

Formale Grundlagen

5. Übungsaufgaben

2008-04-29, Lösungen

1. Die Information 0100101 sollte übermittelt werden. Welche Bitfolge ist zu senden, wenn ein Paritätsbit verwendet werden soll.

Welche der Bitfolgen 01010111, 00010010, 11111111, 01001001 sind korrekte Codewörter? Wieviele Codewörter gibt es?

Wie hängt diese Art der Codierung mit der ASCII-Codierung von Zeichen zusammen.

Lösung: Um die Information 0100101 zu übermitteln ist das Codewort 01001011 zu senden.

00010010 und 11111111 haben eine gerade Anzahl von Einsen und sind deshalb korrekte Codewörter.

Obwohl ein Byte (die kleinste auf gängigen Computern direkt adressierbare Speichereinheit) 8 bit umfaßt, benutzt die ASCII-Codierung nur 7 bit (d.h. 128 Zeichen), um die üblichen Groß- und Kleinbuchstaben, Ziffern, Sonderzeichen und diverse Steuerzeichen darzustellen. Das reichte für die USA, aber für Umleite und andere sprachspezifische Zeichen ist kein Platz mehr. Das achte bit wurde für die Paritätsprüfung verwendet. Erst später wurde der ASCII-Code (in mehreren Varianten) auf 8 bit erweitert.

2. Wir betrachten den folgenden Code: Um eine Nachricht $x_1x_2x_3$ zu übertragen, werden die Prüfziffern $x_4 = x_1 + x_2$, $x_5 = x_1 + x_3$, $x_6 = x_2 + x_3$ angehängt (alle Additionen verstehen sich modulo 2). Betrachten Sie die Bitfolgen dieses Codes als Elemente des Vektorraums \mathbb{B}^6 (Skalare sind hier die Elemente von $\mathbb{B} = \{0, 1\}$, Addition geschieht modulo 2). Finden Sie eine Matrix \mathbf{P} , sodaß $\mathbf{P} \cdot \mathbf{x} = \mathbf{o}$ genau dann wenn \mathbf{x} eine gültiges Codewort ist. Prüfen Sie, ob 101110, 010101 gültige Codewörter sind, indem Sie mit \mathbf{P} multiplizieren.

Lösung: Die gesuchte Gleichung ist genau jene von Beispiel 3.

3. Lösen Sie die Gleichung

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \cdot \mathbf{x} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

in dem Sinne, daß Sie eine Basis des Lösungsraumes angeben (die Lösungsmenge ist ja offensichtlich in linearer Unterraum von \mathbb{B}^6).

Lösung: Hier kann man die Basis sofort hinschreiben, denn die Parameter x_1, x_2, x_3 sind offensichtlich frei wählbar, die anderen müssen dann den Gleichungen entsprechen. Somit ergeben sich die Vektoren $(1, 0, 0, 1, 1, 0)$, $(0, 1, 0, 1, 0, 1)$ und $(0, 0, 1, 0, 1, 1)$. Wir schreiben sie in die Spalten einer Matrix und multiplizieren:

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

4. Das *Hamming-Gewicht* $\|\mathbf{x}\|$ einer Binärfolge \mathbf{x} ist definiert als die Anzahl der Einsen in dieser Folge. Mit dem *Hamming-Abstand* zweier Binärfolgen \mathbf{x} und \mathbf{y} ist dann $\|\mathbf{x} - \mathbf{y}\|$ gemeint. *Der Minimalabstand* eines Codes ist dann der kleinste Hamming-Abstand, den zwei verschiedene Codewörter haben. Bestimmen Sie den Minimalabstand der bisher besprochenen Codes (einfacher Wiederholungscode, doppelter Wiederholungscode, ein Paritätsbit, Beispiel 2. Bestimmen Sie auch die jeweilige Informationsrate.

Erfüllt der Hamming-Abstand tatsächlich alle drei definierenden Bedingungen für einen vernünftigen Abstands begriff?

Lösung: Code mit Parity-Bit: Ändert sich irgendwo ein Bit, so auch das Parity-Bit. Die Minimaldistanz ist daher 2. Deshalb kann man mit einem solchen Code Einfachfehler erkennen, aber nichts korrigieren.

Auch der einfache Wiederholungscode hat Minimaldistanz 2, der zweifache Wiederholungscode hat Minimaldistanz 3, und allgemein der n -fache Wiederholungscode hat Minimaldistanz $n + 1$. Somit kann mit mit Wiederholungs codes jede gewünschte Fehlerkorrektur erreichen; allerdings verschlechtert sich die Informationsrate um denselben Faktor.

Der Code aus Beispiel 2 hat Minimaldistanz 3, wie man leicht (z.B. durch Aufzählen aller 8 Codewörter) feststellt.

5. Überlegen Sie sich geometrisch, wieviele Fehler ein Code mit Minimalabstand d erkennen bzw. korrigieren kann.

Lösung: Man kann $d - 1$ Fehler erkennen und $\lfloor \frac{d-1}{2} \rfloor$ Fehler korrigieren.

6. Die ISBN Nummer (vgl. z.B. <http://en.wikipedia.org/wiki/ISBN>) ist ein Code, der Einfachfehler und auch Vertauschungen zweier benachbarter Ziffern erkennt. Finden Sie auch hier eine Prüfmatrix \mathbf{P} , sodaß die Lösungen von $\mathbf{P} \cdot \mathbf{x} = \mathbf{o}$ genau den gültigen ISBN-Nummern entsprechen.

Lösung: Die letzte Ziffer sollte gleich der gewichteten Summe der restlichen sein. Damit ergibt sich:

$$\mathbf{P} = (1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9 \ -1).$$

Insbesondere stellen wir fest, daß dies ein linearer Code ist.

7. Wie funktioniert RAID-5? Warum reicht hier schon 1 Paritätsbit, um eine Fehlerkorrektur zu bewerkstelligen?

Lösung: Es werden n Festplatten verwendet. Die Information von je $n-1$ bit wird auf $n-1$ Festplatten verteilt, die n -te Platte erhält das Paritätsbit. Laut Codierungstheorie kann man somit Einfachfehler in den $n-1$ Bit erkennen, aber nichts korrigieren. Allerdings hat man hier eine zusätzliche Information, weil man ja weiß, welche Festplatte gegebenenfalls ausgefallen ist, und weil man annehmen kann, daß eine Festplatte korrekt arbeitet, solange sie nicht ausgefallen ist (wegen interner Fehlerkorrektur). Man weiß also zusätzlich, welches Bit jeweils unbrauchbar ist, und kann dieses dann mittels Paritätsbit rekonstruieren.

8. Die Wahrscheinlichkeit, daß ein Bit einer 100 Mbit/s Datenleitung falsch übertragen wird, sei 10^{-8} . Mit welcher Fehlerhäufigkeit ist zu rechnen?

Nun verwenden wir jedes 12. bit als Paritätsbit. Wie lange wird es in etwa dauern, bis ein unerkannter Fehler auftritt?

Lösung: Pro Sekunde werden 10^8 Bit übertragen, jedes mit einer Fehlerwahrscheinlichkeit von 10^{-8} . Somit wird im Mittel jede Sekunde eines falsch übertragen.

Ein Fehler wird nur dann nicht erkannt, wenn in einer 12-Bit-Gruppe mindestens zwei Fehler auftreten. Die Wahrscheinlichkeit dafür ist ca.

$\binom{12}{2}(10^{-8})^2 = 66 \cdot 10^{-16}$. Für die zu erwartende Fehlerzahl pro Sekunde ist dieser Wert mit $10^8/12$ (der Anzahl der pro Sekunden übertragenen 12-Bit Wörter) zu multiplizieren, was $5,5 \cdot 10^{-8}$ ergibt. Für die Fehler pro Jahr multiplizieren wir weiters mit $86400 \cdot 365,25$, was ungefähr 1,7 ergibt. Der Kehrwert davon entspricht der mittleren Dauer bis zum nächsten Fehler, also hier ca. 7 Monate.