مهاجمة نص مشفر (معلوم نصه الصريح) باستخدام خوارزمية جينية

يحيى قاسم إبراهيم
قسم علوم الحاسبات / كلية التربية
جامعة الموصل
yahyaki@yahoo.com

الاستلام **القبول**
2011 / 06 / 26 **2012 / 04 / 03**

Abstract

A Genetic Algorithm is defined as a smart algorithm which can be used to solve the complex matters and improve them, which are used in many fields. Also, Genetic Algorithm is considered as one of the efficient search methods which depends on the natural test and genetics.

In this research, the Genetic Algorithm features were used to determined and find the encrypted key which was used to encrypt a plain-text by firstly using many selected random generated keys, then apply the Genetic Algorithm with its different features such as: selection, crossover and mutation, leading to the most efficient key to attack the cipher-text and return it to its plain-text with very small or non error ratio. Thus the user will be able to use this key to attack other coming texts for the same destination. The GA which is used in this paper were applied on a different English texts.

Results showed the importance of Genetic Algorithm through its potentiality in wide space intelligent search of keys to find the right key, from the other hand, these results showed the accuracy of Genetic Algorithm during its approach.

الخلاصة

تعرف الخوارزمية الجينية بأنها خوارزمية ذكية يمكن استخدامها لإيجاد حل المسائل المعقدة وتحسينها، والتي تدخل في العديد من المجالات. كما تعد الخوارزمية الجينية من طرائق البحث الكفوءة المعتمدة على مبدأ الاختيار الطبيعي وعلم الوراثة.

وتم في هذا البحث الاستفادة من خواص الخوارزمية الجينية لتحديد وإيجاد مفتاح التشفير الذي تم به تشفير نصاً صريحاً مسبقاً من خلال التعامل مع عدد من المفاتيح المنتخبة والمولدة عشوائياً في بادئ الخوارزمية ومن ثم تطبيق الخوارزمية الجينية بعناصرها المختلفة من انتقاء Selection وتقاطع Crossover وطفرة Mutation وصولاً إلى المفتاح الأكثر فعالية في كسر النص المشفر وإرجاعه إلى النص الصريح بنسبة خطأ ضئيلة جداً أو تكاد تكون معدومة. وأن المفتاح الذي سيتم الحصول عليه جراء تطبيق الخوارزمية سيؤهل المستخدم إلى كسر النصوص المشفرة التالية لنفس مصدر النص الصريح الأول دون الحاجة لمعرفة النصوص الصريحة التالية.

وقد تم تطبيق الخوارزمية المقترحة على مجموعة من النصوص الانكليزية المختلفة وقد أظهرت النتائج إلى أهمية الخوارزمية الجينية في الأداء من حيث إمكانياتها في عمليات البحث الذكية خلال مدى واسع من المفاتيح لإيجاد المفتاح الصحيح ومن جانب آخر دقة الخوارزمية الجينية المستخدمة في نتائجها.

1- المقدمة:

قاد ازدياد الصعوبات والمشاكل التقنية وتعقيدها وعدم قدرة الحلول البرمجية التقليدية على استيعاب هذه المشاكل، قاد الباحثين الى التعمق في إيجاد خوارزميات وبرامج كفوءة تساعد في إيجاد الحلول المناسبة والمثالية للمسائل المعقدة وفي سرعة الوصول الى الحلول وخزنها واسترجاعها ومن ثم التوصل الى هيكلية مترابطة لبنية برمجية ذكية ومن هذه الخوارزميات الكفوءة هي الخوارزمية الجينية [1].

وتعتبر مسألة البحث عن مفتاح التشفير للنصوص الانكليزية (المعتمدة في البحث) من المسائل التي تأخذ حيزاً كبيراً من الوقت بسبب كثرة المعالجات وخطوات الحل فيما اذا تم استخدام الخوارزميات التقليدية في عملية البحث هذه، فكان من الضروري استخدام وتطبيق مبادئ الخوارزمية الجينية واعتمادها في الحل.

2- الخوارزمية الجينية Genetic Algorithm:

تعتبر الخوارزمية الجينية أحد أساليب الذكاء الاصطناعي (Artificial Intelligence) المهمة، برزت أهميتها في حل المسائل المعقدة خلال زمن مناسب [2]، وهي واحدة من خوارزميات البحث العامة التي تعتمد على افكار الهندسة الوراثية، ويبدأ حل المسائل المعقدة باستخدام الخوارزميات الجينية بمجتمع عشوائي (Initial Population) يمثل مجموعة الحلول [3]، كل حل تخصص له صلاحية (Fitness Value) معينة ترتبط مباشرة

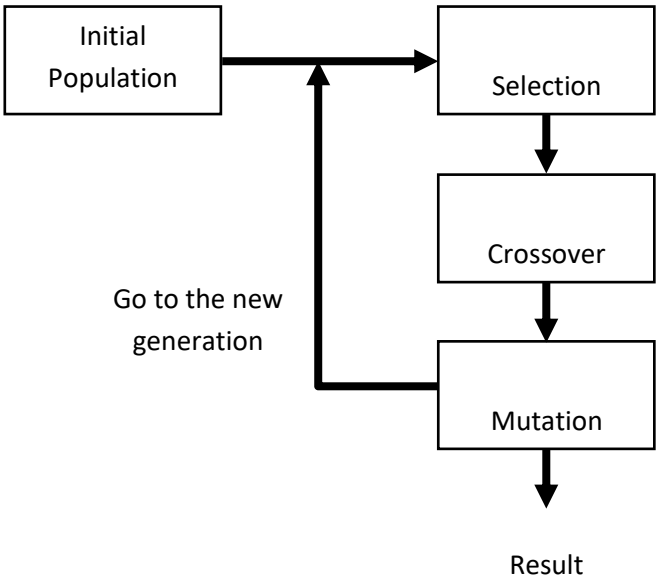
بدالة الهدف (Objective Function) للمسألة المعينة [4]، وبعدها يتم تعديل هذا المجتمع وتوليد مجتمع آخر جديد من خلال تطبيق مجموعة من العوامل الجينية والوراثية (Genetic Operators)، منها الانتقاء (Selection) والتقاطع (Crossover) والطفرة (Mutation) وغيرها من العوامل بصورة متكررة وبالتتابع على اجيال هذا المجتمع لحين تحقق شرط التوقف (Stop Criteria)، لاحظ الشكل رقم (1). [5].

ويتكون المجتمع من عدد من الافراد، حيث يحدد عدد الافراد من قبل مصمم الخوارزمية، فإذا كان حجم المجتمع كبير جداً في بعض الاحيان قد لا يعطي اداءً جيداً للخوارزمية، وكذلك الحال اذا كان حجم المجتمع صغير جداً وغالباً ما يكون احسن حجم متراوحاً ما بين 20–30 وأحياناً ما يكون بين 50–100 ليعطي احسن حل. [6].

ان كل فرد (كروموسوم) يتكون من عدد من القيم ويكون عددها طول الكروموسوم، ويحدد حسب المسألة وقيمة البداية تكون عشوائية ضمن المحددات من المسألة المراد حلها وتسمى قيمة الجينة بـ (Allele). [2] [7].

وقد استخدمت الخوارزمية الجينية بصورة واسعة في مجالات عديدة منها: معالجة الصور (Image Processing) وتمييز الأنماط (Pattern Recognition) وغيرها من المجالات، وقد لاقت نجاحاً واسعاً وعناية كبيرة وبخاصة في السنوات الأخيرة. [8] [9].

وتعتبر الخوارزمية الجينية من تقنيات البحث التي تعتمد التكرار وتعمل بنجاح في العديد من مشكلات الأمثلية الصعبة ولذلك فهي أفضل من التقنيات الأخرى في حالة كون فضاء البحث كبير جداً مثل طرق الحدس (Heuristic) وطريقة البحث الأعمى (Blind Search).[10].



الشكل رقم (1): المخطط العام للخوارزمية الجينية.[10]

تم الاستفادة من خواص الخوارزمية الجينية من حيث السرعة في إعطاء النتائج والمساعدة على تجاوز مراحل عديدة لا يمكن تجاوزها في حالة عدم استخدام الخوارزمية الجينية [3]، حيث تتجه الحلول في الخوارزمية الجينية بشكل سريع الى الحل المثالي وتضمن هذه الطريقة عادة عدم المرور بكل النقاط الموجودة في فضاء البحث الواسع. [4] [6].

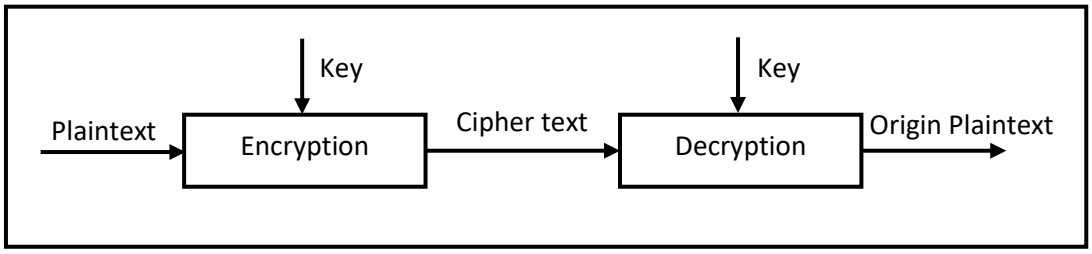
ومن مميزات الخوارزمية الجينية يمكن ان نذكر ما يلي: [1] [4]

1. الخوارزمية الجينية تبحث في المجتمع وهو عبارة عن مجموعة نقاط وليس نقطة واحدة.
2. الخوارزمية الجينية تستعمل دالة الهدف مباشرة ولا تتوسع في معلومات اضافية.
3. الخوارزمية الجينية تستخدم بعض قوانين الاحتمالية ولا تستخدم القوانين التقليدية.
4. الخوارزمية الجينية تعمل مع التشفير لمجموعة المتغيرات وليس مع المتغيرات نفسها.

تم في هذا البحث تطبيق الخوارزمية الجينية على مجموعة من السلاسل الحرفية تمثل السلسلة الحرفية الواحدة مفتاح من مفاتيح التشفير قيد البحث لكسر النص المشفر. وتم تمثيل هذه السلسلة الحرفية بمتجه من الحروف لسهولة إجراء العمليات الجينية والوراثية عليها، ومن ضمنها إجراء عملية التقاطع بين سلسلتين حسب موقع عشوائي لأحد عناصر السلسلتين، وكذلك إجراء عملية الطفرة الوراثية بتغيير احد حروف المتجه عشوائياً حسب موقعه وحسب قيمته.

3- التشفير وفك الشفرة Encryption and Decryption:

التشفير هو احد الوسائل الهامة المستخدمة للحفاظ على المعلومات الخاصة من السرقة أو التلاعب أو العبث فيها من قبل أشخاص غير مخولين. ظهر هذا العلم بعد التطور الحاصل في أنظمة الاتصالات وما وصلت إليه هذه الأنظمة من السرعة مما زاد في إرسال واستقبال الكثير من المعلومات وقواعد البيانات، أدى هذا التطور الايجابي إلى نمو وتكاثر تدخل الغير مخولين لإرسال واستلام هذه المعلومات وظهور مجال جديد يعرف بالقرصنة على هذه المعلومات من قبل الهواة وغيرهم، عليه ظهرت الحاجة لابتكار العديد من طرق التشفير المعقدة والاعتماد على مفاتيح تشفير صعبة المنال أو الكشف.[11]. والشكل (2) يوضح مخطط عملية التشفير وفتح الشفرة بواسطة مفاتيح التشفير.[12].



الشكل رقم (2): التشفير وفتح الشفرة بواسطة المفاتيح. [12]

ومن أبرز أنواع التشفير المستخدمة هي التشفير الانسيابي (Stream Cipher) وهو ان عملية التشفير تجري على مستوى الـ (bit) في النص الصريح أو النص المشفر، بينما يكون التشفير على مستوى البلوك والذي يعرف بـ (Block Cipher) لأن عملية التشفير تجري على مستوى الـ (byte) أو مجموعة من الـ (bits) (block of bits) في النص الصريح أو النص المشفر، وهناك نوعان من خوارزميات التشفير المعتمدة على مفتاح التشفير وهما:

1- Symmetric Algorithms: وهي الخوارزميات التي يكون فيها مفتاح التشفير يمكن الحصول عليه من مفتاح فك الشفرة والعكس صحيح، وفي أغلب هذه الخوارزميات يكون مفتاح التشفير هو نفسه مفتاح فك الشفرة.[11][12].

2- Public Key Algorithms: هذه الخوارزمية مصممة لكي يكون مفتاح التشفير مختلف تماماً عن مفتاح فك الشفرة، ومهما حاول محلل التشفير من عمليات فك الشفرة والحصول على النص الصريح فأنه يبقى هناك الكثير من المعلومات المشفرة، عليه تكون هذه الطريقة في التشفير مأمونة حسابياً (computationally secure) اذا كان بحوزة محلل الشفرة مصادر غير كافية للاعتماد عليها في فك الشفرة. [11][12].

ويعتبر أسلوب التشفير وكسر الشفرة باستخدام طريقة (XOR) من أكثر الطرق شيوعاً في الاستخدام وذلك لسهولة هذه الطريقة وسرعتها:

$Y = P \text{ xor } X$

حيث ان P هي القيم المتاحة، وX هي قيمة مولدة عشوائياً. [13]

3-1 طرق الحصول على النص الصريح ومفتاح فك الشفرة (Attack Methods):

هنالك أربعة طرق عامة وشائعة لمهاجمة النص المشفر، وكل طريقة تفترض على محلل الشفرة ان يكون على دراية كاملة بخوارزمية التشفير المستخدمة في عملية التشفير لكي يستطيع فك الشفرة والحصول على مفتاح التشفير وبالتالي الحصول على النص الصريح. [14] [15] وهذه الطرق هي:

1- Cipher text-only attack: يكون لدى محلل الشفرة في هذه الطريقة العديد من الرسائل المشفرة والتي تم تشفيرها بنفس مفتاح الشفرة:

$C_1=Ek(P_1), C_2=Ek(P_2), \ldots, C_i=Ek(P_i)$ .....(1)

والهدف هو الحصول إما على النصوص الصريحة ($P_1, P_2, \ldots, P_i$) أو على مفتاح التشفير (k) للحصول على النص الصريح لأي نص مشفر جديد P_{i+1} من $C_{i+1}=Ek(P_{i+1})$.

2- Known-plaintext attack: يكون لدى محلل الشفرة العديد من الرسائل المشفرة مع نصوصها الصريحة، وهي الطريقة قيد البحث والتحليل باستخدام الخوارزمية الجينية:

$P_1C_1=Ek(P_1), P_2C_2=Ek(P_2), \ldots, P_iC_i=Ek(P_i)$ .....(2)

والهدف هو الحصول على مفتاح التشفير (k) والحصول على النص الصريح لأي نص مشفر جديد P_{i+1} من $C_{i+1}=Ek(P_{i+1})$.

3- Chosen-plaintext attack: يكون لدى محلل الشفرة ليس العديد من النصوص المشفرة مع النصوص الصريحة فحسب بل يمتلك ويختار النصوص الصريحة التي تم تشفيرها:

$P_1C_1=Ek(P_1), P_2C_2=Ek(P_2), \ldots, P_iC_i=Ek(P_i)$ .....(3)

والهدف هو الحصول على مفتاح التشفير (k) والحصول على النص الصريح لأي نص مشفر جديد P_{i+1} من $C_{i+1}=Ek(P_{i+1})$.

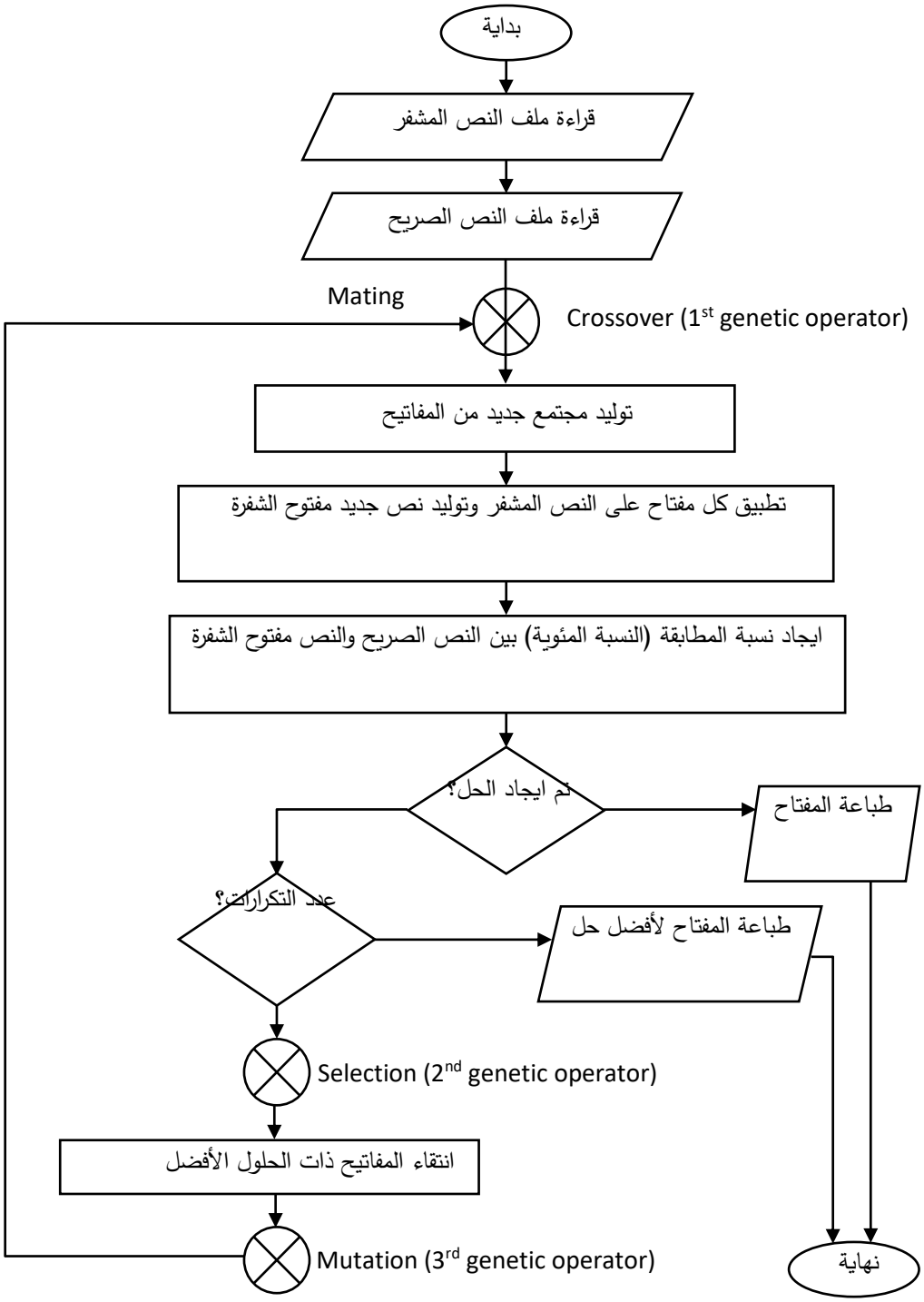
4- Adaptive-chosen-plaintext attack: يكون محلل الشفرة في هذه الطريقة لا يمتلك النصوص الصريحة التي قد تم تشفيرها فقط، بل إنه يمتلك نتائج تحليل للرسائل المشفرة السابقة.

4- الخوارزمية المقترحة:

تم في هذا البحث اقتراح الخوارزمية التالية (والتي تتضمن تطبيق الخوارزمية الجينية على النصوص)، إذ تم التعامل مع كل ملف نصي يمثل ملف النص المشفر أو ملف النص الصريح بعد تحميله للبرنامج بشكل متجه حرفي، وتم استخدام طريقة كسر الشفرة بطريقة خوارزمية الـ XOR لسهولتها أولاً ومن ثم هدف البحث هو إيجاد مفتاح التشفير بغض النظر عن الخوارزمية المستخدمة في التشفير. وخطوات هذه الخوارزمية بالشكل التالي:

1- قراءة الملف النصي المشفر وخزنه في متجه حرفي.
2- قراءة الملف النصي الصريح للنص المشفر وخزنه في متجه حرفي ثاني.
3- توليد المجتمع الابتدائي وملئه بقيم عشوائية من المفاتيح ويكون كل مفتاح على شكل متجه حرفي.
4- اجراء عملية التقاطع على المجتمع الابتدائي وتوليد مجتمع جديد اكبر حجماً.
5- تطبيق مفاتيح التشفير على النص المشفر، وخزن نتائج فك الشفرة في متجهات حرفية مقابلة لكل مفتاح.
6- مقارنة جميع نتائج فك الشفرة مع النص الصريح وإيجاد نسبة المطابقة لكل مفتاح وخزن هذه النسبة لكل مفتاح.
7- اختيار المفاتيح التي تمتلك أعلى نسبة تطابق للنص المشفر التي تم فك الشفرة بها مع النص الصريح.
8- إجراء عملية الطفرة الوراثية على المفاتيح المنتخبة، ويكون اختيار موقع الطفرة عشوائياً وقيمة الجين (الحرفية) عشوائية ايضاً.
9- إعادة تطبيق الخطوة 5 وباقي الخطوات لحين الحصول على المفتاح الأعلى نسبة تطابق أو الوصول الى شرط التوقف للخوارزمية (عدد تكرارات الخوارزمية).

والشكل رقم (3) يوضح المخطط الانسيابي لمراحل تطبيق وعمل الخوارزمية المقترحة:



الشكل رقم (3): المخطط الانسيابي للخوارزمية المقترحة

4-1 تطبيق الخوارزمية المقترحة:

تم في هذا البحث تطبيق الخوارزمية المقترحة على عدد من النصوص المشفرة ولوحظ أهمية وسرعة ودقة الخوارزمية الجينية في الأداء.

إذ تم التعامل مع كل نص مشفر والنص الصريح المقابل له على حدى، وتم أخذ 5 مفاتيح عشوائية ابتدائية، حيث تم إجراء عملية التقاطع الجيني عليها وتكوين 20 مفتاح جديد كعينة أولية للانطلاق بالخوارزمية، ثم تم كسر النص المشفر باستخدام هذه المفاتيح واستخراج نصوص مكسورة الشفرة، ومقارنة هذه النصوص مع النص الصريح واستخراج نسبة التطابق بين النصين باستخدام قانون النسبة المئوية (وهي عملية استخراج درجة اللياقة لكل مفتاح):

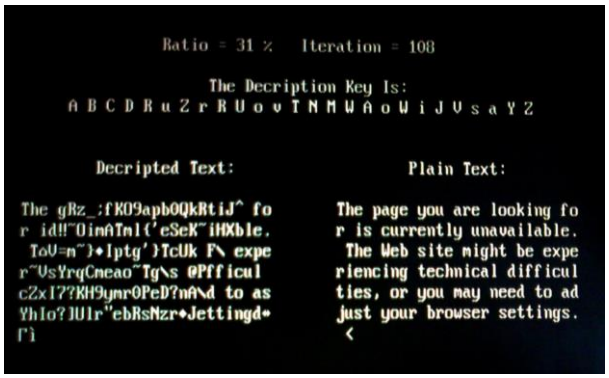
$$\text{النسبة} = \frac{\text{عدد الكلمات المتطابقة بين النصين المشفر والصريح}}{\text{عدد الكلمات الكلية في النص}} \times 100$$

تم ترتيب كل من نسب المطابقة والمفاتيح المقابلة لها –والتي تم استخدامها لفتح الشفرة– والنصوص مكسورة الشفرة ترتيباً تنازلياً، ومقارنة نسبة المطابقة للمفتاح الأول في الترتيب (المفتاح ذات الأعلى نسبة) مع النسبة المطلوبة التي تم إدخالها في بادئ الخوارزمية من قبل المستخدم. فإذا كانت هذه النسبة هي المطلوبة أو أعلى منها بمعنى تم التوصل الى الحل المطلوب والحصول على مفتاح التشفير، فيتم طباعة مفتاح كسر الشفرة مع النص المقابل له.

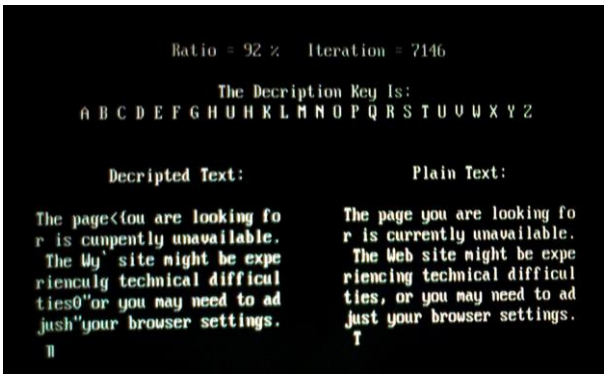
أما اذا كانت هذه النسبة اقل من المطلوبة، عندها تقوم الخوارزمية بأخذ أعلى خمسة نسب للخمسة المفاتيح الأولى حسب الترتيب، ومن ثم تنفيذ عملية الطفرة الوراثية لهذه المفاتيح باختيار موقع عشوائي ضمن المفتاح وتغيير قيمة الجين (الحرف) بحرف عشوائي ايضاً، وإعادة تنفيذ الخوارزمية مرة أخرى لحين الحصول الى المفتاح الذي نسبة المطابقة له مساوية او اكبر للنسبة المطلوبة أو يصل عدد تكرارات الخوارزمية للعدد المحدد من قبل المستخدم في بداية تنفيذ الخوارزمية.

والشكل رقم (4): يوضح نتائج تطبيق هذه الخوارزمية على إحدى النصوص المختارة.

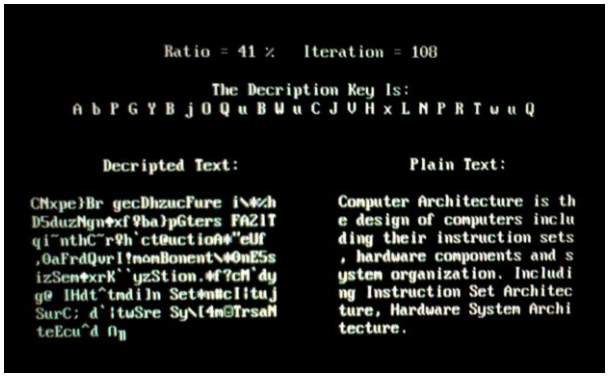
(a)



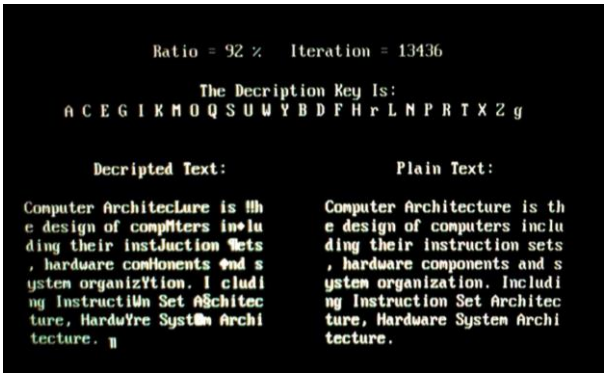
(b)



(c)



(d)



(e)

الشكل (4): يوضح نتائج تطبيق الخوارزمية المقترحة على نموذجين من النصوص المشفرة:

(a) نافذة إدخال النسبة المئوية كقيمة توقف للخوارزمية.

(b) نافذة الحصول على نتائج تطبيق الخوارزمية للنص الأول بقيمة توقف ≤ 30%

(c) نافذة الحصول على نتائج تطبيق الخوارزمية للنص الأول بقيمة توقف ≤ 90%

(d) نافذة الحصول على نتائج تطبيق الخوارزمية للنص الثاني بقيمة توقف ≤ 30%

(e) نافذة الحصول على نتائج تطبيق الخوارزمية للنص الثاني بقيمة توقف ≤ 90%

5- الاستنتاجات:

بعد عملية تطبيق الخوارزمية الجينية المقترحة على بعض من النصوص المشفرة، خلص البحث إلى جملة استنتاجات يمكن إجمالها بما يلي:

1) ان أسلوب تعامل الخوارزمية الجينية مع النصوص والكلمات والحروف (متغيرات السلسلة الحرفية) يقود الى إعطاء مزايا جديدة قد تقود في أحيان كثيرة إلى تغيير وجهات النظر في حل المشاكل المعقدة ذات مساحة البحث الواسعة من خلال التعامل مع هذه المتغيرات.
2) اظهر البحث من خلال التعامل مع النصوص إمكانية التقارب السريع للخوارزمية لإيجاد الحلول المرجوة وبنسبة خطأ قليلة أو تكاد تكون معدومة.
3) حسب ما أظهرته النتائج كما في الشكل (4)، الى ان التغيير الحاصل في النصوص من حيث الحجم لا يؤثر على زمن تنفيذ الخوارزمية بالتأثير الكبير رغم أن عدد تكرارات تنفيذ الخوارزمية يتزايد مع تزايد قيمة إدخال متغير التوقف لها.
4) استخدام عملية الطفرة الوراثية على المفاتيح المنتخبة ساعد وأظهر انتقال عملية البحث من فضاء عينة الى فضاء عينة آخر من مفاتيح كسر الشفرة والتي بالتالي أدت الى الحصول على مفتاح التشفير المطلوب.
5) تم تطبيق الخوارزمية المقترحة بتصميم برنامج بلغة C++ 3.1، حيث تم استخدام النصوص الانكليزية المشفرة فقط ولم يتم التطرق الى النصوص العربية لأن لغة البرمجة المستخدمة لا تدعم الحروف العربية، عليه يُقترح استخدام لغة برمجية تدعم الحروف العربية عملاً مستقبلياً.

المصـادر:

1) الخياط، صباح محمد أحمد وجنان عبد الوهاب الفيضي، (1988)، "الذكاء الاصطناعي: مفاهيمه–تقنياته–اساليب برمجته"، دار حنين، عمان.

2) Bala, J. K. Dejong, (1995), “Hybrid learning using Genetic Algorithms and decision trees for pattern classification”, IJCAI Conference, Montreal.

3) Salem, Abde-Badeeh M. and Abeer M. Mahmoud, (2002), “A Hybrid Genetic Algorithm Decision Tree Classifier”, IJICS, Vol. 2, No. 2, July: 1-12.

4) Prebys, Eric Krevice, (1997), "The Genetic Algorithm In Computer Science", MIT Undergraduate Journal of Mathematics, 165-170.

5) Song, Kai, Andrew Lim and Brian Rodrigues, (2003), "Sexual Selection for Genetic Algorithm", Artificial Intelligence Review, 19:123-152.

6) Koray, Korkut Bilal Alatas and Ali Karci, (2004), "Mining classification rules by using Genetic Algorithms with non-random initial population uniform operator", TURK J. Elec. Engin., Vol. 12, No. 1:43-52.

7) Mitra, Sushmita, (2002), "Data mining in soft computing framework: A Survey", IEEE Transactions on neural networks, Vol. 13, No. 1.

8) Tsun, Chang and Randy Chiao, (2003), "Multi Resolution Genetic Clustering Algorithm for Texture Segmentation", Image and Vision Computing 21:955-966.

9) Huang, Jeffery and Harry Wechsler, (1999), "Eye location using Genetic Algorithm", 2nd International conference on audio and video-based biometrics person authentication.

10) Wang, Wei-Yen and Yi-Hsum Li., (2003), "Evolutionary Learning of BMF Fuzzy-Neural Networks Using a Reduce Form Genetic Algorithm", IEEE, Vol. 33, No. 6, December: 966-976.

11) Bruce, Schneier, (1996), "Applied Cryptography, Protocols, Algorithms and Source Code InC", John Wiley and Sons, Inc. U.S.A.

12) Menezes, A., Van Oorshot, P., and Vanstone, S., (1996), "Handbook of Applied Cryptography", CRC Press.

13) Pallavi K. B., "Heuristic Search Cryptanalysis of The Zodiac 340 Cipher", Sam Jose State University, pp. 11, 2009.

14) Mohamed M., Al-Afari F., Bamatraf M., "Data Hiding by LSB Substitution Using Genetic Optimal Key-Permutation". International Arab Journal of e-Technology. Vol. 2, No. 1. January 2011.

15) Buscema, Massima, (2004), "Genetic doping algorithm (GenD): theory and application", Expert system, May, Vol. 21, No. 2.