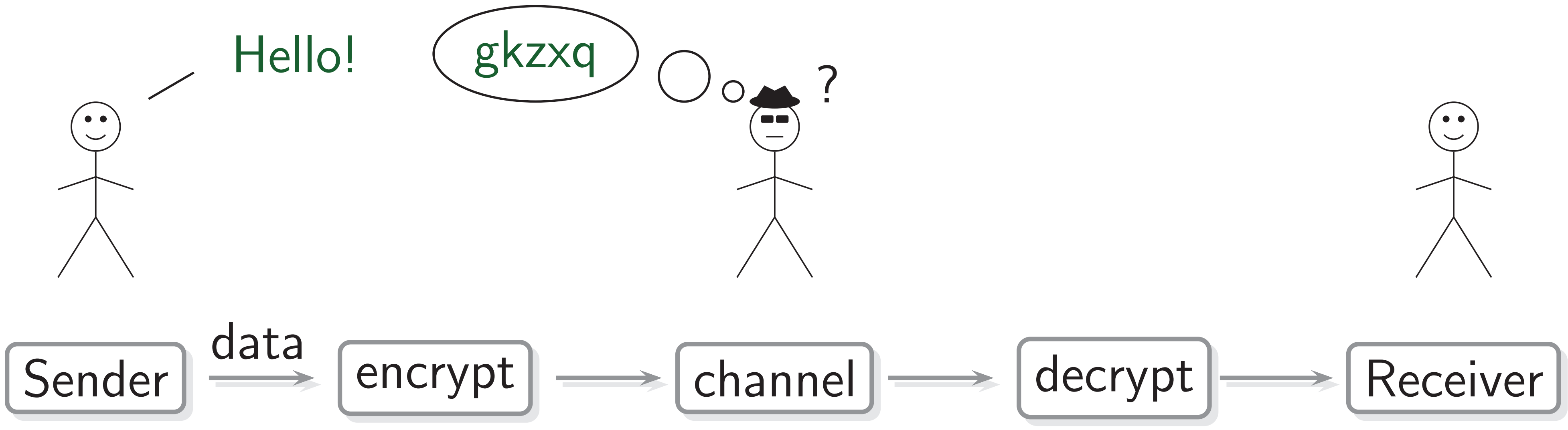
Chapter 7: Cryptography (Ciphers)

Lecture 28

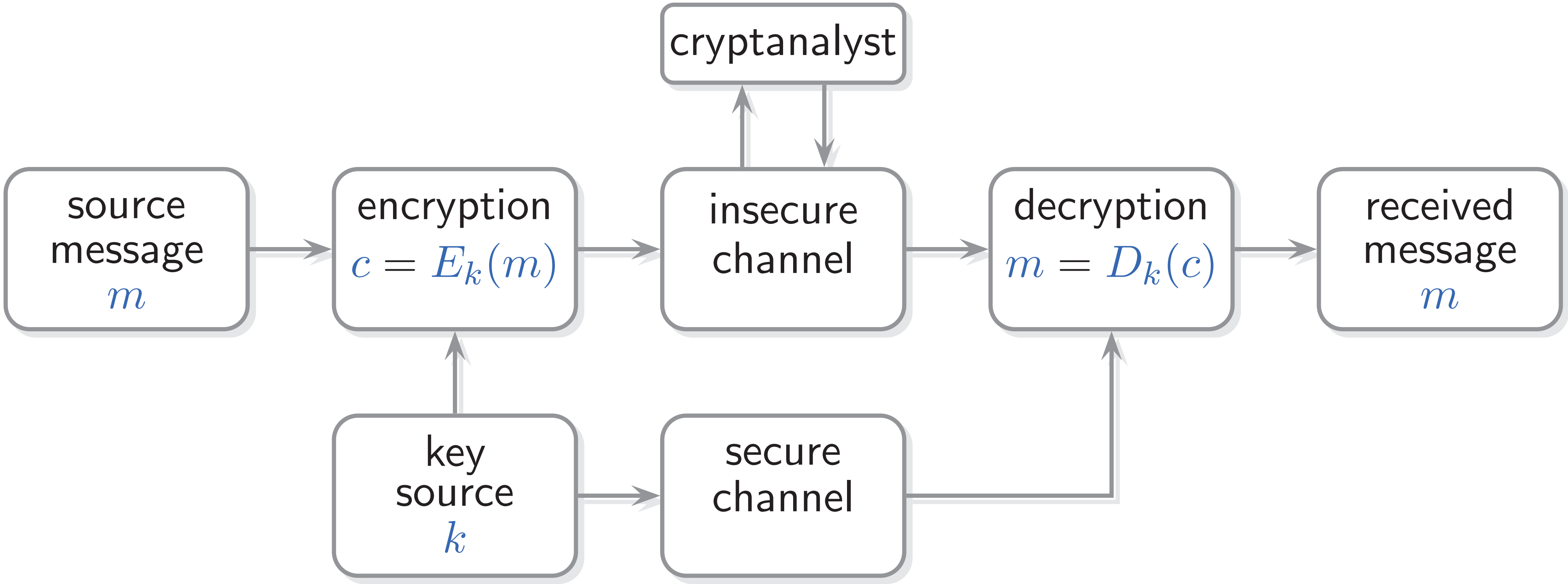


To keep a message (data) secret, we encrypt it.

- The message (Hello!) is called the plaintext.
- Encrypted (gkzxq), it is the ciphertext.

The cryptoanalysist (spy) can be passive (just listening) or active.

Cryptography and cryptoanalysis together form cryptology.
We will just look at cryptography.



$$m = D_k(c) = D_k(E_k(m))$$

Shannon's Maxim: "The enemy knows the system"

In other words, the cryptanalyst knows the cryptosystem's design $\{E_k\}$ and the possible messages M.

Classical cryptosystems

- Caeser ciphers
- Simple (monoalphabetic) substitution ciphers
- Transposition ciphers
- Combined systems
- Polyalphabetic substitution ciphers
- Non-periodic polyalphabetic substitution ciphers
- others...

Caeser ciphers

Cyclicly shift each letter k places forward.

$k = 1$

plaintext	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
ciphertext	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	A

Caeser ciphers

Cyclicly shift each letter k places forward.

$k = 2$

plaintext	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
ciphertext	C	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

Caeser ciphers

Cyclicly shift each letter k places forward.

$k = 3$

plaintext	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
ciphertext	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

Example

For $k = 3$, the plaintext message HELLO is encrypted as KHOOR.

Representing the letters by $\mathbb{Z}_{26}$, we have

$$E_k(i) = i + k \pmod{26}$$
$$D_k(j) = j - k \pmod{26}$$

Julius Caesar used $k = 3$.

Simple (monoalphabetic) substitution ciphers

Permute the letters A, B, . . . , Z (or $\mathbb{Z}_{26}$) by some permutation π.

plaintext	i	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
ciphertext	$\pi(i)$	V	E	D	F	G	K	I	Z	X	L	M	C	Y	A	R	O	B	Q	J	T	S	H	P	U	W	N

Sometimes, a keyword is used to make the code easier to remember.
For instance, we might use the keyword “CODEBREAKING”,
starting at K, and padding Caesar-style with the rest of the letters.

Simple (monoalphabetic) substitution ciphers

Permute the letters A, B, . . . , Z (or $\mathbb{Z}_{26}$) by some permutation π.

plaintext	i	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
ciphertext	$\pi(i)$	P	Q	S	T	U	V	W	X	Y	Z	C	O	D	E	B	R	A	K	I	N	G	F	H	J	L	M

Sometimes, a keyword is used to make the code easier to remember. For instance, we might use the keyword "CODEBREAKING", starting at K, and padding Caesar-style with the rest of the letters.

Representing the letters by $\mathbb{Z}_{26}$, we have

$$E_\pi(i) = \pi(i)$$
$$D_\pi(j) = \pi^{-1}(j)$$

There are $26! \approx 4 \times 10^{26}$ possible keys π.
- but there are many letter-dependencies and non-uniform letter-frequencies.
This type of cipher is therefore easy to break.

Transposition ciphers

- Partition the message into blocks of r letters
- Then apply a fixed permutation π to the letter order of each block.

For instance, let $r = 5$ and

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 4 & 5 & 2 \end{pmatrix}$$

Then

plaintext	T H I S I	S A N E X	A M P L E
ciphertext	H I T I S	A X S N E	M E A P L

Combined systems

These combine transposition and substitution.

This makes them harder to break - but letter frequencies can still be used.

Check lecture slides for animation

Polyalphabetic substitution ciphers

The simplest of these is the Vigenère cipher (1586).

- r different Caesar ciphers are applied periodically, specified by a key.

Example

Let the key be CODE.

Then

key	C	O	D	E	C	O	D	E	C	O	D	E	C	O	D
plaintext	T	H	I	S	I	S	A	N	E	X	A	M	P	L	E
ciphertext	V	V	L	W	K	G	D	R	G	L	D	Q	R	Z	H

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

Polyalphabetic substitution ciphers

The simplest of these is the Vigenère cipher (1586).

- r different Caesar ciphers are applied periodically, specified by a key.

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

Vigenère table

Polyalphabetic substitution ciphers

The simplest of these is the Vigenère cipher (1586).

- r different Caesar ciphers are applied periodically, specified by a key.
- Vigenère ciphers were often used in the 17th–19th centuries.
- They are usually broken by brute force methods, still today.

These sort of ciphers apply substitution alphabets periodically.

- If the length of the key or a period is known,
  then it is easy to break the cipher.

Kasiski's method (1863)

This is a systematic method for finding a key- or period length r.

Let A, B, . . . , Z be represented by $\mathbb{Z}_{26}$
and let f_i be the frequency of letter i in some text m.

The probability of coincidence $P_C(m)$ is the probability that
two randomly chosen letters from a text m are identical.

Example
If m consists of random letters, then $P_C(m) = \dfrac{1}{26} \approx 0.0385$.

Example
If m is an English text, then $p_0 = P(\mathsf{A}) \approx 0.0804, \dots, P(\mathsf{Z}) \approx 0.0009$
and

$$P_C(m) = \sum_{i=0}^{25} p_i^2 \approx 0.0658$$

Let A, B, . . . , Z be represented by $\mathbb{Z}_{26}$ and let f_i be the frequency of letter i in some text m.

The probability of coincidence $P_C(m)$ is the probability that two randomly chosen letters from a text m are identical.

Theorem

For a message m of length n,

$$P_C(m) \approx I_c(m) = \frac{\sum \binom{f_i}{2}}{\binom{n}{2}} = \frac{\left(\sum f_i^2\right) - n}{n^2 - n}$$

where $I_c(m)$ is the index of coincidence.

Proof

Out of the $\binom{n}{2}$ letter pairs, there are $\binom{f_i}{2}$ pairs of letter i.
Also, $\sum \binom{f_i}{2} = \frac{1}{2}\left(\sum f_i^2 - \sum f_i\right) = \frac{1}{2}\left(\sum f_i^2 - n\right)$. $\square$

Let A, B, . . . , Z be represented by $\mathbb{Z}_{26}$
and let f_i be the frequency of letter i in some text m.

The probability of coincidence $P_C(m)$ is the probability that two randomly chosen letters from a text m are identical.

Theorem

For a message m of length n,

$$P_C(m) \approx I_c(m) = \frac{\sum \binom{f_i}{2}}{\binom{n}{2}} = \frac{\left(\sum f_i^2\right) - n}{n^2 - n}$$

where $I_c(m)$ is the index of coincidence.

- $I_c(m)$ does not change if letters or letter positions are permuted.

Let A, B, . . . , Z be represented by $\mathbb{Z}_{26}$
and let f_i be the frequency of letter i in some text m.

The probability of coincidence $P_C(m)$ is the probability that two randomly chosen letters from a text m are identical.

Theorem

For a message m of length n,

$$P_C(m) \approx I_c(m) = \frac{\sum \binom{f_i}{2}}{\binom{n}{2}} = \frac{\left(\sum f_i^2\right) - n}{n^2 - n}$$

where $I_c(m)$ is the index of coincidence.

Example

For $m =$ BAADC, $n = 5$ and $f_0 = 2$, $f_1 = f_2 = f_3 = 1$, so

$$P_C(m) \approx I_c(m) = \frac{\sum \binom{f_i}{2}}{\binom{n}{2}} = \frac{\left(\sum f_i^2\right) - n}{n^2 - n} = \frac{2^2 + 1^2 + 1^2 + 1^2 - 5}{5^2 - 5} = 0.1$$

Assume that a periodic substitution cipher of key length r has been used. Write the n-letter message in r rows of n/r letters (assuming that $r \mid n$), by writing the 1st r letters in the 1st column, the next r letters in the second column, and so on.

- Letters in the same row have had the same substitutions applied
- Letters in the different rows have had different substitutions applied.

There are $r\binom{\frac{n}{r}}{2}$ ways to choose a pair of letters from the same row and the probability of coincidence is approximately 0.0658. (English)

There are $\frac{1}{2}n(n-\frac{n}{r})$ ways to choose a pair of letters from distinct rows and the probability of coincidence is approximately 0.0385. (random)

The number of coincident pairs is thus approximately

$$r\binom{\frac{n}{r}}{2} \times 0.0658 + \frac{1}{2}n(n-\frac{n}{r}) \times 0.0385$$

By definition, it is also $\frac{1}{2}n(n-1)I_c$.

Assume that a periodic substitution cipher of key length r has been used. Write the n-letter message in r rows of n/r letters (assuming that $r \mid n$), by writing the 1st r letters in the 1st column, the next r letters in the second column, and so on.

The number of coincident pairs is thus approximately

$$r\binom{\frac{n}{r}}{2} \times 0.0658 + \frac{1}{2}n(n - \frac{n}{r}) \times 0.0385$$

By definition, it is also $\frac{1}{2}n(n-1)I_c$.
Solving for r, we find that

$$r \approx \frac{0.0273n}{(n-1)I_c - 0.0385n + 0.0658}$$

and that

$$I_c \approx \frac{1}{r}(0.0273) + 0.0385 \quad \text{for} \quad n \to \infty$$

Assume that a periodic substitution cipher of key length r has been used. Write the n-letter message in r rows of n/r letters (assuming that $r \mid n$), by writing the 1st r letters in the 1st column, the next r letters in the second column, and so on.

The number of coincident pairs is thus approximately

$$r\binom{\frac{n}{r}}{2} \times 0.0658 + \frac{1}{2}n(n - \frac{n}{r}) \times 0.0385$$

By definition, it is also $\frac{1}{2}n(n-1)I_c$.
Solving for r, we find that

$$r \approx \frac{0.0273n}{(n-1)I_c - 0.0385n + 0.0658}$$

r	1	2	3	4	5	10	∞
I_c	.066	.052	.048	.045	.044	.041	.0385

Assume that a periodic substitution cipher of key length r has been used. Write the n-letter message in r rows of n/r letters (assuming that $r \mid n$), by writing the 1st r letters in the 1st column, the next r letters in the second column, and so on.

The number of coincident pairs is thus approximately

$$r\binom{\frac{n}{r}}{2} \times 0.0658 + \frac{1}{2}n(n - \frac{n}{r}) \times 0.0385$$

By definition, it is also $\frac{1}{2}n(n-1)I_c$.
Solving for r, we find that

$$r \approx \frac{0.0273n}{(n-1)I_c - 0.0385n + 0.0658}$$

To check that we have found the correct value of r, each row should have coincidence index I_c roughly equal to 0.0658.

- This method only works for very long texts.

Non-periodic polyalphabetic substitution ciphers

These ciphers eliminate (or greatly reduce) periodicity.

Plaintext feedback

key	C	O	D	E	T	H	I	S	I	S	A	N	E	X	A
plaintext	T	H	I	S	I	S	A	N	E	X	A	M	P	L	E
ciphertext	V	V	L	W	B	Z	I	F	M	P	A	Z	T	I	E

Ciphertext feedback

key	C	O	D	E	V	V	L	W	D	N	L	J	H	K	L
plaintext	T	H	I	S	I	S	A	N	E	X	A	M	P	L	E
ciphertext	V	V	L	W	D	N	L	J	H	K	L	V	W	V	P

Text from an external source

key	I	T	W	A	S	T	H	E	B	E	S	T	O	F	T
plaintext	T	H	I	S	I	S	A	N	E	X	A	M	P	L	E
ciphertext	B	A	E	S	A	L	H	R	F	B	S	F	D	Q	X

Non-periodic polyalphabetic substitution ciphers

These ciphers eliminate (or greatly reduce) periodicity.

Rotation ciphers

Vernam ciphers or one-time pad ciphers