Министерство образования и науки Российской Федерации

НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

004
О-753

№ 3315

ОСНОВЫ ТЕОРИИ ИНФОРМАЦИИ И КРИПТОГРАФИИ

Часть 1

Методические указания
к выполнению лабораторных работ
для студентов III курса ФПМИ по специальностям
«Математическое обеспечение и администрирование
информационных систем» (010503),
«Прикладная информатика в менеджменте» (080801)

НОВОСИБИРСК
2007

УДК 004.056.55
О-753

Составитель *Т. А. Гультяева*

Рецензент *Н. Л. Долозов*, канд. техн. наук, доц.

Работа подготовлена на кафедре программных систем и баз данных

СОДЕРЖАНИЕ

Лабораторная работа 1

РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ ПО ТЕМЕ «ОСНОВНЫЕ АСПЕКТЫ ТЕОРИИ ИНФОРМАЦИИ»

Цель работы – приобретение навыка решения практических задач, отражающих основные свойства источников дискретных сообщений.

Указание к работе - ознакомиться с лекционным материалом, а также с литературой [1 – 7].

Задание

В зависимости от номера бригады выбрать задачи и решить их. Общее число задач предполагается равным трем. Отчет должен включать постановку задач и решение с подробным отображением этапов. Вычисления производить при помощи, например, программы MathCad. Третья задача для всех вариантов – 25.

Вариант	Ззадача 1	Задача 2	**Вариант**	Задача 1	Задача 2
1	11	13	**9**	4	22
2	2	14	**10**	7	16
3	5	15	**11**	8	18
4	6	20	**12**	5	19
5	10	12	**13**	10	23
6	9	20	**13**	10	23
7	8	21	**14**	1	17
8	3	19	**15**	5	18

1. На шахматной доске произвольным образом расставлены фигуры. Априори все положения фигур на доске одинаково вероятны. Определить собственную информацию, получаемую от сообщения, что фигура находится в одной из угловых клеток доски.

2. Сколько информации содержится в сообщении о том, что сумма очков на двух подброшенных игральных костях есть четное число?

3. Сколько информации содержится в сообщении о том, что сумма очков на двух подброшенных игральных костях равна 7?

4. Бросаются одновременно две игральные кости. Определить количество информации, содержащееся в сообщении о том, что произведение чисел выпавших очков четно.

5. В цехе работают семь мужчин и три женщины. По табельным номерам наудачу отобраны три человека. Найти количество информации, содержащееся в сообщении: «Все отобранные люди – мужчины».

6. Из шести букв разрезной азбуки составлено слово «машина». Ребенок, не умеющий читать, рассыпал эти буквы и затем собрал в произвольном порядке. Какое количество информации будет содержаться в утверждении, что у него снова получилось слово «машина»?

7. По дискретному каналу передаются сообщения x_1, x_2, x_3. Вследствие действия шумов на выходе канала появляются сигналы y_1, y_2. Вероятности совместного появления заданы в таблице.

$p(x_i, y_k)$	y_k	
x_j	y_1	y_2
x_1	0.4	0.1
x_2	0.2	0.15
x_3	0.1	0.05

Вычислить взаимные информации $I(x_1; y_2)$, $I(x_3; y_1)$, $I(x_2; y_2)$.

8. По двоичному каналу с шумом передаются сообщения x_1, x_2, x_3 с вероятностями 0.2, 0.3 и 0.5. На выходе канала появляются сигналы y_1, y_2, y_3. Вероятности искажения в канале (условные вероятности переходов) $P(y_i \mid x_j)$ даны в таблице.

$P(y_i \mid x_j)$	x_1	x_2	x_3
y_1	3/4	1/8	1/8
y_2	1/8	3/4	1/8
y_3	1/8	1/8	3/4

Вычислить взаимные информации $I(x_1; y_3)$ и $I(x_3; y_1)$.

9. По двоичному каналу с помехами передаются равновероятные и статистически независимые сообщения x_1 и x_2. В результате действия помех они преобразуются в сигналы y_1, y_2, y_3. Условные вероятности переходов $p(y_i \mid x_j)$ заданы в таблице.

$p(y_i \mid x_j)$		y_i		
		y_1	y_2	y_3
x_j	x_1	5/8	2/8	1/8
	x_2	1/8	5/8	2/8

Вычислить взаимные информации $I(x_1; y_3)$ и $I(x_2; y_2)$.

10. Рассматривается ансамбль сообщений, приведенный в таблице.

x_j	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7
$P(x_j)$	1/4	1/4	1/8	1/8	1/16	1/16	1/16	1/16
Код	000	001	010	011	100	101	110	111

Сообщение x_2 поступает в кодер. Вычислить дополнительную информацию об этом сообщении, доставляемую каждым последующим символом на выходе кодера.

11. Сообщения источника x_1, x_2, x_3, x_4 для согласования с каналом кодируются в соответствии с таблицей.

Сообщение x_i	x_1	x_2	x_3	x_4
$p(x_i)$	1/2	1/4	1/8	1/8
Код	000	011	101	100

Пусть на вход кодера поступает сообщение x_3. Вычислить дополнительную информацию об этом сообщении, которую содержит каждый последующий символ на выходе кодера.

12. Алфавит русского языка состоит из 32 букв (если не различать е и ё, ь и ъ), включая промежуток между буквами. Вычислить энтропию однобуквенного текста, считая вероятности появления любой из букв в заданном тексте одинаковыми.

13. Источник вырабатывает ансамбль сообщений

$$X = \begin{Bmatrix} x_1 & x_2 & x_3 & x_4 \\ 0.2 & 0.3 & 0.4 & 0.1 \end{Bmatrix}.$$

Символы в последовательности независимы. Вычислить энтропию источника и определить избыточность.

14. Найти число значений m равномерно распределенной случайной величины Y, при котором ее энтропия будет равна энтропии случайной величины X, заданной в таблице:

j	1	2	3	4	5	6	7	8
$p(x_j)$	1/2	1/4	1/8	1/16	1/32	1/64	1/128	1/128

15. Алфавит источника сообщений состоит из двух букв – x_1 и x_2 с вероятностями 0.6 и 0.4. В последовательности на выходе источника символы статистически зависимы. Условные вероятности переходов $p(x_j^2 \mid x_i^1)$ заданы таблицей.

i – индекс предыдущей буквы	j – индекс последующей буквы	
	1	2
1	0.2	0.8
2	0.7	0.3

Определить энтропию на один символ текста $X^{(1)}X^{(2)}$. Вычислить избыточность источника.

16. Выполнить задачу 15, если условные вероятности переходов $p(x_j^2 \mid x_i^1)$ заданы таблицей.

i – индекс предыдущей буквы	j – индекс последующей буквы			
	1	2	3	4
1	0.0	0.2	0.4	0.4
2	0.2	0.2	0.3	0.3
3	0.25	0.0	0.25	0.5
4	0.2	0.4	0.4	0.0

Безусловные вероятности букв x_1, x_2, x_3, x_4 равны соответственно 0.5; 0.25; 0.125; 0.125.

17. Символы азбуки Морзе могут появиться в сообщении с вероятностями: для точки – 0.51, для тире – 0.31, для промежутка между буквами – 0.12, между словами – 0.06. Определить среднее количество информации в сообщении из 500 символов данного алфавита, считая, что связь между последовательными символами отсутствует.

18. Система радиозонда измеряет давление. Барометр имеет 10 отметок шкалы, и его отсчеты могут изменяться до любого допустимого значения за 0.01 с. Связь между отсчетами отсутствует. Найти энтропию источника за 1 с, если показания барометра будут появляться с вероятностями, представленными ниже.

Отметка шкалы	0	1	2	3	4	5	6	7	8	9
Вероятность	0.05	0.05	0.05	0.05	0.1	0.2	0.3	0.1	0.05	0.05

19. Измеряемое напряжение лежит в пределах от 0 до 6 В. Телеметрический датчик регистрирует приращение напряжения Δu = 0.01 В. Найти наибольшее среднее количество информации, получаемое за 10 независимых отсчетов.

20. На радиорелейной линии для передачи символов 0 и 1 используются радиоимпульсы мощностью 1 и 10 Вт соответственно. Для нормальной работы передатчика необходимо, чтобы отдаваемая им средняя мощность не превышала 3.4 Вт. Найти наибольшее количество информации, которое в среднем переносит один импульс.

21. По дискретному каналу передаются сообщения x_1 и x_2. Вследствие шумов на выходе канала появляются сигналы y_1, y_2, y_3. Вероятности их совместного появления $P(x_i; y_j)$ представлены ниже.

$P(x_i; y_j)$	y_1	y_2	y_3
x_1	1/4	1/16	1/8
x_2	1/8	3/16	1/4

Необходимо найти взаимную информацию $I(X; y_1)$.

22. Задание аналогично 21, но найти $I(x_1; Y)$.

23. Задание аналогично 21, но найти $I(X; Y)$.

24. Сообщение X есть стационарная последовательность независимых символов, имеющих ряд распределения, приведенный ниже.

x_j	x_1	x_2	x_3
$p(x_j)$	0.2	0.1	0.7

Сигнал Y является последовательностью двоичных символов, связанных с сообщением X по следующему правилу:

$$x_1 \to 00,\ x_2 \to 00,\ x_3 \to 1.$$

Найти средние безусловную и условную энтропии сообщения X.

25. Задание аналогично 24.

Вероятности появления символов сообщения X

Вариант	$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$	$p(x_5)$
1	0.1	0.2	0.2	0.2	0.3
2	0.2	0.1	0.2	0.1	0.4
3	0.1	0.3	0.2	0.1	0.3
4	0.2	0.1	0.3	0.2	0.2
5	0.2	0.3	0.1	0.2	0.2
6	0.1	0.1	0.4	0.3	0.1
7	0.1	0.2	0.3	0.1	0.3
8	0.1	0.3	0.2	0.3	0.1
9	0.2	0.1	0.3	0.2	0.2
10	0.1	0.2	0.3	0.2	0.2
11	0.1	0.2	0.4	0.1	0.2
12	0.2	0.2	0.2	0.1	0.3
13	0.2	0.2	0.2	0.0	0.4
14	0.2	0.2	0.2	0.4	0.0
15	0.2	0.2	0.4	0.2	0.0

Сигнал Y является последовательностью двоичных символов, связанных с сообщением X по следующему правилу.

Вариант	$x_1 \rightarrow$	$x_2 \rightarrow$	$x_3 \rightarrow$	$x_4 \rightarrow$	$x_5 \rightarrow$
1	1	0	11	00	1
2	0	00	000	1	11
3	0	0	0	0	000
4	00	1	1	1	0
5	0	0	0	0	00
6	1	00	1	1	0
7	0	0	00	1	1
8	0	0000	0	0	1
9	1	1	0	0	1
10	1	00	000	1	0
11	00	1	111	0	1
12	00	10	1	0	1
13	000	1	111	0	1
14	00	1	10	0	0
15	0	1	1	0	1

Найти средние безусловную и условную энтропии сообщения X .

Контрольные вопросы

1. Доказать следующее свойство функционала энтропии:

$$H(\xi) \geq 0\,,\ H(\xi) = 0\,,\ \text{если } \xi \text{ неслучайно.}$$

2. Доказать аддитивность функционала энтропии.

3. Доказать, что если к алфавиту символов добавить символ с нулевой вероятностью, то энтропия ИДС не изменится.

4. Доказать свойство иерархической аддитивности функционала энтропии:

$$H(X_1, X_2, ..., X_n) = H(X_1) + H(X_2 \mid X_1) + ... + H(X_n \mid X_1, X_2, ..., X_{n-1})$$

5. Доказать свойство симметрии функционала средней взаимной информации.

6. Доказать, что $0 \leq \hat{I}(X,Y) \leq \min(H(X), H(Y))$.

7. Доказать, что равенство нулю в неравенстве $0 \leq \hat{I}(X,Y) \leq \min(H(X), H(Y))$ возможно тогда и только тогда, когда X, Y статистически независимы.

8. Доказать, что $\hat{I}(X,(Y_1,Y_2)) = \hat{I}(X,Y_1) + \hat{I}(X,Y_2)$, если Y_1, Y_2 независимы.

Лабораторная работа 2

ОСНОВНЫЕ МЕТОДЫ ПОБУКВЕННОГО КОДИРОВАНИЯ

Цель работы – освоить основные алгоритмы побуквенного кодирования.

Указание к работе – ознакомиться с лекционным материалом, а также с литературой [5 – 7].

Дискретная форма представления информации является наиболее общей и универсальной. В виде совокупности символов, принадлежащих к ограниченному алфавиту, можно представить как текст или массивы чисел, так и оцифрованные звук и изображение. С учетом этого

очевидно, что должны существовать универсальные методы сжатия данных (цифровой информации), применимые ко всем ее разновидностям. В силу своей универсальности эти методы должны исключать потерю информации (такая потеря может быть допустима при передаче, например, мелкой детали изображения, но неприемлема, когда речь идет, скажем, о коде программы). С другой стороны, в ряде приложений общие методы наверняка не будут наиболее эффективными. Например, в силу особенностей зрительного и слухового восприятия некоторое «огрубление» изображения или звука может оказаться малозаметным, при этом выигрыш в объеме передаваемых данных окажется значительным. В этих случаях уместно использовать специальные методы сжатия с потерями (см. рисунок).

Метод сжатия информации

При кодировании со сжатием без потерь выделяются две разновидности методов. Первая основана на раздельном кодировании символов. Основная идея состоит в том, что символы разных типов встречаются неодинаково часто и если кодировать их неравномерно, так, чтобы короткие битовые последовательности соответствовали часто встречающимся символам, то в среднем объем кода будет меньше. Такой подход, именуемый статистическим кодированием, реализован, в частности, в широко распространенном коде Хаффмана. Именно коды таких типов и рассматриваются в данной лабораторной работе. Все

остальные типы кодирования представлены несколькими вариантами в курсовом проекте.

Очевидно, что посимвольное кодирование не использует такого важного резерва сжатия данных, как учет повторяемости последовательностей (цепочек) символов. Простейший вариант учета цепочек – так называемое «кодирование повторов» или код RLE, когда последовательность одинаковых символов заменяется парой – «код символа + количество его повторов в цепочке». В большинстве случаев цепочки одинаковых символов встречаются нечасто. Однако, например, при кодировании черно-белых растровых изображений, каждая строка которых состоит из последовательных черных или белых точек, такой подход оказывается весьма эффективным (он широко применяется при факсимильной передаче документов). Кроме того, кодирование повторов нередко используется как составной элемент более сложных алгоритмов сжатия. Гораздо более универсален алгоритм, позволяющий эффективно кодировать повторяющиеся цепочки разных символов, имеющие произвольную длину. Такой алгоритм разработан Лемпелем и Зивом, и он применяется в разных версиях в большинстве современных программ-архиваторов. Идея алгоритма состоит в том, что цепочка символов, уже встречавшаяся в передаваемом сообщении, кодируется ссылкой на боле раннюю (при этом указываются «адрес» начала такой цепочки в «словаре» сообщения и ее длина).

Специализированные методы сжатия с потерями информации, естественно принципиально различаются для графики и звука.

К методам сжатия изображений относятся «блочный» алгоритм JPEG, основанный на независимом «огрублении» небольших фрагментов изображений (квадраты 8×8 пикселей). Здесь с ростом степени сжатия проявляется мозаичность изображения. Блочный метод JPEG (разработанный специальной группой международного комитета по стандартизации) получил повсеместное распространение. Достигаемая степень сжатия – в среднем в десятки раз. При волновом сжатии в отличие от блочного изображение как бы «размывается» (чем выше степень сжатия, тем более нечетки границы и детали). При передаче данных получаемое изображение постепенно «проявляется» в деталях. Это позволяет получателю самому выбирать необходимый компромисс между качеством и скоростью получения изображения, что очень

удобно, например, в сети Интернет. К тому же «размытость» не столь резко воспринимается глазом, как потеря качества по сравнению с «мозаичностью». Так что при субъективно близком уровне качества волновой метод дает бóльшую степень сжатия по сравнению с «блочным». Именно такой подход реализован в новом стандарте JPEG 2000. Наконец, фрактальное сжатие основывается на том, что в изображении можно выделить фрагменты, повороты и масштабирование которых позволяет многократно использовать их при построении всей «картинки». Выделение и построение математического описания таких элементов-фракталов – трудоемкая в вычислительном отношении задача. Зато высокая степень сжатия (в сотни раз) и быстрота построения изображения по его фрактальному описанию делают метод очень удобным, когда не требуется быстрота компрессии. Например, этот метод удобно использовать при записи изображений на CD-ROM.

Наконец, методы сжатия звука существенно различаются в зависимости от того, насколько хорошо известны специфические особенности его источника. Примерами источников, чьи особенности решающим образом влияют на характер звука, являются человеческий речевой аппарат и музыкальные инструменты. Для них эффективным способом сжатия звуковой информации является моделирование, когда передаются параметры модели источника, а не характеристика звука.

Задание

1. Реализовать приложение, позволяющее кодировать произвольные сообщения с заданным алфавитом по алгоритму, приведенному в варианте. В программе должна быть также предусмотрена возможность декодирования полученного кодового сообщения. Алфавит приемника $B = \{0;1\}$. Обратите внимание на встречаемость символов в сообщении: оно должно зависеть от распределения вероятности букв.

2. Рассмотреть три возможных распределения вероятностей букв алфавита: равномерное и приведенные в варианте; сравнить полученные коды (вычислить среднюю длину кодовых слов, избыточность кода, проверить неравенство Крафта); сделать выводы по алгоритму.

Вариант	Алфавит источника A	Вероятность появления букв	Алгоритм
		$P_1(A)$	
		$P_2(A)$	
1	$\{A, B, C, D, E, F, 1, 2\}$	$\{0.9, 0.01, 0.01, 0.02, 0.05, 0.01, 0, 0\}$	Хаффмана
		$\{2^{-5}, 2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}, 2^{-5}, 0, 0\}$	
2	$\{2, 20, 200, 0, 00, '\_'\}$	$\{0.4, 0.1, 0.1, 0.05, 0.05, 0.3\}$	Шеннона
		$\{2^{-6}, 2^{-1}, 2^{-2}, 2^{-3}, 2^{-5}, \frac{5}{64}\}$	
3	$\{+, -, *, /, =\}$	$\{0.2, 0.3, 0.1, 0.1, 0.3\}$	Гильбера–Мура
		$\{\frac{3}{32}, 2^{-1}, 2^{-2}, 2^{-3}, 2^{-5}\}$	
4	$\{a, b, c, d\}$	$\{0.01, 0.1, 0.09, 0.8\}$	Шеннона–Фано
		$\{2^{-2}, 2^{-3}, 2^{-1}, 2^{-3}\}$	
5	$\{$а, б, в, г, д, е, ж, з$\}$	$\{0.5, 0.01, 0.09, 0.02, 0.08, 0.2, 0.07, 0.03\}$	Хаффмана
		$\{2^{-4}, 2^{-5}, 2^{-6}, \frac{17}{128}, 2^{-7}, 2^{-2}, 2^{-2}, 2^{-2}\}$	
6	$\{4, 14, 5, 1, 0, 2, 8, '\_'\}$	$\{0.01, 0.1, 0.09, 0.1, 0, 0, 0, 0.7\}$	Шеннона–Фано
		$\{2^{-2}, 2^{-5}, \frac{11}{128}, 2^{-1}, 0, 0, 0, 2^{-3}\}$	
7	$\{$А, Б, В, Г, Д, Е, Ж, З$\}$	$\{0.1, 0.01, 0.09, 0.4, 0.4, 0, 0, 0\}$	Хаффмана
		$\{2^{-2}, 2^{-5}, 2^{-1}, \frac{11}{128}, 2^{-3}, 0, 0, 0\}$	

Окончание табл.

Вариант	Алфавит источника A	Вероятность появления букв	Алгоритм
		$P_1(A)$	
		$P_2(A)$	
8	$\left\{\begin{matrix}11,12,13,14,\\15,16,17\end{matrix}\right\}$	$\{0.5, 0.01, 0.09, 0.02, 0.08, 0.2, 0.1\}$	Гильбера–Мура
		$\left\{2^{-3}, 2^{-4}, 2^{-2}, 2^{-5}, 2^{-7}, 2^{-1}, \frac{3}{128}\right\}$	
9	$\{a, b, d, g, \_\}$	$\{0.1, 0.1, 0.1, 0.4, 0.3\}$	Шеннона
		$\left\{2^{-3}, 2^{-3}, 2^{-1}, 0, 2^{-2}\right\}$	
10	$\{1, 2, 3, 4, 5, 6, 7, 8\}$	$\{0.5, 0.01, 0.09, 0.3, 0.1, 0, 0, 0\}$	Хаффмана
		$\left\{2^{-5}, 2^{-3}, 2^{-1}, \frac{3}{32}, 2^{-2}, 0, 0,\right\}$	
11	$\{a, b, c, d, e, f\}$	$\{0.09, 0.01, 0.1, 0.2, 0.5, 0.1\}$	Гильбера–Мура
		$\left\{2^{-5}, 2^{-1}, 2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}\right\}$	
12	$\{-, 8, 3, =, a\}$	$\{0.88, 0.02, 0.01, 0.05, 0.01, 0.03\}$	Шеннона
		$\left\{2^{-1}, 2^{-5}, 2^{-5}, 2^{-2}, 2^{-3}, 2^{-4}\right\}$	
13	$\{z, w, x\}$	$\{0.01, 0.9, 0.09\}$	Гильбера–Мура
		$\left\{2^{-1}, 2^{-2}, 2^{-2}\right\}$	
14	$\{A, H, z, w\}$	$\{0.35, 0.35, 0.1, 0.1, 0.1\}$	Шеннона–Фано
		$\left\{2^{-5}, 2^{-1}, 2^{-5}, 2^{-4}, \frac{3}{8}\right\}$	
15	$\left\{\begin{matrix}2, 4, 6, 8,\\0, а, б, в\end{matrix}\right\}$	$\{0.4, 0.4, 0.1, 0.05, 0.025, 0, 0.025, 0\}$	Хаффмана
		$\left\{2^{-3}, 2^{-3}, 2^{-1}, 0, 2^{-2}\right\}$	

Контрольные вопросы

1. Алгоритм Хаффмана.
2. Алгоритм Шеннона.
3. Алгоритм Шеннона–Фано.
4. Алгоритм Гильбера–Мура.
5. Алгоритм перевода дробных чисел в двоичную систему счисления.
6. Сжатие без потерь информации.
7. Сжатие с потерей информации.
8. Теорема Шеннона.
9. Оптимальное побуквенное кодирование.
10. Неравенство Крафта.

Лабораторная работа 3

ПОМЕХОУСТОЙЧИВОЕ КОДИРОВАНИЕ

Цель работы – освоить основные алгоритмы помехоустойчивого кодирования.

Указание к работе – ознакомиться с лекционным материалом, а также с литературой [6 – 8].

Задания

1. Реализовать приложение, включающая кодер и декодер с проверкой на четность. В качестве кодирующих преобразований $[x_1, x_2, ..., x_n] \to [y_1, y_2, ..., y_s]$ использовать код, разработанный для своего варианта в лабораторной работе 2 (в случае равновероятных появлений символов алфавита).

На вход кодера подается сообщение $X = [x_1, x_2, ..., x_n]$, получается сообщение $[y_1, y_2, ..., y_{m-1}, \hat{y}_m, y_{m+1}, y_{m+2}, ..., y_{k-1}, \hat{y}_k, ..., \hat{y}_s]$. При этом биты $\hat{y}_m, ..., \hat{y}_s$ являются проверочными. Если длина кодового слова четна, то в окончательном виде оно будет иметь вид

$[y_1, y_2, y_1 \oplus y_2, y_3, y_4, y_3 \oplus y_4, ...]$ – например, в вариантах № 3, 4, 8, 11, 14; иначе: $[y_1, y_2, y_3, y_1 \oplus y_2 \oplus y_3, y_4, y_5, y_6, y_4 \oplus y_5 \oplus y_6, ...]$.

На вход декодера подаются как искаженные кодовые слова, так и неизмененные. Результатом работы приложения должно быть сообщение об обнаруженных ошибках с указанием возможных мест данных ошибок.

Исследовать созданный код на кратность обнаружения и исправления ошибок. Найти кодовое расстояние кода d_0, расстояние Хэмминга, границу Хэмминга, границу Плоткина, граница Варшамова–Гильберта.

2. Реализовать приложение, включающее кодер и декодер, работающие по алгоритму кодирования Хэмминга. Конкретный вид проверочной или порождающей матрицы задан в таблице. При этом считать, что алфавит источника $A = [x \mid x \in Z]$ и $0 \to 00000, 1 \to 00001, 2 \to$ $\to 00010, ..., 24 \to 11111$ для 1-го варианта, для 2-го $0 \to 0000, 1 \to 0001, 2 \to$ $\to 0010, ..., 15 \to 1111$ и т. п.

Схема работы приложения представлена на рисунке.

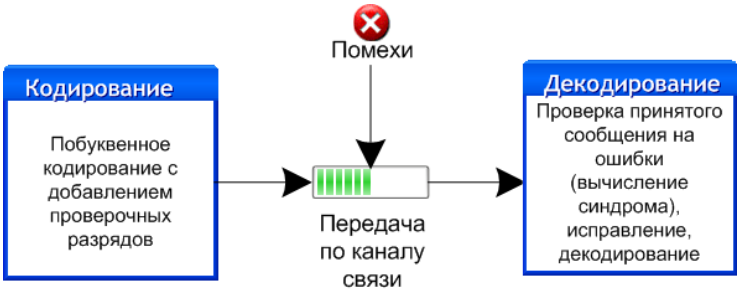

Схема кодирования и декодирования по Хэммингу

Построить таблицу синдромов, применяемую далее в процессе декодирования. Исследовать созданный код на кратность обнаружения и исправления ошибок. Найти кодовое расстояние кода d_0, расстояние Хэмминга, границу Хэмминга, границу Плоткина, граница Варшамова–Гильберта.

Вариант	Проверочная или порождающая матрица (n,k) кода Хэмминга
1	$H_{(9,5)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$
2	$G_{(7,4)} = \left[\begin{array}{cccc\|ccc} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{array}\right]$
3	$H_{(7,4)} = \left[\begin{array}{cccc\|ccc} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$
4	$G_{(7,4)} = \left[\begin{array}{cccc\|ccc} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{array}\right]$
5	$H_{(9,5)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}$
6	$H_{(9,5)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}$
7	$G_{(8,4)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}$

Окончание табл.

Вариант	Проверочная или порождающая матрица (n,k) кода Хэмминга
8	$G_{(9,5)} = \left[\begin{array}{ccccc\|cccc} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{array}\right]$
9	$G_{(7,2)} = \left[\begin{array}{cccccc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right]$
10	$G_{(9,2)} = \left[\begin{array}{ccccccc} 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$
11	$G_{(12,4)} = \left[\begin{array}{cccccccccccc} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{array}\right]$
12	$G_{(12,2)} = \left[\begin{array}{cccccccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$
13	$G_{(7,4)} = \left[\begin{array}{cccc\|ccc} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{array}\right]$
14	$H_{(7,4)} = \left[\begin{array}{cccc\|ccc} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$
15	$G_{(9,5)} = \left[\begin{array}{ccccc\|cccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right]$

Контрольные вопросы

1. Помехоустойчивое кодирование. Теорема Шеннона.

2. Коды с обнаружением ошибок.

3. Коды с исправлением ошибок.

4. Построение таблицы синдромов. С какой целью это делается?

5. Найти порождающую и проверочные матрицы для циклического кода, заданного следующим порождающим многочленом:

1. $g(x) = x^2 + x + 1.$
2. $g(x) = x^3 + x + 1.$
3. $g(x) = x^4 + x + 1.$
4. $g(x) = x^5 + x^2 + 1.$
5. $g(x) = x^6 + x + 1.$
6. $g(x) = x^7 + x + 1.$
7. $g(x) = x^7 + x^3 + 1.$
8. $g(x) = x^8 + x^4 + x^3 + x^2 + 1.$
9. $g(x) = x^9 + x^4 + 1.$
10. $g(x) = x^{10} + x^3 + 1.$
11. $g(x) = x^{11} + x^2 + 1.$
12. $g(x) = x^{12} + x^6 + x^4 + x + 1.$
13. $g(x) = x^{13} + x^4 + x^3 + x + 1.$
14. $g(x) = x^{14} + x^5 + x^3 + x + 1.$
15. $g(x) = x^{15} + x + 1.$

Пару (n,k) подобрать самостоятельно. Проверить основные свойства линейных кодов, свойство линейности полученных кодов.

Лабораторная работа 4

ШИФРЫ ПЕРЕСТАНОВКИ И ЗАМЕНЫ

Цель работы – освоить основные алгоритмы шифрования перестановки и замены.

Указание к работе – ознакомиться с лекционным материалом, а также с литературой [1, 8 – 11].

4.1. Подстановочные шифры

Шифр простой подстановки. В таком шифре производится замена каждой буквы сообщения на некоторый определенный символ (обычно также на букву).

Таким образом, сообщение $M = m_1 m_2 ... m_n$, где $m_1 m_2 ... m_n$ – последовательные буквы, переходит в шифротекст $C = [c_1 c_2 ... c_n] = [\varphi(m_1)\varphi(m_2)...\varphi(m_n)]$, причем функция φ имеет обратную функцию. Ключ является просто перестановкой алфавита (если буквы заменяются на буквы).

Шифр Виженера и его варианты. В шифре Виженера ключ задается набором из T букв. Такие наборы подписываются с повторением под сообщением, и полученные две последовательности складываются по модулю, равному мощности алфавита исходного сообщения (каждая буква рассматриваемого алфавита нумеруется). Таким образом, $c_t = (m_t + k^{\tau}) \bmod |X|$, где $t = (i-1)(T+\tau), i = 1, 2...; \tau = \overline{1,T}$, $|X|$ – мощность алфавита сообщений. Обратное преобразование $m_t = (c_t - k^{\tau} + |X|) \bmod |X|$.

Шифр Виженера с периодом $T = 1$ называется *шифром Цезаря*[1]. Он представляет собой простую подстановку, в которой каждая буква сообщения M сдвигается вперед на фиксированное число мест по алфавиту.

---

[1] 1 век до н.э. Данный шифр описан историком Древнего Рима Светонием. Гай Юлий Цезарь использовал в своей переписке шифр собственного изобретения.

Так называемый *шифр Бофора* и видоизмененный шифр Бофора подобны шифру Виженера. В них сообщения зашифровываются с помощью равенств $c_i = (k_i - m_i) \bmod |X|$ и $c_i = (m_i - k_i) \bmod |X|$ соответственно.

Повторное применение двух или более шифров Виженера будет называться *составным шифром Виженера*. Он имеет уравнение $c_i = (m_i + k_i + ... + s_i) \bmod |X|$, где $k_i, ..., s_i$, вообще говоря, имеют различные периоды. Период их суммы $k_i + ... + s$ будет наименьшим общим кратным отдельных периодов.

Если используется шифр Виженера с неограниченным неповторяющимся ключом, то мы имеем *шифр Вернама*, в котором k_i выбираются случайно и независимо среди чисел $\overline{0, |X|}$. Если ключом служит текст, имеющий смысл, то имеем *шифр* «*бегущего ключа*».

Шифр Виженера с перемешанным один раз алфавитом представляет собой простую подстановку с последующим применением шифра Виженера: $c_i = \varphi(m_i) + k_i$, тогда $m_i = \varphi^{-1}(c_i - k_i)$.

Диграммная, триграммная и n-граммная подстановки. Вместо подстановки одной буквы можно использовать подстановку диграмм, триграмм и т.д. Для диграммной подстановки в общем виде требуется ключ, состоящий из перестановок $|X|^2$ диграмм. Он может быть представлен с помощью таблицы, в которой ряд соответствует первой букве диграммы, а столбец – второй букве, причем клетки таблицы заполнены заменяющими символами (обычно также диграммами).

Матричная система. Имеется один метод подстановки n-грамм, который заключается в применении к последовательным n-граммам некоторой матрицы, имеющей обратную. Предполагается, что буквы занумерованы от 0 до $|X| - 1$ и рассматриваются как элементы некоторого алгебраического кольца. Если к n-грамме сообщения применить матрицу A, то получится n-грамма криптограммы: $A[m_1, m_2, ..., m_n] = [c_1, c_2, ..., c_n]$.

Матрица A является ключом, и расшифровка выполняется с помощью обратной матрицы $[m_1, m_2, ..., m_n] = A^{-1}[c_1, c_2, ..., c_n]$. Обратная матрица будет существовать тогда и только тогда, когда определитель $|A|$ имеет обратный элемент в данном кольце.

4.2. Перестановочные шифры

Перестановка с фиксированным периодом T. В этом случае сообщение делится на группы символов длины T и к каждой группе применяется одна и та же перестановка. Эта перестановка является ключом; она может быть задана некоторой перестановкой первых T целых чисел.

Последовательное применение двух или более перестановок будет называться составной перестановкой. Если периоды этих перестановок равны $T_1,...,T_s$ то, очевидно, в результате получится перестановка периода T, где T – наименьшее общее кратное $T_1,...,T_s$.

Дробные шифры. В этих шифрах каждая буква сначала зашифровывается в две (или более) буквы или в два (или более) числа, затем полученные символы каким-либо способом перемешиваются (например, с помощью транспозиции), после чего их можно снова перевести в первоначальный алфавит. После того как полученный ряд чисел подвергнут некоторой перестановке, его можно снова разбить на пары чисел и перейти к буквам.

Пример. Необходимо предложить вариант шифра матричной системы (триграмма). Пусть алфавит сообщений $X = \{a,b,c,d,e,f\}$. Буква a кодируется 0, $b \to 1,..., f \to 5$. Надо подобрать такую матрицу A, чтобы $|A|^{-1}$ был равен одному из возможных значений 0...5:

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 4 & 3 & 1 \\ 4 & 1 & 2 \end{bmatrix}, \quad |A| = 1.$$

Сообщение для отправки: *badeff* . Делим его на блоки по три символа, кодируем: $M1 = [1\ 0\ 3]$, $M2 = [4\ \ 5\ \ 5]$. Каждый полученный вектор преобразуем в шифр: $M1^T A = [1\ 7\ \ 10]$, $M2^T \cdot A = [9\ \ 36\ \ 31]$. Передаем по каналу связи $C1, C2$; злоумышленник, не зная код матричного шифровании, пытается прочесть сообщение, используя, например

$$B = \begin{bmatrix} 3 & 4 & 0 \\ 3 & 3 & 1 \\ 4 & 3 & 2 \end{bmatrix}$$

(из достоверных источников ему все же стал известен определитель матрицы-кода), и получает $N1 = B^{-1}C1 = \begin{bmatrix} -13 & 10 & 16 \end{bmatrix}$: ничего не вышло, хотя кодировку $a \to 0, b \to 1, ..., f \to 5$ он все же знает! Тот же, кому адресовано наше сообщение, получив его, дешифрует: $M1 = A^{-1}C1 = \begin{bmatrix} 1 & 0 & 3 \end{bmatrix}$, $M2 = A^{-1}C2 = \begin{bmatrix} 4 & 5 & 5 \end{bmatrix}$ и декодирует: *badeff* .

Задания

1. Реализовать приложение, позволяющие шифровать и дешифровать сообщения согласно заданному алгоритму (см. таблицу с вариантами). Конкретную реализацию алгоритма предложить самим. Алфавит сообщений также задан в варианте, кроме того, во всех вариантах для разделения слов используется символ '_', который также надо включать в алфавит.

2. Набор тестов должен содержать не только шифрование различных сообщений, но и демонстрировать шифрование с различными ключами.

3. Отчет должен обязательно содержать несколько тестовых сообщений, зашифрованных и расшифрованных теоретически (например, при помощи программы MathCad) и практически (при помощи вашего приложения).

4. Проанализировать шифр на приведенные в лекции типы криптоаналитического вскрытия.

5. Проанализировать шифр с точки зрения совершенной криптостойкости, т.е. проверить выполнение условий теорем (о совершенной криптостойкости симметричной криптосистемы). Ответить на вопрос об идеальной стойкости данного шифра.

Вариант	Алгоритм шифрования	Алфавит сообщения	Дополнительная информация
1	Шифр Виженера	*A*...*Z*	Период $T = 5$
2	Шифр Бофора	*a*...*z*	–
3	Шифр Цезаря	а...я	–
4	Видоизмененный шифр Бофора	*A*...*Z*	–
5	Составной шифр Виженера	0...9	Периоды $T_1 = 5$, $T_2 = 2$
6	Шифр простой подстановки	А...Я	–
7	Шифр Вернама	•, –	–
8	Диграммная подстановка	*A*...*Z*	–
9	Матричная система	*a*...*h*	(диграммная подстановка)
10	Матричная система	а...к	(триграммная подстановка)
11	Перестановка с периодом	*A*...*Z*	Период $T = 6$
12	Дробные шифры	А...Я	–
13	Шифр Виженера	*a*...*Z*	Период $T = 2$
14	Шифр Бофора	*A*...*Z*	–
15	Шифр Виженера с перемешанным один раз алфавитом	а...я	–

Контрольные вопросы

1. Основные понятия криптографии.
2. Протоколы обмена ключами.
3. Основные аспекты криптоанализа.
4. Общая схема симметричных криптосистем.
5. Математические модели элементарных криптосистем.
6. Криптостойкость симметричных криптосистем.
7. Пессимистическое утверждение Шеннона (теорема).
8. Показатели криптостойкости.
9. Требования, предъявляемые к современным криптографическим системам.

СПИСОК ЛИТЕРАТУРЫ

1. *Богуш Р.П.* Элементы теории информации: Учеб.-метод. комплекс для студ. спец. 1-40 01 01 «Программное обеспечение информационных технологий» / Сост. и общ. ред. Р.П. Богуша. – Новополоцк: УО «ПГУ», 2006. – 160 с.

2. Математические и компьютерные основы криптологии: Учеб. пособие / Ю.С. Харин, В.И. Берник, Г.В. Матвеев, С.В. Агиевич. – Минск, 2003. – 382 с.

3. *Ломакин Д.В.* Прикладная теория информации: Конспект лекций. – Н. Новгород. – 62 с.

4. *Ожиганов А.А., Тарасюк М.В.* Передача данных по дискретным каналам: Учеб.-метод. пособие по дисциплине «Теория информации». – СПб.: СПбГИТМО (ТУ), 199. – 102.

5. *Шульгин В.И.* Основы теории передачи информации: Учеб пособие. – Ч. 1: Экономное кодирование. – Харьков: Нац. аэрокосм. ун-т, 2003. – 102 с.

6. *Потапов В.Н.* Теория информации. Кодирование дискретных вероятностных источников: Учеб. пособие. – Новосибирск: НГУ, 1999. – 71 с.

7. *Лидовский В.В.* Теория информации: Учеб. пособие – М.: Компания Спутник+, 2004. – 111 с.

8. *Шнайер Б.* Прикладная криптография. Протоколы, алгоритмы и исходные тексты на языке С. – Изд. 2-е: Пер. с англ. – М.: Триумф, 2002. – 816 с.

9. *Фомичев В.М.* Дискретная математика и криптология. – М.: Диалог-МИФИ, 2003. – 400 с.

10. *Шеннон К.* Теория связи в секретных системах: Лекции по теории информации и кибернетике. – М.: Изд-во иностр. лит., http://www.enlight.ru/crypto/articles/shannon/shann__i.htm

11. *Погорелов Б.А., Черемушкин А.В., Чечета С.И.* Об определении основных криптографических понятий // Матер. конф. «Математика и безопасность информационных технологий», МаБИТ, 23–24 октябоя 2003 г. – М.: Изд-во МГУ, 2003.

ПРИЛОЖЕНИЕ

Программный код, переводящий число $0 < x < 1$ из десятичной системы исчисления в двоичную, написанный в MathCad 11.0:

$x_0 := 0.35 \quad N := 15$

$$\text{funct}(N,x) := \left| \begin{array}{l} \text{for } i \in 0..N \\ \quad \left| \begin{array}{l} q_i \leftarrow \text{floor}(x_i \cdot 2) \\ x_{i+1} \leftarrow x_i \cdot 2 - \text{floor}(x_i \cdot 2) \end{array} \right. \\ q^T \end{array} \right.$$

$X := \text{funct}(N,x)$

$x_0 = 0.0101100110011101b$

$X =$

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	1	0	1	1	0	0	1	1	0	0	1	1	0	0	1

$$\sum_{i=0}^{N} 2^{-(i+1)} \cdot \left(X^T\right)_i = 0.349990844726563$$

ОСНОВЫ ТЕОРИИ ИНФОРМАЦИИ И КРИПТОГРАФИИ

Часть 1

Методические указания

Редактор *Т. П. Петроченко*
Технический редактор *Н. В. Гаврилова*
Компьютерная верстка *С. Н. Кондратенко*

Подписано в печать 30.03.2007. Формат 60 × 84 1/16. Бумага офсетная.
Тираж 100 экз. Уч.-изд. л. 1,62. Печ. л. 1,75. Изд. № 16. Заказ № .
Цена договорная.

Отпечатано в типографии
Новосибирского государственного технического университета
630092, г. Новосибирск, пр. К. Маркса, 20