Cryptanalysis of Knapsack Cipher Using Genetic Algorithm

Subhi H. Hamdon Najlaa B. Al-Dabbagh Milad J. Saeed
College of Computer Science and Mathematics
University of Mosul, Iraq

Received on: 02/05/2007 Accepted on: 26/09/2007

ABSTRACT

This research offers a new method in Cryptanalysis of knapsack cipher. It focuses on the application of genetic algorithm as a modern way in solving complex problems (problems have a huge numbers of alternate solutions in appropriate time). One of these problems is knapsack problem which is considered one of the known problems in operation researches. Cryptanalysis is done by using a new algorithm that is different from known knapsack breaking algorithm. Genetic algorithm has recently been successfully applied to the cryptanalysis of ciphers, among them Substitution ciphers and Transposition ciphers. This research deals with another type of ciphers called Public-key ciphers, that are high secure ciphers because they are based on NP-Complete problems.

Keywords: Cryptography, Cryptanalysis, Genetic algorithm.

تحليل شفرة نابساك باستخدام الخوارزمية الجينية

صبحي حمادي حمدون نجلاء بديع الدباغ ميلاد جادر سعيد

كلية علوم الحاسبات والرياضيات، جامعة الموصل

تاريخ استلام البحث: 2007/05/02 تاريخ قبول البحث: 2007/09/26

الملخص

قدم هذا البحث طريقة جديدة في التحليل، إذ ركز على كسر شفرة النابساك Knapsack Cipher ، اعتمد على طريقة تطبيق الخوارزمية الجينية بوصفه أسلوبا حديثا في حل المسائل المعقدة (مسائل كبيرة تمتلك كماً هائلاً من الحلول البديلة بزمن مناسب)، وأحدى هذه المسائل هي مسألة النابساك المسماة حقيبة الجندي أيضا التي تعتبر من المسائل المعروفة في بحوث العمليات، إذ يتم تحليل هذه الشفرة باستخدام خوارزمية جديدة تختلف عن خوارزميات الكسر للنابساك المعروفة، وقد استخدمت الخوارزمية الجينية في تحليل الشفرات ومن بينها الشفرات التعويضية و الابدالية ، أما هذا البحث فتناول النوع الأخر من الشفرات وهي شفرات المفتاح العام Public – Key Cipher التي تعتبر من الشفرات العالية السرية لأنها تعتمد على المسائل المسماة NP-Complete Problem .

الكلمات المفتاحية: التشفير، التحليل، الخوارزمية الجينية.

1.المقدمة

علم التشفير Cryptographhgy هو العلم الذي يختص بحماية المعلومات من الأشخاص غير المخولين الذين يحاولون ان يعترضوا المعلومات أو العبث بها [4] . وهو الذي يسمح أيضا بالتغير الجذري للمعلومات لغرض إخفاء محتواها عن طرف ثالث. أما العلم الذي يختص بكسر الشفرة فهو علم تحليل الشفرة Cryptanalysis. ولتحويل النص الصريح إلى نص مشفر يحتاج إلى خوارزمية ومفتاح. وهنالك نوعان من خوارزميات التشفير المعروفة [8]:

- خوارزميات تشفير المفتاح السري التي تستخدم المفتاح نفسه في التشفير وفك الشفرة.
- خوارزميات تشفير المفتاح العام التي تستخدم زوجا من المفاتيح التي تكون مترابطة رياضياً، الأول يستخدم في التشفير والثاني يستخدم لفك الشفرة.

اقترح شفرة نابساك العالمان Merkle and Hellman [3] وتعتبر من المحاولات الأولى لأنظمة المفتاح العام ويطلق عليها أيضا مسالة حقيبة الجندي وشفرة نابساك تتعلق بإيجاد حل لمسألة الجمع الجزئي Subset Sum Problem التي تعرف رياضياً كالأتي:-

توجد مجموعة من الأعداد الصحيحة الموجبة ولتكن A

A={a1,a2,a3,…….an}

ويوجد عدد صحيح موجب وليكن B والمطلوب هل هناك مجموعة جزئية 'A من المجموعة A بحيث أن مجموع عناصر 'A يساوي القيمة B .

إن نابساك الاعتيادية المسماة Super increasing Knapsack من السهولة كسرها أي بمعنى أخر لا يمكن استخدامها لحماية البيانات لذلك اقترح العالمان طريقة لتحويل نابساك البسيطة Simple Knapsack إلى نابساك باب المصيدة Trapdoor Knapsack التي يصعب كسرها والطريقة هي كالأتي:

إذ كان لديك سلسة نابساك البسيطة

A={ a'1,a'2,a'3,…….a'n}

فأن عمليات التحويل إلى سلسلة باب المصيدة هي :-

1. اختيار عدد صحيح U بحيث U>2a'n.
2. اختيار عدد صحيح أخر W بحيث gcd (U,W)=1.
3. إيجاد قيمة W^{-1}.
4. تكون سلسلة باب المصيدة A= WA' mod U.

وبالرغم من أن شفرة نابساك تعتبر من مسائل NP- Complete وهي شفرة ذات سرية عالية، فأن معظم إصدارات هذه الشفرة تم كسرها. حيث أوجد بريكل طريقة لكسر هذه الشفرة عام 1984 [1]. وجاء من بعده شامير ليطور خوارزمية متعددة الحدود الزمنية لكسر شفرة نابساك البسيطة [6].
وقد قدم هذا البحث طريقة أخرى امتازت باتساع تطبيقها (حيث يمكن تطبيقها على كل شفرات النابساك) وسهولة عملها. اعتمدت الخوارزمية الجينية لتحقيق هذا الهدف.

2- الخوارزمية الجينية:

تعرف الخوارزمية الجينية بأنها خوارزمية ذكية يمكن استخدامها لإيجاد وتحسين حل المسائل المعقدة التي تدخل في العديد من المجالات لإعطاء حل ابتدائي للمسألة أو لتحسين حل موجود مسبقاً وهي تطوير لما يعرف ب Evolutionary Programming.

طورت الخوارزمية الجينة على يد العالم جون هولاند John Holland عام 1975 في جامعة ميشغان وكان الهدف الأساسي منها بناء وتحسين العديد من الخوارزميات والبرمجيات والأنظمة وقد اختصرت الزمن والجهد المطلوب لدى مصممي الأنظمة والبرامج وذلك من خلال توفيرها خوارزمية عامة يعتمد عليها في حل مختلف المسائل بدلاً من بناء خوارزمية خاصة لكل مسألة، مع مراعاة التغيرات اللازمة التي تتناسب مع خصوصية كل مسألة من حيث حجم ونوع البيانات المستخدمة وطبيعة دالة الهدف وقيود كل مسألة [2,5].

- استخدمت الخوارزمية الجينية في العديد من المجالات منها:-

1. حل المسائل الصعبة في مجال بحوث العمليات والتحليل العددي.
2. حل مسائل التشفير وكسر الشفرة.
3. تعليم الحاسبة.
4. معالجة الصور.

- **الخطوات العامة للخوارزمية الجينية:**

تتضمن الخوارزمية الجينية عدداً من الخطوات الثابتة لمختلف المسائل ولجميع التطبيقات ويكون الاختلاف في أسلوب صياغة وتطبيق كل خطوة من هذه الخطوات حسب المسألة ومجال تطبيقها، وان هذه الخطوات تكون مترابطة بعضها مع البعض الأخر ولا يمكن تطبيق هذه الخوارزمية على أية مسالة ما لم تطبق جميع هذه الخطوات وألا فستفقد الخوارزمية الجينية قيمتها وفائدتها في إيجاد الحل وتحسينه.

- إنشاء الجيل الابتدائي

إن إنشاء الجيل الابتدائي يعد نقطة الانطلاق في تحسين حالة المسألة المراد تطبيقها. وان جميع المصادر المتوافرة تبين أن عملية بناء الجيل الابتدائي تتم بطريقة عشوائية.

- دالة الهدف وقيمة اللياقة واحتمالية المساهمة

- **دالة الهدف: objective function**

تحسب دالة الهدف الخاصة لمسألة ما لكل مقطع من مقاطع الجيل.

◈ **مدى اللياقة: Fitness Value**

تحسب قيمة مدى اللياقة لكل مقطع من مقاطع الجيل اعتماداً على قيمة دالة الهدف لذلك المقطع. وتحدد قيمة مدى اللياقة مقدار جودة المقطع قياساً بباقي مقاطع الجيل.

◈ **احتمالية المساهمة Probability of Contribution**

تجمع درجات اللياقة لجميع مقاطع الجيل لكي تستخدم في إيجاد احتمالية مساهمة كل مقطع في تكوين الجيل اللاحق والتي تحسب بقسمة درجة اللياقة لكل مقطع على قيمة مجموع درجات اللياقة.

- بعد حساب دالة الهدف ومدى اللياقة واحتمالية المساهمة لجميع مقاطع الجيل الابتدائي تبدأ عملية تكوين الجيل الجديد بتطبيق عدد من العمليات على المقاطع المنتخبة من الجيل الحالي وتتمثل هذه العمليات بالدوال الثلاث آلاتية:-

- اختيار الآباء Parents Selection

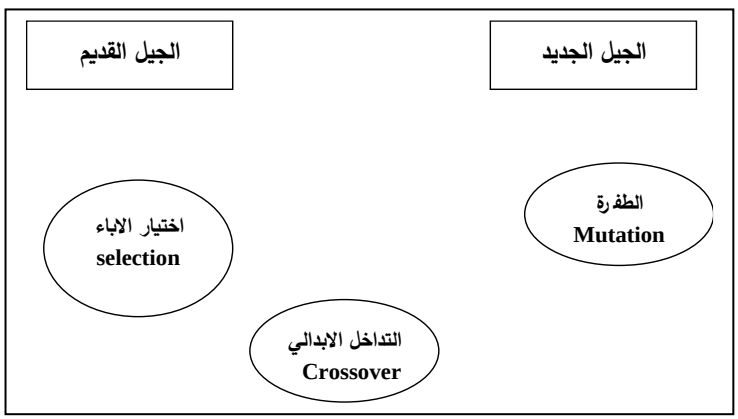
هذه العملية تحدد أي مقطع في هذا الجيل سوف يساهم في توليد الجيل اللاحق وهناك عدد من خوارزميات الاختيار [7] وقد استخدمت طريقة Random selection في مسألة كسر شفرة Knapsack.

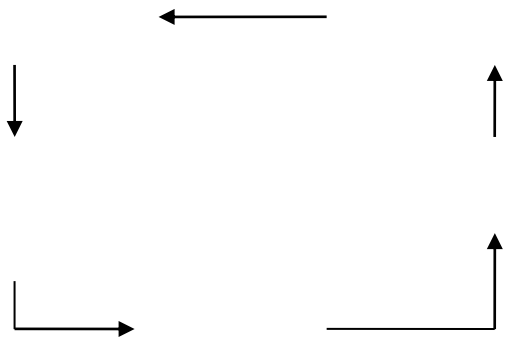
- التداخل الابدالي Crossover

يتكون مقطع جديد ليمثل احد مقاطع الجيل اللاحق من مقطعين من مقاطع الآباء المنتخبة من مقاطع الجيل الحالي باستخدام إحدى طرائق التداخل المعروفة [2].

- التغير ضمن المقطع (الطفرة) Mutation

تتمثل هذه العملية بأجراء تغيير أو تبديل بين قيم محددة ضمن المقطع نفسه لتكوين وإدخال مقاطع تعطي حلولاً جديدة إلى الجيل اللاحق والتي لم يسبق تكوينها في الأجيال السابقة وذلك بهدف توسيع حيز الحلول الممكنة والتمثيل بتكوين اكبر عدد من المقاطع المختلفة ضمن الجيل.





الشكل (1) يوضح دورة حياة الخوارزمية الجينية

- **معيار توقف الخوارزمية الجينية:**

يستمر تكوين الأجيال المتعاقبة بهدف تحسين الحل أي يجعله أكثر اقتراباً من الحل الامثل إلى أن يتحقق شرط التوقف الذي يعتمد على معيار توقف الخوارزمية ويختلف من مسألة إلى أخرى، ومن أهم معايير التوقف وأكثرها استخداماً:

- التحديد المسبق لعدد الأجيال المتكونة.
- التحديد المسبق لعدد الأجيال التي لم يحدث تكوينها أي تحسين للحل.
- قد تحدد نسبة معينة تمثل عدد المقاطع المتشابهة قياساً بالعدد الكلي لمقاطع الجيل, إذ تفحص هذه النسبة بعد تكوين كل جيل فإذا بلغت نسبة المقاطع المتشابهة القيمة المحددة مسبقاً تتوقف الخوارزمية.

3- تحليل شفرة نابساك:

اعتمدت طريقة التحليل على ثلاث خطوات رئيسية:

أ- التمثيل ودالة الهدف

الهيكل التمثيلي لمشكلة النابساك سهل التوليد لان المسألة تقترح صيغة طبيعية الشكل وهو شكل البت الثنائي التي ربما تكون هي أحسن هيكلية وعليه فأن الصيغة العامة ل Trapdoor KnapSack تكون مثلاً كالأتي:

إضافة القيم الموجودة في المواقع 1, 3, 5, 7	1	0	1	0	1	0	1	0
إضافة القيم الموجودة في المواقع 2, 3, 5, 6	0	1	1	0	1	1	0	0

عندما يتم التمثيل للخوارزمية الجينية من الضروري إضافة دالة الهدف التي تستخدم لتحديد أفضل التمثيلات لشفرة النابساك. أما الهيكل الأساسي لهذه الدالة فهو سهل التحديد. مع هذه الإجراءات العامة يوجد مدى واسع للمتغيرات المحتملة، في هذا البحث فأن دالة الهدف الحقيقية يجب أن يكون لها ثلاث خصائص:

- مدى الدالة يجب أن يكون بين الصفر والواحد، إذ أن الواحد يمثل التطابق التام مع مجموع الهدف للنابساك.
- الكروموسومات التي تنتج مجموع اكبر من الهدف يجب أن تكون بصورة عامة لها مدى اللياقة اقل من الكروموسومات التي تنتج مجموع اقل من الهدف، بهذه الطريقة فان الحلول غير العملية هي التي تنتج مجموع أكثر من الهدف المطلوب بينما الحلول العملية لها فرصة اكبر من أن تتماشى مع الخوارزمية.
- يجب أن تكون هناك صعوبة في إنتاج قيمة لياقة عالية ، وان الاختلافات الصغيرة بين الكروموسومات الحالية ومجموع الهدف يجب أن تضخم، وهذا ينجز باستخدام الجذر ألتربيعي في حساب الدالة المختارة.

إن دالة الهدف الحقيقة التي تستعمل في هذا البحث تستخدم لتحديد ما يأتي:

1. حساب أعلى اختلاف قد يحدث بين الكروموسومات ومجموع الهدف

MaxDiff = max (Target, Full sum – Target)

حيث Fullsum هو مجموع كل مكونات النابساك.

2. تحدد قيمة الكروموسوم الحالي ويطلق عليه المجموع Sum .
3. إذا كان المجموع اقل أو يساوي الهدف فاللياقة للكروموسوم تحسب كما يأتي:

Fit= 1-sqrt(|sum – target | / Target)

4. إذا كان المجموع اكبر من الهدف فان اللياقة للكروموسوم تحسب كما يأتي:

$$Fit = 1 - \sqrt[6]{(|sum - Target|)/(MaxDiff)}$$

ب- عملية التزاوج

في هذه المرحلة لدينا عددا من الكروموسومات التي يطلق عليها المجتمع وكل واحدة من هذه الكروموسومات لها قيمة اللياقة، إذ أن الخوارزمية تتقدم باختيار عشوائي لاثنين من الكروموسومات لكي يتم التزاوج بينهما، ومن ثم يتم تأييد الكروموسومات التي لها قيمة لياقة عالية إذ أن الكروموسومات التي لها مدى لياقة عالية لها فرصة اكبر من حيث اختيارها لتوليد الأبناء في الجيل القادم لأنه عادة يستخدم الأبوان لتوليد اثنين من الأبناء باستخدام طريقة تداخل إبدالي لقيمة واحدة.

ج- عملية الطفرة

هذه المرحلة تساهم في توليد أبناء جدد لكي تجنب الخوارزمية الوقوع في نقطة الحل المثلى المحلية Local Optimal وقد تم استخدام المتمم Complement التي سوف تطبق بعد أن تتم مقارنة قيمة الاحتمالية المختارة عشوائياً مع القيمة المدخلة من التنفيذ.

4- الخوارزمية الكاملة:

هذه المعالجات تتعاون لخلق خوارزمية جينية متكاملة، تبدأ بنص مشفر باستخدام شفرة نابساك وهي سلسلة من الأرقام كل رقم يمثل مجموع النابساك الصعب، الهدف من الخوارزمية الجينية هو لترجمة كل رقم إلى التمثيل الصحيح للنابساك والذي يمثل الاسكي للنص الصريح وعليه فان النقاط التالية تُنفذ لكل رقم مشفر وهي كالأتي:

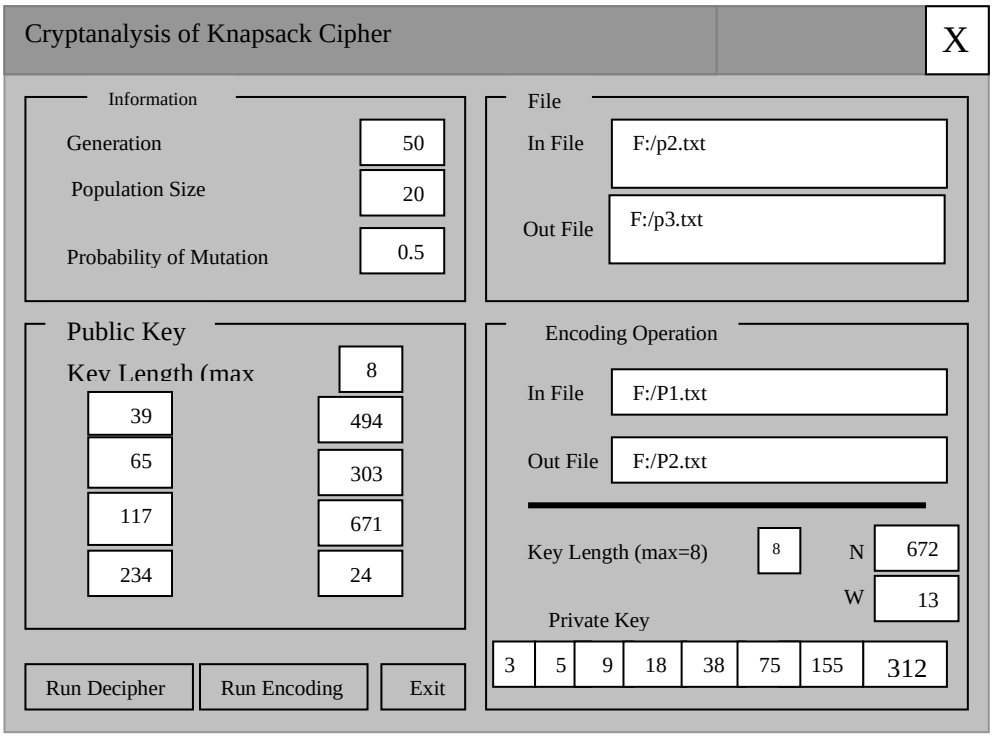
1. توليد مجتمع عشوائي للكروموسومات (سلسلة ثنائية من الاصفار والواحدات).
2. حساب مدى اللياقة لكل كروموسوم في المجتمع.
3. بناء قاعدة للاختيار العشوائي بالاعتماد على اللياقة.
4. تطبيق عملية التزاوج على الآباء المختارة.
5. تطبيق عملية الطفرة على الأبناء.

إن هذه العملية سوف تتوقف بعد عدد ثابت من الأجيال وأفضل كروموسوم هو الذي سوف يستخدم لفك شفرة النص المشفر(العدد).

5- النتائج والتطبيق العملي:

تمت برمجة خوارزمية التحليل باستخدام لغة Delphi البرمجية الواجهة الرئيسية للبرنامج موضحة بالشكل (2) ويمكن توضيح استخدام الخوارزمية بالمثال العملي الأتي:

1- لترميز كل حرف باستخدام رمز الاسكي تم اختيار نابساك البسيطة Super Incresing Knapsack المكونة من ثمانية أعداد {3, 5, 9, 18, 38, 75, 155, 312}أما قيم N و W فهي N=672 و W=13



الشكل (2) واجهة البرنامج

2- تم تكوين نابساك الصعب Hard Knapsack وكانت

{ 39, 65, 117, 234, 494, 303, 671, 24}

3- اختير النص الواضح التالي ليتم تشفيره

AN OPERATING SYSTEM IS APROGRAM THAT ACTS AS NINTEREDIARY BETWEEN AUSER OF COMPUTER AND THE COMPUTER HARDWARE

4- النص المشفر كان:-

{710,1087,1126,1165,827,1230,710,1282,944,1087,892,1269,1438,1269,1282,827,1061,944,1279,710,1165,1230,1126,892,1230,710,1061,1282,905,710,1282,710,775,1282,1269,710,1269,710,1087,944,1087,1282,827,1230,827,788,944,710,1230,1438,736,827,1282,1386,827,827,1089,710,1231,1269,827,1230,1126,853,710,775,1126,1061,1165,1231,1282,827,1230,710,1087,788,1282,905,827,775,1126,1061,1165,1231,1282,827,1230,905,710,1230,788,1386,710,1230,827}

5- أما النتائج فموضحة بالجدول (1):-

Key	Sum	Fitness	Char	Best Gen
10000010	710	1.0000	A	35
01110010	1087	1.0000	N	8
11110010	1126	1.0000	O	8
00001010	1165	1.0000	P	10
10100010	827	1.0000	E	13
01001010	1230	1.0000	R	24
10000010	710	1.0000	A	5
00101010	1282	1.0000	T	2
10010010	944	1.0000	I	6
01110010	1087	1.0000	N	9
11100010	892	1.0000	G	1
11001010	1269	1.0000	S	29
10011010	1438	1.0000	Y	21
11001010	1269	1.0000	S	46
00101010	1282	1.0000	T	3
10100010	827	1.0000	E	15
10110010	1061	1.0000	M	6
10010010	944	1.0000	I	3
11001010	1269	1.0000	S	17
10000010	710	1.0000	A	15
00001010	1165	1.0000	P	16
01001010	1230	1.0000	R	11
11110010	1126	1.0000	O	4
11100010	892	1.0000	G	1
01001010	1230	1.0000	R	3
10000010	710	1.0000	A	11
10110010	1061	1.0000	M	1
00101010	1282	1.0000	T	2
00010010	905	1.0000	H	1
10000010	710	1.0000	A	8
00101010	1282	1.0000	T	1
10000010	710	1.0000	A	3
11000010	775	1.0000	C	35
00101010	1282	1.0000	T	3
11001010	1269	1.0000	S	17
10000010	710	1.0000	A	15
11001010	1269	1.0000	S	12
10000010	710	1.0000	A	24
01110010	1087	1.0000	N	27

10010010	944	1.0000	I	4
01110010	1087	1.0000	N	4
00101010	1282	1.0000	T	10
10100010	827	1.0000	E	26
01001010	1230	1.0000	R	23
10100010	827	1.0000	E	6
00100010	788	1.0000	D	1
10010010	944	1.0000	I	17
10000010	710	1.0000	A	17
01001010	1230	1.0000	R	33
10011010	1438	1.0000	Y	1
01000010	736	1.0000	B	6
10100010	827	1.0000	E	3
00101010	1282	1.0000	T	1
11101010	1386	1.0000	W	12
10100010	827	1.0000	E	9
10100010	827	1.0000	E	2
01110010	1087	1.0000	N	5
10000010	710	1.0000	A	16
10101010	1321	1.0000	U	12
11001010	1269	1.0000	S	6
10100010	827	1.0000	E	25
01001010	1230	1.0000	R	49
11110010	1126	1.0000	O	1
01100010	853	1.0000	F	16
10000010	710	1.0000	A	51
11000010	775	1.0000	C	20
11110010	1126	1.0000	O	55
10110010	1061	1.0000	M	12
00001010	1165	1.0000	P	1
10101010	1231	1.0000	U	3
00101010	1282	1.0000	T	6
10100010	827	1.0000	E	15
01001010	1230	1.0000	R	4
10000010	710	1.0000	A	44
01110010	1087	1.0000	N	2
00100010	788	1.0000	D	1
00101010	1282	1.0000	T	7
00010010	905	1.0000	H	1
10100010	827	1.0000	E	1
11000010	775	1.0000	C	17
11110010	1126	1.0000	O	16
10110010	1061	1.0000	M	2
00001010	1165	1.0000	P	34
10101010	1321	1.0000	U	6
00101010	1282	1.0000	T	3
10100010	827	1.0000	E	15
01001010	1230	1.0000	R	17
00010010	905	1.0000	H	9
10000010	710	1.0000	A	20
01001010	1230	1.0000	R	29
00100010	788	1.0000	D	6

11101010	1386	1.0000	W	8
10000010	710	1.0000	A	8
01001010	1230	1.0000	R	13
10100010	827	1.0000	E	3

الجدول (1) النتائج

6- الاستنتاجات و التوصيات :

استخدمت الخوارزمية الجينية في تحليل شفرة نابساك التي تعتبر أداة جديدة وقوية للتحليل حيث جهزت طريقة ناجحة جداً لإيجاد التمثيل الثنائي الصحيح لمجموع نابساك الصعبة(Hard Knapsack)، وهذه الخوارزمية يمكن أن تستعمل لأي حجم من النابساك، أي أن حجم النابساك الذي سوف تطبق عليه الخوارزمية يعتمد على حجم الذاكرة المتوافرة للإلة المستخدمة.

هنالك عدة مجالات مفتوحة للأعمال المستقبلية، فعلى سبيل المثال ممكن تحسين كفاءة الخوارزمية إذا تمت معرفة بعض حروف النص ومن جهة أخرى يمكن تطبيق الخوارزمية الجينية لكسر شفرات أكثر تعقيداً.

المصادر

[1] Brickell, E. (1984) "Solving low Density Knapsacks", Advances in Cryptolog3/: proceedings of CRYPTO 88. NewYork: Plenum Press. Pp.25-37.

[2] Delman B. (2004) "Genetic Algorithms in Cryptography", A thesis Submitted in partial Fulfillment of the Requirements of the Degree of Master of Science in Computer Engineering, Rochester, New York, July.

[3] Fed piper (1982) “Cipher System The Protection of Communications”

[4] Gurpreet Dhillon. (2007) **Principles of Information Systems Security**, John Wiley & Sons.

[5] Rawlins, G. (1991) **Foundations of Genetic Algorithms** ,los Altos CA: Morgan Kaufmann Publishers.

[6] Shamir, A. (1984) "A polynomial – Time Algorithm for Breaking the Basic Merkle- Hellman Cryptosystem. IEEE Transactions on Information Theory. IT30:PP.699-704.

[7] Spillman R.; Janssen M.; Nelson B. and Kepner M. (1993) “Use of A Genetic Algorithm in analysis of simple Substitution Ciphers”,Department of Computer Science Pactific Lutheran , University Tacoma WA 98447 USA.

[8] Wiley,J. &Sons (1996) “Applied Cryptography”.