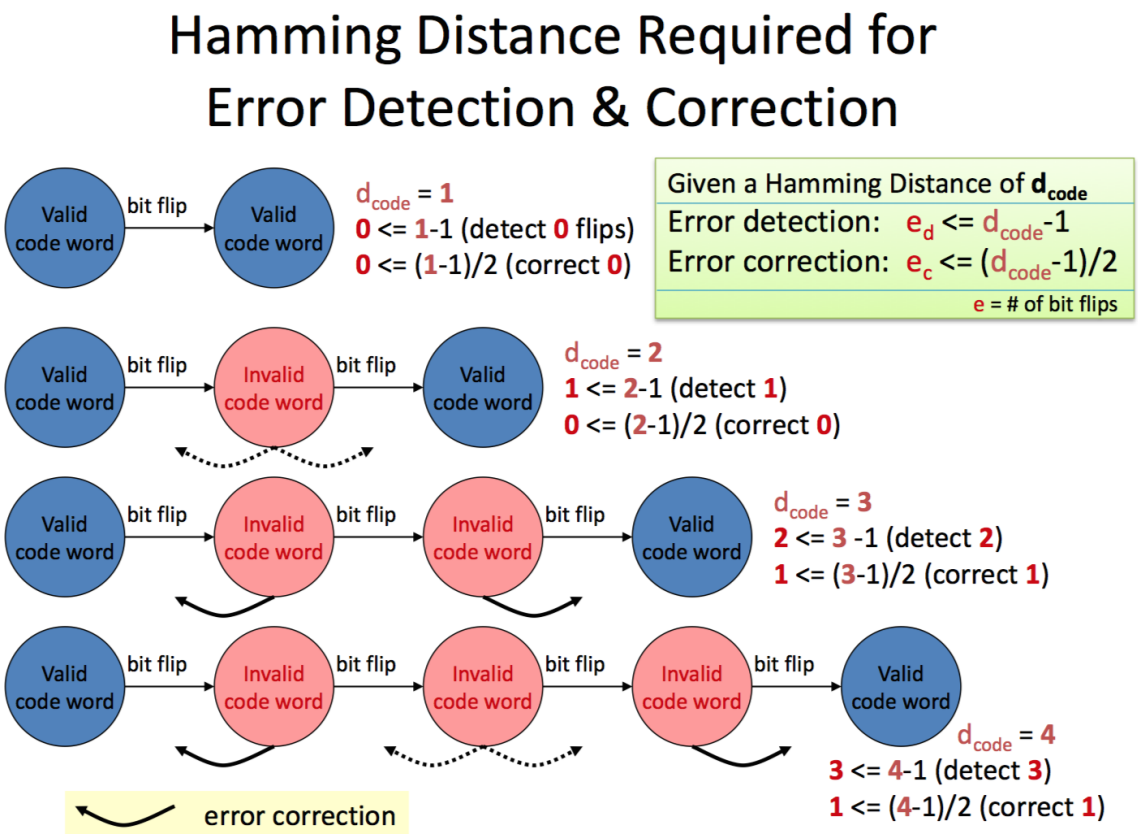
王贤义 320210931221
计算机基地班

1.海明距离 (Hamming distance)：两个码字之间不同对应比特的数目，试证明下述性质成立。

（1）为了检查出d个错（比特错），可以使用海明距离为 d+1 的编码

（2）为了纠正d个错，可以使用海明距离为 2d+1 的编码



例如当汉明距离为3时（第三个图例），意味着任意两个编码字之间至少有3个位不同。假设10111和10000是合法码字，而收到了10110。如果假设只有一个位被损坏，可以得出结论收到的可能是10111的损坏，因此可以纠正一个位的错误。然而，如果假设一个或两个位可能被损坏，则将无法确定10110应该是10111（一个1变成0）还是10000（两个0变成1）。

可以纠正一个位的错误（如果假设只有这种错误会发生），但无法区分来自一个编码字的一个位错误和来自另一个编码字的两位错误，因此无法处理两位错误。如果发生了两位错误，则能够检测到发生了一些错误，但是会错误地假设它是来自不同编码字的一个位错误。

同理，使用第四个图中示例的距离为4的编码，可以检测到两位错误。但如果收到中间的码字，可以知道（至少）两个位被损坏，但无法确定应该是左边的蓝色码字还是右边的蓝色码字。

2.试用数学方法证明CRC校验码可能存在的不足（接收端能够整除但不能检测到发送出错）

假设K是数据串的长度，则$S(D)$定义成如下多项式

$$s(D) = s_{K-1}D^{K-1} + s_{K-2}D^{K-2} + \ldots + s_0$$

其中D为幂次项，假设校验码的长度是L，那么CRC余数可以表示为如下式子

$$c(D) = c_{L-1}D^{L-1} + \ldots + c_1 D + c_0$$

那么传输的整个数据串可以表示为$x(D) = s(D)D^L + c(D)$,即

$$x(D) = s_{K-1}D^{L+K-1} + \ldots + s_0 D^L + c_{L-1}D^{L-1} + \ldots + c_0$$

一个CRC中的L次生成多项式$g(D)$定义如下

$$g(D) = D^L + g_{L-1}D^{L-1} + \ldots + g_1 D + 1$$

根据CRC校验的原理有

$$c(D) = Remainder[s(D)D^L / g(D)]$$

令$Z(D)$为$s(D)D^L$除以$g(D)$所得到的商，那么$c(D)$就可以表达成如下的式子

$$s(D)D^L = g(D)z(D) + c(D)$$

对上面两式子进行模二减$c(D)$，会发现模二减与加是一致的。

$$x(D) = s(D)D^L + c(D) = g(D)z(D)$$

因此，所有的码字$x(D)$都可被$g(D)$整除。

现在假设$x(D)$被发送方发送并且接受方收到$y(D)$，如果将错误序列表示为多项式$e(D)$，则有如下式子，其中的$+$的含义是模二加

$$y(D) = x(D) + e(D)$$

传输中的每个误差都对应于 $e(D)$ 中的一个非零系数。

在接收方，$REM[y(D)/g(D)]$会被计算，又因为$x(D)$能被$g(D)$整除，则有如下式子：

$$REM[y(D)/g(D)] = REM[e(D)/g(D)]$$

如果没有错误发生，那么$e(D) = 0$并且上式的余数为0。但是当错误发生时，如果接收方计算的余数为0那么就检测不到错误。也就是当且仅当下式成立时

$$e(D) = g(D)z(D)$$

有$e(D) \neq 0$且接收方检测不到错误。（其中$z(D)$是非0多项式）

首先，当只有一个错误发生时，即$e_i = 1$,$e(D) = D^i$。因为$g(D)$至少有两个非零项，即D^L和1这两项，那么$g(D)z(D)$就必须同时有至少两个非零项（$z(D)$为非0）。因此$g(D)z(D)$不可能等同于D^i,因此对于所有的i,所有的单错误都能被检测到。

同理，$g(D)$中最高位和最低位之间相距L，则$g(D)z(D)$中最高项和最低项之间也相距L（$z(D)$为非0）。因此，对于$e(D)$这个码字来说，突发错误的长度至少为$L+1$时，不能被检测到。

其次，当两个错误发生时，例如在位置i和j发生错误，则有：

$$e(D) = D^i + D^j = D^j(D^{i-j} + 1), i > j$$

由上面的论述，D^j不能被$g(D)$和$g(D)$中的任何因数整除（一个错误发生的情况）。因此当且仅当$D^{i-j} + 1$能被$g(D)$整除时$e(D)$的错误无法被检测出。对于任意L阶二项式$g(D)$，总会存在一个最小的n使得$D^n + 1$能被$g(D)$整除。在有限域理论中可知这个最小的n不会大于$2^L - 1$。此外，对于所有$L > 0$，存在一特定的多项式叫做本原多项式，此时这个对应的n等于$2^L - 1$。因此，如果$g(D)$是L级本原多项式，并将长度限制在最大$2^L - 1$,那么$D^{i-j} + 1$不能被$g(D)$整除，也就是错误能被检测到，否则将无法检测到。

3.尝试分析生成多项式对检错能力的影响（可以结合CRC-16、CRC-32等国际标准）

$CRC-16$的生成多项式如下：

$$g(D)=D^{16}+D^{15}+D^2+1$$

$CRC-CCITT$的生成多项式如下：

$$g(D)=D^{16}+D^{12}+D^5+1$$

CRC的编码结果有2^K种不同，它们都是$g(D)$的倍数。信道中可能发生的非全0错误共有$2^n-1=2^{K+L}-1$种。当错误能被$g(D)$整除，也即错误自身是编码器的可能输出之一时，这样的错误将骗过接收端。这种错误的个数是2^K-1个，占总错误的比例为$\frac{2^K-1}{2^{K+L}-1}\approx 2^{-L}$，故不能检测出的错误占总可能错误的$2^{-L}$。例如：$CRC-16$和$CRC-CCITT$不能检出的错误只是总可能错误的$\frac{1}{2^{16}}$，约为六万分之一，而$CRC-32$的为$\frac{1}{2^{32}}$。