Перелік посилань

1.Абазина Е.С. Цифровая стеганография: состояние и перспективы / Е.С. Абазина, А.А. Ерунов // Системы управления, связи и безопасности. – 2016. – № 2. – С. 182–201.

2.Колесов В.В. Применение дискретных хаотических алгоритмов в широкополосных телекоммуникационных системах / В.В. Колесов, А.И. Полубехин, Е.П. Чигин, А.Д. Юрин // Вестник СибГУТИ. – 2016. – № 3. – С. 77–92.

3.Барабаш О. В. Методи пошуку оптимальних маршрутів графа структури розгалуженої інформаційної мережі за заданим критерієм оптимальності при різних обмеженнях / О. В. Барабаш, І. П. Саланда, А. П. Мусієнко // Наукові записки Українського науково-дослідного інституту зв'язку. -К.: УНДІЗ, 2016. - №2 (42). - С 99-106.

Оптимальне кодування як засіб підвищення захищеності передачі шифрованих даних

Гончар Р. М., Нагребецький О.В., Орленко В.С., Чешун В.М.
Хмельницький національний університет

Постійне збільшення обсягів інформації в кіберпросторі і зростання її цінності зумовлює зацікавленість конкуруючих сторін і зловмисників у незаконному заволодінні нею, що створює постійну появу нових загроз щодо цілісності і конфіденційності інформації і актуальність заходів її захисту. Одним із основних способів захисту даних є шифрування, про що свідчить поява великої кількості методів та алгоритмів шифрування з різними функціональними можливостями і принципами дії (алгоритми DES-базовий, подвійний і потрійний DES, IDEA, ГОСТ 28147, Діффі-Хелмана, RSA тощо [1]) та тенденція до їх постійного вдосконалення.

Підвищення криптостійкості алгоритмів шифрування досягається як розробкою нових їх реалізацій, так і модернізацією-вдосконаленням існуючих або їх комбінуванням.

Проведені дослідження показують, що підвищення криптостійкості алгоритмів шифрування можна досягти попередньою підготовкою вхідних даних, в ході якого забезпечується порушення статистичних даних повторюваності символів вхідного тексту, тобто, збільшення характеристик його ентропії. Одним із варіантів такої підготовки вхідного тексту може бути застування методів оптимального нерівномірного кодування.

Для демонстрації можливості збільшення криптостійкості алгоритмів шифрування попередньою підготовкою вхідних даних оптимальним кодуванням обрано два класичних методи:

– кодування Хаффмена як класичний метод оптимального

ентропійного нерівномірного кодування даних, що дає стабільний оптимальний код на виході у відповідності до статистичних властивостей алфавіту вхідного тексту;

– шифри зсуву (заміни) як найпростіший варіант шифрування даних, що дозволяє наочно спостерігати вплив попередньої підготовки даних на криптостійкість алгоритму шифрування.

Недоліком шифрів зсуву є збереження статистичних характеристик появи символів первинного алфавіту (вхідного тексту) у вторинному алфавіті (зашифрованому тексті), що зумовлює їх низьку криптостійкість

Таким чином, криптостійкість шифрів зсуву може бути підвищена зміною властивостей алфавіту шифрування, для чого застосовується оптимальне кодування Хаффмена.

Стосовно кодування Хаффмена можна сказати наступне:

– оптимальне кодування Хаффмена є різновидом ефективного кодування;

– оптимальне кодування Хаффмена відноситься до класу ентропійних кодів;

– коди Хаффмена є нерівномірними кодами;

– коди Хаффмена є префіксними кодами;

– головною відмінністю коду Хаффмена від коду Шеннона-Фано є те, що він завжди дає оптимальний варіант ентропійного кодування за наявними статистичними даними;

– коди Хаффмена, як і інші методи ефективного кодування, безпосередньо не призначені для вирішення задач шифрування і захисту даних;

– стиснення даних оптимальним кодуванням Хаффмена дає позитивний ефект для захисту даних через зменшення розмірів повідомлень, що передаються (менша імовірність втрат інформації при передачі);

– відхилення варіанту кодування Хаффмена від примітивного кодування символів також ускладнює дешифрування тексту.

Для визначення перспектив застосування оптимального кодування Хаффмена в задачах криптографічного захисту дослідимо принципи цього кодування.

Побудова оптимального нерівномірного коду за методикою Хаффмена виконується за загальним алгоритмом, що включає наступні етапи:

– всі символи, що кодуються, упорядковуються в порядку зменшення ймовірностей;

– останні два символи впорядкованої множини (вони повинні мати найменші значення ймовірностей) замінюються допоміжним символом, значення ймовірності для якого визначається сумарною ймовірністю елементів, що замінюються;

– всі елементи нової множини знову упорядковуються на зменшення

ймовірностей;

– виконання попередніх двох операці повторюється до отримання єдиного допоміжного символу.

Для визначення кодових комбінацій символів виконується зворотний аналіз виконаних об'єднань.

Тобто, двом останнім символам, при об'єднанні яких було отримано символ з ймовірністю 1, присвоюються значення коду 0 і 1. Після цього розглядаються символи попереднього рівня, які прийняли участь в утворенні останніх допоміжних символів. Аналогічним чином їм ставляться у відповідність значення 0 та 1, які дописуються в молодший розряд кодових комбінацій.

Завершення кодування Хаффмена відбувається після досягнення етапу, на якому кодові комбінації будуть співставлені у відповідність всім символам вхідного алфавіту.

Отриманий за наведеним алгоритмом методикою нерівномірний код є оптимальним кодом Хаффмена для використаного в тексті алфавіту.

Для визначення перспектив застосування оптимального нерівномірного коду Хаффмена для збільшення криптостійкості алгоритмів шифрування розглянемо приклад практичного застосування методу кодування Хаффмена.

В якості прикладу дослідимо застосування методу Хаффмена для довільного набору з 100 латинських символів:

ADBCBADCSDCASCZDSMMMCCMMM
MAWSDSASWZDSMWUSCCMMMMEEE
WWBWBUDWSDWUSCCCCMMMMCWSD
SASCZDSMMMMMEWWZDSMMZDSMM

Потужність застосованого в наданому для кодування тексті алфавіту дорівнює 10, оскільки в ньому використано 10 символів: A, B, C, D, E, M, S, U, W, Z.

Спочатку визначимо рекомендовану розрядність двійкових кодів для заданого алфавіту при використанні примітивних (рівномірних) кодів:

$$n = \log_2 10 = 3.322.$$

Відповідно, для результату 3.322 отримуємо мінімально-достатню розрядність примітивного коду 4.

Для аналізу використаємо простий варіант послідовного кодування символів двійкового алфавіту двійковими числами (табл. 1.1).

Таблиця 1.1 – Примітивний код для кодування вхідного тексту

Символ алфавіту	A	B	C	D	E	M	S	U	W	Z
Код	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001

Із застосуванням наведеного в таблиці 1.1 примітивного рівномірного коду початковий текст перетворюється в масив двійкових кодів загальною розрядністю 400 біт:

0000 0011 0001 0010 0001 0000 0011 0010 0110 0011
0010 0000 0110 0010 1001 0011 0110 0101 0101 0101
0010 0010 0101 0101 0101 0101 0000 1000 0110 0011
0110 0000 0110 1000 1001 0011 0110 0101 1000 0111
0110 0010 0010 0101 0101 0101 0101 0100 0100 0100
1000 1000 0001 1000 0001 0111 0011 1000 0110 0011
1000 0111 0110 0010 0010 0010 0010 0101 0101 0101
0101 0010 1000 0110 0011 0110 0000 0110 0010 1001
0011 0110 0101 0101 0101 0101 0101 0100 1000 1000
1001 0011 0110 0101 0101 1001 0011 0110 0101 0101

Дослідження статистичних властивостей вхідного тексту дозволяє визначити ймовірності появи в ньому символів алфавіту (табл. 1.2).

Таблиця 1.2 – Статистичні характеристики алфавіту вхідного тексту

Символ алфавіту	A	B	C	D	E	M	S	U	W	Z
Кількість повторів символу	6	4	14	12	4	25	16	3	11	5
Статистична імовірність	0.06	0.04	0.14	0.12	0.04	0.25	0.16	0.03	0.11	0.05

Наявність статистичних даних щодо імовірностей входження символів алфавіту до тексту, що подається на кодування, дозволяє застосувати алгоритм кодування Хаффмена. На рис. 1.1 зображене кодове дерево Хаффмена, побудоване за даними таблиці 1.2.

На основі кодового дерева Хаффмена формуємо кодові комбінації оптимального нерівномірного коду Хаффмена для кодування тексту (табл. 1.3).

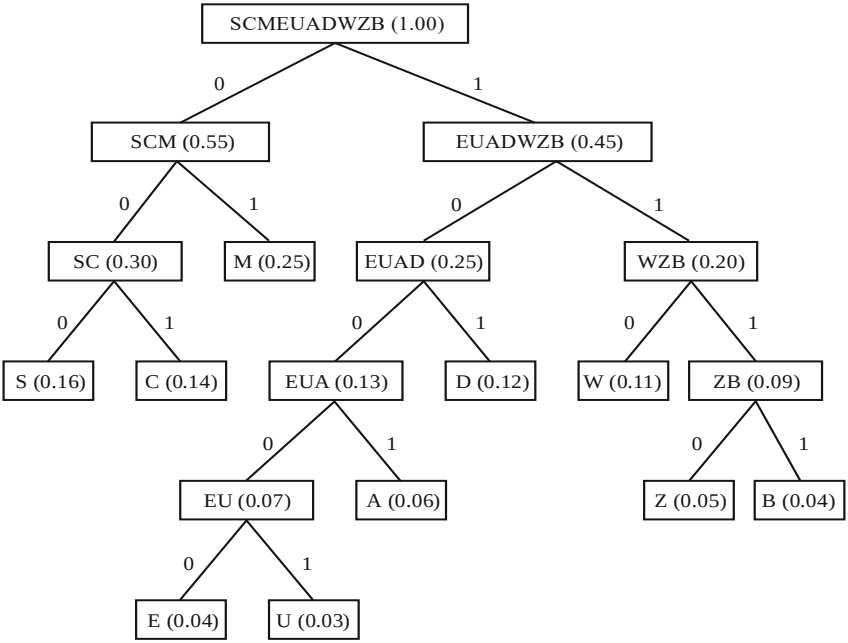

Рисунок 1.10 – Кодове дерево Хаффмена

Таблиця 1.3 – Оптимальний код Хаффмена для кодування алфавіту тексту

Символ алфавіту	A	B	C	D	E	M	S	U	W	Z
Код Хафмена	1001	1111	001	101	10000	01	000	10001	110	1110

Із застосуванням наведеного в таблиці 1.3 оптимального нерівномірного коду Хаффмена початковий текст перетворюється в масив двійкових кодів загальною розрядністю 304 біт:

1001 101 1111 001 1111 1001 101 001 000 101
001 1001 000 001 1110 101 000 01 01 01
001 001 01 01 01 01 1001 110 000 101
000 1001 000 110 1110 101 000 01 110 10001
000 001 001 01 01 01 01 10000 10000 10000
110 110 1111 110 1111 10001 101 110 000 101
110 10001 000 001 001 001 001 01 01 01
01 001 110 000 101 000 1001 000 001 1110
101 000 01 01 01 01 01 10000 110 110
1110 101 000 01 01 1110 101 000 01 01

Першочергово відзначимо позитивний ефект стиску даних із застосуванням оптимального нерівномірного кодування Хаффмена – замість 400 біт даних сам текст в оптимальному кодуванні займає лише 304 біти.

Визначимо коефіцієнт стиснення коду:

$$k = \frac{304}{400} * 100\% = 76\%.$$

Отриманий коефіцієнт є досить високим показником для несистематизованого тексту і дозволяє стверджувати, що ризик ушкодження оптимального коду тексту порівняно з початковим примітивним зменшується майже на чверть (24% пропорційно зменшенню розрядності коду). Це підтверджує ефективність використання оптимального кодування Хаффмена за призначенням для кодування тексту даного прикладу.

Для визначення перспектив застосування оптимального кодування Хаффмена для підвищення ефективності захисту даних з використанням шифрів зсуву перетворимо отриманий нерівномірний код початкового тексту в рівномірні кодові комбінації, потрібні нам для реалізації шифру зсуву.

Початково формуємо потокове (нерозривне послідовне) представлення тексту кодом Хаффмена. Для наочності і зручності подальшого аналізу коди Хаффена окремих символів з непарними номерами позицій в тексті виділено жирним шрифтом (застосовано виділення жирним символів через один для можливості наочного розрізнення кодів символів в бітовій послідовності):

1001101**1111**001**1111**1001**101**001**000**101**001**1001**000**001**1110**101**000**01**0**101**001**
001**01**01**01**1**001**110**000**101**000**1001**000**110**1110**101**000**01**110**10001**000**001**001**
01**01**01**1**0000**10000**10000**110**110**1111**110**1111**10001**101**110**000**101**110**10001**0**
00001**001**001**001**01**01**01**001**110**000**101**000**100**1000**001**1110**101**000**01**0**101**01**0**
110000**110**110**1110**101**000**01**01**1110**101**0000**01**.

Розіб'ємо отримане потокове представлення тексту кодом Хаффмена на чотирьохрозрядні комбінації (тетради) у відповідності із розмірністю початкового варіанту примітивного кодування:

1001 101**1** **111**0 01**11** **11**10 01**10** **1**001 **000**1 01**00** **1**100
1**000** 001**1** **110**1 01**00** **0**01**0** **1**01**0** **01**00 1**0**1**0** 1**0**1**0** 1**100**
1110 **000**1 01**00** **0**100 1**000** 110**1** **110**1 01**00** **0**011 **10**10
001**0** **00**00 1**001** 01**01** 01**01** 1000 0**100** **00**10 000**1** **10**11
0**111** **1**110 **1111** 1000 1**101** 110**0** **00**10 1**110** 1000 1**000**
001**0** **01**00 1**001** 01**01** 01**01** 001**1** **1**000 0**101** 000**1** **001**0
00**00** **1**111 0**101** 000**0** **1**01**0** **1**01**0** **1**100 00**11** **0**110 **1110**
101**0** **00**01 **0**111 10**10** **1**000 **01**01

Аналіз отриманого розбиття дозволяє побачити певні особливості утворення рівномірного коду з нерівномірного:

– лише невелика кількість отриманих кодових комбінацій

отриманого рівномірного коду відповідає кодовим комбінаціям символів у реалізації коду Хаффмена (однократно зустрічаються коди 1001, 1110 і 1111);

– визначені в попередньому аналізі кодові комбінації коду Хаффмена 1001, 1110 і 1111 в рівномірному коді наявні не лише як коди символів A, B і Z відповідно – аналогічні кодові комбінації утворюються з фрагментів кодів інших символів;

– в більшості випадків кодові комбінації коду Хаффмена при переході від нерівномірного кодування до рівномірного дробленням бітової послідовності зазнають дроблення на частини між декількома кодовими комбінаціями рівномірного коду (кожна з комбінацій 101**1, 111**0, 01**11**, **11**10**,** 01**10**, як і більшість інших, утворені ділять між собою фрагменти двох кодових комбінацій коду Хаффмена);

– наслідком попередньої властивості є те, що більшість кодових комбінацій рівномірного коду, отриманого дробленням бітової послідовності коду Хаффмена, містять фрагменти декількох кодових комбінацій коду Хаффмена (кожна з комбінацій 101**1, 111**0, 01**11, 11**10, 01**10,** як і більшість інших, утворені з фрагментів двох кодових комбінацій коду Хаффмена, а комбінації **0**01**1** і **0**01**1,** як приклад, утворені з фрагментів трьох кодових комбінацій);

– серед кодових комбінацій рівномірного коду, отриманого дробленням бітової послідовності коду Хаффмена, зустрічаються кодові комбінації, які є фрагментами кодових комбінацій коду Хаффмена більшої розрядності (кодова комбінація 1000 в одному місці отримана урізанням з комбінації 10000, а в іншому - з комбінації 10001). При цьому можливе виокремлення частини розрядів як зі сторони молодших розрядів, так і зі сторони старших або з середньої частини довгої кодової комбінації коду Хаффмена.

Більш детальний аналіз рівномірних кодів, отриманих дробленням бітової послідовності оптимального нерівномірного коду Хаффмена на тетради, дозволяє визначити збільшення кількості наявних кодових комбінацій порівняно з початковим примітивним кодуванням.

В наведеному нижче представленні підкреслено різні види кодових комбінацій, отриманих дробленням бітової послідовності оптимального нерівномірного коду Хаффмена на тетради:

1001 1011 1110 0111 1110 **0110** 1001 **0001 0100 1100**
1000 0011 1101 0100 **0010 1010** 0100 1010 1010 1100
1110 0001 0100 0100 1000 1101 1101 0100 0011 1010
0010 **0000** 1001 **0101** 0101 1000 0100 0010 0001 1011
0111 1110 **1111** 1000 1101 1100 0010 1110 1000 1000
0010 0100 1001 0101 0101 0011 1000 0101 0001 0010
0000 1111 0101 0000 1010 1010 1100 0011 0110 1110
1010 0001 0111 1010 1000 0101

З прикладу чітко видно, що в отриманих дробленням бітової послідовності оптимального нерівномірного коду Хаффмена на тетради кодових комбінаціях наявні всі 16 можливих кодових комбінацій довжиною 4 біти, а не 10, як це було в початковому примітивному коді.

В таблиці 1.4 наведено статистичні дані щодо частоти появи різних кодових комбінацій рівномірного коду, отриманих дробленням бітової послідовності оптимального нерівномірного коду Хаффмена на тетради.

Таблиця 1.4 – Статистичні характеристики фінального рівномірного коду

№ з/п	Код	Кількість повторів коду	Статистична імовірність
1.	0000	3	0,039474
2.	0001	5	0,065789
3.	0010	6	0,078947
4.	0011	4	0,052632
5.	0100	8	0,105263
6.	0101	7	0,092105
7.	0110	2	0,026316
8.	0111	3	0,039474
9.	1000	8	0,105263
10.	1001	4	0,052632
11.	1010	8	0,105263
12.	1011	2	0,026316
13.	1100	4	0,052632
14.	1101	4	0,052632
15.	1110	6	0,078947
16.	1111	2	0,026316

Порівняння статистичних даних таблиць 1.2 і 1.4 свідчить про руйнування застосованою процедурою оптимального нерівномірного кодування Хаффмена статистичних залежностей між кодовими комбінаціями, що співставляються символам вхідного тексту при примітивному кодуванні, а також про здатність відповідної процедури змінювати кількість використовуваних комбінацій рівномірного коду і характер їх формування.

Невідповідність кількості кодових комбінацій застосовуваного коду кількості символів кодованого алфавіту та руйнування статистичних залежностей і ентропійних взаємозв'язків між символами є передумовою підвищення криптостійкості шифрування зсувом і може бути позитивно застосоване для криптографічних методів шифрування загалом.

Перелік посилань

1 Мережні інформаційні технології: навчальний посібник / О. А. Мясіщев, В. М. Джулій, С. Р. Красильников, В. М. Чешун. – Хмельницький : ХНУ, 2012. – 422 с.

2 Курко А. М. Введення в теорію інформації: посібник до вивчення дисципліни / А. М. Курко, В. Я. Решетник – Тернопіль : Тернопільський національний технічний університет імені Івана Пулюя, 2017. – 108 с.

3 Беркман Л.Н. Основні поняття та теореми теорії інформації: навчальний посібник для самостійної роботи студентів ВНЗ / Беркман Л.Н., Комарова Л.О., Чумак О.І. – Київ: ДУТ ННІТІ, 2015. – 91 с.

4 Класифікація основних методів кодування і кодів [Електронний ресурс] / Портал «stud.com.ua». – Режим доступу: https://stud.com.ua/171429/tehnika/klasifikatsiya_osnovnih_metodiv_koduvannya_kodiv (дата звернення 30.10.2020). – Назва з екрана.

5 Захист інформації в комп'ютерних системах та мережах: навчальний посібник / С.Г.Семенов, А.О.Подорожняк, О.І.Баленко, С.Ю.Гавриленко. – Х.: НТУ «ХПІ», 2014. – 251 с.

6 Тарнавський Ю. А. Технології захисту інформації: підручник / Ю. А. Тарнавський; КПІ ім. Ігоря Сікорського. – Київ : КПІ ім. Ігоря Сікорського, 2018. – 162 с.

7 Помехоустойчивое кодирование и декодирование в дискретных КПС [Електронний ресурс] / Портал «wiki». – Режим доступу: https://ru.bmstu.wiki/Помехоустойчивое_кодирование_и_декодирование_в_дискретных_КПС (дата звернення 30.10.2020). – Назва з екрана.

8 Остапов С. Е. Технології захисту інформації : навчальний посібник / С. Е. Остапов, С. П. Євсеєв, О. Г. Король. – Х. : Вид. ХНЕУ, 2013. – 476 с.

9 Майданюк В. П. Кодування та захист інформації. навчальний посібник / В. П. Майданюк. – Вінниця:ВНТУ, 2009. – 164 с.

10 Бойко Ю. М. Теоретичні аспекти підвищення завадостійкості й ефективності обробки сигналів в радіотехнічних пристроях та засобах телекомунікаційних систем за наявності завад : монографія / Ю. М. Бойко, В. А. Дружинінін, С. В. Толюпа. - Київ : Логос, 2018. - 227 с.

11 Прикладна криптологія : системи шифрування : підручник / О. Г. Корченко, В. П. Сіденко, Ю. О. Дрейс. – К. : ДУТ, 2014. – 448 с.

12 Кодування джерел інформації та каналів зв'язку: навчальний посібник / [Беркман Л.Н., Бондарчук А.П., Гайдур Г.І., Чумак Н.С.]. – Київ: ННІТІ ДУТ, 2018. – 91 с.

13 Alencar & Marcelo S Information theory / Alencar & Marcelo S. – NEWYORK : Momentum Press , LLC, 2015. – 178 p.

14 Galić I. Image compression with B-tree coding algorithm enhanced by data modelling with Burrows-Wheeler transformation / Irena Galić, Časlav Livada, Branka Zovko-Cihlar. //Journal for Control, Measurement, Electronics, Computing and Communications. – 2017. – Volume 57, Issue 1. – P. 76-88.