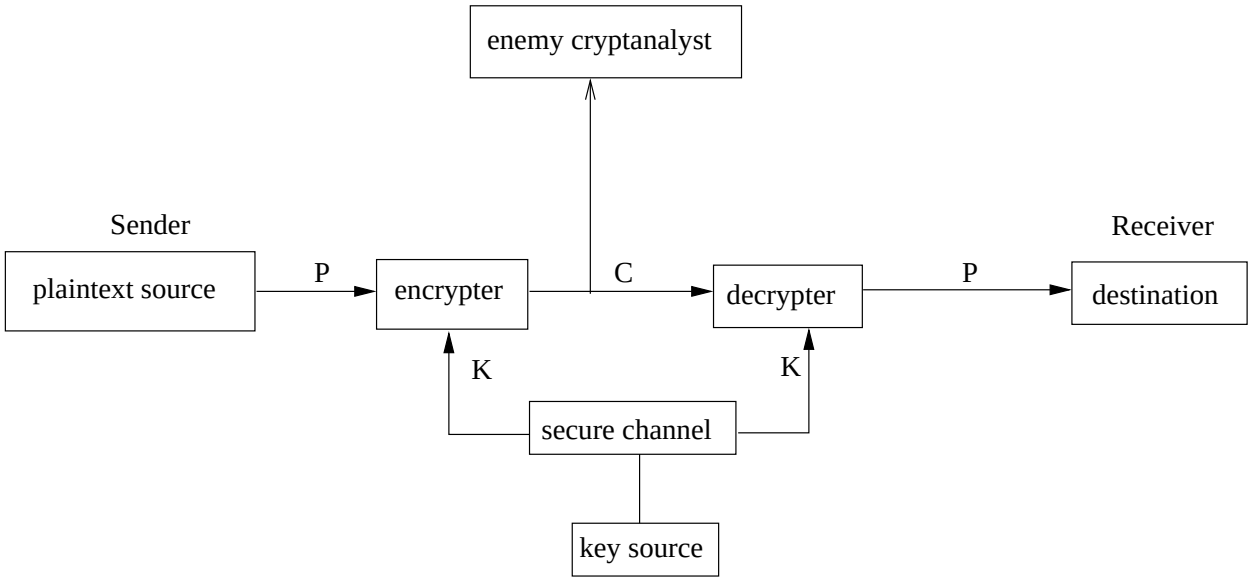
Shannon's Model of a Secrecy System

Diagram as in [Sha49].



- Encryption encodes a message so its meaning is not obvious.
- For symmetric encryption $P = D(K, E(K, P))$.
- For asymmetric encryption $P = D(K_D, E(K_E, P))$.
- Security of a cryptosystem *should rest entirely in the secrecy of the key*, and not in the secrecy of the algorithm (Kerckhoffs).
- Cryptographers design their algorithms to resist the following increasingly aggressive attacks [Susan Landau]:
  - **Ciphertext-only**: adversary has access to encrypted comms.
  - **Known-plaintext**: adversary has some (plaintext, ciphertext).
  - **Chosen-text**: the adversary chooses
    - ▷ the plaintext to be encrypted
    - ▷ the ciphertext to be decrypted (chosen ciphertext)
    - ▷ the plaintext to be decrypted depending on ciphertext received from previous requests (adaptive chosen plaintext)

Monoalphabetic Substitution Cipher

Caesar's cipher is a very simple permutation, for e.g., rot13. An example is shift by 3 (rot3).

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c

So the secret message *"attack at dawn"* is encrypted as

a	t	t	a	c	k		a	t		d	a	w	n
$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
d	w	w	a	f	n		d	w		g	d	z	q

Keyspace: 25.

Can be broken using cipher text only.

Might give a demo using vigenere-encrypt.cgi with the above as example.

Monoalphabetic Substitution Cipher

In general, the secret key is a table which is a permutation π that maps each symbol of PT onto a symbol of CT, for e.g.,

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
p	a	n	d	o	r	s	b	x	c	e	f	g	h	i	j	k	l	m	q	t	u	v	w	y	z

Show how "attack at dawn" is encrypted now using mono-encrypt.cgi.

Keyspace: $26!$ (permutations). At 1 decipherment per μs, it'd take over 10^3 years to cycle through. Constructing an easy to remember permutation is easy with a keyword, such as "pandorasbox" in above.

Unfortunately, one can use frequency of English letters

e	12.5%
t	9.25%
a	8.04%
o	7.60%
i	7.26%
n	7.09%

and pairs to break this cipher using a *ciphertext only* attack. Most common *digram*: **th**, most common *trigram*: **the**.

Example of monoalphabetic cipher cryptanalysis

Example from [Kip99]:

```
ETNAN XFWN LYK Y RYETNA QF EBWKXF LTX KYQP ETQK YPHQWN QK RXA DXB KXF
DXB PXFE LYKT DXBAKNMR LNMM KX DXBA RNNE KCNMM MQUN TNMM
QR QF VNP LQET Y ZQAM UNNI DXBA KTXNK XF
```

Set $N = E$ as N is the most frequent letter; now search for *the*, 3 times N is preceded by T, set $T = H$, 2 times TN is preceded by E, set $E = T$. Notice the lone Y, set $Y = A$. The first word could be *there*, so set $A = R$. If $A = R$, then $R \neq R$, maybe $R = F$.

- Same letters in plaintext encrypt to same letters in ciphertext.

Try http://localhost/info/cgi-bin/mono.cgi for a demo.

Playfair Cipher

≈ 1854 by Sir Charles Wheatstone.

5×5 matrix of letters constructed using a keyword [Sta99].

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

ATTACKATDAWN
RSSRDERSBRNY

In general, insert "i" between successive identical letters to avoid needing to encrypt pairs such as "tt".

- Great advance over simple monoalphabetic ciphers.
- Used by the British Army in WW I.
- Frequency analysis more difficult.
- Flatter distribution than plaintext, nevertheless plenty of structure.
- [Sta99, Fig. 2.7]

Vigenère Cipher

Invented circa 1520. Applied arithmetic to ciphers.

```
whatanicedaytoday
cryptocryptocrypt
-----------------
yyyitb..........
```

Use the Vigenère tableau to encrypt or decrypt messages. It's like n instances of Cæsar's cipher. Or, it's addition modulo 26 where $a = 0, \ldots, z = 25$.

- Keyspace: 26^n, n is the number of symbols in the key.
- It evens out the frequency disparity in the plaintext alphabet.

Vigenère's tableau (part of)

	abcdefghijklmnopqrstuvwxyz
a	abcdefghijklmnopqrstuvwxyz
b	bcdefghijklmnopqrstuvwxyza
c	cdefghijklmnopqrstuvwxyzab
d	defghijklmnopqrstuvwxyzabc

Cryptanalysis of Vigenère Cipher

Easy to break [Fri84, pg. 17], [Sta99, pg. 40], [Pfl96, pg. 35]:

- Find key length. **Kasiski**: Identical sequences of plaintext at integral multiples of keyword length $\equiv$ identical ciphertext sequences. Look for common factors.
  - English uses several endings and beginnings disproportionately often.
  - Words such as *of*, *and*, *to* etc. appear in high frequency.
- Divide the cipher text into key length sized blocks. All elements corresponding to the same relative position within each block form a *monoalphabetic cipher*. Break each for every position of the block.
- Cipher-text only attack.

Cryptanalysis of Vigenère Cipher

Use the **Index of Coincidence**. It is defined as the probability that two randomly selected letters in a ciphertext are identical.

$$IC = \frac{\sum \binom{n_i}{2}}{\frac{1}{2}n(n-1)}$$

where n_i is the # of occurrences of symbol i.

Example: Consider the ciphertext

```
WSPGM HHEHM CMTGP NROVX WISCQ TXHKR
VESQT IMMKW BMTKW CSTVL TGOPZ XGTQM
CXHCX HSMGX WMNIA XPLVY GROWX LILNF
JXTJI RIRVE XRTAX WETUS BITJM CKMCO
TWSGR HIRGK PVDNI HWOHL DAIVX JVNUS
JX
```

The counts of the various letters are:

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
3	2	7	2	4	1	8	9	10	5	5	5	11	5	5	5	3	8	8	12	2	8	9	13	1	1

The total # of letters is 152. Thus

$$\begin{aligned} IC &= \frac{\sum_{i=0}^{25} n_i(n_i-1)}{n(n-1)} \\ &= \sum 3 \cdot 2 + 2 \cdot 1 + \cdots 1 \cdot 0 + 1 \cdot 0 \\ &= \frac{1048}{152 \cdot 151} \\ &= .0457 \end{aligned}$$

Index of Coincidence

IC is a measure of the variation between frequencies in a distribution [Pfl96, Section 2.3]. If λ represents a plaintext symbol, then $\sum_\lambda P_\lambda = 1$. Let's find the variation of a given distribution from a **flat** distribution $P_\lambda = 1/|\lambda|$.

$$\begin{aligned} var &= \sum_{\lambda=a}^{\lambda=z} \left(P_\lambda - \tfrac{1}{26}\right)^2 \\ &= \sum_{\lambda=a}^{\lambda=z} P_\lambda^2 - \tfrac{1}{26} \end{aligned}$$

Now, $\sum P_\lambda^2 \approx \sum \frac{n_i}{n} \cdot \frac{n_i - 1}{n-1} \approx IC = var + const$!

- IC is a predictor of key length when it is small. It cannot discriminate well for large key lengths.

keylen	1	2	3	4	5	10	large
IC	.068	.052	.047	.044	.044	.041	.038

Permutation Cipher

Columnnar transposition.

Consider the plain text "howareyoudoing". Write this as two blocks of seven characters each

h	o	w	a	r	e	y
o	u	t	o	d	a	y

The cipher text is the plaintext read in column order. So the cipher text is "hoouwtaordeayy".

- Same letter frequencies as original text.
- Can be broken using a form of frequency analysis.
- Can be broken with a KPA.

Generic permutation.

Here the permutation is on the position of PT symbols in the corresponding CT. For e.g., HELLOWORD might be transformed into LWHOEROLD. An example permutation is

$$\Pi = \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 3 & 5 & 1 & 4 & 2 & 7 & 9 & 6 & 9 \end{array}$$

Hill Cipher

Lester Hill, 1929.

$$C = KP \mod 26,\ 0 \leq k_{ij} < 26$$

- Strong against ciphertext only attack, but easily broken under known plaintext attack i.e., given a set of (P, C) pairs, solve for K.

"Perfect" Substitution Cipher

- Use an infinite *nonrepeating* sequence as key. Confounds both Kasiski and Index of Coincidence.
  - One-time pad. Difficulties: 1) Need for synchronization between sender and receiver 2) Need for unlimited number of keys.
  - Long Random Number Sequences. Can be used at bit level [Gilbert Vernam, 1918]. Difficulty: (pseudo) random $\not\Rightarrow$ unpredictable.

    A **Linear Congruential Generator** is of the form

    $$r_{i+1} = (a \times r_i + b) \bmod n$$

    where a, b, n are constants. It's totally **linear**!
    For e.g., given the random sequence 958833456, 396607904, 2147285887 for $n = 2^{31} - 1$, we have the equations

    $$\begin{array}{rcll} 396607904 & = & a \times 958833456 + b & \bmod 2147483647 \\ 2147285887 & = & a \times 396607904 + b & \bmod 2147483647 \\ \hline 396805664 & = & a \times 562225552 & \bmod 2147483647 \\ a & = & 16807 & \end{array}$$

  - Long Sequences from Books.
    - ▷ Digits from the phone book. (might have some *non-uniformity*)

Rotor Machines

Lessons

- Compress before you encrypt.

References

[Fri84] William Friedman. *Military Cryptanalysis II.* Aegean Park Press, 1984.

[Kip99] Rudolph Kippenhahn. *Code Breaking – A History and Exploration.* The Overlook Press, Peter Mayer Publishers Inc., Lewis Hollow Road, Woodstock, New York 12498, 1999.

[Pfl96] Charles P. Pfleeger. *Security in Computing.* Prentice Hall, iind edition, 1996.

[Sha49] Claude E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28:656–715, October 1949.

[Sta99] William Stallings. *Cryptography and Network Security.* Prentice-Hall Inc., 2nd edition, 1999.