DEVELOPMENT OF UNIVERSAL DATA BLOCKS ERROR-CORRECTING TECHNIQUE BASED ON REED–SOLOMON CODES

Garnyshev I.N.[1], Kazantsev S.V.[2], Malkov R.Yu.[3], Semenov I.D.[4], Iudin S.V.[5]
(Russian Federation) Email: Kazantsev355@scientifictext.ru

[1]Garnyshev Igor Nikolaevich - Network Engineer,
DATA NETWORK ADMINISTRATION DEPARTMENT,
TINKOFF BANK;
[2]Kazantsev Sergei Vladimirovich - Senior Engineer,
NETWORK DEPARTMENT,
SBERBANK;
[3]Malkov Roman Yurevich – Expert,
CLOUD SOLUTIONS DEPARTMENT,
TECHNOSERV CLOUD,
MOSCOW;
[4]Semenov Ivan Dmitrievich - Senior Engineer,
NETWORK DEPARTMENT.
SERVERS.COM LIMASSOL, CYPRUS;
[5]Iudin Stepan Vyacheslavovich - Network Administrator,
DEPARTMENT OF TECHNICAL SUPPORT AND INFORMATION SYSTEMS INFRASTRUCTURE DEVELOPMENT,
SPORTMASTER, MOSCOW

Abstract: *the article analyzes the principles of digital data coding and decoding systems based on Reed-Solomon codes development. The proposed generalized scheme for representing Reed-Solomon error correcting codes. The constructed mathematical model of the finite field, which can be used for developing systems for the practical application of the Reed-Solomon code. A universal Reed-Solomon error coding algorithm has been developed, which is based on polynomials with two attributes and error-locator polynomial.*
Keywords: *block code, finite field, Reed-Solomon codes, error correcting coding systems, algebraic coding theory, modular arithmetic, error-locator polynomial.*

ПОСТРОЕНИЕ УНИВЕРСАЛЬНОЙ МЕТОДИКИ КОРРЕКЦИИ ОШИБОК В БЛОКАХ ДАННЫХ НА ОСНОВЕ КОДОВ РИДА-СОЛОМОНА

Гарнышев И.Н.[1], Казанцев С.В.[2], Мальков Р.Ю.[3], Семенов И.Д.[4], Юдин С.В.[5]
(Российская Федерация)

[1]Гарнышев Игорь Николаевич - сетевой инженер,
Отдел администрирования сетей передачи данных,
Тинькофф Банк;
[2]Казанцев Сергей Владимирович - главный инженер,
Департамент сетей передачи данных,
Сбербанк;
[3]Мальков Роман Юрьевич – эксперт,
Центр компетенций по облачным решениям,
Техносерв,
г. Москва;
[4]Семенов Иван Дмитриевич - старший инженер,
Департамент сетей передачи данных,
Servers.com Лимассол, Кипр;
[5]Юдин Степан Вячеславович - администратор сети,
Департамент технического обеспечения и развития инфраструктуры информационных систем,
Спортмастер, г. Москва

Аннотация: *в статье проведен анализ принципов построения систем кодирования и декодирования блоков цифровых данных на базе кодов Рида-Соломона. Предложена обобщенная схема представления кодов Рида-Соломона как систем кодирования с механизмом коррекции ошибок. Построена математическая модель конечного поля, которая может быть использована при разработке систем практического применения кода Рида-Соломона. Разработан универсальный алгоритм декодирования кодов Рида-Соломона, который базируется на полиномах двух переменных и полиномиальном индикаторе ошибок.*
Ключевые слова: *блочный код, конечное поле, код Рида-Соломона, помехоустойчивых систем кодирования, теория алгебраического кодирования, модулярная арифметика, полиномиальный индикатор ошибок.*

Введение

На сегодняшний день коды Рида-Соломона рассматриваются как базовый подход построения помехоустойчивых систем кодирования и декодирования цифровых данных [1-3]. Коды Рида-Соломона относятся к группе недвоичных циклических кодов, которые позволяют исправлять ошибки в блоках передаваемых данных. Характерно, что элементами кодового вектора являются группы бит, что обуславливает преимущество указанных кодов при построении систем восстановления данных с цифровых носителей, хранилищ архивной информации и помехоустойчивой передачи данных.

Анализ последних исследований и публикаций в данной области позволил обобщить представления о принципах характеризации конечного поля кода Рида-Соломона [3, 4], основах теории алгебраического кодирования [5, 6] и модулярной арифметики [7, 8]. Также рассмотрены методы представления конечного поля в виде полинома [9-11] и построения полиномиального индикатора ошибок [12-14].

Целью работы стало построение комплексной методологии по созданию помехоустойчивых систем кодирования и декодирования на основе кодов Рида-Соломона.

1. Характеризация конечного поля кода Рида-Соломона

На базовом уровне коды Рида-Соломона [3, 4] рассматриваются как помехоустойчивые системы кодирования, предназначенные для работы с конечным алфавитом (finite alphabet), который содержит большое количество элементов, т.н. расширенным алфавитом (large alphabet). При этом, следует отметить, что принципы построения кодов Рида-Соломона лежит в основе теории алгебраического кодирования [3-6], в рамках которой могут быть предложены методы исправления ошибок в блоках оцифрованных данных путем построения соответствующего математического аппарата (рис. 1).

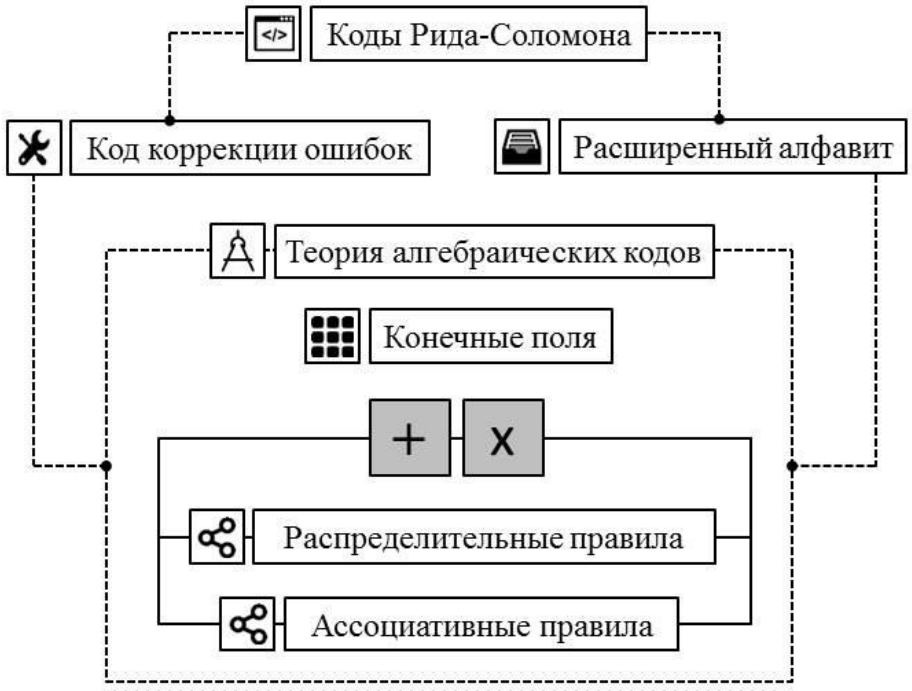

Рис. 1. Диаграмма представления кодов Рида-Соломона как помехо-устойчивых систем кодирования для расширенного алфавита

Описание кода и методов декодирования, а также конечный алфавит символов являются составными элементами структуры, которая называется конечным полем (finite field). Базовыми операциями для конечного поля являются операции сложения и умножения, которые подчиняются ассоциативным и распределительным правилам (рис. 1). Для построения математического аппарата в данном случае также следует ввести понятие нейтрального элемента как элемента, который оставляет любой другой элемент набора неизменным при применении бинарной операции, в частности элементов «0» и «1». Также, следует отметить, что у элемента «0» есть аддитивная инверсия (additive inverse), т.е. инверсия относительно операции сложения, а у элемента «1» — аддитивная инверсия и мультипликативная инверсия.

Простейший пример конечного поля в рамках предложенной методологии может быть представлен в рамках модулярной арифметики (modular arithmetic) и методики сравнения по модулю [7, 8]. Рассмотрим набор целых чисел $[0, 1, \dots i, j, \dots (I-1)]$ для которого величина I одновременно является как модулем умножения, так и модулем сложения, причем, очевидно, что в данном случае поиск и само обнаружение

факта наличия мультипликативных инверсий элементов является нетривиальной задачей. Для ненулевого элемента x можно рассмотреть ряд произведений $[0 \cdot x, 1 \cdot x \dots i \cdot x, j \cdot x, , \dots (I-1) \cdot x]$. В рамках доказательства от противного рассмотрим тот случай, когда два элемента ряда идентичны по модулю I. В таком случае $(i-j) \cdot x$ кратно m, но это невозможно, поскольку x, как и разница $(i-j)$ меньше m. Таким образом, произведения различны по модулю m, и одно из них равно «1». Данный подход наилучшим образом подходит для помехоустойчивого кодирования цифровых данных, в то время как расширение предложенной математической модели актуально только для криптографических задач, что выходит за рамки данной работы. Поскольку в практических приложениях рассматриваются двоичные данные, конечные поля включают в себя 2^I элементов.

2. Помехоустойчивое кодирование на базе кодов Рида-Соломона

Конечное поле в математической форме может быть представлено через полином $F(q)$, где q — элемент конечного поля, причем полный набор составляет Q элементов:

$$\begin{cases} F(q) = f_0 + f_1 \cdot z + \dots + f_m \cdot z^m + \dots + f_M \cdot z^M \\ z^m = \prod_{z=1}^{m}(z) \end{cases}, \quad (1)$$

причем следует заметить, что z^m являются целыми числами, которые сами по себе не входят в набор элементов $F(q)$. При этом полиномы как функции нескольких могут быть определены аналогичным образом и в дальнейшем при работе с ними применяются операции умножения и деления полиномов.

На основе данного представления конечного поля в виде полинома могут быть определены коды Рида-Соломона [9-11]. Рассмотрим набор элементов $x: \{x_1, x_2, \dots x_n, \dots x_N\}$ конечного поля $F(q)$. Для $M \leq N$ можно построить полином для $F(q)$ причем код Рида-Соломона будет представлен через набор выходных значений, который определяет каждый из переходов через элементы конечного алфавита:

$$U\big(F(q)\big): \{u(x_1); u(x_2) \dots, u(x_n) \dots, u(x_N)\}, \quad (2)$$

с учетом выполнения следующего условия

$$M \leq N \leq Q \quad (3)$$

Математическое же представление кода Рида-Соломона, которое может быть использовано в практических приложениях, подразумевающих кодирования и декодирование цифровых данных, может быть выражено как система уравнений:

$$\begin{cases} c_1 = \{f_1(x_1), f_1(x_2), \dots f_1(x_n), \dots f_1(x_N)\} \\ c_2 = \{f_2(x_1), f_2(x_2), \dots f_2(x_n), \dots f_2(x_N)\} \\ \dots \\ c_m = \{f_m(x_1), f_m(x_2), \dots f_m(x_n), \dots f_m(x_N)\} \\ \dots \\ c_M = \{f_M(x_1), f_M(x_2), \dots f_M(x_n), \dots f_M(x_N)\} \end{cases}. \quad (4)$$

Причем с учетом того, что код Рида-Соломона является линейным кодом, следует отдельно отметить, что:

$$a_1 \cdot c_1 \cdot c_1 + a_2 \cdot c_2 + \dots + a_q \cdot c_q = \big(g(x_1), g(x_2), \dots g(x_n), \dots g(x_N)\big), \quad (5)$$

где

$$\begin{cases} a_1, a_2, \dots a_q \in F(q) \\ g(x_n) = \sum_{q=1}^{Q}\Big(a_q \cdot f(x_n)\Big) \end{cases} \quad (6)$$

Представленные полиномы образуют векторное пространство $F(q)$ размерности M. Каждый полином генерирует свое кодовое слово, поскольку их разность будет полиномом степени меньшей M, т.к. ни один из полиномов не является нулевым. Кодовое слово имеет вес $w \geq N - K + 1$, т.к. полином степени K или меньшей имеет не более $(K+1)$ нулевых элементов. Таким образом, минимальное расстояние кода Рида-Соломона также может быть определено через N и K: $D = N - K + 1$.

3. Интерполяция при декодировании кодов Рида-Соломона

Математическая форма представления процесса получения приемником информации, может быть выражена в виде вектора $\bar{r}$, который является суммой вектора $\bar{c}$: $\{c_n\}$ и вектора ошибок, возникающих при декодировании $\bar{e}$, который характеризуется максимальным количеством ненулевых элементов T:

$$\begin{cases} \bar{r} = \bar{c}(N,K) + \bar{e}(T) \\ T = \dfrac{N-K}{2} \end{cases}. \qquad (7)$$

Полином P при этом может быть представлен как функция двух переменных x и y на которую наклатываются соответствующие ограничения [12-14]:

$$\begin{cases} P = P_0(x) + y \cdot P_1(x) \\ \begin{cases} P(x_n, r_n) = 0 \\ n \in [1; N] \end{cases} \\ \begin{cases} deg\big(P_0(x)\big) \le N - T - 1 \\ deg\big(P_1(x)\big) \le N - T - K \end{cases} \end{cases}, \qquad (8)$$

что позволяет ввести граничные значения для степеней обоих полиномов:

$$\begin{cases} L_0 = N - T - 1 \\ L_1 = N - T - K' \end{cases} \qquad (9)$$

а также на основе системы уравнений (7) граничное значение для суммы степеней обоих полиномов:

$$L = deg\big(P_0(x)\big) + deg\big(P_1(x)\big) \le N - 1 \qquad (10)$$

Таким образом систему уравнений, необходимую для определения полинома, можно сформировать как произведение двух матриц равное нулевой матрице:

$$RX \times P = 0 \qquad (11)$$

где

$$RX = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{L_0} & r_1 & r_1 \cdot x_1 & \dots & r_1 \cdot x_1^{L_1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{L_0} & r_2 & r_2 \cdot x_2 & \dots & r_2 \cdot x_2^{L_1} \\ & & & & & \dots & & & \\ 1 & x_n & x_n^2 & \dots & x_n^{L_0} & r_n & r_n \cdot x_n & \dots & r_n \cdot x_n^{L_1} \\ & & & & & \dots & & & \\ 1 & x_N & x_N^2 & \dots & x_N^{L_0} & r_N & r_N \cdot x_N & \dots & r_N \cdot x_N^{L_1} \end{bmatrix}, \qquad (12)$$

$$P = \begin{bmatrix} P_{0;0} \\ P_{0;1} \\ \dots \\ P_{0;L_0} \\ P_{1;0} \\ P_{1;1} \\ \dots \\ P_{1;L_1} \end{bmatrix}. \qquad (13)$$

Таким образом, вектор $\bar{r}$ включает не более T ошибок, в случае которых $\bar{r} \neq \bar{c}$. Из этого следует, что $P\big(x_n, u(x_n)\big) = 0$ в по меньшей мере $(N-T)$ случаях для элементов входного набора x: $\{x_1, x_2, \dots x_n, \dots x_N\}$. Поскольку $deg\left(P\big(x, u(x)\big)\right) \le L_0 = N - T - 1$ справедливо равенство:

$$u(x) = -\frac{P_0(x)}{P_1(x)} \qquad (14)$$

Другой подход, который может быть использован при декодировании цифровых данных заключается в получении P из $P_1 \cdot (y - u(x))$. P_1 дает нулевые значения для x, которые соответствуют ошибкам и может быть опеределен как полиномиальный индикатор ошибок (error-locator polynomial).

Выводы

В результате проведенного исследования был разработан математический аппарат на базе кодов Рида-Соломона, который в дальнейшем может быть использован для построения систем кодирования и декодирования блоков цифровых данных. В частности были предложены:

- обобщенная схема представления кодов Рида-Соломона как систем кодирования для расширенного конечного алфавита, в рамках которых подразумевается возможность коррекции ошибок;
- математическая модель конечного поля как базы для построения и практического применения кода Рида-Соломона;
- универсальный алгоритм декодирования кодов Рида-Соломона, построенный на основе полинома как функции двух переменных и полиномиального индикатора ошибок.

Предложенная методология может быть эффективно использована при построении комплексной методологии построения помехоустойчивых систем кодирования и декодирования, в которых используется код Рида-Соломона.

Список литературы / References

1. *Sungkar M. & Berger T.*, 2018. Discrete Reconstruction Alphabets in Discrete Memoryless Source Rate-Distortion Problems. 2018 IEEE International Symposium on Information Theory (ISIT). doi:10.1109/isit.2018.8437835.
2. *Lei W., Yizhou G., Fucai Z. & Yong W.*, 2018. The Method to Recognize Linear Block Code Based on the Distribution of Code Weight, 2018 13th APCA International Conference on Control and Soft Computing (CONTROLO). doi:10.1109/controlo.2018.8439758.
3. *Stamplecoskie S.*, 2006. A study of the concatenated Reed Solomon: convolutional coding performance used in WiMAX. Ottawa: Defence R&D Canada. Ottawa.
4. *Kythe D.K. & Kythe P.K.*, 2012. Algebraic and stochastic coding theory. Boca Raton, FL: CRC Press.
5. *Berlekamp E.R.*, 2011). Algebraic coding theory. New Jersey: World Scientific.
6. *Neubauer A., Freudenberger Jürgen & Kühn Volker*, 2007. Coding theory: algorithms, architectures, and applications. Chichester, England: John Wiley.
7. *Sato N.*, 2009. Modular arithmetic. Ottawa: Canadian Mathematical Society = Société mathématique du Canada.
8. *Hunter D.J.*, 2017. Essentials of discrete mathematics. Burlington, MA: Jones & Bartlett Learning.
9. *Bierbrauer J.*,2018. Singleton bound and Reed-Solomon codes. Introduction to Coding Theory, 71–80. doi: 10.1201/9781482296372-4.
10. *Kao M.-Y.*, 2008. Error-Control Codes, Reed–Muller Code. Encyclopedia of Algorithms, 281–281. doi: 10.1007/978-0-387-30162-4_128.
11. *Mishra V.*, 2016. Efficient data administration with reed-Solomon code. International Journal of Scientific Research and Management. doi: 10.18535/ijsrm/v4i12.03.
12. The polynomial method in error-correcting codes. (2016). University Lecture Series Polynomial Methods in Combinatorics, 37–49. doi: 10.1090/ulect/064/04.
13. *Caruso F., Orsini E., Sala M. & Tinnirello C.*, 2017. On the Shape of the General Error Locator Polynomial for Cyclic Codes. IEEE Transactions on Information Theory, 63 (6), 3641–3657. doi: 10.1109/tit.2017.2692213.
14. *Lee C.-D.*, 2011. Weak General Error Locator Polynomials for Triple-Error-Correcting Binary Golay Code. IEEE Communications Letters, 15(8), 857–859. doi: 10.1109/lcomm.2011.060811.110688.