Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра теоретических основ компьютерной безопасности и криптографии

Коды Рида-Маллера

АВТОРЕФЕРАТ

дипломной работы

студента 6 курса 631 группы

специальности 10.05.01 Компьютерная безопасность

факультета компьютерных наук и информационных технологий

Донских Дениса Сергеевича

Научный руководитель

доцент, к.ф.-м.н. _____ А.Н. Гамова

18.01.2018 г.

Заведующий кафедрой

профессор, к.ф.-м.н. _____ В.Н. Салий

18.01.2018 г.

Саратов 2018

ВВЕДЕНИЕ

В области математики и теории информации линейный код – это важный тип блокового кода, использующийся в схемах определения и коррекции ошибок. Линейные коды, по сравнению с другими кодами, позволяют реализовывать более эффективные алгоритмы кодирования и декодирования информации.

В процессе хранения данных и передачи информации по сетям связи неизбежно возникают ошибки. Контроль целостности данных и исправление ошибок – важные задачи на многих уровнях работы с информацией(в частности, физическом, канальном, транспортном уровнях модели OSI).

В данной работе мы подробно опишем коды Рида-Маллера, реализуем декодирование кодов Рида-Маллера, а так же создадим исскуствнное воздействие помех на передаваемое сообщение и попробуем обнаружить, и исправить ошибки, в процессе передачи сообщения.

Дипломная работа состоит из введения, 4 разделов, заключения, списка использованных источников и 2 приложений. Общий объем работы – 75 страниц, из них 42 страницы – основное содержание, включая 15 рисунков и 2 таблицы, список использованных источников из 21 наименования.

КРАТКОЕ СОДЕРЖАНИЕ

1. Основные понятия

В подразделе 1.1 «Кодирования информации» рассматриваются основные понятия кодов, кодирования и декодирования, а так же цели кодирования. *Код* представляет собой систему условных знаков (символов) для передачи, обработки и хранения информации (сообщения). *Кодирование* – это процесс представления информации (сообщения) в виде кода. Все множество символов, используемых для кодирования, называется алфавитом кодирования *Декодирование* представляет собой процесс обратного преобразования кода к форме исходной символьной системы, т.е. получение исходного сообщения.

В подразделе 1.2 «Коды обнаружения и исправления ошибок» рассматриваются корректирующие коды, принцип их работы и разновидности. *Корректирующие коды* – коды, служащие для обнаружения или исправления ошибок, возникающих при передаче информации под влиянием помех, а также при её хранении. Для этого при записи (передаче) в полезные данные добавляют специальным образом структурированную избыточную информацию, а при чтении (приеме) её используют для того, чтобы обнаружить или исправить ошибки. Естественно, что число ошибок, которое можно исправить, ограничено и зависит от конкретного применяемого кода.

В подразделе 1.3 «Блоковые коды» дается определение блоковых кодов. Пусть кодируемая информация делится на фрагменты длиной k бит, которые преобразуются в *кодовые слова* длиной n бит. Тогда соответствующий блоковый код обычно обозначают (n, k). При этом число $R = \frac{k}{n}$ называется *скоростью кода.*

В подразделе 1.4 «Линейные коды с избыточностью» рассматриваются двоичные коды с избыточностью, основное правило кодирования векторов и операции над ними. В двоичных кодах с контролем ошибок (синонимы: «помехоустойчивые коды», «корректирующие коды», «коды обнаружения и исправления ошибок», «коды с избыточностью») кодирование и декодирование заключается в преобразованиях строк из нулей и единиц. Формула $y = x \odot G$

задаёт правило кодирования исходного вектора $x \in V_k$ посредством его преобразования в вектор $y \in V_n$.

2. Коды Рида-Маллера

В подразделе 2.1 «Булевы функции и многочлены Жегалкина» объясняются основные понятия булевых функций и многочленов Жегалкина.

В подразделе 2.2 «Свойства кодов Рида-Маллера» рассматриваются основные теоремы, леммы и следствия, касающиеся кодов Рида-Маллера.

3. **Алгоритмы кодирования и декодирования**

В подразделе 3.1 «Коды Рида-Маллера 0-го порядка» описан принцип кодирование и декодирования кодов Рида-Маллера 0-го порядка. Кодирование сообщения проводится посимвольно, и $RM(0, m)$ – просто код с 2^m-кратным повторением каждого символа x. Кодовых слов всего два: нулевой вектор и вектор из единиц.

Если в полученном слове не все символы одинаковы, мы заключаем, что при передаче сообщения имелись ошибки. Декодирование при этом очевидно: если большая часть битов в полученном векторе – единицы, то считаем, что $x = 1$, если нули – считаем, что $x = 0$. Если ошибочными были меньше половины битов, то есть не более $2^{m-1} - 1$, то ошибки будут исправлены.

В подразделе 3.2 «Коды Рида-Малера 1-го порядка» рассматривается кодирование кодов Рида-Маллера 1-го порядка. Зафиксируем $m \geq 1$ и составим матрицу $G_1(m)$ из m строк и 2^m столбцов, где j-й столбец (при нумерации от 0) соответствует двоичному представлению числа j, записанному в m битах.

$$G_1(1) = (0\ 1), \quad G_1(4) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Кодирование проводится по основному для линейного кода правилу при $G = G_{1,m}$. Размер исходного слова при этом равен $m + 1$, по числу строк матрицы $G_{1,m}$. Так, код $RM(1, 4)$ переводит слово $x = 10011$ в кодовое слово $y = xG = 1001\ 1001\ 1001\ 1001$.

В подразделе 3.3 «Кодовое расстояние» определяется кодовое расстояние кодов Рида-Маллера. Корректирующая способность всякого кода характеризуется его *кодовым расстоянием*, определяемым как минимальное расстояние Хэмминга $\rho(u, v)$ между векторами u, v этого кода, то есть как число несовпадающих компонент этих векторов. Чем больше это расстояние, тем проще распознать кодовое слово, если оно слегка искажено. Другими словами кодовое расстояние кода вычисляется как минимальный вес его ненулевого кодового слова

Для $RM(1, m)$ минимальный вес ненулевого кодового слова, он равен 2^{m-1}. Это и есть кодовое расстояние d данного кода, при этом $\left[\frac{d-1}{2}\right] = 2^{m-2} - 1$, так что $RM(1, m)$ обнаруживает до $2^{m-1} - 1$ ошибок в слове и исправляет до $2^{m-2} - 1$ ошибок (включительно), это очень хороший показатель.

В подразделе 3.4 «Декодирование кодов Рида-Малера 1-го порядка» описывается алгоритм декодирования кодов Рида-Маллера 1-го порядка. Декодирование кодов Рида-Маллера 1-го порядка происходит по принципу максимального правдоподобия. Это означает выбор из всех возможных кодовых слов того слова y, которое находится на минимальном расстоянии Хэмминга от принятого слова y', и лишь затем – восстановление исходного слова x из его кода y.

Рассмотрим применение принципа максимального правдоподобия к декодированию кода $RM(1, m)$.

Пусть y' – принятое слово длины 2^m, возможно, содержащее ошибки, и пусть Y' означает вектор, полученный из y' заменой всех компонент 0 компонентами −1:

$$Y' = 2y' - 1.$$

Если c – какое-то кодовое слово (будем обозначать это как $c \in RM(1, m)$), то оно, как отмечалось выше, является строкой двоичной матрицы Адамара A_m или её дополнения $\bar{A}_m$. Тогда соответствующий вектор

$$C = 2c - 1$$

есть строка матрицы H_m или $-H_m$ (будем обозначать это как $C \in H_m$ или $C \in -H_m$). Расстояние $\rho(y', c)$ равно расстоянию $\rho(Y', C)$.

При вычислении скалярного произведения $\langle Y', C\rangle$каждая несовпадающая компонента даст слагаемое -1, а каждая совпадающая даст 1. Таким образом,

$$\langle Y', C\rangle = N_0 - N_1 = N - 2\rho(Y', C),$$

где N_0 – число совпадений компонент Y'и $C, N_1 = \rho(Y', C)$ – число несовпадений, $N = N_0 + N_1$ – длина слова, $N = 2^m$.

Там, где значение $\rho(Y', C)$ минимально, скалярное произведение $\langle Y', C\rangle$ максимально.

Алгоритм декодирования принятого вектора Y' следующий:

1) Умножить матрицу Адамара H_m на столбец Y' (или, что даст то же самое, умножить строку Y' на матрицу Адамара).

2) Найти максимальную по абсолютной величине компоненту полученного вектора. Пусть её номер i.

3) Если эта компонента положительна, определить кодовое слово y, ближайшее к y', как равное i-й строке двоичной матрицы Адамара A_m. В противном случае y будет дополнением к этой строке (i-й строкой $\overline{A}_m$).

После этого для восстановления исходного сообщения x по $y \in RM(1, m)$ достаточно воспользоваться уравнениями из системы $y = x \odot G$ с номерами $0, 1, 2, 4, \dots, 2^{m-1}$:

$$y_0 = x_0,\ \ y_{2^i} = x_0 \oplus x_{m-i},\ \ i = 0, 1, 2, \dots, m-1\ ,$$

то есть

$$\begin{gathered} y_0 = x_0, \\ y_1 = x_0 \oplus x_m, \\ y_2 = x_0 \oplus x_{m-1}, \\ y_4 = x_0 \oplus x_{m-2}, \end{gathered}$$

и так далее.

Складывая равенство $y_0 = x_0$ с каждым из остальных, получаем для компонент вектора x равенства

$$x_0 = y_0,\ \ x_{m-i} = y_0 \oplus y_{2^i},\ \ i = 0, 1, 2, \dots, m-1\ .$$

В зависимости от знака наибольшей компоненты скалярного произведения мы получим y_0, равное 0 (при $y \in \overline{A}_m$) или 1 (при $y \in A_m$). В первом случае вектор x будет подвектором y, во втором – дополнением к подвектору y.

В подразделе 3.5 «Коды Рида-Маллера r-го порядка» описано кодирование кодов Рида-Маллера r-го порядка. Определим теперь код $RM(r,m)$ при произвольном $r < m$. На основе матрицы $G_1(m)$ выстроим цепочку матриц $G_2(m), \dots, G_r(m)$ по следующему правилу: строки матрицы $G_i(m)$ – это всевозможные поточечные (покомпонентные, побитовые) произведения по i строк из $G_1(m)$.

Число строк матрицы $G_i(m)$ равно C_m^i. Вес каждой строки матрицы $G_i(m)$ равен 2^{m-i}. Кодирование проводится по основному для линейного кода правилу при $G = G_{r,m}$

В подразделе 3.6 «Декодирование кодов Рида-Маллера r-го порядка» рассматривается алгоритм декодирования кодов Рида-Маллера r-го порядка. Чтобы декодировать вектор y, полученный из вектора x, надо решить систему уравнений относительно x.

Для $G = G_{r,m}$ эта система имеет вид:

$$y[N] = x[K(r)] \odot G[K(r), N],$$

где $K(r) = K_0 \cup K_1 \cup \dots \cup K_r$ – множество индексов строк матрицы G, $N = \{0, 1, \dots, 2^m - 1\}$ – множество индексов столбцов матрицы G.

Так как $K(r) = K(r-1) \cup K_r$, эту систему можно переписать в виде

$$y[N] = x[K(r-1)] \odot G[K(r-1), N] + x[K_r] \odot G[K_r, N]. \quad (3.1)$$

При решении системы уравнений (3.1) вначале определяют старшие компоненты $x[K_r]$. Для определения каждого $x_i = x[i], i \in K_r$, получается – как именно, будет показано ниже – набор из 2^{m-r} независимых уравнений вида $x_i = \beta_j$, где каждое β_j вычисляется как некоторая линейная функция от компонент вектора y.

Если все значения $\beta_j, j \in 1:2^{m-r}$, равны между собой, то их общее значение, естественно, и будет искомым значением x_i,. Но из-за возможных ошибок вектор y может не быть кодовым словом и не удовлетворять (3.1), тогда среди β_j, могут быть и нули, и единицы. В связи с этим решение принимается по мажоритарному принципу: если единиц среди β_j больше, чем нулей, то присваиваем $x_i := 1$, если наоборот, то $x_i := 0$. Будет обнаружено до $2^{m-r} - 1$ ошибок (сообщать об ошибке будем тогда, когда не все β_j одинаковы) и исправлено до $2^{m-r-1} - 1$ ошибок (исправления корректны, когда число искажённых символов меньше, чем неискажённых).

После того, как $x[K_r]$ определены, их значения подставляют в решаемую систему уравнений (3.1) и приходят к системе

$$y^{(r-1)}[N] = x[K(r-1)] \odot G[L(r-1), N],$$

где $y^{(r-1)}[N] = y[N] \oplus x[K_r] \odot G[K_r, N]$. Заметим, что $G[K_r, N] = G_r(m)$.

В этой системе тоже определяют старшие компоненты, теперь это $x[K_{r-1}]$, и вновь понижают число уравнений и неизвестных системы. Наконец, приходят к равенству $y^{(0)} = x[0] \odot G[0, N] = x_0 G_0(m)$, которое определяет $x_0 = x[0]$ по мажоритарному принципу.

4. Программная реализация.

На языке Java была создана программа «client», реализующая кодирование и передачу сообщений, и программа «server», которая осуществляет прием сообщений и декодирование кодов Рида-Маллера, а также обнаруживает и исправляет ошибки, возикающие при передаче сообщений.

ЗАКЛЮЧЕНИЕ

В данной работе были рассмотрены коды обнаружения и исправления ошибок, а именно коды Рида-Маллера. Рассмотрены коды Рида-Малера 0-го, 1-го, r-го порядка, рассмотрены алгоритмы декодирования кодов Рида-Маллера. По изученным алгоритмам была создана программа клиент-сервер, которая способна обнаруживать и исправлять ошибки, возникающие при передаче информации под воздействием помех.

Таким образом, все поставленные задачи были полностью решены.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Мак-Вильямс, Ф. Дж. Теория кодов, исправляющих ошибки. / Ф. Дж. Мак-Вильямс, Н. Дж. А. Слоэн. М.: Связь, 1979. 549 с.
2. Сидельников В. М. Теория кодирования [Электронный ресурс] / В. М. Сидельников. М.: Физматлит, 2006. 298 с. Загл. с экрана. Яз. Рус.
3. Логачев, О. А. Булевые функции в теории кодирования и криптологии / О. А. Логачев, А. А.Сальников, В. В.Ященко. М.: МЦНМО, 2004. 367 с.
4. Питерсон У. Коды, исправляющие ошибки [Электронный ресурс] / У. Питерсон, Э. Уэлдон. М.: Мир, 1986. 251 с. Загл. с экрана. Яз. Рус.
5. Сагалович, Ю. Л. Введение в алгебраические коды / Ю. Л. Сагалович. М.: Мир. 2011. 302 с.
6. Морелос-Сарагоса Р. Искуство помехоустойчивого кодирования. Методы, алгоритмы, применение [Электронный ресурс] / Р. Морелос-Сарагоса. М.: Техносфера, 2005. 320 с. Загл. с экрана. Яз. Рус.
7. Берлекэмп Э. Алгебраическая теория кодирования [Электронный ресурс] / Э. Берлекэмп. М.: Мир, 1991. 477 с. Загл. с экрана. Яз. Рус.
8. Блэйхут, Р. Теория и практика кодов, контролирующих ошибки / Р Блэйхут. М.: Мир, 1986. 576 с.
9. Афанасьев В. Б. Кодирование в радиоэлектронике [Электронный ресурс] / В. Б. Афанасьев. М.: Радио и связь, 1986. 176 с. Загл. с экрана. Яз. Рус.
10. Хорстманн Кей С. Java. Библиотека профессионала / С. Хортсман Кей, Г. Корнелл. М. : ООО Вильямс, 2014. 598 с.
11. Золотарев В. В. Искусство помехоустойчивого кодирования [Электронный ресурс] / В. В. Золотарев, Г. В. Овечкин. М.: Техносфера, 2005. 321 с. Загл. с экрана. Яз. Рус.
12. Лидовский В. В. Теория информации [Электронный ресурс] / В. В. Лидовский. М.: Наука, 2003. 112 с. Загл. с экрана. Яз. Рус.

13. Биккенин Р. Р. Помехоустойчивость систем [Электронный ресурс] /Р. Р. Биккенин, С. Д. Хворов. М.: Наука и технологии, 2011. 271 с. Загл. с экрана. Яз. Рус.

14. Волков Л. Н. Системы цифоровой связи [Электронный ресурс] / Л. Н. Волков. М.: Экотрендз, 205. 395 с. Загл. с экрана. Яз. Рус.

15. Садовский Л. Е. Коды и математика [Электронный ресурс] / Л. Е. Садовский, М. Н. Аршинов. М.: Наука, 1993. 144 с. Загл. с экрана. Яз. Рус.

16. Конопелько В. К. Теория прикладного кодирования [Электронный ресурс] / В. К. Конопелько, В. А. Липницкий. М.: БГУИР, 2004. 398 с. Загл. с экрана. Яз. Рус.

17. Глушков В. А. Теория электрической связи [Электронный ресурс] / В. А. Глушков, А. В. Дормидонтов, А. Г. Нестеренко. М.: Мир, 2008. 452 с. Загл. с экрана. Яз. Рус.

18. Дмитриев В. И. Прикладная теория информации [Электронный ресурс] / В. И. Дмитриев. М.: Высшая школа, 1999. 485 с. Загл. с экрана. Яз. Рус.

19. Вишневский В. М. Беспроводные системы передачи информации [Электронный ресурс] / В. М. Вишневский М.: Техносфера, 2005. 592 с. Загл. с экрана. Яз. Рус.

20. Королев А. И. Коды и устройства помехоустойчивого кодирования информации [Электронный ресурс] / А.И. Королев. М.: Москва, 2010. 279 с. Загл. с экрана. Яз. Рус.

21. Матвеев Б. В. Основы корректирующего кодирования [Электронный ресурс] / Б. В. Матвеев. М.: БХВ-Петербург, 2016. 192 с. Загл. с экрана. Яз. Рус.