College Cryptografie

Cursusjaar 2003

Informatietheorie

29 januari 2003

Claude E. Shannon

Informatiekanaal

Entropie

Equivocatie

Markov ketens

Entropie Markov keten

Unicity distance

Binair symmetrisch kanaal

Kansberekening ud

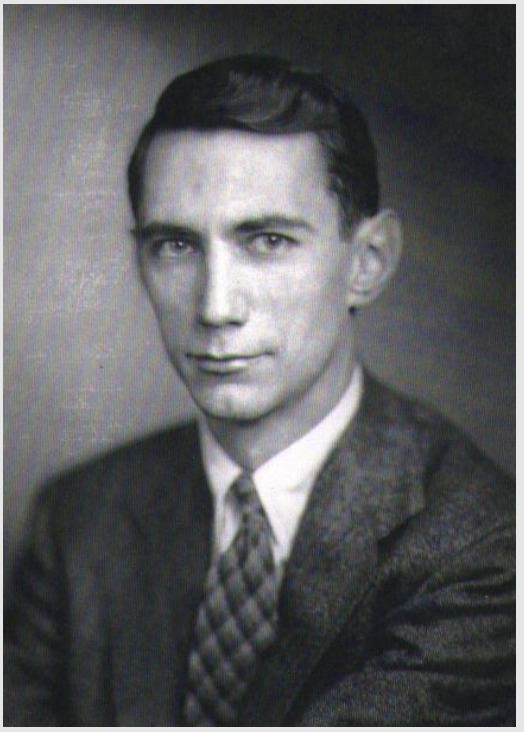
ONDERWERPEN

Claude Elwood Shannon (1916–2001)
AT&T Bell Telephones (1941–1972)

Grondslagen informatietheorie 1948
A Mathematical Theory of Communication
Bell System Technical Journal

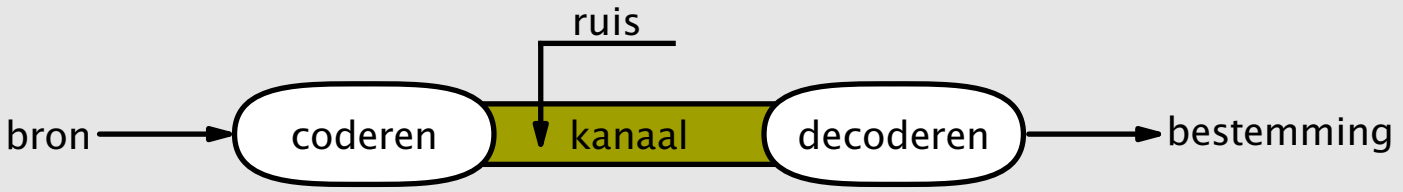
Toepassing in de cryptografie 1949
Communication theory of secrecy systems
Bell System Technical Journal

Foto plm. 1950



Claude Elwood Shannon

Claude E. Shannon



probleemstellingen informatieoverdracht:

- volledig ondanks ruis
- meeluisteren verhinderen
- zo compact mogelijk

Informatiekanaal

p_i = kans op symbool $s_i \in \{S\}$

$I(s_i)$ = informatiewaarde symbool s_i

$|S|$ = aantal symbolen in S

$$I(s_i) = \log_2 \left(\frac{1}{p_i} \right) = -\log_2 p_i$$

entropie van $\{S\} = -\sum_{i=1}^{|S|} p_i \log_2 p_i$ (bits per symbool)

grenswaarden $0 \leq H(S) \leq \log_2 |S|$

analogie met entropie in de thermodynamica (warmteleer)

	p	$-\log_2 p$	$-p \log_2 p$		p	$-\log_2 p$	$-p \log_2 p$
A	.074	3.763	.277	N	.080	3.653	.290
B	.010	6.682	.065	O	.075	3.732	.281
C	.031	5.027	.154	P	.027	5.227	.140
D	.042	4.558	.193	Q	.004	8.158	.029
E	.130	2.944	.383	R	.076	3.722	.282
F	.028	5.142	.146	S	.061	4.031	.247
G	.016	5.932	.097	T	.092	3.444	.316
H	.034	4.883	.165	U	.026	5.265	.137
I	.074	3.766	.277	V	.015	6.028	.092
J	.002	9.252	.015	W	.016	6.002	.094
K	.003	8.400	.025	X	.005	7.758	.036
L	.036	4.799	.174	Y	.019	5.692	.110
M	.025	5.337	.132	Z	.001	9.995	.010

$H_{26} = 4.700$

$H_{engels} = 4.167$

(bits per symbool)

Entropie Engels

combinatie (i,j) met j afhankelijk i en kans $p_{ij} = p_i p_{j|i}$

$$\text{entropie } H(I,J) = -\sum_{i,j}^{|S|} p_{ij} \log p_{ij}$$

$$H(I,J) = \underbrace{-\sum_i p_i \log p_i}_{H(I)} \underbrace{-\sum_i p_i \sum_j p_{j|i} \log p_{j|i}}_{H(J|I)}$$

als i en j onafhankelijk $\quad p_{ij} = p_i p_j \quad \rightarrow \quad H(I,J) = H(I) + H(J)$

grenswaarden $\quad 0 \leq H(J|I) \leq H(J) \leq \log_2 |S|$

Equivocatie

informatiekanaal met I als zender en J als ontvanger

equivocatie term (kanaalverlies, noise entropy)

$$H(J|I) = -\sum_i p_i \sum_j p_{j|i} \log p_{j|i}$$

perfecte communicatie

$$p_{j|i} = \begin{cases} 1 & \text{als } j = i \\ 0 & \text{als } j \neq i \end{cases} \quad \rightarrow \quad \begin{cases} H(J|I) = 0 & \text{equivocatie is nul} \\ H(I,J) = H(I) & \text{geen verlies} \end{cases}$$

imperfecte communicatie

$$0 < p_{j|i} < 1 \;\rightarrow\; H(J|I) > 0 \;\rightarrow\; H(I,J) = H(I) + H(J|I) > H(I)$$

definitie wederzijdse informatie

$$W(I,J) = H(J) - H(J|I) \quad (I = \text{zender},\ J = \text{b.v. cryptoanalist})$$

perfecte waarneming (bekende sleutel)

$$p_{j|i} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \rightarrow\ H(J|I) = 0 \ \rightarrow\ W(I,J) = H(J) = H(I)$$

imperfecte waarneming (cryptoanalyse)

$$0 < p_{j|i} < 1 \ \rightarrow\ 0 < H(J|I) < H(J) \ \rightarrow\ 0 < W(I,J) < H(J)$$

geen waarneming (onbreekbaar cryptosysteem)

$$p_{j|i} = p_j \ \rightarrow\ H(J|I) = H(J) \ \rightarrow\ W(I,J) = 0$$

markov keten lengte 1

$$\text{kans op QU is } p(x|Q) = \begin{cases} x = U & 1 \\ x \neq U & 0 \end{cases}$$

markov keten lengte 2

$$\text{kans op QUx is } p(x|QU) = \begin{cases} x = A & 0.37 \\ x \in E, I & 0.30 \\ x = O & 0.03 \\ x \notin A, E, I, O & 0.00 \end{cases}$$

markov keten lengte n

kans op $x_1 x_2 \ldots x_n x_{n+1}$ met $x_{n+1} = y$ is $p(y|x_1 x_2 \ldots x_n)$

bereken entropie markov keten

$$p_x = \sum_{i\ldots j}^{|S|} p_{i\ldots j} p_{x|i\ldots j}$$

$$H(I\ldots JX) = H(I\ldots J) + H(X|I\ldots J)$$

bepaal kansen uit tellingen $p(j|i) = \lim_{n\to\infty} \#ij/\#i$

entropie berichten van L letters

$$H(M^L) = H_1(S_1) + H_2(S_2|S_1) + H_3(S_3|S_1S_2) + \geq L.H_L(S_L|S_1 \ldots S_{L-1})$$

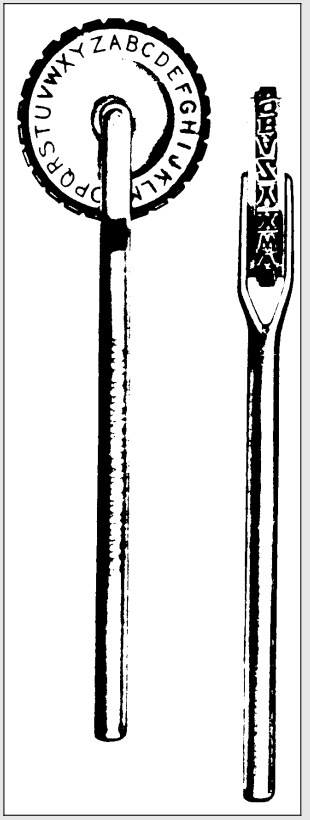
engelse taal, 26 letter alfabet, in bits per symbool

$$H_0 = 4.7 \quad (= \log_2 26)$$
$$H_1 = 4.2$$
$$H_2 = 3.6$$
$$H_\infty = 1.2\text{–}1.5$$

rate of the language H_∞ redundantie $D = H_0 - H_\infty$

Alfabet rundown

L=2		L=4		L=8	
AB	0.044	ABIG	0.139	ABIGPMZM	0.000
BC	0.000	BCJH	0.000	BCJHQNAN	0.000
CD	0.002	CDKI	0.008	CDKIROBO	0.000
DE	0.142	DELJ	0.041	DELJSPCP	0.000
EF	0.088	EFMK	0.000	EFMKTQDQ	0.000
..	..	..	..	..	..
RS	0.081	RSZX	0.000	RSZXGDQD	0.000
ST	0.254	STAY	0.658	STAYHERE	0.998
TU	0.046	TUBZ	0.000	TUBZIFSF	0.000
UV	0.007	UVCA	0.061	UVCAJGTG	0.000
..	..	..	..	..	..
ZA	0.002	ZAHF	0.003	ZAHFOLYL	0.000
H_2 = 3.182		H_4 = 1.719		H_8 = 0.022	



Entropie in Caesar systeem

known plaintext $W((M,C),K) = H(K) - H(K|M,C)$

$$H(M,C,K) = H(M|C,K) + H(C,K)$$
$$H(M,C,K) = H(K|M,C) + H(M,C)$$

$$H(M|C,K) = 0$$
$$H(C,K) = H(C) + H(K|C)$$
$$H(M,C) = H(C) + H(M|C)$$

$$H(K|M,C) = H(K|C) - H(M|C)$$

ciphertext only $W(M,C) = H(M) - H(M|C) \geq H(M) - H(K)$

Key appearance

Ciphertext-only analyse: sleutel bekend? Als $H(K|C) \approx 0$

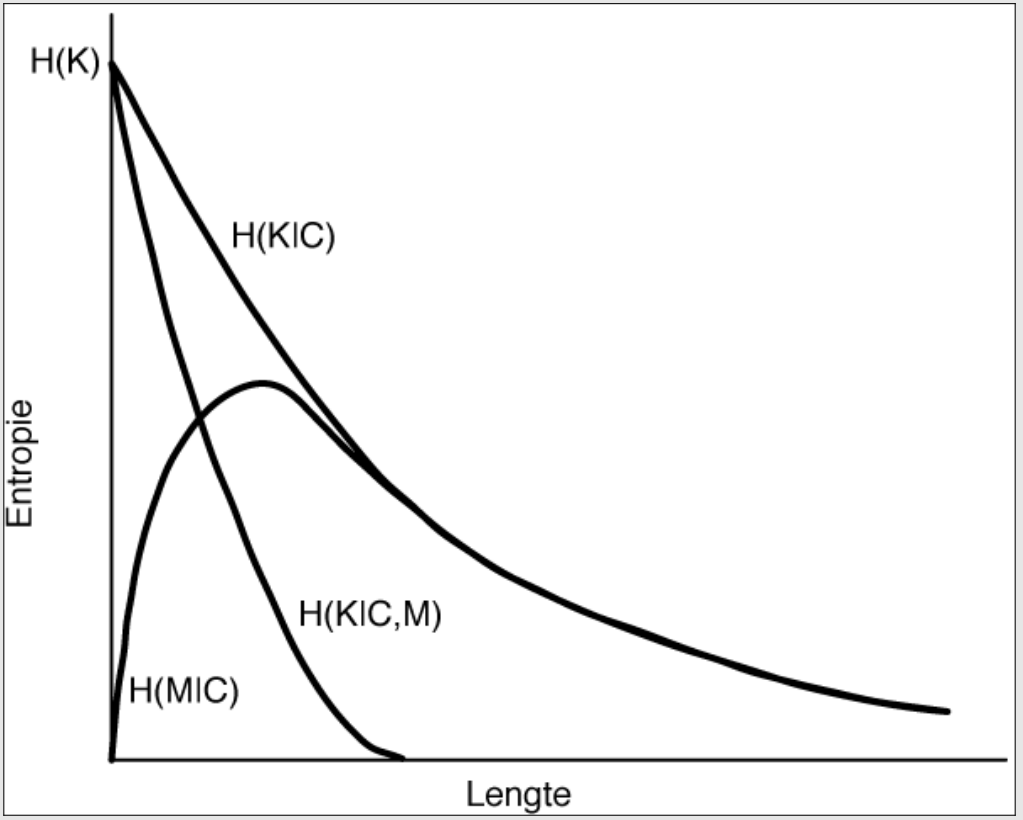
$$H(K|C) = H(C,K) - H(C)$$

$$\begin{aligned} M \overset{K}{\leftrightarrow} C &\rightarrow H(C,K) = H(M,K) \\ K \text{ onafhankelijk } M &\rightarrow H(M,K) = H(M) + H(K) \\ &\rightarrow H(K|C) = H(M) + H(K) - H(C) \end{aligned}$$

$$\begin{aligned} \text{Markov model} &\rightarrow H(M) \geq L \cdot H_\infty \text{ en } H(C) \leq L \cdot H_0 \\ &\rightarrow H(K|C) \geq H(K) + L \cdot H_\infty - L \cdot H_0 \end{aligned}$$

$$H(K|C) \approx 0 \quad \rightarrow \quad L \approx \frac{H(K)}{H_0 - H_\infty}$$ L heet *unicity distance*

Unicity distance



Unicity distance

$$L \geq \frac{H(K)}{H_0 - H_\infty}$$

$$= \frac{\log_2 26!}{4.7 - 1.2}$$

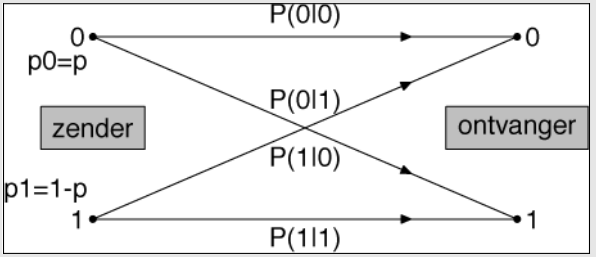
$$L \approx 25 \text{ letters}$$

monoalfabeet

$L \geq \frac{H(K)}{H_1 - H_i}$			
i	H_i	$L = 20$ $n \approx 2.5 10^{18}$	$L = 8 \times 8$ $n = 4^{16}$
1	4.2	∞	∞
2	3.6	111	55
3	3.2	65	32
7	2.8	45	22
∞	1.2	20	10

transpositie

Voorbeelden unicity distance



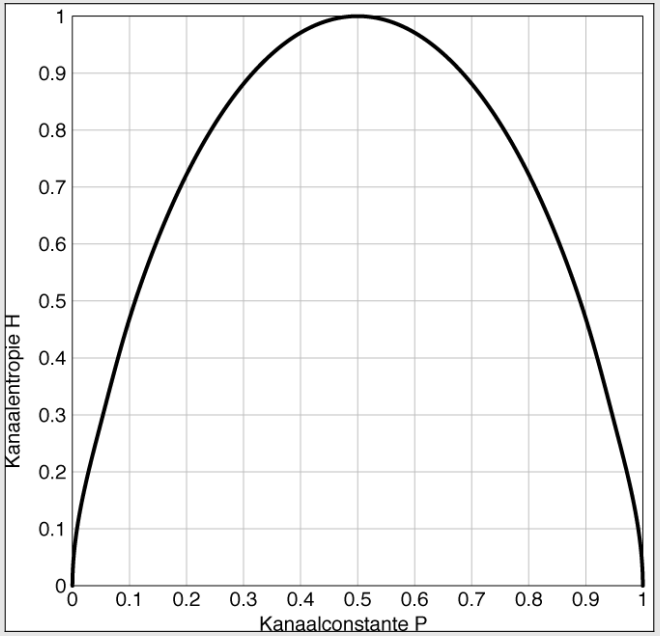
kanaalmatrix

$$P = \begin{pmatrix} p_{0|0} & p_{1|0} \\ p_{0|1} & p_{1|1} \end{pmatrix} = \begin{pmatrix} P & 1-P \\ 1-P & P \end{pmatrix} \text{ en } \vec{p} = (p, 1-p)$$

uitvoerdistributie

$$P \cdot \vec{p} = (1 - p - P + 2pP,\ p + P - 2pP)$$

Binair symmetrisch kanaal



$$H(\text{ontvanger}|\text{zender}) =$$

$$= -\sum_{i=0}^{1} p_i \sum_{j=0}^{1} p_{j|i} \log p_{j|i}$$

$$= -P \log P - (1-P) \log(1-P)$$

Kanaalentropie

kans op een zeker aantal oplossingen

$$\underline{M} = \#[k \in \{K\} : D(C,k) \in \{M\}] \quad \text{met} \quad \underline{M} \in \{1, \ldots, |K|\}$$

kans op de goede bij meerdere oplossingen

$$\underline{M} = m \quad \rightarrow \quad p_m = \frac{1}{m}$$

kans op één of meer oplossingen

$$P = \sum_{m=1}^{|K|} \frac{1}{m} p(\underline{M} = m)$$

kans op een oplossing

$$p_k = \begin{cases} 1 & \text{goede sleutel } k \\ 2^{-L.D} & \text{willekeurige sleutel } k \end{cases}$$

schakel echte oplossing uit

$$P(\underline{M}') \text{ met } \underline{M}' = \underline{M} - 1$$

kans bij binomiale verdeling

$$p(\underline{M}' = m') = \binom{|K| - 1}{m'} p_k^{m'} (1 - p_k)^{|K|-1-m'}$$

voeg echte oplossing weer toe

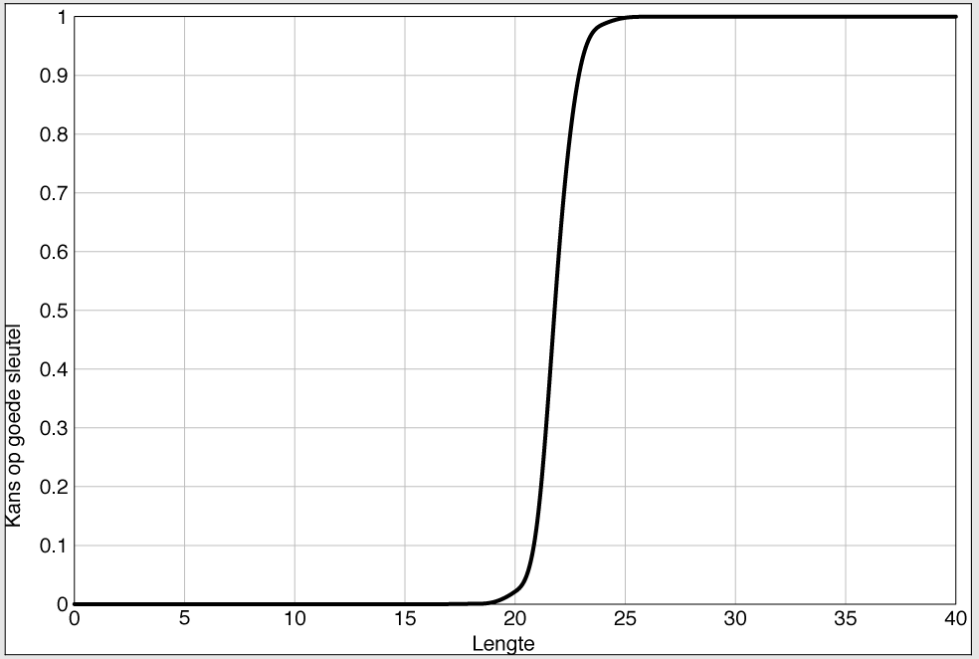
$$m' = m - 1 \quad \rightarrow \quad P = \sum_{m=1}^{|K|} \frac{1}{m} \binom{|K|-1}{m-1} p_k^{m-1} (1-p_k)^{|K|-m}$$

gebruik binomium van Newton

$$P = \frac{1}{p_k |K|} (1^{|K|} - (1-p_k)^{|K|})$$

benadering omdat kans klein

$$p_k \ll 1 \quad \rightarrow \quad P \approx \frac{1-e^{-\mu}}{\mu} \text{ met } \mu = p_k |K|$$



$$P = \frac{1 - e^{-\mu}}{\mu}$$

$$\mu = p_k |K|$$

$$p_k \approx 2^{-L(H_0 - H_\infty)}$$

$$H_0 = 4.7 \quad H_\infty = 2.0$$

$$|K| = 8.4_{10}{}^{17}$$

$$ud \approx 20 < L < 25$$

Sleutelkans