INF247

Introduction to Cryptanalysis of Symmetric Ciphers

via Teams by the end of January

after combined via Teams and ordinary classes at UiB.

5 mandatory ass.

written exam in a form of assignment.

---

course prerequisites

1) INF247 not first in Cryptology

//

Cryptography +	Cryptanalysis
how to protect data with math. tool like ciphers, digital sign. etc... (AES, RSA, ..)	how to find weaknesses in those tools we use AES because no weaknesses were found

2) Cryptanalysis is mathematically based.

Number Theory : arithmetic with residues

Algebra : polynomial arithmetic, matrices, how to solve systems of linear equations

Probability Theory : conditional probability, random variables, expectations, var. binomial destribution

normal distribution.

3) common sense.

Short Intro to Cryptography

A alphabet

Latin $|A| = 26$

A B C D ... X Y Z

binary $|A| = 2$

0, 1

Latin character may be written binary by 5-bit strings as $26 < 2^5 = 32$

A B C ...

00000 00001 00010 ...

$32 - 26 = 6$ ...

ASCII code 7-bit (8-bit)

Encryption Algorithm (Cipher)

M message space, set of all possible strings over A of finite length.

C cipher-text space, set of strings over some other alphabet A' finite length. (not every string over A' is a valid cipher-text)

K key space, binary strings of a fixed length.

e.g. AES-128

key-space all 128-bit strings

$|K| = 2^{128}$.

$E_e : M \to C$

↑ encryption key $\in K$

$D_d : C \to M$

↑ decryption key $\in K$

$D_d(E_e(m)) = m$ ←

for any $e \in K$ there exists $d \in K$ s.t.
for any $m \in M$.

---

the keys should be kept secret.

1) $e = d$ or d is easy to compute from e.
encryption is called symmetric.
e.g. AES

2) $e \neq d$ and d is difficult to compute given e
asymmetric encryption
e public (not secret)
d secret (private)
RSA.

Two parties Alice, Bob
choose encryption/decryption algorithms

define their symmetric key $k = e = d$

---

important practical problem how distribute k.

solved by asymmetric cryptography or DH protocol.

assume that symmetric k is centrally distributed by some authority.

---

symmetric cipher key system.

long-term keys (changed monthly)
daily keys
message keys

also encryption algorithm (cipher) was kept secret.

now commercial cipher as AES are public

secret cipher key (128-bit string)
public IV (initial value)
↑ plays role of message key.

Why do we need change keys?

E_k encryption function

$m_1, \ldots, m_N$ messages ← adversary knows

$c_1, \ldots, c_N$ cipher-texts ↓

his goal to recover cipher key k.

$C_{N+1} = E_{(k)}(m_{N+1})$

C_{N+1}: known; m_{N+1}: unknown

1) $\begin{cases} E_k(m_1) = c_1 \\ \vdots \\ E_k(m_N) = c_N \end{cases}$

N equations in one variable k.

→ k may leak through side channels.

3) k may be compromised (sold by an operator)

⇒ cipher key has to be changed regularly.

Two classes of symmetric ciphers:

1) block ciphers

2) stream ciphers.

Block ciphers

$m = m_1 \; m_2 \; \cdots \; m_N$

blocks of characters over A of fixed length t (block size)

e.g. AES-128 $A = \{0,1\}$ $t = 128$.

$C = c_1 \; c_2 \; \cdots \; c_N$

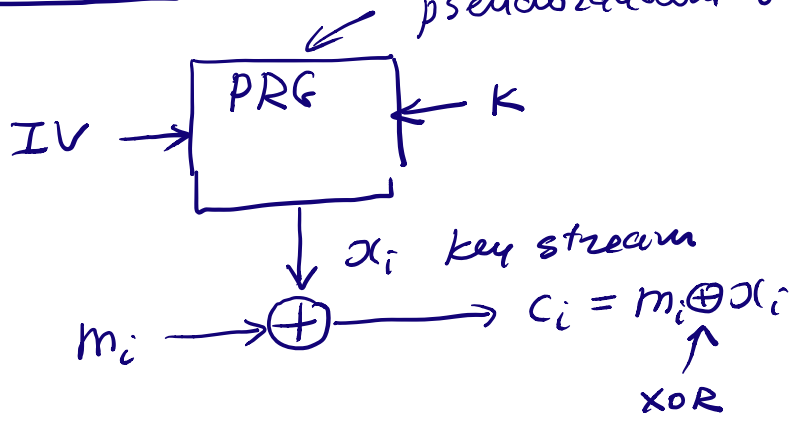
blocks

$$c_i = E_k(m_i)$$

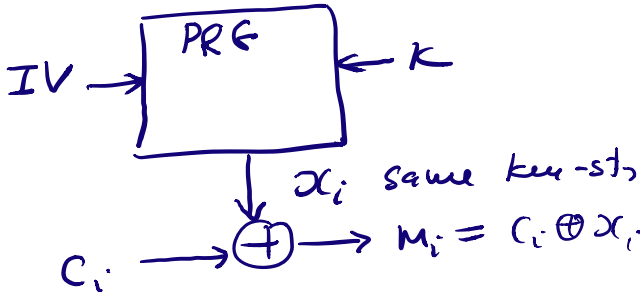
Stream ciphers

$$c_i = E_{k,i}(m_i)$$

Example. Binary stream cipher



IV is sent in clear (without encryption) before cipher-text



Block ciphers may work as stream ciphers.

CBC block cipher mode

(Cipher Block Chaining)

$$m = m_1 \; m_2 \; \cdots \; m_N$$

↑ blocks of bits

$$C = C_0 \; C_1 \; \cdots \; C_N$$

C_0 public IV

$$C_i = E_K(m_i \oplus C_{i-1}) \quad i = 1, \ldots, N.$$

Substitution Cipher.

on Latin alphabet A

$M = m_1 m_2 \cdots m_N$

↑ alphabet characters

block size is 1.

key substitution on A (permutation on A)

S

$C = c_1 c_2 \cdots c_N$

$c_i = S(m_i)$

Example.

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
x	n	y	a	R	p	o	g	z	q	w	B	t	s	p	l	z	c	v	m	u	e	k	j	d	i

pl-text

m = THE DATA ENCRYPTION STANDARD

c = m g k a x m x R s y c d t m z p s r m x s q x c a

↑ cipher-text.

cryptanalysis: given cipher-text, recover pl-text.

brute force (exhaustive search)

over all substitution S on A

compute $S^{-1}(C) \stackrel{?}{=}$ sensible English text

$\Rightarrow S$ is cipher key.

$|A| = 26$ # subst. on A is $26! \approx 4 \cdot 10^{26}$

How to solve the problem faster.

$\alpha \in A$ f_α frequency of α in the cipher-text

$$\left(\frac{\#\text{ appearances of } \alpha}{\text{text length}} \right)$$

F_α frequency of α in the plain-text.

according to encryption algorithm:

$$F_\alpha = f_{S(\alpha)}$$

$S(\alpha)$: encryption of α.

$$F_T = f_{m=S(T)} = \frac{4}{25} = 0.16$$

English language constants:

q_α frequency of α in long English texts:

$q_E \approx 0.1231\ldots$

$q_T \approx 0.0959\ldots$

$q_A \approx 0.0805\ldots$

...

$f_{S(\alpha)} = F_\alpha \approx q_\alpha$ if pl-text is long enough.

$\Downarrow$

recover S by comparing f_β, q_α

... the Lecture Notes (at Mitt)

cipher-text produced with subst. cipher
of length 290 characters:

THERE ARET R
xlivi evixq somzR wsjqv ...

Find char. frequencies in the cipher-text

i	0.148	← E	$q_E \approx 0.1231$
x	0.11	← T	$q_T \approx 0.0959$
w	0.107		
...			

THERE IS
ARE

use digrams

TH	0.0315
HE	0.0251
- - -	

tree search: