ОБ ОДНОМ МЕТОДЕ ПОМЕХОЗАЩИЩЕННОГО КОДИРОВАНИЯ

Бодян Г. К.
Технический Университет Молдовы
Шт. чел Маре, 168, Кишинев – MD2012, Молдова
Тел.: (3732)237505; e-mail: gbodean@mail.md

Аннотация – Матроидный (М-) код – новый класс (линейных) корректирующих кодов, который способен восстановить исходные данные при потере (искажении) до половины переданных символов. Построение М-кода относится к задачам полиномиальной сложности. Предложен метод построения М-кода, основанный на поиске однородных (U-) матроидов среди циклоклассов сопряженных элементов векторного пространства над расширенным полем Галуа. Установлены границы существования U-матроидов.

I. Введение

В системах обработки, хранения и передачи дискретной информации под воздействием внешних и внутренних факторов часто происходит искажение данных. Причинами искажения данных могут быть царапины на поверхности компакт-дисков, несанкционированный доступ к информации, шумы в коммуникационных каналах, собственные шумы в приборах и устройствах СВЧ техники.

Для исправления различного рода ошибок применяется помехоустойчивое кодирование. Наибольшее распространение получило блоковое кодирование. Среди блоковых кодов выделим класс кодов с многоразрядными символами. В этом, важном для практики, случае поток данных представляет собой последовательность кодовых слов (векторов), состоящих из информационных и контрольных (проверочных) символов.

Корректирующая способность (эффективность) ρ кодов определяется числом исправленных ошибочных символов. Для линейных блоковых кодов установлен теоретический предел эффективности [1]: $\rho \le (n-k)/2$, где n – длина вектора, k - количество информационных символов (в слове).

В данной работе представлен метод помехоустойчивого кодирования с $\rho \le 1/2$. Метод, названный матроидное кодирование (или М-код), предусматривает линейное преобразование исходных символов, формирование выходного вектора, анализ полученного кода, распознавание ошибки и восстановление (если это возможно) исходных данных.

II. Основная часть

Для характеризации матроида воспользуемся понятием F-*представимости* [2]. Пусть $S=\{1,2,\ldots,n\}$, $V=F^k$- набор векторов-солбцов длины k над полем **F**. Определим отображение $\varphi: S \to V$, которое можно представить матрицей **A** размерности $k \times n$, при этом столбцы матрицы **A** есть $\varphi(1),\ldots, \varphi(n)$. Говорят, что матроид M от S *представим* над полем **F** (или F-*представим*), если отображение φ сохраняет ранг.

С точки зрения помехозащищенного кодирования представляет интерес F-представимость максимального ранга, т.е. случай *однородного* (U-) *матроида* ранга k. Причем представление U-матроидов должно быть реализовано над расширенным полем многочленов $\mathbf{GF}^k(2^m)$, где k - размерность поля, 2^m – характеристика поля, m - разрядность символов. Тогда матрица **A** будет определять линейное преобразование вида

$$\mathbf{v} = \mathbf{x} \bullet \mathbf{A}, \qquad (1)$$

где $\mathbf{x} = \langle x_1, \ldots, x_k \rangle$ - вектор исходных символов, $\mathbf{v} = \langle v_1, \ldots, v_n \rangle$ - результирующее кодовое слово, $\mathbf{A} = [a_{ij}]$, $a_{ij} \in \mathbf{GF}(2^m)$, $i=1,\ldots,k$; $j=1,\ldots,n$.

В случае U-представимости преобразование (1) задает систему линейных уравнений, в которой любая подсистема из k уравнений образует *линейно-независимый базис*.

Теперь возникает вопрос: каково должно быть соотношение между k и n? Очевидно, что для восстановления всех k информационных символов необходимо иметь, по крайней мере, еще k дополнительных базисных уравнений.

Таким образом, задача построения М-кода сводится к поиску однородного матроида U_k над полем $\mathbf{GF}^k(2^m)$. Приведем пример. Пусть $k=2$ и $m=2$. Нетрудно показать, что матрица

$$\mathbf{A}_{2\times 4} = \begin{bmatrix} 2 & 3 & 2 & 3 \\ 3 & 0 & 1 & 3 \end{bmatrix} \qquad (2)$$

представляет однородный матроид, т.е. любое подмножество ранга 2 матрицы (2) есть базис в векторном пространстве над $\mathbf{GF}^2(2^2)$. Это означает, что в системе уравнений

$$\mathbf{v} = \mathbf{x} \bullet \mathbf{A} = \begin{cases} 2x_1 + 3x_2 = v_1; \\ 3x_1 = v_2; \\ 2x_1 + x_2 = v_3; \\ 3x_1 + 3x_2 = v_4; \end{cases} \qquad (3)$$

все подсистемы ранга 2 являются линейно-независимыми. В общем случае, для заданного матроида U_k могут быть $C_{2k}^k = \frac{(2k)!}{k!\,k!}$ систем линейно-независимых уравнений (ЛЗУ). Все операции в (3) выполняются по модулю $g(x)$ $g(x) = \sum_{i=0}^{k} a_i x^i$, a_i – элементы расширенного поля $\mathbf{GF}(2^m)$ с *характеристическим полиномом* $p(x) = \sum_{i=0}^{m} b_i x^i$, $b_i \in$ GF(2). Для приведенного примера характеристическим является полином $p(x)=1+x+x^2$, а $g(x)$, называемый *генератором поля* (или *кода*), выбирают среди *неприводимых* (*примитивных*) *полиномов* вида $g(x) = 1 + \sum_{i=1}^{k} a_i x^i$.

Величина k является важной характеристикой матроидного кода. Она определяет его эффективность (корректирующую способность) ρ. Код, сгенерированный системой (3), способен восстановить до 2 ошибочных символов. Для этого (декодером) применяется одна из $C_4^2 = 6$ подсистем ЛЗУ.

Другой оценкой корректирующего кода является скорость передачи информации R, определяемая отношением k/n, которая для М-кода постоянна и равна ½. Следует отметить, что для кодов, корректирующих пакеты ошибок, характерны меньшие значения, как для корректирующей способности ρ, так и для скорости передачи R [1, 3].

2003 13th Int. Crimean Conference "Microwave & Telecommunication Technology" (CriMiCo'2003). 8-12 September, Sevastopol, Crimea, Ukraine
 ISBN: 966-7968-26-X. IEEE Catalog Number: 03EX697

Одна из основных проблем построения матроидного кода является поиск однородных матроидов U_k над $GF^k(2^m)$. Действительно, для нахождения U-матроидов ранга k потребуется, в общем, перебрать $C(k,m)=\mathbf{C}_{2^{mk}}^{2^k}$ комбинаций векторов. Например, для фиксированного m=4 имеем следующие значения переборной сложности: $C(2, 4)\approx 1.75\cdot 10^8$, $C(3, 4)\approx 6.535\cdot 10^{18}$ и $C(4, 4)\approx 8.436\cdot 10^{33}$. Для практически важных случаев, т.е. больших m и k, поиск однородных матроидов переборным методом становится нецелесообразным (из-за полиномиальной сложности задачи поиска)!

Один из подходов уменьшения "переборной сложности" - использование технологии *жадных* алгоритмов. Применительно к задаче поиска U–матроида "жадность" будет проявляться в (локальном) поиске *нормальных базисов* ограниченной длины [4]. Распространим понятие нормального базиса на поле многочленов $GF^k(2^m)$ и назовем такой базис *расширенным*. Расширенный нормальный базис (РНБ) имеет вид:

$$\{\gamma^1, \gamma^2, \gamma^4, \ldots, \gamma^\theta\},\ \theta = 2^{2k-1}, \qquad (4)$$

где γ - *порождающий* (*примитивный*) элемент поля $GF^k(2^m)$. Тогда задача поиска однородного матроида сводится к задаче поиска РНБ, сложность которой определяется мощностью семейства циклоклассов вида (4).

В нижеприведенной таблице представлены значения числа циклоклассов для тех случаев, когда в соответствующем векторном пространстве над $GF^k(2^m)$ существует РНБ. Символ Ø указывает на отсутствие РНБ, а символом √ отмечены случаи наличия РНБ в ограниченном диапазоне циклоклассов (из-за нехватки ресурсов ЭВМ).

Число циклоклассов и границы существования однородных матроидов над $GF^k(2^m)$

Numbers of cycloclasses and the bounds of uniform matroids over $GF^k(2^m)$

k \ m	2	3	4	5	6	7
2	4	27	104	426	1709	6835
3	12	170	1235	10099	80941	646410
4	Ø	Ø	15725	258281	4146454	66199674
5	Ø	Ø	Ø	6908009	222056530	√
6	Ø	Ø	Ø	Ø	√	√
7	Ø	Ø	Ø	Ø	Ø	√

Предложенный способ поиска U-матроидов на несколько десятков порядков меньше, чем поиск прямым перебором.

III. Заключение

В работе представлен метод построения матроидного кода, позволяющий восстановить исходное сообщение при потере до половины переданной информации.

IV. Список литературы

[1] *Питерсон У., Уэлдон Э.* Коды, исправляющие ошибки. – М.: Мир, 1976.

[2] *Айгнер М.* Комбинаторная теория. – М.: Мир, 1982.

[3] *MacWilliams F.J. and Sloane N.J.A.* The Theory of Error Correcting Codes. – North Holland, 1996.

[4] *Муттер В.М.* Основы помехоустойчивой телепередачи информации. – Л.: Энергоатомиздат, 1990.

A TECHNIQUE OF ERROR-CORRECTING CODING

Bodean G. C.
Moldova Technical University
168 St. cel Mare, Chişinau, Moldova, MD2012
phone (3732) 237505
e-mail: gbodean@mail.md

Abstract – The matroid (M-) code is a new type of error-correcting codes capable of restoring the original data despite half of the transmitted symbols being lost or distorted. M-coding is a complicated problem. A technique is suggested based on searching for uniform (U-) matroids among cycloclasses of the vector space transforms over the extended Galois field. The bounds of the U-matroid existence have been established.

I. Introduction

Error-correcting coding is widely used to restore the damaged data. For this purpose data streams are broken into series of codewords which consist of data and check symbols. Each symbol is a sequence of m bits. The efficiency (correcting capability) ρ of an error-correcting code (ECC) is determined by the number of erroneous symbols that have been corrected. An important division of ECCs are block codes. The efficiency of a linear block code has the following theoretical limits [1]: $\rho\le(n-k)/2$, where n is the codeword length and k is the number of information symbols.

The present paper deals with the development of a new ECC technique with $\rho\le½$. This technique, called matroid or M-coding, is based on a linear transformation of the original symbols, output vector generation, analysis of the codewords obtained, error recognition and restoration of the original data.

II. Main part

A notion of *F-representability* is used to define matroids [2]. For the ECC purposes, the *F*-representability of the maximal rank is of prime interest, i. e. the case of a *k*-rank *uniform* (U-) *matroid*. U-matroids should be represented over the extended Galois field $GF^k(2^m)$, in which case the matrix **A** would define the linear transformation (1) with the original codeword **x** and the output vector **v**.

The expression (1) defines a system of linear equations where any subsystem of k equations is a *linear-independent basis*. The number n of equations in (1) is taken as equal to double k, i. e. $n=2k$, in order to achieve a maximum efficiency of $\rho=1/2$.

Hence the tasks of M-coding are reduced to searching for a uniform matroid $U(k,2k)$ or U_k over $GF^k(2^m)$. As an example of the obtained U_2 matroid see (2) and (3) where the operations are performed modulo polynomial $p(x)$ over **GF**(2) and the vectors are generated by polynomials $g(x)$ over $GF(2^m)$. The combination of $\mathbf{C}_{2k}^{k}$ linear-independent subsystems (LIS) generally has a U_k-matroid. The value k defines the matroid code efficiency. Another important ECC feature is the rate R which has a constant value and for the M-code equals ½. The known cluster-correcting codes have smaller values of ρ and R [1, 3]; however, such performance is difficult to achieve. This is because the search for a U_k-matroid is a complicated task, i. e. it is NP-complete.

The technology of greedy algorithms should provide a way to decrease the enumerative complexity. The uniform matroids should be found among the *cycloclasses* of the length n. These cycloclasses are defined by (4), in which case the searching complexity is reduced drastically compared to enumeration procedures. The table above shows the cycloclass values for the vector spaces where uniform matroids are found; the symbol Ø denotes the absence of U-matroids, and the symbol √ denotes occurrences of such matroids within a restricted range of cycloclasses (due to insufficient computer resources).

III. Conclusion

A technique for the matroid coding allowing for the restoration of original message even with up to half of the transmitted data being lost is described and analyzed in this work.

2003 13th Int. Crimean Conference "Microwave & Telecommunication Technology" (CriMiCo'2003). 8-12 September, Sevastopol, Crimea, Ukraine

ISBN: 966-7968-26-X. IEEE Catalog Number: 03EX697