

Course: Data Security and Cryptology

אבטחת מידע וקריפטולוגיה 61767

מטרת הקורס לספק לסטודנטים ידע בסיסי בנושא הקריפטולוגיה בדגש על אופי המדעי של הקריפטולוגיה המודרנית. הסטודנטים ייחשפו למגוון מערכות כגון מערכות הצפנה ומערכות חתימה דיגיטלית, ילמדו כיצד לנתח את בטיחותן, וכיצד לעשות בהן שימוש במגוון יישומים.

תיאור מהלך הקורס והדרישות: הקורס מוקדש לרעיונות בסיסיים ותפיסות בקריפטולוגיה מודרנית ו מימושים לאבטחת מידע. במהלך הקורס נדון באלגוריתמי הצפנה קלאסיים, צפנים סימטריים, הצפנת מפתח ציבורית, יישומים ופרוטוקולים מתקדמים.

Course: Data Security and Cryptology

1. מבוא לצפנים קלאסיים.
2. ניתוח הצפנים הקלאסיים: לינארי, מעריכי, Vigenere, צפני החלפה, צפני תמורה. צפני זרם (stream ciphers). צופן היל (Hill cipher).
3. מערכת פנקס חד-פעמי (One-time pad).
4. שימוש בתורת האינפורמציה. אנטרופיה.
5. סודיות מושלמת (Perfect secrecy). Unicity Distance.
6. צפני בלוקים. רשתות Feistel. תקן הצפנה (DES) Data Encryption Standard. האלגוריתם DES3.
7. התקפה Meet-in-the-Middle. מצבי יסוד הפעולה: ECB ו-CBC.
8. תקן הצפנה מתקדם (Advanced Encryption Standard (AES).
9. פרוטוקול החלפת מפתחות Diffie-Hellman. בעיית הלוגריתם הדיסקרטי. צופן El-Gamal.
10. מבוא למערכות מפתח פומבי. אלגוריתם RSA.
11. הצפנה בעקום אליפטי (Elliptic Curve methods). (אופציונלי)
12. פונקציית גיבוב (Hash) ושיטות חתימה דיגיטלית.
13. פרוטוקול אפס ידיעה (Zero-Knowledge). (אופציונלי)
14. בקרת גישה, קובץ הגנה, אימות משתמש.
15. מדיניות אבטחה, מודלים של אבטחה.

Course: Data Security and Cryptology

Recommended literature:

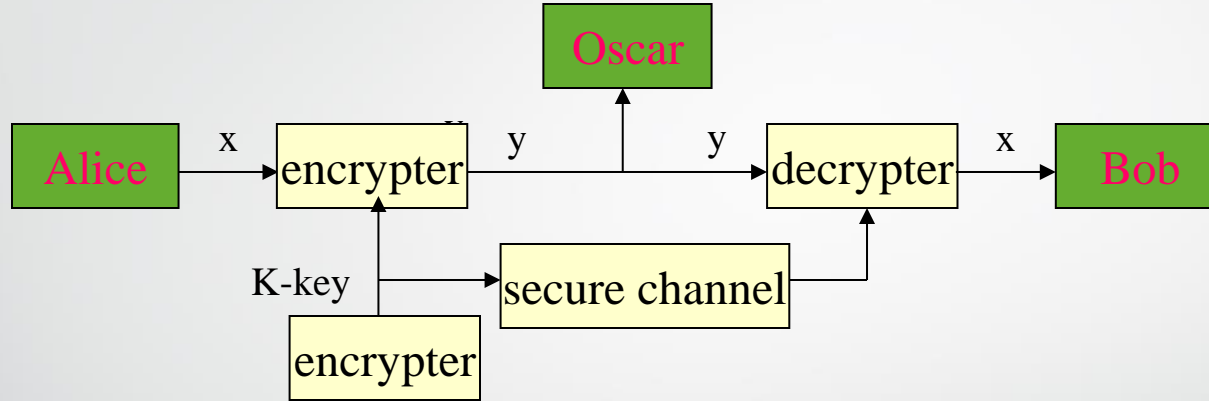
1. Stinson, D. R., Cryptography: Theory and Practice, (3rd ed.), Chapman & Hall, 2006.
2. Katz J. and Lindell Y., Introduction to Modern Cryptography, Second Edition, Chapman & Hall/CRC Cryptography and Network Security Series, 2007.
3. Paar C. and Pölz J., Understanding Cryptography Springer, ISBN: 978-3-642-04100-6, (2010).



Terminology

Plaintext	Original message
Ciphertext	Encrypted or coded message
Encryption	Convert from plaintext to ciphertext (enciphering)
Decryption	Restore the plaintext from ciphertext (deciphering)
Key	Information used in cipher known only to sender/receiver
Cryptography	Study of algorithms used for encryption
Cipher	A particular algorithm (cryptographic system)
Cryptanalysis	Study of techniques for decryption without knowledge of plaintext
Cryptology	Areas of cryptography and cryptanalysis

Communication Model



1. Two parties – Alice and Bob
2. Reliable communication line
3. Shared encryption scheme: E , D , k
4. Goal: send a message m confidentially

Cryptography

Cryptography is used for other purposes besides ensuring confidentiality:

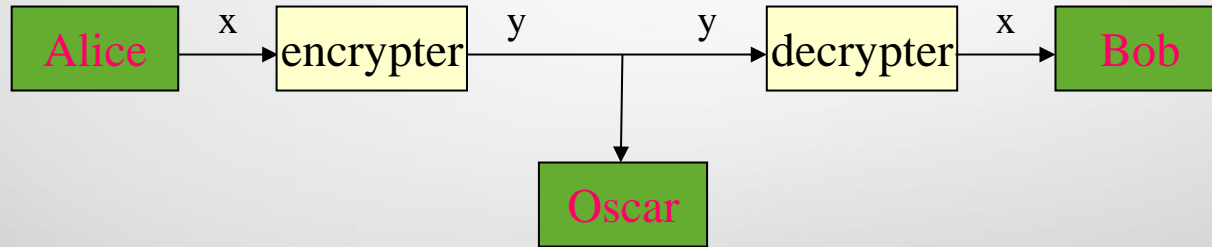
- **Authentication** – ability to identify the sender and nobody can't give himself out to be a sender
- **Integrity** - a “copy” of a signed digital message is identical to the original
- **Nonrepudiation** – the sender can't say that it's not he who send

Eavesdropping

Attackers can try to get the information they need in various ways.

Passive eavesdropping:

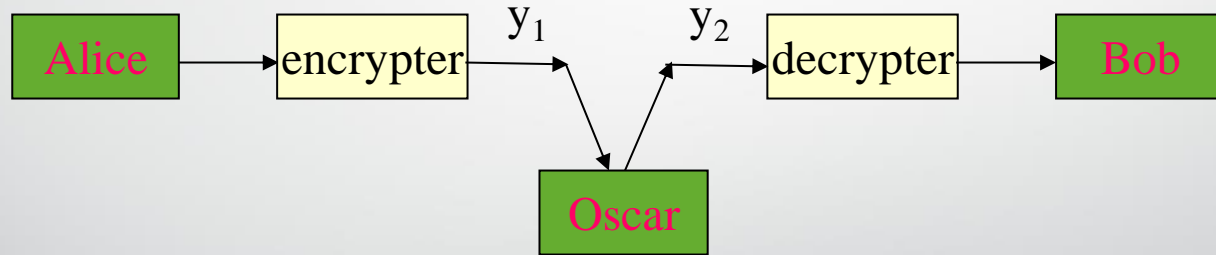
The attacker can only listen to the communication:



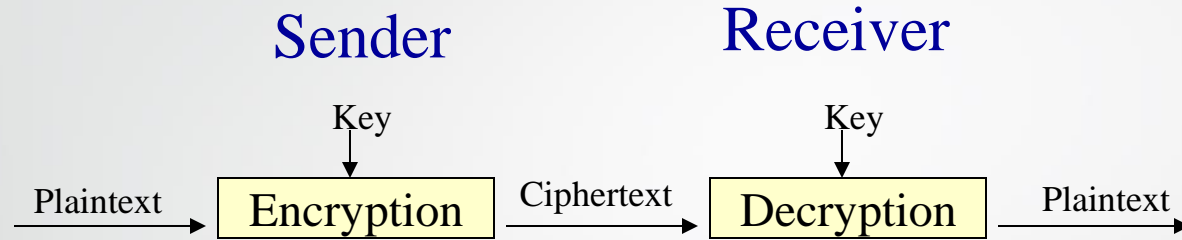
Eavesdropping

Active eavesdropping:

The attacker can modify the communication.



Encryption - Decryption



1. Plaintext - M – the message space
2. Ciphertext - C
3. K - the key space

$\forall k \in K$

$$E_k(M)=C, D_k(C)=M \iff D_k = (E_k)^{-1}$$

Encryption - Decryption

Standards:

1. E_k - invertible for each k
2. For each k_2 there is k_1 such that $D_{k_2} = E_{k_1}^{-1}$
3. If $k_2 \neq k_1$ then there exists M such that
$$E_{k_1}(M) \neq E_{k_2}(M)$$

Cipher Securities is based on Key but not on Algorithm.

Algorithm is public !!!

**Cryptosystem = possible M + possible C + Keys +
+ Algorithms of Encryption and Decryption**

Formal Definition of Cryptosystem

A cryptosystem is a five-tuple (P, C, K, E, D) , where the following conditions are satisfied:

1. P is a finite set of possible plaintexts,
2. C is a finite set of possible ciphertexts,
3. K , the keyspace, is a finite set of possible keys,
4. For each $k \in K$, there is an encryption rule $e_k \in E$ and a corresponding decryption rule $d_k \in D$.

Each E_k and D_k are functions such that

$D_K(E_K(X)) = X$ for every plaintext $X \in P$.

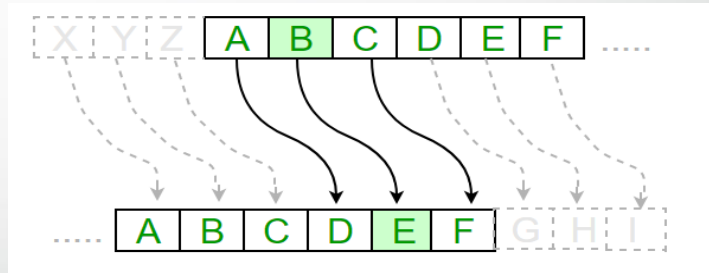
Classical Cryptosystems

The first known algorithmic encryption is the **Caesar Cipher**.

a b c d e f gw x y z

Julius Caesar encrypted his messages by substituting each letter in the text by the third letter thereafter (cyclically):


$a \rightarrow d$	$w \rightarrow z$
$b \rightarrow e$	$x \rightarrow a$
$c \rightarrow f$	$y \rightarrow b$
$d \rightarrow g$	$z \rightarrow c$



Thus, the *caesar cipher* is encrypted to *fdhvdu flskhu*.

Weakness: Everyone who knows the encryption scheme can decrypt any message.

Examples from Literature and History

- Simple substitution ciphers:
 - Arthur Conan Doyle “Sherlock Holmes tale: The Adventure of the Dancing Men”

 - Edgar Allan Poe “The Gold Bug”
- Greeks: **scytale** cipher
- **Morse code**
- WWII: **Enigma machine**
- **ASCII** (American Standard Code for Information Interchange): Each character is encoded by numbers from 0 to 127 in binary format



Classical Cryptosystems

Shift Cipher:

Letters of the alphabet are assigned a number as below in the case when the alphabet is the English one. So the alphabet size (n) equals to 26.

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

Classical Cryptosystems

Shift Cipher:

Here $P = C = K = Z_n$. The plaintext P , the ciphertext C and the key K are elements of the ring Z_n

For $0 \leq k \leq n - 1$, define

encryption: $E_k(x) = (x + k) \bmod n = y$

and

decryption: $D_k(y) = (y - k) \bmod n = x \quad (x, y \in Z_n).$

For the particular key $k = 3$, the cryptosystem is often called the **Caesar Cipher**.

Shift Cipher

Example: Let the key $k = 17$

Plaintext: $X = A T T A C K = (0, 19, 19, 0, 2, 10).$

Ciphertext : $Y = (0+17 \bmod 26, 19+17 \bmod 26, \dots)$

$Y = (17, 10, 10, 17, 19, 1) = R K K R T B$

Attacks on Shift Cipher

- **Exhaustive Search:** Try all possible keys. $|K|=26$. can be easily broken by attackers. The problem is the small set of keys. Nowadays, for moderate security $|K| \geq 280$, for recommended security $|K| \geq 2100$.
- **Letter frequency analysis:** will be discussed later.

Brute-force or exhaustive search: example

Brute-force attack: Try every possible key on ciphertext until intelligible translation into plaintext obtained

Key	Message
0:	LZWJWAKFGGLZWDJSFYMSYWTMLXJWFUZ
1:	KYVIVZJEFFKYVICREXLRXVSLKWIVETY
2:	JXUHUYIDEEJXUHBQDWKQWURKJVHUDSX
3:	IWTGTXHCDDIWTGAPCVJPVTQJIUGTCRW
4:	HVSFSWGBCCHVSFZOBUIOUSPIHTFSBQV
5:	GURERVFABBGUREYNATHNTROHGSEAPU
6:	FTQDQUEZAAFTQDXMZSGMSQNGFRDQZOT
7:	ESPCPTDYZZESPWCWLYRFLRPMFEQCPYNS
8:	DROBOSCXYDROBVKXQEKQOLEDPBXMR
9:	CQNANRBWXXCQNAUJWPDJPNKDCOANWLQ
10:	BPMZMQAVWWBPMZTIVOC IOMJCBNZMVKP
11:	AOLYLPZUVVAOLYSHUNBHNLIBAMYLUJO
12:	ZNKXKOYTUUZNXRGTMAGMKHAZLXKTIN
13:	YMJWJNXSTTYMJWQFSLZFLJGZYKWSHM
14:	XLIVIMWRSSXLIVPERKYEKIFYXJVIRGL
15:	WKHUHLVQRRWKHUODQJXDJHEXWIUHQFK
16:	VJGTGKUPQQVJGTNCPIWCIGDWVHTGPEJ
17:	UIFSFJTOPPUIFSMBOHVBFHFCVUGSFODI
18:	THEREISNOOTHERLANGUAGEBUTFRENCH
19:	SGDQDHRMNNSGDQKZMFTZFDATSEQDMBG
20:	RFCPCGQLMMRFPCPJYLESYECZSRDPCLAF
21:	QEBOBFPKLLQEBOIXKDRXDBYRQCOBKZE
22:	PDANAEOKKKPDANHWJCQWCAXQPBNAJYD
23:	OCZMZDNIJJOCZMGVIBPVBZWPOAMZIXC
24:	NBYLYCMHIINBYLFUHAOUAYVONZLYHWB
25:	MAXKXBLGHHMAXKETGZNTZXUNMYKXGVA

M = THERE IS NO OTHER LANGUAGE BUT FRENCH.†



Kerckhoffs' Principle

While assessing the strength of a cryptosystem, one should always assume that the enemy knows the cryptographic algorithm used.

The security of the system, therefore, should be based on:

- the quality (strength) of the algorithm but not its obscurity
- the key space (or key length)

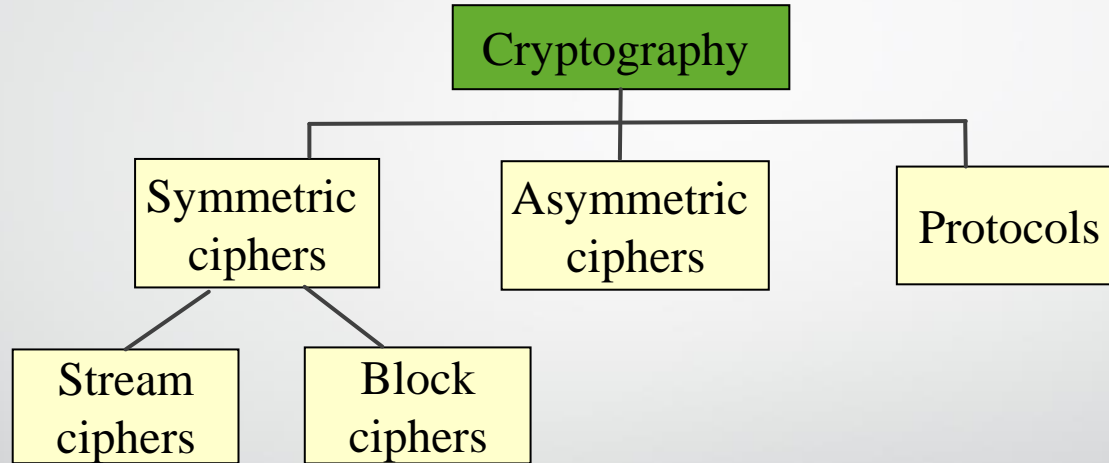


Security of Algorithms

We are safe if:

- The cost required to break an algorithm is greater than the value of the encrypted data.
- The time required to break an algorithm is longer than the time the encrypted data must remain secret.
- The amount of data encrypted with a single key is less than the amount of data needed to break the algorithm.

Classification



Classification 1

Restricted Algorithms

- A *restricted* cryptographic algorithm is one whose security requires keeping the algorithm secret.
- An example: German Enigma.
- Restricted algorithms are impractical with a large or changing group of users.
- They also don't permit quality control or standardization.
- Modern cryptographic algorithms are not restricted; they may even be published.
- Their security derives from the use of one or more *keys*.

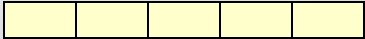
Classification 2

**There are two types of key-based algorithms:
symmetric and public-key**

- With *symmetric algorithms*, the decryption key can be calculated from the encryption key.
- With most symmetric algorithms, the same key is used for encryption and decryption.
- The sender and receiver must agree on a key before they can communicate secretly.
- The key must be kept secret from others.

Classification 3

There are two types of symmetric algorithms:

- **Stream algorithms** or stream ciphers operate on the plaintext one bit (or byte) at a time.
- **Block algorithms** or block ciphers operate on groups of bits called blocks. 

Public-Key Algorithms 1

With *public-key* or *asymmetric* algorithms, different keys are used for encryption and decryption.

Moreover, it is not feasible to compute the decryption key from the encryption key.

This means the encryption key can be made public.

So, it is often called the *public key*, and the decryption key is called the *private key*. Public key encryption and decryption can be denoted as follows:

$$E_{K_1}(M) = C \text{ and } D_{K_2}(C) = M.$$

Public-Key Algorithms 2

- Each one can encrypt. **Public Key** - known to anyone in the systems with assurance;
- But only a receiver can decrypt. **Private Key** – known only to the owner;
- As a rule, it is impossible (difficult) calculate k_2 from k_1 . Usually, k_1 is public and k_2 is private.
- May be quite the reverse – for Signature Algorithm.

Hybrid

Combines strengths of both methods. Asymmetric distributes symmetric key, also known as a *session key*. Symmetric provides bulk encryption.

Example: SSL negotiates a hybrid method.

There two algorithms: a **symmetric ES** and a **public key EP** ones.

1. Alice encrypts a key *session key* of ES by means of EP and send it to Bob: $\text{EncryptedKey} = \text{EP}_{K_1}(k)$;
2. Bob encrypts $k = \text{DP}_{K_2}(\text{EncryptedKey})$;
3. This point Alice and Bob are aware of a shared key k and are able to connect using the symmetric algorithm ES.



One-Time Pads

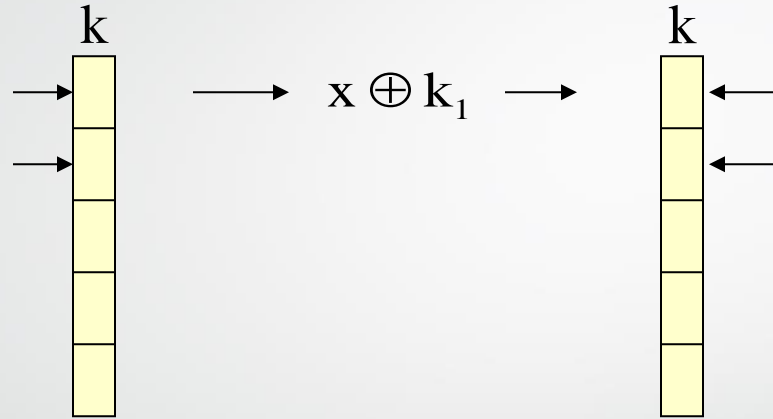
- A classical *one-time pad* is a large set of truly random key letters, written on sheets of paper and glued together in a pad.
- The sender uses each key letter on the pad to encrypt exactly one plaintext character.
- Each key letter is used exactly once, for only one message.
- The sender destroys used parts of the pad. The receiver has an identical pad and uses it to decrypt the ciphertext.
- The idea has been provided by Major Joseph Mauborgne in 1917.



One-Time Pads

- U.S. Army Major Joseph Mauborgne and AT&T's Gilbert Vernam developed a cipher in 1917
- Uses a one time arrangement of a key string that is as long as the plaintext
- Plaintexts are assumed to be short
- Also known as **One-Time Pad** cipher
- Key is used only once but characters in key may not be distinct

One Time Pad



$$k_i = (a \cdot k_{i-1} + b) \bmod n$$

RNG

$k_0 = \text{TIMESTAMP}$

Random Number Generators

True random number generators (TRNGs) are characterized by the fact that their output cannot be reproduced.

Pseudorandom Number Generators (PRNG)

Often are computed recursively in the following way:

s_0 = seed value,

$s_{i+1} = f(s_i)$.

A popular example is the **Linear congruential generator**

s_0 = seed value,

$s_{i+1} = (a*s_i + b) \bmod n$, where a, b, n are integer values.

Note that PRNGs are not random in a true sense because they can be computed and are thus completely deterministic.

One-Time Pad: Example

- Suppose the message consists of n characters

$$x = (x_1, x_2, \dots, x_n).$$

- We choose a key consisting of n random characters

$$k = (k_1, k_2, \dots, k_n).$$

- Each character is viewed as an integer in the range 0...25.
- The one-time pad encryption function is defined by:

$$E_k(x) = (x_i + k_i) \% 26$$

- The decryption function is defined by:

$$D_k(y) = (y_i - k_i) \% 26$$

One-Time Pad: Example

Example [Garrett]:

Suppose the plaintext is “IMPOSSIBLE” coded by integers:

$$x = (8, 12, 15, 14, 18, 18, 8, 1, 11, 4)$$

Let the key be

$$k = (8, 13, 24, 19, 9, 1, 0, 7, 20, 3)$$

Then

$$E_k(x) = ((8 + 8) \% 26, (12 + 13) \% 26, \dots, ((4 + 3) \% 26) = (16, 25, 13, 7, 1, 19, 8, 8, 5, 7) = \text{“QZNHBTIIFH”}$$

The Linear (Affine) Cipher

The Linear (Affine) Cipher :

$$E_k(x) = (ax + b) \bmod n$$

The key $k = (a, b)$ and $a, b, x, y \in \mathbb{Z}_n$. **Example:** $k = (a, b) = (13, 4)$

- INPUT = (8, 13, 15, 20, 19) \Rightarrow ERRER
- ALTER = (0, 11, 19, 4, 17) \Rightarrow ERRER
- There is no one-to-one map between plaintext and ciphertext space.
What went wrong?

Decryption: $D_k(y) = a^{-1}(y - b) \bmod n$ is possible only if a^{-1} is exist

$$\Leftrightarrow \gcd(a, n) = 1.$$

Hence the Affine Cipher has $\phi(n)$ possible keys ($\phi(26) = 12$).

The Linear (Affine) Cipher

An example:

$$E_k(x) = (7x + 3) \bmod 26$$

and

$$D_k(y) = (15y - 19) \bmod 26$$

‘one to one’ – the necessary property,
each function with this property may be used

The Linear (Affine) Cipher

Attack types:

- *Ciphertext only*: exhaustive search or frequency analysis
- *Known plaintext*: two letters in the plaintext and corresponding ciphertext letters would suffice to find the key.

Example: plaintext IF=(8, 5) and ciphertext PQ=(15, 16)

$$8a + b \equiv 15 \pmod{26}$$

$$5a + b \equiv 16 \pmod{26} \quad \Rightarrow a = 17 \text{ and } b = 9$$

What happens if we have only one letter of known plaintext?

The Exponential Cipher

$$E_k(x) = (ax^b) \bmod n \text{ and } D_k(y) = (a^{-1}y)^c \bmod n$$

Here $c = b^{-1} \bmod \theta(n)$.

Therefore a key (a,b) is admissible only if

- $a^{-1} \bmod n$ is exist $\Leftrightarrow \gcd(a,n) = 1$,
- $\gcd(b, \theta(n)) = 1$.

Hence are exist $\theta(n) \cdot \theta(\theta(n))$ possible keys.

$$12 = 2 \cdot 2 \cdot 3 \Rightarrow \text{teta}(12) = 2 \cdot 2 \Rightarrow 12 \cdot 4 = 48$$

Monoalphabetic Substitution Ciphers

Each letter in the alphabet is replaced (substituted) by another letter. More precisely, a permutation of the alphabet is chosen and applied to the plaintext.

The **Shift**, **Affine** and **Exponential** ciphers are examples of substitution ciphers.

Since ciphertext preserves the statistic of the language used in the plaintext, the frequency analysis is an effective way of Breaking substitution ciphers.

<http://www.sherlockian.net/canon/stories/danc.html>

Monoalphabetic Substitution Ciphers

Example: The key is a permutation:

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

Encryption:

Plaintext: monoalphabetic substitution

Ciphertext: cvgvpnkopdrehuwydweheyehvg

Decryption:

Ciphertext: cvgvpnkopdrehuwydweheyehvg

Plaintext: monoalphabetic substitution

Substitution Ciphers: Security

The number of possible keys is $26! = 4 \cdot 10^{26} = 1.3 \cdot 2^{88}$.

Therefore, the key can be represented with 89 bits.

Clearly, it is impractical to search all the key space exhaustively, and the probability of guessing the key is very low.

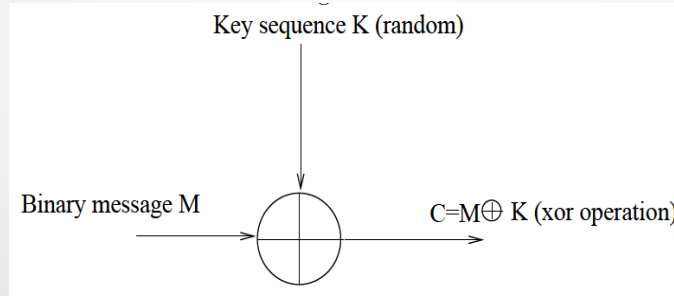
Therefore, it seems that this cipher is secure.

Actually, all classical ciphers are special cases of this general model. It can be broken by a statistical attack.

Stream Ciphers

Here the plaintext, the ciphertext and the key stream consist of individual bits (Bytes)

Vernam Cipher



encryption: $E_k(x) = (x \oplus k)$

decryption: $D_k(y) = (y \oplus k)$

where $x, y, k \in \mathbb{Z}_2$.

Cryptanalysis

Cryptanalysis is the science of recovering the Plaintext of a message without access to the key(s).
An attempted cryptanalysis is called an *attack*.
It's a good idea for a cryptographer to assume that the cryptanalyst knows the cryptographic algorithm.
It should also be assumed that eavesdroppers have access to communications between the sender and receiver.

Cryptanalysis

- **Ciphertext Only Attack :**

The cryptanalyst has the ciphertext of several messages to work with and attempts to deduce the corresponding plaintexts or the key(s).

- **Known Plaintext Attack :**

The cryptanalyst has access to both the ciphertext and the plaintext of several messages and attempts to deduce the key(s) or an algorithm to infer the plaintexts.

- **Chosen Plaintext Attack :**

The cryptanalyst not only has access to the ciphertext and plaintext of several messages, but he chooses the plaintext that gets encrypted. He attempts to deduce the key(s) or an algorithm to infer the plaintexts.

Cryptanalysis

- **Chosen Ciphertext Attack :**

The cryptanalyst can choose different ciphertexts to be decrypted and has access to the decrypted plaintext. He attempts to deduce the key(s).

- **Rubber Hose Cryptanalysis:**

The cryptanalyst can also use theft, bribery, extortion, etc. to obtain the key(s).

Type of Attack	Known to Cryptanalyst
Ciphertext Only	<ul style="list-style-type: none"> • Encryption algorithm • Ciphertext
Known Plaintext	<ul style="list-style-type: none"> • Encryption algorithm • Ciphertext • One or more plaintext-ciphertext pairs formed with the secret key
Chosen Plaintext	<ul style="list-style-type: none"> • Encryption algorithm • Ciphertext • Plaintext message chosen by cryptanalyst, together with its corresponding ciphertext generated with the secret key
Chosen Ciphertext	<ul style="list-style-type: none"> • Encryption algorithm • Ciphertext • Ciphertext chosen by cryptanalyst, together with its corresponding decrypted plaintext generated with the secret key
Chosen Text	<ul style="list-style-type: none"> • Encryption algorithm • Ciphertext • Plaintext message chosen by cryptanalyst, together with its corresponding ciphertext generated with the secret key • Ciphertext chosen by cryptanalyst, together with its corresponding decrypted plaintext generated with the secret key

Security of Algorithms

- A cryptographic algorithm is *unconditionally secure* if it is impossible for a cryptanalyst to recover the plaintext, regardless of how much ciphertext they have.
- Only a “one-time pad” is unbreakable in this sense. Every other cryptosystem is breakable by trying every possible key one by one.

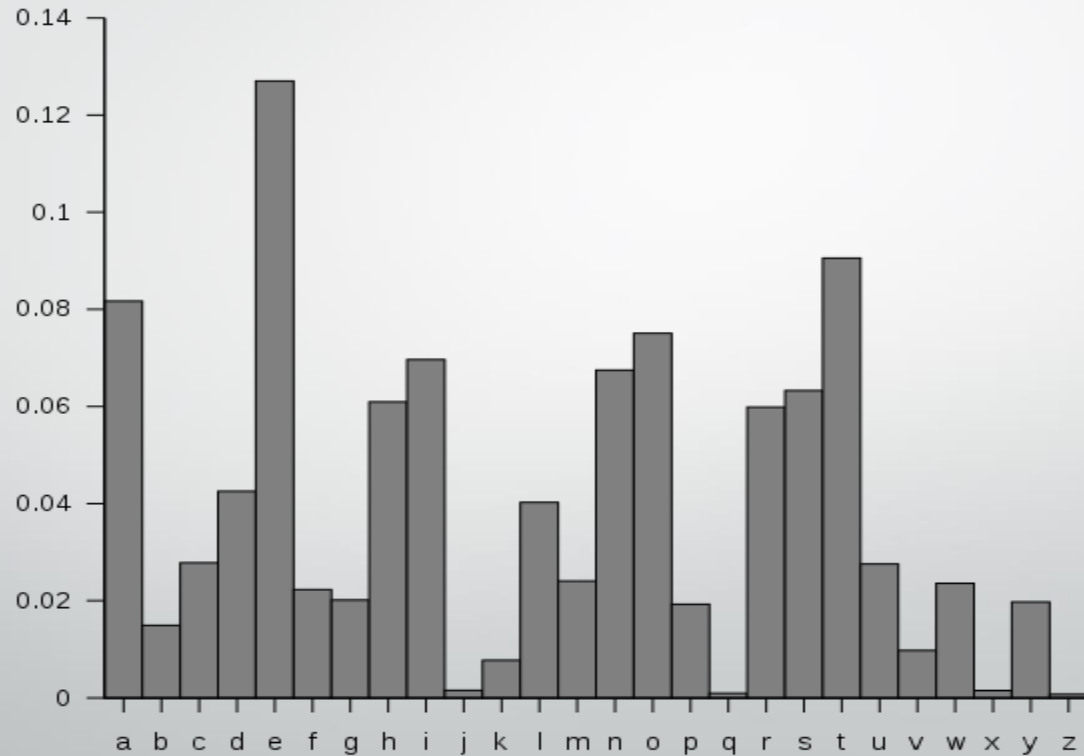
This is called a *brute-force* attack.

- An algorithm is *computationally secure* if it cannot be broken with available resources. The complexity of an attack is generally taken as the maximum of these three factors:
 - ✓ **The amount of data needed as input to the attack;**
 - ✓ **The time needed to perform the attack;**
 - ✓ **The amount of storage needed for the attack.**

English letter frequencies

Letter	Number	Letter	Number	Letter	Number
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		

English letter frequencies



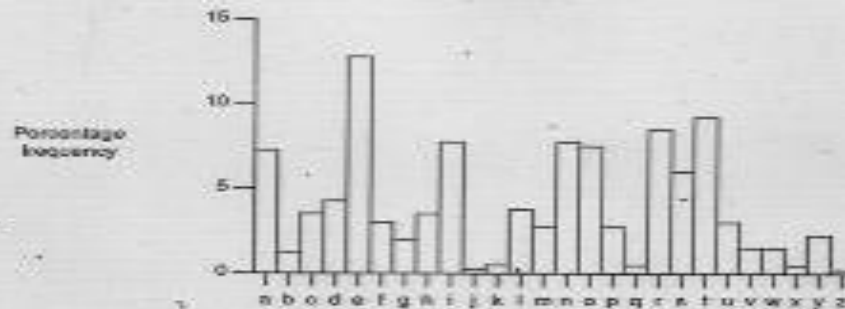


Figure 3.1 English character frequencies

Example

Suppose we have to decipher:

WZDUY ZZYQB OTHTX ZDNZD KWQHB BYQBP WZDUY ZXZDSS

We note that:

Z occurs 8 times

D occurs 5 times

Y occurs 4 times

W, Q, B occurs 3 times each

Presuming the language is English, we note that the most frequently occurring letters in English text (see Appendix A) are, in order,

E, T, R, I, N, O, A

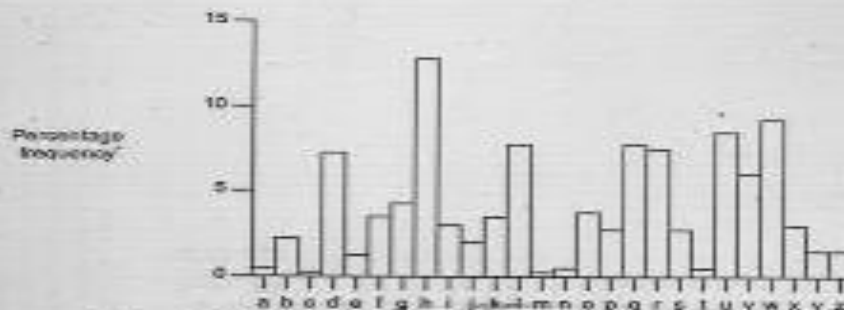


Figure 3.2 Encryption character frequencies with $i=3$

English words frequencies

The most frequent English word is THE

Word	Frequency	Word	Frequency	Word	Frequency
THE	6.421%	A	2.092%	I	0.945%
OF	4.028%	IN	1.778%	IT	0.930%
AND	3.150%	THAT	1.244%	FOR	0.770%
TO	2.367%	IS	1.034%	AS	0.764%

Frequency Analysis

- Discovered by Arabs (approx. 9th century)
- Statistically, it is possible to determine how often each letter appears in an “average” text
- Frequency table, other useful observation:
 - ST, NG, TH, and QU are common pairs of letters (bigrams), while NZ and QJ are rare.
 - What is one the most common trigram? THE
 - Letters that often appear at the beginnings of words.

Frequency Analysis

Example (S. Singh, The Code Book, 1999)

Identifying common letters, bigrams and trigrams...

PCQ VMJYPD LBYK LYSO KBXBJXWXV BXV ZCJPO EYPD KBXB
JYUXJ LBJOO KCPK. CP **LBO** LBCMXPV XPV IYJKL PYDBL, QB
OP KBO BXV OPVOV **LBO** LXRO CI SX'XJMI, KBO JCKO XPV EY
KKOV **LBO** DJCMPV ZOICJO BYS, KXUYPD: 'DJOXL EYPD, **X** LBC
MKXPV XPV CPO PYDBLK **Y** BXNO ZOOP JOACMPLYPD LC UCM
LBO IXZROK CI FXKL XDOK XPV **LBO** RODOPVK CI XPAYOPL E
YPDK. SXU Y SXEO KC ZCRV XK LC AJXNO X IXNCMJ **CI** UCMJ
SXGOKLU?' OFYRCDMO, LXROK IJCS **LBO** LBCMXPV XPV CP
O PYDBLK

First guess: **LBO** is **THE**

Frequency Analysis

- Assuming **LBO** represents **THE** we replace **L** with **T**, **B** with **H**, and **O** with **E** and get
- PCQ VMJYPD **TH**YK **TY**SE **KHXHJXWXV** **HXV** ZCJ**PE** EYPD
KHXHJYUXJ **TH**JEE KCPK. CP **THE** **TH**CMKXPV XPV IYJKT
PYD**HT**, QHEP KHO **HXV** EPVEV **THE** LXRE CI SX'XJMI, **KHE**
JCKE XPV EYKKOV THE DJCMPV ZEICJE HYS, KXUYPD: 'DJ
EXT EYPD, ICJ X LHCMKXPV XPV CPE PYDHLK Y **HXNE** **ZE**
EP JEACMPTYPD **TC** UCM **THE** IXZ**REK** CI FXKL XDEK XPV
THE REDEPVK CI XPAYE**PT** EYPDK. SXU Y **SXEE** KC ZCRV X
K **TC** AJXNE X IXNCMJ CI UCMJ SXGEKTU?' EFYRCDME, **T**
XREK IJCS **THE** **LH**CMKXPV XPV CPE PYDBTK
- *More guesses...?*



The Solution

Now during this time Shahrazad had borne King Shahriyar three sons. On the thousand and first night, when she had ended the tale of Ma'aruf, she rose and kissed the ground before him, saying: 'Great King, for a thousand and one nights I have been recounting to you the fables of past ages and the legends of ancient kings. May I make so bold as to crave a favor of your majesty?'

Epilogue, Tales from the Thousand and One Nights

The Iterative Attack

Since usually algorithm E is known and $M = C$,
if we know a cipher symbol y_0 then it is possible
to calculate the sequence:

$$y_0 \rightarrow y_1 = e(y_0) \rightarrow y_2 = e(y_1) \rightarrow \dots \rightarrow y_j = e(y_{j-1}) \rightarrow y_0 = e(y_j)$$

The sequence is returned to first element because C is
a finite ring, and view of the fact that a function
encryption is “one to one” we have $x_0 = y_j$.

The Iterative Attack: example

$$y_0 \rightarrow y_1 = e(y_0) \rightarrow y_2 = e(y_1) \rightarrow \dots \rightarrow y_j = e(y_{j-1}) \rightarrow y_0 = e(y_j)$$

An example:

$$y_0 = e(x)$$

Affine cipher

$$E_k(x) = (7x + 3) \bmod 26$$

$$K = (1, 0)$$

$$E(x) = 1 * x + 0 = x$$

$$y_{12} = e(y_{11})$$

a	b	mod	ciphertext
7	3	26	10

y_0	0	10
y_1	1	21
y_2	2	20
y_3	3	13
y_4	4	16
y_5	5	11
y_6	6	2
y_7	7	17
y_8	8	18
y_9	9	25
y_{10}	10	22
y_{11}	11	1
y_{12}	12	10

Unhidden messages

A message x is called unhidden if $E_k(x)=x$.

In case of the linear cipher

$$(a \cdot x + b) = x \bmod n.$$

It implies

$$(a-1) \cdot x = -b \bmod n$$

The equation has $g = \gcd(a-1, n)$ solutions if g divides b , and no solution otherwise.

Block Ciphers

- In the substitution ciphers, changing one letter in the plaintext changes exactly one letter in the ciphertext.
- This greatly facilitates finding the key using frequency analysis.
- Block ciphers prevents this by encrypting a block of letters simultaneously.
- Many of the modern (symmetric) cryptosystems are block ciphers. DES operates on 64 bits of blocks while AES uses 128 bits of blocks (192 and 256 are also possible).

Example: Hill Cipher

- The key is an $m \times m$ matrix whose entries are integers in \mathbb{Z}_{26} .

The Hill Cipher

The Hill Cipher (1929)

A multidimensional linear cipher (m equations)

Key $K_{[m \times m]}$ - a square matrix, $\vec{x} = (x_1, x_2, \dots, x_m)$

$$E_K(\vec{x}) = \vec{x} \cdot K \bmod n$$

and

$$D_K(\vec{y}) = \vec{y} \cdot K^{-1} \bmod n$$

It is possible if

$$\exists K^{-1} \bmod n \Leftrightarrow \gcd(\det K, n) = 1$$

The Hill Cipher

Example 1:

$$(x_1, x_2) \rightarrow (y_1, y_2)$$

$$N=26, \quad m=2$$

$$\begin{aligned} y_1 &= 11 \cdot x_1 + 3 \cdot x_2 \\ y_2 &= 8 \cdot x_1 + 7 \cdot x_2 \end{aligned} \Rightarrow K = \begin{pmatrix} 11 & 3 \\ 8 & 7 \end{pmatrix}$$

$$|K| = (77 - 24) \bmod 26 = 1 \quad \Rightarrow \exists K^{-1}$$

i.e. the matrix K is admissible

The Hill Cipher

Encryption:

the plaintext: july \implies (9,20,11,24)

$$(9,20) \begin{pmatrix} 11 & 8 \\ 3 & 7 \end{pmatrix} = (3,4)$$

$$(11,24) \begin{pmatrix} 11 & 8 \\ 3 & 7 \end{pmatrix} = (11,22)$$

(3,4,11,22) = [DELW] - the cipher text

The Hill Cipher

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad A^{-1} = \frac{1}{|A|} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Decryption:

$$K = \begin{pmatrix} 11 & 8 \\ 3 & 7 \end{pmatrix} \Rightarrow K^{-1} = \begin{pmatrix} 7 & 18 \\ 23 & 11 \end{pmatrix}$$

the ciphertext: [DELW] = (3,4,11,22)

$$\begin{pmatrix} 3 & 4 \\ 11 & 22 \end{pmatrix} \cdot \begin{pmatrix} 7 & 18 \\ 23 & 11 \end{pmatrix} = \begin{pmatrix} 9 & 20 \\ 11 & 24 \end{pmatrix}$$

\Rightarrow (9,20,11,24) = [july] - the plain text

The Hill Cipher

Example 2: Let $m = 3$ and the key matrix be

$$M = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 11 & 9 & 8 \end{pmatrix}$$

and the plaintext be $ABC = (0, 1, 2)$ then the encryption operation is a vector-matrix multiplication

$$(0, 1, 2) \times \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 11 & 9 & 8 \end{pmatrix} \equiv (0, 23, 22) \bmod 26 \Rightarrow \text{AXW (ciphertext)}$$

$$M^{-1} = \begin{pmatrix} 22 & 5 & 1 \\ 6 & 17 & 24 \\ 15 & 13 & 1 \end{pmatrix}$$

The Hill Cipher

Example 2: Let $m = 3$

$$M = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 11 & 9 & 8 \end{pmatrix}$$

and the plaintext be $ABC = (0, 1, 2)$ then the encryption operation is a vector-matrix multiplication

$$(0, 1, 2) \times \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 11 & 9 & 8 \end{pmatrix} \equiv (0, 23, 22) \bmod 26 \Rightarrow \text{AXW (ciphertext)}$$

$$M^{-1} = \begin{pmatrix} 22 & 5 & 1 \\ 6 & 17 & 24 \\ 15 & 13 & 1 \end{pmatrix}$$

$$M = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix},$$

$$M^{-1} = \frac{1}{|M|} \begin{bmatrix} ei - fh & ch - bi & bf - ce \\ fg - di & ai - cg & cd - af \\ dh - eg & bg - ah & ae - bd \end{bmatrix}$$

Cryptanalysis of the Hill Cipher

- Number of keys k = number of invertible $m \times m$ matrices with coefficients from \mathbf{Z}_{26} .

Does anyone know the formula?

- If p is prime, the alphabet is \mathbf{Z}_p then

$$k = \prod_{i=0}^{m-1} (p^m - p^i)$$

- If $p = 29$ and

m	3	4	5	10
k	$1.4 \cdot 10^{13}$	$2.4 \cdot 10^{23}$	$3.5 \cdot 10^{36}$	$1.7 \cdot 10^{146}$

Cryptanalysis of the Hill Cipher

- Easily broken with known plaintext attack.
- Permutation Cipher = Hill Cipher, where
the key is a permutation matrix.
- Both ciphers are insecure.

The Permutation Cipher

The Permutation (Transposition) Cipher is the analog of the Substitution Cipher to Polyalphabetic Cryptosystem.

It was analyzed in 1563 by Giovanni Porta.

As with , it is more convenient to use, since there are no algebraic operations. But it is a particular case of the Hill cipher.

The Permutation Cipher

Let $m \in \mathbf{Z}^+$, let $P = C = (\mathbf{Z}_{26})^m$, let $K = S_m$.

For a key (i.e. a permutation) p_π we define

$$E(x_1, x_2, \dots, x_m) = p(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(m)})$$

and

$$D(y_1, y_2, \dots, y_m) = p^{-1}(y_{\pi^{-1}(1)}, y_{\pi^{-1}(2)}, \dots, y_{\pi^{-1}(m)})$$

where p^{-1} is the inverse permutation to p .

The Vigenère Cipher

Let $m \in \mathbf{Z}^+$, let $P = C = K = (\mathbf{Z}_{26})^m$.

For a key $K = (k_1, k_2, \dots, k_m)$, we define

$$E_K(x_1, x_2, \dots, x_m) = (x_1 + k_1, x_2 + k_2, \dots, x_m + k_m)$$

and

$$D_K(y_1, y_2, \dots, y_m) = (y_1 - k_1, y_2 - k_2, \dots, y_m - k_m)$$

where all operations are modulo 26.

Considered “unbreakable” for 300 years

(broken by Babbage, Kasiski 1850's)

The Vigenère Cipher

Uses Caesar's cipher with various different shifts, in order to hide the distribution of the letters. The key defines the shift used in each letter in the text.

A key word is repeated as many times as required to become the same length as the plaintext.

The result is added to the plaintext as follows:

Plaintext: **vigenere's cipher**

vig ene res cip her

Key: **key key key key key**

Ciphertext: **fme orc biq mmn rip**

($a=0, b=1, \dots, z=25, \text{ mod } 26$).

This cipher was considered very secure in the 19'th century, and was still used in the first world war...