Тема 22. Кодування стійке до завад

В середині 40-х років 20 ст. Річард Геммінг працював у Лабораторіх Белла (Bell Labs) на обчислювальній машині Ball Model V. Це була електромеханічна машина, яка використовувала релейні блоки, швидкість яких була дуже низька: один оберт за декілька секунд. Дані вводились в машину за допомогою перфокарт і тому в процесі зчитування часто відбувались помилки. В робочі дні використовувались спеціальні коди, які виявляли та виправляли знайдені помилки, при цьому оператор дізнавався про помилку по спеціальному світовому сигналу, виправляв та запускав машину. У вихідні дні, коли не було операторів, при виникненні помилки машина автоматично виходила з програми та запускала іншу.

Геммінг часто працював у вихідні, і тому все більше дратувався, тому що часто повинен був перевантажувати свою програму через ненадійність перфокарт. Протягом декількох років він проводив багато часу над побудовою ефективних алгоритмів виправлення помилок. У 1950 році він опублікував спосіб, який сьогодні називають код Геммінга.

22.1. Коди стійкі до перешкод

Розглянемо один частковий випадок рівномірного двійкового кодування, коли $A = B = \{0, 1\}$. Розглянемо схему рівномірно кодування $\sigma_{k,n}$:

$$\alpha_1 \to \beta_1,$$
$$\alpha_2 \to \beta_2,$$
$$\dots$$
$$\alpha_{2^k} \to \beta_{2^k},$$

де α_i, β_i – відповідно слова довжиною k та n, $n>k$. Говорять, що схему $\sigma_{k,n}$ задає код $V = \{\beta_1, \beta_2, \dots, \beta_{2^k}\} \subset E_2^n$.

Введемо додаткові позначення. Елементи множини E_2^n (двійкові вектори довжиною n) позначатимемо великими латинськими буквами $X, Y, Z, \dots$, а їх компоненти – відповідними малими буквами з індексами. Зокрема, елементарні коди позначатимемо як традиційно ($\beta_1, \beta_2, \dots$), так і X, Y, Z залежно від контексту.

Означення 22.1. **Нормою** $\|X\|$ двійкового вектора $X = x_1x_2\dots x_n$ називають число, яке дорівнює кількості його одиничних компонент. Отже:

$$\|X\| = \sum_{i=1}^{n} x_i$$

Припустимо, що в каналі зв’язку діє **джерело адитивних перешкод**, яке описують множиною $P(n, t)$. Її елементи – двійкові вектори-помилки. При чому на n переданих поспіль двійкових символів припадає не більше ніж t помилок. Це означає, що коли на вході каналу зв'язку передано повідомлення α, то на виході можна отримати будь-яке слово з множини $\{\alpha \oplus \gamma \mid \gamma \in P(n, t), |\alpha| = |\gamma|\}$, де $\alpha \oplus \gamma$ - поразрядне додавання за mod 2.

Позаяк проблема локалізації інформації (тобто розділення закодованого повідомлення на елементарні коди) у моделі рівномірного кодування тривіальна, то виявлення помилок полягає у відшуканні незбігу локалізованої групи n символів, яке не відповідає жодному з елементарних кодів. Якщо через помилку елементарний код перейде в інший елементарний код, то помилку не буде виявлено. Іноді можна виправити помилку. Якщо групу бітів локалізовано правильно, то для цього необхідно й достатньо, щоб помилкова група була «синонімом» єдиного елементарного коду.

Канал зв'язку називають **надійним**, якщо будь-які помилки можна виявити чи виправити відповідно до заданої мети декодування. Далі наведено головні положення побудови кодів, які забезпечують надійність найпростіших каналів зв'язку.

Означення 22.2. **Віддалю Геммінга** називають функцію $\rho(X, Y)$ двох змінних, означену на множині E_2^n:

$$\rho(X,Y) = \sum_{i=1}^{n} (x_i \oplus y_i).$$

Тобто, віддаль Геммінга дорівнює кількості розрядів, у яких вектори X та Y не збігаються.

Означення 22.3. **Скалярний добуток** двійкових векторів $X, Y \in E_2^n$ означається як:

$$\langle X,Y \rangle = \sum_{i=1}^{n} x_i y_i$$

Відповідно він дорівнює кількості розрядів, у яких X та Y збігаються й дорівнюють 1.

Легко перевірити таке співвідношення:

$$\rho(X,0) = \|X\| = \sum_{i=1}^{n} x_i , \tag{22.1}$$

де 0 – n-вимірний вектор із нульовими компонентами;

$$\rho(X,Y) = \|X \oplus Y\|, \tag{22.2}$$

$$\rho(X \oplus Z, Y \oplus Z) = \rho(X,Y), \tag{22.3}$$

$$\rho(X,Y) = \|X\| + \|Y\| - 2\langle X,Y \rangle . \tag{22.4}$$

Для віддалі Геммінга виконуються аксіоми метрики:

- $\rho(X, Y) \geq 0$, причому $\rho(X, Y) = 0$ в тому лише випадку, коли $X = Y$;
- $\rho(X, Y) = \rho(Y, X)$;
- $\rho(X, Y) + \rho(Y, Z) \geq \rho(X, Z)$ (нерівність трикутника).

Метрика Геммінга – зручне математичне поняття для формулювання умов надійності кодування в разі адитивних помилок.

Означення 22.4. Нехай систему $\sigma_{k,n}$ задано кодом $V = \{\beta_1, \beta_2, \ldots, \beta_{2^k}\}$. **Кодовою віддалю** для коду V називають величину:

$$\rho(V) = \min \{\rho(X,Y) \mid X, Y \in V, X \neq Y\}.$$

Теорема 22.1. Якщо в каналі зв'язку діє джерело адитивних перешкод $P(n, t)$, то правдиві такі твердження.

1. Для виявлення будь-яких помилок необхідно й достатньо, щоб $\rho(V) > t$.
2. Для виправлення будь-яких помилок необхідно й достатньо, щоб $\rho(V) > 2t$.

Доведення. 1. Нехай $\rho(V) > t$. Якщо $X \in V$, $Z \in P(n, t)$, $Z \neq 0$, то, використовуючи спочатку рівність (22.3), а потім – (22.1), можемо записати $\rho(X, X \oplus Z) = \rho(X \oplus X, X \oplus X \oplus Z) = \rho(0, Z) = \|Z\| \leq t$. Отже, $X \oplus Z \notin V$, і помилку виявлено.

Навпаки, нехай $\rho(X, Y) \leq t$ й $X, Y \in V$, $X \neq Y$. Тоді, застосувавши рівність (22.2), маємо $\|X \oplus Y\| = \rho(X, Y) \leq t$; отже, $Z = X \oplus Y \in P(n, t)$. Звідси випливає, що $X \oplus Z = Y$, тобто помилку в елементарному коді Y виявити не можна.

2. Нехай $\rho(V) > 2t$. Якщо $X \in V$, $Z \in P(n, t)$, то X – єдиний елементарний код із V, який міг перейти внаслідок помилки в $X \oplus Z$. Справді, припустимо, що існує такий двійковий вектор $Y \neq X$, що $Y \in V$ та $Y \oplus Z_1 = X \oplus Z$ для якогось $Z_1 \in P(n, t)$. Додавши до обох частин останньої рівності $X \oplus Z_1$, отримаємо $Y \oplus X = Z_1 \oplus Z$. Але, згідно з рівністю (22.2) можемо записати $\|X \oplus Y\| = \rho(X, Y) > 2t$, а $\|Z_1 \oplus Z\| \leq \|Z_1\| + \|Z\| \leq 2t$. Одержали суперечність.

Навпаки, нехай $\rho(X, Y) \leq 2t$ для якихось різних $X, Y \in V$. Тоді $\|X \oplus Y\| \leq 2t$, й існують такі двійкові вектори Z_1, Z_2, що $\|Z_1\| \leq t$, $\|Z_2\| \leq t$ (тобто вони належать $P(n, t)$) та $X \oplus Y = Z_1 \oplus Z_2$. Додавши до обох частин останньої рівності $Y \oplus Z_1$, одержимо $X \oplus Z_1 = Y \oplus Z_2 = W$. Отже, у разі отримання спотвореного елементарного коду W неможливо виявити, що було передано насправді: X чи Y. ►

Доведене твердження має геометричну інтерпретацію.

Означення 22.5. Множину $S_t(X) = \{Z \mid \rho(X, Z) \le t\}$ називають **кулею радіусом *t* з центром у точці *X***.

Теорема 22.2 (без доведення). Якщо в каналі зв'язку діє джерело адитивних перешкод $P(n, t)$, то правдиві такі твердження.

1. Для виявлення будь-яких помилок необхідно й достатньо, щоб для будь-якого $X \in V$ куля $S_t(X)$ не містила інших елементарних кодів, окрім X.

2. Для виправлення будь-яких помилок необхідно й достатньо, щоб для будь-яких $X, Y \in V$ було виконано умову $S_t(X) \cap S_t(Y) = \varnothing$.

Означення 22.6. Рівномірне кодування $\sigma_{k,n}: \alpha_i \to \beta_i$ $(i = 1, 2,\ldots, 2^k)$ називають **систематичним**, якщо можна виділити множину k розрядів $I = \{i_1,\ldots,i_k\} \subset \{1,2,\ldots,n\}$, які називаються **інформаційними**, так, що коли $\beta_i = x_1\ldots x_n$ $(i = 1, 2,\ldots, 2^k)$, то $\alpha_i = x_{i1}\ldots x_{ik}$. Решту розрядів у такому разі називають **контрольними**.

22.2. Коди з самоконтролем

Як було показано в теоремі 22.1 для автоматичного виявлення помилок достатньо, щоб кодова віддаль $\rho(V)$ була більша за кількість можливих помилок t на n переданих бітів інформації. Розглянемо кодування таблиці символів ASCII, в якій представлено 256 записів. Для кодування елементів таблиці використовується один байт – вісім бітів. Легко бачити, що кодова віддаль $\rho(V)$ для таблиці ASCII буде складати 1 (наприклад, відстань між кодом 00000000 та 00000001 складає якраз 1). Таким чином, якщо $t = 1$, то ми не зможемо автоматично виявляти помилки. Отже, $\rho(V)$ повинно бути не менше 2. Як цього можна досягнути?

На практиці використовується наступна схема. До кожного інформативного байту (8 розрядів) додається ще один розряд, який називається контрольним. Значення контрольного розряду встановлюється таким чином, щоб загальна кількість одиничних розрядів була парною (враховуючи як інформативні розряди, так і контрольні). Тоді, замість 8 розрядів буде передаватись 9. В попередньому прикладі ми отримуємо коди 000000000 та 000000011 (контрольний розряд позначений жирним). Як бачимо, в даному прикладі відстань між кодами вже становить 2, що призведе до автоматичного виявлення помилки.

Такий підхід дозволяє виявляти помилки, які відбулись в одному розряді, але не працює для помилок в більшій кількості розрядів. Наприклад, якщо ми хочемо передати код 00000000, то повинні додати ще контрольний розряд 0, і, зрештою, відправити повідомлення 000000000. Якщо помилки відбулись, припустимо, в двох розрядах і ми отримали, наприклад, код 100010000. В такому випадку кількість одиниць є парною і контрольний розряд містить 0, що є коректним з точки побудови коду. Отже, помилки не зможуть бути виявлені. Звісно, також даний підхід не дозволяє з'ясувати в якому саме розряді відбулась помилка.

22.3. Коди з самовиправленням

Нижче розглянемо коди, які дозволяють виправлення при одиничній помилці (тобто коли $t = 1$).Можна показати, що кількість контрольних розрядів k повинна бути обрана такою, щоб виконувалась рівність: $2^k \ge k + n + 1$ або $k \ge \log_2(k + n + 1)$, де n – кількість інформативних розрядів. Мінімальні значення k для заданого n наведені в таблиці 22.1

Діапазон n	Мінімальне k
1	2
2 – 4	3
5 – 11	4
12 – 26	5
27 – 57	6

Табл. 22.1

Перевіримо справедливість наведеної кількості контрольних кодів для найпростіших випадків. Розглянемо $n = 1$. Зрозуміло, що можливі тільки два коди: X_1=0 та X_2=1. В даній

системі кодова віддаль $\rho(V)$=1. За теоремою 22.1 $\rho(V)$ повинно бути не менше 3 ($\rho(V)$>2), щоб код дозволяв автоматичне виправлення одиничних помилок. Якщо до кодів X_i додати тільки один контрольний розряд, то відстань складатиме лише 2 (X_1=0**0** та X_2=1**1**). Отже, необхідно додати мінімум 2 контрольних розряди, щоб уможливити автоматичну корекцію одиничних помилок: X_1=0**00** та X_2=1**11**. Дійсно, припустимо ми передавали код X_1 і відбулась помилка в другому розряді – ми отримали код X'=0**10**. Знаючи, що помилка можлива тільки в одному розряді, ми легко бачимо, що код X' міг бути отриманий тільки з коду X_1, адже від коду X_2 його віддаляє дві помилки (потрібно змінити два розряди в X_2 щоб отримати X').

Тепер розглянемо n = 2. Тут можуть існувати чотири різні коди: X_1=00, X_2=01, X_3=10, X_4=11. За теоремою 22.1 $\rho(V)$ знову повинно бути не менше 3, тобто коди повинні відрізнятись як мінімум трьома розрядами. З табл. 22.1 видно, що кількість контрольних розрядів повинна дорівнювати 3, щоб дозволити автоматичну корекцію. Нижче наводиться одне з можливих приписувань значень контрольним розрядам, яка задовольняє умови теореми 22.1.

X_1: 00**000**
X_2: 01**011**
X_3: 10**101**
X_4: 11**110**

Відповідні значення відстаней між кодами становитимуть: $\rho(X_1,X_2)$ = 3, $\rho(X_1,X_3)$ = 3, $\rho(X_1,X_3)$ = 4, $\rho(X_2,X_3)$ = 4, $\rho(X_2,X_4)$ = 3, $\rho(X_3,X_4)$ = 3. Отже, $\rho(V)$ = 3.

Загалом можлива наступна схема визначення значень контрольних розрядів у випадку виправлення одиночних помилок. Припишемо кожному з розрядів (інформативним та контрольним) свій власний номер від 1 до n+k. Запишемо ці номери у двійковій системі. Оскільки $2^k > n+k$, то кожний номер можна представити k-розрядним двійковим числом.

Припустимо далі, що всі n+k розрядів коду розбиті на контрольні групи, які частково перетинаються, причому так, що одиниці у двійковому представленні номеру розряду вказують на його приналежність до певної контрольної групи. Наприклад, розряд №5 належить до першої та третьої контрольним групами, тому що у двійковому представленні його номеру 5_{10} = …000101_2 1й та 3й розряди містять одиниці.

Серед n+k розрядів при цьому є k розрядів, кожний з яких належить тільки одній контрольній групі: 1_{10} = …000001_2, 2_{10} = …000010_2, 4_{10} = …000100_2, … Ці k розрядів будемо вважати контрольними. Інші n розрядів, кожний з яких належить принаймні двом контрольним групам, будуть інформативними розрядами. В кожній контрольній групі буде по одному контрольному розряду. При цьому в кожний контрольний розряд помістимо таку цифру (0 або 1), щоб загальна кількість одиниць в його контрольній групі була парною.

Розглянемо випадок, коли n=8. За табл. 22.1 маємо, що k=4. Нехай початковий інформативний байт (без контрольних розрядів) є 01101011. Позначимо k_i i-й контрольний розряд та n_i i-й інформативний розряд.

№ розряду	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100
	k_1	k_2	n_1	k_3	n_2	n_3	n_4	k_4	n_5	n_6	n_7	n_8
Інформаційне кодове слово			0		1	1	0		1	0	1	1
k_1	**1**		0		1		0		1		1	
k_2		**0**	0			1	0			0	1	
k_3				**1**	1	1	0					1
k_4								**1**	1	0	1	1
Кодове слово з контрольними розрядами	**1**	**0**	0	**1**	1	1	0	**1**	1	0	1	1

Як видно, всі контрольні групи перекриваються між собою. Наприклад, перша група контролює розряди n_1, n_2, n_4, n_5 та n_7. Друга група – n_1, n_3, n_4, n_6 та n_7. Очевидно, що вони перетинаються.

Припустимо тепер, що при передачі даного коду 100111011011 відбулась помилка в 11-му розряді і ми отримали код 100111011001. Виконавши перевірку парності одиниць в кожній контрольній групі, ми виявили, що кількість одиниць стала непарною в групах №1, №2 та №4, а в групі №3 лишилась парною. Це, по-перше, вказує на наявність помилки, а, по-друге, означає, що номер помилкового розряду містить у двійковому представленні одиниці на першому, другому та четвертому місці та нуль на третьому місці, адже в третій групі помилку не виявлено (а тому помилковий розряд не входить у третю групу).

№ розряду	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	Парність
	k_1	k_2	n_1	k_3	n_2	n_3	n_4	k_4	n_5	n_6	n_7	n_8	
Передане слово	**1**	**0**	0	**1**	1	1	0	**1**	1	0	1	1	
Отримане слово	**1**	**0**	0	**1**	1	1	0	**1**	1	0	0	1	
k_1	**1**		0		1		0		1		0		Помилка
k_2		**0**	0			1	0			0	0		Помилка
k_3				**1**	1	1	0					1	Вірно
k_4								**1**	1	0	0	1	Помилка

Побудований код, звісно, не спроможний виявляти подвійні помилки. Наприклад, якщо помилка при передачі відбудеться в розрядах №3 та №6, так що отримали слово 101110011011.

№ розряду	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	Парність
	k_1	k_2	n_1	k_3	n_2	n_3	n_4	k_4	n_5	n_6	n_7	n_8	
Передане слово	**1**	**0**	0	**1**	1	1	0	**1**	1	0	1	1	
Отримане слово	**1**	**0**	1	**1**	1	0	0	**1**	1	0	1	1	
k_1	**1**		1		1		0		1		1		Помилка
k_2		**0**	1			0	0			0	1		Вірно
k_3				**1**	1	0	0					1	Помилка
k_4								**1**	1	0	1	1	Вірно

За запропонованою схемою виявиться, що помилка відбулась в розряді під номером $0101_2 = 5_{10}$. Передане слово – 100111011011, отримане слово – 101110011011, виправлене слово – 101100011011. Результат виявляється ще більше віддаленим від початкового слова, ніж отриманий код.

Для випадку автокорекції одиничних помилок можна додати можливість виявлення подвійних помилок. Для цього до коду потрібно додати ще один додатковий розряд, який би контролював парність в усіх розрядах коду (щоб кількість одиниць в усьому коді, враховуючи новий доданий розряд, була парною). При цьому попередні контрольні розряди не повинні враховувати новий доданий розряд при розрахунку парності одиниць в контрольних групах.