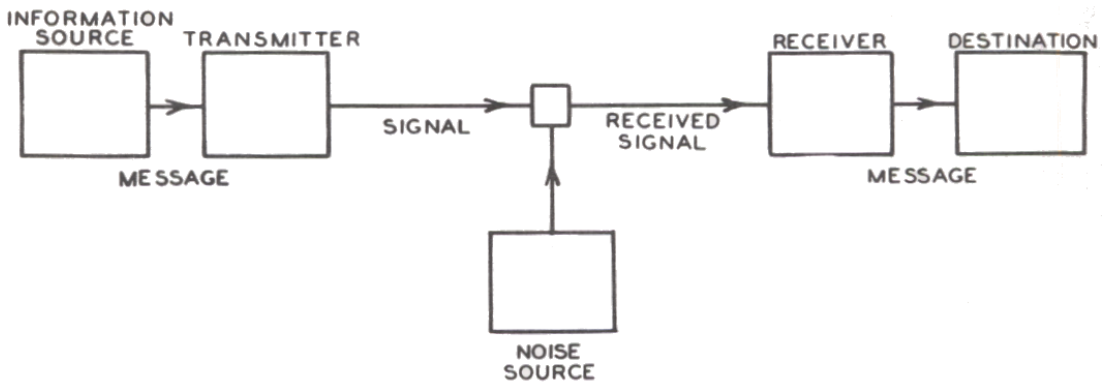
ChaosSeminar - Informationstheorie

Alexander.Bernauer@ulm.ccc.de
Stephanie.Wist@ulm.ccc.de

18. November 2005

Motivation

- schnelle Übertragung von Daten über gestörten Kanal



Shannon48

Was ist Information?

- Information ist Abnahme von Unsicherheit
- ohne Zufall keine Unsicherheit
- Unsicherheit vorher ist Information nacher

Übersicht

Stichprobenraum

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$$

- $\omega_1, \omega_2, \ldots, \omega_n$: Elementarereignisse
- $E \subseteq \Omega$: Ereignis
- $\emptyset$: unmögliches Ereignis
- Ω : sicheres Ereignis

Zufallsvariable

$$X : \Omega \to \{x_1, x_2, \ldots, x_L\}$$

Wahrscheinlichkeitsfunktion

$$f_X : \{x_1, x_2, \ldots, x_L\} \rightarrow \mathbb{R}$$

- $f_X(x_i) = P(X = x_i),\ i = 1, 2, \ldots, L$
- $\forall i : f_X(x_i) \geq 0$
- $\sum_{i=1}^{L} f_X(x_i) = 1$

Unabhängigkeit

- Ereignisse A und B stochastische unabhängig
  $\Leftrightarrow P(A \cap B) = P(A) \cdot P(B)$
- Zufallsvariablen X und Y stochastisch unabhängig
  $\Leftrightarrow \forall x_i, y_j : f_{XY}(x_i, y_j) = f_X(x_i) \cdot f_Y(y_j)$

bedingte Wahrscheinlichkeit

$$f_{X|Y}(x_i|y_j) := \frac{f_{XY}(x_i, y_j)}{f_Y(y_j)}$$

- $f_Y(y_j) > 0$

Erwartungswert

$$E\left[F(X)\right] := \overline{F(X)} := \sum_{i=1}^{L} F(x_i) f(x_i)$$

- F: Funktion einer Zufallsvariablen
- f: Wahrscheinlichkeitsfunktion der Zufallsvariablen

Erwartungen an ein Informationsmaß

- maximale Unsicherheit bei Gleichverteilung
- Addition von Informationen
- Stetigkeit auf den Wahrscheinlichkeiten

Entropie

$$H(X) = -K \sum_{x \in X} f_X(x) log f_X(x)$$

- $K > 0$
- $\lim_{f_X(x) \to 0} f_X(x) log f_X(x) := 0$

Information - Einheiten

Austausch der Basis:

$$log_{b_1} x = \frac{log_{b_2} x}{log_{b_2} b_1}$$

$b = 2$	binary digits (bits)
$b = e$	natural digits (nats)
$b = 10$	decimal digits

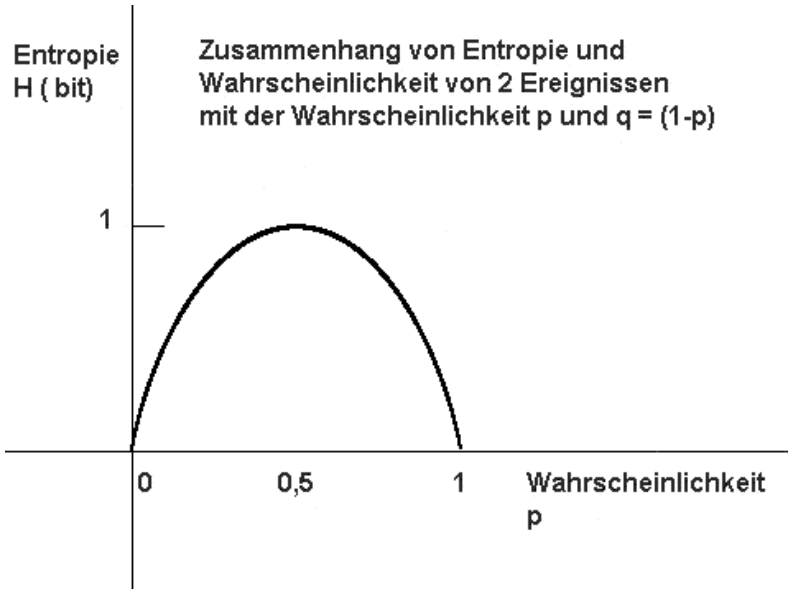
Entropie - Schranken

- $X : \Omega \rightarrow \{x_1, x_2, \ldots, x_L\}$

$$0 \leq H(X) \leq logL$$

- Gleichheit nach links $\Leftrightarrow \exists i : \ f_X(x_i) = 1$
- Gleichheit nach rechts $\Leftrightarrow X$ gleichverteilt

binäre Entropie



wikipedia

$$h(p) = -(p \log p + (p-1) \log p - 1)$$

bedingte Entropie

$$H(X|Y=y) = -\sum_{x \in X} f_{X|Y}(x|y) log f_{X|Y}(x|y)$$

$$H(X|Y) = \sum_{y \in Y} f_Y(y) H(X|Y=y)$$

- $H(X) = -\sum_{x \in X} f_X(x) log f_X(x)$

bedingte Entropie - Eigenschaften

- $0 \leq H(X|Y) \leq logL$
- Verminderung der Entropie $H(X|Y) \leq H(X)$

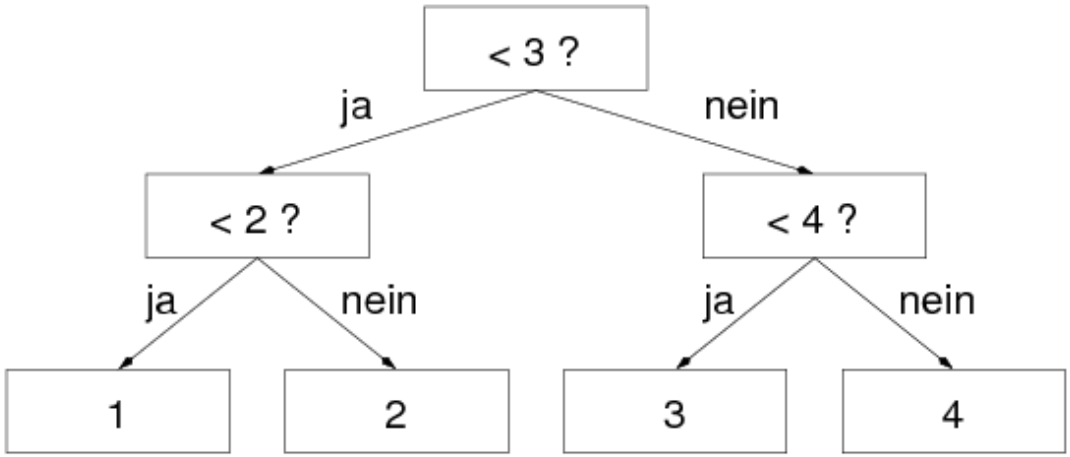
Addition von Informationen

$$H(X_1X_2\ldots X_N) = H(X_1) + H(X_2|X_1) + \ldots + H(X_N|X_1X_2\ldots X_{N-1})$$

Entscheidungsbäume

Die Entropie ist die mittlere Tiefe eines Entscheidungsbaumes

- Beispiel: ich denke mir eine Zahl...



wechselseitige Information

- Die Information, die ich über A erhalte, indem ich B beobachte
- A, B Ereignisse eines oder zweier Zufallsexperimente

$$I(A;B) := log\frac{P(A|B)}{P(A)}$$

- $P(A) \neq 0; P(B) \neq 0$

Eigeninformation

$$I(A) := I(A;A) = log\frac{P(A|A)}{P(A)} = -log\,P(A)$$

- $P(A|A) = 1$

Entropie

Der Erwartungswert der Eigeninformation der Ereignisse einer Zufallsvariablen

$$H(X) := E\left[I(X = x)\right] = E\left[-log f_X(x)\right] = -\sum_{x \in X} f_X(x) log f_X(x)$$

Transinformation

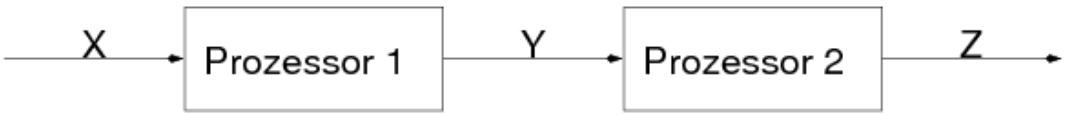
Die Information, die ich über X erhalte, in dem ich Y beobachte

$$I(X;Y) := E\left[I(X=x;Y=y)\right]$$

$$I(X;Y) := H(X) - H(X|Y)$$

- $H(X|Y)$: Äquivokation

Satz der Datenverarbeitung



$$I(X;Z) \leq \begin{cases} I(X;Y) \\ I(Y;Z) \end{cases}$$

$\Rightarrow$ Information kann nicht durch Datenverarbeitung erhöht werden.

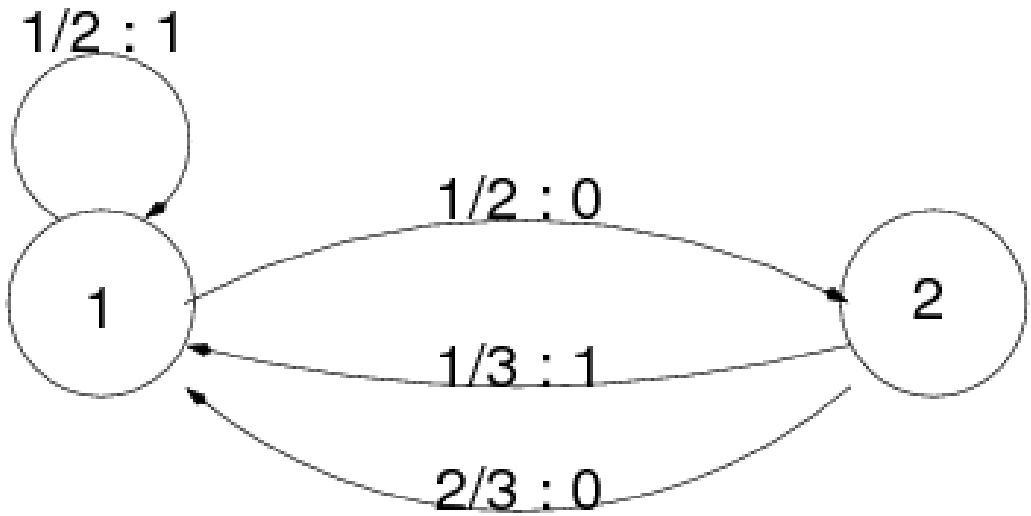
Fragen

- Wie gut kann man komprimieren?
- Was kann man alles komprimieren?
- Wie sehen Kompressionsalgorithmen aus?

Quelle

- Eine Quelle erzeugt einen Strom von Symbolen
- Strom entspricht den Ausgaben eines Markoff Prozesses

Markoff-Prozesse



- hier: Beschränkung auf diskrete, gedächtnisfreie Quellen

Redundanz

- Quelle Q mit L Quellsymbolen

$$R := 1 - \frac{H(Q)}{logL}$$

Quellcodierung

Quellcodiertheorem

Um eine Quelle Q mit Wahrscheinlichkeit 1 perfekt rekonstruieren zu können, sind $H(Q)$ bit pro Quellensymbol notwendig und hinreichend

Konsequenzen

- Quellcodierung ist Entfernung von Redundanz
- Mehrfachkomprimierung bringt nichts
- gleichverteilten Zufall kann man schlecht komprimieren

Beispiel

- $p(a) = \frac{1}{2}$
- $p(b) = \frac{1}{4}$
- $p(c) = \frac{1}{4}$
- $H(Q) = 1.5,\ log_2 3 = 1.58,\ R = 5.3\%$
- $a \rightarrow 0$
- $b \rightarrow 10$
- $c \rightarrow 11$
- $p(0) = \frac{1}{2}$
- $p(1) = \frac{1}{2}$
- $H(Q') = 1,\ log_2 2 = 1,\ R = 0\%$

Fragen

- Kann man fehlerfrei übertragen?
- Wie schnell kann man übertragen?
- Wie wirken sich Störungen auf dem Kanal aus?

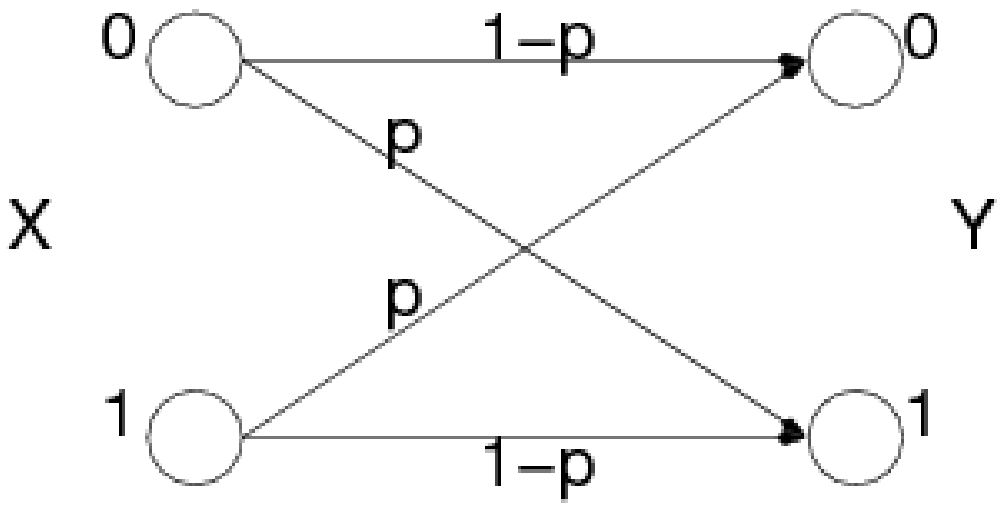
Kodierer

Rate

$$R := \frac{k}{n}$$

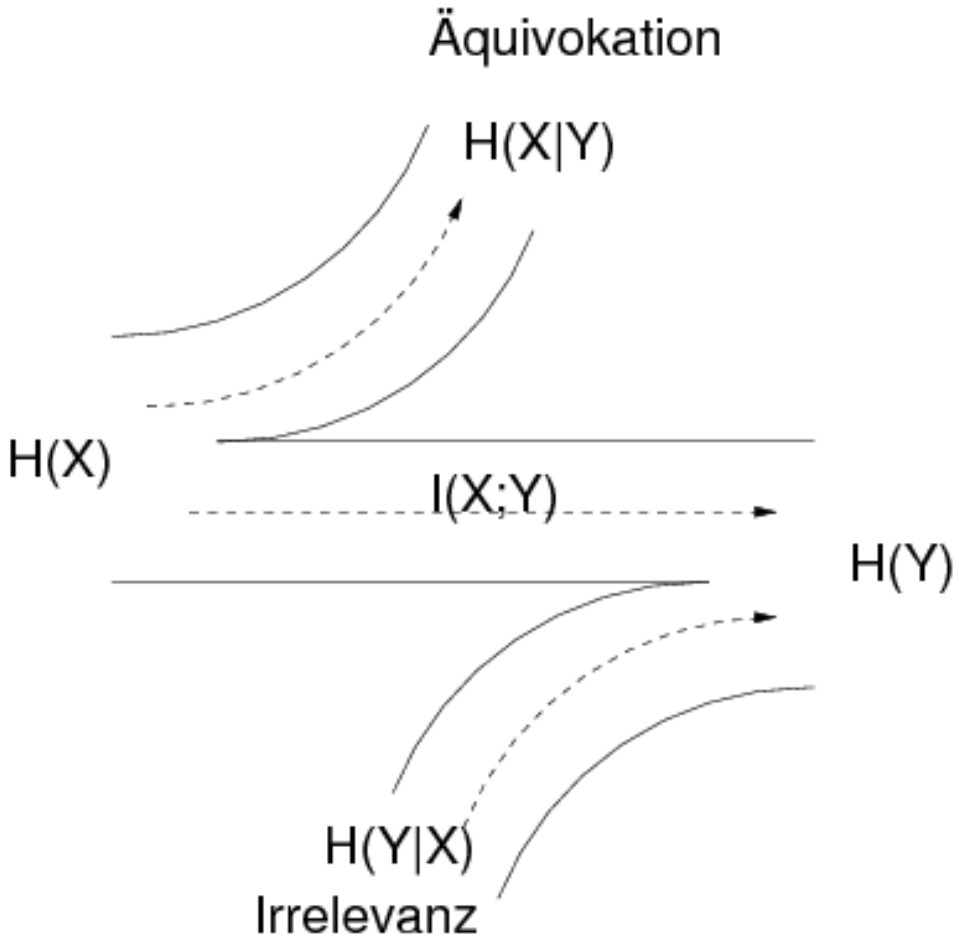
- Für den Kanal erscheint der Kodierer wie eine Quelle mit Entropie R

Kanal



- hier: Beschränkung auf diskrete, gedächtnisfreie Kanäle

Bergersches Diagramm



Kapazität

$$C := \max_{f_X} I(X;Y)$$

- $I(X;Y) = H(X) - H(X|Y)$

Kanalcodiertheorem - Teil 1

- Kanal mit Kapazität C, Quelle mit Entropie H.

Wenn $H < C$, so existiert ein Code mit beliebig kleiner Fehlerwahrscheinlichkeit

Kanalcodiertheorem - Teil 2

- Kanal mit Kapazität C, Quelle mit Entropie H.

Wenn $H > C$, so existiert kein Code mit einer Äquivokation unter $H - C$

Beispiel - Kapazität

Ein Kanal mit additiver, weißer, Gauss verteilter Störung hat die Kapazität $C = W \cdot log_2 \frac{P+N}{N}$

- W: Bandbreite
- P: Sendeleistung
- N: Rauschleistung

Bester Code?

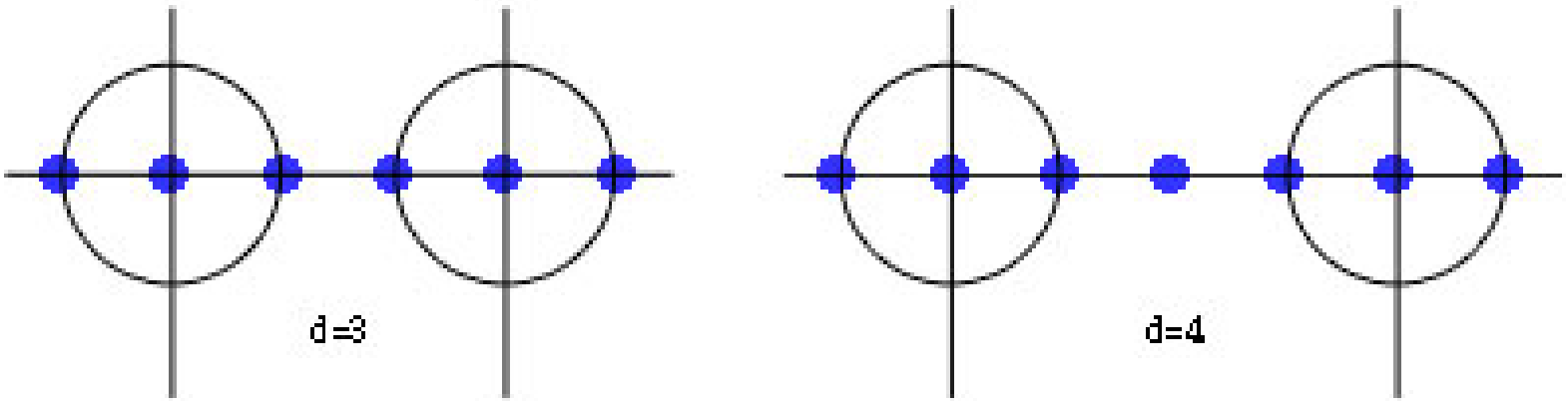
- Fehlerrate: $10^{-8} \approx 2^{-24} = 1MB$ (Ethernet)
- Fehlerrate: $10^{-4} \approx 2^{-12} = 1kB$ (Mobilfunk)
- $\longrightarrow$ unterschiedliche Codes für verschiedene Einsatzzwecke geeignet.

Beschreibung eins Codes $\mathcal{C}$

Code $\mathcal{C} = (n, k, d)$

- n: Codewortlänge
- k: Anzahl der codierten Bits
- d: Hammingabstand der Codewörter

Hammingabstand



Hammingabstand d ist minimale Distanz zweier Codewörter
Fehlererkennung für maximal $d-1$ Fehler
Fehlerbehebung für maximal $\frac{d-1}{2}$ Fehler

Beispiel - ParityCheck-Code

Code $\mathcal{C} = (3, 2, 2)$

- Codewörter $\{000, 011, 101, 110\}$
- Rate $R = \frac{k}{n} = \frac{2}{3}$
- Hammingabstand $d = 2$
- $\longrightarrow d - 1 = 1$ Fehler können erkannt werden
- $\longrightarrow \frac{d-1}{2} = 0$ Fehler können behoben werden
- empfangenes Codewort 111 wird als verfälscht erkannt

Beispiel - Wiederholungscodes

Code $\mathcal{C} = (5, 1, 5)$

- Codewörter $\{00000, 11111\}$
- Rate $R = \frac{k}{n} = \frac{1}{5}$
- Hammingabstand $d = 5$
- $\longrightarrow d - 1 = 4$ Fehler können erkannt werden
- $\longrightarrow \frac{d-1}{2} = 2$ Fehler können behoben werden
- empfangenes Codewort 00101 wird zu 00000

Reed-Solomon-Codes

- entdeckt 1960 von Irving S. Reed und Gustave Solomon
- es gilt: $\mathcal{C} = (n(= p-1), k, d = n-k+1)$
- MDS-Code $\rightarrow$ Korrektur aus beliebigen k Stellen des Codeworts
- Anwendungen: Fehlerkorrektur von Audio-CDs, Mobilfunk, DVB, Kommunikation mit Raumsonden

Beispiel CD

- Reed-Solomon-Kodierung mit Interleaver (CIRC)
- bis zu 3.500 Bit lange Fehler korrigierbar
- entspricht Kratzer von etwa $2,4mm$ Spurlänge
- bei Audio-Cds können sogar 12.000 Bit (etwa $8.5mm$) kompensiert werden

Restklassenringe

p: Primzahl
$\mathbb{Z}_p$ ist ein Körper
Rechenvorschrift: a+b mod p, a*b mod p
$\rightarrow$ Es gibt p verschiedene Elemente $(0, .., p-1)$
$\rightarrow 0 <= a+b, a*b < p$
$\rightarrow a^0 = a^{p-1} = 1$

Eigenschaften Körper

- Assoziativität
- Neutrales Element
- Inverses Element
- Kommutativität
- Distributivgesetz

Beispiel: Restklassenring p=7

+	0	1	2	3	4	5	6
0	0	1	2	3	4	5	6
1	1	2	3	4	5	6	0
2	2	3	4	5	6	0	1
3	3	4	5	6	0	1	2
4	4	5	6	0	1	2	3
5	5	6	0	1	2	3	4
6	6	0	1	2	3	4	5

*	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6
2	0	2	4	6	1	3	5
3	0	3	6	2	5	1	4
4	0	4	1	5	2	6	3
5	0	5	3	1	6	4	2
6	0	6	5	4	3	2	1

Primitives Element

α ist primitives Element
$\Leftrightarrow \{\alpha^0 = \alpha^{p-1}, \alpha^1, \alpha^2, .., \alpha^{p-2}\} = \{1, 2, .., p-1\}$
Beispiel: $p = 7$, $\alpha = 5$
$\{\alpha^0 = 1, \alpha^1 = 5, \alpha^2 = 5^2 = 4, \alpha^3 = 5^3 = 6, \alpha^4 = 5^4 = 2, \alpha^5 = 5^5 = 3\}$
Zu jedem Primkörper gibt es mindestens ein primitives Element

Polynomdarstellung

- $(A_0, A_1, .., A_{k-1}) = A(x)$
  $= A_0 + A_1 x + A_2 x^2 + .. + A_{k-1} x^{k-1} \rightarrow grad(A(x)) < k$
- $(a_0, a_1, .., a_{n-1}) = a(x)$
  $= a_0 + a_1 x + a_2 x^2 + .. + a_{n-1} x^{n-1} \rightarrow grad(a(x)) < n$

Kodierung

Kodierung:

- $A_0, .., A_{k-1}$: Information
- $a_0, .., a_{n-1}$: Codewort
- Es gilt: $a_i = A(\alpha^i) bzw. A_j = n^{-1} a(\alpha^{-j})$

Beispiel für $p = 7, \alpha = 5$

Es sei: $n = p - 1 = 6,\ k = 2,\ d = n - k + 1 = 5$
$\{\alpha^0 = 1, \alpha^1 = 5, \alpha^2 = 4, \alpha^3 = 6, \alpha^4 = 2, \alpha^5 = 3\}$
Information: $(23) \rightarrow A_0 = 2, A_1 = 3 \rightarrow A(x) = 2 + 3x$
Berechnung von $a_0, .., a_5$:

- $a_0 = A(\alpha^0 = 1) = 2 + 3 = 5$
- $a_1 = A(\alpha^1 = 5) = 2 + 3 * 5 = 2 + 1 = 3$
- $a_2 = A(\alpha^2 = 4) = 2 + 3 * 4 = 2 + 5 = 0$
- $a_3 = A(\alpha^3 = 6) = 2 + 3 * 6 = 2 + 4 = 6$
- $a_4 = A(\alpha^4 = 2) = 2 + 2 * 3 = 2 + 6 = 1$
- $a_5 = A(\alpha^5 = 3) = 2 + 3 * 3 = 2 + 2 = 4$

$\rightarrow$ Das erzeugte Codewort ist: (530614)

Dekodierung

Empfangen: $r(x) = a(x) + f(x)$
wobei f(x) der Fehler sein soll (f(x) = 0 $\rightarrow$ kein Fehler)
Dekodierung erfolgt über $R_j = n^{-1} * r(\alpha^{-j})$
Ist $grad R(x) < k$, dh. $R_k = R_{k+1} = .. = R_{n-1} = 0$, so ist r(x) ein gültiges Codewort,
R(x) = A(x) die Informationi
Wenn nicht, ist min. 1 Fehler aufgetreten. (Fehlererkennung)
Fehlerbehebung zB. über Fehlerstellenpolynom C.

Fehlerstellenpolynom

Gesucht: Polynom C(x) mit möglichst geringem Grad, für das gilt:
$C(x) = \prod_{f_j != 0}(\alpha^j - x)$
$\rightarrow$ *grad*$(C(x)) = \#$*Fehlerstellen*

Naive Berechnung

Sei $e \leq \frac{d-1}{2}$ die Anzahl der Fehler

- $\rightarrow C(x) = 1 + C_1 x + C_2 x^2 + .. + C_e x^e$
- Sei $S_0 = R_k, .., S_e = R_{n-1}$
- dann kann $C(x)$ über Schlüsselgleichungen $0 = \sum_{i=0}^{e-1} C_i S_{j-i}$ mit $j = 2e - 1, .., e$ berechnet werden
- Die Nullstellen von $C(x)$ sind die Stellen der inkorrekten Symbole

Beispiel

Gesendet: $a(x) = (530614)$
Empfangen: $r(x) = (530600)$

Beispiel

Gesendet: $a(x) = (530614)$
Empfangen: $r(x) = (530600)$
(korrekte) Annahme: 2 Fehler

Beispiel

Gesendet: $a(x) = (530614)$
Empfangen: $r(x) = (530600)$
(korrekte) Annahme: 2 Fehler $\rightarrow C(x) = 1 + C_1 x + C_2 x^2$

Berechnung von R

$n = 6, n^{-1} = 6, \{\alpha^0 = 1, \alpha^1 = 5, \alpha^2 = 4, \alpha^3 = 6, \alpha^4 = 2, \alpha^5 = 3\}$
$r(x) = 5 + 3x + 6x^3$ $R_j = n^{-1} r(\alpha^{-j})$

- $R_0 = r(1) = 6(5 + 3 + 6) = 6 * 0 = 0$
- $R_1 = r(3) = 6(5 + 2 + 1) = 6 * 1 = 6$
- $R_2 = r(2) = 6(5 + 6 + 6) = 6 * 3 = 4$
- $R_3 = r(6) = 6(5 + 4 + 1) = 6 * 3 = 4$
- $R_4 = r(4) = 6(5 + 5 + 6) = 6 * 2 = 5$
- $R_5 = r(5) = 6(5 + 1 + 1) = 6 * 0 = 0$
- $\rightarrow R_2, R_3, R_4 != 0 \rightarrow$ kein gültiges Codewort!

Berechnung von C

- $S_0 = R_2 = 4$
- $S_1 = R_3 = 4$
- $S_2 = R_4 = 5$
- $S_3 = R_5 = 0$

Zu lösende Gleichungen:

- $0 = C_0 S_3 + C_1 S_2 + C_2 S_1 = 0 + 5C_1 + 4C_2$
- $0 = C_0 S_2 + C_1 S_1 + C_2 S_0 = 5 + 4C_1 + 4C_2$
- $\rightarrow C_1 = 5, C_2 = 6$

$C(x) = 1 + 5x + 6x^2$ $C(x) = 0 \leftrightarrow x = 2, x = 3$
$\alpha^4 = 2, \alpha^5 = 3 \rightarrow$ Fehler an Stellen 4 und 5

Specials

- Faltungscodes (Zustandsmaschine, Trellis-Codierung)
- Kanalcodes, die je nach Bedarf zusätzliche Informationen schicken
- geschachtelte Kanalcodes (innerer Code, äusserer Code)

verwendete Software

- vim
- LaTeX(Beamer Klassen)
- xfig
- ImageMagick
- Gimp
- Gimp