\Big(1\underbrace{0\ldots0}_{n}\Big)_G=\Big(11\underbrace{0\ldots0}_{n}\Big)_G. \tag{2}$$

Then as

$$a=2^n+1=2^{n+1}-2^n+1=\Big(11\underbrace{0\ldots0}_{n-1}1\Big)_G=\Big(11\underbrace{0\ldots0}_{n-2}11\Big)_G. \tag{3}$$

Theorem 1. *Every natural number has exactly two G*-representation, *i.e. for any natural number a there exists a set of zeros and ones $(a_1,a_2,\ldots,a_n)$ such that (1)* where $\sigma_1=0$, $\sigma_k=a_1+\ldots+a_{k-1}$, *and there are exactly two such decompositions.*

Proof. It is obvious that for any natural number a there exists a natural number n, such that $2^{n-1}<a\leq 2^n$. Let us use the method of mathematical induction on the number n. For $n=1$ we have $1<a\leq 2$ and the statement is obvious for the numbers 1 and 2 (see the example above).

Let us assume the truth of the statement for $n=k$, i.e. $a\in(a^{k-1};2^k]$ has the decomposition

$$a=2^k+\sum_{i=1}^{k}\left[(-1)^{1+\sigma_i}a_i 2^{k-i}\right]\equiv(1a_1\ldots a_k)_G, \tag{4}$$

where $\sigma_i=a_1+\ldots+a_{i-1}$, and there are exactly two of them.

Consider $n=k+1$, i.e. the number $a\in(2^k;2^{k+1}]$. If $a=2^{k+1}$, then $a=\left(1\underbrace{0\ldots0}_{k+1}\right)_G=\left(11\underbrace{0\ldots0}_{k+1}\right)_G$.

Let $2^k<a<2^{k+1}$, then $0<a-2^k\equiv u<2^k$. By assumption, the number u has G-representation

$$u=(1c_1c_2\ldots c_m)_G, \tag{5}$$

where $m<k$.

Then $a=2^k+u=2^{k+1}-2^k+u=\left(11\underbrace{0\ldots0}_{k-m}c_1\ldots c_m\right)_G$. Since the number $u\equiv a-2^k$ satisfies the inequality $0<u<2^k$, then by assumption, it has exactly two G-representation. Therefore, exactly two G-representation has the number a. According to the principle of mathematical induction, the statement is proved for any natural number a.

Definition 2. *The canonical G-representation of a natural number is the G-representation, which has the minimum number of digits and the maximum number of zeros at the same time.*

For example, the canonical G-representation of numbers:

$$16=(10000)_G=(110000)_G, a=2^n-1=\left(1\underbrace{0\ldots0}_{n-1}1\right)_G=\left(1\underbrace{0\ldots0}_{n-2}11\right)_G \tag{6}$$

is the first of the G-representations.

2.2. G-representation of the fractional part of a real number

Theorem 2 [14]. *For any number $x\in[0;\frac{1}{2}]$ there exists a sequence of zeros and ones (α_n) such that*

$$a=\frac{\alpha_1}{2}+\sum_{k=2}^{\infty}\frac{\alpha_k(-1)^{\sigma_k}}{2^k}=\frac{\alpha_1}{2}+\sum_{k=2}^{\infty}\frac{\alpha_k}{2^{k-\sigma_k}(-2)^{\sigma_k}}\equiv\Delta^G_{\alpha_1\alpha_2\ldots.\alpha_n\ldots}, \tag{6}$$

where $\sigma_k=a_1+\ldots+a_{k-1}$.

Corollary. *For any number $x\in[0;1]$ there exists a sequence of zeros and ones (α_n) such that*

$$x=\frac{1}{2}\alpha_0+\frac{\alpha_1}{2}+\sum_{k=2}^{\infty}\frac{\alpha_k(-1)^{\sigma_k}}{2^k}\equiv\Delta_{\alpha_0\alpha_1\ldots\alpha_n\ldots}. \tag{7}$$

The symbolic notation $\Delta^G_{\alpha_1\alpha_2\ldots.\alpha_n\ldots}$ is called the G-representations of the number $x\in[0;\frac{1}{2}]$ and its expansion into the series (1).

2.3. G_2-representation of numbers in the interval $[0;g_0]$

G-representations is a special case of a more general two-symbol mapping with two bases of different signs. Let us recall its definition. Let two bases be fixed $g_0 \in (0;\frac{1}{2}]$, $g_1 \equiv g_0 - 1$ and numbers: $\delta_0 = 0, \delta_1 = g_0$.

Theorem 3 [15, 16]. *For any number $x \in [0;g_0]$ there is a sequence of zeros and ones (α_n) such that*

$$x = \delta_{\alpha_1} + \sum_{k=2}^{\infty} (\delta_{\alpha_k} \prod_{j=1}^{k-1} g_{\alpha_j}) \equiv \Delta^{G_2}_{\alpha_1\alpha_2\ldots\alpha_k\ldots}, \delta_{\alpha_k} = \alpha_k g_{1-\alpha_k}. \tag{8}$$

Corollary 1. *If $g_0 = \frac{1}{2}$, then $\delta_{\alpha_k} = \frac{\alpha_k}{2^k}$,*

$$\prod_{j=1}^{k-1} g_{\alpha_j} = \frac{(-1)^{\alpha_1 + \ldots + \alpha_{k-1}}}{2^{k-1}} \tag{9}$$

and the series (8) takes the form (6).

The symbolic notation $\Delta^{G_2}_{\alpha_1\alpha_2\ldots\alpha_k\ldots}$ of a number x and its decomposition into an alternating series (8) is called the G_2-representations, and α_k is its k^{th} digit.

Lemma 1. *Lebesgue measure of a set*

$$C \equiv C[G_2; \overline{s_1 \ldots s_m}] = \{x : x = \Delta^{G_2}_{\alpha_1\alpha_2\ldots\alpha_n\ldots}, \overline{\alpha_k \ldots \alpha_{k+m-1}} \neq \overline{s_1 \ldots s_m} \,\forall k \in N\} \tag{9}$$

numbers from the interval $[0;g_0]$, G_2-representations of which does not use the set of digits $s_1 s_2 \ldots s_m$ as consecutive digits of the G_2-representations of the number, is equal to zero.

Proof. Consider sets

$$E = \left\{x : x = \Delta^{G_2}_{\alpha_1\alpha_2\ldots\alpha_n\ldots}, \overline{\alpha_{km+1}\alpha_{km+2}\ldots\alpha_{km+m}} \neq \overline{s_1 \ldots s_m} \,\forall k \in N\right\}.$$

It is obvious that $C \subset E$. Let us prove that the Lebesgue measure $\lambda(E) = 0$. From which it follows that $\lambda(C) = 0$. To do this, let us denote by $F_0 \equiv [0;g_0]$, F_k the union of all cylinders of rank km, among the interior points of which there are points of the set E, and $\overline{F}_{k+1} \equiv F_k \setminus F_{k+1}$. Then $F_k = \overline{F}_{k+1} \cup F_{k+1}$, $\lambda(F_{k+1}) = \lambda(F_k) - \lambda(\overline{F}_{k+1})$, $E \supset F_k \supset F_{k+1}$ $\forall k \in N$ and $\lambda(E) = \lim\limits_{k\to\infty} \lambda(F_k)$.

Let us express

$$\lambda(F_k) = \frac{g_0 \lambda(F_k)}{\lambda(F_{k-1})} \cdot \frac{\lambda(F_{k-1})}{\lambda(F_{k-2})} \cdot \ldots \cdot \frac{\lambda(F_1)}{\lambda(F_0)} = \prod_{i=1}^{k} \frac{\lambda(F_i)}{\lambda(F_{i-1})} = \prod_{i=1}^{k} \frac{\lambda(F_i) - \lambda(\overline{F}_i)}{\lambda(F_{i-1})} = \\ = \prod_{i=1}^{k} \left(1 - \frac{\lambda(\overline{F}_i)}{\lambda(F_{i-1})}\right). \tag{10}$$

Whence

$$\lambda(E) = \lim_{k\to\infty} \lambda(F_k) = g_0 \prod_{k=1}^{\infty} \frac{\lambda(F_k)}{\lambda(F_{k-1})} = g_0 \prod_{k=1}^{\infty} \left(1 - \frac{\lambda(\overline{F}_k)}{\lambda(F_{k-1})}\right). \tag{11}$$

Since

$$\left|\Delta^{G_2}_{a_1\ldots a_k i}\right| = \left|g_i\right|\left|\Delta^{G_2}_{a_1\ldots a_k}\right| \tag{12}$$

then

$$c_1 \le \frac{\left|\Delta^{G_2}_{a_1\ldots a_k i}\right|}{\left|\Delta^{G_2}_{a_1\ldots a_k}\right|} \le c_2, \tag{13}$$

where $c_1 = min\{g_0; -g_1\}$, $c_2 = max\{g_0; -g_1\}$, and therefore

$$0 < c_1^m \le \frac{\left|\Delta^{G_2}_{\alpha_1\ldots\alpha_{km} s_1\ldots s_m}\right|}{\left|\Delta^{G_2}_{\alpha_1\ldots\alpha_{km}}\right|} = \prod_{i=1}^{m}\left|g_{s_i}\right| \le c_2^m < 1. \tag{14}$$

Then the ratio $\frac{\lambda(\overline{F}_k)}{\lambda(F_{k-1})}$ is separate from 0, and therefore the difference $1 - \frac{\lambda(\overline{F}_k)}{\lambda(F_{k-1})}$ is separate from 1, i.e. the necessary condition for the convergence of the infinite product is not met, and therefore it converges to zero.

Corollary 2. *The set C is a null set of Cantor type with a self-similar structure.*

Theorem 4. *Almost every number from the interval $[0; g_0]$ in its G_2-representations use each of the sets of digits of the alphabet as consecutive digits an infinite number of times.*

Proof. Let $(s_1, \ldots, s_m)$ be an arbitrary ordered set of zeros and ones, H be the set of all numbers $[0; g_0]$, in the G_2-representation of which the set of digits $s_1 s_2 \ldots s_m$ appears an infinite number of times as consecutive digits, D be the set of all numbers that use this set as consecutive digits of the G_2-representation only a finite number of times.

Let us prove that H is a set of full Lebesgue measure, that is, that $\lambda(H) = g_0$. To do this, it is enough to prove that $\lambda(\overline{H}) = 0$, where $\overline{H} = [0; g_0] \setminus H$. It is obvious that

$$\overline{H} = \bigcup_{n=1}^{\infty} D_n, \tag{15}$$

where $D_n = \left\{x : x = \Delta^{G_2}_{\alpha_1\ldots\alpha_n\ldots}, \alpha_{k+1}\ldots\alpha_{k+m} \ne s_1\ldots s_m \,\forall\, k \ge n\right\}$.

Let us calculate $\lambda(D_n)$. Since

$$D_n = \bigcup_{\alpha_1\in A} \ldots \bigcup_{\alpha_n \in A} [\Delta^{G_2}_{\alpha_1\ldots\alpha_n} \cap D], \Delta^{G_2}_{\alpha_1\ldots\alpha_n} \cap D = \Delta^{G_2}_{\alpha_1\ldots\alpha_n} \cap H, \tag{16}$$

then according to the previous lemma 1 $\lambda\left(\Delta^{G_2}_{\alpha_1\ldots\alpha_n} \cap D\right) = 0$, and therefore, $\lambda(D_n) = 0$. Therefore $\lambda(\overline{H}) = 0$, since

$$\lambda(\overline{H}) \le \sum_{n=1}^{\infty} \lambda(D_n) = 0. \tag{17}$$

The theorem is proved.

2.4. Probabilistic theory of binary representation

The probability theory of real numbers in their binary representation deals with solving problems related to probability distributions on sets of numbers defined by the properties of the binary representation of their elements [17].

Theorem 5. *If the random variable $\xi=\Delta^{G_2}_{\xi_1\xi_2\ldots}$ has a uniform distribution on the interval $[0;g_0]$, then the digit (ξ_n) of its G_2-representation are independent and have distributions $P(\xi_n=0)=g_0$, $P(\xi_n=1)=-g_1$.*

Theorem 6. *If the distribution function $F_\eta(x)$ of a random variable $\eta=\Delta^{G_2}_{\eta_1\ldots\eta_k\ldots}$ with independent binary digits η_k has a positive derivative at the points $2^{-1},2^{-2},\ldots$, then η has an exponential distribution on a $[0;1]$ with density*

$$f(x)=\frac{\beta\left(e^{\beta x}\right)}{e^{\beta}-1}.-\propto<\beta<\propto, \tag{18}$$

where the distributions *of digits η_k are given by the formulas*

$$\begin{aligned} P(\eta_k=0)&=p_{0k}=\frac{1}{1+e^{\frac{\beta}{2^k}}},\\ P(\eta_k=1)&=p_{1k}=\frac{e^{\frac{\beta}{2^k}}}{1+e^{\frac{\beta}{2^k}}}. \end{aligned} \tag{19}$$

2.5. Fractal theory of binary representation

The essence of the fractal theory of numbers in a given system of their coding consists in studying fractal properties of a set of numbers, determined by conditions (restrictions) on their representation (use of digits). These conditions can be formulated in terms of prohibitions on the use of digits and their combinations, in terms of frequencies of digits or combinations of digits in the representation of a number, etc.

The methods of fractal theory are based on the ideas of self-similarity, in particular self-similarity and self-affinity, tools of the theory of metric dimensions, and measures of fractional orders. The Q_2-representation of numbers is effectively used in fractal theory [17].

Let the G_2-representation of numbers in the interval $[0;g_0]$ be given, W_G is the set of all G_2-cylinders, $[0;g_0]\supset E$ is a fixed set, $1\geq\alpha$ is a positive number, $\varepsilon>0$. Let $l^\alpha_\varepsilon\equiv\inf\limits_{|\omega_i|\leq\varepsilon}\left\{\sum\limits_i|\omega_i|^\alpha : E\subset\underset{i}{U}\,\omega_i,\omega_i\in W_G\right\}$ be given, where $|\omega_i|$ is the length of the cylinder ω_i, and the lower bound is taken as all possible coverings of the set E by G_2-cylinders whose lengths do not exceed ε.

It is obvious that the numbers $l^\alpha_\varepsilon(E)$ and $m^\alpha_\varepsilon(H)$, generally speaking, are different ($l^\alpha_\varepsilon(E)\geq m^\alpha_\varepsilon(H)$, since the class of all possible covers of the set E by intervals includes the class of covers by G_2-cylinders). Let us prove that despite this, the functions $H^\alpha(E)$ and $L^\alpha(E)=\lim\limits_{\varepsilon\to\propto}l^\alpha_\varepsilon(E)$ of the variable α take the values 0 and ∞ simultaneously. And this is a reason to believe that when determining the Hausdorff-Bezikovitch dimension of an arbitrary set $E\subset[0;g_0]$ one can limit oneself to covers of the set by G_2-cylinders, which follows from the statement.

Theorem 7. For arbitrary $E\subset[0;g_0]$, $0<\alpha\leq1$, $\varepsilon>0$, the double inequality holds

$$m^\alpha_\varepsilon(E)\leq l^\alpha_\varepsilon(E)\leq2\left(2^m+1\right)\cdot m^\alpha_\varepsilon(E), \tag{20}$$

m is the smallest natural number satisfying the inequality $g_0^m\leq1-g_0$.

3. Two-symbol system in some variants of information encoding

The proposed two-symbol system has great prospects in mathematics, but its possibilities in the field of coding were not considered. Let us try to highlight those areas of its application in information coding for further research:

1. Representation in the two-symbol system can be used as a simplified (primitive) variant of cryptographic information closure, using the numbers recorded in this way as components of a series of pseudorandom numbers.
2. The two-symbol system involving Q_2-representation has a chance to be used in the interests of steganographic closure (hiding) of information in video images or other open data streams of both static and dynamic types based on fractal theory. In this case, the pixels carrying information (part of the message code) are placed in the plane of the picture randomly.

It should be noted that this representation can be used as a primary code. But for such codes, one of the important properties is conciseness—the minimum number of bits of representation of a particular state of an object or process values. In this sense, the two-symbol system is inferior to the classical positional binary code, because it uses additional digits with a negative sign in the representation of a number, which introduces redundancy. This increases presentation time, and memory load, and increases channel transmission time. However, when using a two-symbol system, the Hamming distance (a measure of dissimilarity of code combinations—the number of bits in which the symbols of one of the compared combinations differ from the other) between the combinations may increase slightly (on average). It is a consequence of the introduction or presence of redundancy. Combinations appear that are inherent in this particular representation and those that cannot exist in this representation. For example, the number 5 can be written:

$$5=2^3-2^2+1=(1101)_G=2^3-2^2+2-1=(1111)_G, \tag{21}$$

while the classic 1001 cannot be used and is a prohibited combination within the proposed two-symbol system. From this position, it can be suggested that there is a chance of using a two-symbol system as a simpler, not very efficient, but conditionally noise-resistant code. Additional research is needed to assess this possibility. First of all, this concerns the quantitative ratio between the allowed and forbidden combinations that arise in a two-symbol system. This is also related to the value of the possible code distance (the minimum Hamming distance between any two allowed code combinations). The greater the number of forbidden combinations, the potentially greater the code distance between the permitted combinations (d_k), and the greater the error rate (number of distorted symbols in the accepted code combination) that the code allows to detect (y_{detect}) or correct ($y_{correct}$):

for detection

$$d_{k\,detect} \geq y_{detect}+1; \tag{22}$$

for error correction

$$d_{k\,correct} \geq 2\,y_{correct}+1. \tag{23}$$

The number of allowed values with the required code distance is important, and it increases with increasing code redundancy.

The given example of the representation of the number 5 also indicates the possibility of using the specified representation as a variant in the binary-decimal code.

Conclusions

The information society requires effective information protection. Coding plays a key role in ensuring the confidentiality, integrity, and availability of information, and the basis of coding rules are mathematical methods and models. The authors of the article proposed directions for applying mathematical theory related to two-symbol number coding systems in the field of information coding.

We see directions for further exploration in a more detailed study of the possibilities of implementing this system, using the G-representation not only of a natural number, but also of the fractional part of a real number.

Declaration on Generative AI

While preparing this work, the authors used the AI programs Grammarly Pro to correct text grammar and Strike Plagiarism to search for possible plagiarism. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

References

[1] R. Bose, Information theory, coding and crytography: 3/e, 2016.

[2] V. T. Busel, Large explanatory dictionary of the modern ukrainian language, Perun, 2005.

[3] E. Ukkonen, Approximate string-matching with q-grams and maximal matches, Theor. Comput. Sci. 92(1) (1992) 191–211. doi:10.1016/0304-3975(92)90143-4

[4] S. Kumar, T. Utsab, A. Raychoudhury, Image compression using approximate matching and run length, Int. J. Adv. Comput. Sci. Appl. 2(6) (2011). doi:10.14569/ijacsa.2011.020617

[5] A. Gupta, A. Bansal, V. Khanduja, Modern lossless compression techniques: Review, comparison and analysis, in: 2nd International Conference on Electrical, Computer and Communication Technologies, ICECCT, 2017, 1–8. doi:10.1109/icecct.2017.8117850

[6] P. Venkatram, A new lossless data compression algorithm exploiting positional redundancy, arXiv, 2021. doi:10.48550/arXiv.2107.13801

[7] S. Yamagiwa, Stream-based lossless data compression, in: Sublinear Computation Paradigm, 2021, 391–410. doi:10.1007/978-981-16-4095-7_16

[8] A. Serkov, et al., Noise-like signals in wireless information transmission systems, Adv. Inf. Syst. 1(2) (2017) 33–38. doi:10.20998/2522-9052.2017.2.06

[9] A. Pieshkin, et al., The comparative assessment of corrective parameters for antinoise convolutional and block codes, in: International Conference on Information and Telecommunication Technologies and Radio Electronics, UkrMiCo, 2018, 1–4. doi:10.1109/ukrmico43733.2018.9047530

[10] B. Zhurakovskyi, et al., Enhancing information transmission security with stochastic codes, in: Classic, Quantum, and Post-Quantum Cryptography, CQPC, vol. 3829, 2024, 62–69.

[11] R. Chernenko, et al., Encryption method for systems with limited computing resources, in: Cybersecurity Providing in Information and Telecommunication Systems, vol. 3288 (2022) 142–148.

[12] M. Pratsiovytyi, Two-symbol image systems (coding) of real numbers, in: Students' Physical and Mathematical Sketches, vol. 9, 2010, 6–26.

[13] M. Pratsiovytyi, Geometry of real numbers in their encodings by means of an infinite alphabet as the basis of topological, metric, fractal and probabilistic theories, Sci. J. Natl. Pedag. Dragomanov Univ. 1(14) (2013) 189–216.

[14] M. Pratsiovytyi, et al., G-representation of real numbers and some of its applications, Nonlinear Oscil. 25(4) (2022) 377–387.

[15] I. Lysenko, Y. Maslova, M. Pratsiovytyi, The binary number system with different bases and special functions associated with it, in: Proceedings of the Institute of Mathematics of the NAS of Ukraine, vol. 16(2) (2019) 50–62.
[16] M. Pratsiovytyi, I. Lysenko, Y. Maslova, Geometry of number series: A series as a model of a real number in a new two-symbol number coding system, in: Proceedings of the Institute of Mathematics of the NAS of Ukraine, vol. 15(1), 2018, 132–146.
[17] M. Pratsiovytyi, Two-symbol encoding systems of real numbers and their application, Kyiv, Naukova Dumka, 2022.