Software Security

LECTURE 2 – CRYPTO BASICS

The Basic Terminology

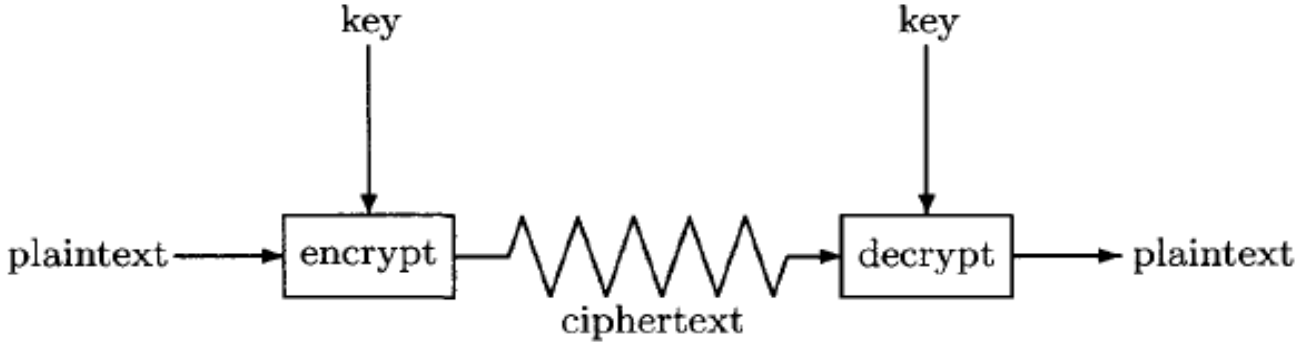
- Cryptology – the art and science of making and breaking “secret codes”.
- Cryptography – the making of “secret codes”.
- Cryptanalysis – the breaking of “secret codes”.
- Crypto – a synonym for any or all of the above (and more), when the precise meaning should be clear from context.

More Terminology

- A ***cipher*** or ***cryptosystem*** is used to ***encrypt*** data.
- The original unencrypted data is known as ***plaintext***, and the result of encryption is ***ciphertext***.
- We ***decrypt*** the ciphertext to recover the original plaintext.
- A ***key*** is used to configure a cryptosystem for encryption and decryption.

Keys

In a **symmetric** cipher, the same key is used to encrypt and to decrypt, as illustrated by the black box cryptosystem in the following figure:



Crypto as a Black Box

In symmetric key crypto, the key is known as a ***symmetric key***.

Keys

There is also a concept of ***public key*** cryptography where the encryption and decryption keys are different. Since different keys are used, it's possible to make the encryption key public. (Public key crypto is also known as asymmetric crypto, in reference to the fact that the encryption and decryption keys are different.)

In public key crypto, the encryption key is known as the ***public key***, whereas the decryption key, which must remain secret is the ***private key***.

Kerckhoff's Principle

- A cipher "must not be required to be secret, and it must be able to fall into the hands of the enemy without inconvenience", that is, the design of the cipher is not secret.

- What is the point of Kerckhoff's Principle? After all, it must certainly be more difficult for Trudy to attack a cryptosystem if she doesn't know how the cipher works. So, why would we want to make Trudy's life easier?

Problems of “Secrecy”

There are (at least) a couple of problems with relying on a secret design for your security:

- The details of “secret” cryptosystem seldom, if ever, remain secret for long. Reverse engineering can be used to recover algorithms from software, and even algorithms embedded in a damage-resistant hardware are sometimes subject to reverse engineering attacks and exposure.
- Cryptographers will not consider a crypto-algorithm worthy of use until it has withstood extensive public analysis by many cryptographers over an extended period of time.

Classic Crypto

- Simple Substitution Cipher
- Double Transposition Cipher
- One-Time Pad
- ADFGVX Cipher
- The Vigenère cipher
- Playfair cipher

Simple Substitution Cipher

In the simplest case, the message is encrypted by substituting the letter of the alphabet ***n*** places ahead of the current letter. For example, with **n = 3**, the substitution – which acts as a key – is

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

where (by convention) the plaintext is lowercase, and the ciphertext is uppercase. In this example, the key could be given as “3” since the amount of the shifts is, in effect, the key.

Simple Substitution Cipher

Using the key 3, we can encrypt the plaintext message

fourscoreandsevenyearsago

By looking up each plaintext letter in the plaintext row and then substituting the corresponding letter in the ciphertext row, or by simply replacing each letter by the letter that is three positions ahead of it in the alphabet.

For the particular plaintext, the resulting ciphertext is

IRXUVFRUHDAGVHYHABHDUVDIR.

Simple Substitution Cipher

To decrypt this simple substitution, we look up the ciphertext letter in the ciphertext row and replace it with the corresponding letter in the plaintext row, or we can shift each ciphertext letter backward by three.

The simple substitution with a shift of three is known as the ***Caesar's cipher***.

Simple Substitution Cipher

If we limit the simple substitution to shifts of the alphabet, then the possible keys are $n \in \{0, 1, 2, \dots, 25\}$. Suppose Trudy intercept the ciphertext message

CSYEVIXIVQMREXIH

and she suspects that it was encrypted with a simple substitution cipher using a shift by ***n***. Then she can try each of the 26 possible keys, "decrypting" the message with each considered key and checking whether the resulting putative plaintext makes sense.

In general, a simple substitution cipher can employ any permutation of the alphabet as a key, which implies that there are $26! \approx 2^{88}$ possible keys.

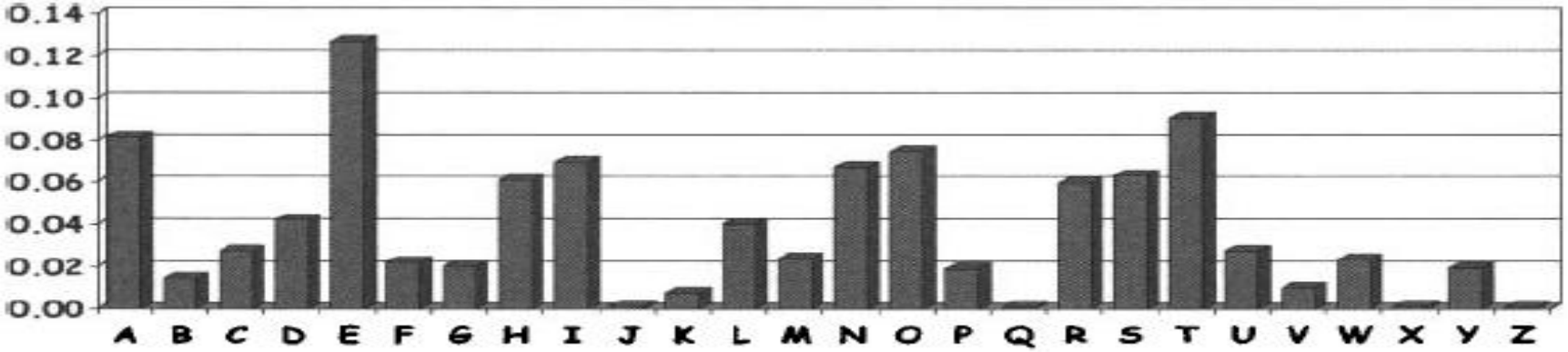
Cryptanalysis of a Simple Substitution

Suppose Trudy intercepts the following ciphertext, which she suspects was produced by a simple substitution cipher, where there could be any permutation of the alphabet:

```
PBFPVYFBQXZTYFPBFEQJHDXXQVAPTPQJKTOYQWIPBVWLXTOXBTFXQWA
XBVCXQWAXFQJVWLEQNTOZQGGQLFXQWAKVWLXQWAEBIPBFXFQVXGTVJV
WLBTPQWAEBFPBFHCVLXBQUFEVWLXGDPEQVPQGVPPBFTIXPFHXZHVFAG
FOTHFEFBQUFTDHZBQPOTHXTYFTODXQHFTDPTOGHFQPBQWAQJJTODXQH
FOQPWTBDHHIXQVAPBFZQHCFWPFHPBFIPBQWKFABVYYDZBOTHPBQPQJT
QOTOGHFQAPBFEQJHDXXQVAVXEBQPEFZBVFOJIWFFACFCCFHQWAUVWFL
QHGFXVAFXQHFUFHILTTAVWAFFAWTEVOITDHFHFQAITIXPFHXAFQHEFZ
QWGFLVWPTOFFA
```

Cryptanalysis of a Simple Substitution

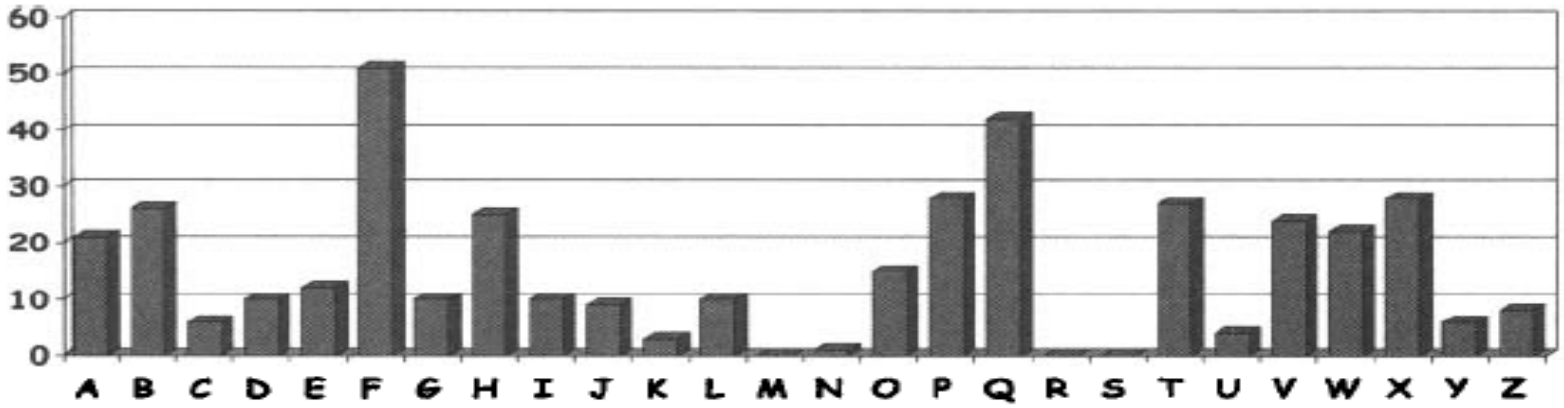
Since it's too much work for Trudy to try all 2^{88} possible keys, can she be more clever? Assuming the plaintext is English, Trudy can make use of the English letter frequency counts together with the frequency counts for the ciphertext.



English Letter Frequency Counts

Cryptanalysis of a Simple Substitution

The frequency counts for the ciphertext is shown in the following figure:



Ciphertext Frequency Counts

Cryptanalysis of a Simple Substitution

From the ciphertext frequency counts, we see that “F” is the most common letter in the encrypted message and, according to figure of English letter frequency counts, “E” is the most common letter in the English language. Trudy therefore surmises that is likely that “F” has been substituted for “E”. Continuing in this manner, Trudy can try likely substitutions until she recognizes words, at which point she can be confident in her guess.

Definition of Secure

We say that a cryptosystem is secure if the best-known attack requires as much work as an exhaustive key search.

A secure cipher with a small number of keys could be easier to break than an insecure one with a large number of keys.

In practice, we must select a cipher that is secure and has a large enough key space so that an exhaustive key search is impractical.

Double Transposition Cipher – simple form

To encrypt with a double transposition cipher, we first write the plaintext into an array of a given size and then permute the rows and the columns according to specified permutations. For example, suppose we write the plaintext **attackatdawn** into a 3×4 array:

$$\begin{bmatrix} \mathtt{a} & \mathtt{t} & \mathtt{t} & \mathtt{a} \\ \mathtt{c} & \mathtt{k} & \mathtt{a} & \mathtt{t} \\ \mathtt{d} & \mathtt{a} & \mathtt{w} & \mathtt{n} \end{bmatrix}$$

Double Transposition Cipher

If we transpose (or permute) the rows according to $(1,2,3) \rightarrow (3,2,1)$ and then transpose the columns according to $(1,2,3,4) \rightarrow (4,2,1,3)$, we obtain

$$\begin{bmatrix} a & t & t & a \\ c & k & a & t \\ d & a & w & n \end{bmatrix} \longrightarrow \begin{bmatrix} d & a & w & n \\ c & k & a & t \\ a & t & t & a \end{bmatrix} \longrightarrow \begin{bmatrix} n & a & d & w \\ t & k & c & a \\ a & t & a & t \end{bmatrix}$$

The ciphertext is then read from the final array: **NADWTKCAATAT**

Double Transposition Cipher

For the double transposition, the key consists of the size of the matrix and the row and the column permutations. Anyone who knows the key can simply put the ciphertext into the appropriate sized matrix and undo the permutations to recover the plaintext. E.g., to decrypt **NADWTKCAATAT**, the ciphertext is first put into a 3×4 array. Then the columns are numbered as $(4,2,1,3)$ and rearranged into $(1,2,3,4)$,

$$\begin{bmatrix} \mathrm{N} & \mathrm{A} & \mathrm{D} & \mathrm{W} \\ \mathrm{T} & \mathrm{K} & \mathrm{C} & \mathrm{A} \\ \mathrm{A} & \mathrm{T} & \mathrm{A} & \mathrm{T} \end{bmatrix} \longrightarrow \begin{bmatrix} \mathrm{D} & \mathrm{A} & \mathrm{W} & \mathrm{N} \\ \mathrm{C} & \mathrm{K} & \mathrm{A} & \mathrm{T} \\ \mathrm{A} & \mathrm{T} & \mathrm{T} & \mathrm{A} \end{bmatrix} \longrightarrow \begin{bmatrix} \mathrm{A} & \mathrm{T} & \mathrm{T} & \mathrm{A} \\ \mathrm{C} & \mathrm{K} & \mathrm{A} & \mathrm{T} \\ \mathrm{D} & \mathrm{A} & \mathrm{W} & \mathrm{N} \end{bmatrix}$$

And we see that we have recovered the plaintext **attackatdawn**.

One-Time Pad

The one-time pad, which is also known as the Vernam cipher, is a provably secure cryptosystem.

For simplicity, let us consider an alphabet with only eight letters and their binary representation.

letter	e	h	i	k	l	r	s	t
binary	000	001	010	011	100	101	110	111

It is important to note that the mapping between letters and bits is not secret.

One-Time Pad

Suppose that Alice, who recently got the job as a spy, wants to use a one-time pad to encrypt the plaintext message **heilhilter**. She first consults the previous table to convert the plaintext letters to the bit string

001 000 010 100 001 010 111 100 000 101.

The one-time pad key consists of a randomly selected string of bits that is the same length as the message. The key is the XORed with the plaintext to yield the ciphertext. Since $x \otimes y \otimes y = x$, decryption is accomplished by XOR-ing the same key with the ciphertext.

One-Time Pad

Now suppose that Alice has the key

111 101 110 101 111 100 000 101 110 000

which is the proper length to encrypt her message above. Then to encrypt, Alice computes the ciphertext as

	h	e	i	l	h	i	t	l	e	r
plaintext:	001	000	010	100	001	010	111	100	000	101
key:	111	101	110	101	111	100	000	101	110	000
ciphertext:	110	101	100	001	110	110	111	001	110	101
	s	r	l	h	s	s	t	h	s	r

Converting these ciphertext bits back into letters, the ciphertext message to be transmitted is **srlhssthsr**.

One-Time Pad

When her fellow spy, Bob, receives Alice's message, he decrypts it using the same shared key and thereby recovers the plaintext:

	s	r	l	h	s	s	t	h	s	r
ciphertext:	110	101	100	001	110	110	111	001	110	101
key:	111	101	110	101	111	100	000	101	110	000
plaintext:	001	000	010	100	001	010	111	100	000	101
	h	e	i	l	h	i	t	l	e	r

The ADFGVX cipher

In cryptography, the ADFGVX cipher was a field cipher used by the German army during WWI. ADFGVX was in fact an extension of an earlier cipher called ADFGX. The cipher is named after the six possible letters used in the cipher text: A, D, F, G, V and X. These letters were chosen deliberately because they sound very different from each other when transmitted via Morse code. The intention was to reduce the possibility of operator error.

The ADFGVX cipher

Rules of ADFGVX encryption:

- Remove spaces and punctuation marks from message
- Then substitute each letter or number with a pair of letters from the column's and the row's headings
- Next, use a transposition operation on the pair of letters using a key word (which the receiver knows)
- Rearrange the columns of the new arrangement in alphabetical order
- Finally, arrange the letters from consecutive columns

The ADFGVX cipher

Example The cipher for “Attack now1”

Substitution is below 26 letters 10 digits

Cipher is AV FA FA AV GF XG AG DD FG DV

	A	D	F	G	V	X
A	e	i	u	n	a	B
D	m	o	v	h	1	l
F	t	d	s	w	8	7
G	P	r	c	x	2	g
V	f	o	y	k	q	6
X	5	z	3	9	4	0

The Vigenère cipher

- Vigenère cipher starts with a 26 x 26 matrix of alphabets in sequence. First row starts with ‘A’, second row starts with ‘B’, etc.
- Like the ADFGVX cipher, this cipher also requires a keyword that the sender and receiver know ahead of time
- Each character of the message is combined with the characters of the keyword to find the cipher text character

The Vigenère cipher

	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

The Vigenère cipher

המחרוזת "שלום רב" לפי המפתח "סלמנדרה" תוצפן למחרוזת "מאצד פו"

ת	ש	ר	ק	צ	פ	ע	ס	נ	מ	ל	כ	י	ט	ח	ז	ו	ה	ד	ג	ב	א
ת	ש	ר	ק	צ	פ	ע	ס	נ	מ	ל	כ	י	ט	ח	ז	ו	ה	ד	ג	ב	א
ש	ר	ק	צ	פ	ע	ס	נ	מ	ל	כ	י	ט	ח	ז	ו	ה	ד	ג	ב	א	ת
ר	ק	צ	פ	ע	ס	נ	מ	ל	כ	י	ט	ח	ז	ו	ה	ד	ג	ב	א	ת	ש
ק	צ	פ	ע	ס	נ	מ	ל	כ	י	ט	ח	ז	ו	ה	ד	ג	ב	א	ת	ש	ר
צ	פ	ע	ס	נ	מ	ל	כ	י	ט	ח	ז	ו	ה	ד	ג	ב	א	ת	ש	ר	ק
פ	ע	ס	נ	מ	ל	כ	י	ט	ח	ז	ו	ה	ד	ג	ב	א	ת	ש	ר	ק	צ
ע	ס	נ	מ	ל	כ	י	ט	ח	ז	ו	ה	ד	ג	ב	א	ת	ש	ר	ק	צ	פ
ס	נ	מ	ל	כ	י	ט	ח	ז	ו	ה	ד	ג	ב	א	ת	ש	ר	ק	צ	פ	ע
נ	מ	ל	כ	י	ט	ח	ז	ו	ה	ד	ג	ב	א	ת	ש	ר	ק	צ	פ	ע	ס
מ	ל	כ	י	ט	ח	ז	ו	ה	ד	ג	ב	א	ת	ש	ר	ק	צ	פ	ע	ס	נ
ל	כ	י	ט	ח	ז	ו	ה	ד	ג	ב	א	ת	ש	ר	ק	צ	פ	ע	ס	נ	מ
כ	י	ט	ח	ז	ו	ה	ד	ג	ב	א	ת	ש	ר	ק	צ	פ	ע	ס	נ	מ	ל
י	ט	ח	ז	ו	ה	ד	ג	ב	א	ת	ש	ר	ק	צ	פ	ע	ס	נ	מ	ל	כ
ט	ח	ז	ו	ה	ד	ג	ב	א	ת	ש	ר	ק	צ	פ	ע	ס	נ	מ	ל	כ	י
ח	ז	ו	ה	ד	ג	ב	א	ת	ש	ר	ק	צ	פ	ע	ס	נ	מ	ל	כ	י	ט
ז	ו	ה	ד	ג	ב	א	ת	ש	ר	ק	צ	פ	ע	ס	נ	מ	ל	כ	י	ט	ח
ו	ה	ד	ג	ב	א	ת	ש	ר	ק	צ	פ	ע	ס	נ	מ	ל	כ	י	ט	ח	ז
ה	ד	ג	ב	א	ת	ש	ר	ק	צ	פ	ע	ס	נ	מ	ל	כ	י	ט	ח	ז	ו
ד	ג	ב	א	ת	ש	ר	ק	צ	פ	ע	ס	נ	מ	ל	כ	י	ט	ח	ז	ו	ה
ג	ב	א	ת	ש	ר	ק	צ	פ	ע	ס	נ	מ	ל	כ	י	ט	ח	ז	ו	ה	ד
ב	א	ת	ש	ר	ק	צ	פ	ע	ס	נ	מ	ל	כ	י	ט	ח	ז	ו	ה	ד	ג

The Vigenère cipher

- Easiest way to handle Vigenère cipher is to use arithmetic modulo 26
- This approach dispenses with the need for the table
- Keyword is converted to numbers and corresponding numbers in message and keyword are added modulo 26
  - $C_i = E_k(M_i) = (M_i + K_i) \bmod 26$
  - $M_i = D_k(Ci) = (C_i - K_i) \bmod 26$

Playfair cipher

- Each plaintext letter is replaced by a di-gram in this cipher
- Number of di-grams is 26 x 26 = 676
- User chooses a keyword and puts it in the cells of a 5 x 5 matrix. I and J stay in one cell. Duplicate letters appear only once.
- Alphabets that are not in the keyword are arranged in the remaining cells from left to right in successive rows in ascending order

Playfair cipher

Keyword “Infosec”

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

Playfair cipher

Rules:

- Group plaintext letters two at a time
- Separate repeating letters with an x
- Take a pair of letters from plaintext
- Plaintext letters in the same row are replaced by letters to the right (cyclic manner)
- Plaintext letters in the same column are replaced by letters below (cyclic manner)
- Plaintext letters in different row and column are replaced by the letter in the row corresponding to the column of the other letter and vice versa

Playfair cipher

- E.g., Plaintext: “CRYPTO IS TOO EASY”
- Keyword is “INFOSEC”
- Grouped text: CR YP TO IS TO XO EA SY
- Ciphertext: AQ VT YB NI YB YF CB OZ

To decrypt, the receiver reconstructs the 5 x 5 matrix using the keyword and then uses the same rules as for encryption