УДК 621.317

# ОСОБЛИВОСТІ ВИКОРИСТАННЯ ЗАВАДОЗАХИЩЕНИХ КОДІВ ДЛЯ ЗАКРИТТЯ ІНФОРМАЦІЇ ПРИ ПЕРЕДАВАННІ КОЛЕКТИВНИМИ ЛІНІЯМИ ЗВ'ЯЗКУ

***Кривогубченко С.Г., Компанець М.М., Кулик А.Я.***
***Вінницький державний технічний університет***

*The loading of shared lines, and first of all of telephone network, constantly grows. The large volumes of the information, which are transferred with the help of modems in conditions of a high level of industrial handicapes, require the effective protocols and algorithms of an exchange connected from coding, protected from handicapes.*

Завантаження ліній колективного користування, і в першу чергу телефонної мережі, постійно зростає. Великі обсяги інформації, що передаються за допомогою модемів в умовах високого рівня промислових завад, потребують ефективних протоколів та алгоритмів обміну, пов'язаних із завадозахищеним кодуванням.

Повідомлення, що має передаватися до лінії зв'язку, складається з інформаційної P та службової B частин

$$L = P + B . \qquad (1)$$

Перша з них вміщує ту інформацію, що має передаватися, а друга – службові повідомлення протоколу, маршрутизацію тощо.

Принцип завадозахищеного кодування полягає у створенні надлишковості – формуванні додаткових контрольних розрядів. В залежності від кодової відстані $d$ код може мати різне функціональне призначення:

$$d = r + s + 1 \quad (r \geq s) , \qquad (2)$$

де $r$ – кількість помилок, що визначаються;
$s$ – кількість помилок, що виправляються.
Для виправлення однієї помилки потрібен код із кодовою відстанню $d = 3$, а для виправлення двох помилок – із кодовою відстанню $d = 5$.

Таким чином після виконання завадозахищеного кодування довжина повідомлення **L** збільшується за рахунок контрольних розрядів **K**:

$$M = L + K . \qquad (3)$$

Найчастіше використовується блоковий принцип кодування, коли початкове повідомлення **L** розбивається на ряд блоків $L_1, L_2, L_3, \ldots, L_n$ і до кожного з них додається відповідна комбінація контрольних розрядів $K_1, K_2, K_3, \ldots, K_n$.

Оскільки більша частина інформації лініями колективного користування передається модемами з використанням персональних комп'ютерів, то виникають певні особливості, які необхідно враховувати:

- послідовні інтерфейси, що використовуються для передавання інформації до лінії зв'язку, здебільшого працюють з байтами. Навіть якщо кількість розрядів можна програмувати як для BIC Intel 8251, то порт все рівно доповнює слово нулями і до лінії зв'язку передається повних вісім розрядів;
- на сигнали, що передаються до лінії зв'язку суттєво впливають зовнішні чинники, які змінюються в залежності від пори року, часу доби, погодних умов, завантаженості лінії тощо і мають випадковий характер;
- використання лінійних (послідовних) кодів типу Манчестер II, AMI, BNZS, HDB3 тощо недоцільне, оскільки вони лише фіксують помилки, не виправляючи їх, вимагають суттєвих апаратних витрат і складно реалізуються програмно;
- аналіз похибок, що виникають під час передавання і спотворюють інформацію проведений недостатньо, причому проводити його необхідно для кожної конкретної місцевості окремо з урахуванням індивідуальних особливостей мережі.

Це призвело до того, що розмір блока кодової комбінації практично не визначався.

Попередній аналіз прийнятої інформації, що передається без використання завадозахищених кодів телефонною мережею міста Вінниці (Україна), показав, що на швидкостях 1200 та 2400 біт/с спотворюється приблизно один біт зі ста.

Використання завадозахищених кодів (циклічних, Хеммінга тощо) потребує оцінки кількості інформативних розрядів та кількості помилок, що мають виправлятися. Співвідношення інформаційних і контрольних розрядів для слів різної довжини наведене у таблиці 1.

*Таблиця 1 - Співвідношення інформаційних та контрольних розрядів*

| Довжина слова | $d = 3$ | | $d = 5$ | |
|---|---|---|---|---|
| | Кількість контрольних розрядів | Загальна кількість розрядів | Кількість контрольних розрядів | Загальна кількість розрядів |
| 1 | 2 | 3 | | |
| 2 | 3 | 5 | 4 | 6 |
| 3 | 3 | 6 | 5 | 8 |
| 4 | 3 | 7 | 9 | 13 |
| 5 | 4 | 9 | | |
| 6 | 4 | 10 | | |
| 7 | 4 | 11 | | |
| 8 | 4 | 12 | | |
| 9 | 4 | 13 | | |
| 10 | 4 | 14 | | |
| 11 | 4 | 15 | | |
| 12 | 5 | 17 | | |

Імовірність безпомилкового обміну інформацією розглядається для найпростішого випадку (незалежність помилок, що виникають у симетричному каналі).

Імовірність безпомилкового передавання буде визначатися:

$$p_{пр} = 1 - p_{пом}\,. \tag{4}$$

де $p_{пом}$ – імовірність помилки.

Для системи, що використовує принцип кодування із виправленням помилок, за формулою Бернуллі [1]:

$$p_{пр} = 1 - \sum_{i=k+1}^{m-1} C_m^i p_0^i (1 - p_0)^{m-1} - p_0^m\,, \tag{5}$$

де $p_0$ – імовірність помилки;

$k$ - кількість помилок, що виправляються кодом;

$m$ – загальна кількість символів.

За іншою методикою [2], розробленою Л. Пуртовим,:

$$p_{пр} = 1 - \frac{m}{k} p_0\,. \tag{6}$$

Результати розрахунків за обома методиками зведені до таблиці 2. Вони показують, що вже при восьми символах методика визначення імовірності за формулою Бернуллі не працює, оскільки імовірність помилки значно перевищує одиницю.

Розрахунки проводились з урахуванням попереднього аналізу імовірності спотворення символів ($p_0 \approx 0{,}01$). Аналіз підтверджує, що формувати код для виправлення двох і більше помилок недоцільно, оскільки імовірність спотворення за рахунок зростання кількості символів перекриває кількість виправлених помилок.

На ефективність передавання суттєво впливає час обміну. За несприятливих умов швидкість передавання не перевищує 4800 біт/с, а реально становить 2400 біт/с. Для цієї швидкості час передавання одного кілобіта (128 байт інформації) для слів різної довжини з урахуванням принципів роботи послідовних інтерфейсів наведений у таблиці 3.

Враховуючи особливості роботи послідовних інтерфейсів, доцільно перед передаванням інформації здійснювати перепаковування байтів, доповнюючи кількість інформаційних і контрольних розрядів до восьми, додаючи необхідну кількість з наступного байта. При цьому час передавання суттєво скорочується. Результати таких розрахунків за аналогічних умов зведені до таблиці 4.

Для наочності доцільно графіки розрахованих параметрів розташувати в одній системі координат. Їх аналіз дозволяє зробити певні висновки:

- для передавання інформації необхідно будувати код для виправлення однієї помилки;
- найбільш ефективним буде передавання напівбайтами з додаванням трьох контрольних розрядів до чотирьох інформаційних;

*Таблиця 2 - Імовірність безпомилкового передавання інформації для слів різної довжини*

| Кількість інформаційних розрядів | Загальна кількість розрядів | $d = 3$ | | $d = 5$ | |
|---|---|---|---|---|---|
| | | За формулою Бернуллі | За формулою Пуртова | За формулою Бернуллі | За формулою Пуртова |
| 1 | 3 | 0,99985 | 0,97 | | |
| 2 | 5 | 0,997 | 0,95 | 0,99993 | 0,985 |
| 3 | 6 | 0,983 | 0,94 | 0,89 | 0,96 |
| 4 | 7 | 0,88 | 0,93 | Методика не працює | 0,935 |
| 5 | 9 | Методика не працює | 0,91 | | |
| 6 | 10 | | 0,9 | | |
| 7 | 11 | | 0,89 | | |
| 8 | 12 | | 0,88 | | |
| 9 | 13 | | 0,87 | | |
| 10 | 14 | | 0,86 | | |
| 11 | 15 | | 0,85 | | |
| 12 | 17 | | 0,83 | | |

*Таблиця 3 - Час передавання слів різної довжини*

| Кількість розрядів | Час передавання, с ($d = 3$) | Кількість розрядів | Час передавання, с ($d = 5$) |
|---|---|---|---|
| 1 + 2 | 1,7 | | |
| 2 + 3 | 1,7 | 2 + 4 | 1,7 |
| 3 + 3 | 1,14 | 3 + 5 | 1,28 |
| 4 + 3 | 0,85 | 4 + 9 | 1,7 |
| 5 + 4 | 1,37 | | |
| 6 + 4 | 1,14 | | |
| 7 + 4 | 0,98 | | |
| 8 + 4 | 0,85 | | |
| 9 + 4 | 0,76 | | |
| 10 + 4 | 0,68 | | |
| 11 + 4 | 0,62 | | |
| 12 + 5 | 0,98 | | |

*Таблиця 4 - Час передавання слів різної довжини за умови перепакування байтів*

| Кількість розрядів | Час передавання, с ($d = 3$) | Кількість розрядів | Час передавання, с ($d = 5$) |
|---|---|---|---|
| 1 + 2 | 1,28 | | |
| 2 + 3 | 1,07 | 2 + 4 | 1,28 |
| 3 + 3 | 0,85 | 3 + 5 | 1,14 |
| 4 + 3 | 0,75 | 4 + 9 | 1,39 |
| 5 + 4 | 0,77 | | |
| 6 + 4 | 0,71 | | |
| 7 + 4 | 0,67 | | |
| 8 + 4 | 0,64 | | |
| 9 + 4 | 0,62 | | |
| 10 + 4 | 0,6 | | |
| 11 + 4 | 0,58 | | |
| 12 + 5 | 0,6 | | |

- перед передаванням необхідно перепаковувати інформацію, доповнюючи розряди з наступного байта до восьми.

Програмна реалізація дозволяє уникнути фіксованості коду, тобто розташування контрольних розрядів на певних місцях кодової комбінації. Виходячи з цього, виникає можливість досить простого шифрування інформації методом перестановки. При цьому можуть бути реалізовані певні особливості:

- можуть використовуватись алгоритми завадозахищеного кодування (циклічний, Хеммінга тощо), кожний з яких передбачає власний алгоритм формування контрольних розрядів;
- може змінюватись порядок передавання напівбайтів (молодший – старший чи навпаки);
- слово, що передається, може складатися з семи розрядів, а може доповнюватися до восьми з наступного байта. У будь-якому випадку паралельний інтерфейс доповнює кількість розрядів слова, що передається, до восьми і лише після цього передає інформацію до лінії зв'язку;
у сформованому байті можуть переставлятися розряди.

Кількість можливих неповторюваних кодових комбінацій в кожному випадку визначається кількістю у ній одиниць та нулів:

$$N_r = C_8^r = \frac{8!}{r! \cdot (8-r)!} \qquad (7)$$

де $r$ – кількість одиниць у кодовій комбінації.

У відповідності з цим, імовірність визначення кодової комбінації становить:

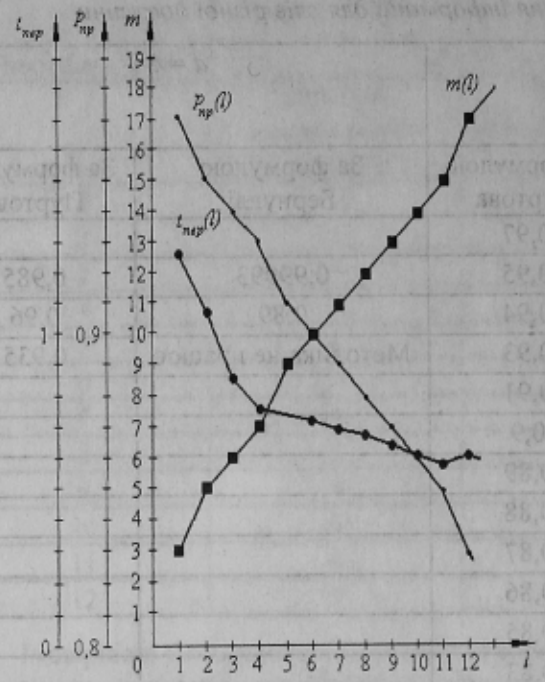


*Рисунок 1 – Залежність параметрів від довжини інформаційного слова*

$$p_r = \frac{1}{N_r} \quad (8)$$

Результати розрахунків цих параметрів для різних кодових комбінацій подані у таблиці 5.

З точки зору криптоаналітика кодові комбінації рівноімовірні, оскільки йому невідомі характер документа (текстовий, графічний, вимірювані значення тощо), формат, в якому спаковані дані за допомогою редактора чи іншого програмного продукту, а також інші початкові відомості. Виходячи з цього, імовірність розкриття байта становить:

$$p_b = p_r \cdot p_s = \frac{1}{N_r \cdot N_s} \quad (9)$$

де $p_s$ – імовірність появи одного символу;

$N_s$ – кількість символів в алфавіті передавання.

З урахуванням вищезгаданих особливостей формування завадозахищеного коду формула (9) набуде вигляду

$$p_б = \frac{1}{k_k \cdot k_n \cdot k_з \cdot N_r \cdot N_s} \quad (10)$$

*Таблиця 5 - Чисельні параметри для завадозахищених кодів*

| Кількість одиниць у кодовій комбінації, $n$ | $N_r$ | $p_r$ | $p_b$ | $p_б$ |
|---|---|---|---|---|
| 0 | 1 | 1 | $4 \cdot 10^{-3}$ | $5 \cdot 10^{-4}$ |
| 1 | 8 | 0,125 | $5 \cdot 10^{-4}$ | $6 \cdot 10^{-5}$ |
| 2 | 28 | 0,036 | $1,4 \cdot 10^{-4}$ | $1,75 \cdot 10^{-5}$ |
| 3 | 56 | 0,0178 | $6,95 \cdot 10^{-5}$ | $8,6 \cdot 10^{-6}$ |
| 4 | 70 | 0,014 | $5,46 \cdot 10^{-5}$ | $6,8 \cdot 10^{-6}$ |
| 5 | 56 | 0,0178 | $6,95 \cdot 10^{-5}$ | $8,6 \cdot 10^{-6}$ |
| 6 | 28 | 0,036 | $1,4 \cdot 10^{-4}$ | $1,75 \cdot 10^{-5}$ |
| 7 | 8 | 0,125 | $5 \cdot 10^{-4}$ | $6 \cdot 10^{-5}$ |
| 8 | 1 | 1 | $4 \cdot 10^{-3}$ | $5 \cdot 10^{-4}$ |

де $k_k$ – коефіцієнт, який показує кількість завадозахищених кодів, що можуть бути використані. В найпростішому випадку можна використати два коди – циклічний і Хеммінга;

$k_n$ – коефіцієнт, який показує кількість комбінацій під час передавання складових байта (молодший потім старший чи навпаки);

$k_з$ – коефіцієнт, який відповідає можливості комбінування слова (доповнювати розряди до восьми чи ні).

Для персонального комп'ютера алфавіт становить $2^8 = 256$ символів. Чотири інформаційних і три контрольних символи визначають шістнадцять кодових комбінацій (цифри 0H … FH). Але за рахунок доповнення кодової комбінації до восьми розрядів і їх перестановки можна отримати будь яке число з ряду 00H … FFH (0000B … 1111B).

Тоді, з урахуванням того, що коефіцієнти набувають значень 2:

$$p_б = \frac{1}{2 \cdot 2 \cdot 2 \cdot 256 \cdot N_r} = \frac{1}{2048 \cdot N_r} \quad (11)$$

Найпростішими для дешифрування є випадки, коли отримані комбінації становлять 00H чи FFH. Якщо не здійснювались інші дії крім доповнень та перестановок, то із впевненістю можна сказати, що вони характеризують відповідно цифри 0H та FH. Імовірність $p_б = 5 \cdot 10^{-5}$ пояснюється лише малою імовірністю

появи цих комбінацій і додатковими заходами, використаними для шифрування з метою ще більшого зниження імовірності появи цих комбінацій.

Найбільшу складність для дешифрування складають комбінації із рівною кількістю нулів та одиниць. Вони охоплюють 70 можливих кодових комбінацій. Тому доцільно розглянути ці два екстремальні випадки.

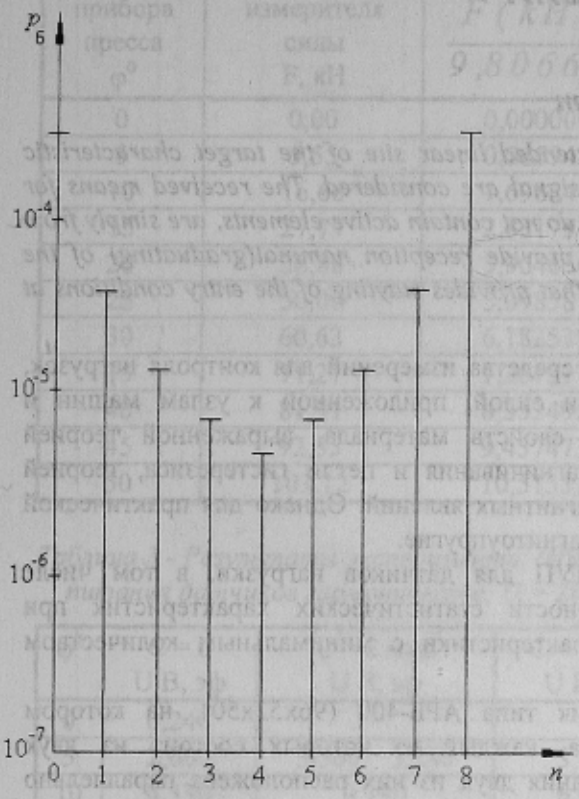


*Рисунок 2 – Імовірність дешифрування байта в залежності від його складу*

Кількість інформації, що отримується після оброблювання повідомлення можна визначити як різницю ентропій повідомлень:

$$I = H_1 - H_0 \quad (12)$$

$$H_0 = -p_0 \log_2 p_0 \quad (13)$$

$$H_1 = -p_1 \log_2 p_1 \quad (14)$$

Оскільки, після перестановок у кодових комбінаціях, їх змін не відбувається, то імовірності апріорна $p_0$ та апостеріорна $p_1$ складають одиницю.

Тоді:

$I = -1 \cdot \log_2 1 + 1 \cdot \log_2 1 = 0$

Тобто після дешифрування цих цифр інформація не отримується. Те, що цифри ідентифікуються як 0Н та FH було відомо і так, а відповіді на жодне із запитань, що стосуються особливостей, визначених вище (який алгоритм кодування використовувався, в якому порядку передавались інформаційні напівбайти тощо) отримати з цього неможливо.

Найбільше інформації надає дешифрування кодової комбінації з рівної кількості одиниць та нулів.

В першому наближенні алгоритм дешифрування можна розглядати як перебір можливих комбінацій. Оскільки чотири одиниці то чотири нулі у байті створюють найбільшу кількість кодових комбінацій, випадати вони будуть пропорційно частіше (при незалежності кодових комбінацій). Для попередніх розрахунків можна скористатися даними з таблиці. Імовірність $6{,}8 \cdot 10^{-6}$ передбачає розгляд приблизно 147059 кодових комбінацій, що займають 143 Кбайти пам'яті. Для визначення другого інформаційного напівбайта потрібно приблизно стільки ж. Тобто 280 Кбайт оперативної пам'яті потрібно лише для утримування можливих кодових комбінацій одного байта. Тобто пам'ять персонального комп'ютера дозволяє вмістити до 2000 переданих байт (при обсягу ОЗП 512 Мбайт).

З одного боку це багато, а з іншого одна сторінка тексту, набраного у редакторі Word версії 1997 року займає приблизно 30 Кбайт, причому більша частина цього обсягу – службові повідомлення, які не містять інформації для користувача. Час, який буде витрачений на дешифрування інформації з мережі колективного користування за рахунок старіння зводить її цінність практично до нуля.

Розглянуті розрахунки стосуються ситуації, коли не змінюється алгоритм перестановки. Якщо ввести ще декілька степенів свободи і змінювати алгоритм перестановки від байта до байта, принципово змінювати алгоритм через певну їх кількість (геометричний, матричний, гіперкубічний, транспозиції) тощо, то ту інформацію, що користувач психологічно може довірити лініям колективного користування, за реальний час дешифрувати дуже складно. У відповідності із теорією ненадійності кодів точка одиничності знаходиться на відстані:

$$H(K) = \log_2 b! \quad (15)$$

де $b$ – відстань транспозиції.

Транспозиція за рахунок доповнення розрядів займає 7 байтів. Виходячи з формули (15) точка одиничності знаходиться на відстані 12 байт. В межах цієї відстані алгоритм повинен змінюватись.

Таким чином для передавання інформації колективними лініями зв'язку використання завадозахищених кодів і досить простого алгоритму перестановки дає впевненість у збереженні її конфіденційності, без збільшення обсягу і зменшення завадозахищеності.

ЛІТЕРАТУРА

1. Shennon C. A mathematical theory of communication // Bell System Techn. J., 27(1948), № 3, pp. 379 – 423, 27(1948), № 4, pp. 623 – 656.

2. Элементы теории передачи дискретной информации / под ред. Пуртова Л.П. – М.: связь, 1972.Shennon C. Communication theory of secrecy systems // Bell System Techn. J., 28(1949), № 4, pp. 656 – 715.