SEGURETAT I PROTECCIÓ DE DADES (GEINF-GDDV-GEB, UdG)

CURS 24-25

ACTIVITAT 1 – DESENVOLUPAMENT D'UN CRIPTOSISTEMA HÍBRID ITERATIU AMB SUBSTITUCIÓ POLIALFABÈTICA + TRANSPOSICIÓ

PES: 10% DEL TOTAL DE L'ASSIGNATURA

DATA LÍMIT ENVIAMENT → dissabte 12 d'octubre, mitjanit (lliurament via Moodle).

DESCRIPCIÓ DEL CRIPTOSISTEMA

PROCEDIMENT DE SUBSTITUCIÓ POLIALFABÈTICA

Es proposa un sistema **basat en el xifrat afí** (vist a classe), on **F(x) = (ax + c) mod L** , amb 0<a<L, 0<=c<L, a i L coprimers, sent L la mida de l'alfabet en clar (usarem l'alfabet anglès on L=26).

a = *serà el nombre entre 2 i L-1 més proper a la llargada de la paraula a encriptar que sigui coprimer a L*

Exemples per a diferents paraules del missatge en clar (suposant L=26):

- **alhambra** : llargada = 8 -> **a =7**, ja que és el coprimer a L més proper a 8 (9 també ho és, però en cas d'ambigüitat triarem el més petit)
- **peach**: llargada = 5 -> **a =5**, ja que 5 és coprimer a L
- **i**: llargada = 1 -> **a =1**, ja que 1 és coprimer a L
- **aviron**: llargada = 6 -> **a =5**, ja que és el coprimer a L més proper a 6 (7 també ho és, però en cas d'ambigüitat triarem el més petit)

c = *codi corresponent a la primera lletra de la paraula, suposant a=0, b=1...*

Exemples:

- **alhambra** → c =0
- **peach** → c =15
- **i** → c =8
- **aviron** → c =0

Notem que, tal i com està definit, aquest sistema, basat en el xifrat afí, és polialfabètic i admet la funció de desxifrat. **Anomenarem X a aquesta funció de xifrat.**

PROCEDIMENT DE TRANSPOSICIÓ

Correspondrà a un procediment força senzill que descriurem tot seguit.

Disposarem tot el missatge a transposar, horitzontalment, en una matriu amb un nombre determinat de columnes (anomenarem K a aquest nombre). Proporcionarem una permutació P de K. La sortida de la transposició correspondrà a l'escriptura del text per columnes en l'ordre donat per la permutació P. Vegem-ne un exemple:

Permutació P= ***{ 3, 2, 5, 1, 4}*** (per tant, K=5)

Missatge en clar: *y entre tantas falsedades muchas de mis mentiras ya son verdades*

1	2	3	4	5
y		e	n	t
r	e		t	a
n	t	a	s	
f	a	l	s	e
d	a	d	e	s
	m	u	c	h
a	s		d	e
	m	i	s	
m	e	n	t	i
r	a	s		y
a		s	o	n
	v	e	r	d
a	d	e	s	

Missatge transposat segons permutació: *e aldu inssee etaamsmea vdta eshe iynd yrnfd a mra antssecdst ors*

Anomenarem **T_P a aquesta funció de transposició** (sent P la permutació)

CRIPTOSISTEMA ITERATIU

El criptosistema que haureu de desenvolupar estarà basat en la concatenació dels 2 procediments que acabem de descriure. Aquesta concatenació es realitzarà un cert nombre N de vegades (N >=1). Així, podem definir f(m), sent m un cert missatge, com a:

$f_P(m) = T_P(X(m))$ (primer xifrem per substitució i després transposem el resultat del xifrat)

Llavors definim l'algorisme del criptosistema com a **$F_{P,N}(m) = f_P\text{^}N(m)$**
(és a dir, apliquem N vegades el xifrat polialfabètic + la transposició)

IMPORTANT: per fer-ho una mica més interessant, **aplicarem, a cada iteració, un shift a l'esquerra a la permutació**. A l'exemple:

- 1a iteració. Permutació= { 3, 2, 5, 1, 4}
- 2a iteració. Permutació= { 2, 5, 1, 4, 3}
- 3a iteració. Permutació= { 5, 1, 4, 3, 2}
- ...

CONSIDERACIONS

- Usarem fitxers de text codificats en UTF-8.
- L'alfabet constarà de L= 26 caràcters (a..z).
- Es consideraran equivalents les majúscules i les minúscules.
- Les lletres amb accents gràfics (accent habitual, dièresi, circumflex, ...) s'assimilaran a les corresponents lletres sense aquests accents.
- La ç s'assimilarà a la c, la ñ a la n i la "beta" que apareix en textos en alemany, al dígraf "ss"
- Es recomana realitzar un preprocés on es modifiqui el text original aplicant-hi els canvis indicats als punts anteriors
- Els textos contindran separadors de paraules (espai en blanc, salt de línia, signes de puntuació, apòstrof, ...). Aquests caràcters no es xifraran a l'algorisme de substitució

- Les xifres del 0 al 9 no són separadors, però tampoc seran xifrades a la substitució
- La transposició tindrà en compte absolutament tots els caràcters

FEINA A FER: PRIMERA PART

Heu de **crear un criptosistema basat a la descripció donada**. Aquest criptosistema constarà de:

a) Programa de **xifrat per substitució** (amb el corresponent desxifrat)
b) Programa de **xifrat per transposició** (amb el corresponent desxifrat)
c) **Programa principal corresponent al criptosistema**, que permeti xifrar i desxifrar

Els **ínputs** del programa principal seran:

- Nom del fitxer que contingui el missatge en clar (o xifrat si es tracta del procés de xifrat)
- Nom del fitxer que contindrà el missatge xifrat (o desxifrat si es tracta del procés de desxifrat)
- Indicació de si es vol xifrar o desxifrar
- **Nombre N d'iteracions**
- **Permutació inicial P** (que implícitament ens dona el valor K)

Tot plegat ho heu de programar preferentment en llenguatge **Python** (tot i que s'hi admeten altres llenguatges).

FEINA A FER: SEGONA PART

Cal que realitzeu **un ampli joc de proves.** Caldrà que hi apareguin textos de 3 mides:

- Un text **curt**: amb unes poques paraules
- Un text **mitjà**: d'una pàgina aproximadament
- Un text **llarg**: corresponent a un llibre, una obra literària...

Per a les proves (especialment la segona i la tercera), podeu baixar-vos textos en diferents idiomes, de forma totalment gratuïta i legal, i en format txt, a la web www.gutenberg.org

Cada prova ha de partir d'un text en clar, ha d'obtenir el seu xifrat i, a partir d'aquest xifrat, ha d'obtenir el seu corresponent desxifrat.

Es demana que useu un mínim de 3 idiomes diferents per al text llarg.

UNA PROVA OBLIGATÒRIA

Demanem que feu el xifrat i desxifrat (i adjunteu els resultats a la memòria) d'un text en català de llargada mitjana que teniu al Moodle. Cal que hi useu

- **N = 5** iteracions
- **P = {4, 2, 3, 1}**

FEINA A FER: TERCERA PART

PENDENT

FEINA A FER: QUARTA PART

Heu de fer un document, a mode de memòria d'aquesta activitat, al qual cal que hi figuri:

- Una descripció breu de com s'ha implementat tot plegat
- Una descripció detallada de tots els textos de prova utilitzats. No cal incloure els mitjans ni els llargs, però sí els curts (text en clar, xifrat, i desxifrat). Dels mitjans i els llargs només cal que citeu els que heu usat, indicant títol, idioma, nombre de caràcters i paraules, si ha funcionat bé el vostre algoritme, ...
- Una comparativa dels índexs de coincidència obtinguts per a cada text de prova usat, tant per al seu text en clar com per al seu text xifrat. Cal comparar els valors obtinguts i raonar els resultats en les següents dimensions (usant els coneixements assolits a la classe de teoria):
    - Entre textos en clar i els seus corresponents xifrats (després de tot el procés de substitucions i transposicions)
    - Entre un text en clar curt, un de mitjà i un de llarg dins el mateix idioma
    - Entre els textos llargs dels diferents idiomes usats

QUÈ CAL LLIURAR AL MOODLE ?

- Codi de les aplicacions
- Memòria (en format pdf)
- 2 fitxers txt, corresponents al text en català penjat al Moodle un cop xifrat i al mateix text un cop desxifrat (usant els paràmetres N i P indicats en aquest enunciat)

QUÈ ES VALORARÀ ?

- Correctesa de la **implementació** dels diferents programes.
- Completesa i correctesa del càlcul de **l'índex de coincidència**
- Completesa i correctesa de la **memòria**, en especial de les conclusions a les quals haureu arribat a partir de les proves i els resultats obtinguts pel que fa a l'índex de coincidència

CAL TENIR EN COMPTE

Es tracta d'una pràctica individual. Qualsevol còpia detectada suposarà una qualificació de zero en aquesta activitat.