مجلة ابن الهيثم للعلوم الصرفة والتطبيقية المجلد21 (2) 2008

معالجة النصوص المشفرة انسيابيا باستخدام النص المشفر فقط

قاسم محمد حسين
قسم علوم الحاسبات, كلية الشيخ محمد الكسنزان الجامعة

الخلاصة

يتناول هذا البحث دراسة إمكانية مهاجمة بعض أنظمة التشفير الانسيابي من خلال الاستفادة من خواص اللغة المستعملة في كتابة النص الصريح الذي تم تشفيره ، ومن الخواص العشوائية العالية لمفتاح التشفير (أي تحويل نقاط القوة إلى نقاط ضعف) من خلال استثمار حالة التوازن لعدد الواحدات مع عدد الأصفار للاستفادة منها في تقليص عدد احتمالات المفتاح وبالاعتماد على النص المشفر فقط.

المقدمة

يتم مهاجمة النظام الشفري الانسيابي بالطرائق التقليدية من خلال معرفة الخوارزمية ودراسة نقاط الضعف ، ومن ثم حلها بالطرائق الرياضية الملائمة من خلال الحصول على أحد المستلزمات الاتية اعتمادا على درجة تعقيد الخوارزمية(1).

أ- الحصول على نص صريح ونص مشفر بطول معين.
ب- الحصول على رسائل عديدة مشفرة على المفتاح نفسه.
ج- المهاجمة باستخدام النص المشفر لرسالة واحدة.

من أهم المعضلات التي تواجه محلل الشفرة في مهاجمة الأنظمة الشفرية هو صعوبة الحصول على الخوارزمية التفصيلية المستخدمة في النظام لأن هذه الخوارزميات سرية ولا يسمح بنشرها أو الاطلاع عليها.

كذلك تستخدم هذه الخوارزميات مولدات ذا عشوائية عالية لتوليد مفاتيح التشفير ، ويكون معظم الاستخدام للمفاتيح لمرة واحدة ، لذا برزت الحاجة إلى التفكير بابتكار طرائق وأساليب جديدة لمهاجمة هذا النوع من الأنظمة الشفرية. فدفعنا ذلك الى القيام بهذا البحث.

مفاهيم أساسية: سنتطرق الى بعض المفاهيم الاساسية (2)

التشفير (Cryptography) هو علم تصميم الأنظمة الشفرية التي تحول النص الصريح إلى نص مشفر (نص غير مفهوم عند قراءته).

تحليل الشفرة (Cryptanalysis) إيجاد الطرائق التي يتم فيها تحويل النص المشفر إلى نص واضح بدون توفر المفتاح أو جزء منه.

النص الصريح (Plain text) : النص الأصلي للرسالة المكتوبة وتحتوي المعلومات المطلوب تشفيرها (التي تكون مفهومة عند قرائتها).

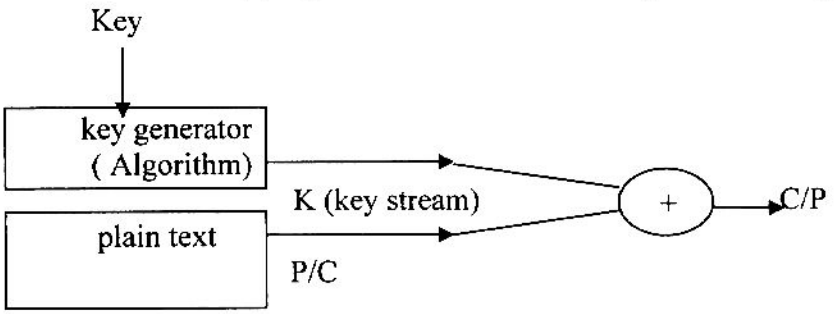
النص المشفر (Cipher text) : نص الرسالة بعد إجراء عملية التشفير (تكون غير مفهومة عند قرائتها).

المفتاح (Key) : وهو مجموعة الرموز او الحروف التي تتعامل مع النص الصريح على وفق صيغة رياضية لتوليد النص المشفر.

خوارزمية التشفير : هي مجموعة الخطوات الرياضية التي تستخدم لتوليد المفتاح الشفري ومعاملته مع النص الواضح لتوليد النص المشفر.

انظمة التشفير الانسيابي

أن نظام التشفير الانسيابي يكون بصورة عامة بالشكل الاتي:- (3).



اذا تمثل :

P : النص الصريح. K : سلسلة المفاتيح المولدة التي تستخدم في التشفير C : النص المشفر.

أن قوة أنظمة التشفير الانسيابي تكمن في وجوب امتلاك مخرجات خوارزمياتها (المفاتيح) مواصفات معينة وهي : (4)

1. أن تكون دورتها طويلة (Long period) . لضمان عدم تكرار المفتاح.
2. امتلاكها لخصائص شبه عشوائية. لتفادي انعكاس الخواص الإحصائية للنص الصريح في النص المشفر، ومنع محلل الشفرة من التوصل إلى أجزاء المفتاح إذا عرف جزء منها.
3. ذو تعقيد خطي كبير (Large linear complexity) ، لعرقلة إيجاد مكافئ لتوليد المفتاح.

الاختبارات الإحصائية للعشوائية

هنالك عدد من الاختبارات الإحصائية التي تستخدم في فحص عشوائية المفتاح المولد في خوارزميات التشفير ، اذا أن عدم اجتياز المفتاح الشفري لها يدل على ضعفه والعكس غير صحيح (5).

ومن تلك الاختبارات:

1. **اختبار التكرار** (The Frequency Test): وهو يفحص عدد الواحدات في المفتاح المولد ، والذي يجب أن لا يختلف عن (n/2) . اذا أن n هو عدد العناصر الثنائية.

2. **اختبار التسلسل** (Serial Test): يقوم بفحص تكرارات التمثيل الثنائي لمقاطع المفتاح وهي 00 ,01 , 10 , 11 التي يجب أن تكون متوازنة.

3.**اختبار بوكر** (Poker Test) :يستخدم لفحص سلسلة المفتاح بعد تجزئتها الى مقاطع ، فإذا كانت السلسلة عشوائية فان عدد المقاطع تكون مستقلة وكذلك تكرار المقاطع تكون متساوية تقريبا.

4.**اختبار الـ Run** : يقوم هذا الاختبار بفحص أطوال المقاطع للواحدات او الأصفار المتشابهة في المفتاح المولد .

5. **اختبار الارتباط الذاتي** (Autocorrelation Test) : يستخدم في قياس عدد المواقع التي تتطابق فيها سلسلة من المفتاح مع نفسها عند تزحيفها بمقدار معين.

تضمن البحث بعض الجوانب النظرية والعملية والنتائج المتحققة.

الجانب النظري

أ. الخاصية العشوائية للمفتاح

لنفرض أن

$$C_i = p_i \oplus K_i \quad i= 1,2,3, \ldots, m$$

اذا أن C_i : النص المشفر لعنصر ثنائي واحد (bit).

P_i : النص الصريح لعنصر ثنائي واحد ، K_i : عنصر ثنائي من المفتاح.

ان سلسلة المفاتيح المولدة (K_i) يجب أن تستقر بعد مدة مناسبة من البدء على عشوائية جيدة اذ يتوازن فيها عدد الأصفار (0's) مع عدد الوحدات (1's) المنتجة من المولد.

وفي التمثيل السباعي للحروف (ASCII) يكون عدد الاحتمالات المستخدمة للمفتاح هو (2^7) أي (128) احتمال لكل حرف . ولان المفتاح المستخدم يجب أن يكون ذو عشوائية جيدة أي يقترب عدد الأصفار من عدد الواحدات فان هناك كثيرا من الاحتمالات تستبعد لعدم إعطائها الخاصية العشوائية الجيدة للمفتاح التي يتراوح فيها عدد الـ 1'S

أو عدد الـ 0'S بين اقل من (35%) واكثر من (65%) وحسب طول المفتـــاح المستخدم ، وهذا معناه استبعاد أكثر من (70 %) من احتمالات المفتاح.

مثال(1)

اذا كان لدينا النص المشفر ZQVQPU وبعد معاملته مع المفتاح التالي:

011001111110010010000111O111

يظهر لنا أن النص الصريح هو :□ THE وكالآتي:

C = ZQVQPU ⇨ 1100111 0110001 1010101 1010111

K ⟹ 0110011 1111001 0010000 1110111

P = THE □ ⇦ 1010100 1001000 1000101 010000

ان هذا المفتاح يعتمد ويؤخذ بنظر الاعتبار لان نسبة الـ 1's في المفتـــاح هـــي (57%) وهي نسبة مقبولة, لذا يعتمد هذا النص كأحد احتمالات النص الصريح.

مثال 2

إذا كان لدينا النص المشفر QCWXBX وبعد معاملته مع المفتاح الاتي:

001000001000100100000100100 1

يظهر لنا أن النص الصريح هو : □ THE وكما يأتي:

C = QCWXBX ⇨ 1100111 0110001 1010101 1010111

K = ⟹ 0010000 0100010 0100000 1001001

P = THE □ ⇦ 1010100 1001000 1000101 010000

نلاحظ أن النص الصريح مقبول ، ولكن عند احتساب نسبة الـ 1's في المفتـــاح نجـــدها تساوي 25%. وهي نسبة غير جيدة لذلك يرفض النص الصريح أعلاه.

ب – الخواص اللغوية للنص الواضح.

أولا : تلاصق الحروف:

لكل لغة طبيعية خواص إحصائية من حيث تكرار حروف الأبجدية في كتابة النـــصوص الصريحة والترابط بين الحروف لتكوين الكلمات. فعلى سبيل المثال أن الحرف E هـــو الأكثر استخداما في اللغة الإنكليزية والحرف (أ) هو الأكثر استخداما في اللغة العربية.

كما أن هنالك العديد من الحروف لا تأتي سوية (متجاورة) ، فعلى سبيل المثال لا يأتي حرفي BF بشكل متجاور في اللغة الإنكليزية (أي لا توجد كلمة مقبولة في اللغة الإنكليزية يأتي فيها الحرفين B و F متجاورين). أن هذا يساعد في استبعاد العديد من الاحتمالات التي ترد في اختيار النصوص الواضحة عند الربط بين الحروف لتكوين الكلمات.

ثانيا : التعويض

على الرغم من أن عدد احتمالات المفتاح هو 128 احتمال لكل حرف يمكن أن تتعامل مع النص المشفر لإيجاد 128 احتمالا لكل حرف من النص الصريح الا انه يستخدم تمثيل لـ((32)) احتمال فقط وبذلك تستطيع اختصار عدد الاحتمالات لكل حرف من (2^7) الى (2^5), أي من (128) حالة الى (32) حالة فقط لكل حرف من النص المفتوح .

الجانب العملي

لغرض التطبيق العملي لهذا البحث :

أ. استخدام نظام تشفير يولد مفتاح عشوائي اجتاز الاختبارات الإحصائية كأنموذج للعمل عليها:(موضح في الملحق رقم 5).

ب-دراسة خواص المفتاح :

تمت دراسة الخاصية العشوائية للمفتاح المستخدم في التشفير من خلال اخذ عشرة مفاتيح يختلف فيها محتوى مكونات خوارزمية النظام الشفري في كل مرة ، ومن ثم إجراء دراسة إحصائية لعدد 1's لاطوال مختلفة من المفتاح(1 ، . . . ، 40) عنصر ثنائي وتم تمثيلها بيانيا,اذ تستقر العشوائية بعد(28) عنصرا تكون متراوحة بين 35% و70% وكما موضحه في الملحق رقم (1) .

ج- دراسة خواص النص الصريح :

أجريت دراسة لتحديد العلاقات بين الأحرف المتجاورة لحروف أبجدية اللغة الإنكليزية وتحديد الحالات التي لا تأتي بعض الحروف مجاورة لحروف اخرى , وقد استعين بقاموس المورد الذي أدخلت جميع كلماته الى الحاسبة ,و أمكن تحديد (197) حالة من

اصل (676) حالة لا تظهر فيها حروف معينة مجاورة للأخرى مثـل ....,KK,ZX . عند دراسة الخواص الإحصائية للغة الإنكليزية تبين أن إمكانية استبعاد الحروف الثنائيـة المتلاصقة تصل إلى 29% ، مثل:

CB , CF, CJ , GK , GV , KC , VB , VJ , WF, ZG, . . .

ويرينا ملحق رقم (2) العلاقة الثنائية بين الحروف للغة الانكليزية.

د- دراسة جداول التعويض

اعتمد جدول التعويض الموضح في الملحق رقم (3) .

اعداد البرامج على الحاسبة الالكترونية: اعدت برامج على الحاسبة الالكترونيـة لتنفيـذ العمليات بدقة وسرعة.

و- **النتائج المتحققة** : تم العمل على نص مشفر بنظام تشفير انسيابي ، وتم العمل علـى أيجاد نصه الصريح باستخدام النص المشفر دون معرفة تفاصيل خوارزميـة التـشفير ، بالاستفادة مما ورد من الدراسات وتم التوصل إلى تحديد الحرفين الأول والثاني بنجاح, اذ تقلص عدد الاحتمالات الى 83 احتمال فقط من اصل 676 احتمال(ملحـق رقـم 4) . وتحديد أربعة حروف من خلال حصر 643 احتمال مقبول من اصل 226576 (26^4) ولغاية الحرف السادس 1825 احتمال مقبول من اصل (26^6).اذ يمكن لمحلل الشفرة أن يقلل من الاحتمالات باستبعاد الحالات التي لا تكون كلمة مقبولة ، وتكون النتائج افضل عندما يصبح لدينا كلمة او كلمات متعددة, اذ نبدأ بالتعامل ضمن سياقات الجملة من ناحية القواعد والمعنى. كذلك استقراء او معرفة الموضوع الذي تتضمنه الرسالة ، وعلى الرغم من هذا الانحسار في عدد الاحتمالات المطلوب تدقيقها الا أن الاحتمالات التي لا تستبعد يبقى كبيرا.

الاستنتاجات

أ- كلما زادت عشوائية المفتاح المستخدم (نسبة الـ 1's تقترب من 50%) كلمـا سـهل الحل. اذ يقل عدد الاحتمالات التي تظهر النص الواضح بتقليل مدى الاحتماليـة. فعلـى سبيل المثال ، لو كانت الاحتمالية هي 50% على مستوى 14 عنصرا فان عدد الحروف التي تظهر تبلغ 14 حرفا فقط لكل حالة.

ب-أن زيادة طول المقطع الذي يجري عليه الاختيار (يحتوي علـى حـروف اكثـر)) يؤدي إلى سهولة أيجاد الحل,اذ يصبح تحديد مدى الاحتمالية اكثر دقة واقل مـدى,اذ أن التعامل مع الكلمات يكون اسهل من التعامل مع الحروف, وان الزيادة في الطول سـيؤدي إلى ظهور كلمات كاملة وحتى جمل.

ج-ضرورة الإلمام باللغة المستخدمة في كتابة النص الصريح لتسهيل عملية اختيار الاحتمالات المقبولة عند أيجاد النص الصريح .

د- أن مكننة طريقة الحل باستخدام الحاسبة الإلكترونية سيختصر الزمن بشكل كبير.

ه- أن هذه الطريقة يمكن أن تكون مساعدة عند استخدام طرائق تعتمد على توافر جـزء من النص الصريح في مهاجمة الأنظمة الشفرية الانسيابية.

التوصيات

أ- إدخال تطبيق الاختبارات الأخرى للعشوائية لتقليل الاحتمالات, اذ أخذنا بالاعتبار فقـط اختبار التكرار.

ب- ضرورة تضمين البرامج إمكانية التدقيق اللغوي للنص الصريح الناتج ، أي ما يشابه الترجمة الآلية في تبيان فيما إذا كانت الجملة الناتجة مقبولة قواعديا وهذا يؤدي الى تقليل الجهد الذي يبذله المحلل في تحديد الحالات التي يتم قبولها.

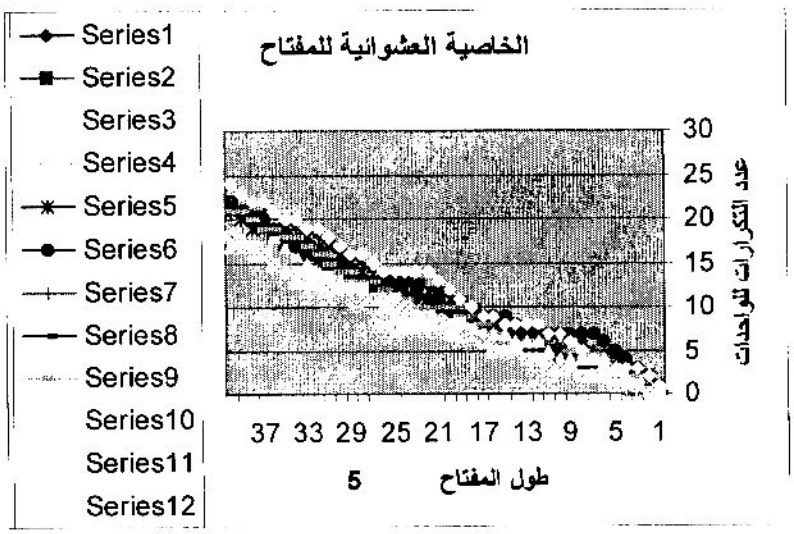
المصادر

1. Meneze , Alfred, J. ; Paul, C.; Oorschot and Scott, A. Vanston (2000). Boca Raton ., Handbook Of Applied Cryptography. FL CRC Press
2. Aki, S. G. and Meijer, H. (1995) , A Fast Pseudo Random Generator With Application To Cryptography.
3. Robshow, M. J. (1995). Stream Cipher. RSA Laboratories Technical Report .
4. Anderson , R. J. (1993). Faster Attack On Certain Stream Cipher, Electronical letters.Fort Washington. PA: A nderson CO.
5. Beker , H. and Piper, F. (1982). Cipher System.. Northwood Publication, London.

5. Beker , H. and Piper, F. (1982). Cipher System.. Northwood Publication, London.
6. Anderson, R. J. (1990). Solving A Class Of Stream Cipher, Cryptologia
7. Blum, M. and Mial, S. (1984). – Random – bits. SIAM Journal on computing.
8. Colic, J. (1994).Linear Cryptanalysis Stream Cipher.
9. Erdman, E. D.(1992) Empirical Tests Of Binary Keystream, master's thesis.

الملاحق

ملحق رقم 1



ملحق رقم (2)

التوزيع التكراري الثنائي لنص واضح في اللغة الانكليزية طوله (10000) حرف

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
2	20	41	37	1	10	19	3	29	1	10	82	27	156	1	19	0	87	82	117	9	18	9	2	21	1
14	1	0	0	47	0	0	0	8	1	0	10	0	0	18	0	0	9	3	1	17	0	0	0	12	0
42	1	6	1	48	1	0	46	20	0	13	12	1	0	60	1	0	12	3	31	10	0	0	0	3	0
37	16	8	10	64	10	6	14	53	2	1	8	9	8	30	7	1	13	25	41	13	2	13	0	7	0
98	20	60	118	45	32	18	23	40	3	6	54	47	121	36	35	5	175	134	81	8	24	39	15	17	1
22	2	5	2	20	14	2	5	26	1	0	7	4	2	42	4	0	19	6	36	8	1	3	0	2	0
21	3	2	2	31	3	3	25	17	0	0	6	3	6	17	2	0	17	8	16	7	0	3	0	2	0
88	2	3	1	261	2	1	4	72	0	0	2	4	1	43	2	0	8	5	23	7	0	4	0	4	0
19	7	52	28	28	16	22	2	0	0	5	37	28	189	55	7	1	26	89	89	10	20	2	2	0	5
2	0	0	0	4	0	0	0	0	0	0	0	0	0	46	0	0	0	0	0	5	0	0	0	0	0
5	1	1	0	22	1	0	2	10	0	0	2	1	5	3	1	0	1	5	3	0	0	2	0	1	0
47	5	5	27	70	8	2	3	53	0	3	52	5	2	36	6	0	3	16	16	11	3	4	0	37	0
49	9	2	1	64	2	0	2	28	0	0	1	8	1	30	17	0	4	10	08	10	0	2	0	5	0
49	8	37	108	64	11	83	11	40	2	5	9	6	10	51	6	1	5	48	12	8	4	11	0	10	0
14	14	17	19	6	86	9	8	10	1	2	31	48	132	26	21	0	101	30	49	78	16	32	1	4	0
26	1	0	0	37	1	0	7	12	0	0	20	2	0	28	11	0	33	5	9	8	0	1	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0
65	7	15	19	146	2	9	9	60	1	8	11	18	15	66	9	0	12	43	48	11	6	8	0	19	0
62	12	23	8	74	13	4	40	63	1	5	11	15	10	57	24	1	7	47	124	25	2	21	0	5	0
63	9	11	6	95	8	3	297	11	1	1	14	10	4	105	6	0	35	38	50	20	1	21	0	18	0
10	7	14	8	11	2	12	1	8	0	0	28	10	33	1	12	0	40	37	36	0	0	1	0	1	0
9	0	0	0	65	0	0	0	19	0	0	0	0	0	5	4	0	0	0	0	0	0	0	0	0	0
42	1	1	1	31	1	0	33	33	0	0	2	1	8	21	1	0	3	4	3	0	0	1	0	1	0
2	0	2	0	1	0	0	0	2	0	0	0	0	0	1	5	0	0	0	3	0	0	0	0	0	0
17	7	7	5	12	6	2	7	12	1	1	4	7	3	23	6	0	4	17	19	1	1	9	0	1	0
2	0	0	0	4	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1

ملحق رقم (3)

جداول التعويض

أ. **التمثيل السباعي (ASCII).**

A = 1000001	B=1000010	C= 1000011	D=1000100
E = 1000101	F=1000110	G= 1000111	H=1001000
I = 1001001	J=1001010	K= 1001011	L=1001100
M = 1001101	N=1001110	O= 1001111	P=1010000
Q = 1010001	R=1010010	S= 1010 011	T=1010100
U = 1010101	V=1010110	W=1010111	X=1011000
Y = 1011001	Z=1011010		
□ = 0100000			

ب – التمثيل الخماسي

A=00000	B= 00001	C=00010	D=00011	E=00100	F=00101
G= 00110	H=00111	I=01000	J=01001	K=01010	L=01011
M=01100	N=01101	O=01110	P=01111	Q=10000	R=10001
S=10010	T=10011	U=10100	V=10101	W=10110	X=10111
Y=11000	Z=11001				

ملحق رقم (4)

احتمالات الحرفين الاول والثاني

PA PL PI RE RA RI RO RI SU ST SE SA SO SI TE
TA TO TI **TH** US UN UM VE VO WE WA WO YE YA ZE
AV AT AS AR AP AG AF AN AM AL AK BU BY BE BA
BO BI CU CE CA CO DU DE DA DO DI EV ET ES ER
EQ EF ED EC EN EM EK FE FA FO GR GO HE HA HO
HI IT IS IR IN IM JE KE

* الحالة الصحيحة هي **TH**

ملحق رقم (5)

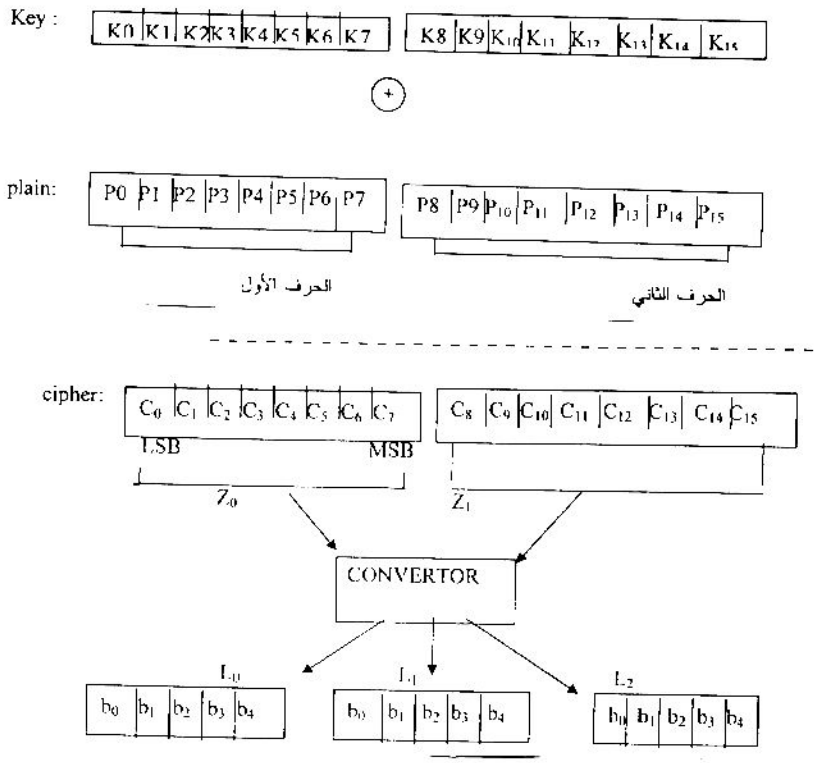
النظام الشفري المستخدم يتضمن القيام بالخطوات الاتية :

أ. دراسة تحويل النص الصريح الى مشفر وبالعكس :

يستخدم النظام خاصية تشفير كل حرفين من النص الصريح بعد معاملتها بعلامة (XOR) مع المفتاح الى ثلاثة حروف مجفرة من خلال الخطوات الاتية وكما موضحه في الشكل رقم (2).

اولا: تحويل حرفي النص الصريح الى التمثيل الثنائي وبواقع (7) عناصرثنائية (Bit) لكل حرف, اذ يتكون لدينا (14) عنصرا.

ثانيا: معاملة الـ(14) عنصرا الناتجة من الفقرة (اولا) مع المفتاح العشوائي المتولد داخل النظام الشفري باستخدام العلاقة(XOR)



شكل (1)

ليتكون لدينا (14) عنصرا يمثل النص المشفر بالتمثيل الثنائي.

ثالثا : يتم تحويل النص المشفر من التمثيل الثنائي الى الحروف باستخدام المعادلات الاتية:

$$L_0 = Z_0 \bmod 26$$
$$L_1 = Z_1 \bmod 26$$

$$L_2 = \frac{Z_0 - L_0}{26} + 5 * \frac{Z_1 - L_1}{26}$$

اذا $0 \le Z_0 \le 127$, Z_1

ملاحظة : Z_0 , Z_1 تمثل قيم العناصر في الفقرة ثانيا .

وكمثال على ذلك :

اذا كانت TH تمثل النص الصريح فيعامل الحرفين TH معا والحرفين E معا.

Plain :	T	H	
	1010100	1001000	
Key :	1100111	0110001	⊕
Cipher :	0110011	1111001	
	$Z_0 = 51$	$Z_1 = 121$	

$L_0 = Z_0 \bmod 26 = 51 \bmod 26 = 25 \Longrightarrow Z$

$L_1 = Z_1 \bmod 26 = 121 \bmod 26 = 17 \Longrightarrow R$

$$L_2 = \frac{Z_0 - L_0}{26} + 5 * \frac{Z_1 - L_1}{26} = \frac{51 - 25}{26} + 5 * \frac{121 - 17}{26}$$

$L_2 = 1 + 5 * 4 = 21 \Longrightarrow V$

Cipher = ZRV

Plain : E □
1000101 0100000
Key : 1010101 1010111
⊕

Cipher : 0010000 1110111
Z_0= 16 Z_1 = 119

$L_0 = 16 \Longrightarrow Q$
$L_1 = 15 \Longrightarrow P$

$$L_2 = \frac{16-16}{26} + 5 * \frac{119-15}{26} = 0 + 5(4) = 20 \Longrightarrow U$$

وبذلك يشفر النص الصريح THE□ إلى النص المشفر ZRVQPU.

IBN AL- HAITHAM J. FOR PURE & APPL. SCI VOL.21 (2) 2008

Attack Stream Cipher By Using Cipher System Only

K. M. Hussein
College of Al-Shaikh Muhammed AL-Kaznazan/ Computer Science Department

Abstract

Attack stream cipher system , using cipher text only , depends on the characteristics of plain text language and the randomness of the key , that used in encryption , without having detailed knowledge of cipher algorithm by benefiting from the balance between 0's and 1' in the key to reduce the probability of key space.