Hill cipher.

Hill cipher was developed by its inventor Lester Hill in 1929. Hill cipher is known to be the first polygraphic cipher. The method is based on linear matrix transformation of a message space. Given a plaintext message $p = (p_1, p_2, \dots)$ where p_i is a letter in some alphabet and invertible $m \times m$ matrix H, Hill cipher represents p_i by numeric value $x_i \in Z_n (Z_n = \{0,1, \dots, n-1\})$ and encrypts plaintext as $y = H \cdot x \pmod{n}$, where x and y are plaintext and ciphertext column vectors. Similarly, y is decrypted as $x = H^{-1} \cdot y \pmod{n}$, where H^{-1} is the inverse of H. That is, $H \cdot H^{-1} = H^{-1} \cdot H = I$ holds, where I is the identity matrix.

The following exemplifies Hill cipher for $n = 26$,

$$H = \begin{pmatrix} 6 & 24 & 1 \\ 13 & 16 & 10 \\ 20 & 17 & 15 \end{pmatrix}, H^{-1} = \begin{pmatrix} 8 & 5 & 10 \\ 21 & 8 & 21 \\ 21 & 12 & 8 \end{pmatrix}.$$

The plaintext $x = (ACT) = (0\ 2\ 19)$ is encrypted as $y = H \cdot x \pmod{26} = (15\ 14\ 7) = (POH)$. Likewise, the ciphertext $y = (15\ 14\ 7)$ is decrypted as $x = H^{-1} \cdot y \pmod{26} = (0\ 2\ 19) = (ACT)$.

Hill cipher is vulnerable to cryptanalysis. The cryptanalyst usually sits in tight loop and tries to possess the plaintext of some messages and the corresponding ciphertext of those messages to deduce the key. If cryptanalyst succeeds with gaining access to the flow of messages then he can easily decrypt any new messages encrypted with the same key. The system can be obviously broken, knowing only $m$distinct plaintext and ciphertext pairs (x, y) and by computing $H = Y \cdot X^{-1} (\text{mod } n)$, where X and Y are the matrices composed of m columns of x and y, respectively. Whenever X is invertible the opponent can obviously compute the unknown key as $H = Y \cdot X^{-1} (\text{mod } n)$ and consequently break the cipher. If the X is not invertible then cryptanalyst keeps on collecting m plaintext and ciphertext pairs until the resulting matrix is invertible. When m is unknown, cryptanalyst might try the procedure for $m = 2,3,4$ until the key is found.

The affine Hill cipher was proposed to overcome this drawback. The affine Hill cipher is a secure variant of Hill cipher in which the concept is extended by mixing it with an affine transformation. Similar to the Hill cipher the affine Hill cipher is polygraphic cipher, encrypting/decrypting m letters at a time. Given key matrix H and vector V, in affine Hill cipher the encryption expression is represented by $y = H \cdot x + V \pmod{n}$. Similarly, the decryption is performed by $x = H^{-1} \cdot (y - V) (\text{mod } n)$. The following example illustrates the way encryption and decryption is performed in affine Hill cipher. The following example exemplifies affine Hill cipher. Let $n = 26$,

$$H = \begin{pmatrix} 6 & 24 & 1 \\ 13 & 16 & 10 \\ 20 & 17 & 15 \end{pmatrix}, H^{-1} = \begin{pmatrix} 8 & 5 & 10 \\ 21 & 8 & 21 \\ 21 & 12 & 8 \end{pmatrix} \text{ and } V = \begin{pmatrix} 5 \\ 0 \\ 7 \end{pmatrix}.$$

The encryption $x = (ACT) = (0\ 2\ 19)$is possessed by

$$y = H \cdot x + V \pmod{26} = \begin{pmatrix} 6 & 24 & 1 \\ 13 & 16 & 10 \\ 20 & 17 & 15 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \\ 19 \end{pmatrix} + \begin{pmatrix} 5 \\ 0 \\ 7 \end{pmatrix} \pmod{26} = \begin{pmatrix} 20 \\ 14 \\ 14 \end{pmatrix}.$$

Likewise, the decryption is performed as follows:

$$x = H^{-1} \cdot (x - V) \pmod{26} = \begin{pmatrix} 8 & 5 & 10 \\ 21 & 8 & 21 \\ 21 & 12 & 8 \end{pmatrix} \cdot \begin{pmatrix} 15 \\ 14 \\ 7 \end{pmatrix} \pmod{26} = \begin{pmatrix} 0 \\ 2 \\ 19 \end{pmatrix}.$$

Suppose Alice chooses affine Hill cipher to send confidential message to Bob. Firstly, she selects a pair (H, V) to encrypt the plaintext. Then she sends ciphertext as well as (H, V) to Bob. When Bob receives ciphertext and a pair (H, V) he creates H^{-1}, and then decrypts the ciphertext with (H^{-1}, V).

In 2000, Saeednia proposed an interesting modification of Hill cipher. The main idea behind of his algorithm is to modify the key matrix each time Hill cipher is implemented. Encrypting a message by a one-time used matrix would make the algorithm more secure compared to the original Hill cipher and affine Hill cipher.

Assume Alice decides to send confidential message of size $m \times s$ to Bob, and she chooses Saeednia's algorithm to encrypt the message. Then she firstly selects random permutation π of size m and creates $m \times m$ permutation matrix P_π by permuting the rows of same size identity matrix. Such a matrix is always row equivalent to an identity matrix. Then she creates its inverse P_π^{-1} by permuting the columns of the identity matrix. Likewise, P_π^{-1} is column equivalent to the row-permuted matrix. After that, she creates one-time used matrix H_π from the key matrix H as $H_\pi = P_\pi^{-1} \cdot H \cdot P_\pi$. She further encrypts x as $y = H_\pi \cdot x \pmod{n}$ and sends a pair (y, π') to Bob where $\pi' = H \cdot \pi \pmod{n}$. Upon receipt of the message, Bob computes permutation π from π' and H^{-1} as follows $\pi = H^{-1} \cdot \pi' \pmod{n}$. Bob next calculates $(H^{-1})_\pi = P_\pi^{-1} \cdot H^{-1} \cdot P_\pi$, and decrypts the ciphertext as $x = (H_\pi)^{-1} \cdot y \pmod{n}$ keeping in mind that $(H_\pi)^{-1} = (H^{-1})_\pi$.

It should be noticed that the permutation of any pair of rows (or columns) of matrix H yields a matrix whose inverse is obtained by the permutationof the same columns (or rows) of H^{-1}. This is the reason why Bob does not need to use transposition algorithm to find $(H_\pi)^{-1}$; it can be easily obtained from equality $(H_\pi)^{-1} = (H^{-1})_\pi$. This observation essentially decreases computational cost of Saeednia's algorithm.

Below we exemplify encryption and decryption with Saeednia's algorithm for

$$\pi = \binom{123}{213}, H = \begin{pmatrix} 6 & 24 & 1 \\ 13 & 16 & 10 \\ 20 & 17 & 15 \end{pmatrix} \text{ and } x = \begin{pmatrix} 0 \\ 2 \\ 19 \end{pmatrix}.$$

By using permutation matrix $P_\pi = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and its inverse $P_\pi^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ we obtain

$$H_\pi = P_\pi^{-1} \cdot H \cdot P_\pi = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 6 & 24 & 1 \\ 13 & 16 & 10 \\ 20 & 17 & 15 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 16 & 13 & 10 \\ 24 & 6 & 1 \\ 17 & 20 & 15 \end{pmatrix}.$$

After that we encrypt the plaintext as follows

$$y = H_\pi \cdot x \ (\text{mod } n\ ) = \begin{pmatrix} 16 & 13 & 10 \\ 24 & 6 & 1 \\ 17 & 20 & 15 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 2 \\ 19 \end{pmatrix} (\text{mod } 26) = \begin{pmatrix} 8 \\ 5 \\ 13 \end{pmatrix}.$$

Decryption is carried out as follows

$$(H_\pi)^{-1} = (H^{-1})_\pi = P_\pi^{-1} \cdot H^{-1} \cdot P_\pi = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 8 & 5 & 10 \\ 21 & 8 & 21 \\ 21 & 12 & 8 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 21 & 8 & 21 \\ 8 & 5 & 10 \\ 21 & 12 & 8 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 21 & 21 \\ 5 & 8 & 10 \\ 12 & 21 & 8 \end{pmatrix}.$$

$$x = (H_\pi)^{-1} \cdot y \ (\text{mod } 26) = \begin{pmatrix} 8 & 21 & 21 \\ 5 & 8 & 10 \\ 12 & 21 & 8 \end{pmatrix} \cdot \begin{pmatrix} 8 \\ 5 \\ 13 \end{pmatrix} (\text{mod } 26) = \begin{pmatrix} 0 \\ 2 \\ 19 \end{pmatrix}.$$

It was reported in that Saeednia's algorithm has the same problem as the original Hill cipher. By collecting$m$ pairs of (π, π') and simultaneously solving m equations $\pi = H \cdot \pi'$, a cryptanalyst can obtain the key matrix H. Further, he can use P_π and P_π^{-1} to calculate H_π. In the same paper it was noticed that Saeednia's algorithm is time consuming since it is based on frequent use of matrix operations. Matrix operations require a lot of time to compute when the matrix size is large enough.

The main steps of Hill cipher are indicated below:

Encryption	Decryption
1. Select invertible matrix $\boldsymbol{H}$.	1. Calculate $\boldsymbol{H^{-1}}$.
2. Calculate $\boldsymbol{y = H \cdot x}$ (mod $\boldsymbol{n}$).	2. Calculate $\boldsymbol{x = H^{-1} \cdot y}$ (mod $\boldsymbol{n}$).
3. Alice sends ($\boldsymbol{y}$, $\boldsymbol{H}$) to Bob.	

The affine Hill cipher is represented by the following steps:

Encryption	Decryption
1. Select invertible matrix $\boldsymbol{H}$.	1. Calculate $\boldsymbol{H^{-1}}$.
2. Select vector $\boldsymbol{V}$.	2. Calculate $\boldsymbol{x = H^{-1}(y - V)}$.
3. Calculate $\boldsymbol{y = H \cdot x + V}$ (mod $\boldsymbol{n}$).	
Alice sends $\boldsymbol{H}, \boldsymbol{V}, \boldsymbol{y}$ to Bob.	

The main steps of Saeednia'smethod are as follows:

Encryption

1. Select random permutation $\boldsymbol{\pi}$.

2. Select matrix $\boldsymbol{H}$.

3. Calculate permutation matrix $\boldsymbol{P_\pi}$ by permuting the rows of identity matrix $\boldsymbol{I}$.

4. Calculate permutation matrix $\boldsymbol{P_\pi^{-1}}$ by permuting the columns of identity matrix $\boldsymbol{I}$.

5. Calculate $\boldsymbol{H_\pi = P_\pi^{-1} \cdot H \cdot P_\pi}$.

6. Calculate $\boldsymbol{y = H_\pi \cdot x}$ (mod $\boldsymbol{n}$)

7. Calculate $\boldsymbol{\pi' = H \cdot \pi}$ (mod $\boldsymbol{n}$)

Alice sends $\boldsymbol{\pi'}, \boldsymbol{H}, \boldsymbol{y}$ to Bob.

Decryption

1. Calculate $\boldsymbol{H^{-1}}$.

2. Calculate $\boldsymbol{\pi = H^{-1} \cdot \pi'}$ (mod $\boldsymbol{n}$).

3. Calculate permutation matrix $\boldsymbol{P_\pi}$ by permuting the rows of identity matrix $\boldsymbol{I}$.

4. Calculate permutation matrix $\boldsymbol{P_\pi^{-1}}$ by permuting the columns of identity matrix $\boldsymbol{I}$.

5. Calculate $\boldsymbol{(H^{-1})_\pi = P_\pi^{-1} \cdot H \cdot P_\pi}$.

6. Calculate $\boldsymbol{x = (H_\pi)^{-1} \cdot y}$ (mod $\boldsymbol{n}$)

Polyalphabetic ciphers. All polyalphabetic techniques have the following features in common:

1. A set of related monoalphabetic substitution rules are used.
2. The key determines which particular rule is chosen for a given transformation.

The best-known and one of the simplest polyalphabetic algorithms is referred to as the Vigenere cipher. In this scheme, the set of related monoalphabetic substitution rules consists of the 26 Caesar ciphers, with shifts of 0 to 25. Each cipher is denoted by a key letter, which is ciphertext letter that substitutes for the plaintext letter *a*. Thus, a Caesar cipher with a shift of 3 is denoted by the key value *d*.

The process of encryption is easy: Given a key letter *x* and the plaintext letter *y*, we pick up a letter that is at the intersection of the row and column pointed out by *x* and *y*, respectively. In this case the cipherletter is V.

Obviously, to encrypt the plaintext a key is needed that is as long as the message. Usually, the key is a

repeating keyword. For example, if the keyword is DECEPTIVE, the message “we are discovered save yourself” is encrypted as follows:

Key: deceptivedeceptivedeceptive

Plaintext: wearediscoveredsaveyourself

Ciphertext: ZICVTWQNGRZGVTWAVZHCQYGLMGL

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

Decyption is equally simple. The key letter again identifies the row. The position of the ciphertext letter in that row determines the column, and the plaintext letter is at the top of that column.

A cryptanalyst catches the ciphertext and tries to determine whether the ciphertext was created using a monoalphabetic substitution or a Vigenere type cipher. A simple test can be performed to make sure on this matter. If the statistical properties of the ciphertext are similar to those of a language used to

encrypt the plaintext, e.g., English, then plaintext was encrypted using monoalphabetic substitution. Otherwise plaintext was encrypted using a polyalphabetic cipher.

If, on the other hand, polyalphabetic substitution like Vigenere cipher is suspected, then progress depends on determining the length of the keyword. Let us concentrate on how the keyword length can be determined. The important insight that leads to a solution is the following: if two identical sequences of plaintext letters occur at a distance that is an integer multiple of the keyword length, they will generate identical ciphertext sequences. In the foregoing example, two instances of the sequence RED are separated by nine character positions. Consequently, in both cases, R is encrypted using key letter E, E is encrypted using key letter P, and d is encrypted using key letter T. Thus, in both cases the ciphertext would detect the repeated sequence VTW.

An analyst looking at only the ciphertext would detect the repeated sequences VTW at a displacement of 9 and make the assumption that the key is either 3 or 9 letter in length. If the message is long enough then there will be such repeated fragments. By calculating the factors in the displacements of the various sequences, the analyst should be able to make a good guess of the keyword length. Solution of the cipher now depends on an important insight. If the keyword length is *N*, then the cipher, in effect, consists of *N* monoalphabetic substitution ciphers. For example, if the keyword is DECEPTIVE then the letters in positions 1, 10, 19, and so on are all encrypted with the same monoalphabetic cipher. Thus we can use the known frequency characteristics of the plaintext language to attack each of the monoalphabetic ciphers separately.

Another example of Vinegere cipher is illustrated below

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

Plaintext: *this is a message*

Keyword: *keykeykeykeykeyke*

Ciphertext: CLFBDFBDYJQBBWYQI

In this example, *t* , the first letter of the plaintext is intersected with the alphabet pointed by the letter *k* . The result is the ciphertext letter C. Such a process, though slow to do by hand, can be done very quickly on a computer. The complexity of the cipher can be raised by increasing the number of unique letters in the keyword, and hence increasing the number of alphabets used. As a result, it is important for cryptanalysits to learn how many alphabets are being used.

Questions and problems

1. Use Hill cipher with the key matrix $K^{-1} = \begin{bmatrix} 3 & 17 \\ 8 & 25 \end{bmatrix}$ to decrypt the ciphertext CIKKGEUWEROY.
2. Assume the Vigenere cipher with the key EXAM is used to encrypt the text “The winter is here”. What will be the ciphertext?
3. Use the Vigenere ciher with the keyword ABCD to decrypt the ciphertext **CSASTP**KVSIQUTGQU**CSASTP**IUAQJB.