МЕТОД ОБЕСПЕЧЕНИЯ СВЯЗНОСТИ МАССИВОВ ЦИФРОВЫХ ДАННЫХ ПОСРЕДСТВОМ ЧЛБ (N, K)-КОДА

Терентьев А.И.

Московский государственный технический университет гражданской авиации (МГТУ ГА),
Россия, г. Москва, Кронштадтский бульвар, д. 20
terentyev-doc@yandex.ru

Аннотация: Предложен отличный от технологии блокчейн метод, позволяющий обеспечить целостность массива цифровых данных и неизменность хронологического порядка его элементов. Метод основан на кодировании элементов массива цифровых данных числовым линейным блоковым корректирующим кодом (ЧЛБ-кодом) над кольцом конечных десятичных дробей.

Ключевые слова: технологии связных данных, технология блокчейн, числовой линейный блоковый корректирующий код (ЧЛБ-код).

Введение

Традиционно научным сообществом уделяется особое внимание обсуждению новых идеи, новации и проблем развития современных крупномасштабных систем. В частности, в области финансовых и экономических систем высокую популярность получила технология блокчейн (англ. blockchain), импульсом к распространению которой послужило создание и развитие сети криптовалюты Биткоин (англ. Bitcoin). Согласно основополагающему документу этой сети [1] осуществление связи блоков данных, входящих в формируемую и постоянно удлиняющуюся рекуррентную последовательность (цепь), обеспечивается посредством специальной криптографической процедуры, основанной на свойствах функции хэширования и использовании древовидного хешировании блоков данных. Такой способ создания связного множества цифровых данных, обеспечивающего сохранение предыстории и хронологического порядка формирования (добавления) его элементов, устойчиво отождествляется в настоящее время с технологиями блокчейн и распределенного реестра, что неформально закреплено в современном определении термина «блокчейн». Однако, несмотря на широкую известность и высокие ожидания в отношении перспектив применения технологии блокчейн [2, 3], в том числе для обработки больших массивов данных, следует отметить, что темпы ее внедрения оставляют желать лучшего. Это обусловлено как практическими, так теоретическими проблемами. В частности, в статье [4] авторами отмечается, что имеется много нерешенных вопросов в области терминологии, унификации и реализации технологии блокчейн. Следует добавить, что на практике часто происходит смешение или отождествление технологии блокчейн с другими сопутствующими методиками, процессами и технологиями, направленными на обеспечение целостности, доступности и достоверности различных массивов данных, которые в терминах дискретной математики [5] можно представить как множества M цифровых элементов (чисел). Такие множества могут быть заданы посредством перечисления всех своих элементов: $M = \{m_1, m_2, ..., m_k\}$, или посредством указания порождающей процедуры (алгоритма) или свойства P, на основании которого элементы принадлежат множеству: $M = \{m|P(m)\}$. При этом элементами множества M могут являться множества, а также гибридные и иные сложно структурированные объекты, включающие, в том числе, информационную и служебную составляющие. Множество M будем считать связным, если на нем установлено бинарное или иное отношение R, обладающее свойством связности. Соответственно данные, которые представляются элементами такого множества, будут являться связанными между собой. На связном множестве M может быть установлено отношение порядка, что единственным образом обеспечит нумерацию его элементов. Например, если множество M рассматривать как рекуррентную последовательность, то для любой пары ее элементов будет выполнятся отношение строго порядка m_iRm_j, а для каждой пары соседних элементов (m_{i-1}, m_i) или (m_i, m_{i+1}) будет выполняться отношение доминирования $m_{i-1}Rm_i$ и m_iRm_{i+1} [6].

Проведенные автором исследования показывают, что кроме технологии блокчейн существуют иные способы обеспечения связности и упорядоченности множества цифровых данных, которые представляют дополнительные возможности по обеспечению стойкости такого множества к модификации, в том числе изменению хронологического порядка и целостности его элементов. В частности, представляется перспективным использовать в качестве порождающей процедуры предложенный далее метод, основанный на кодировании блоков цифровых данных посредством числового корректирующего кода. Поскольку средством обработки цифровых данных является

электронное техническое устройство (электронная вычислительная машина), то элементы корректирующего числового кода должны иметь вид конечных десятичных дробей. В связи с этим в качестве конкретных кодов для задания порождающих процедур упорядоченных связных множеств цифровых данных (чисел) предлагается использовать равномерные разделимые систематические числовые линейные блоковые корректирующие коды (ЧЛБ-коды) над кольцом D конечных десятичных дробей [7, 8].

1 Выбор, определение и способы задания ЧЛБ (n, k)-кода

Код принято называть числовым, если его комбинациями являются упорядоченные последовательности чисел ограниченной длины. Числовой код называется корректирующим (исправляющим ошибки), если он позволяет обнаруживать и исправлять искаженные по тем или иным причинам числа в его кодовых комбинациях. Если код способен только обнаруживать факт искажения какого-либо из чисел его кодовой комбинации, без определения позиции такого числа, то такой код можно считать только помехоустойчивым.

Для обеспечения связности массивов данных, которые можно представить как упорядоченные последовательности цифровых элементов (чисел ограниченной длины), предлагается использовать равномерный разделимый систематический ЧЛБ-код над кольцом D конечных десятичных дробей, которым называется множество последовательностей длины n, элементами которых являются действительные числа, имеющие вид конечных десятичных дробей, причем проверочные числа являются заданными линейными функциями от информационных чисел [7, 8]. Выбор такого кода также обусловлен тем, что он гармонично сочетается с принципами блокового представления информации и блокового построения цепи блокчейн.

Последовательность

$$u = (a1, ..., ak, b1, ..., br), \quad (1)$$

где $a_1, ..., a_k$ - последовательность информационных чисел,

k - количество информационных чисел,

$b_1, ..., b_r$ - последовательность проверочных чисел,

r - количество проверочных чисел, причем $n=k+r$ (n – длина кодовой комбинации),

называется кодовой комбинацией ЧЛБ-кода. Все кодовые комбинации u образуют множество U_k, которое и есть равномерный разделимый систематический ЧЛБ-код над кольцом D. Кодовая комбинация u является n-мерным ($n = k + r$) числовым вектором арифметического пространства U_n над кольцом D, а множество U_k всех числовых векторов длины $n > k$, является k-мерным арифметическим подпространством n-мерного арифметического пространства U_n над кольцом D. Базис подпространства U_k состоит из k базисных числовых векторов длины n. Запишем все базисные векторы в виде трапецеидальной вещественной матрицы размерности $k \times n$

$$G = \begin{pmatrix} 1 & 0 & \dots 0 & c_{11} & \dots & c_{1r} \\ 0 & 1 & \dots 0 & c_{21} & \dots & c_{2r} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots 1 & c_{k1} & \dots & c_{kr} \end{pmatrix} \quad (2)$$

Ранг такой матрицы равен k. Сокращенная форма записи $G = [\, E_k \,\vdots\, C_{k\times r} \,]$, где вертикальная штриховая линия условно разделяет матрицу G на единичную матрицу E_k размерности $k \times k$ и матрицу $C_{k\times r}$ размерности $k \times r$. Матрица $C_{k\times r}$ называется матрицей коэффициентов c_{ij}, т.к. ее столбцы определяют правила формирования проверочных чисел кодовой комбинации u и являются числовыми коэффициентами при соответствующих слагаемых (информационных числах). Следовательно, ЧЛБ-коды можно задавать и правилами формирования проверочных чисел

$$b_j = \sum_{i=1}^{k} c_{ij} a_i \quad (3)$$

Или

$$\begin{aligned} b1 &= c11\ a1 + c21\ a2 + ... + ck1\ ak \\ b2 &= c12\ a1 + c22\ a2 + ... + ck2\ ak \\ br &= c1r\ a1 + c2r\ a2 + ... + ckr\ ak \end{aligned} \quad (4)$$

где $c_{11},...,c_{kr}$ – элементы матрицы C коэффициентов c_{ij};

$i=1,2,...,k$ – номер строки матрицы G;

$j=1,2,...,r$ - номер столбца матрицы $C_{k\times r}$.

Очевидно, что любой вектор k-мерного векторного пространства U_k разлагается по векторам базиса этого пространства и притом единственным образом. Базис порождает линейное пространство или ЧЛБ-код. В связи с этим матрицу G вида (2) называют порождающей матрицей ЧЛБ-кода.

Вид элементов порождающей матрицы G существенно влияет на свойства и возможности ЧЛБ-кодов. Классификация, приведенная в [8], учитывает расширение множества значений, из которого выбираются возможные элементы матрицы G. В частности, если элементами матрицы G в одном случае могут быть числа из множества $M1$, а в другом случае числа из множества $M2$, причем, если $M1 \subset M2$, то можно утверждать, что все свойства, которыми обладает код $G(M1)$, присущи и коду $G(M2)$, но не наоборот. Предлагается классифицировать ЧЛБ-коды в порядке возрастания кодового уровня:

- *G2* ЧЛБ - код, порождающая матрица G которого бинарная (т.е. ее элементы могут принимать только значения "0" или "1");
- *GZ* ЧЛБ - код, элементы порождающей матрицы G которого принадлежат множеству целых чисел;
- *GD* ЧЛБ - код, элементы порождающей матрицы G которого принадлежат множеству конечных десятичных дробей;
- *GQ* ЧЛБ - код, элементы порождающей матрицы G которого принадлежат множеству рациональных чисел;
- *GR* ЧЛБ - код, элементы порождающей матрицы G которого принадлежат множеству действительных чисел.

Вид порождающей матрицы существенно влияет на параметры кода. Так, например, *G2* ЧЛБ-коды не могут являться кодами с максимально достижимым кодовым расстоянием (МДР-кодами).

Корректирующая способность ЧЛБ-кода определяется его минимальным кодовым расстоянием d. Код позволяет исправлять ошибки в v и менее позициях если

$$d \geq 2v+1. \qquad (5)$$

2 Метод обеспечения связности линейного множества цифровых данных

Связность линейного упорядоченного множества цифровых данных обеспечивается посредством вычисления для выбранных информационных элементов (блоков) последовательности проверочных элементов и последующего их использования при проверке целостности и хронологического порядка соответствующих информационных элементов (блоков).

Информационными элементами кодовых комбинаций равномерного разделимого систематического ЧЛБ (n, k)-кода (также как в технологии блокчейн) должны являться блоки цифровых данных заданной длинны, а не составляющие их отдельные наборы цифровых данных. Каждый информационный блок представляются как одно число (конечная десятичная дробь).

Характеристики ЧЛБ (n, k)-кода, используемого в качестве основы для конкретной порождающей процедуры, в том числе количество информационных и проверочных элементов его кодовых комбинаций, определяют число блоков множества M цифровых элементов (чисел), подлежащих кодированию в конкретный момент времени. При этом последним, т.е. k-тым информационным элементом формируемой кодовой комбинации будет являться текущий (последний) блок, добавляемый в множество M.

При создании и добавлении в следующий момент времени к множеству M нового числового элемента (блока цифровых данных), нумерация информационных элементов в новой формируемой кодовой комбинаций ЧЛБ (n, k)-кода сдвигается в сторону нового блока цифровых данных, который на текущем этапе будет являться k-тым информационным элементом. Для определенных таким образом информационных элементов будут вычислены по правилам (4) соответствующие им проверочные элементы. Структура формируемой кодовой комбинации u_i будет иметь следующий вид:

$$ui = (ai\text{-}(k\text{-}1),1, ai\text{-}(k\text{-}2),2, ..., ai\text{-}(k\text{-}k),k, bi,1, bi,2, ... , bi,r), \qquad (6)$$

где $a_{i-(k-1),1}$, $a_{i-(k-2),2}$, …, $a_{i-(k-k),k}$ – последовательность чисел, представляющих k блоков цифровых данных, являющихся информационными элементами кодовой комбинации u_i;

i-(k-1), …, i-(k-k) – индексы чисел *a*, указывающие порядковый номер (позицию) элементов в последовательности *M*, составляющих последовательность информационных элементов (чисел) кодовой комбинации u_i;

1, …, k – индексы чисел *a*, указывающие их порядковый номер (позицию) в последовательности информационных элементов (чисел) кодовой комбинации u_i;

$b_{i,1}$, $b_{i,2}$, …, $b_{i,r}$ – последовательность проверочных элементов кодовой комбинации u_i, формируемых для информационных чисел $a_{i-(k-1),1}$, …, $a_{i-(k-k),k}$ по правилам ().

Если мощность множества *M* меньше *k*, что вполне возможно в самом начале его формирования, то несуществующие в такой момент времени информационные элементы с индексами $i \leq 0$ могут заменяться любой заранее обусловленной константой или нулями. Это позволит вычислить проверочные элементы для такой кодовой комбинации.

Абстрактная иллюстрация предложенного метода определения и формирования последовательности информационных элементов для одномерного линейного упорядоченного множества *M* (последовательности или цепи) представлена на рис. 1.

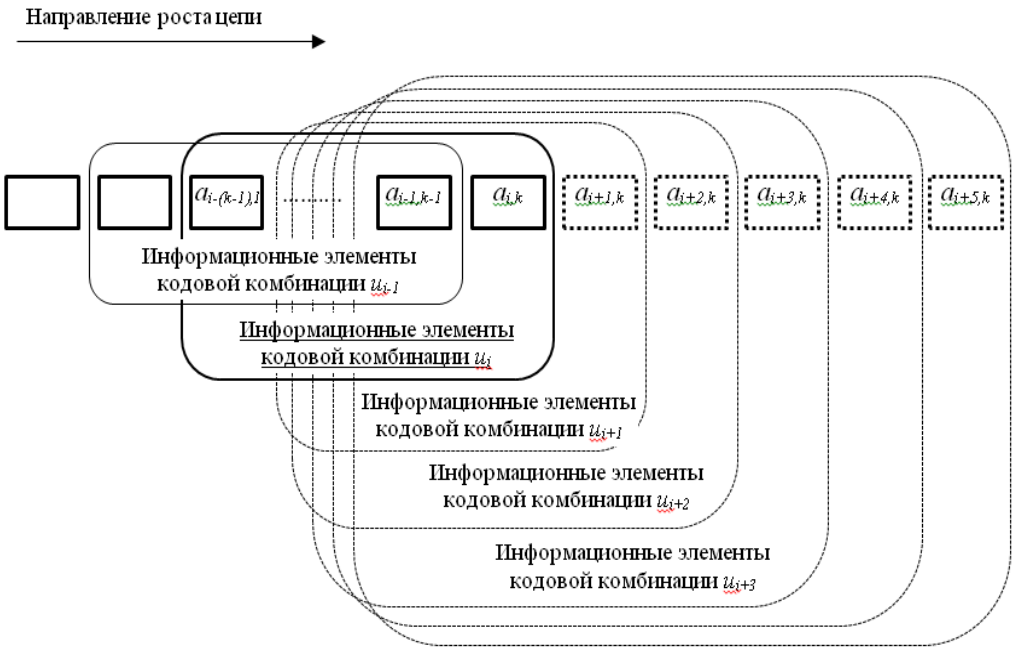

Рис. 1. Иллюстрация принципа формирования последовательностей информационных элементов кодовых комбинаций и $\in U_k$

3 Структура блока данных и варианты построения упорядоченной последовательности связного множества *M* (цепи)

Проверочные элементы сформированных кодовых комбинаций для последующего хранения и использования могут записываться (помещаться) как дополнительные служебные данные непосредственно в *k*-тый информационный блок (рис. 2 а) или записываться как самостоятельные блоки данных в отдельную, синхронизированную с исходным (порождающим их) множеством M цифровых элементов(чисел), последовательность блоков (цепь), как это показано на рис. 2 б. При этом такая самостоятельная служебная последовательность проверочных блоков (элементов) может храниться отдельно от основной информационной цепи.

Кроме этого, в зависимости от решаемой практической задачи, возможен вариант, когда последовательность проверочных элементов (чисел) может оформляться в собственный служебный блок, который может добавляться как равноправный отдельный блок к единой последовательности информационных и проверочных блоков, в которой информационные и проверочные блоки будут чередоваться.

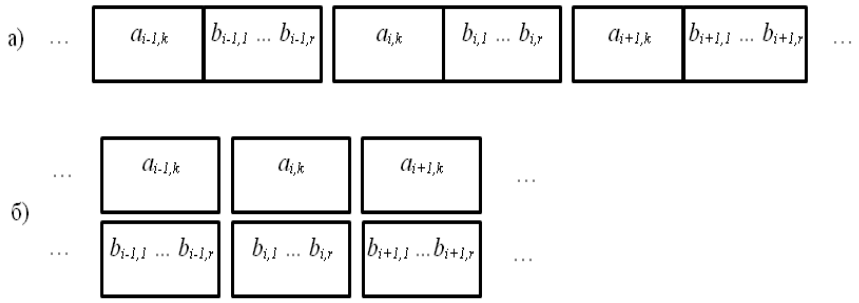

Рис. 2. Запись последовательности проверочных элементов в i-тый информационный блок (а) и формирование отельной цепи проверочных блоков, содержащих последовательности проверочных элементов (б)

4 Стойкость к модификации и влияние избыточности

С позиции условного потребителя обеспечение связности различных массивов цифровых данных, которые являются источниками критической для пользователя в конкретный момент времени информации, необходимо для исключения или существенного затруднения возможности их модификации и формирования у пользователя уверенности в целостности, достоверности и актуальности используемых данных. Это одно из основных требований для всех современных информационных, автоматизированных и иных систем, используемых обществом в век цифровой трансформации. Обеспечение целостности, доступности и достоверности различных массивов данных является в современном цифровом мире необходимым условием для формирования высокого уровня доверия к ним со стороны потенциальных потребителей.

Стойкость к модификации связного множества (массива) цифровых данных, защищенного посредством предложенного метода кодирования числовым кодом, непосредственно связана с корректирующей способностью конкретного ЧЛБ (n, k)-кода, который используется для этих целей. Чем выше минимальное кодовое расстояние d, тем большей кратности ошибки (искажения) он может обнаруживать и, соответственно, исправлять. Вместе с тем, необходимые значения минимального кодового расстояния достигаются за счет выбора оптимальной порождающей матрицы, а также введением необходимого количества проверочных элементов, которые определяют избыточность любого корректирующего кода.

Важным классом ЧЛБ-кодов являются коды с максимально достижимым кодовым расстоянием (ЧЛБ МДР-коды), которые имеют наименьшую избыточность [8, 9]. Количество проверочных чисел у таких кодов $r = d - 1$ (где d – минимальное кодовое расстояние ЧЛБ-кода). Вместе с тем, даже для МДР-кодов, исправляющих только одиночную ошибку ($d = 2$), необходимо наличие двух проверочных элементов ($r = 2$), что влечет за собой увеличение мощности связного множества цифровых данных за счет обязательного введения и хранения соответствующих служебных элементов. Это в свою очередь увеличивает объем (размер) физического представления на различных носителях всего массива цифровых данных.

Учитывая изложенное, необходимо выбирать оптимальное значение корректирующей способности ЧЛБ-кода, используемого для построения связного множества цифровых данных, исходя из принципа разумной достаточности. Это является отдельной исследовательской задачей. При этом существуют различные способы снижения (регулирования) избыточности. В частности, можно проводить не сплошное поступательное каскадное кодирование всех информационных элементов последовательности M один за другим, как это было предложено в рассмотренном методе (рис. 1), а осуществлять выбор информационных элементов с увеличенным шагом, при котором глубина рекурсии, а следовательно, и избыточность всей совокупности кодовых комбинаций будут меньше. Однако это в свою очередь снизит силу взаимосвязи информационных элементов последовательности M и ее стойкость к модификации.

Условная сила S установленной связи (т.е. гипотетическая стойкость к изменению или нарушению связи) зависит от сложности T порождающей процедуры, p – глубины рекурсии, n-мерности структуры множества M и мощности множества M:

$$S = \partial\,(T, p, n, |M|), \qquad (7)$$

где ∂ – функция соответствующих аргументов.

Интуитивно понятно: чем алгоритмически сложнее порождающая процедура и глубже рекурсия, тем сильнее связь между элементами множества, полученными в результате этой процедуры; чем

больше n, тем сильнее связаны элементы множества. При больших значениях $|M|$ и n (где $|M|$ – мощность множества M) изменение установленного ранее порядка или замена (модификация) элемента, с сохранением установленной связности множества M, может является трудновыполнимой задачей, а при $n > \mu$ эта задача может быть теоретически не выполнимой (где μ – некоторое пороговое значение, сложным образом зависящее от мощности множества M, алгоритмической сложности порождающей процедуры и глубины рекурсии).

Высокая стойкость связного предложенным методом множества M цифровых данных к модификации обусловлена тем, что модифицированный (каким-либо образом искаженный) элемент множества M является информационным элементом одновременно k кодовых комбинаций $u \in U_k$. Причем информационные элементы в каждой из этих кодовых комбинаций в свою очередь взаимосвязаны с другими информационными элементами. Таким образом, помимо кодовой комбинации u_i каждый ее информационный элемент еще входи в последовательность информационных элементов $k - 1$ смежных кодовых комбинаций, за исключением первых $k - 1$ и последних $k - 1$ информационных блоков последовательности M. Это каскадным образом обеспечивает одновременную, прямую или опосредованную, связь всех информационных элементов (блоков), входящих на данный момент времени в связную последовательность (цепь) M.

Таким образом, модификация одного информационного элемента нарушает целостность всей цепи M, что гарантирует обнаружение такого факта. Для сокрытия факта модификации необходим пересчет и новая перезапись всех кодовых комбинаций цепи, что является практически не выполнимой задачей. Стойкости к модификации также будет способствовать формирование отельной цепи проверочных блоков, содержащих только последовательности проверочных элементов как это показано на рис. 2 б, которая будет храниться и использоваться отдельно от основной информационной последовательности (цепи). При этом необходимо отметить, что последовательность проверочных элементов не допускает своей модификации в пределах корректирующей способности используемого для ее вычисления ЧЛБ-кода, поскольку проверочные элементы также являются полноценными элементами кодовой комбинации корректирующего кода.

Заключение

Предлагаемая порождающая процедура на основе ЧЛБ-кода относится к виду математических порождающий процедур. Использование таких процедур и свойств числовых корректирующих кодов представляется, по мнению автора, перспективным при построении связных упорядоченных множеств цифровых данных, имеющих высокую степень устойчивости к умышленной и случайной модификации. Это обусловлено, в том числе тем, что использование ЧЛБ-кодов открывает новые возможности по сравнению с традиционной технологией блокчейн, в которой взаимосвязанность информационных элементов (блоков) в цепи обеспечивается посредством функции хэширования. В частности, одним из основных свойств функции хэширования, которое лежит в основе организации и стимулирования майнинга Биткоина и других криптовалют, является то, что хэш значение не несет в себе никакой информации о цифровых данных, для которых оно вычислено. В случае использования ЧЛБ-кода для построения связного множества цифровых данных ситуация совершенно противоположная, а именно проверочные элементы кодовой комбинации содержат в себе некоторую необходимую информацию о информационных элементах, по которым они были вычислены, что позволяет определить модифицированные (искаженные) элементы этой кодовой комбинации, включая сами проверочные элементы. Это важное свойство позволяет не только автоматически восстановить, т.е. вычислить в пределах корректирующей способности ЧЛБ-кода модифицированный элемент кодовой комбинации, а также определить место и направление атаки, что представляется ценным при анализе и осуществлении противодействия возможным деструктивным воздействиям и минимизации их последствий. Кроме этого, достоинством ЧЛБ-кодов является то, что они допускают простое сложение массивов цифровых данных с другими однородными (согласованными) массивами цифровых данных с сохранением установленных свойств связности. Это способность обусловлена свойством ЧЛБ-кодов, при котором сумма кодовых комбинаций также является комбинацией ЧЛБ-кода.

Дополнительный синергетический эффект может иметь место также при одновременном применении предложенной технологии и технологий, основанных на методах криптографии (хэширования). В этом случае порождающая процедура, сочетающая в себе несколько различных методов, будет относится к виду комбинированных процедур.

Литература

1. https://bitcoin.org/bitcoin.pdf «Bitcoin: A Peer-to-Peer Electronic Cash System» Satoshi Nakamoto satoshin@gmx.com www.bitcoin.org
2. *Генкин А.* Блокчейн: Как это работает и что ждет нас завтра / Артем Генкин, Алексей Михеев. – М.: Альпина Паблишер, 2018. – 592 с.
3. *Тапскотт, Дон.* Технология блокчейн: то, что движет финансовой революцией сегодня / Дон Тапскотт, Алекс Тапскотт; [пер. с англ. К. Шашковой, Е. Ряхиной]. – Москва: Эксмо, 2018. – 448 с. – (Top Economics Awards).
4. *Будзко, Владимир И.; Милославская, Наталья Г.* Вопросы практического применения технологии блокчейна. Безопасность информационных технологий, [S.l]. Т. 26, № 1. С. 36 – 45, 2019. ISSN 2074-7136. Доступно на: <https://bit.mephi.ru/index.php/bit/article/view/1178>. Дата доступа: 19 feb. 2019. doi:http://dx.doi.org/10.26583/bit.2019.1.04.
5. *Горбатов В.А.* Фундаментальные основы дискретной математики. Информационная математика. – М.: Наука. Физматлит, 1999. – 544 с.
6. *Терентьев А.И.* Виды порождающих процедур связного множества цифровых данных / А.И. Терентьев // Гражданская авиация на современном этапе развития науки, техники и общества [Текст]: сборник тезисов докладов / Московский государственный технический университет гражданской авиации; редколлегия: Б. П. Елисеев (главный редактор) [и др.]. – М.: ИД Академии Жуковского, 2021. – 600 с.
7. *Хохлов Г.И.* Числовые линейные блоковые корректирующие коды / Г.И. Хохлов // Электронная техника. сер.10. Микроэлектронные устройства. 1991, вып.2.
8. *Терентьев А.И.* Элементы теории и практики числовых линейных блоковых корректирующих кодов. –М.: Альтекс, 2000. – 204 с.: ил.
9. *Терентьев А.И.* Некоторые результаты исследования границ для кодового расстояния двоичных, q–ичных и ЧЛБ (n, k)–кодов // Безопасность информационных технологий. 2007. № 4. 76–80 с.