Information security

Eng.taghreed harfoush

Frequency of English Letters (%)

Frequency	Letter	Frequency	Letter	Frequency	Letter
3.06	C	1.54	B	8.04	A
2.30	F	12.51	E	3.99	D
7.26	I	5.49	H	1.96	G
4.14	L	0.67	K	0.16	J
7.60	O	7.09	N	2.53	M
6.12	R	0.11	Q	2.00	P
2.71	U	9.25	T	6.54	S
0.19	X	1.92	W	0.99	V
		0.09	Z	1.73	Y

The Most Frequent English Digrams per 200 Letters

Digram	Frequency	Digram	Frequency	Digram	Frequency
TH	50	AT	25	ST	20
ER	40	EN	25	IO	18
ON	39	ES	25	LE	18
AN	38	OF	25	IS	17
RE	36	OR	25	OU	17
HE	33	NT	24	AR	16
IN	31	EA	22	AS	16
ED	30	TI	22	DE	16
ND	30	TO	22	RT	16
HA	26	IT	20	VE	16

Ref: http://www.fortunecity.com/skyscraper/coding/379/lesson1.htm

The Most Frequent English Trigrams per 200 Letters

Trigram	Freq	Trigram	Frequ	Trigram	Freq
THE	89	TIO	33	EDT	27
AND	54	FOR	33	TIS	25
THA	47	NDE	31	OFT	23
ENT	39	HAS	28	STH	21
ION	36	NCE	27	MEN	20

Ref: http://www.fortunecity.com/skyscraper/coding/379/lesson1.htm

Monoalphabetic Cryptanalysis

- Step1: count the occurrences of each letter in the cipher text (calculate the frequency of the letters)
- خطوة ١: أعد تكرار حدوث كل حرف بالنص المشفر ومن ثم أحسب نسبة تكرار هذا الحرف وأقارنها مع الإحصائيات الموجودة باللغة الانكليزية مثلاً أكثر شيء مكرر في النص المشفر e فعندها نعرف أن هذا الحرف يقابله a في النص الأصلي.
- Step2: Match against the statistics of English
  - Letters frequencies
    - High: E T A O N I R S H
    - Medium: D L U C M
    - Low: P F Y W G B V
    - Rare: J K Q X Z
  - Most frequent digrams
  - Most frequent trigrams

- خطوة ٢: أبدل الرمز بما وجدته من خلال الإحصائيات.
- تكرارات الأحرف :
- الأعلى : $E\ T\ A\ O\ N\ I\ R\ S\ H$
- المتوسط: $D\ L\ U\ C\ M$
- المنخفض: $P\ F\ Y\ W\ G\ B\ V$
- النادر: $J\ K\ Q\ X\ Z$
- أكثر ثنائيات مكررة: $TH\ \ ER\ \ ON$
- أكثر ثلاثيات مكررة: $THE\ \ AND$
-

- مثال: قم بكسر الشيفرة التالية : fqjcb rwjwj vnjax bnkhj whxcq nawjv nfxdu mbvnu ujbbf nnc
- الحل :
- في البداية أول خطوة هي معرفة كم مرة تكرر كل حرف ، نستطيع أن نعرف ما هو الحرف الذي تكرر أكثر من غيره، ومنه قد يكون هو الحرف E .
- الآن نبدأ بعد الحروف .....
- نعود إلى الحرف الأكثر تكراراً وهي j و n حيث تكرر كل منهم ٧ مرات.
- إن الفرق بين الحرف الأول j والحرف E هو ٥ ، وبالتالي المفتاح هو ٥ ، ونبدأ بفك الشيفرة:
- نطرح من كل حرف ٥ حروف :والناتج يكون alexw mrere ajevs نتوقف هنا بالطبع ، لأن النص ليس له معنى بتاتاً. نأخذ الحرف الثاني تكراراً ، وهو الحرف N .
- الفرق بينه وبين E هو ٩ ، إذاً المفتاح في هذه الحالة يكون ٩.
- الآن نطرح من كل حرف ٩ حروف .....
- وهكذا ليخرج لدينا : whats inana mearo sebya nyoth ernam ewoul dsmel lassw eet
- وبعد ترتيب الكلمات نبدأ بمعرفة الجمل فيكون لدينا :
- what's in a name a rose by any other name would smell as sweet

خوارزميات Polyalphabetic Ciphers :

Vigenère Cipher

- M = {m1,m2,m3 ...}
- K = {k1, k2,k3, ...,kt}
- C = {c1,c2,c3,...}
  - ci = mi + ki mod 26
  - The index I in ki is taken modulo t.

- في هذه الخوارزمية يكون المفتاح أكثر من رمز .

- write the plaintext out
- write the keyword repeated above it
- use each key letter as a Caesar cipher key
- encrypt the corresponding plaintext letter

Example:

keyword *deceptive*

key: deceptivedeceptivedeceptive

plaintext: wearediscoveredsaveyourself

ciphertext:ZICVTWQNGRZGVTWAVZHCQYGLMGJ

مثال:

- من أجل المفتاح deceptive

- key: deceptive deceptive deceptive
- plaintext: wearedisc overedsav eyourself
- ciphertext: ZICVTWQNG RZGVTWAVZ HCQYGLMGJ

- ملاحظة : بتقسيم شيفرة Vigenere إلى مجموعات حسب طول المفتاح يكون لأول عنصر من كل مجموعة الإزاحة نفسها وهكذا فالعنصر الثاني .....
- أمن خوارزمية فيجينير:
- هناك عدة نصوص مشفرة من أجل كل نص واضح ولهذا السبب فان تكرارات الأحرف غامضة، ولكن ليس بشكل كلي.

كسر الخوارزمية

- الفكرة:
- أعرف طول المفتاح
- أقسم الشيفرة إلى مجموعات حسب طول المفتاح .
- ملاحظة:
- كيف أعرف أن طول المفتاح صحيح ؟؟
- نرى المقاطع المكررة كم أبعادها عن بعضها البعض ثم نأخذ القاسم المشترك الأكبر لها.
- في مثالنا يبدأ الأول بـ ٣ والثاني بـ١٢ طول المفتاح ١٢-٣=٩

الآلية:

- خطوة ١: تحديد m طول كلمة المفتاح keyword ، لدينا طريقتين :
  - اختبار Kasiski test: تقنية تهدف إلى إيجاد طول المفتاح المستعمل في التشفير "القاسم المشترك الأكبر للأبعاد بين المقاطع المتشابهة".
  - و معيار التصادف index of coincidence : I_c تستخدم للتأكد أن طول المفتاح الذي توصلنا إليه صحيح.
- أقسم الرسالة إلى سلاسل حسب طول المفتاح وأحسب I_cلكل سلسلة فإذا كان قريباً من النسبة الموضوعة للغة الانكليزية فقد توصلنا إلى طول مفتاح صحيح .
- خطوة ٢: استخدام التحليل الإحصائي Frequency Analysis أو Chi square .

Kasiski Test

- تحديد m طول المفتاح كما يلي :
- المراقبة : كل مقطعين متطابقين من النص الواضح Plaintext سيتم تشفيرهم إلى نفس النص المشفر Ciphertext أينما ظهروا .
- δ مواقع البداية للمقاطع المتشابهة حيث $\delta \mod m = 0$ وبالعكس.
- البحث: البحث عن النص المشفر .
- تسجيل المسافة بين مواضع بداياتهم مثل $\delta 1, \delta 2, \ldots,$ عندها m يجب أن تقسم كل الـ $\delta i's$ القاسم المشترك الأكبر gcd من أجل كل الـ $\delta i's$.

- مثال على اختبار Kassiski :

- Ciphertext:
- CHRGQPWOEIRULYANDOSHCHRIZKEBUSNOFKYWR OPD
- CHRKGAXBNRHROAKERBKSCHRIWK
  - السلاسل الجزئية "CHR" موجودة في المواقع ٠، ٢٠، ٤٠، ٦٠ .
  - المسافات (٢٠، ٢٠، ٢٠).
- GCD=20
  - عندها الطول المتوقع للمفتاح هو ٢٠ .

Index of Coincidence (IC)

معيار التصادف:

- معيار التصادف هو عبارة عن احتمال تطابق حرفين مختارين من عينة ما .
- IC تم التعارف عليها من قبل William friedman في The Index of coincidence and its Applications in Cryptography(1920).
- الشكل العام لحساب الـIC :

$$\frac{\Sigma(fi * (fi-1))}{N(N-1)}$$

- حيث : $0 \leq i \leq 25$
- f_i :تكرار الحرف الذي ترتيبه i من الأبجدية في العينة المعطاة (النص المشفر).
- N: عدد الأحرف في العينة .
-

- أمثلة:
- معيار التصادف IC للنص " THE INDEX OF COINCIDENCE".
- يعطى كما يلي :
- c(3*2)+ d(2*1)+ e(4*3)+ f(1*0)+ h(1*0)+ i(3*2)+ n(3*2)+ o(2*1)+ t(1*0)+ x(1*0) = 34
- يتم تقسيمها على :
- N*(N-1) = 21*20 = 420
- والتي تعطينا :
- IC of 34/420 = 0.0809

- معيار التصادف للنص " BMQVSZFPJTCSSWGWVJLIO".
- يعطى كما يلي :

- b(1*0)+ c(1*0)+ f(1*0)+ g(1*0)+ i(1*0)+ j(2*1)+ l(1*0)+ m(1*0)+ o(1*0) + p(1*0)+ q(1*0)+ s(3*2)+ t(1*0)+ v(2*1)+ w(2*1)+ z(1*0) = 12

- يتم تقسيمها على :

- N*(N-1) = 21*20 = 420

- والتي تعطينا :

- IC of 12/420 = 0.0286

- معيار التصادف IC لسلسلة محارف باللغة الانكليزية يساوي تقريباً :

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

- الآن نقوم بكتابة النص المشفر C (Ciphertext) كما يلي :

$$
\begin{aligned}
c_1 &= c_1 c_{m+1} c_{2m+1} \cdots \\
c_2 &= c_2 c_{m+2} c_{2m+2} \cdots \\
&\cdots \\
c_m &= c_m c_{2m} c_{3m} \cdots
\end{aligned}
$$

- حيث S_1 أول حرف من كل كتلة ،S_2 ثاني حرف من كل كتلة وهكذا ....
- اذا كانت C1,C2,.....C3 تم تثبيتها على أن m هو طول المفتاح عندها كل Ic(Si) يجب أن يكون تقريباً مساوي ٠,٠٦٥.

One-Time Pad

الشفرة الآمنة

- إذا استخدمت مفتاح عشوائي على طول الرسالة عندها الخوارزمية (Cipher) ستكون آمنة One Time Pad لأنه لا يوجد تكرار للمفتاح . وبالتالي عند فك التشفير ممكن تظهر رسالتين مقروءتين فبذلك لا نعرف أيها الرسالة الأصلية.
- غير قابلة للكسر وذلك لأن النص المشفر ليس له علاقات إحصائية مع النص الواضح (Plaintext).
- من أجل أي نص واضح نحتاج إلى مفتاح عشوائي بنفس الطول .
- "من الصعب توليد عدد كبير من المفاتيح "

Transposition Ciphers

- تعتمد على إخفاء الرسالة عن طريق إعادة ترتيب تتالي أحرف الرسالة ، وذلك دون تبديل أحرف الرسالة بأحرف أخرى .
- مثال :
- (Rail-Fence Cipher)
- النص الواضح هو :
- HELLO WORLD
- يتم إعادة ترتيبه كما يلي :

H L O O L

E L W R D

- عندها يصبح النص المشفر :

HLOOL ELWRD