КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

ФАКУЛЬТЕТ РАДІОФІЗИКИ ЕЛЕКТРОНІКИ ТА КОМП'ЮТЕРНИХ СИСТЕМ

Лабораторна робота №2

з курсу «Цифровий зв'язок»

# ЕФЕКТИВНЕ ЗАВАДОСТІЙКЕ КОДУВАННЯ ІНФОРМАЦІЇ В ЦИФРОВИХ КАНАЛАХ ЗВ'ЯЗКУ

Виконав

студент 4 курсу

кафедри квантової радіофізики

Лукашенко Валерія

КИЇВ

2016

Мета роботи: Вивчити методи ефективного кодування інформації. Вивчити можливості кодів щодо корекції помилок. Отримати практичні навички роботи з ефективними та завадостійкими кодами.

Хід роботи

1. Запустити на персональному комп'ютері програму "Ефективне кодування інформації" та дочекайтися завершення формування нових завдань.

2. Вибрати в головному меню пункт завдання "Префіксність". Виконати завдання згідно з вказівками у вікні програми.

| Символ | Код |
|---|---|
| T | 10 |
| A | 00 |
| Y | 110 |
| G | 010 |
| V | 011 |
| C | 1110 |
| L | 11111 |

Таб. 1 Префіксний код

Повідомлення: 111001110011011010

Декодоване повідомлення: 1110 011 10 011 011 010

C V T V V Q

| Символ | Код |
|---|---|
| T | 10 |
| A | 00 |
| Y | 110 |
| G | 011 |
| V | 1110 |
| C | 11101 |
| L | 11111 |

Таб. 2 Непрефіксний код

Новий код: 11101110111111110

Варіант 1 декодованого за непрефіксним кодом повідомлення:

1110 1110 11111 1110

V V L V

Варіант 2 декодованого за непрефіксним кодом повідомлення:

11101 110 11111 1110

C Y L V

3. Вибрати в головному меню пункт завдання "Арифметичне кодування". Виконати відповідне кодування та декодування запропонованих послідовностей.

Повідомлення: CDBADDBACB

Заносимо у допоміжну таблицю імовірності появи кожного символу у слові та відповідні межі, які займають символи у загальному інтервалі (0,1).

| Символ | Ймовірність | Нижня межа | Верхня межа |
|---|---|---|---|
| A | 0,2 | 0 | 0,2 |
| B | 0,3 | 0,2 | 0,5 |
| C | 0,2 | 0,5 | 0,7 |
| D | 0,3 | 0,7 | 1 |

Таб. 3 Допоміжна таблиця кодування

Інтервал символу «A» буде займати у вхідному потоці від 0 до 20%, «B» відповідно від 20% до 50% і т.д. Першим у повідомленні йде символ «C». Отже, перший інтервал (0,5-0,7). Приймемо його за інтервал (0,1). Тоді наступний символ «D» займає в ньому інтервал (0,64-0,7)б що відповідає 70% та 100% інтервалу (0,5-0,7). Таким чином пророблюємо з усіма символами у повідомленні і записуємо результати у таблицю кодування. З останнього інтервалу обираємо будь-яке число. Воно буде однозначно представляти наше повідомлення.

| Символ | Інтервал, що він займає | Нижня межа | Верхня межа |
|---|---|---|---|
| C | 0,2 | 0,5 | 0,7 |
| D | 0,06 | 0,64 | 0,7 |
| B | 0,018 | 0,652 | 0,67 |
| A | 0,0036 | 0,652 | 0,6556 |
| D | 0,00108 | 0,65452 | 0,6556 |
| D | 0,000324 | 0,655275 | 0,6556 |
| B | 0,0000972 | 0,6553408 | 0,655438 |
| A | 0,00001944 | 0,6553408 | 0,65536024 |
| C | 0,000003888 | 0,65535052 | 0,655354408 |
| B | 0,0000011664 | 0,6553512976 | 0,655352464 |

Таб. 4 Таблиця кодування

Число, що однозначно описує наше повідомлення має бути в інтервалі (0,6553512976; 0,655352464). Обираємо 0,655351533

Для того, щоб декодувати арифметичне кодування потрібно отримане число віднести до одного з наявних кодувань. Наприклад, число 0,97 у інтервал символу «D», тому перший символ для такого числа буде «D». Символ «D» відповідає інтервалу (0,7-1). Розділимо цей інтервал між іншими числами відповідно до їх інтервалів у допоміжній таблиці. Тоді, «A» займатиме інтервал (0,7 – 0,76), «B» займатиме (0,76-0,85), «C» займатиме (0,85-0,91), «D» - (0,91-1). Отже другий символ це також «D». Передане повідомлення: DD.

Отриманий код: 0,19406.

1) A (0 – 0,2)

| Символ | Ймовірність | Нижня межа | Верхня межа |
|---|---|---|---|
| A | 0,04 | 0 | 0,04 |
| B | 0,06 | 0,04 | 0,1 |
| C | 0,04 | 0,1 | 0,14 |
| D | 0,06 | 0,14 | 0,2 |

Таб. 5 Допоміжна таблиця для декодування

2) D (0,14-0,2)

| Символ | Ймовірність | Нижня межа | Верхня межа |
|---|---|---|---|
| A | 0,012 | 0,14 | 0,152 |
| B | 0,018 | 0,152 | 0,17 |
| C | 0,012 | 0,17 | 0,182 |
| D | 0,018 | 0,182 | 0,2 |

Таб. 6 Допоміжна таблиця для декодування

3) D (0,182-0,2)

| Символ | Ймовірність | Нижня межа | Верхня межа |
|---|---|---|---|
| A | 0,0036 | 0,182 | 0,1856 |
| B | 0,0054 | 0,1856 | 0,191 |
| C | 0,0036 | 0,191 | 0,1946 |
| D | 0,0053 | 0,1946 | 0,2 |

Таб. 7 Допоміжна таблиця для декодування

4) C (0,191-0,1946)

| Символ | Ймовірність | Нижня межа | Верхня межа |
|---|---|---|---|
| A | 0,00072 | 0,191 | 0,19172 |
| B | 0,00108 | 0,19172 | 0,1928 |
| C | 0,00072 | 0,1928 | 0,19352 |
| D | 0,00108 | 0,19352 | 0,1946 |

Таб. 8 Допоміжна таблиця для декодування

5) D (0,1935-0,1946)

Декодоване повідомлення: ADDCD

4. Вибрати в головному меню пункт завдання "Коди Хаффмана". Знайти за методом Хаффмана коди усіх запропонованих символів, виходячи із заданої ймовірності їх появи, що відображена у відповідній таблиці.

| Символ | Ймовірність |
|---|---|
| A | 0,5344 |
| B | 0,2495 |
| C | 0,0098 |
| D | 0,0358 |
| E | 0,0574 |
| F | 0,0077 |
| G | 0,0135 |
| H | 0,0919 |

Таб. 9 Ймовірності появи символів у повідомленні

Для того, щоб знайти відповідні коди необхідно розмістити символи від найбільш ймовірного до найменш ймовірного. Беремо два найменш ймовірних символи і об'єднуємо їх в один, додаючи їх ймовірності. Знову сортуємо за ймовірностями символи. Тка робимо доки не залишиться 2 символи, ймовірності яких у сумі дають 1.

| Символ | Ймовірність | Допоміжні імовірності | | | | | |
|---|---|---|---|---|---|---|---|
| A | 0,5344 | 0,5344 | 0,5344 | 0,5344 | 0,5344 | 0,5344 | 0,5344 |
| B | 0,2495 | 0,2495 | 0,2495 | 0,2495 | 0,2495 | 0,2495 | 0,4656 |
| H | 0,0919 | 0,0919 | 0,0919 | 0,0919 | 0,1242 | 0,2161 | |
| E | 0,0574 | 0,0574 | 0,0574 | 0,0668 | 0,0919 | | |
| D | 0,0358 | 0,0358 | 0,0358 | 0,0574 | | | |
| G | 0,0135 | 0,0175 | 0,031 | | | | |
| C | 0,0098 | 0,0135 | | | | | |
| F | 0,0077 | | | | | | |

Таб. 10 Розрахунок допоміжних ймовірностей

Тепер побудуємо кодове дерево для кода Хаффмана. З точки, що відповідає імовірності 1, опустимо 2 гілки з ймовірностями 0,5344 та 0,4656. Більшій імовірності присвоюємо 1, меншій – 0. Продовжуємо розгалужувати гілки та присвоювати їм значення 0 чи 1 за ймовірностями далі. Досягаючи ймовірності, що співпадає з однією з реальних імовірностей появи символа присвоюємо гілці кінцеве значення цього символа. Опускаючись вниз по намальованому дереву по кожній гільці можна отримати коди відповідних символів на кінцях гілок.

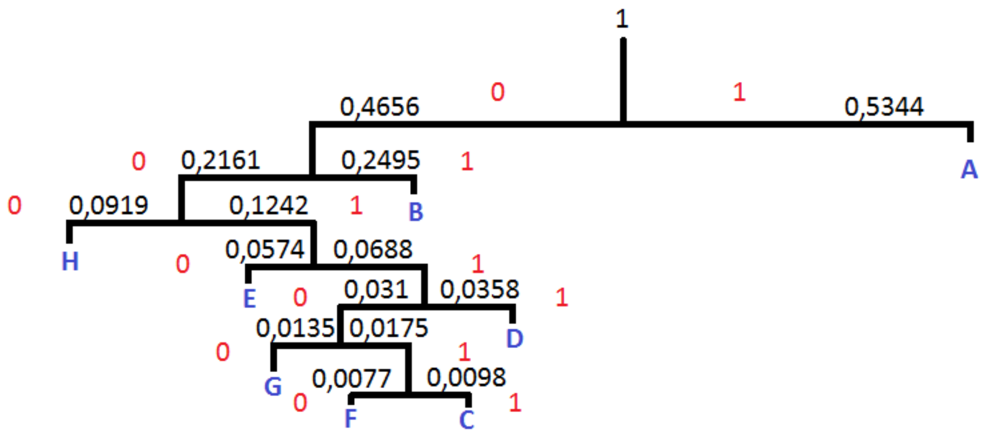


Рис. 1 Кодове дерево для кода Хаффмана

Отримані за допомогою дерева коди записуємо в таблицю.

| Символ | Код |
|---|---|
| A | 1 |
| B | 01 |
| C | 0011011 |
| D | 00111 |
| E | 0010 |
| F | 0011010 |
| G | 001100 |
| H | 000 |

Таб. 11 Коди Хаффмана для символів

5. Вибрати в головному меню пункт завдання "Коди Хеммінга". Відновити правильне повідомлення.

Прийняті кодові слова: \$1B5 \$0E9 \$1C8 \$01A \$17F \$0A1

Розташування у кодовому слові інформаційних та перевірочних бітів показано у таб. 12

| № позиції біта | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Призначення біта | $r_2$ | $i_1$ | $r_3$ | $i_5$ | $r_1$ | $i_2$ | $i_4$ | $i_3$ | $r_4$ |

Таб. 12 Розташування інформаційних і перевірочних бітів

Маємо усічений код Хеммінга (9,5), де 5 бітів – інформаційні, а 4 – перевірочні. Перевірочні бути – це біти парності. У даному випадку вони розраховуються за такими формулами:

$$r_1 = i_1 \oplus i_2 \oplus i_4 \oplus i_5$$
$$r_2 = i_1 \oplus i_3 \oplus i_4$$
$$r_3 = i_2 \oplus i_3 \oplus i_4$$
$$r_4 = i_5$$

В декодері в режимі виправлення помилок розраховуються синдроми – сполучення результатів перевірки на парність відповідних символів кодової групи, що характеризує певну конфігурацію помилок $s(s_1, s_2, s_3, s_4)$.

$$s_1 = r_1 \oplus i_1 \oplus i_2 \oplus i_4 \oplus i_5$$
$$s_2 = r_2 \oplus i_1 \oplus i_3 \oplus i_4$$
$$s_3 = r_3 \oplus i_2 \oplus i_3 \oplus i_4$$
$$s_4 = r_4 \oplus i_5$$

| Слово | $r_2$ | $i_1$ | $r_3$ | $i_5$ | $r_1$ | $i_2$ | $i_4$ | $i_3$ | $r_4$ |
|---|---|---|---|---|---|---|---|---|---|
| $1B5 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 |
| $0E9 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 |
| $1C8 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| $01A | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 |
| $17F | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $0A1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |

Таб. 13 Кодові слова у двійковій системі

Визначимо які синдроми яким помилкам відповідають. Для цього робимо нульовими всі біт окрім одного, для якого шукаємо синдром.

| Помилковий біт | $r_1$ | $r_2$ | $r_3$ | $r_4$ | $i_1$ | $i_2$ | $i_3$ | $i_4$ | $i_5$ |
|---|---|---|---|---|---|---|---|---|---|
| Синдром | 1000 | 0100 | 0010 | 0001 | 1100 | 1010 | 0110 | 1110 | 1001 |

Таб. 14 Конфігурації синдромів та відповідні помилкові біти

Розрахуємо за кодовими словами синдроми та визначимо помилкові біти.

| $r_2$ | $i_1$ | $r_3$ | $i_5$ | $r_1$ | $i_2$ | $i_4$ | $i_3$ | $r_4$ | $r'_1$ | $r'_2$ | $r'_3$ | $r'_4$ | $s_1$ | $s_2$ | $s_3$ | $s_4$ | Помилка |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | $i_3$ |
| 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | $i_1$ |
| 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | - |
| 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | $r_2$ |
| 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | $r_2$ |
| 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | $r_2$ |

Таб. 15 Конфігурації синдромів та помилкові біти

Врахуємо помилкові біти та запишимо інформаційні біти та відповідні їм шістандцяткові числа

| $i_1$ | $i_2$ | $i_3$ | $i_4$ | $i_5$ | Число |
|---|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 1 | $17 |

| | | | | | |
|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 1 | $09 |
| 1 | 1 | 0 | 0 | 0 | $18 |
| 0 | 1 | 1 | 0 | 0 | $0C |
| 0 | 1 | 1 | 1 | 1 | $0F |
| 1 | 0 | 0 | 0 | 1 | $11 |

Таб. 16 Скореговані слова

Відновимо значення цих слів за таб. 17

| Код символів | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
| 1 | Q | R | S | T | U | V | W | X | Y | Z | + | - | < | = | > | ? |

Таб. 17 Кодова таблиця символів, що передаються

Отже, отримані кодові слова відповідають наступному повідомленню: XJYMPR

Висновок

У даній роботі було вивчено методи ефективного кодування інформації та можливості кодів щодо корекції помилок.

У роботі було декодовано повідомлення за допомогою префіксного кодування і отримано наступні символи: CVTVVQ. Було показано, що при використанні непрефіксного кодування нема однозначності коду. Так, для запропонованого непрефіксного коду повідомлення можна було декодувати, як мінімум двома способами: VVLV та CYLV. І хоча код забезпечує високу однозначність коду, ефективність його посередня, адже коди приписуються символам випадковим чином, тому є висока вирогідність, що символу з високою ймовірністю появи буде приписаний довгий код, а символу з низькою ймовірністю появи короткий код. По-перше, в цьому випадку зменшується швидкість передачі, а по-друге збільшується вірогідність появи помилок, адже доволі часто передається довга послідовність.

За допомогою арифметичного кодування було закодовано наступне повідомлення: CDBADDBACB. Число, що однозначно описує це повідомлення знаходиться в межах (0,6553512976; 0,655352464). Для кодування було обрано наступне число з діапазону: 0,655351533. Також за допомогою арифметичного декодування з числа 0,19406 було одержано наступне повідомлення: ADDCD. Цей код є однозначним. Довжина отриманого числа залежить від довжини повідомлення, що є доволі оптимально. Код залежить від ймовірності появи символів. Однак, по-перше, перетворення десяткового числа в двійкову послідовність вимагає доадаткових ресурсів, по-друге, відсутність прямого співвідношення між кодом і символом та необхідність застосування непрямого алгоритму кодування та декодування призводить до збільшення не тільки часу передачі і отримання інформації, а й до збільшення вірогідності помилок.

За допомогою кодування Хаффмана було визначено коди для наступних символів: A, B, C, D, E, F, G, H (див. таб. 11). Це однозначний і оптимальний код, який має пряму відповідність між кодами та символами, а також враховує частоту появи символів у послідовності. Найефективніший некорегуючий код з розглянутих у роботі. Однак, немає можливості слідкувати за помилками та виправляти їх.

За допомогою корегуючих кодів Хеммінга було знайдено помилки у отриманих словах ($1B5 $0E9 $1C8 $01A $17F $0A1), встановлено правильні слова ($17 $09 $18 $0C $0F $11) та відновлено правильне повідомлення: XJYMPR. Не дивлячись на те, що з 9 бітів лише 5 несуть корисну інформацію, а 4 біти несуть інформацію про помилки, цей код дуже оптимально використовує парність як маркер помилки та операцію виключаючого АБО для визначення парності. Код однозначно виявляє помилки і дозволяє динамічно їх виправляти.