Οικονομικό Πανεπιστήμιο Αθηνών
Τμήμα Πληροφορικής
ΠΜΣ

Κρυπτογραφία και Εφαρμογές

Μαριάς Ιωάννης

marias@aueb.gr

Μαρκάκης Ευάγγελος

markakis@gmail.com

Περίληψη

- Συμμετρικά κρυπτοσυστήματα
- Block ciphers (κρυπτογράφηση τμήματος)
  - ✓ Substitution Ciphers (Κώδικες αντικατάστασης)
    - Μονοαλφαβητική αντικατάσταση
    - Πολυαλφαβητική αντικατάσταση
  - ✓ Transposition ciphers (Κώδικες αντιμετάθεσης)
- Stream ciphers (κρυπτογράφηση ροής)
  - ✓ Synchronous, asynchronous stream ciphers
  - ✓ One-time pad
- Κρυπτανάλυση συστημάτων

Κρυπτοσυστήματα

Ένα κρυπτοσύστημα καθορίζεται από μία πλειάδα (P, C, K, E, D), όπου:

P: Χώρος μηνύματος (plaintext ή message space), π.χ. Ελληνική γλώσσα ή αριθμητικά ψηφία

C: Χώρος κρυπτογραφήματος (ciphertext space), για απλότητα συνήθως P = C

K: Χώρος κλειδιών (key space)

E: Συναρτήσεις κρυπτογράφησης (encryption functions)

D: Συναρτήσεις αποκρυπτογράφησης (decryption functions)

Αρχή του Kerchoff: Η ασφάλεια ενός συστήματος δεν πρέπει να εξαρτάται από τη μυστικότητα του αλγορίθμου κρυπτογράφησης

- Θεωρούμε ότι ο αλγόριθμος είναι δημόσια γνωστός
- Το κλειδί πρέπει να διατηρείται μυστικό

Επιθυμητές ιδιότητες

1. **Σύγχυση (confusion):** Η ικανότητα του αλγορίθμου ώστε ο αντίπαλος να μην μπορεί να προβλέψει ποιες μεταβολές θα συμβούν στο ciphertext αν αλλάξουμε κάτι στο plaintext

2. **Διάχυση (diffusion)** Το να μπορεί μία μικρή αλλαγή σε ένα τμήμα του plaintext να αλλάζει όσο το δυνατόν περισσότερο το ciphertext

Μοντέλα αξιολόγησης ασφάλειας:

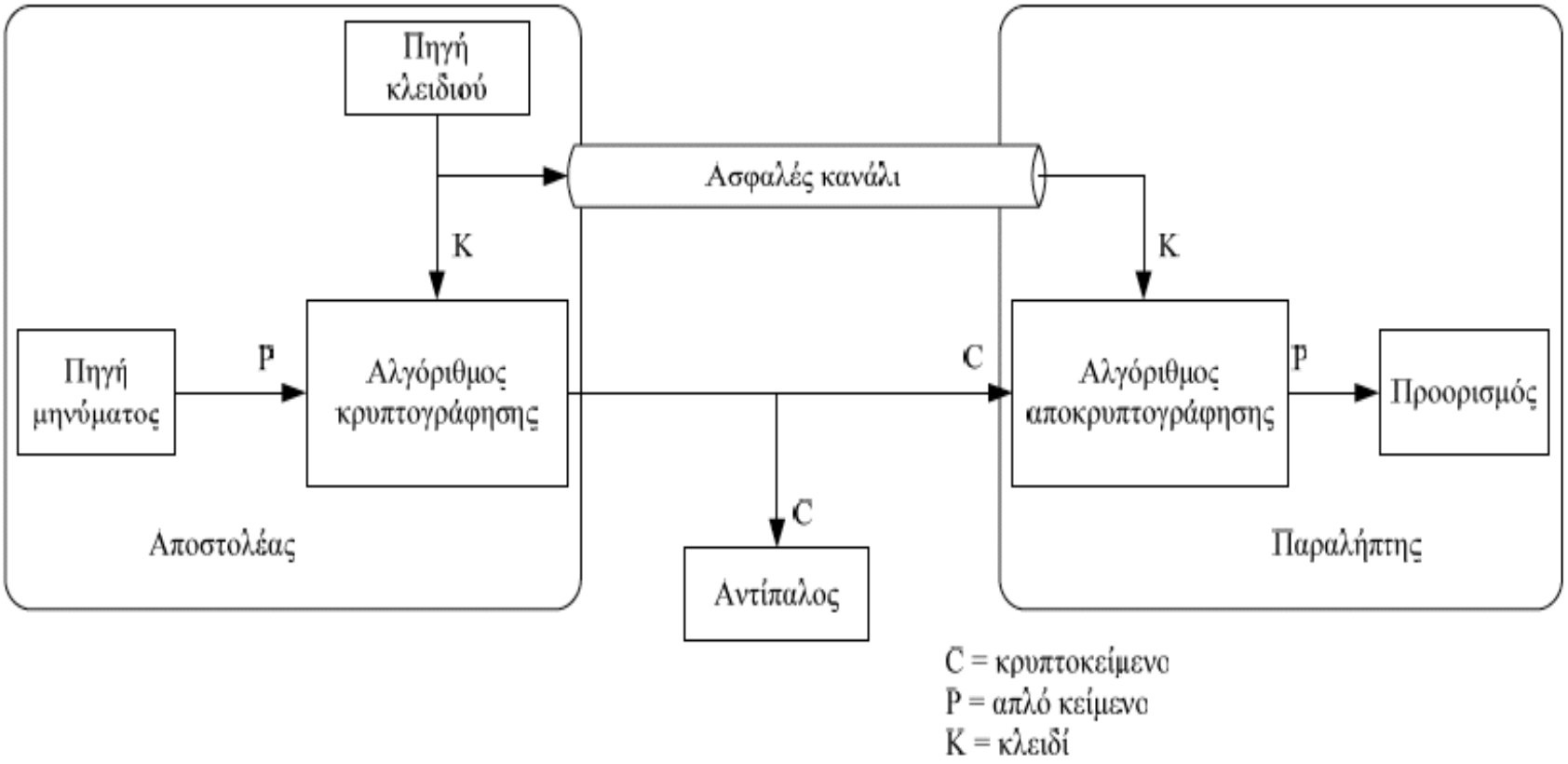
- Unconditionally secure
- Computationally secure
- Secure under complexity-theoretic assumptions

Συμμετρικά Κρυπτοσυστήματα

Στα συμμετρικά κρυπτοσυστήματα, για κάθε ζεύγος (e_k, d_k), είναι υπολογιστικά **εφικτό** να προσδιοριστεί το d_k γνωρίζοντας μόνο το e_k, ή να προσδιοριστεί το e_k από το d_k (αποστολέας και παραλήπτης χρησιμοποιούν το ίδιο κλειδί)

Αναφέρεται και ως *συμβατική*, *single-key* ή *private key* cryptography

Συμμετρική Κρυπτογραφία



Μέθοδος Διανομής Κλειδιών (key distribution):

- Κρίσιμο θέμα.
- Όλα τα μέρη γνωρίζουν το encryption scheme αν μάθουν και το κλειδί
- Κοινά συμφωνημένη μέθοδος για ασφαλή ανταλλαγή κλειδιών
- Θα δούμε σε επόμενες διαλέξεις κάποιες μεθόδους

Δύο κύριες κλάσεις από symmetric cryptosystems:

block ciphers → Κρυπτογράφηση τμήματος

stream ciphers → Κρυπτογράφηση ρεύματος

Κρυπτογράφηση τμήματος: το αρχικό κείμενο διασπάται σε *blocks* σταθερού μήκους, και κρυπτογραφείται ένα block τη φορά. Κάθε ciphertext είναι ίσου μήκους. Κατηγορίες κρυπτογραφικών πράξεων:

1. Αναδιάταξη (transposition): επιδρά στη θεση των συμβόλων του plaintext
2. Αντικατάσταση (substitution): επιδρά στα σύμβολα του plaintext
    1. Μονοαλφαβητική αντικατάσταση
    2. Πολυαλφαβητική αντικατάσταση

Κρυπτογράφηση ρεύματος: η συνάρτηση κρυπτογράφησης αλλάζει για κάθε σύμβολο του αρχικού κειμένου. Είναι *block cipher* με block length ίσο με μονάδα

Shift Cipher (ή κώδικας μετατόπισης)

Απεικονίζουμε το αλφάβητο με αριθμούς (A->0, ..., Z-> 25)

$P = C = K = Z_{26}$

Encryption: Για κλειδί $k \in K$, $e_k(x) = (x + k) \bmod 26$

Decryption: Για ciphertext y, $d_k(y) = (y - k) \bmod 26$

Παράδειγμα (έστω *k=3*)

THEHUNSARECOMING	(plaintext)
19 7 4 7 20 13 18 0 17 4 2 14 12 8 13 6	(αντιστοίχιση)
22 10 7 10 23 16 21 3 20 7 5 17 15 11 16 9	(μετατόπιση)
WKHKXQVDUHFRPLQJ	(ciphertext)

Shift Cipher (ή κώδικας μετατόπισης)

- Τύπου: substitution cipher, μονοαλφαβητική αντικατάσταση

- Κάθε γράμμα κρυπτογραφείται σε ένα μοναδικό γράμμα του αλφαβήτου

- Για k=3 αναφέρεται ως Caesar cipher

- Χρησιμοποιήθηκε από τον Ιούλιο Καίσαρα για να επικοινωνεί με τους στρατηγούς του

Shift Cipher (ή κώδικας μετατόπισης)

Κρυπτανάλυση του shift cipher:

Πόσο ασφαλές είναι το κρυπτοσύστημα?

Μειονεκτήματα:

1. Αριθμός διαφορετικών κλειδιών: 26
   - Ουσιαστικά 25 γιατί το k=0 δεν συνίσταται!
   - Υπερβολικά μικρός αριθμός! Μπορεί να σπάσει με brute force σε ciphertext-only attack

2. Μονοαλφαβητική αντικατάσταση: μπορεί κανείς γνωρίζοντας τη συχνότητα εμφάνισης κάθε γράμματος να μαντέψει το κείμενο

Affine Cipher (ή γραμμικός κώδικας)

Γενίκευση του shift cipher

Και πάλι τύπου: substitution cipher, μονοαλφαβητικής αντικατάστασης

Αντιστοίχισε αλφάβητο σε αριθμούς (A->0, ..., Z-> 25)

Encryption: Για κλειδί $k = (a,b)$, με $a, b \in Z_{26}$, $e_k(x) = (ax + b) \bmod 26$

Πώς θα αποκρυπτογραφήσουμε;

Θα πρέπει να λύσουμε την $y = ax + b \bmod 26 \Leftrightarrow ax = y - b \bmod 26$

1.Πρέπει δηλαδή να λύνουμε εξισώσεις της μορφής $ax = z \bmod 26$

2.Πρέπει να υπάρχει μοναδική λύση

3.Ισχύει μόνο αν gcd(a, 26) = 1

Πιο γενικά: Αν αντί για Z_{26} είχαμε Z_n

$P = C = Z_n$ $K = \{(a,b) : b \in Z_n, gcd(a,n) = 1\}$
Encryption: $e_k(x) = (ax + b) \bmod n$
Decryption: $d_k(y) = a^{-1}(b - y) \bmod n$

Affine Cipher (ή γραμμικός κώδικας)

Κρυπτανάλυση του affine cipher

Ίδια μειονεκτήματα με το shift cipher:

1. Αριθμός κλειδιών:
   - Στο Z_n υπάρχουν φ(n) τιμές του *a* έτσι ώστε *gcd(a,n)=1*.
   - Άρα συνολικά $n \cdot \varphi(n)$ διαφορετικά κλειδιά (εξακολουθεί να είναι πολύ μικρός αριθμός)

2. Ο γραμμικός κώδικας υπόκειται και αυτός στις αδυναμίες της μονοαλφαβητικής αντικατάστασης

Μονοαλφαβητική αντικατάσταση

Αν αυξήσουμε το χώρο των κλειδιών? π.χ. πιο πολύπλοκη συνάρτηση $e_k(x)$?

Δεν μας βοηθάει.
Αδυναμία της μονοαλφαβητικής αντικατάστασης: το κρυπτοκείμενο διατηρεί τις συχνότητες εμφάνισης των γραμμάτων

Στην ελληνική γλώσσα:

Γράμμα	Συχνότητα εμφάνισης (%)	Γράμμα	Συχνότητα εμφάνισης (%)
α	12	ν	7,9
β	0,8	ξ	0,6
γ	2	ο	9,8
δ	1,7	π	5,024
ε	8	ρ	5,009
ζ	0,5	σ	4,9
η	2,9	τ	9,1
θ	1,3	υ	4,3
ι	7,8	φ	1,2
κ	4,2	χ	1,4
λ	3,3	ψ	0,2
μ	4,4	ω	1,6

Πίνακας 3.2 Συχνότητα εμφάνισης των γραμμάτων της ελληνικής γλώσσας (Πηγή: Simon Singh, «Κώδικες και Μυστικά»).

Πιο συχνά γράμματα:
α, ο, τ, ε, ν, ι

Μονοαλφαβητική αντικατάσταση

Αν αυξήσουμε το χώρο των κλειδιών? π.χ. πιο πολύπλοκη συνάρτηση $e_k(x)$?

Δεν μας βοηθάει.
Αδυναμία της μονοαλφαβητικής αντικατάστασης: το κρυπτοκείμενο διατηρεί τις συχνότητες εμφάνισης των γραμμάτων

Στην αγγλική γλώσσα:

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

Πίνακας 3.3 Συχνότητα εμφάνισης των γραμμάτων της αγγλικής γλώσσας (Πηγή: Lewand, 2000)

Πιο συχνά γράμματα: e, t, a, o, i

Μονοαλφαβητική αντικατάσταση

Ακόμα και αν αυξήσουμε τον χώρο των κλειδιών, αν βρούμε σωστά τα πιο δημοφιλή γράμματα θα μπορέσουμε να βρούμε το κλειδί που χρησιμοποιήθηκε

Παραδείγματα:

1. Shift cipher: αρκεί να μαντέψουμε σωστά ένα γράμμα. Αν c είναι η κρυπτογράφηση του e, τότε $c = e + k \bmod 26$

2. Affine cipher: αρκεί να μαντέψουμε σωστά 2 γράμματα και να λύσουμε σύστημα 2 γραμμικών εξισώσεων mod n (mod 26)
Π.χ. Έστω ciphertext

```
FMXVEDKAPHFERBNDKRXRSREFMORUDSDKDVSHVUFEDKAPRKDL
YEVLRHHRH
```

Πιο συχνά γράμματα: R (8), D (7), E, H, K (5)

Μονοαλφαβητική αντικατάσταση

1η προσπάθεια: Αν υποθέσουμε ότι R ⇔ e και D ⇔ t, τότε
$e_k(4) = 17 \bmod 26$ και $e_k(19) = 3 \bmod 26$

Άρα θέλουμε να λύσουμε το σύστημα:
$4a + b = 17 \bmod 26$
$19a + b = 3 \bmod 26$
Λύση: *a = 6, b=19. Αλλά gcd(a, 26) = 2 >1, μαντέψαμε λάθος*

2η προσπάθεια: R ⇔ e και K ⇔ t,

Λύνοντας παίρνουμε *a = 3, b = 5, gcd(3, 26)=1*

⇒ υπολογίζουμε την d_k και αποκρυπτογραφούμε

⇒ αν το τελικό κείμενο βγάζει νόημα βρήκαμε το σωστό κλειδί

Αλλιώς συνεχίζουμε να μαντεύουμε…

Going beyond monoalphabetic substitution…

Polygram substitution

- Επεξεργάζεται και αντικαθιστά ομάδες χαρακτήρων (π.χ., δύο συνεχόμενα γράμματα, digrams ή trigrams ή n-grams)
- Για αλφάβητο Α των 26 χαρακτήρων υπάρχουν 26^2 «διγράμματα»
- Επομένως $(26^2)!$ Κλειδιά
- Καταστρέφεται έτσι η συχνότητα εμφάνισης κάθε γράμματος
- Όμως μπορούμε να μετρήσουμε συχνότητες από ομάδες συμβόλων (υπάρχουν στατιστικά στοιχεία για τα πιο συχνά δι-γράμματα και τρι-γράμματα)

Going beyond monoalphabetic substitution…

Homophonic substitution

• Για κάθε σύμβολο α του αλφαβήτου, συσχέτισε ένα σύνολο Η(α) από strings m συμβόλων, έτσι ώστε τα σύνολα Η(α) να είναι ξένα μεταξύ τους (για να είναι καλά ορισμένη η αποκρυπτογράφηση). Το κλειδί αποτελείται από όλα τα σύνολα Η(α)

• Για την κρυπτογράφηση: για κάθε εμφάνιση του συμβόλου α, μπορούμε να παίρνουμε τυχαία απόφαση για να διαλέξουμε ποιο string από το Η(α) θα χρησιμοποιήσουμε

• Για την αποκρυπτογράφηση: πρέπει για κάθε string c να προσδιοριστεί α τέτοιο ώστε $c \in H(\alpha)$

Going beyond monoalphabetic substitution…

Homophonic substitution – Παράδειγμα

Έστω $A = \{a, b\}$, m=2

Ορίζουμε $H(a) = \{00, 10\}$, και $H(b) = \{01, 11\}$, $H(a) \cap H(b) = \varnothing$

Το plaintext *ab* μπορεί να κρυπτογραφηθεί ως ένα από τα επόμενα: *0001, 0011, 1001, 1011*.

Το πεδίο ορισμού της decryption function (για μηνύματα 2 συμβόλων) αποτελείται από τα ακόλουθα ξένα σύνολα, κάθε ένα των τεσσάρων στοιχείων από 4-bitstreams

{0000, 0010, 1000, 1010} → aa

{0001, 0011, 1001, 1011} → ab

{0100, 0110, 1100, 1110} → ba

{0101, 0111, 1101, 1111} → bb

Μειονέκτημα: Αν οι πληθικοί αριθμοί |H(α)| είναι σχετικά μικροί, τότε θα υπάρχει αρκετή επανάληψη και θα μπορέσουμε να μετρήσουμε συχνότητες εμφάνισης γραμμάτων

Going beyond monoalphabetic substitution…

Πολυαλφαβητική αντικατάσταση

Κάθε σύμβολο μπορεί να κρυπτογραφείται σε περισσότερα σύμβολα

Vigenère cipher

Χωρίζουμε το κείμενο σε κομμάτια μεγέθους m, για κάποιο m > 0.

$P = C = K = (Z_{26})^m$. *Για κλειδί* $k = (k_1, k_2, \ldots, k_m)$:

Encryption: $e_k(x_1, \ldots, x_m) = (x_1+k_1 \bmod 26, x_2+k_2 \bmod 26, \ldots, x_m+k_m \bmod 26)$

Decryption: $d_k(y_1, \ldots, y_m) = (y_1-k_1 \bmod 26, y_2-k_2 \bmod 26, \ldots, y_m-k_m \bmod 26)$

Παράδειγμα: Έστω k = DHK = (3,7,10) και έστω το μήνυμα:

```
THISCIPHERISCERTAINLYNOTSECURE

  THI SCI PHE RIS CER TAI NLY NOT SEC URE
+ DHK DHK DHK DHK DHK DHK DHK DHK DHK DHK
= WOS VJS SOO UPC FLB WHS QSI QVD VLM XYO
```

16ος Αιώνας 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

tabula recta, Johannes Trithemius

Η ιστορία είναι άδικη απέναντι στον L. B. Alberti που ανακάλυψε πρώτος τη μέθοδο

- Πίνακας αντικατάστασης λατινικών χαρακτήρων
- Αναπαράσταση των μετατοπίσεων mod 26
  - Διαστάσεις 26x26
  - Κάθε γραμμή / στήλη ξεκινά απαρίθμηση γραμμάτων από το γράμμα που τις αντιστοιχεί
- Ο αποστολέας επιλέγει το plaintext
- Ο αποστολέας επιλέγει μυστική λέξη και παράγει ακολουθία ίδιου μήκους με το κείμενο
  - π.χ. Μυστική λέξη KEY οπότε η ακολουθία μπορεί να είναι KEYKEYKEYKEYKEYK
  - Εναλλακτικά μπορούμε να προσθέσουμε χαρακτήρες στο κείμενο ώστε να είναι πολλαπλάσιο του 3
- το κρυπτοκείμενο προκύπτει από το περιεχόμενο του πίνακα που τέμνει η γραμμή του κειμένου και η στήλη του κλειδιού

Κρυπτανάλυση του Vigenère Cipher

- Λίγο πιο δύσκολη από ότι στην μονοαλφαβητική αντικατάσταση
- Έστω ciphertext $c = c_1 c_2 \ldots c_s$, όπου το s αρκετά μεγάλο
- Ιδέα: Αν βρούμε το μήκος του κλειδιού m, τότε αναγόμαστε σε πολλές μονοαλφαβητικές αντικαταστάσεις στα υποσύνολα:
  - $y_1 = \{c_1, c_{m+1}, c_{2m+1}, \ldots\}$
  - $y_2 = \{c_2, c_{m+2}, c_{2m+2}, \ldots\}$
  - …
  - $y_m = \{c_m, c_{2m}, c_{3m}, \ldots\}$
- 2 γνωστές μέθοδοι
  - Μέθοδος Kasiski
  - Έλεγχος του δείκτη σύμπτωσης (index of coincidence)

Κρυπτανάλυση του Vigenère Cipher - Μέθοδος Kasiski (1863)

- Στηρίζεται στο γεγονός ότι επαναλαμβανόμενα μοτίβα θα τύχουν κρυπτογράφησης με το ίδιο τμήμα του κλειδιού πάνω από 1 φορά
- Στην ελληνική γλώσσα μοτίβα όπως «στο», «από», «ένα», «του» εμφανίζονται αρκετά συχνά
- Χρησιμοποιούμε συνήθως μοτίβα με τουλάχιστον 3 ή 4 χαρακτήρες που επαναλαμβάνονται τουλάχιστον 3 φορές
- Ο Oscar παρατηρεί κάθε μοτίβο και σημειώνει τις αποστάσεις από την 1η εμφάνιση
- Αν έχουν κρυπτογραφηθεί με το ίδιο τμήμα του κλειδιού, οι αποστάσεις πρέπει να είναι 0 (mod m).
- Άρα το μήκος του κλειδιού είναι διαιρέτης των αποστάσεων

Κρυπτανάλυση του Vigenère Cipher - Μέθοδος Kasiski (1863)

```
CHREEVOAHMAERATBIAXXWTNXBEEOPHBSBQMQEQERBW
RVXUOAKXAOSXXWEAHBWGJMMQMNKGRFVGXWTRZXWIAK
LXFPSKAUTEMNDCMGTSXMXBTUIADNGMGPSRELXNJELX
VRVPRTULHDNQWTWDTYGBPHXTFALJHASVBFXNGLLCHR
ZBWELEKMSJIKNBHWRJGNMGJSGLXFEYPHAGNRBIEQJT
AMRVLCRREMNDGLXRRIMGNSNRWCHRQHAEYEVTAQEBBI
PEEWEVKAKOEWADREMXMTBHHCHRTKDNVRZCHRCLQOHP
WQAIIWXNRMGWOIIFKEE
```

- Το μοτίβο CHR εμφανίζεται 5 φορές στις θέσεις 1, 166, 236, 276, 286
- Αποστάσεις από την 1η εμφάνιση: 165, 235, 275, 285
- gcd = 5
- Υποψήφιο μήκος κλειδιού m = 5
- Κάνουμε 5 μονοαλφαβητικές αποκρυπτογραφήσεις για να ελέγξουμε αν όντως m = 5

Κρυπτανάλυση του Vigenère Cipher - Index of Coincidence

- Για ένα string $x = x_1x_2\ldots x_n$ μίας γλώσσας έστω f_i ο αριθμός εμφανίσεων του i-οστού γράμματος, i=0,…,25.
- Ορισμός: Ο δείκτης σύμπτωσης (index of coincidence) $I_c(x)$ του x είναι η πιθανότητα 2 τυχαία γράμματα του x να συμπίπτουν

$$I_c(x) = \frac{\sum_{i=0}^{25} f_i(f_i - 1)}{n(n-1)}$$

- Έστω p_i οι συχνότητες εμφάνισης κάθε γράμματος στην αγγλική γλώσσα. Τότε για ένα σχετικά «τυχαίο» string x της αγγλικής γλώσσας περιμένουμε ότι

$$I_c(x) \approx \sum_{i=0}^{25} p_i^2 \approx 0.065$$

Κρυπτανάλυση του Vigenère Cipher - Index of Coincidence

- Για ένα τελείως τυχαίο string x, $I_c(x) = 0.038$
- Ιδέα: έστω ότι με τη μέθοδο Kasiski υποψιαζόμαστε ότι το κλειδί είναι μήκους m. Ο δείκτης σύμπτωσης μας βοηθάει να επιβεβαιώσουμε την ορθότητα της μεθόδου Kasiski
- Χωρίζουμε το ciphertext $c = c_1c_2c_3\ldots$ στις υποακολουθίες
  - $y_1 = c_1, c_{m+1}, c_{2m+1},\ldots$
  - $y_2 = c_2, c_{m+2}, c_{2m+2},\ldots$
  - …
  - $y_m = c_m, c_{2m}, c_{3m},\ldots$
  - Αν έχουμε μαντεψει το m σωστά, κάθε y_i είναι μία μονοαλφαβητική αντικατάσταση, άρα οι συχνότητες διατηρούνται και θα πρέπει να έχει δείκτη σύμπτωσης κοντά στο 0.065
- Υπολογίζουμε τους δείκτες $I_c(y_i)$, $i = 1,\ldots,m$
- Αν είναι όλοι κοντά στο 0.065, μάλλον έχουμε βρει το μήκος του κλειδιού

Κρυπτανάλυση του Vigenère Cipher - Index of Coincidence

```
CHREEVOAHMAERATBIAXXWTNXBEEOPHBSBQMQEQERBW
RVXUOAKXAOSXXWEAHBWGJMMQMNKGRFVGXWTRZXWIAK
LXFPSKAUTEMNDCMGTSXMXBTUIADNGMGPSRELXNJELX
VRVPRTULHDNQWTWDTYGBPHXTFALJHASVBFXNGLLCHR
ZBWELEKMSJIKNBHWRJGNMGJSGLXFEYPHAGNRBIEQJT
AMRVLCRREMNDGLXRRIMGNSNRWCHRQHAEYEVTAQEBBI
PEEWEVKAKOEWADREMXMTBHHCHRTKDNVRZCHRCLQOHP
WQAIIWXNRMGWOIIFKEE
```

- Με m=1, δείκτης = 0.045
- Με m=2, δείκτες = {0.046, 0.041}
- Με m=3, δείκτες = {0.043, 0.050, 0.047}
- Με m=4, δείκτες = {0.042, 0.039, 0.045, 0.040}
- Με m=5, δείκτες = {0.063, 0.068, 0.069, 0.061, 0.072}, άρα σταματάμε εδώ

Hill Cipher (*Hill 1929*)

Ιδέα: Κάθε σύμβολο του ciphertext εξαρτάται και από τα m σύμβολα του plaintext αν έχουμε blocks μεγέθους m, παίρνοντας γραμμικούς συνδυασμούς

Π.χ. με m=2 και $x = (x_1, x_2)$. Έστω

$y_1 = 11\, x_1 + 3\, x_2 \bmod 26$

$y_2 = 8\, x_1 + 7\, x_2 \bmod 26$

Πιο περιεκτικά: $(y_1, y_2) = (x_1, x_2) \begin{pmatrix} 11 & 8 \\ 3 & 7 \end{pmatrix} \pmod{26}$

Για γενικό m, κάθε y_i θα γίνεται $k_{1,i}\, x_1 + k_{2,i} x_2 + \ldots + k_{m,i} x_m$

Hill Cipher *(Hill 1929)*

Hill cipher

Χωρίζουμε το κείμενο σε κομμάτια μεγέθους m, για κάποιο m > 0.

$P = C = (Z_{26})^m$.

Χώρος κλειδιών = το σύνολο των αντιστρέψιμων m x m πινάκων mod 26. Αν K ένα τέτοιο κλειδί και $x = (x_1, \ldots, x_m)$:

Encryption: $e_k(x_1, \ldots, x_m) = xK$

Decryption: $d_k(y_1, \ldots, y_m) = yK^{-1}$

Προσοχή: ο πίνακας K πρέπει να επιλεγεί ώστε να είναι αντιστρέψιμος

- Στο R ο K είναι αντιστρέψιμος αν και μόνο αν det K ≠ 0
- Στο Z_{26} ο K είναι αντιστρέψιμος αν και μόνο αν gcd(det K, 26) = 1
  - Αν υπάρχει ο K^{-1} τότε det(I) = 1 = det K det K^{-1}. Άρα έχει λύση η εξίσωση det K x = 1 mod 26 ⇒ gcd(det K, 26) = 1

Κρυπτανάλυση του Hill cipher

- Είναι δύσκολο να σπάσει με ciphertext-only attack
- Άρα πιο ασφαλές κρυπτοσύστημα από όλα όσα έχουμε δει μέχρι τώρα
- Μπορεί όμως να σπάσει με known plaintext attacks
- Αν εχουμε m ζεύγη plaintexts-ciphertexts
  - Ανάγεται στην επίλυση συστήματος γραμμικών εξισώσεων

Κρυπτανάλυση του Hill cipher

- Έστω ότι ο Oscar έχει μαντέψει το m και έχει m ζεύγη plainexts - ciphertexts
- π.χ. $x_j = (x_{1,j}, \ldots, x_{m,j})$, $j=1,\ldots,m$
- και $y_j = (y_{1,j}, \ldots, y_{m,j})$, $j=1,\ldots,m$, με $y_j = e_K(x_j) \Rightarrow y_j = x_j K$
- Άρα τελικά έχουμε το σύστημα $Y = X \cdot K$, με $X = (x_{i,j})$, $Y = (y_{i,j})$
- Αν ο Oscar υπολογίσει τον αντίστροφο X^{-1}, μπορεί να βρει το κλειδί K
- Αν ο πίνακας X δεν είναι αντιστρέψιμος?
  - Τότε ο Oscar θα προσπαθήσει να βρει διαφορετικά m plaintexts ώστε ο X να είναι αντιστρέψιμος
- Αν ο Oscar δεν μάντεψε σωστά το m?
  - Θα φανεί αν δοκιμάσει κάποιο νέο ζεύγος plaintext-ciphertext
  - Μπορεί και να κάνει και brute force για την τιμή του m

Transposition ciphers – Κώδικες που βασίζονται σε λειτουργίες αναδιάταξης

Το κείμενο χωρίζεται σε τμήματα μήκους m

Σε κάθε block εφαρμόζεται μία απλή αντιμετάθεση π

Στην αποκρυπτογράφηση εφαρμόζεται απλή αντιμετάθεση $d=\pi^{-1}$ η οποία αντιστρέφει την π.

Permutation cipher (known since 1563 – Giovanni Porta)

$P = C = (Z_{26})^m$.

$K = \{\pi: \pi \text{ είναι permutation του } Z_m\}$

Encryption: $e_\pi(x_1,\ldots,x_m) = (x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(m)})$, όπου $\pi \in K$

Έστω $d = \pi^{-1}(\cdot)$

Decryption: $d_\pi(y_1,\ldots,y_m) = (y_{d(1)}, y_{d(2)}, \ldots, y_{d(m)})$

Transposition ciphers – Κώδικες που βασίζονται σε λειτουργίες αναδιάταξης

Παράδειγμα

Έστω permutation cipher με m=6 και αντιμετάθεση π = (6 4 1 3 5 2)

Το μήνυμα m=ΠΕΡΑΣΕ κωδικοποιείται ως c=ΕΑΠΡΣΕ

Η αποκωδικοποίηση γίνεται με την αντίστροφη αντιμετάθεση d=(3 6 4 2 5 1)

Είναι ειδική περίπτωση του Hill cipher

Ορισμός: Ένας πίνακας που σε κάθε γραμμή και κάθε στήλη έχει ακριβώς ένα 1 και τα υπόλοιπα στοιχεία 0 λέγεται permutation matrix

Έστω π το κλειδί ενός permutation cipher.

Αν θέσουμε $k_{i,j} = 1$, *αν* $i = \pi(j)$ και 0 διαφορετικά, τότε

Encryption με permutation cipher ⇔ encryption με Hill cipher και $K = (k_{i,j})$

Product Ciphers

Είναι η σύνθεση $t \geq 2$ μετασχηματισμών:

$$E_k^1 \circ E_k^2 \circ \ldots \circ E_k^t \qquad \text{για plaintext } x, \text{ ciphertext} = E_k^1 \ldots (E_k^{t-1}(E_k^t(x)))$$

- ✓ Συνήθως πιο ασφαλείς από τους απλούς κώδικες
- ✓ Χρησιμοποιούνται σχεδόν σε όλα τα σύγχρονα συμμετρικά κρυπτοσυστήματα, όπου κάθε E^i είναι είτε substitution είτε transposition cipher
- ✓ Πρακτικοί και αποδοτικοί

Παράδειγμα

Έστω $P=C=K$ το σύνολο όλων των binary string μήκους 6 ($|P|=64$), και $x=(x_1x_2...x_6)$

Έστω

$E_k^1(x) = x \oplus k$,όπου $k \in K$,

$E^2(x) = (x_4x_5x_6x_1x_2x_3)$

Εδώ ο E_k^1 είναι ένας polyalphabetic substitution cipher και ο E^2 είναι ένας transposition cipher

Το $E_k^1 \, o \, E_k^2$ είναι ένας product cipher

Κρυπτογράφηση ρεύματος

Μέχρι τώρα υποθέσαμε ότι κάθε σύμβολο από το plaintext κρυπτογραφείται με το ίδιο κλειδί

Ιδέα λειτουργίας ενός *stream cipher:*

1. δέχεται ως είσοδο το plaintext string $x_1x_2x_3...$
2. παράγει μία ακολουθία κλειδιών $z_1, z_2, z_3, ...$
3. το παραγόμενο ciphertext string είναι $c_1c_2c_3...$ όπου $c_i = E_{z_i}(x_i)$.

Η ακολουθία συμβόλων $z_1z_2z_3 ... z_i$, καλείται *keystream*

- ✓ Synchronous stream cipher: τα z_i δεν εξαρτώνται από το plaintext
- ✓ Asynchronous stream cipher: μπορεί να υπάρχει εξάρτηση από τα σύμβολα του plaintext
- ✓ Periodic stream cipher: όταν για κάποιο d, $z_{i+d} = z_i$
- ✓ π.χ. Ο Vigenere cipher είναι περιοδικός με περίοδο ίση με το μήκος του κλειδιού

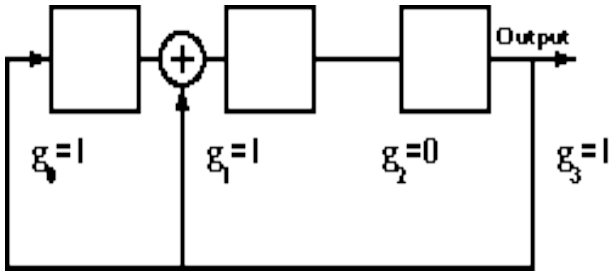
Παράδειγμα synchronous stream cipher: Linear Feedback Shift Register

- ✓ Συνήθως σε stream ciphers δουλεύουμε με δυαδικό αλφάβητο
- ✓ Encryption/decryption αντιστοιχούν σε λειτουργίες XOR:
  - ✓ $e_z(x) = (x + z) \bmod 2$
  - ✓ $d_z(y) = e_z(y) = (y + z) \bmod 2$
- ✓ Keystream: επιλέγουμε πρώτα ένα κλειδί $K = (k_1, \ldots, k_m, c_0, \ldots, c_{m-1})$, $c_i \in Z_2$
- ✓ Αρχικοποιούμε $z_i = k_i$, $i = 1, \ldots, m$, και κατόπιν:

$$z_{i+m} = \sum_{j=0}^{m-1} c_j z_{i+j} \bmod 2$$

Παράδειγμα synchronous stream cipher: Linear Feedback Shift Register

- ✓ Γραμμική αναδρομική εξίσωση
- ✓ Αν είμαστε προσεκτικοί θα έχουμε μεγάλη περίοδο (2^m-1)
- ✓ Μπορεί να υλοποιηθεί εύκολα σε hardware
  - ✓ Shift και XOR operations



- ✓ Κρυπτανάλυση: Σχετικά ασφαλές σε ciphertext-only attack
- ✓ Λόγω της γραμμικής αναδρομής, μπορεί να αναχθεί σε επίλυση συστήματος γραμμικών εξισώσεων σε known plaintext attack

Παράδειγμα asynchronous stream cipher: Autokey cipher

✓ Προτάθηκε από τον Vigenere

✓ Είναι shift cipher αλλά κάθε z_i γίνεται ίσο με το προηγούμενο σύμβολο του plaintext

Autokey cipher

$P = C = K = Z_{26}$

Για κλειδί $k \in K$, ορίζουμε $z_1 = k$, $z_i = x_{i-1}$

Encryption: $e_z(x) = (x + z) \bmod 26$

Decryption: Για ciphertext y, $d_z(y) = (y - z) \bmod 26$

Κρυπτανάλυση: εύκολη λόγω του χαμηλού αριθμού διαφορετικών κλειδιών

Vernam cipher/One-time Pad (σημειωματάριο μίας χρήσης)

- ✓ Προτάθηκε από τον Vernam (1917)
- ✓ Ιδέα: το κλειδί είναι ίσο με το μήκος του plaintext και δεν ξαναχρησιμοποιούμε *ποτέ ξανά* το ίδιο κλειδί
- ✓ Μπορούμε να θεωρήσουμε ότι έχουμε ένα σημειωματάριο όπου κάθε σελίδα περιέχει μία ακολουθία αριθμών (κλειδί)
- ✓ Διανομή κλειδιών: παράγονται αρχικά 2 αντίγραφα του σημειωματάριου και μοιράζονται σε Alice/Bob
- ✓ Encryption: Η Alice επιλέγει τυχαία μία σελίδα, και κρυπτογραφεί με κώδικα Vigenere και με βάση το κλειδί της σελίδας
- ✓ Στέλνει στον Bob ciphertext και αριθμό σελίδας, και σκίζει τη σελίδα από το σημειωματάριο της για να μην ξαναχρησιμοποιηθεί

Vernam cipher/One-time Pad (σημειωματάριο μίας χρήσης)

Πιο αποδοτική υλοποίηση:

Vernam cipher

$P = C = K = (Z_2)^m = \{0, 1\}^m$

Κλειδί: τυχαίο binary m-bit string $k = (k_1, k_2, ..., k_m)$

Encryption: $e_k(x) = (x_1 \oplus k_1, x_2 \oplus k_2, ..., x_m \oplus k_m)$ (πρόσθεση mod 2)

Decryption: $d_k(y) = (y_1 \oplus k_1, y_2 \oplus k_2, ..., y_m \oplus k_m)$

✓ U.S. Patent 1,310,719, (1919), 1η υλοποίηση: ηλεκτρολογική με χρήση διάτρητων καρτών και teletype machines

✓ Στην αρχική υλοποίηση τα κλειδιά μπορούσαν να επαναληφθούν

✓ J. Mauborgne: Ιδέα για τυχαία κλειδιά

✓ Όταν το κλειδί επιλέγεται τυχαία και δεν ξαναχρησιμοποιείται, το one-time pad μπορεί να δειχθεί ότι είναι θεωρητικά unbreakable

Vernam cipher/One-time Pad (σημειωματάριο μίας χρήσης)

Πιο αποδοτική υλοποίηση:

Vernam cipher

$P = C = K = (Z_2)^m = \{0, 1\}^m$

Κλειδί: τυχαίο binary m-bit string $k = (k_1, k_2, ..., k_m)$

Encryption: $e_k(x) = (x_1 \oplus k_1, x_2 \oplus k_2, ..., x_m \oplus k_m)$ (πρόσθεση mod 2)

Decryption: $d_k(y) = (y_1 \oplus k_1, y_2 \oplus k_2, ..., y_m \oplus k_m)$

- ✓ Χρησιμοποιήθηκε ευρέως κατά το Β' παγκόσμιο πόλεμο, αλλά και αργότερα
- ✓ Υπάρχουν ενδείξεις για χρήση από: KGB, NSA, FBI, Guevara-Castro,…
- ✓ Πρακτικά ανεφάρμοστο στα σύγχρονα συστήματα
  - ✓ Πολύ μεγάλο μήκος κλειδιού
  - ✓ Προβληματική η διανομή κλειδιών (πρέπει σε κάθε session να υπάρχει συγχρονισμός για το κλειδί)
  - ✓ Δύσκολη η δημιουργία μεγάλων ακολουθιών από τέλεια τυχαίους αριθμούς

Block ciphers vs Stream ciphers

Μειονεκτήματα των stream ciphers

✓ Κάθε σύμβολο κρυπτογραφείται χωριστά. Η πληροφορία βρίσκεται σε ένα μόνο σύμβολο του κρυπτοκειμένου ⇒ χαμηλή διάχυση

✓ Συγχρονισμός των 2 γεννητριών μεταξύ Alice/Bob για το keystream

Πλεονεκτήματα των stream ciphers

✓ Ταχύτητα κρυπτογράφησης, εφαρμογές σε κρυπτογράφηση τηλεφωνικών συνδιαλέξεων και γενικότερα σε δεδομένα τηλεσυνδυάσκεψης

✓ χαμηλή διάχυση ⇒ λάθη μετάδοσης σε κάποια σύμβολα δεν επηρεάζουν το υπόλοιπο ciphertext