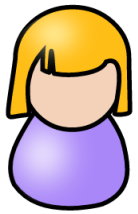
Ancient ciphers

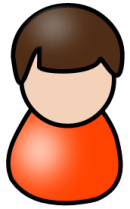
How to encrypt a message?





PLAY

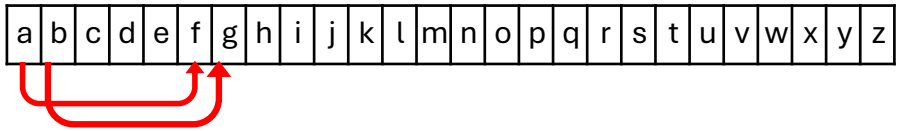




Idea 1: rotation (shift) cipher

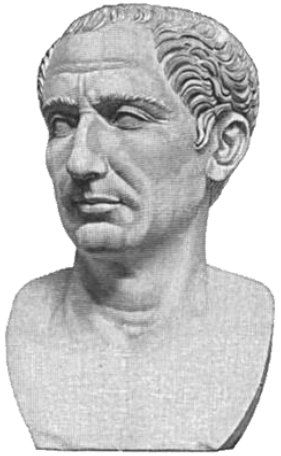
- **C = X + K mod 26**
- **C, X, K are number of corresponding letters.**

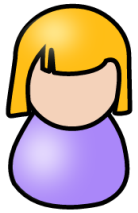
0	1	2	3	4	5	6	7	8	9	10	11	12
A	B	C	D	E	F	G	H	I	J	K	L	M
13	14	15	16	17	18	19	20	21	22	23	24	25
N	O	P	Q	R	S	T	U	V	W	X	Y	Z



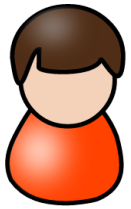


Number of Possible Key?





rotate by 5, K = 5

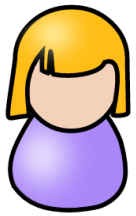


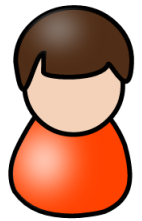
rotate by 5, K = 5

Idea 2: substitution cipher

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

hpd mevdshzjdo mt hpzo smknod und hm zqtmni ohkfdqho uemkh hpd tzdcf mt smigkhdn oszdqsd uqf dqazqddnzqa, uqf hm zqhnmfksd hpdi hm hpd xzfd nuqad mt dbszhzqa uggczsuhzmqo mt smigkhuhzmq uqf hdspqmcmal zq omszdhl. xd xzcc dbgcuzq hpuh smigkhdn oszdqsd zo qmh vkoh uemkh gnmanuiizqa, ekh uemkh smigkhuhzmquc hpzqrzqa uqf hpd ohkfl mt ucamnzhpio. zigmnhuqh hdnio uqf smqsdgho zq hpd tzdcf xzcc ed zqhnmfksdf, uo u gndjzdx hm xpuh xzcc ed cdunqdf zq mhpdn smknodo. xd xzcc dbgcuzq hpd imhzjuhzmq tmn junzmko uogdsho mt mkn iuvmn, zqsckfzqa hpd smnd smknod odwkdqsd zq mkn sknnzskcki, iuhp ndwkzndidqho, okggmnhzqa undu, sm-mgo, dhs.





Try all possible substitutions? 26! = 4×10^{26}

Attack: frequency analysis

hpd mevdshzjdo mt hpzo smknod und hm zqtmni ohkfdqho uemkh hpd tzdcf mt smigkhdn oszdqsd uqf dqazqddnzqa, uqf hm zqhnmfksd hpdi hm hpd xzfd nuqad mt dbszhzqa uggczsuhzmqo mt smigkhuhzmq uqf hdspqmcmal zq omszdhl. xd xzcc dbgcuzq hpuh smigkhdn oszdqsd zo qmh vkoh uemkh gnmanuiizqa, ekh uemkh smigkhuhzmquc hpzqrzqa uqf hpd ohkfl mt ucamnzhpio. zigmnhuqh hdnio uqf smqsdgho zq hpd tzdcf xzcc ed zqhnmfksdf, uo u gndjzdx hm xpuh xzcc ed cdunqdf zq mhpdn smknodo. xd xzcc dbgcuzq hpd imhzjuhzmq tmn junzmko uogdsho mt mkn iuvmn, zqsckfzqa hpd smnd smknod odwkdqsd zq mkn sknnzskcki, iuhp ndwkzndidqho, okggmnhzqa undu, sm-mgo, dhs.

Observation: substitution preserves frequency!!

Letter	Relative frequency in the English language
a	8.167%
b	1.492%
c	2.782%
d	4.253%
e	12.702%
f	2.228%
g	2.015%
h	6.094%
i	6.966%
j	0.153%
k	0.772%
l	4.025%
m	2.406%
n	6.749%
o	7.507%
p	1.929%
q	0.095%
r	5.987%
s	6.327%
t	9.056%
u	2.758%
v	0.978%
w	2.360%
x	0.150%
y	1.974%
z	0.074%

Letter	Relative frequency in the English language
e	12.702%
t	9.056%
a	8.167%
o	7.507%
i	6.966%
n	6.749%
s	6.327%
h	6.094%
r	5.987%
d	4.253%
l	4.025%
c	2.782%
u	2.758%
m	2.406%
w	2.360%
f	2.228%
g	2.015%
y	1.974%
p	1.929%
b	1.492%
v	0.978%
k	0.772%
j	0.153%
x	0.150%
q	0.095%
z	0.074%

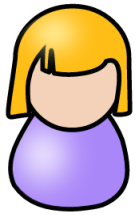


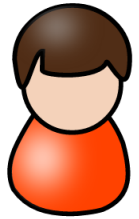
d	10.85%
h	10.08%
m	9.11%
z	8.53%
q	7.56%
u	6.20%
n	6.01%
s	5.43%
k	5.04%
o	5.04%
c	3.68%
g	3.10%
p	3.10%
i	3.10%
f	2.91%
a	2.13%
t	1.94%
x	1.74%
e	1.36%
j	0.78%
b	0.58%
l	0.58%
v	0.58%
r	0.39%
w	0.19%
y	0

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

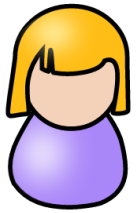
Attacking substitution cipher

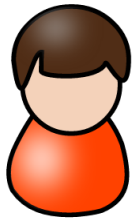
the objectipes og this course are to ingorm students about the gield og comwuter enfineerinf, and to introduce them to the yide ranfe og evcitinf awwlications og comwutation and technolofk in societk. ye t yill evwlain that comwuter enfineerinf is not just about bgcuzq hpuh wroframminf, but about comwutational thinqinf and the studk og alforithms. imwortant terms and concewts in the gield yill be introduced, as a wrepiey to yhat yill be learned in other courses. ye yill evwlain the motipation gor parious aswects og our major, xd xzcc includinf the core course sexuence in our curriculum, math vmn, rexuirements, suwwortinf area, co-ows, etc. mkn sknnzskcki, iuhp ndwkzndidqho, okggmnhzqa undu, sm-mgo, dhs.





The objectives of this course are to inform students about the field of Computer Engineering, and to introduce them to the wide range of exciting applications of computation and technology in society. We will explain that Computer Engineering is not just about programming, but about computational thinking and the study of algorithms. Important terms and concepts in the field will be introduced, as a preview to what will be learned in other courses. We will explain the motivation for various aspects of our major, including the core course sequence in our curriculum, math requirements, supporting area, co-ops, etc.





Vigenère (*vee-zhuh-ner*) cipher

Idea 3: rotation by different shifts

Plaintext	Plaintext value P	Key	Key value K	Ciphertext Value E	Ciphertext
C	2	K	10	12	M
R	17	E	4	21	V
Y	24	Y	24	22	W
P	15	K	10	25	Z
T	19	E	4	23	X
O	14	Y	24	12	M
G	6	K	10	16	Q
R	17	E	4	21	V
A	0	Y	24	24	Y
P	15	K	10	25	Z
H	7	E	4	11	L
Y	24	Y	24	22	W

Attacking the Vigenère cipher

- If the key length is known
  - Every 3rd letter still follows the letter frequency

Plaintext	Plaintext value P	Key	Key value K	Ciphertext Value E	Ciphertext
C	2	K	10	12	M
R	17	E	4	21	V
Y	24	Y	24	22	W
P	15	K	10	25	Z
T	19	E	4	23	X
O	14	Y	24	12	M
G	6	K	10	16	Q
R	17	E	4	21	V
A	0	Y	24	24	Y
P	15	K	10	25	Z
H	7	E	4	11	L
Y	24	Y	24	22	W

How to find the key length?

- Try key length of 1, 2, 3,…
    - If key length is correct: following letter frequency

Friedman test

Index of coincidence I : probability of two letters the **same** when drawing **randomly**

$I = \sum_{i=0}^{25} p_i^2$, p_i is the frequency of the $i\ th$ letter

English: $I = 0.065$

Uniform distribution $I = 0.038$ (why?)

I is invariant to rotation

Probability

- *Event*: a particular occurrence in some experiment
  - $\Pr[E]$: probability of event E

- Law of total probability: say E_1, …, E_n are a *partition* of all possibilities. Then for any A:

$$\Pr[A] = \sum_i \Pr[A \; and \; E_i] = \Sigma_i \Pr[A \mid E_i] \cdot \Pr[E_i]$$

Friedman test

Plaintext	Plaintext value P	Key	Key value K	Ciphertext Value E	Ciphertext
C	2	K	10	12	M
R	17	E	4	21	V
Y	24	Y	24	22	W
P	15	K	10	25	Z
T	19	E	4	23	X
O	14	Y	24	12	M
G	6	K	10	16	Q
R	17	E	4	21	V
A	0	Y	24	24	Y
P	15	K	10	25	Z
H	7	E	4	11	L
Y	24	Y	24	22	W

If the key length guess is correct:

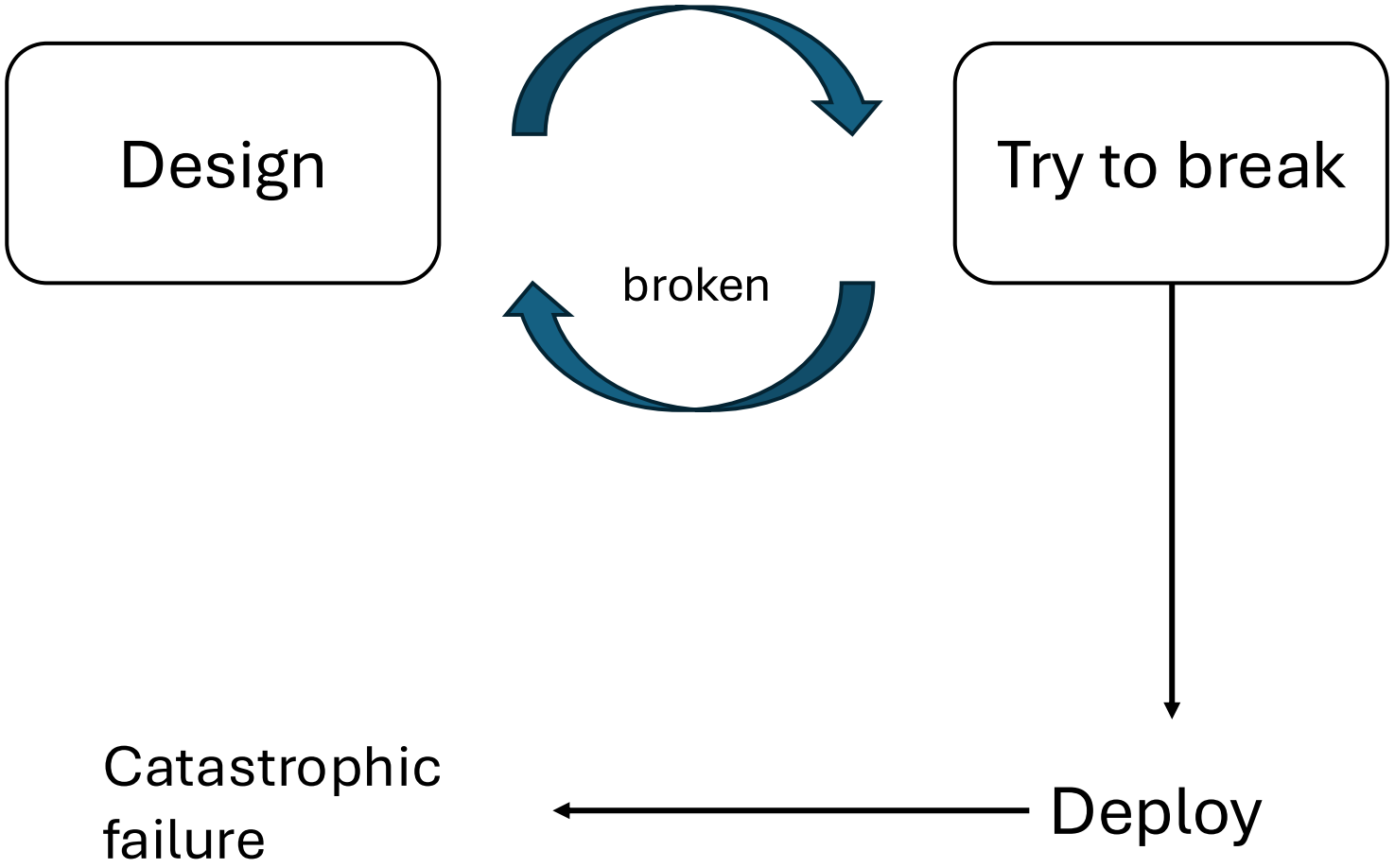
- l is the key length and n is the cipher length.
- The probability that both letters are selected from the same key group and are the same is $\frac{n-l}{l\,(n-1)} \times 0.065$.
- The probability that both letters are selected from different key groups and are the same letter $\frac{n(l-1)}{l\,(n-1)} \times 0.038$. **WHY?**
- $I = \frac{n-l}{l\,(n-1)} \times 0.065 + \frac{n(l-1)}{l\,(n-1)} \times 0.038 \approx 0.038 + \frac{1}{l} \times 0.027$

Some other old things

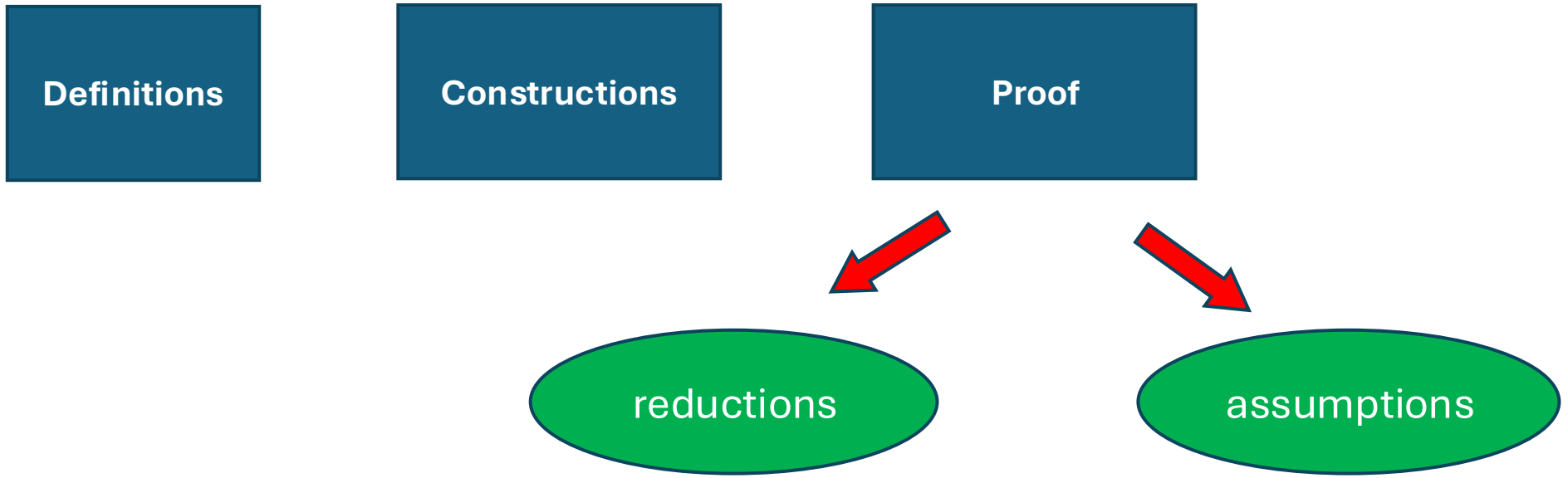


Enigma Machine

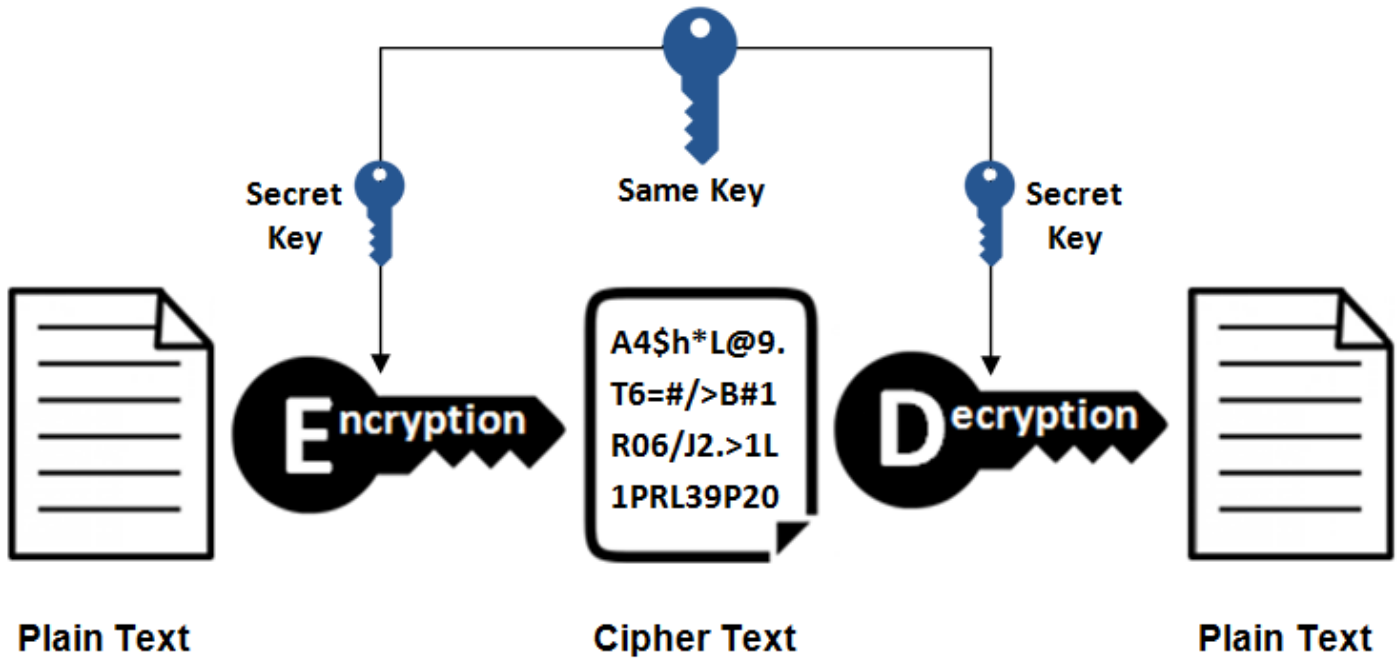
Classical cryptography



Modern cryptography



Symmetric Encryption



Definitions for encryption

Algorithms:

- ***k***: secret key, ***m***: message, ***c***: ciphertext
- $k = \mathbf{Gen}(\,)$
- $c = \mathbf{Enc}(k, m)$
- $m = \mathbf{Dec}(k, c)$

Notation

- $\mathcal{K}$ (key space) – set of all possible keys
- $\mathcal{C}$ (ciphertext space) – set of all possible ciphertexts
- $\mathcal{M}$ (message space) – set of all possible messages

Definitions for encryption

- Algorithms:
  - ***k***: secret key, ***m***: message, ***c***: ciphertext
  - $k = \mathrm{Gen}(\,)$
  - $c = \mathrm{Enc}(k, m)$
  - $m = \mathrm{Dec}(k, c)$

- **Correctness**: for every message $m \in \mathcal{M}$, any secret key $k \in \mathcal{K}$

$$m = \mathrm{Dec}(k, \mathrm{Enc}(k, m))$$

Probability

- *Event*: a particular occurrence in some experiment
  - $\Pr[E]$: probability of event E

- Conditional probability: probability that one event occurs, *given that* some other event occurred
  - $\Pr[A \mid B] = \Pr[A \; and \; B]/\Pr[B]$

- Two r.v.'s X, Y are *independent* if
  for all x, y: $\Pr[X = x \mid Y = y] = \Pr[X = x]$

Definitions for encryption

- Algorithms:
  - ***k***: secret key, ***m***: message, ***c***: ciphertext
  - $k = \mathrm{Gen}(\,)$
  - $c = \mathrm{Enc}(k, m)$
  - $m = \mathrm{Dec}(k, c)$

- **Perfect secrecy**: for any distribution over the message space $\mathcal{M}$, for every message $m \in \mathcal{M}$ and every ciphertext $c \in \mathcal{C}$,

$$\Pr[\, M = m \mid C = c \,] = \Pr[M = m]$$

Example

- $\mathcal{M}: \{0,1\}^n, \mathcal{K}: \{0,1\}^n, \mathcal{C}: \{0,1\}^n$

- for any distribution over the message space $\mathcal{M}$: English words encoded as ASCII code and then binary string

- for every message $m \in \mathcal{M}: m = 0000000$
- every ciphertext $c \in \mathcal{C}: c = 0100100$

$$\Pr[M = 0000000 | C = 0100100] = \Pr[M = 0000000]$$

Example: rotation cipher (k = 3)

- Rotation shifts every letter by the same *k*, so repeated letters stay repeated.
- The ciphertext **xx** has two identical letters; the message **ab** does not .
- no key can map **ab** to **xx**:

  If $C = xx, \Pr[M = ab | C = xx] = 0 \neq \Pr[M = ab]$

- So the rotation cipher is **not** perfectly secret — the ciphertext leaks whether the two letters are the same.

An equivalent definition

- Algorithms:
  - ***k***: secret key, ***m***: message, ***c***: ciphertext
  - $k = \mathrm{Gen}()$
  - $c = \mathrm{Enc}(k, m)$
  - $m = \mathrm{Dec}(k, c)$

- **Perfect secrecy**: for any distribution over the message space $\mathcal{M}$, for every message $m \in \mathcal{M}$ and every ciphertext $c \in \mathcal{C}$,

$$\Pr[C = c | M = m] = \Pr[C = c]$$

Proof

- Bayes's theorem: $\Pr[A \mid B] \cdot \Pr[B] = \Pr[B \mid A] \cdot \Pr[A]$
- $\Pr[\, M = m \mid C = c \,] \Pr[C = c] = \Pr[\, C = c \mid M = m \,] \Pr[M = m]$
- If $\Pr[\, M = m \mid C = c \,] = \Pr[M = m]$,
  then $\Pr[\, C = c \mid M = m] = \Pr[C = c]$

One-time pad

- rotation by different shifts
- Key length the same as message length

Rotation vs modular addition

Plaintext **r** with key **k = 24** gives ciphertext **p**

Rotation

Slide the alphabet 24 places forward:

r → s → t → … → o → p

Same as sliding back 2 places,
since 24 ≡ −2 (mod 26)

Modular addition

Index the letters a = 1 … z = 26:

r = 18, k = 24

18 + 24 = 42

42 mod 26 = 16 = p

Rotating by k and adding k mod 26 are two views of the same operation.

XOR

- $A \oplus B$

input		output
0	0	0
0	1	1
1	0	1
1	1	0

Equivalent to addition mod 2

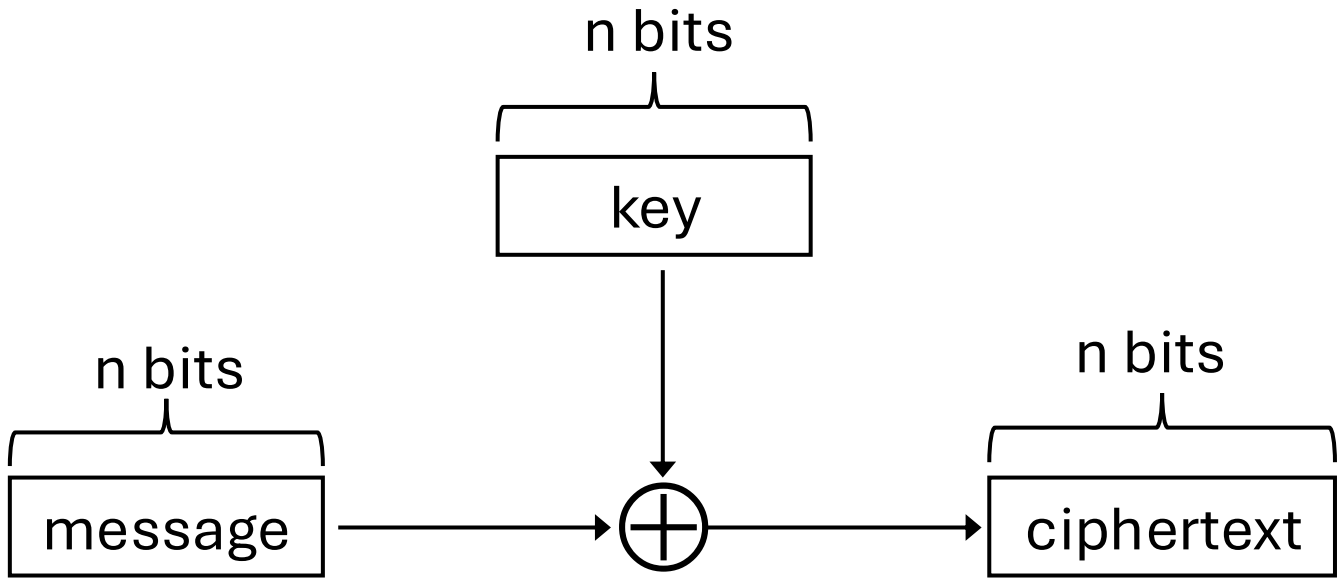
One-time pad

- Let $\mathcal{M} = \{0,1\}^n$
- $\mathrm{Gen}()$: choose a uniform key k $\{0,1\}^n$
- $\mathrm{Enc}_k(m)$: $c = k \oplus m$
- $\mathrm{Dec}_k(c)$: $m = k \oplus c$

- Correctness:

$$\mathrm{Dec}_k(\,\mathrm{Enc}_k(m)\,) = k \oplus (k \oplus m)$$
$$= (k \oplus k) \oplus m = m$$

One-time pad



Theorem

- Theorem 1: One-time pad encryption is perfectly secure.

Perfect secrecy: for any distribution over the message space $\mathcal{M}$, for every message $m \in \mathcal{M}$ and every ciphertext $c \in \mathcal{C}$,

$$\Pr[C = c | M = m] = \Pr[C = c]$$

Proof

- $\Pr[C = c | M = m] = \Pr[k \oplus m = c | M = m]$
$= \Pr[k = m \oplus c \mid M = m] = 2^{-n}$

Probability

- Law of total probability: say E_1, …, E_n are a *partition* of all possibilities. Then for any A:

$$\Pr[A] = \sum_i \Pr[A \text{ and } E_i] = \Sigma_i \Pr[A \mid E_i] \cdot \Pr[E_i]$$

Proof

- $\Pr[C = c] = \sum_{m \in \mathcal{M}} \Pr[C = c | M = m] \Pr[M = m]$

  $= \sum_{m \in \mathcal{M}} 2^{-n} \Pr[M = m]$

  $= 2^{-n} \sum_{m \in \mathcal{M}} \Pr[M = m]$

  $= 2^{-n}$

Another definition: perfect indistinguishability

Experiment:

1. Adversary A chooses any two messages $m_0, m_1 \in \mathcal{M}$
2. Generate a key k using $\mathrm{Gen}()$, pick a random bit $b \in \{0,1\}$, send attacker $c = \mathrm{Enc}(k, m_b)$
3. Attacker guesses *b'*
4. Experiment outputs 1 (success) if *b*=*b'*

Perfectly indistinguishable: for every adversary *A*

Pr[Experiment outputs 1] = ½

Perfect indistinguishability

- Lemma 2 (2.7 in the book): An encryption scheme is perfectly secret if and only if it is perfectly indistinguishable.

Lemma

For every two messages $m_0, m_1 \in \mathcal{M}$ and any ciphertext $c \in \mathcal{C}$,

$$\Pr[\mathrm{Enc}_k(m_0) = c] = \Pr[\mathrm{Enc}_k(m_1) = c]$$

where the probability is over the random choice of k.

An encryption scheme is perfectly secure if and only if the equation above holds

Using the same key twice?

- $m_1 \oplus m_2$ is information about m_1, m_2

- Is this significant?
    - No longer perfectly secret!
    - $m_1 \oplus m_2$ reveals where m_1, m_2 differ
    - Frequency analysis
    - Exploiting characteristics of ASCII…