# 11. СТИСНЕННЯ ПОВІДОМЛЕНЬ ПРИ ПЕРЕДАЧІ ДАНИХ

**11.1. Теоретичні положення**

Стиснення повідомлень застосовується для прискорення передачі та обробки повідомлень, зменшення витрат на обробку, зберігання та пошук інформації, а також зменшення об’єму пам’яті ЕОМ .

Під стисненням повідомлень будемо розуміти операцію, у результаті якої даному повідомленню ставиться у відповідність повідомлення меншої довжини.

Для стиснення повідомлень при передачі даних використовують способи, які дозволяють повністю відновити початковий стан повідомлень або з частковою втратою інформації. Останні використовуються, головним чином, при цифровій обробці сигналів та зображень, тобто графічної інформації ( креслень, графіків, діаграм тощо ), і тому розглядатися у цьому розділі не будуть.

Різниця між способами стиснення повідомлень, які застосовуються при передачі даних, та способами ( архіваторами ), які використовуються при архівації повідомлень в ЕОМ, полягає у тому, що при передачі даних оперують з інформаційними масивами значно меншого обсягу ( від 32 біт до 2 кбіт ), а при архівації – з великими масивами ( до 10 і більше кбайт ). Викликане це тим, що при передачі даних, як правило, вводяться жорсткі обмеження на час обробки ( затримки ) повідомлень у передавальному та приймальному пристроях системи передачі даних.

Ефективність стиснення визначається коефіцієнтом стиснення *K*ст =  *N*1/*N*2 , де  *N*1 та *N*2 – відповідно кількість бітів ( байтів ) у первинному і стисненому масивах.

Способи стиснення повідомлень з повним відновленням їх початкового стану можна розділити на: лінійні, матричні, комбіновані та каскадні.

До **лінійних способів стиснення** належать способи, у яких стиснення виконується по рядках або стовпцях. Це способи:

–  з використанням символу *ρ*;

–  з використанням символів *ρ* і *K*;

–  з використанням символів *X*, *Y*, *Z*;

–  з вилученням символів, що повторюються;

–  кодовий;

–  адаптивного кодування;

–  зонний.

***Спосіб з використанням символу ρ***  полягає у тому, що замість вилучених розрядів масиву вводиться знак розділення *ρ*, який дозволяє відокремити елементи у згорнутому масиві. При розгортанні замість знаку *ρ* поновлюються всі пропущені розряди, які були до елемента, що міститься безпосередньо за *ρ* у стисненому масиві. Запис знаків, що стоять після *ρ*, виконується з кінця рядка ( стовпця ). Використання цього способу буде ефективним у тому разі, коли інформаційний масив подається у вигляді рядків або стовпців, що розташовані у зростаючому порядку, має однакові значення елементів в одних і тих же розрядах ( що характерне для техніко-економічної інформації ).

Недоліком цього способу стиснення є неможливість його застосування до впорядкованих масивів, у яких розряди, що повторюються, зустрічаються не на початку рядків ( стовпців ).

***Спосіб з використанням символів ρ та K***  полягає у тому, що замість вилучених розрядів масиву вводяться знак розділення *ρ*, який дозволяє відокремити елементи у згорнутому масиві, та знак кінця рядка ( стовпця ). У цьому разі символом *ρ*  позначається одне повторення елементів, яке може бути розташоване у будь-якому місці

рядка ( стовпця ). Розгортання масиву ведеться від *K* до *K*. При фіксованій довжині рядка (стовпця) всі розряди, які містяться між *K* разом з пропущеними розрядами, повинні утворювати повний рядок ( стовпець ).

***Спосіб з використанням символів X*, *Y*, *Z*** застосовується у разі наявності у рядку ( стовпці ) декількох ділянок, що повторюються. Спеціальні символи *X*, *Y*, *Z*, що вводяться, вказують на число пропусків. При цьому необхідність у символі *K*, що визначає кінець рядка ( стовпця ), відпадає.

Розгортання масивів виконується з їх початку або кінця, а заповнення відповідної кількості пропусків замість додаткових символів *X*, *Y*, *Z* здійснюється переносом відповідної кількості символів, які розташовані на однойменних розрядах попереднього рядка (стовпця ).

***Спосіб з вилученням символів*, *що повторюються*,** застосовується при стисненні різного роду специфікацій, переліків однойменних товарів тощо. Стиснення виконується за допомогою введення двох додаткових символів, один з яких вказує на повторення, а другий – на кількість повторень ( кількість літер або цифр, що повторюються, у десятковій системі числення ). Відновлення даних при розгортанні виконується зверху вниз і справа наліво.

***Кодовий спосіб*** грунтується на переході від однієї системи числення ( наприклад, двійкової ) до іншої ( наприклад, четвіркової, вісімкової тощо ), яка б давала виграш у кількості елементів у повідомленні.

***Спосіб з використанням адаптивного кодування*** використовується, головним чином, при стисненні повідомлень статистичного характеру, які містять як текстову частину, так і цифрові дані. За цим способом всі повідомлення за допомогою аналізатора інформації поділяються на категорії, якими визначається різна цінність пові-домлень і з якою достовірністю її дозволяється передавати. Тоді, за-мість використання тільки одного коду заданої довжини для кодування всіх повідомлень, який би забезпечив однакову достовірність передачі в

сіх повідомлень, застосовуються декілька кодів різної довжини з відповідною спроможністю виявляти та виправляти помилки. Це дає змогу зменшити середню довжину кодових комбінацій *n*сер, якими кодуються окремі знаки повідомлень різної категорії. Якщо позначити: *N*1, *N*2, *N*3 – кількість знаків ( кодових комбінацій ) у повідомленнях, відповідно, першої, другої та третьої категорій; *N* = *N*1+*N*2+*N*3 – загальна кількість знаків ( кодових комбінацій ) у повідомленнях всіх категорій; *n*1, *n*2, *n*3 – довжина кодових комбінацій, якими кодуються знаки повідомлень, відповідно, першої, другої та третьої категорій, тоді

|  |  |  |
| --- | --- | --- |
| *n*cep= | *N*1*n*1+*N*2*n*2+*N*3*n*3 |  |
| N | *.* |

***Зонний спосіб стиснення*** застосовується при передачі пові-домлень з обмеженим числом знаків у них, наприклад, телеграфних. Це дає змогу замість передачі 128 або 256 знаків довжиною у 1 байт обмежитися передачею 48...64 знаків довжиною у 1/2  байта. Це стає можливим, якщо з 16 двійкових тетрад ( чотирибітових ком-бінацій ), наприклад, 12 виділити на передачу знаків повідомлень, а решту з 4 комбінацій  –  на позначення зон, в яких розташовані відпо-відні знаки. Як правило, у кодовому слові вторинного алфавіту перші 4 розряди будуть становити зону, а 4 інші – знак. Кількість комбінацій вторинного алфавіту у даному разі буде *N*  =  4 ×12 = 48.

Для виконання зонного стиснення повідомлень потрібно знаки ( літери і цифри ) вторинного алфавіту розбити на зони, по 12 у кож-ній зоні. При цьому, якщо у тексті зустрічаються поряд знаки, які належать до однієї зони, її номер вказується тільки перед першим знаком, а запис наступних знаків обмежується записом їх знаковою частиною.

Для того, щоб знаки, які мають однакові зони, утворювали більш довгі послідовності при створенні кодових слів, у вторинному алфавіті необхідно враховувати статистичні характеристики алфавіту,

з якого утворюються повідомлення. Бажано також врахувати і ймовірності різних сполучень деяких літер. Ефективність розбивки на зони визначається за допомогою коефіцієнту стиснення *K*ст = *N*1/ *N*2 , де *N*1 та *N*2 –  відповідно кількість байтів у первинному і стисненому масивах. У таблиці 11.1 наведений один з прикладів розбивки знаків вторинного алфавіту на зони ( у дужках поданий запис двійкових тетрад у шістнадцятковій системі числення ).

Таблиця 11.1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Код  знака | Код зони | | | |
| 1100(С) | 1101(D) | 1110(E) | 1111(F) |
| 0000(0) | Пробіл | М | Ф | - |
| 0001(1) | О | П | Ш | 1 |
| 0010(2) | Е | Є | Щ | 2 |
| 0011(3) | А | З | Ц | 3 |
| 0100(4) | Р | К | Ч | 4 |
| 0101(5) | Л | Д | Ж | 5 |
| 0110(6) | Т | Я | Х | 6 |
| 0111(7) | Н | У | Ю | 7 |
| 1000(8) | В | Ь | , | 8 |
| 1001(9) | І | Б | . | 9 |
| 1010(А) | И | Й | : | ! |
| 1011(В) | С | Г | ; | ? |

До **матричних способів стиснення**  належать способи, у яких стиснення елементів інформаційного масиву виконується з використанням матричного принципу заміни елементів, що повторюються. Це способи: зі зберіганням атрибутів у вигляді бітової матриці та із заміною елементів, що повторюються, типовими матрицями.

***Спосіб зі зберіганням атрибутів у вигляді бітової матриці*** грунтується на скінченому числі атрибутів, які виносяться у першу

частині ( шапку ) матриці, тіло якої є набір двійкових елементів, що позначають за допомогою “1” – наявність, а за допомогою ”0” – від-сутність атрибута. Шапка бітової матриці і її тіло можуть зберігатися на різних ділянках пам’яті ЕОМ.

***Спосіб з заміною елементів*, *що повторюються*, *типовими матрицями*** грунтується на заміні елементів, що повторюються на деякій визначеній площі інформаційного масиву, одиночними елементами, які несуть ознаки цих ділянок масиву. Ці одиночні елементи відповідають деяким типовим матрицям певних розмірів, які визначаються заздалегідь для даного інформаційного масиву і відомі передавальному і приймальному пристроям.

**Комбіновані способи стиснення** поєднують у собі одночасне використання для стиснення інформаційного масиву двох чи більше лінійних або / та матричних  способів.

З комбінованих способів найбільше поширення набув ***лінійно-матричний спосіб стиснення***, в якому виконується одночасне стиснення інформаційного масиву лінійним способом з використанням символів *X*, *Y*, *Z*  та матричним способом з заміною елементів, що повторюються, типовими матрицями.

До **каскадних способів стиснення** належать способи, у яких стиснення інформаційного масиву виконується послідовно, з використанням декількох ( лінійних  та / або  матричних ) способів стиснення. З цих способів найбільше поширення набув зонно-кодовий спосіб стиснення, коли послідовно виконується спершу зонне, а потім – кодове стиснення.

Використання комбінованих та каскадних способів дає можливість значно підвищити коефіцієнт стиснення, але при цьому, звичайно, збільшується час обробки інформаційного масиву як при згортанні ( стисненні ) – у передавальному, так і при розгортанні – у приймальному пристроях.

**11.2. Приклади розв’язання задач**

**Задача  11.2.1**

Стиснути лінійними способами з використанням символу *ρ* та з використанням символів *X*, *Y*, *Z*інформаційний масив:

10011000

10011011

10011010

10011110

10011110.

Порівняти ефективність стиснення цими способами.

***Розв’язання.*** Стиснутий інформаційний масив буде мати вигляд:

– при стисненні способом з використанням символу *ρ*:

10011000

*ρ*11*ρ*0*ρ*11

0*ρ* ,

коефіцієнт стиснення: *K*ст.1 = 40 / 18 =  2,2(2);

– при стисненні способом з використанням символів *X*, *Y*, *Z*, коли *Х* = 2, *Y*= 3, *Z*= 5 : 10011000

*YY*11*ZX*0*Z*

1*XZY* ,

коефіцієнт стиснення: *K*ст.2 =  40 / 20  = 2 .

Ефективність стиснення для вказаного інформаційного масиву вища при використанні першого способу: *K*ст.1 >  *K*ст.2 .

Задача  11.2.2

Стиснути лінійними способами з використанням символів *ρ*  і *K* та з використанням символів *X*, *Y*, *Z*інформаційний масив: 00011001

10011011

10010011

00010010

00010010.

Порівняти ефективність стиснення цими способами.

***Розв’язання.*** Стиснутий інформаційний масив буде мати вигляд:

–  при стисненні способом з використанням символів *ρ* та *K* :

00011001

*K*1*ρ*11*Kρ*0

011*K*0*ρ*0*K*

*ρK* ,

коефіцієнт стиснення: *K*ст.1 = 40 / 26  =  1,538 ;

–  при стисненні способом з використанням символів *X*, *Y*, *Z*, коли *Х* = 2, *Y* = 3, *Z* = 5:

00011001

1*Z*1*Z*0*Y*0*Y*

*Y*0*ZY* ,

коефіцієнт стиснення: *K*ст.2 =  40 / 20  = 2 .

Ефективність стиснення для вказаного інформаційного масиву вища при використанні другого способу: *K*ст.2 > *K*ст.1 .

Задача  11.2.3

Визначити ефективність стиснення способом з використанням адаптивного кодування масиву статистичної інформації загальним обсягом  *N*= 1000 знаків, який містить: повідомлення першої категорії ( цифрові повідомлення від 1000 і вище ) обсягом *N*1= 100 знаків, для яких при передачі було б гарантоване виправлення двох помилок; повідомлення другої категорії ( цифрові повідомлення від одиниць до 999 ) обсягом *N*2 = 200 знаків, для яких при передачі було б гарантоване виправлення однієї помилки; повідомлення третьої категорії ( текстові повідомлення ) обсягом *N*3 = 700 знаків, для яких при передачі було б гарантоване виявлення однієї помилки. Кожний знак кодується первинним семиелементним ( *k* = 7 ) кодом.

***Розв’язання.*** Вибираємо коди для кодування повідомлень:

першої категорії – код БЧХ з кодовою відстанню *dmin* = 5, який дозволяє виправити дві помилки. Параметри коду: *k*= 7, *r*= 8, *n*1=  *k*+ *r*=  7 + 8  =  15;

другої категорії – код Хеммінга з кодовою відстанню *dmin* = 3, який дозволяє виправити одну помилку. Параметри коду: *k* = 7, *r* = 4, *n*2 =  *k*+ *r* =  7 + 4  = 11;

третьої категорії – код з перевіркою на парність з кодовою відстанню *dmin* = 2, який дозволяє виявити одну помилку. Параметри коду: *k* = 7, *r* = 1, *n*3 =  *k*+ *r*=  7 + 1  =  8.

Визначимо середню довжину кодової комбінації:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *n*cep= | *N*1*n*1+*N*2*n*2+*N*3*n*3 | = | 100×15 + 200×11+ 700×8 | = | 9,3. |
| N | 1000 |

Коефіцієнт стиснення визначаємо з урахуванням того, що без використання способу стиснення за допомогою адаптивного кодування, всі повідомлення треба було б кодувати тільки кодом БЧХ довжиною  *n*1= 15 елементів.  Тоді *K*ст = 15 / 9,3 = 1,613.

Задача  11.2.4

Розгорнути інформаційний масив:

10100001

*XX*1*Y*01*Z*0

1*Z*11*X*0*Z* ,

якщо він був стиснутий лінійним способом з використанням символів *X*, *Y*, *Z* ***,*** при *X* = 2, *Y* = 3, *Z* = 5. Визначити коефіцієнт стиснення.

***Розв’язання.*** Розгортання згорнутого масиву виконуємо з першого рядка, який переписуємо без змін, а потім робимо пропуски у розрядах наступних рядків згідно з заданими значеннями символів *X*, *Y*, *Z*:

1 0 1 0 0 0 0 1 10100001

. . . . 1 . . . 10101001

0 1 . . . . . 0 → 01101000

1 . . . . . 1 1 11101011

. . 0 . . . . . 11001011.

Коефіцієнт стиснення *K*ст = 40 / 23  =  1,739.

Задача  11.2.5

Стиснути лінійним зонним та каскадним зонно-кодовим способами такий текст: «Україна здобула державну незалежність у 1991 році.». Порівняти ефективність стиснення цими способами.

***Розв’язання.*** Для стиснення за цими способами необхідно побудувати таблицю розбивки алфавіту на зони. Для цього використаємо табл.11.1. Тоді стиснутий текст буде мати вигляд:

при лінійному зонному способі стиснення: 1101 0111 0100 1100 0100 0011 1001 0111 0011 0000 1101 0011 0101 1100 0001 1101 1001 0111 1100 0101 0011 0000 1101 0101 1100 0010 0100 1110 0101 1100 0011 1000 0111 1101 0111 1100 0000 0111 0010 1101 0011 1100 0011 0101 0010 1110 0101 1100 0111 1001 1011 0110 1101 1000 1100 0000 1101 0111 1100 0000 1111 0001 1001 1001 0001 1100 0000 0100 0001 1110 0011 1100 1001 1110 1001 ;

коефіцієнт стиснення *K*ст.1 = ( 50 × 8 ) / ( 75 × 4 ) = 1,3(3);

при каскадному зонно-кодовому стисненні: D74C439730D35 C1D97C530D5C24E5C387D7C072D3C352E5C79B6D8C0D7C0F1991C041E3C9E9;

коефіцієнт стиснення *K*ст.2 = ( 50 × 8 ) / 75 = 5,3(3).

Ефективність стиснення для вказаного інформаційного масиву вища при використанні другого способу вища: *K*ст.2 > *K*ст.1 .

Задача  11.2.6

Стиснути лінійним кодовим способом при *q* = 4 та матричним способом з заміною елементів, що повторюються, типовими матрицями інформаційний масив:

00011001

10011001

00010001

10010010

00010010

10010010.

Порівняти ефективність стиснення цими способами.

***Розв’язання.*** Стиснутий інформаційний масив буде мати вигляд:

при стисненні лінійним кодовим способом з *q* = 4 :

0121

2121

0101

2102

0102

2102,

коефіцієнт стиснення: *K*ст.1 =  48 / 24  =  2 ;

при стисненні матричним способом з заміною елементів, що повторюються, типовими матрицями, якщо типові матриці *L* та *M* мають розміри – *L*= 3 × 2, *M*= 4 × 3:

00011001

1*L*1*L*001*M*

0100*L*1 ;

коефіцієнт стиснення: *K*ст.2 = 48 / 22 =  2,181.

Ефективність стиснення для вказаного інформаційного масиву вища при використанні другого способу: *K*ст.2 > *K*ст.1 .

**11.3. Задачі**

**11.3.1.**  Згідно з варіантами, поданими в таблиці 11.3.1, стиснути лінійними способами: з використанням символу *ρ*; з використанням символів *ρ*  і *K*; з використанням символів *X*, *Y*, *Z*( *X*= 2, *Y*= 3, *Z*= 5 ) ; кодовими з *q* = 4 та *q* = 16 інформаційний масив *А*. Показати процес розгортання стиснутого інформаційного масиву. Визначити коефіцієнт стиснення та порівняти ефективність стиснення цими способами.

Таблиця 11.3.1

|  |  |  |  |
| --- | --- | --- | --- |
| №  варіанта | Інформаційний масив *А* | №  варіанта | Інформаційний масив *А* |
| 1 | 11010001  11010001  11010100  11010101  11010110 | 6 | 01011000  01010011  01011010  01011010  01011000 |
| 2 | 10111001  10111001  10111100  10111101  10111010 | 7 | 00011001  10011001  00010001  10010010  00010010  10010010 |
| 3 | 01011000  01011001  10011011  00010011  10010010  00010010 | 8 | 00011001  10011001  00010001  10011001  10011001  10011001 |
| 4 | 01011000  10011001  10011011  01011010  11010010  01010000 | 9 | 000110011001  100110011001  001000110101  011011100111  100110111001  100110111001 |
| 5 | 10011000  00011011  01011010  01111110  11111111.  11011011 | 10 | 00011001  10001001  10010001  10010011  00010011  10010011 |

**11.3.2.** Згідно з варіантами, поданими в таблиці 11.3.2, визначити ефективність стиснення способом з використанням адаптивного кодування масиву статистичної інформації загальним обсягом *N* знаків, який містить: повідомлення першої категорії ( цифрові повідомлення від 1000 і вище ) обсягом *N*1  знаків, для яких при передачі було б га-рантоване виправлення трьох помилок; повідомлення другої категорії ( цифрові повідомлення від одиниць до 999 ) обсягом *N*2 знаків, для яких при передачі було б гарантоване виправлення двох помилок; повідомлення третьої категорії ( текстові повідомлення ) обсягом *N*3 знаків, для яких при передачі було б гарантоване виправлення однієї помилки. Кожний знак кодується двійковим первинним *k*-кодом.

Таблиця 11.3.2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| №  варіанта | Масив статистичної інформації | | | | Довжина первинного коду, *k* |
| *N* | *N*1 | *N*2 | *N*3 |
| 1 | 1500 | 200 | 300 | 1000 | 8 |
| 2 | 2500 | 300 | 700 | 1500 | 8 |
| 3 | 3500 | 500 | 1000 | 2000 | 7 |
| 4 | 7000 | 1000 | 2000 | 4000 | 7 |
| 5 | 10000 | 2000 | 3000 | 7000 | 7 |

**11.3.3.** Згідно з варіантами, поданими в таблиці 11.3.3, стиснути матричним способом з заміною елементів, що повторюються, типовими матрицями та комбінованим лінійно-матричним способом інфор-маційний масив *А*. Показати процес розгортання стиснутого масиву. Визначити коефіцієнт стиснення та порівняти ефективність стиснення цими способами.

Таблиця 11.3.3

|  |  |  |  |
| --- | --- | --- | --- |
| №  варіанта | Інформаційний  масив *А* | №  варіанта | Інформаційний  масив *А* |
| 1 | 11010001  11010001  11010100  11010101  11010110 | 6 | 01011000  01010011  01011010  01011010  01011000 |
| 2 | 10111001  10111001  10111100  10111101  10111010 | 7 | 00011001  10011001  00010001  10010010  00010010  10010010 |
| 3 | 01011000  01011001  10011011  00010011  10010010  00010010 | 8 | 00011001  10011001  00010001  10011001  10011001  10011001 |
| 4 | 000110011001  100110011001  000100011001  100110111001  100110111001  100110111001 | 9 | 000110011001  100110011001  001000110101  011011100111  100110111001  100110111001 |
| 5 | 10011000  00011011  01011010  01111110  11111111  11011011 | 10 | 00011001  10001001  10010001  10010011  00010011  10010011 |

**11.3.4.** Згідно з варіантами, поданими в таблиці 11.3.4, стиснути лінійним зонним та каскадним зонно-кодовим способами заданий текст. Визначити коефіцієнт стиснення та порівняти ефективність стиснення цими способами. При стисненні користуватися таблицею 11.1.

Таблиця 11.3.4

|  |  |
| --- | --- |
| №  варіанта | Текст |
| 1 | Вперше людина побувала у космосі у 1961 році. |
| 2 | Стиснення повідомлень використовується для прискорення процесів обробки та пошуку інформації. |
| 3 | Сучасні електронно-обчислювальні машини побудовані на дискретних елементах. |
| 4 | Мультиплікативні завади, які діють у радіоканалах, впливають на якість прийому сигналів. |
| 5 | 9.03.1814 року народився Тарас Григорович Шевченко. |