

Mittlerer Informationsgehalt oder Entropie

mit Gewichtung der n Einzelelemente durch jeweilige Häufigkeit $p(x_i)$:

$$H = \sum_{i=1}^n p(x_i) \cdot H_i = \sum_{i=1}^n p(x_i) \cdot \text{ld} \frac{1}{p(x_i)} = - \sum_{i=1}^n p(x_i) \cdot \text{ld} p(x_i)$$

Sei Q eine Quelle ...

Die **Shannon-Entropie** von Q ist $H_1(Q) = - \sum_{i=1}^n \Pr[q_i] \log_2 \Pr[q_i]$

- fragt nach der durchschnittlichen Anzahl an Bits, die gebraucht werden, um die Wörter aus einer Quelle zu speichern.

Entropie bei Gleichverteilung

Ist Q eine gleichverteilte Quelle von n Elementen, dann ist $H_1(Q) = \log_2(n)$.

Entropie bei Ungleichverteilung

Ist Q nicht gleichverteilt, so ist $H_1(Q) < \log_2(n)$. Egal wie Q verteilt ist, stets gilt $0 \leq H_1(Q) \leq \log_2(n)$.

Entropie bei unabhängigen Zufallsquellen

Für zwei Quellen Q und R bezeichnet QR die Menge zufälliger Paare (q, r) , von Elementen q aus Q und r aus R .

Sind Q und R zwei voneinander unabhängige Quellen, dann ist $H_1(QR) = H_1(Q) + H_1(R)$.

Es bezeichne Q^i die Quelle, die durch i -faches und unabhängiges Ziehen eines Elements aus Q definiert ist. $H_1(Q^i) = i * H_1(Q)$.

Min-Entropie

- Die Shannon-Entropie betrachtet die **durchschnittliche** Bitanzahl. Es besteht aber die Möglichkeit, dass der Großteil der Werte unterdurchschnittlich und nur wenige überdurchschnittlich sind. Das führt - bezogen auf die Passwortsicherheit - zu optimistischen Schlussfolgerungen. Deshalb wird alternativ zur Shannon-Entropie auch die Min-Entropie betrachtet.

Die **Min-Entropie** von Q ist $H_\infty(Q) = \min_i (-\log_2(p_i)) = -\log(\max(p_i))$.

Ist Q gleichverteilt, gelten $H_\infty(Q) = H_1(Q) = \log_2(n)$ und $0 \leq H_\infty(Q) \leq H_1(Q) \leq \log_2(n)$, außerdem $H_\infty(Q^i) = i * H_\infty(Q)$.

Beispiel

Zeichensatz	Symbolanzahl	Entropie pro Symbol
Arabische Ziffern (0-9)	10	3.322 bit
Hexadezimalzahlen (0-9, A-F)	16	4.000 bit
Alphabet (a-z oder A-Z) (case insensitive)	26	4.700 bit
Alphanumerisch (a-z oder A-Z, 0-9) (case in.)	36	5.170 bit
Alphabet (a-z, A-Z) (case sensitive)	52	5.700 bit
Alphanumerisch (a-z, A-Z, 0-9) (case sen.)	62	5.954 bit
Alle druckbaren ASCII Zeichen	96	6.570 bit
Alle erweiterten druckbaren ASCII Zeichen	218	7.768 bit
Diceware word list	7776	12.925 bit