РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ

Кафедра информационной безопасности

Баранова Е.К.

Методические указания к выполнению

ЛАБОРАТОРНОЙ РАБОТЫ ПО ДИСЦИПЛИНЕ

“Методы и средства защиты компьютерной информации”

Тема: **Корректирующие коды**

Москва 2007

1. Некоторые виды корректирующих кодов

Понятие о корректирующих кодах

Обрабатываемая информация обычно представляется различными комбинациями из двух символов 0 и 1, соответственно, любой процесс кодирования состоит из преобразования чисел и слов в соответствующие комбинации 1 и 0. Введем некоторые понятия из теории кодирования.

Код - это есть совокупность всех комбинаций из определенного количества символов, которые избраны для представления информации. Каждая такая комбинация называется *кодовой комбинацией*. Общее число кодовых комбинаций в данном коде может быть равно или меньше числа всех возможных комбинаций из данного количества символов.

Коды подразделяются на равномерные и неравномерные.

Равномерные - такие коды, в которых все комбинации имеют одинаковое количество знаков.

Неравномерные - такие коды, в которых количество знаков может быть различным. Примером такого кода может служить известный телеграфный код Морзе.

С помощью n двоичных знаков, очевидно можно получить 2^n кодовых комбинаций. В зависимости от того все возможные 2^n кодовые комбинации задействованы для представления информации или нет, коды подразделяются на *простые* и *корректирующие (избыточные)*.

Простые - такие коды, в которых используются все возможные 2^n комбинации, полученные с помощью n двоичных знаков. В таком коде всякая ошибка, состоящая в изменении 0 на 1 или 1 на 0 превращает одну информационную комбинацию в другую. Для обнаружения и исправления ошибки в таком коде необходима дополнительная информация.

Пример 1. Пусть n=3 , тогда количество возможных кодовых комбинаций 2^n = 8. Простой код для n=3 будет иметь вид:

0 000
1 001
2 010
3 011
4 100
5 101
6 110
7 111

Корректирующие - такие коды, в которых лишь некоторая часть всех возможных 2^n комбинаций, полученных с помощью n двоичных знаков, используется для представления информации. В таком коде все остальные кодовые комбинации являются запрещенными и их появление, свидетельствует о наличии ошибки.

> *Пример 2.* Пусть n=3, но из всех возможных кодовых комбинаций, представленных в предыдущем примере, только четыре изображают числа от 0 до 3, а остальные считаются запрещенными. Такой корректирующий код будет иметь вид:
>
> 0 000
> 1 011
> 2 101
> 3 110

Очевидно, что любая ошибка в таком коде превращает информационную комбинацию в запрещенную.

В свою очередь, корректирующие коды можно разделить на *систематические* и *несистематические*.

Систематические - такие n-значные коды, которые содержат постоянное количество m информационных и $k = n - m$ избыточных знаков, причем эти знаки занимают одни и те же позиции во всех кодовых комбинациях.

Несистематические - такие коды, в которых знаки закодированного числа или слова разделить на информационные и контрольные невозможно.

Основными характеристиками корректирующих кодов служат их *избыточность* и *корректирующая способность.*

Избыточность - кода определяется по формуле $k = n - m$, где n - общее число знаков в коде, m - число информационных знаков, необходимых для изображения N чисел или слов в простом коде, определяемое по формуле
$m = \log_2 N$, k - число контрольных знаков. Например, для кода из *примера 2* $m = \log_2 4 = 2$, $k = 1$.

Часто для характеристики кода применяют понятие *относительная избыточность,* которая определяется из соотношения $R = k / m$..

Корректирующая способность кода количественно может быть определена вероятностью обнаружения или исправления ошибок различных типов. Разработка кодов, имеющих максимальную корректирующую способность при заданной избыточности, а также кодов, обеспечивающих заданную корректирующую способность при минимальной избыточности - одна из важнейших задач теории кодирования.

Корректирующая способность кода связана с понятием кодового расстояния. Прежде чем сформулировать определение кодового расстояния, введем понятие о весе кодовой комбинации.

Вес, W(A), кодовой комбинации *A* определяется количеством содержащихся в ней двоичных единиц.

Пример 3.

Для $A = 111001$, $W(A) = \Sigma a_i = 4$.

Кодовое расстояние между двумя кодовыми комбинациями определяется числом позиций, в которых их элементы не совпадают.

Это означает, что кодовое расстояние между комбинациями *A* и *B* , равно весу некоторой третьей комбинации *C,* полученной поразрядным сложением двух этих комбинаций в соответствии со следующей формулой:

$$W(C) = W(A \oplus B) = \Sigma (a_i \oplus b_i)$$

Пример 4. Пусть имеется кодовая комбинация A =111001 и кодовая комбинация B = 100101, тогда

$$\begin{array}{r} 111001 \\ \oplus\ 100101 \\ \hline \end{array}$$

C = 011100 $W(C) = 3$, кодовое расстояние между A и B

Минимальное кодовое расстояние α, данного кода, есть минимальное расстояние между двумя любыми комбинациями в этом коде. Если, например, в данном коде есть хотя бы одна пара комбинаций, которые отличаются друг от друга только в одной позиции, то минимальное расстояние данного кода $\alpha = 1$. Так для кода из *примера 1,* $\alpha = 1$, а для кода из *примера 2*, $\alpha = 2$.

Рассмотрим некоторые корректирующие коды.

Код с проверкой на четность

Простейший корректирующий код - код с проверкой на четность, который образуется добавлением к группе информационных разрядов одного избыточного, значение которого выбирается таким образом, чтобы сумма единиц в кодовой комбинации (т.е. вес кодовой комбинации) была всегда четна.

Пример 5. Рассмотрим код с проверкой на четность, образованный добавлением контрольного разряда к простому коду из *примера 1*.

	Информационные разряды	Контрольный разряд
0	000	0
1	001	1
2	010	1
3	011	0
4	100	1
5	101	0
6	110	0
7	111	1

Таким образом, если в простом коде число 4 имеет изображение
100, то в коде с проверкой на четность оно будет изображаться
комбинацией 1001.

Минимальное кодовое расстояние кода с проверкой на четность $\alpha = 2$. Такой код обнаруживает все одиночные ошибки и групповые ошибки нечетной кратности, так как четность количества единиц в этом случае будет также нарушаться.

Следует отметить, что при кодировании целесообразно число единиц в кодовой комбинации делать нечетным, и осуществлять контроль на нечетность, в этом случае любая комбинация, в том числе и изображающая нуль, будет иметь хотя бы одну единицу, что дает возможность отличить полное отсутствие информации от передачи нуля.

Код Хэмминга[1]

Код Хэмминга представляет собой систематический код, имеющий большую относительную избыточность, нежели код с проверкой на четность и предназначен либо для исправления одиночных ошибок (при $\alpha = 3$), либо для исправления одиночных и обнаружения без исправления двойных ошибок (при $\alpha = 4$). *N* - значный код Хэмминга имеет *m* - информационных разрядов и *k* - контрольных. Число контрольных разрядов должно удовлетворять соотношению:

$$k >= log_2 (n+1),$$

откуда $$m <= n - log_2 (n+1).$$

Код Хэмминга строится таким образом, что к имеющимся информационным разрядам кодовой комбинации добавляется вычисленное по вышеприведенной формуле количество контрольных разрядов, которые формируются путем подсчета четности суммы единиц для определенных групп информационных разрядов. При приеме такой кодовой комбинации из полученных информационных и контрольных разрядов путем аналогичных подсчетов четности составляют корректирующее число, которое равно нулю, при отсутствии ошибки, либо указывает номер ошибочного разряда.

Рассмотрим подробнее процесс кодирование для кода с $\alpha = 3$.

Пусть первый контрольный разряд имеет нечетный порядковый номер, установим его при кодировании таким образом, чтобы сумма единиц всех разрядов с нечетными порядковыми номерами была равна 0. Такая операция может быть записана в виде соотношения:

$$E_1 = a_1 \oplus a_3 \oplus a_5 \oplus a_7 \ \ldots = 0,$$

где a_1, a_3, a_5, a_7 - двоичные символы, размещенные в разрядах с номерами

1, 3, 5,7,…

Появление единицы во втором разряде (справа) корректирующего числа означает ошибку в тех разрядах кодовой комбинации, порядковые номера которых (2, 3, 6, 7, …) в двоичном изображении имеют единицу во втором справа разряде. Поэтому вторая операция кодирования, позволяющая найти второй контрольный разряд, имеет вид:

$$E_2 = a_2 \oplus a_3 \oplus a_6 \oplus a_7 \ \ldots = 0,$$

[1] Хэмминг Ричард (R.Hamming) - лауреат премии Тьюринга 1968 г. Нью Джерси, за работу по кодам, исправляющим ошибки, которую он проделал для AT& Bell Laboratories. С 1976 г. преподавал в Naval Postgraduate School, Монтерей, Калифорния. Автор работ по теории веростностей и комбинаторике.

Рассуждая аналогичным образом:

$$E_3 = a_4 \oplus a_5 \oplus a_6 \oplus a_7 \ \ldots = 0,$$

$$E_4 = a_8 \oplus a_9 \oplus a_{10} \oplus a_{71} \ \ldots = 0,$$

После приема кодовой комбинации, совместно со сформированными контрольными разрядами, выполняются те же операции подсчета, что были описаны выше, а полученное число:

$$E_k \ E_{k-1} \ldots E_2 \ E_1$$

считается корректирующим, причем при $E_k \ E_{k-1} \ \ldots \ E_2 \ E_1 = 0$ ошибки отсутствуют, а при наличии ошибок неравными нулю оказываются те суммы E_i , в образовании которых участвовал ошибочный разряд. Корректирующее число при этом будет равно порядковому номеру этого разряда.

Выбор места для контрольных разрядов в каждой из кодовых комбинаций определяется таким образом, чтобы контрольные разряды участвовали только в одной операции подсчета четности. Это упрощает процесс кодирования., такими позициями являются целые степени двойки: 1, 2, 4, 8, 16,… .

Пример 6. Составим шестизначный код Хэмминга $n = 6$, $k >= log_2 7$, $k = 3$, $m = n - k = 3$

Цифра	**Простой код**	**Код Хэмминга**					
		6	**5**	**4**	**3**	**2**	**1**
0	000	0	0	0	0	0	0
1	001	0	0	0	1	1	1
2	010	0	1	1	0	0	1
3	011	0	1	1	1	1	0
4	100	1	0	1	0	1	0
5	101	1	0	1	1	0	1
6	110	1	1	0	0	1	1
7	111	1	1	0	1	0	0

Принят код: 111100 *исправлено* 110100 - *ошибка по корректирующемучислу в разряде 4;*

111010 *исправлено* 101010 - *ошибка по корректирующемучислу в разряде 5;*

100000 *исправлено* 000000 - *ошибка по корректирующему числу в разряде 6.*

К *k* - контрольным разрядам может еще добавляться *(k+1)*-й разряд, обеспечивающий дополнительный контроль по четности всей кодовой комбинации.

При проверке информации после ее приема возможны три случая.

- Отсутствие ошибок - корректирующее число равно нулю, общая четность суммы единиц кодовой комбинации правильна.
- Одиночная ошибка - контроль общей четности кодовой комбинации обнаруживает ошибку, корректирующее число указывает номер искаженного разряда (если корректирующее число равно нулю - ошибка произошла в разряде общей четности).
- Двойная ошибка - корректирующее число не равно нулю, контроль общей четности кодовой комбинации не обнаруживает ошибки.

Циклический код

Циклические коды являются разновидностью систематических кодов и поэтому обладают всеми их свойствами. Характерной особенностью циклического кода, определяющей его название, является то, что если *n* - значная кодовая комбинация $a_0\ a_1\ a_2\ \ldots\ a_{n-1}\ a_n$ принадлежит данному коду, то и комбинация $a_n\ a_0\ a_1\ a_2\ \ldots\ a_{n-1}$, полученная циклической перестановкой знаков, также принадлежит этому коду.

Идея построения циклических кодов базируется на использовании *неприводимых многочленов*. Неприводимым называется многочлен, который не может быть представлен в виде произведения многочленов низших степеней, то есть такой многочлен, который делится только на самого себя или на 1.

Неприводимые многочлены при построении циклических кодов играют роль так называемых образующих полиномов, от вида которых, собственно и зависят основные характеристики полученного кода: избыточность и корректирующая способность. В таблице 1 указаны неприводимые многочлены со степенями *k=1,2,3,4.*

Таблица 1

k=1	$x+1$	*11*
k=2	x^2+x+1	*111*
k=3	x^3+x+1	*1011*
	x^3+x^2+1	*1101*

$k=4$	x^4+x+1	10011
	x^4+x^3+1	11001
	$x^4+x^3+x^2+x+1$	11111

Основные принципы кодирования в циклическом коде заключаются в следующем. Двоично-кодированное n - разрядное число представляется полиномом $(n-1)$ - ой степени некоторой переменной x, причем коэффициентами полинома являются двоичные знаки соответствующих разрядов. Запись, чтение и передача кодовых комбинаций в циклическом коде производится начиная со старшего разряда. В соответствии с этим правилом, в дальнейшем сами числа и соответствующие им полиномы будем записывать так, чтобы старший разряд оказывался справа.

0 1 2 3 4 5

Пример 7. Число 110101 (нумерация разрядов, согласно выше приведенного правила, ведется слева направо от 0 до 5) будет представлено полиномом пятой степени: $1 + x + x^3 + x^5$.

Следует отметить, что циклическая перестановка разрядов в двоичном представлении числа соответствует умножению полинома на x, при котором x^n
Заменяется 1 и переходит в начало полинома.

Пример 8. Осуществим умножение полинома, полученного в предыдущем примере на x полученный полином $x + x^2 + x^4 + x^6$, преобразуем, заменив x^6 на 1. Окончательно получим $1 + x + x^2 + x^4$ соответствует числу 111010.

Циклический код n-значного числа, как всякий систематический код, состоит их m информационных и k контрольных знаков, причем последние занимают k младших разрядов. Так как последовательная передача кодовых комбинаций производится, как мы уже указывали, начиная со старших разрядов, контрольные знаки передаются в конце кода.

Образование кода осуществляется с помощью так называемого *порождающего полинома*, $P(x)$, степени k и именно видом этого полинома определяются все свойства кода - избыточность и корректирующая способность.

Кодовым полиномом , $F(x)$, является полином степени меньшей $(m+k)$, если он делится без остатка на порождающий полином $P(x)$. После передачи сообщения, декодирование состоит в выполнении деления полинома $H(x)$, соответствующего принятому коду на $P(x)$. При отсутствии ошибок $H(x)=F(x)$

И деление выполняется без остатка. Наличие ненулевого остатка указывает на то, что при передаче или хранении произошли искажения информации.

Для получения систематического циклического кода используется следующее соотношение:

$$F(x) = x^k G(x) \oplus R(x), \qquad (*)$$

где $G(x)$ - полином, представляющий информационные символы *(информационный полином),*

$R(x)$ - остаток от деления $x^k G(x)$ на $P(x)$.

Пример 9. Рассмотрим кодирование 8-значного числа 10110111.

Пусть для кодирования задан порождающий полином 3-ей степени

$P(x) = 1 + x + x^3$

Делим $x^3 G(x)$ на $P(x)$, $G(x) = 1 + x^2 + x^3 + x^5 + x^6 + x^7$

$x^3 G(x) = x^3 + x^5 + x^6 + x^8 + x^9 + x^{10}$

$$\begin{array}{l|l} x^{10} + x^9 + x^8 + x^6 + x^5 + x^3 & 1 + x + x^3 \\ \underline{x^{10} + x^8 + x^7} & \hline x^7 + x^6 + x^2 \\ x^9 + x^7 + x^6 + x^5 + x^3 & \\ \underline{x^9 + x^7 + x^6} & \\ x^5 + x^3 & \\ \underline{x^5 + x^3 + x^2} & \\ R(x) = x^2 & \end{array}$$

Используя соотношение *(*)* находим $F(x)$

$$F(x) = (x^3 + x^5 + x^6 + x^8 + x^9 + x^{10}) \oplus x^2$$

Таким образом окончательно кодовая комбинация

Соответствующая $F(x)$ имеет вид : 00010110111

001

контрольные разряды ⟶ 001 10110111

Практически, применяемая процедура кодирования еще более проста, так как нас интересует только остаток, а частное в конечном результате не используется, то можно производить последовательное вычитание по модулю 2 делителя из делимого и полученных разностей до тех пор, пока разность не будет иметь более низкую степень, чем делитель. Эта разность и есть остаток. Такой алгоритм может быть реализован аппаратно с помощью k - разрядного сдвигающего регистра, имеющего

обратные связи. Очевидно, что полученный таким способом циклический код будет являться системтическим.

Однако существует и второй вариант получения циклического кода, когда очередная кодовая комбинация получается путем умножения кодовой комбинации *C(x)* простого *n*-значного кода на образующий полином *P(x)*.

При втором способе образования циклических кодов информационные и контрольные символы в комбинациях циклического кода не отделены друг от друга, что затрудняет процесс декодирования. Поэтому этот способ кодирования применяется реже, чем первый. Рассмотрим пример кодирования с использованием второго варианта, причем при выполнении операций будем использовать непосредственно записи исходных кодовых комбинаций в двоичном виде.

Пример 10. Дан порождающий полином вида *P(1,0)=1101*. Требуется построить циклический код из простого четырехзначного кода вторым способом. В качестве примера для построения используем исходную комбинацию *C(1,0)=0011*. Операция умножения этой комбинации на образующий полином запишется следующим образом:

```
  0011
  1101
 ------

  0011
 0011
0011
-------
```

0010111 - это и есть циклический код для 0011

Простой четырехсимвольный код, C(x)	*Циклический (7,4)-код C(x) P(x), где P(1,0)=1101*
0000	*0000000*
0001	*0001101*
0010	*0011010*
*0011 **	*0010111 **
0100	*0110100*
0101	*0111001*
0110	*0101110*
0111	*0100011*
1000	*1101000*
1001	*1100101*
1010	*1110010*
1011	*1111111*
1100	*1011100*
1101	*1010001*
1110	*1000110*
1111	*1001011*

Рассмотренные два варианта построения циклических кодов для простоты реализации могут использовать так называемое матричное представление [Березюк 78, с.169-184][2]

Проблема обнаружения различного типа ошибок с помощью циклического кода, как уже указывалось, сводится к нахождению нужного порождающего полинома. В наши задачи не входит рассмотрение этого вопроса, достаточно полно основные принципы подбора порождающих полиномов изложены в [Березюк78, с.174-184].

2. Порядок выполнения работы и требования к отчету

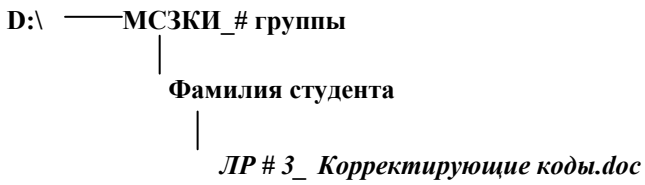
Цель лабораторной работы

Ознакомление с общими принципами построения и использования корректирующих кодов для контроля достоверности информации, распространяемой по телекоммуникационным каналам.

Для выполнения лабораторной работы на компьютере необходимо установить файл *Код Хэмминга.rar* и файл *Циклический код.rar*

Порядок выполнение работы

1. На диске *D:* создать дерево директорий следующего вида:



2. Скопировать файл ***Код Хэмминга.rar*** в личную папку. Ознакомиться со сведениями о программе кодирования информации с использованием кода Хэмминга (папка **Документация**). Запустить программу кодирования (папка **Исполняемый модуль**, файл ***Hamming.exe***).

3. Закодировать с помощью кода Хэмминга предложенный алфавит (варианты указаны в *приложении 1*). Кодовую таблицу сохранить в файле *ЛР # 3_ Корректирующие коды.doc*.

4. В каждую строку таблицы с закодированной информацией внести одиночную ошибку, зафиксировать в кодовой таблице результат дешифрования.

---

[2] [Березюк78] Березюк Н.Т., Андрущенко А.Г., Мощинский С.С. и др. Кодирование информации (двоичные коды). - Харьков, “Вища школа”, 1978.

5. В последние две строки таблицы с закодированной информацией внести двойные ошибки, зафиксировать в кодовой таблице результат дешифрования.

6. В первые две строки таблицы с закодированной информацией внести тройные ошибки, зафиксировать к кодовой таблице результат дешифрования.

7. Проанализировать полученные результаты и сформулировать аргументированные выводы.
   Описать полученный код Хэмминга :
   - количество контрольных и информационных разрядов и их номера;
   - избыточность кода;
   - относительная избыточность;
   - минимальное кодовое расстояние.
   - оценить корректирующую способность полученного кода.
8. Составить из предложенного алфавита слово длиной не менее 5 символов и закодировать его с помощью полученного кода Хэмминга. Подсчитать длину исходного текста (кодировка *ASCII*) и закодированного текста (код Хэмминга). Сделать выводы.

Примечание:
Наибольшее распространение для представления символьных данных получил американский стандартный код для информационного обмена - ***ASCII*** введён США 1963г. Согласно этому стандарту каждому символу поставлено в соответствие число от 0 до 255. Символы от 0 до 127 – латинские буквы, цифры и знаки препинания – составляют постоянную часть таблицы. Остальные символы используются для представления национальных алфавитов. Конкретный состав этих символов определяется кодовой страницей.

Пример
ASCII-код символа A = $65_{10} = 41_{16} = 01000001_2$;
ASCII-код символа G = $71_{10} = 47_{16} = 01000111_2$;
ASCII-код символа Z = $90_{10} = 5A_{16} = 01011010_2$.
ASCII-код символа C = $67_{10} = 43_{16} = 01000011_2$

A	C	C	A
01000001	01000011	01000001	01000011
<1 байт>	<1 байт>	<1 байт>	<1 байт>

9. Скопировать файл ***Циклический код.rar*** в личную папку. Ознакомиться со сведениями о программе кодирования информации с использованием циклического кода (папка **Документация**). Запустить программу кодирования (папка **Исполняемый модуль**, файл ***CyclicCode74.exe***). Проделать п.п.3-8 для циклического кода.

10. Добавить к отчету о выполнении лабораторной работы титульный лист, содержащий:
- название университета,
- факультета (*Социологии и информационных технологий*),
- кафедры (*Информационной безопасности*),
- учебной дисциплины,
- номер и название лабораторной работы,
- фамилию и инициалы студента,

- город и год выполнения лабораторной работы.

Пронумеровать страницы отчета.

11. Сохранить отчет о выполнении лабораторной работы в папке, созданной при выполнении п. 1.
12. Завершить работу с ОС *Windows*.

Приложение 1

Номера вариантов	**Исходный алфавит**
1,5,9,13,17	Кириллица А...Р
2,6,10,14,18,22	Кириллица О..Я
3,7,11,15,19,23,27	Латиница A..N
4,8,12,16,20,24,28	Латиница O..Z
21,25,29,26,30	Десятичные цифры 0..9