КОД ХЭММИНГА

ИСТОРИЯ И НАЗНАЧЕНИЕ

В середине 1940-х годов Ричард Хэмминг работал в знаменитых Лабораториях Белла на счётной машине Bell Model V. Это была электромеханическая машина, использующая релейные блоки, скорость которых была очень низка: один оборот за несколько секунд. Данные вводились в машину с помощью перфокарт, и поэтому в процессе чтения часто происходили ошибки. В рабочие дни использовались специальные коды, чтобы обнаруживать и исправлять найденные ошибки, при этом оператор узнавал об ошибке по свечению лампочек, исправлял и запускал машину. В выходные дни, когда не было операторов, при возникновении ошибки машина автоматически выходила из программы и запускала другую.

Хэмминг часто работал в выходные дни, и все больше и больше раздражался, потому что часто был должен перезагружать свою программу из-за ненадежности перфокарт. На протяжении нескольких лет он проводил много времени над построением эффективных алгоритмов исправления ошибок. В 1950 году он опубликовал способ, который на сегодняшний день известен как код Хэмминга.

Код Хэмминга используется для обнаружения и исправления ошибок в двоичных сообщениях (в прикладных программах в области хранения данных, особенно в RAID 2);

КОНТРОЛЬНЫЕ БИТЫ

В исходное сообщение добавляется избыточность – контрольные биты (обозначаются как ε), которые будут контролировать правильность передачи каждого символа в сообщении.

Место расположение контрольных бит в исходном определяется по формуле :

№ -номер (положение)

№ε = 2^i , где i = 0,1,2,3,4 и т.д. тогда № ε = 1,2,4,8,16,32,64,128,256 и т.д.

1 **2** 3 **4** 5 6 7 **8** 9 10 11 12 13 14 15 **16** 17 18

Формула кодового слова такова:

$\boldsymbol{\varepsilon_0}\ \boldsymbol{\varepsilon_1}\ a_1\ \boldsymbol{\varepsilon_2}\ a_2\ a_3\ a_4\ \boldsymbol{\varepsilon_3}\ a_5\ a_6\ a_7\ a_8\ a_9\ a_{10}\ a_{11}\ \boldsymbol{\varepsilon_4}$

Понятно, что количество ε в кодовой последовательности зависит от количества символов в исходной.

Например: в последовательности из 4 символов их 3 ($\boldsymbol{\varepsilon_0}\ \boldsymbol{\varepsilon_1}\ a_1\ \boldsymbol{\varepsilon_2}\ a_2\ a_3\ a_4$)

МАТРИЦА ХЭММИНГА

Используется для построения кодовой последовательности (определения значений контрольных бит), а также для проверки на стороне получателя используется проверочная матрица Хэмминга.

Выбор размера матрицы зависит от количества символов в исходном сообщении.

Например, если у нас исходная последовательность из 4 символов, то ε будет в количестве 3 и матрица будет (7х3) :

3– высоту матрицы определяет количество ε. (ε_0, ε_1, ε_2 см. слайд 3),

7- количество символов с добавленными ε

$$
H = \begin{array}{c}
\left[\begin{array}{ccccccc}
0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 1
\end{array}\right]
\begin{array}{c} h_0 \\ h_1 \\ h_2 \end{array} \\
\begin{array}{ccccccc} \varepsilon_0 & \varepsilon_1 & a_1 & \varepsilon_2 & a_2 & a_3 & a_4 \end{array}
\end{array}
$$

1 – 001	4 – 100	7 – 111
2 – 010	5 – 101	
3 – 011	6 – 110	

ТРАНСПОРТИРОВАННОЕ КОДОВОЕ СЛОВО

1) Построение кодового слова: рассмотрим на примере

Пусть дано информационное слово a = (1 0 0 1) – его надо зашифровать

1 2 3 4

$a_1=1,\ a_2=0,\ a_3=0,\ a_4=1$

То предварительно кодовое слово будет выглядеть:

$\varepsilon_0\ \varepsilon_1\ 1\ \varepsilon_2\ 0\ 0\ 1$

Далее определяем значение контрольных бит по формулам:

$\varepsilon_0 = a_1 \oplus a_2 \oplus a_4 = 1 \oplus 0 \oplus 1 = 0$

$\varepsilon_1 = a_1 \oplus a_3 \oplus a_4 = 1 \oplus 0 \oplus 1 = 0$

$\varepsilon_2 = a_2 \oplus a_3 \oplus a_4 = 0 \oplus 0 \oplus 1 = 1$

$\oplus$ - XOR

$1 \oplus 1=0$	$0 \oplus 0=0$
$1 \oplus 0=1$	$1 \oplus 0=1$

Составляем кодовое слово, оно будет называться **транспонированное кодовое слово**:

$\mathbf{R^t} = \mathbf{0}\ \mathbf{0}\ 1\ \mathbf{1}\ 0\ 0\ 1$ которое будем передавать

ВЫЧИСЛЕНИЕ ОШИБКИ

2) На приёмной стороне осуществляется проверка:

Ошибка (синдром ошибки) вычисляется по формуле:

$S = H \times R^t$, где S- это синдром,

H – проверочная матрица Хэмминга (такая же, что использовалась для построения кодовой последовательности),

R^t – принятое транспонированное кодовое слово (со слада5). $\mathbf{R^t}$**=** **0 0** 1 **1** 0 0 1 .

Итак, для того чтобы определить есть ли в принятой последовательности ошибка умножаем матрицу на строку:

$$H = \begin{array}{c} h_o \\ h_1 \\ h_2 \\ \\ \end{array} \begin{array}{c} \left[\begin{array}{ccccccc} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right] \\ \begin{array}{ccccccc} \boldsymbol{\varepsilon_0} & \boldsymbol{\varepsilon_1} & a_1 & \boldsymbol{\varepsilon_2} & a_2 & a_3 & a_4 \end{array} \end{array} \quad x \quad (\boldsymbol{\mathit{0\,0\,1\,1\,0\,0\,1}})$$

$0*\mathbf{\mathit{0}} \oplus 0*\mathbf{\mathit{0}} \oplus 0*\mathbf{\mathit{1}} \oplus \underline{1*\mathbf{\mathit{1}}} \oplus 1*\mathbf{\mathit{0}} \oplus 1*\mathbf{\mathit{0}} \oplus \underline{1*\mathbf{\mathit{1}}} = 0$

$0*\mathbf{\mathit{0}} \oplus 1*\mathbf{\mathit{0}} \oplus \underline{1*\mathbf{\mathit{1}}} \oplus 0*\mathbf{\mathit{1}} \oplus 0*\mathbf{\mathit{0}} \oplus 1*\mathbf{\mathit{0}} \oplus \underline{1*\mathbf{\mathit{1}}} = 0$ S = (000) → Значит

$1*\mathbf{\mathit{0}} \oplus 0*\mathbf{\mathit{0}} \oplus \underline{1*\mathbf{\mathit{1}}} \oplus 0*\mathbf{\mathit{1}} \oplus 1*\mathbf{\mathit{0}} \oplus 1*\mathbf{\mathit{0}} \oplus \underline{1*\mathbf{\mathit{1}}} = 0$ ошибок нет

ВЫЧИСЛЕНИЕ ОШИБКИ

А если бы нам передали вместо 0011001 что-то вроде 0001001

У нас синдром бы получился:

$0*\mathbf{0} \oplus 0*\mathbf{0} \oplus 0*\mathbf{0} \oplus \underline{1*\mathbf{1}} \oplus 1*\mathbf{0} \oplus 1*\mathbf{0} \oplus \underline{1*\mathbf{1}} = 0$

$0*\mathbf{0} \oplus 1*\mathbf{0} \oplus \underline{1*\mathbf{0}} \oplus 0*\mathbf{1} \oplus 0*\mathbf{0} \oplus 1*\mathbf{0} \oplus \underline{1*\mathbf{1}} = 1$ S = (011) → Значит

$1*\mathbf{0} \oplus 0*\mathbf{0} \oplus \underline{1*0} \oplus 0*\mathbf{1} \oplus 1*\mathbf{0} \oplus 1*\mathbf{0} \oplus \underline{1*\mathbf{1}} = 1$ есть ошибка

А чтобы определить в каком из символов произошло искажение, необходимо воспользоваться проверочной матрицей Хэмминга:

Найдём вычисленную комбинацию синдрома (011) в матрице – это третий символ или а1(см. слайд 6). 00**0**1001 Можем исправить: так код двоичный, то меняем на противоположное значение 0 → 1 и получаем верную комбинацию 0011001 .

ЗАДАНИЕ

1. Закодируйте слово 1101. Запишите решение в отдельном файле под пунктом 1.
2. Проверьте соответствует ли закодированное слово 1010111 исходному 1101. Запишите решение с помощью матрицы в этом же файле под пунктом 2.
3. Закодируйте любое слово из 4 символов. Запишите решение в отдельном файле под пунктом 3.
4. Умышленно сделайте одну ошибку. Передайте слово соседу для нахождения ошибки.