

ПОМІХОСТІЙКІ ШТРИХОВІ КОДИ

М. Ф. Козак, В. В. Ніколенко

УжДУ, каф. кібернетики та прикладної математики, Ужгород, вул. Підгірна, 46
тел. (03122)43325, jvtech@mail.uzhgorod.ua

ВСТУП

Штриховому кодуванню інформації, що надзвичайно широко використовується в країнах заходу, в нашій країні приділяється дуже мало уваги. Такий стан речей є незаслуженим – штрихові коди покликані полегшити та прискорити рутинну роботу людини в найрізноманітніших галузях людської діяльності. Широка сфера застосування штрихових кодів спричинила велику кількість найрізноманітніших типів штрихкодів. Нижче буде описано процес створення нового поміхостійкого типу штрихового коду Нем-8. Потреба в створенні цього коду викликана комп'ютеризацією контролю за переміщенням персоналу на підприємстві. Для цього кожного співробітника потрібно забезпечити пластиковою картою-перепусткою, на якій окрім фото та звичайної інформації про робітника, повинен був бути штриховий код для автоматичного зчитування інформації про робітника сканерами, що підключені до комп'ютера. Сформульовано вимоги до штрихового коду: простота та економічність у виготовленні, можливість кодувати до 8 чисел, і висока стійкість до помилок. Остання вимога найважливіша – при зчитуванні сканером штрихкоду, часто з'являються помилки. Причинами цього є ряд факторів: простота скануючих пристроїв, рух картки через сканер із прискоренням (робітник сам повинен протягувати її через сканер, а швидкість руху руки може бути нерівномірною), пошкодження штрихового коду у виді забруднення, подряпин, підтертостей. Помилки, що виникають внаслідок цього, є причинами того, що штриховий код не вдається розкодувати, або, що ще гірше, він розкодується неправильно, вони не дають змоги використовувати штрихові коди вже існуючих типів і виникає потреба в створенні нового поміхостійкого типу штрихкоду, основною особливістю якого є висока стійкість до помилок.

В ході роботи по створенню нового типу штрихового коду було вивчено та порівняно основні типи існуючих штрихкодів [3,4,5,6], на їх основі описано основні методи побудови нових типів штрихкодів, отримано математичний апарат, що дає змогу оцінювати стійкість коду до помилок, його надлишковість [1], та на основі цього апарату створено програмний пакет для дослідження стійкості до помилок різних типів штрихових кодів. Результатом цієї роботи став штриховий код Нем-8 основною відмінністю якого, від вже існуючих типів штрихових кодів, є його висока стійкість до помилок, яка була досягнута використанням в

штриховому кодуванні методів корекції помилок, що базуються на алгоритмі Хеммінга.

1. Способи побудови штрихових кодів

У штриховому кодуванні існує два способи графічного представлення бінарних значень '0' та '1'. Перший спосіб: значення '0' та '1' кодуються відповідно двома кольорами - білим та чорним. Наприклад: бітова послідовність 10110011101100011 буде мати наступне штрихове представлення (мал. 1). В цьому способі штрихи, що відповідають '0' та '1' мають однакову ширину. В разі якщо в бінарній послідовності йдуть одне за одним кілька однакових п значень '0' чи '1' їм буде відповідати білий чи чорний штрих п-кратної ширини.



Мал. 1

Другий спосіб представлення бінарних значень '0' та '1' у штриховому коді є спосіб, коли біти задані не різними кольорами, а різними значеннями ширини штрихів. Тобто є чотири атомарні графічні символи: два вузькі штрихи та два широкі, білого та чорного кольору. В такому штриховому коді білі та чорні штрихи чергуються, а значенням '0' та '1' відповідають відповідно широкі та вузькі штрихи. В цьому разі наведена вище бінарна послідовність буде мати вигляд (мал.2).

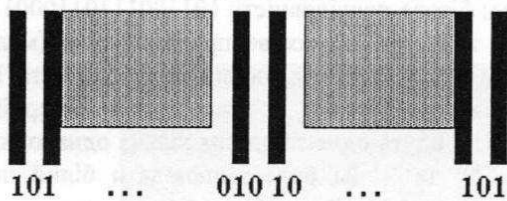


Мал. 2

Кожен із цих варіантів має як переваги, так і недоліки. В першому варіанті штриховий код буде коротшим, в наслідок того, що всі біти кодуються однаковими по ширині штрихами. По цій же причині в першому варіанті штриховий код бінарної послідовності зі сталим числом бітів буде мати сталий розмір, в той час як у другому варіанті розмір штрихового коду буде залежати від співвідношення нулів та одиниць. По цій причині перший варіант використовується якщо розмір інформації, що кодується, сталий, а другий варіант, коли кількість інформації, що кодується, змінна.

Розглянемо інші особливості побудови штрихових кодів, які також використовуються для

класифікації штрихових кодів. Однією з них є наявність чи відсутність контрольних штрихів (бітів). В випадку відсутності контрольних штрихів, при нерівномірній швидкості зчитування штрихкоду, є ймовірність, що цей код буде інтерпретовано неправильно. Щоб цьому запобігти, на початку та в кінці коду розміщується набір із принаймні двох контрольних штрихів. Після зчитування сканером цього коду, обчислювальна машина може судити про зміну швидкості сканування штрихового коду і відповідно корегувати процес декодування. Прикладом застосування контрольних штрихів може бути штриховий код типу EAN-13. В ньому контрольні штрихи наявні не тільки на початку та в кінці, а і в середині коду (мал. 3).



Мал.3

Ще однією особливістю при побудові штрихового коду є наявність чи відсутність контрольної суми. Для гарантування правильності декодування штрихового коду в деяких типах штрихового коду до інформації, що кодується, додається деяка контрольна сума, яка функціонально залежить від кодуваної інформації. Ця контрольна сума кодується в штриховий код разом з основною інформацією, а при декодуванні ЕОМ знову вираховує контрольну суму цього коду і порівнює із заданим. Зрозуміло, що неспівпадання цих двох контрольних сум свідчить про неправильно зісканований штрихкод. Поряд із контрольними числами, які кодуються разом з інформацією, нерідко в бітовій послідовності використовуються також біти контролю парності. Всі вищезазначені методи, покликані мінімізувати ймовірність неправильного інтерпретування штрихового коду, що сканується. В цьому штрихові коди чимось нагадують бітові послідовності сигналів, що передаються через цифрові канали зв'язку, тощо. Але штрихові коди мають одну дуже важливу відмінність, бітова послідовність, яка графічно представлена штрихкодом, не повинна містити велику кількість підряд розміщених однакових бітів, тому що графічне представлення такої послідовності, утвореної з одних тільки "0", перетвориться на тривалу темну смугу на штрихкоді, виміряти довжину якої для сканера буде складно. Тому, як правило, на штрихкоди накладається обмеження на кількість підряд розміщених однакових бітів, найчастіше це не більш ніж 3 однакові елементи. Все вище описане приводить до збільшення надлишковості інформації у штриховому коді, але в той самий час збільшує надійність коду. Надлишковість інформації присутня практично в

кожному типі штрихового коду в більшій чи меншій мірі, але у випадку штрихового кодування міру надлишковості можна умовно вважати мірою надійності (стійкості до помилок) штрихового коду.

2. Типи помилок, що виникають при штриховому кодуванні

В ході дослідження та створення штрихових кодів особливу увагу приділяють типам помилок, що виникають, та факторам, які стають причинами цих помилок. Найчастіше - це нерівномірна швидкість сканування штрихового коду, яка може бути наслідком багатьох причин: а) ручне сканування, дуже часто рука людини, що тримає сканер чи носій штрихкоду, рухається з прискоренням; б) траєкторія руху штрихкоду відносно сканера нелінійна; в) короткі зупинки штрихкоду під час сканування. Запобігти таким помилкам можна ще на стадії створення нового типу штрихового коду, ввівши в нього додатково контрольні біти-штрихи (мал.3), чи розмістивши поряд зі штрихкодом шкалу синхронізації сканера (мал. 4).



Мал.4

Нерівномірна швидкість сканування спричиняє помилкові ситуації, які досить важко виправити – порушуються пропорції в ширині штрихів і, як наслідок, в бінарному коді після сканування та декодування з'являються зайві біти, виявити котрі дуже складно.

Ще одним фактором, що спричиняє помилки при декодуванні штрихового коду, є забруднення та пошкодження штрихкоду. Такі проблеми виникають значно рідше ніж попередні, передусім, тому що з'являються, як правило, тільки при тривалому та частому використанні носія зі штрихкодом (посвідчення особи, бібліотечна справа, і т. п.). В таких випадках виникають три типи помилок: а) написи на штрихкоді, подряпини, перегини; б) плями; в) підтертості та іншого роду зменшення контрасту штрихового коду. В першому випадку пошкодження спричиняють заміну 1-2 бітів при декодування на протилежні. У другому випадку в бітовій послідовності з'являється тривала послідовність однакових бітів. В третьому випадку з'являється висока ймовірність неправильного декодування штрихового коду.

Усі ці три випадки формують специфічні задачі по розпізнаванню:

1. Локалізація та видалення написів із штрихової поверхні (мал.5).
2. Оцінка розмірів, локалізація та видалення плям (у разі якщо пляма не покриває повністю штрихкод). (мал.6).

3. Очистка штрихкоду та збільшення його контрастності (мал.7).



Мал.5



Мал. 6



Мал. 7

3. Нем-8

Нем-8 штриховий код створений для використання на посвідчені особи в межах великого підприємства. Цей штрихкод дозволяє кодувати 8-ми значні порядкові номери. Головною особливістю цього типу штрихкоду є використання методу Хеммінга для виявлення та корекції помилок. Серед існуючих типів штрихових кодів використовуються найрізноманітніші методи контролю за помилками, але методи корекції помилок, або не використовуються, або слабо реалізовані. Нижче описується процес створення штрихового коду Нем8.

Вимоги до штрихкоду, що були сформульовані до його створення:

Вимога 1. Мінімальна кількість символів, яку повинен містити штрихкод – 6. Можливо збільшити це число до 7 або 8, ввівши резервне або контрольне число.

Вимога 2. Алфавіт нашого штрихового коду складається тільки з 10 цифр 0,1, ...,9. Тому на кодування одного символу мінімально потрібно 4 біти $2^4=16$ ($2^3<10$) (насправді ця кількість буде дещо більшою, за рахунок уведення контрольних бітів та накладення деяких додаткових умов. Однією з таких умов буде наступна).

Вимога 3. Для запобігання злиття двох сусідніх бітових рядів, необхідно щоб кожен із таких рядів розпочинався з '0' та закінчувався '1'-цею.

Вимога 4. Велика кількість підряд розміщених однакових символів збільшує ймовірність помилки. Тому накладається ще одна умова: в бітовому ряді не повинно бути підряд більш ніж 3 однакових бітів.

Вимога 5. Як зазначалося у вимозі 2, в наслідок вимог 3-4, на кодування одного символу потрібно 6 біт.

Штриховий код Нем-8 забезпечено

властивістю самовиправлення помилок, яка базується на методі Хеммінга. Для цього необхідно з'ясувати, яку кількість бітів потрібно відвести під перевірочні. Виходячи з відомої формули $2^k > n+1$, де n - загальна кількість бітів, k - кількість перевірочних бітів, $m=n-k$ - кількість інформаційних бітів, отримуємо наступну таблицю:

Табл. 1.

K	N	m=k-n	K	N	m=k-n
1	1	0	5	31	26
2	3	1	6	63	57
3	7	4	7	127	120
4	15	11	8	255	247

Мінімальне число бітів у бітовому ряді 6, якщо під інформаційні відвести 26 біт, то можна було б кодувати $26/6=4$ символи. Така кількість є недостатньою, тому виділимо

Під інформаційні виділено 57 бітів (при 26 можна кодувати тільки $26/6=4$ символи), таким чином можна кодувати до $57/6=9$ символів, що є навіть надлишково для сформульованих вимог.

В таблиці 2 вказано всі можливі 6-ти значні бітові ряди, що задовольняють нашим умовам.

Табл. 2.

1	000101	5	001101	9	010111
2	000111	6	010001	10	011001
3	001001	7	010011	11	011011
4	001011	8	010101	12	011101

Цієї кількості бітових рядів достатньо, щоб кожному символів алфавіту штрихового коду було поставлено у відповідність по одному з них. Але розгляд матриці кодових відстаней цих бітових рядів, вказує на деякі недоліки набору. В різі виникнення однієї помилки в штриховому коді, вона буде виправлена за допомогою алгоритму Хеммінга. В разі виникнення двох помилок, метод Хеммінга перестане бути ефективним, в такій ситуації принаймні повинно бути виявлено наявність помилки. Перевірку на наявність подвійних помилок забезпечується вводом восьмого контрольного числа. Воно функціонально залежить від перших семи, і ця залежність вибирається так, щоб при зміні одного з перших семи чисел, восьме контрольне також обов'язково змінилось. Якщо при виникненні двох помилок, вони обидві знаходяться в одному бітовому ряді, то наявність такої помилки буде виявлено контрольним числом, але якщо ці помилки з'являться в різних бітових рядах, тобто будуть змінені два числа із семи, то можливі такі комбінації чисел, коли помилка не буде виявлена взагалі. Для запобігання таких помилок набір бітових рядів потрібно підібрати так, щоб матриця кодових відстаней цього набору не містила одиниць. Тоді при виникненні двох помилок у різних бітових рядах, ці новоутворені бітові ряди не будуть належати нашому наборові, що буде свідчити про помилку. В разі, якщо матриця кодових відстаней не містить одиниць,

то, для того щоб помилковий код не виявлявся, необхідно, щоб помилки виникли по дві в двох різних бітових рядах. Тобто вивід із матриці кодових відстаней одиниць, значно покращує стійкість коду до можливих помилок. Але для того щоб вивести з матриці кодових відстаней одиниць, потрібно скоротити набір бінарних рядів, але тоді цього набору буде замало щоб поставити один бітовий ряд у відповідність одному символу нашого алфавіту. З'являється необхідність у збільшення довжини бітових рядів з 6-ти до 7-ми бітів. При 7-ми бітових рядах можна кодувати $57/7=8$ символів. Це задовольняє вимогам. В таблиці 3 вказано всі можливі бітові ряди, що задовольняють сформульованим умовам.

Табл. 3.

1	0001001	9	0011011	17	0110001
2	0001011	10	0011101	18	0110011
3	0001101	11	0100011	19	0110101
4	0010001	12	0100101	20	0110111
5	0010011	13	0100111	21	0111001
6	0010101	14	0101001	22	0111011
7	0010111	15	0101011		
8	0011001	16	0101101		

Матриця кодових відстаней новоутворених бітових рядів також містить одиниці. Після відбору з 22-х бітових рядів десять таких, щоб їх матриця кодових відстаней не містила одиниць отримаємо таблицю 4.

Табл. 4.

0	0001001	4	0011101	8	0110011
1	0010001	5	0100111	9	0110101
2	0010111	6	0101011		
3	0011011	7	0101101		

При використанні цих бітових рядів Нем-8 повинен мати наступну структуру:
 0xxxxx10xxxxx10xxxxx10xxxxx10xxxxx10xxxxx10xxxx
 xxx10xxxxx1nnnnnn, де n – контрольні біти.

Ця структура містить поки що один невеликий недолік - останні шість бітів можуть бути причиною деяких помилок. Наприклад: а) вище ми накладали обмеження на кількість підряд розміщених однакових бітів, на комбінацію pnnnnn в даному випадку таке обмеження не поширюється; б) в разі, якщо останній байт чи байти будуть рівні "1", то при графічному представленні цього коду вони будуть втрачені, оскільки їм буде відповідати білий колір і вони зіллються з фоном штрихкоду. Щоб запобігти цим небажаним явищам, вводяться додаткові нульові біти в кінець коду. Кінцева структура коду:

0xxxxx10xxxxx10xxxxx10xxxxx10xxxxx10xxxxx10xxx
 xxx10xxxxx101nn01nn01nn010

ВИСНОВКИ

Стійкість до помилок штрихкоду Нем-8 було перевірено за допомогою спеціально розробленої прикладної програми для ПК, яка

генерувала помилки та намагалася їх виявити чи виправити. Результати отримані за допомогою комп'ютера, підтвердили наші теоретичні розрахунки: при появі одного помилкового біта він виправляється завжди, при появі двох таких бітів будь-де в коді в 100% випадків фіксується наявність помилки та в 59% випадків виправляється. Помилки, які не виявляються, можуть з'явитись тільки при появі 4 помилкових бітів. Усе це свідчить про високу стійкість створеного штрихового коду.

Для порівняння штрихкоду Нем-8 з існуючими типами підрахуємо його ентропію. Нем-8 складається з 71 біта, кількість комбінацій чисел, які можна утворити 10 мільйонів (10^7), оскільки восьме число залежить тільки від перших семи, ми його не враховуємо. Кількість інформації, що несе в собі штрихкод Нем-8 рівна $I(a)=\log(10^7)$. В [3] наведена формула для знаходження величини надлишковості коду:

$$r = 1 - \frac{I(a)}{n \cdot \log L} \quad (1)$$

алфавіт двозначний $L=2$, кількість бітів $n=71$, тому

$$r = 1 - \frac{\log 10^7}{71 \cdot \log 2} = 0.67 \quad (2)$$

Коефіцієнт надлишковості штрихкоду EAN13:

$$r = 1 - \frac{\log 10^{12}}{95 \cdot \log 2} = 0.58 \quad (3)$$

Надлишковість в штриховому коді не є критичною, завдяки цій надлишковості коди забезпечуються більшою стійкістю до помилок, що при їх застосуванні більш важливо ніж надлишковість.

ЛІТЕРАТУРА

1. Шинон К. Работы по теории информации и кибернетике / Пер. с англ. под ред. Добрушина Р.Л., Лупанова О.Б. -М.: Издательство иностранной литературы, 1963. 830 с.
2. Котов П.А. Повышение достоверности передачи цифровой информации. -М.: Связь, 1966. 184 с.
3. Боховский В. А. Кибернетические системы в медицине. -М.: Наука, 1973. 328 с.
4. Cracking as an art, [Barcodes][Instant Access] / Lesson C(1). www.annon.penet.fi
5. Handbuch zum barcode-lesegerat / ABP-Programmierhandbuch fur Einbauscanner.
6. Barcodes FAQ. : www.azalea.com
7. Barcodes. : www.incom.ru/barcodes/s1.html