Symmetric Key Encryption

Göran Andersson

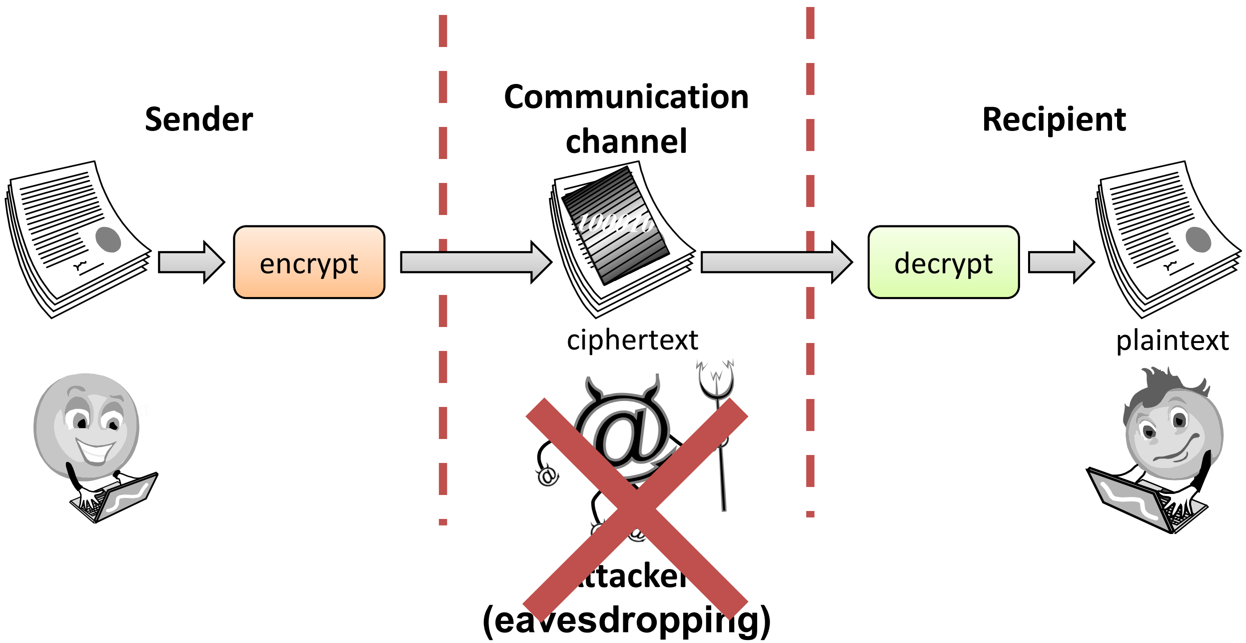
goeran@kth.se

Cryptographic Concepts

Short Review

Encryption and Decryption

- The message M is called the **plaintext.**
- Alice encrypts M using an algorithm E that outputs a **ciphertext C for M.**
- Bob decrypts C using an algorithm D that outputs the plaintext M.



Encryption and Decryption

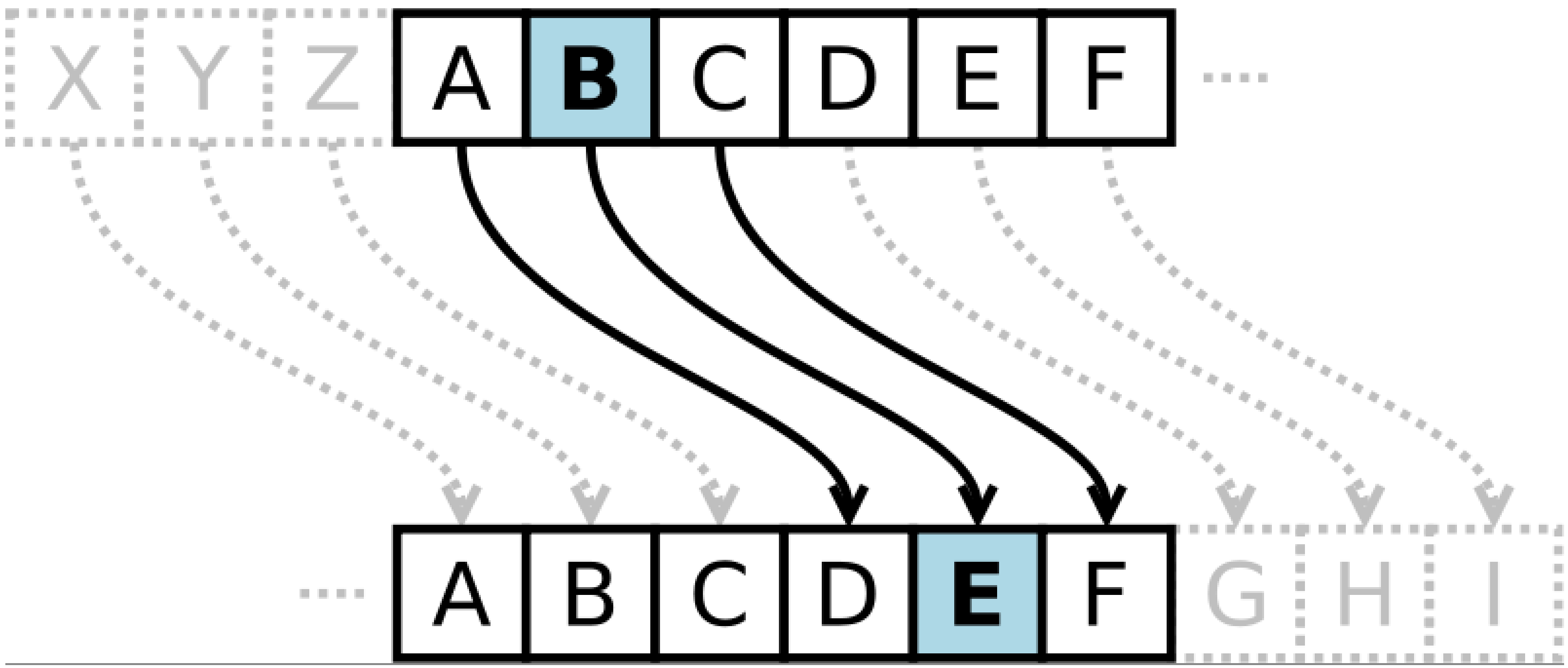
As equations:

$$C = E(M)$$

$$M = D(C)$$

Example Caesar Cipher

- Replace each letter with the one “three over” in the alphabet.



Public domain image from http://commons.wikimedia.org/wiki/File:Caesar3.svg

Example Caesar Cipher

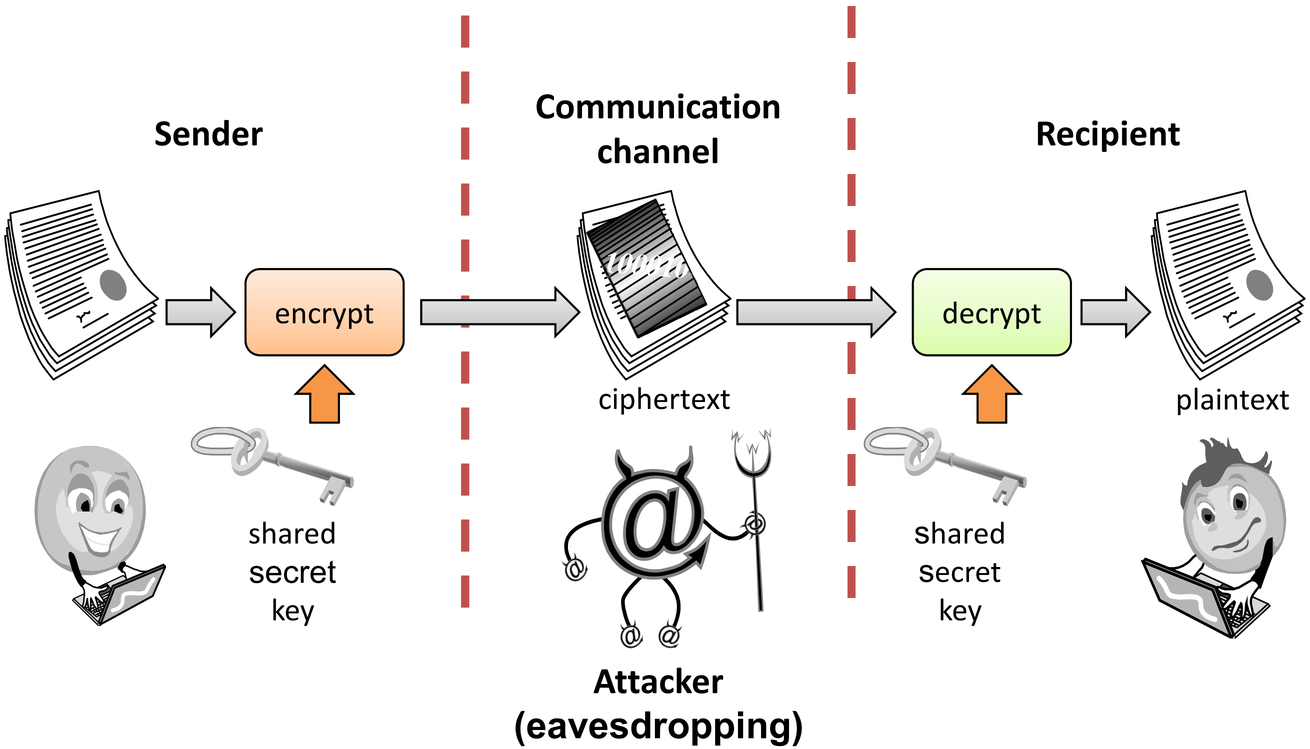
As equations:

$$C_i = E(M_i) = \text{symbol}(M_i) + 3 \pmod{26}$$

$$M_i = D(C_i) = \text{symbol}(C_i) - 3 \pmod{26}$$

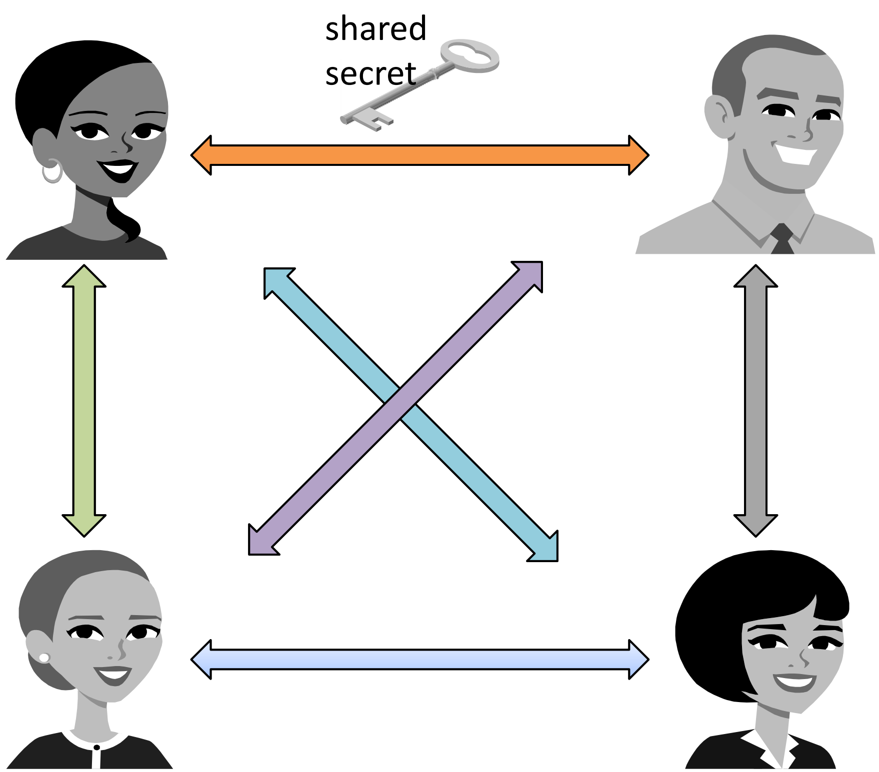
Symmetric Cryptosystems

- Alice and Bob share a secret key, which is used for both encryption and decryption.



Symmetric Key Distribution

- Requires each pair of communicating parties to share a (separate) secret key.



Complete graph

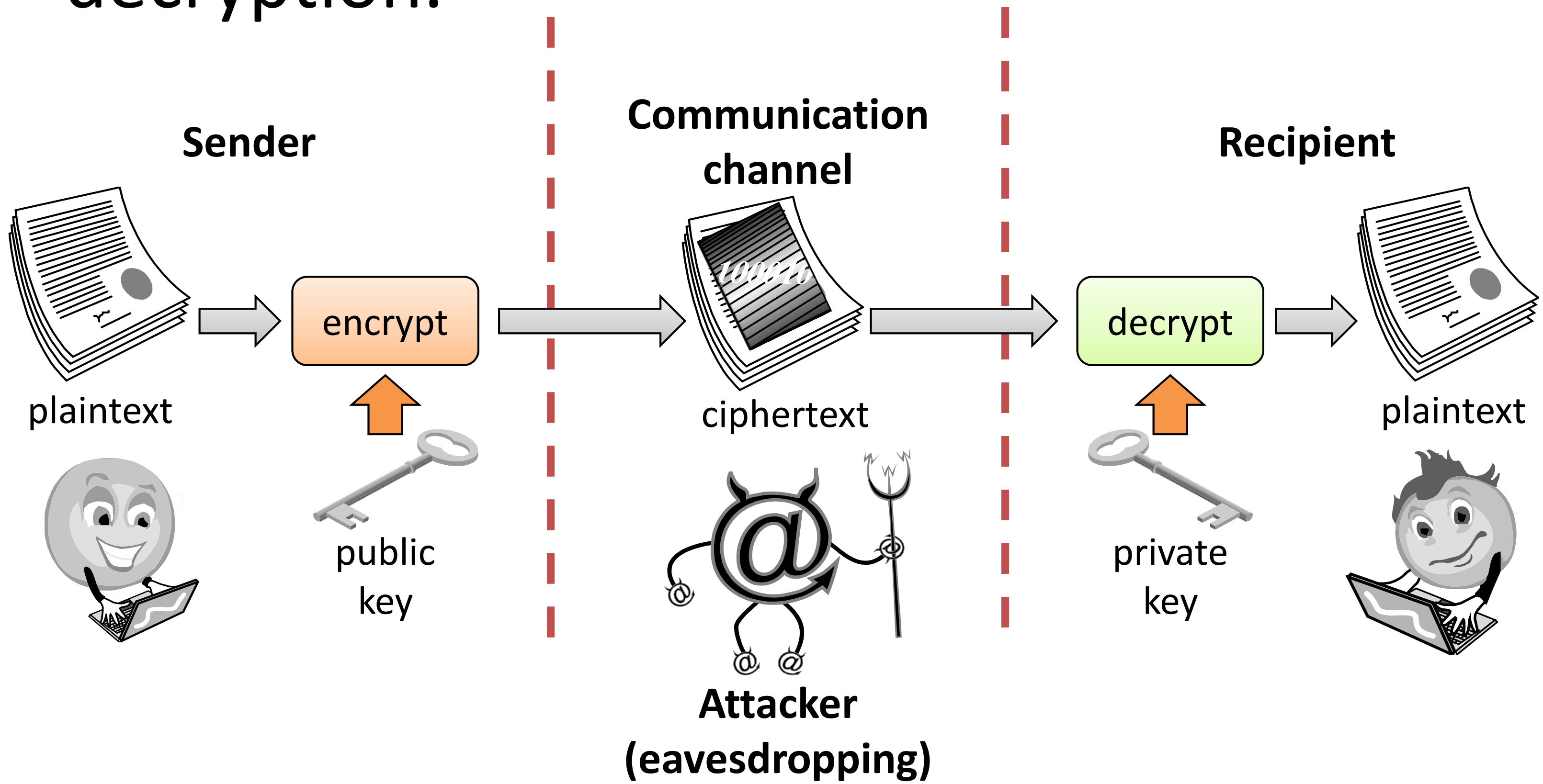
$$\binom{n}{2} \text{ keys}$$

Public-Key Cryptography

- Bob has two keys: a **private key,** S_B, which Bob keeps **s**ecret, and a **public key,** P_B, which Bob broadcasts widely.
- Alice encrypts using Bob's public key, P_B,
- $C = E_{P_B}(M)$
- Bob then uses his private key to decrypt the message
- $M = D_{S_B}(C)$.

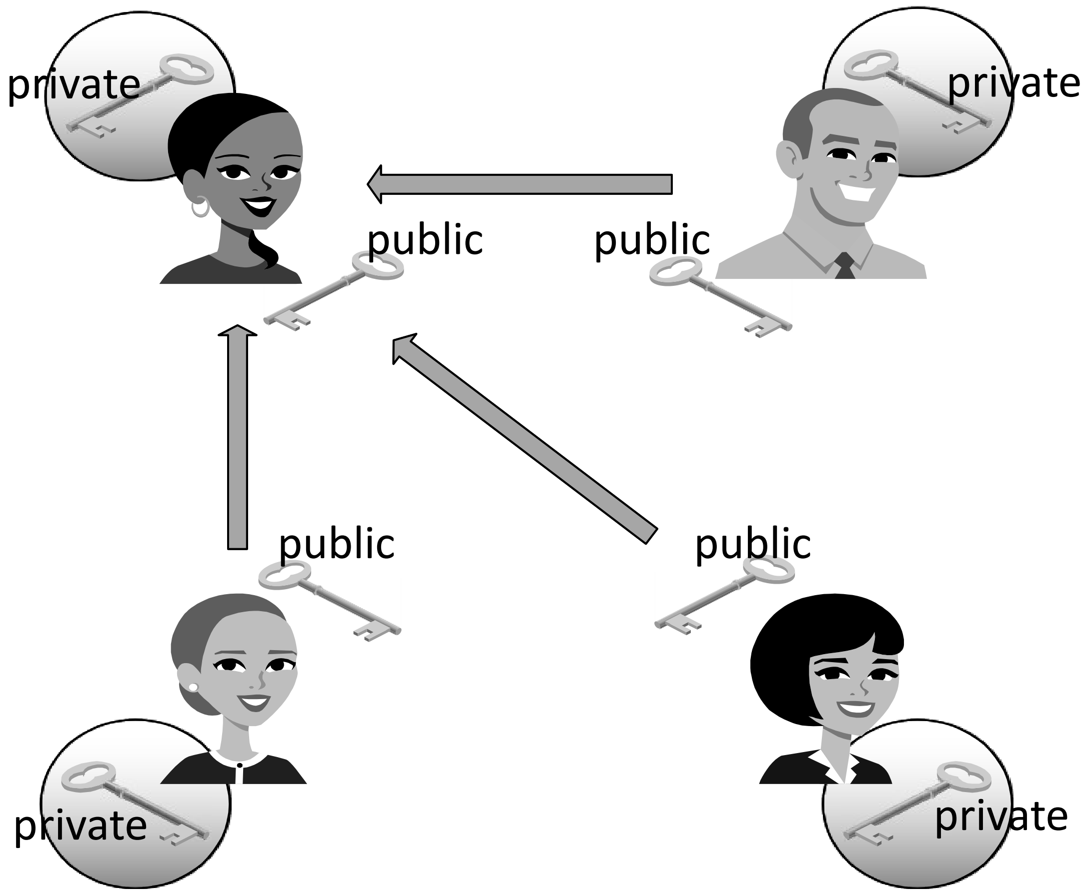
Public-Key Cryptography

- Separate keys are used for encryption and decryption.



Public Key Distribution

- Only one key pair is needed for each participant



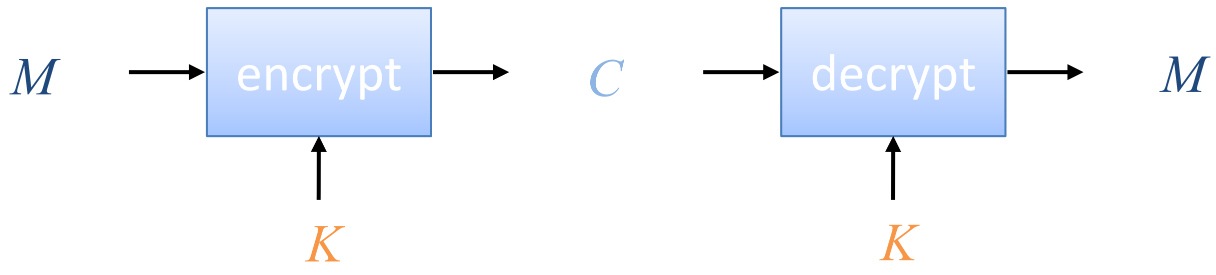
$$2\,n \quad \text{keys}$$

$$n > 5 \Rightarrow 2\,n < \binom{n}{2}$$

Symmetric Key Encryption

Symmetric Cryptosystem

- Scenario
  - Alice wants to send a message (plaintext M) to Bob.
  - The communication channel is insecure and can be eavesdropped
  - If Alice and Bob have previously agreed on a symmetric encryption scheme and a secret key K, the message can be sent encrypted (ciphertext C)
- Issues
  - What is a good symmetric encryption scheme?
  - What is the complexity of encrypting/decrypting?
  - What is the size of the ciphertext, relative to the plaintext?



Basics

- Notations
  - Secret key K
  - Encryption function $E_K(M)$
  - Decryption function $D_K(C)$
  - Plaintext length typically the same as ciphertext length
  - Encryption and decryption are permutation functions (bijections) on the set of all n-bit arrays
- Efficiency
  - functions E_K and D_K should have efficient algorithms
- Consistency
  - Decrypting the ciphertext yields the plaintext
  - $D_K(E_K(M)) = M$

Entropy of Natural Language

- Information content (entropy, H) of English: 1.25 bits per character (byte) n-bit arrays that are English text:

$$H(2^{1.25\, n/8}) = \log_2(2^{1.25\, n/8})$$
$$= 1.25\, n/8 \approx 0.156n = \alpha n$$

- For a natural language, constant $\alpha < 1$ such that there are $2^{\alpha n}$ messages among all n-bit arrays
- Redundancy

$D = (1 - \alpha)$

- Fraction (probability) of valid messages in English

$$2^{\alpha n} / 2^n = 1/2^{(1-\alpha)n} \approx 0.56^n$$

- Brute-force decryption
  - Try all possible 2^k decryption keys
  - Stop when valid plaintext recognized
- Given a ciphertext, there are 2^k possible plaintexts
- Expected number of valid plaintexts

$$2^k/2^{(1-\alpha)n}$$

- Expected unique valid plaintext, (no spurious keys) achieved at unicity distance

$$2^k/2^{(1-\alpha)U} = 1 \Leftrightarrow U = k/(1-\alpha) = H(2^k)/D$$

- This is minimum #bits to get the key by brute-force

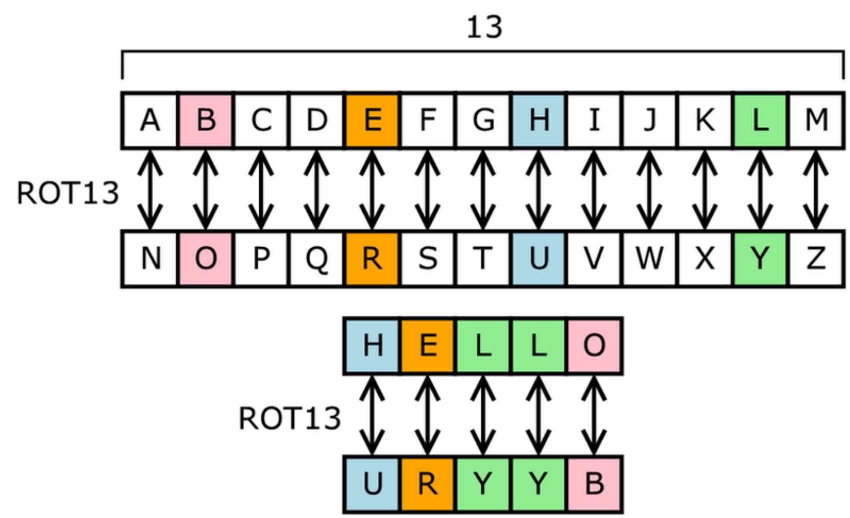
Problem

- Find the unicity distance for a 256 bit key using English.

Substitution Ciphers

- Each letter is uniquely replaced by another.
- There are $26! \approx 4 \times 10^{26}$ possible substitution ciphers.

- One popular substitution “cipher” for some Internet posts is ROT13.



Public domain image from http://en.wikipedia.org/wiki/File:ROT13.png

Frequency Analysis

- Letters in a natural language, like English, are not uniformly distributed.
- Knowledge of letter frequencies, including pairs and triples can be used in cryptologic attacks against substitution ciphers.

a:	8.05%	b:	1.67%	c:	2.23%	d:	5.10%
e:	12.22%	f:	2.14%	g:	2.30%	h:	6.62%
i:	6.28%	j:	0.19%	k:	0.95%	l:	4.08%
m:	2.33%	n:	6.95%	o:	7.63%	p:	1.66%
q:	0.06%	r:	5.29%	s:	6.02%	t:	9.67%
u:	2.92%	v:	0.82%	w:	2.60%	x:	0.11%
y:	2.04%	z:	0.06%				

Letter frequencies in the book *The Adventures of Tom Sawyer*, by Twain.

Substitution Boxes

- Substitution can also be done on binary numbers.
- Such substitutions are usually described by substitution boxes, or S-boxes.

	00	01	10	11
00	0011	0100	1111	0001
01	1010	0110	0101	1011
10	1110	1101	0100	0010
11	0111	0000	1001	1100

(a)

	0	1	2	3
0	3	8	15	1
1	10	6	5	11
2	14	13	4	2
3	7	0	9	12

(b)

Figure 8.3: A 4-bit S-box (a) An S-box in binary. (b) The same S-box in decimal.

S-box

	A	B	C	D	E	F	…						
A	PY	BI	HF	SS	PA	SV	II	RH	ZM	MJ	RV	GX	O
B	MP	TJ	YR	IV	TQ	LF	TC	WV	VX	MN	MB	OI	X
C	CQ	KG	YP	AN	QP	EN	TH	RI	ZD	VY	OP	LT	T
D	EE	PC	KP	MS	RC	WC	NV	CU	GV	TI	XS	GP	D
E	FB	AF	ZW	DV	DR	OO	OR	JZ	IP	RK	KY	SL	H
F	GL	AP	ZT	US	BE	RA	YG	TK	BA	SF	WP	WH	N
…	FI	XV	DJ	ZR	SG	WF	JQ	DX	KT	RG	SC	VF	H
	WX	VQ	QJ	XZ	ZC	WR	FC	HL	CX	YV	LN	TW	Z
	XL	OJ	VU	UA	HY	CS	OL	HK	IC	EV	IK	QQ	E
	XB	GC	ZU	FD	MU	CF	NC	ZS	NS	KD	TF	WM	S

One-Time Pads

- There is one type of substitution cipher that is absolutely unbreakable.
  - The **one-time pad** was invented in 1917
  - We use a block of shift keys, $(k_1, k_2, \ldots, k_n)$, to encrypt a plaintext, M, of length n, with each shift key being chosen uniformly at random.
- Since each shift is random, every ciphertext is equally likely for any plaintext.

Vigenère Cipher, Published 1586

Encryption:

message = "ATTACKATDAWN"
key = "LEMON"

$\mathbb{m} = \{0, 19, 19, 0, 2, 10, 0, 19, 3, 0, 22, 13\}$

$\mathbb{k} = \{11, 4, 12, 14, 13, 11, 4, 12, 14, 13, 11, 4\}$

$\mathbb{c} = (\mathbb{k} + \mathbb{m}) \bmod 26$

$\{11, 23, 5, 14, 15, 21, 4, 5, 17, 13, 7, 17\}$

cipher = "LXFOPVEFRNHR"

Decryption:

$(\mathbb{c} - \mathbb{k}) \bmod 26$

Problem

- Find the Vigenère Cipher
  - message = ATTACKNOW
  - key = APPLE

Pseudo Random Number Generators

Desired properties for a PRNG

- Uniform distribution
- Independent numbers
- Very long period
- A simple PRNG is a linear congruential generator

$$x_{i+1} = a\,x_i + b \bmod m$$

- $a \in [1, m-1]$, $b \in [0, m-1]$

Linear Congruential Generator

$$x_{i+1} = a\, x_i + b \bmod m$$

- The period of this generator is at most $\phi(m)$
- Choosing a as a primitive root to m gives max period
- x_0 and b has to be chosen properly
- If m is a prime p then $\phi(p) = p - 1$ is the max period
- If $b = 0$ then $x_i = 0$ can be avoided and all values $[1,\ p - 1]$ are represented

Linear Congruential Generator

$$x_{i+1} = a\,x_i + b \bmod m$$

- If $m = p = 7$ then the primitive roots are $\{3, 5\}$
- $a = 3$ gives max period $\phi(7) = 7 - 1 = 6$
- $b = 0$ and $x_0 \neq 0$ results in a uniform sequence $[1, 6]$
- I.e. we've got a die
- However this sequence is completely predictable
- We have to use very large p and change a
- Then use $r_i = x_i / p$ that results in $r_i \in (0, 1)$
- Finally use $y_i = \text{floor}(6\, r_i + 1)$ for our die

Problem

- Make a primitive die in the interval [1,10]

Stream Cipher

- Key stream
  - Pseudo-random sequence of bits $S = S[0], S[1], S[2], \ldots$
  - Can be generated on-line one bit (or byte) at the time
- Stream cipher
  - XOR the plaintext with the key stream $C[i] = S[i] \oplus M[i]$
  - Suitable for plaintext of arbitrary length generated on the fly, e.g., media stream
- Synchronous stream cipher
  - Key stream obtained only from the secret key K
  - Works for unreliable channels if plaintext has packets with sequence numbers
- Self-synchronizing stream cipher
  - Key stream obtained from the secret key and q previous ciphertexts
  - Lost packets cause a delay of q steps before decryption resumes

Stream Cipher (Binary Pad)

Encryption:

message = "This secret message cannot be revealed!"

$m = \{84, 104, 105, 115, 32, 115, 101, 99, 114, 101, 116, 32, 109, 101, 115, 115, 97, 103, 101, 32, 99, 97, 110, 110, 111, 116, 32, 98, 101, 32, 114, 101, 118, 101, 97, 108, 101, 100, 33\}$

$s = \{149, 134, 11, 33, 12, 64, 182, 249, 26, 136, 199, 132, 171, 64, 36, 54, 149, 200, 110, 145, 84, 25, 217, 179, 246, 145, 149, 210, 81, 13, 254, 55, 93, 6, 46, 23, 133, 78, 224\}$

$c = m \oplus s$

$\{193, 238, 98, 82, 44, 51, 211, 154, 104, 237, 179, 164, 198, 37, 87, 69, 244, 175, 11, 177, 55, 120, 183, 221, 153, 229, 181, 176, 52, 45, 140, 82, 43, 99, 79, 123, 224, 42, 193\}$

cipher = "ÁîbR,3Ó   hí³¤Æ%WEô¯ ±7x·Ý   åµ°4–   R+cO{à∗Á"

Decryption:

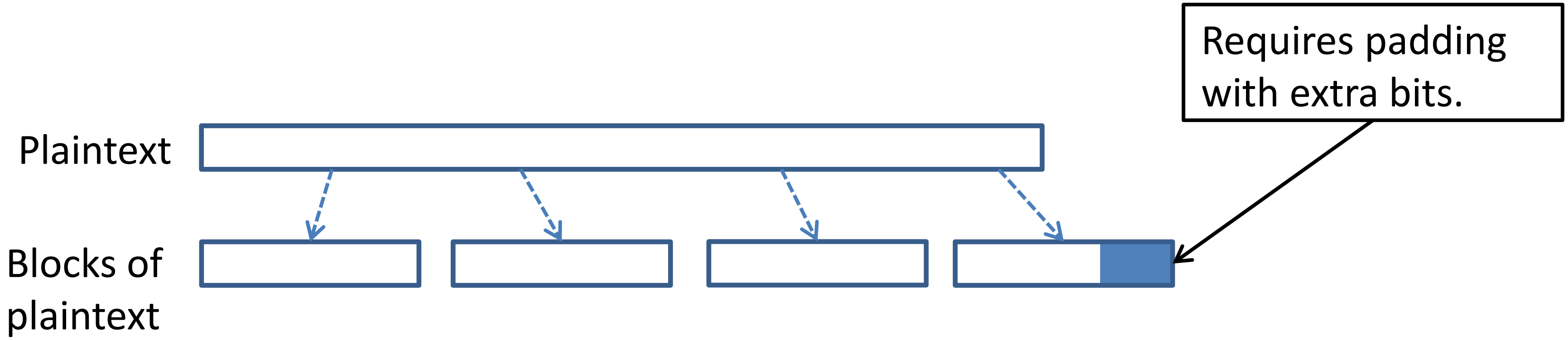
$r = c \oplus s$

Key Stream Generation

- RC4
  - Designed in 1987 by Ron Rivest for RSA Security
  - Trade secret until 1994
  - Uses keys with up to 2048 bits
  - Simple algorithm

Block Ciphers

- In a **block cipher:**
  - Plaintext and ciphertext have fixed length b (e.g., 128 bits)
  - A plaintext of length n is partitioned into a sequence of m **blocks**, $M[0], \ldots, M[m-1]$, where $n \leq bm < n + b$
- Each message is divided into a sequence of blocks and encrypted or decrypted in terms of its blocks.



Padding

- Block ciphers require the length n of the plaintext to be a multiple of the block size b
- Padding the last block needs to be unambiguous (cannot just add zeroes)
- When the block size and plaintext length are a multiple of 8, a common padding method (PKCS5) is a sequence of identical bytes, each indicating the length (in bytes) of the padding
- Example for $b = 128$ (16 bytes)
  - Plaintext: “Roberto” (7 bytes)
  - Padded plaintext: “Roberto999999999” (16 bytes), where 9 denotes the number and not the character
- We need to always pad the last block, which may consist only of padding

The Hill Cipher

- Block cipher invented 1929
- English letters are treated as numbers $\bmod 26$
- The key $\boldsymbol{K}$ is an invertible $n \times n$ matrix $\bmod 26$
- Message partitioned in n-block (column vectors) and padded
- Encryption: $\boldsymbol{C} = \boldsymbol{K} \cdot \boldsymbol{M} \bmod 26$
- Decryption: $\boldsymbol{M} = \boldsymbol{D} \cdot \boldsymbol{C} \bmod 26$, where $\boldsymbol{D} = \boldsymbol{K}^{-1} \bmod 26$

- If $\boldsymbol{K}^{-1}$ and $d = \det(\boldsymbol{K})$ are known then
  $\boldsymbol{D} = [(d^{-1} \bmod 26)\ (d\ \boldsymbol{K}^{-1})] \bmod 26$

The Hill Cipher

Example:

$$\boldsymbol{K} = \begin{pmatrix} 1 & 0 & 11 \\ 11 & 16 & 24 \\ 7 & 17 & 1 \end{pmatrix}$$

$$d = 433 \Rightarrow d^{-1} \equiv_{26} 23$$

$$\text{check} : 433 \times 23 = 9959 \equiv_{26} 1$$

$$d\,\boldsymbol{K}^{-1} = \begin{pmatrix} -392 & 187 & -176 \\ 157 & -76 & 97 \\ 75 & -17 & 16 \end{pmatrix}$$

$$23\,d\,\boldsymbol{K}^{-1} = \begin{pmatrix} -9016 & 4301 & -4048 \\ 3611 & -1748 & 2231 \\ 1725 & -391 & 368 \end{pmatrix} \equiv_{26} \begin{pmatrix} 6 & 11 & 8 \\ 23 & 20 & 21 \\ 9 & 25 & 4 \end{pmatrix} = \boldsymbol{D}$$

The Hill Cipher

message = "CATANDHOUND"

$$\boldsymbol{M} = \begin{pmatrix} 2 & 0 & 7 & 13 \\ 0 & 13 & 14 & 3 \\ 19 & 3 & 20 & 1 \end{pmatrix}$$

$$\boldsymbol{C} = \boldsymbol{K} \cdot \boldsymbol{M} = \begin{pmatrix} 1 & 0 & 11 \\ 11 & 16 & 24 \\ 7 & 17 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 & 7 & 13 \\ 0 & 13 & 14 & 3 \\ 19 & 3 & 20 & 1 \end{pmatrix} \equiv_{26} \begin{pmatrix} 3 & 7 & 19 & 24 \\ 10 & 20 & 1 & 7 \\ 7 & 16 & 21 & 13 \end{pmatrix}$$

cipher = "DKHHUQTBVYHN"

$$\boldsymbol{R} = \boldsymbol{D} \cdot \boldsymbol{C} = \begin{pmatrix} 6 & 11 & 8 \\ 23 & 20 & 21 \\ 9 & 25 & 4 \end{pmatrix} \cdot \begin{pmatrix} 3 & 7 & 19 & 24 \\ 10 & 20 & 1 & 7 \\ 7 & 16 & 21 & 13 \end{pmatrix} \equiv_{26} \begin{pmatrix} 2 & 0 & 7 & 13 \\ 0 & 13 & 14 & 3 \\ 19 & 3 & 20 & 1 \end{pmatrix}$$

Problem

- Find the inverse key to

$$\begin{pmatrix} 0 & 15 \\ 1 & 5 \end{pmatrix}$$

Transposition Cipher

- Plaintext shuffled around according to permutation
- The encryption key π consists of permutation cycles
- The decryption key is the inverse permutation π^{-1}
- Encryption: $C = \pi(M)$
- Decryption: $M = \pi^{-1}(C)$

Example:

$$M = \text{"CATANDHOUND"}$$
$$\pi = (1, 6, 11, 9, 8)\,(4, 7, 5)$$
$$C = \pi(M) = \text{"OATNHCAUDND"}$$
$$\pi^{-1} = (1, 8, 9, 11, 6)\,(4, 5, 7))$$
$$M = \pi^{-1}(C) = \text{"CATANDHOUND"}$$

Explanation, C:

$$M_1 \to M_6 \to M_{11} \to M_9 \to M_8 \to M_1$$
$$M_4 \to M_7 \to M_6 \to M_4$$
$$M_2,\ M_3,\ M_{10} \text{ fixed}$$