Entropy:

$$H(X) = H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

if X has n outputs

1) $H(X) \leq \log_2(n)$

2) $H(X) = \log_2(n)$ iff X is uniform.

a random English letter has

$4.7 \approx \log_2(26)$ bits

of entropy.

Letters at random: 4.7 bits

frequency chart: 4.132 bits/letter.

bigrams: 3.56 bits/letter

full analysis: English has ~ 1.5 bits/letter.

redundancy 3.2 bits/letter

70%

Aoccdrnig to rscheearch at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a toatl mses and you can sitll raed it wouthit a porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe.

This is readable!

Xkkyrosl

xkkyorsl

Unicity Distance

Dfn: The <u>unicity distance</u> for a language + cipher is the length of cipher text to break a cipher (on average).

Ex: Caesar cipher

CWU → GAY

↓

VPN

not enough info!

UD of caesar is 2.

Ex: Vigenère cipher

ciphertext ABCDE

not breaking this.

Prop: $UD = \frac{\#\text{ bits in key}}{\text{redundancy of language}}$

Sketch of pf/ message length n, redundancy r has nr bits of redundancy.

If key has $< nr$ bits, can read the message.

Caesar cipher

25 possible keys

$\log(25) \approx 4.64$ bits

$UD \approx \frac{4.64}{3.2} \approx 1.45$

Simple substitution

$26! = 2^{88}$ possible keys

88 bits

$UD = \frac{88}{3.2} \approx 27.5$

Vigenere cipher

26^n keys

entropy

$\log(26^n) = n \log(26)$

$= 4.7n$

$UD = \frac{4.7n}{3.2} \approx 1.47n$.

One-time pad

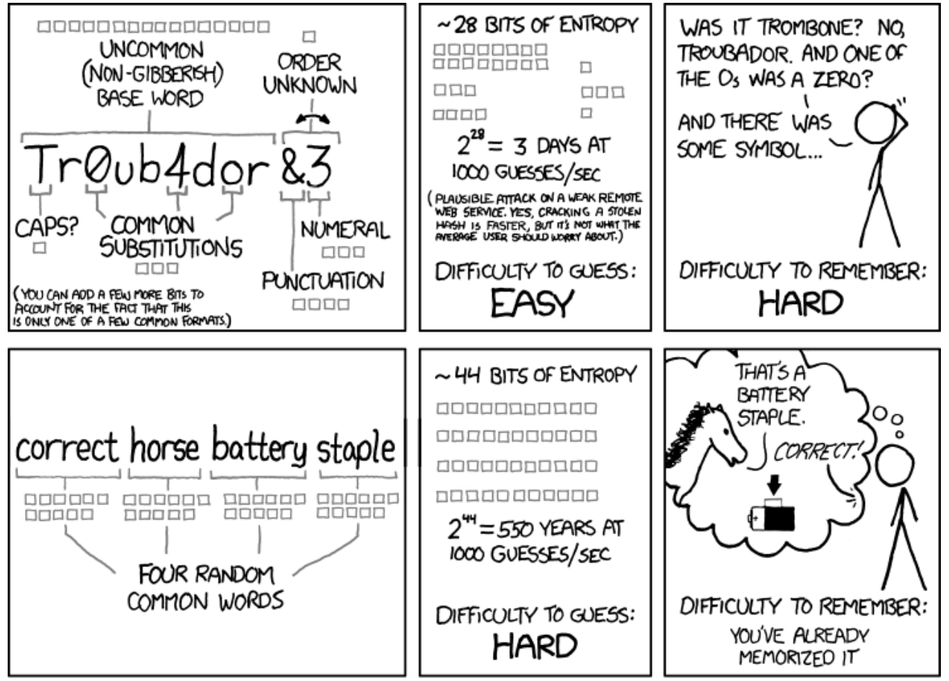
26^n keys

$4.7n$ bits

$UD = \frac{4.7n}{3.2} \approx 1.47n$

But message is length $n < 1.47n$.

xy3 7X 1042



§2.3 Inconvenience.

Diffusion and Confusion

Dfn: Good diffusion if changing 1 char of plaintext changes several of ciphertext.

Good	bad
Autokey-ish	monoalph
Hill	Vigenère

Dfn: Good confusion if each part of ciphertext depends on many parts of key.

Good	bad
Hill-ish	monoalph
	vigenère
	autokey

Complexity and big-O notation

Dfn: $f(x), g(x)$ pos fns

$f = O(g)$ if $\exists c, C > 0$ s.t.

$f(x) \leq c\, g(x) \quad \forall x \geq C$

Prop: If $\lim_{x \to \infty} \frac{f(x)}{g(x)}$ is finite,

then $f(x) = O(g(x))$

Ex: • If $f(x)$ is bounded,

$f(x) = O(1)$.

- $x = O(x^2)$
- $3x^2 = O(x^2)$
- $7x^2 + 5x + 3 = O(x^2)$
- $x^n = O(2^x)$ for any n.

Complexity: # of steps it takes to compute something.

f: # bits $\longrightarrow$ # steps

if algorithm takes $f(k)$ steps for k bits, and $f(k) = O(g(k))$ we say runtime of $O(g(k))$.

$O(k^e)$ = polynomial time

$O(k)$ = linear time

$O(1)$ = constant time.

} polynomial time fast

$O(\log(k))$ logarithmic time very fast

$O(2^{ck})$ exponential time (slow)

$O(2^{\varepsilon k})\ \forall \varepsilon > 0$ subexponential. (medium speed)

e.g. $O(2^{\sqrt{k}})$

Caesar cipher message length n.

encrypt $O(n)$

break: $O(1)$

Vigenère cipher

encrypt: $O(n)$

Kasiski: $O(n^3)$ naively

$O(n^2)$ cleverly

IOC: $O(n)$

Hill cipher

encrypting: $O(n^2)$

breaking brute force $O(n^3 26^{n^2})$

clever $O(n\ 13^n)$

Very secure against ciphertext-only attacks

Very vulnerable to known-plaintext attack.

(I think $O(n^2)$?)