Random Cryptographic System

Plaintext Message Space $M = \{m_1, m_2, \dots, m_N\}$
Key Space $K = \{k_1, k_2, \dots, k_S\}$
Ciphertext Message Space $C = \{c_1, c_2, \dots, c_Q\}$

$$\boxed{c = E_k(m)}$$

The encrypting transformation corresponding to the key k

$$\boxed{m = D_k(c)}$$

The decrypting transformation corresponding to the key k

$$\boxed{\text{Three random variables} \begin{cases} M = \text{the chosen plaintext} \\ K = \text{the chosen key} \\ C = \text{the resulting ciphertext} \end{cases}}$$

Assumptions

$$\boxed{P(M = m_i) = p_i \text{ and } P(K = k_i) = q_i}$$

We choose the key K independently of the message M. Since we have the $C = E_K(M)$, the ciphertext is a random variable which depends on M and K. We also have $M = D_K(C)$. Therefore,

$$\boxed{H(K, C) = H(K, M)}$$

$H(K|C)$ = remaining uncertainty about the key after intercepting ciphertext

$H(K|C) = 0$ means that the ciphertext determines the key

$$
\begin{aligned}
H(K) + H(M) &= H(K, M) \\
&= H(K, C) \\
&= H(C) + H(K|C)
\end{aligned}
$$

For a ciphertext only attack

$$\boxed{H(K|C) = H(K) + H(M) - H(C)}$$

Unicity Distance

Definition The *unicity distance* is the smallest number of characters in the ciphertext that uniquely determines the plaintext.

Since we assume that the ciphertext uniquely determines the plaintext, the key must also be determined. Therefore

$$\boxed{H(K|C) = 0}$$

or in other words

$$\boxed{H(K) = H(C) - H(M)}$$

Unicity Distance for Caesar

Assume that we have just intercepted N letters of ciphertext that was encrypted using a Caesar shift. How large does N have to be (on average) in order the uniquely determine the shift? Assume that the entropy of the english language is 3.2 bits.

We begin with the identity:

$$\boxed{H(K) = H(C) - H(M)}$$

Assuming that each of the 26 keys is equally likely, we have

$$\boxed{H(K) = \log_2 26 \approx 4.7}$$

Assuming that each of the 26^N ciphertexts is equally likely, we have

$$\boxed{H(C) = \log_2 26^N = N \log_2 26 \approx 4.7N}$$

Therefore,

$$\boxed{4.7 = 4.7N - 3.2N \Rightarrow N \approx 3.13}$$

E	S	P	Y	L	E	T	G	P	D	L	C	P	C	P	D	E	W	P	D	D
F	T																			
G	U	R																		
H	V																			
I	W	T																		
J	X																			
K	Y																			
L	Z																			
M	A	X	G																	
N	B																			
O	C	Z																		
P	D																			
Q	E																			
R	F																			
S	G																			
T	H	E	N	A	T	I	V	E	S	A	R	E	R	E	S	T	L	E	S	S
U	I																			
V	J																			
W	K																			
X	L																			
Y	M																			
Z	N																			
A	O																			
B	P																			
C	Q																			
D	R	O	X																	

Monoalphabetic Substitution

Assume that we have intercepted N letters of a ciphertext message that was encoded using a Monoalphabetic substitution and that the entropy of english is 2 bits.

Length of text	5	10	15	20	30	40	50
# of distinct letters	4	8	11	12	14	16	18

For instance, a typical english sample of 30 letters contains about 14 di erent letters. Thus the key for a Monoalphabetic substitution only permutes 14 letters. Therefore the number of keys is

$$26 \times 25 \times \cdots \times 13$$

and not 26!.

Assuming that each key is equally likely, we have

$$\boxed{H(K) = \log_2(26 \times 25 \times \cdots \times 13) \approx 59.54}$$

Assuming that each of the 26^N ciphertexts is equally likely, we have

$$\boxed{H(C) = \log_2 26^N = N \log_2 26 \approx 4.7N}$$

Therefore,

$$\boxed{59.54 = 4.7N - 2N \Rightarrow N \approx 22.05}$$