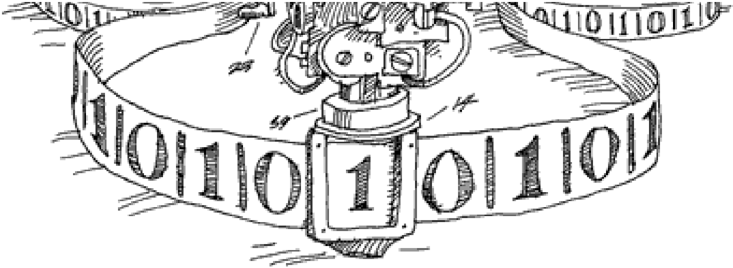
Karlsruher Institut für Technologie

Tutorium Theoretische Grundlagen der Informatik

Simon Bischof

Institut für Kryptographie und Sicherheit



Organisatorisches

- ÜB7: Abgabe am 1.2., 12:00 Uhr; Abholung ab 8.2. (im Tut oder danach bei den Übungsleitern)
- Hauptklausur: 22.02., 8:00 Uhr
- Anmeldung ab sofort bis 15.2.
- Nachklausur: 10.04., 11:30 Uhr
- Klausur geht 60 min
- Es gibt 60 Punkte, 20 sind zum Bestehen hinreichend
- Keine Hilfsmittel erlaubt

Präfix Codes

- Sei Q ein Alphabet und $f : Q \to \{0, 1\}^*$ eine Kodierung.

Präfix Codes

Karlsruher Institut für Technologie

- Sei Q ein Alphabet und $f : Q \rightarrow \{0, 1\}^*$ eine Kodierung.
- Achtung: Evtl. ist die Dekodierung nicht eindeutig!

Präfix Codes

- Sei Q ein Alphabet und $f : Q \to \{0, 1\}^*$ eine Kodierung.
- Achtung: Evtl. ist die Dekodierung nicht eindeutig!
- Falls für alle Codewörter $c_1 c_2 \ldots c_n \in f(Q)$ und für $k < n$ das Wort $c_1 c_2 \ldots c_k$ kein Codewort ist, so heißt die Kodierung Präfix-Code.
- Präfix-Codes sind immer eindeutig.

Präfix Codes

- Sei Q ein Alphabet und $f : Q \rightarrow \{0, 1\}^*$ eine Kodierung.
- Achtung: Evtl. ist die Dekodierung nicht eindeutig!
- Falls für alle Codewörter $c_1 c_2 \dots c_n \in f(Q)$ und für $k < n$ das Wort $c_1 c_2 \dots c_k$ kein Codewort ist, so heißt die Kodierung Präfix-Code.
- Präfix-Codes sind immer eindeutig.
- Für einen Präfixcode gilt: Die mittlere Codewortlänge ist größer oder gleich der Entropie der Quelle

Shannon-Fano

- Shannon-Fano konstruiert einen (nicht immer optimalen) Präfix-Code.

Shannon-Fano

- Shannon-Fano konstruiert einen (nicht immer optimalen) Präfix-Code.

- Sortiere die vorkommenden Symbole nach ihrer Häufigkeit.

Shannon-Fano

Karlsruher Institut für Technologie

- Shannon-Fano konstruiert einen (nicht immer optimalen) Präfix-Code.

- Sortiere die vorkommenden Symbole nach ihrer Häufigkeit.
- Bestimme nun den Punkt, an dem die Reihe aus Symbolen aufgeteilt werden muss, so dass die aufsummierten Wahrscheinlichkeiten der beiden entstehenden Gruppen möglichst gleich sind.

Shannon-Fano

- Shannon-Fano konstruiert einen (nicht immer optimalen) Präfix-Code.

- Sortiere die vorkommenden Symbole nach ihrer Häufigkeit.
- Bestimme nun den Punkt, an dem die Reihe aus Symbolen aufgeteilt werden muss, so dass die aufsummierten Wahrscheinlichkeiten der beiden entstehenden Gruppen möglichst gleich sind.
- Hänge nun die beiden entstehenden Gruppen als Blätter an eine neue Wurzel.

Shannon-Fano

- Shannon-Fano konstruiert einen (nicht immer optimalen) Präfix-Code.

- Sortiere die vorkommenden Symbole nach ihrer Häufigkeit.
- Bestimme nun den Punkt, an dem die Reihe aus Symbolen aufgeteilt werden muss, so dass die aufsummierten Wahrscheinlichkeiten der beiden entstehenden Gruppen möglichst gleich sind.
- Hänge nun die beiden entstehenden Gruppen als Blätter an eine neue Wurzel.
- Verfahre nun rekursiv: Ersetze die Gruppen jeweils durch den Baum der beim Anwenden des Verfahrens auf sie jeweils entsteht, solange, bis alle Blätter einzelne Symbole sind.

Shannon-Fano

- Shannon-Fano konstruiert einen (nicht immer optimalen) Präfix-Code.

- Sortiere die vorkommenden Symbole nach ihrer Häufigkeit.
- Bestimme nun den Punkt, an dem die Reihe aus Symbolen aufgeteilt werden muss, so dass die aufsummierten Wahrscheinlichkeiten der beiden entstehenden Gruppen möglichst gleich sind.
- Hänge nun die beiden entstehenden Gruppen als Blätter an eine neue Wurzel.
- Verfahre nun rekursiv: Ersetze die Gruppen jeweils durch den Baum der beim Anwenden des Verfahrens auf sie jeweils entsteht, solange, bis alle Blätter einzelne Symbole sind.
- Am Ende: Wege nach links: 0, nach rechts: 1

Kanalkodierungen

- Daten werden über einen Kanal geschickt

Kanalkodierungen

- Daten werden über einen Kanal geschickt
- Kanal verändert (zufällig) einzelne Bits

Kanalkodierungen

- Daten werden über einen Kanal geschickt
- Kanal verändert (zufällig) einzelne Bits
- Wie sichere ich die Daten?

Grundlagen

- Hammingdistanz für $x, y \in \{0,1\}^n$:
  $d(x,y) := \sum_{i=1}^{n} (1 - \delta_{x_i,y_i}) = \#\{i = 1, \ldots, n | x_i \neq y_i\}$

Grundlagen

- Hammingdistanz für $x, y \in \{0,1\}^n$:
  $d(x,y) := \sum_{i=1}^{n} (1 - \delta_{x_i,y_i}) = \#\{i = 1,\dots,n | x_i \neq y_i\}$
- Hammingkugel um x mit Radius ϱ:
  $B_\varrho(x) := \{y \in \{0,1\}^n | d(x,y) \leq \varrho\}$

Grundlagen

- Hammingdistanz für $x, y \in \{0,1\}^n$:
  $d(x,y) := \sum_{i=1}^{n} (1 - \delta_{x_i,y_i}) = \#\{i = 1, \dots, n | x_i \neq y_i\}$
- Hammingkugel um x mit Radius ϱ:
  $B_\varrho(x) := \{y \in \{0,1\}^n | d(x,y) \leq \varrho\}$
- Maximum-Likelyhood-Decoding: Dekodiere empfangenes Wort y als Codewort x mit $d(x,y)$ minimal.

Grundlagen

- Hammingdistanz für $x, y \in \{0,1\}^n$:
  $d(x,y) := \sum_{i=1}^{n} (1 - \delta_{x_i,y_i}) = \#\{i = 1, \ldots, n | x_i \neq y_i\}$
- Hammingkugel um x mit Radius ϱ:
  $B_\varrho(x) := \{y \in \{0,1\}^n | d(x,y) \leq \varrho\}$
- Maximum-Likelyhood-Decoding: Dekodiere empfangenes Wort y als Codewort x mit $d(x,y)$ minimal.
- Rate für Code mit M Wörtern der Länge n: $R = \frac{\log_2 M}{n}$

Grundlagen

- Hammingdistanz für $x, y \in \{0,1\}^n$:
  $d(x,y) := \sum_{i=1}^{n} (1 - \delta_{x_i,y_i}) = \#\{i = 1, \ldots, n | x_i \neq y_i\}$
- Hammingkugel um x mit Radius ϱ:
  $B_\varrho(x) := \{y \in \{0,1\}^n | d(x,y) \leq \varrho\}$
- Maximum-Likelyhood-Decoding: Dekodiere empfangenes Wort y als Codewort x mit $d(x,y)$ minimal.
- Rate für Code mit M Wörtern der Länge n: $R = \frac{\log_2 M}{n}$
- Mit Maximum-Likelyhood-Decoding werde gesendetes x_i mit WK P_i falsch dekodiert. Wahrscheinlichkeit für falsche Dekodierung:
  $P_C = \frac{1}{M} \sum_{i=1}^{M} P_i$

Kanalkodierungen

Hier gibt es vor allem zwei Möglichkeiten:

Kanalkodierungen

Hier gibt es vor allem zwei Möglichkeiten:

- Block-Codes: Codeworte gleicher Länge, unabhängige Kodierung von aufeinanderfolgenden Blöcken
- Faltungs-Codes: Codeworte evtl. beliebig lang, Kodierung abhängig vom Vorgeschehen

Kanalkodierungen

Hier gibt es vor allem zwei Möglichkeiten:

- Block-Codes: Codeworte gleicher Länge, unabhängige Kodierung von aufeinanderfolgenden Blöcken
- Faltungs-Codes: Codeworte evtl. beliebig lang, Kodierung abhängig vom Vorgeschehen

In der VL: Nur Block-Codes

Eigenschaften von Block-Codes

- Code immer über endl. Alphabet Q. Oft: $Q = \{0, 1\}$.
- Der Block-Code ist eine Teilmenge $C \subseteq Q^n$ für festes $n \in \mathbb{N}$.
- Code heißt für $\#Q = 1$ trivial (nur ein Codewort!), $\#Q = 2$ binär, $\#Q = 3$ terniär, ...

Eigenschaften von Block-Codes

- Code immer über endl. Alphabet Q. Oft: $Q = \{0, 1\}$.
- Der Block-Code ist eine Teilmenge $C \subseteq Q^n$ für festes $n \in \mathbb{N}$.
- Code heißt für $\#Q = 1$ trivial (nur ein Codewort!), $\#Q = 2$ binär, $\#Q = 3$ terniär, ...
- Minimaldistanz $m(C) := \min\limits_{c_1, c_2 \in C, c_1 \neq c_2} d(c_1, c_2)$
- Rate $R(C) := \frac{\log(\#C)}{\log(\#Q^n)} = \frac{\log(\#C)}{n \cdot \log(\#Q)}$
- Ein Code C mit $m(C)$ ungerade heißt perfekt, falls für alle $y \in Q^n$ genau ein $x \in C$ existiert mit $d(x, y) \leq \frac{m(C)-1}{2}$

Eigenschaften von Block-Codes

- Code immer über endl. Alphabet Q. Oft: $Q = \{0, 1\}$.
- Der Block-Code ist eine Teilmenge $C \subseteq Q^n$ für festes $n \in \mathbb{N}$.
- Code heißt für $\#Q = 1$ trivial (nur ein Codewort!), $\#Q = 2$ binär, $\#Q = 3$ terniär, ...
- Minimaldistanz $m(C) := \min_{c_1, c_2 \in C, c_1 \neq c_2} d(c_1, c_2)$
- Rate $R(C) := \frac{\log(\#C)}{\log(\#Q^n)} = \frac{\log(\#C)}{n \cdot \log(\#Q)}$
- Ein Code C mit $m(C)$ ungerade heißt perfekt, falls für alle $y \in Q^n$ genau ein $x \in C$ existiert mit $d(x, y) \leq \frac{m(C)-1}{2}$
- Ein Block-Code C kann entweder bis zu $m(C) - 1$ Fehler erkennen oder bis zu $\lfloor \frac{m(C)-1}{2} \rfloor$ Fehler korrigieren.

Lineare Codes

Normalerweise: Beliebiger endlicher Körper $\mathbb{F}_q$. Hier immer:
$\mathbb{F} = \mathbb{F}_2 = \{0, 1\}$.

- Ein linearer $[n, k]$-Block-Code C ist ein Untervektorraum von $\mathbb{F}^n$ der Dimension k.

Lineare Codes

Normalerweise: Beliebiger endlicher Körper $\mathbb{F}_q$. Hier immer: $\mathbb{F} = \mathbb{F}_2 = \{0, 1\}$.

- Ein linearer $[n, k]$-Block-Code C ist ein Untervektorraum von $\mathbb{F}^n$ der Dimension k.
- Hamming-Gewicht: $wgt(x) := d(x, 0)$. Es ist außerdem $d(x, y) = wgt(x - y)$.

Lineare Codes

Normalerweise: Beliebiger endlicher Körper $\mathbb{F}_q$. Hier immer:
$\mathbb{F} = \mathbb{F}_2 = \{0, 1\}$.

- Ein linearer $[n, k]$-Block-Code C ist ein Untervektorraum von $\mathbb{F}^n$ der Dimension k.
- Hamming-Gewicht: $wgt(x) := d(x, 0)$. Es ist außerdem $d(x, y) = wgt(x - y)$.
- Beschreibe C als Kern einer $\mathbb{F}^{(n-k)\times n}$-Matrix (Parity-Check- oder Prüf-Matrix): $C = Kern(H) = \{x \in \mathbb{F}^n | Hx = 0\}$
- ...oder als Bild einer $\mathbb{F}^{n\times k}$-Matrix (Generatormatrix): $C = Bild(G) = \{Gx | x \in \mathbb{F}^k\}$.

Lineare Codes

Normalerweise: Beliebiger endlicher Körper $\mathbb{F}_q$. Hier immer: $\mathbb{F} = \mathbb{F}_2 = \{0, 1\}$.

- Ein linearer $[n, k]$-Block-Code C ist ein Untervektorraum von $\mathbb{F}^n$ der Dimension k.
- Hamming-Gewicht: $wgt(x) := d(x, 0)$. Es ist außerdem $d(x, y) = wgt(x - y)$.
- Beschreibe C als Kern einer $\mathbb{F}^{(n-k)\times n}$-Matrix (Parity-Check- oder Prüf-Matrix): $C = Kern(H) = \{x \in \mathbb{F}^n | Hx = 0\}$
- ...oder als Bild einer $\mathbb{F}^{n\times k}$-Matrix (Generatormatrix): $C = Bild(G) = \{Gx | x \in \mathbb{F}^k\}$.
- $s = Hx$ heißt das Fehlersyndrom von $x \in \mathbb{F}^n$.

Lineare Codes

Normalerweise: Beliebiger endlicher Körper $\mathbb{F}_q$. Hier immer: $\mathbb{F} = \mathbb{F}_2 = \{0, 1\}$.

- Ein linearer $[n, k]$-Block-Code C ist ein Untervektorraum von $\mathbb{F}^n$ der Dimension k.
- Hamming-Gewicht: $wgt(x) := d(x, 0)$. Es ist außerdem $d(x, y) = wgt(x - y)$.
- Beschreibe C als Kern einer $\mathbb{F}^{(n-k)\times n}$-Matrix (Parity-Check- oder Prüf-Matrix): $C = Kern(H) = \{x \in \mathbb{F}^n | Hx = 0\}$
- ...oder als Bild einer $\mathbb{F}^{n\times k}$-Matrix (Generatormatrix): $C = Bild(G) = \{Gx | x \in \mathbb{F}^k\}$.
- $s = Hx$ heißt das Fehlersyndrom von $x \in \mathbb{F}^n$.
- Für gegebenes s heißt (falls eindeutig) das $e \in \mathbb{F}^n$ mit $wgt(e) = min\{wgt(x) | x \in \mathbb{F}^n \setminus \{0\}, Hx = s\}$ der Coset-Leader von s.

Hamming-Codes

- Hamming-Codes sind $[2^k - 1, 2^k - k - 1]$-Codes, für die je zwei Spalten der Prüfmatrix linear unabhängig sind.

Hamming-Codes

- Hamming-Codes sind $[2^k - 1, 2^k - k - 1]$-Codes, für die je zwei Spalten der Prüfmatrix linear unabhängig sind.
- Es gilt $R(C) = \frac{2^k-k-1}{2^k-1} = 1 - \frac{k}{2^k-1}$ und $m(C) = 3$.
- Hamming-Codes sind perfekt.