Передача данных через каналы с помехами

Упрощенная модель цифровой системы связи

В общем случае выглядит так:

Есть источник данных (пример флешка), из него на выходе получаем информационные символы. Далее эти символы поступают в кодер источника, который сопоставляет им кодовые символы и на выходе из него – кодовые символы. Далее идет канал с шумом. После эти данные с шумом поступают на декодер источника. Затем в получатель поступают декодированные кодовые символы = информационные символы.

С математического языка:

Мы будем работать с дискретными последовательностями

Будем считать что информационная последовательность состоит из элементов GF(2)={0,1}

Источник генерирует последовательность из 0 и 1, которая поступает на вход кодера, а из него получаем битовые кодовые символы, затем они поступают в канал.

Будем рассматривать двоичный симметричный канал.

Допустим поступил 0, а получить можем либо 0 либо 1 с вероятность ошибки p(переходная вероятность), аналогично для 1.

Чтобы улучшить качество передачи данных можно как вариант продублировать данные.

Пусть $p = 10^{-3}$

Из источника выходит последовательность 0011, они поступают в кодер и они дублируются: 000 000 111 111.

Информационные символы: 0 1

С помощью кодера превратили в: 000 111

Рассмотрим 000, на выходе из канала возможно 8 вариантов. В первых 4 вариантах возможно исправить ошибку(000 001 010 100).

Рассмотрим другой вариант:

Выходящие из источника данные склеим в пары: 00 01 10 11

Кодируем: 00000 10110 01011 11101

Надо будет посчитать вероятность ошибки

Из канала пришло 10000(ошибка в 1), декодер скажет пришло 1 слово.

Данный декодер тоже исправляет 1 ошибку.

k – число информационных символов

n – число символов в кодовом слове

R = k/n – скорость кода

Из кодирования источника пример алгоритм Хаффмана(убирали избыточность).

Канальное кодирование добавляет избыточность.

Для ДСК с переходной вероятностью p

Вводится понятие припускной способности: C = 1 – h(p),

Где $h(x) = -x\log_2 x - (1-x)\log_2(1-x)$ – энтропия двоичного ансамбля

При скорости передачи R < величины C может быть обеспечено сколь угодно малая вероятность ошибки кодирования за счет увеличения длины используемых кодов => увеличение сложности кодирования и декодирования.

Если R > C, то надежность передачи невозможна.

Вес Хэмминга

Если x – кодовое слово, то ω(x) **–** вес Хэмминга и определяется как число ненулевых элементов в x.

В двоичном случае – число 1.

А расстояние Хэмминга между двумя кодовыми словами x и y -> d(x,y) и определяется как кол-во элементов слова, которые отличаются друг от друга.

d(x,y)= ω(x+y)

$\mathrm{d_{min}} = \mathrm{mind(x, y)}$ – минимальное расстояние кода

Линейный код – код, в котором сумма двух любых кодовых слов тоже является кодовым словом.

C – множество кодовых слов.

Для любых x, y ∈ C: (x+y) ∈ C

Линейный q-ичный код (n, k) – код C называют любое k-мерное подпространство пространства F_q^n всевозможных векторов длины n.

Порождающей матрицей (n, k) – кода называется матрица размера k на n, где строки - базисные вектора.

Кодовые слова – линейные комбинации базисных векторов.

G – порождающая матрица

m – информационное слово

c – кодовое слово

c = m*G

Предположим, что для некоторого вектора h=(h1, .. hn) все кодовые слова удовлетворяют:

$$\left(\underset{c_i}{\rightarrow}, \underset{h}{\rightarrow}\right) = c_1h_1 + \cdots + c_nh_n = 0$$

$$Gh^T = 0$$

n-k будет проверок