Use Cutting Enzymes to Encode the Secret Message

Saadoun H. A.

College of Science

University of Al Mosul

Received on : 24/6/2013 **Accepted on 12/2/2014**

ABSTRACT

Researchers begin to explore new methods to manage the huge amount of data shared over the Internet, in order to use them more efficiently and safely, One of the ways that is raised nowadays is Steganography, The art of information hiding within electronic media. Steganography ways vary in their approaches for hiding information.

During this research designed algorithm that using restriction enzymes to cutting sequence of DNA, the fragments resulted was used to building symbols table (character , number , special character,…),by using symbols table the plain message (text ,sound ,image ,video) coding to DNA and sent it to recipient, the recipient build symbols table same way in the sender site and use function to known restriction enzymes, using restriction enzymes to cutting secret sequence DNA , by using the fragments of DNA and symbols table can retrieve the plain message (text ,sound ,image ,video).

Keyword: Algorithms , Enzymes , DNA

استخدام الانزيمات القاطعة لترميز الرسالة السرية

سعدون حسين عبدالله

كلية العلوم

جامعة الموصل

تاريخ استلام البحث : 2013\6\24 **تاريخ قبول البحث : 2014\2\12**

الملخص

اتجه الباحثون إلى استكشاف طرائق جديدة لإدارة الكم الهائل من البيانات المُشتركة عبر الانترنت من أجل استخدامها بشكل آمن وأكثر فاعلية, ومن تلك الطرائق التي برزت في العصر الحالي هي الكتابة المغطّاة (Steganography) وهو فن إخفاء المعلومات ضمن وسط، تتفاوت طرائق الكتابة المغطّاة في نهجها لإخفاء المعلومات.

خلال هذا البحث تم تصميم خوارزمية تستخدم الانزيمات القاطعة بأنواعها المختلفة لقطع قطع محددة من سلسلة DNA واستخدام القطع الناتجة لبناء جدول الرموز(حروف، ارقام، حروف خاصة،...)، بالاعتماد على هذا الجدول وعلى رسالة المرسلة (نصية، صوتية، صورية، فديوية) يتم بناء(تصنيع) سلسلة DNA سرية حاوية للرسالة السرية ، ترسل سلسلة الـ DNA الناتج الى المستلم وهو بدوره يعتمد على الخوارزمية السابقة لبناء الجدول ويستخدم دالة لمعرفة الانزيمات التي تقطع سلسلة DNA السرية بعدها قطعها بهذه الانزيمات ومعرفة القطع وبالاستعانة بجدول الرموز يتم استرجاع الرسالة السرية(نصية، صوتية ، صورية ، فديوية).

الكلمات المفتاحية: خوارزميات ، الأنزيمات ، DNA

1. مقدمة Introduction

في عالمنا هذا ذي الرقمية العالية تلعب شبكة الانترنت دوراً مهماً في نقل البيانات وتبادلها ومع ذلك ولأنها شبكة عالمية ووسيط عام فإن بعض البيانات (وخاصةً السرية منها) قد تتعرض إلى السرقة, أو النسخ, أو التعديل, أو التدمير. ولما كانت المعلومات هي ثروة أي منظمة مهما كان نوعها فهذا يجعل من المسائل الأمنية الأولوية القصوى, لاسيما للمنظمة التي تتعامل مع بيانات سرية[3,2,1]. وعلى هذا أُقترح حقلان بهدف تعزيز أمنية الاتصال: (التشفير Cryptography وإخفاء المعلومات Information/Data Hiding) وبالرغم من أن كليهما وجدا على حد سواء لحماية البيانات السرية فإن الفرق الأساسي بينهما هو في ظهور البيانات المنقولة[3].

2. إخفاء المعلومات Information Hiding

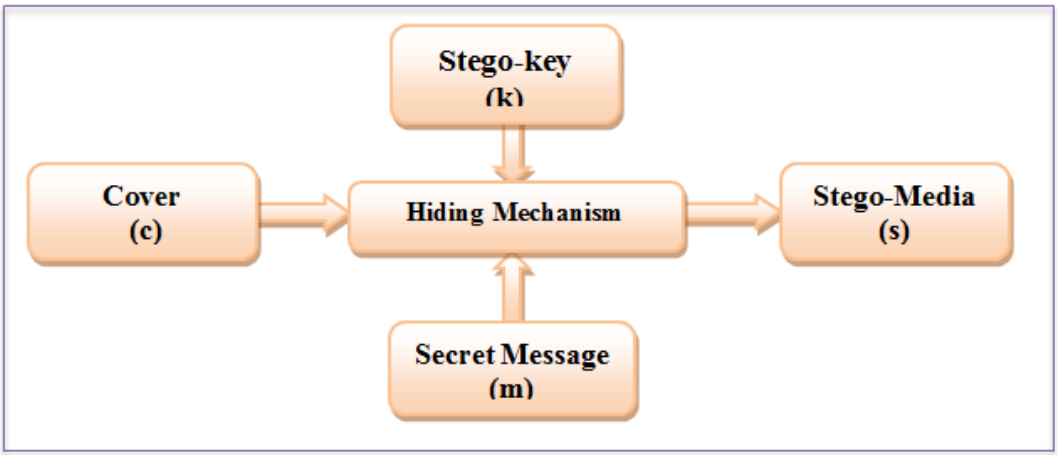
يعد أمن المعلومات اليوم احد أهم عوامل تكنولوجيا المعلومات والاتصالات. ووفقا لذلك فنحن بحاجة إلى القيام بإجراءات تحمي المعلومات السرية[4,5] .

إن إخفاء المعلومات (أو إخفاء البيانات) هي واحدة من التقنيات المتجددة باستمرار التي تهدف إلى توفير الأمن عن طريق إخفاء المعلومات السرية في محتويات الوسائط المتعددة بتغيير بعض المكونات غير الضرورية في الملف المضيف أو الملف الغطاء[7,6,5] .

ومع تقبل المجتمع للتقنيات الرقمية لإرسال الرسائل واستقبالها على نطاق واسع, ظهرت الحاجة إلى تقنيات توفر الأمنية لهذا التناقل وأبرز هذه التقنيات هي الكتابة المُغطّاة (Steganography) [3] .

هناك أربعة عناصر في الإخفاء تمثل بالشكل (1) وهي كالآتي [9,8].

- الناقل (Carrier): يسمى غطاء "Cover" ويشار إليه بالحرف c.
- الرسالة السرية (Secret message): التي سوف يتم إخفائها ويشار إليها بالحرف m.
- وسط الإخفاء (Stego-media): هو إخراج الناتج بعد الإخفاء ويشار إليه بالحرف s.
- مفتاح الإخفاء (Stego-key): ويشار إليه بالحرف k.

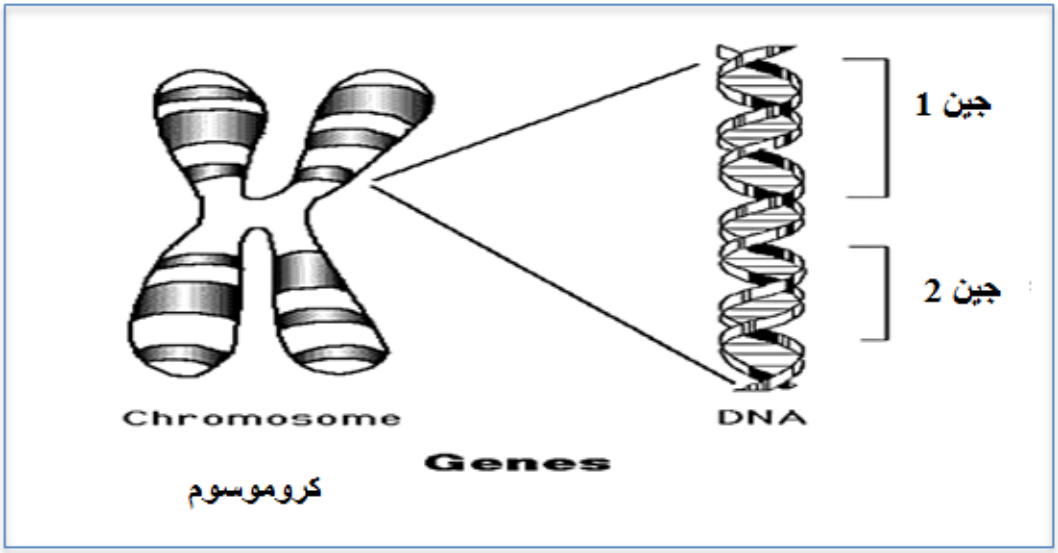


الشكل(1) عناصر الإخفاء

3. الحامض النووي الرايبي منقوص الأوكسجين Deoxyribonucleic Acid (DNA)

يمكن تعريفه بأنه التركيبة الكاملة للتعليمات الخاصة بتكوين الكائن الحي، ويحتوي على البصمات التي تحدد كل مكونات الخلية وأنشطتها طوال حياة الكائن الحي، وهذه العوامل الوراثية موجودة على أشرطة من الحامض

النووي (DNA) الحلزونية الشكل, فضلا عن جزيئات البروتين، وهما معا يكونان وحدات تسمى الكروموسومات، وعلى هذه الكروموسومات توجد المورثات أو الجينات (Genes) وهي التي تحدد كل صفات الكائن الحي. والشكل (2) يوضح أن جزءاً صغيراً من الكروموسوم قد يحوي مجموعة من الجينات المختلفة [11,10] .



شكل (2) : يمثل مقطع للمادة الوراثية (DNA) والتي تتضمنها الجينات

إن الحامض النووي (DNA) يتكون من شريطين ملتفين على بعضهما بحيث يشبهان السلم الملتوي، وأنه يتكون من أربعة أنواع من القواعد النتروجينية هي, الأدنين (A) والثايمين (T) والسايتوسين (C) والكوانين (G), وتتكرر هذه القواعد ملايين أو مليارات المرات في جميع أجزاء الحامض النووي (DNA)[11,12].

اكتشف الحامض النووي (DNA) أول مرة العالمان (Watson) و (Crick) سنة (1953) وقام (Coenberg) بعزل أول أنزيم في أنبوبة اختبار للحامض النووي (DNA) في سنة (1958). وفي عام (1966) بيّن كل من (Nirenberg) و (Mathaei) و(Ochoa) أن كل ثلاث نيوكليوتيدات (رموز من DNA) تشفر حامضاً أمينياً وحيداً من مجموع الحوامض الامينية. كما تم نشر تفاصيل لخوارزمية (Needleman–Wunsch) الخاصة بمقارنة السلاسل سنة (1970), وتبعها الإعلان عن إنشاء بنك بيانات البروتين (Protein Data Bank) في سنة (1973) وكان يحوي معلومات وبيانات لأسماء الجينات وطرائق تحليلها. وفي سنة (1972) أجريت أول تجربة ناجحة لاستنساخ الحامض النووي (DNA) في كاليفورنيا, وفي مطلع الثمانينات حصلت شركة (Cetus) الأمريكية على براءة إختراع لكونها استطاعت مضاعفة الحامض النووي باستخدام تقنية (PCR). وفي سنة (1990) تم إكمال مشروع الجينوم البشري الذي يعد ثروة علمية هائلة واكتشافاً رائعاً. أما في عام (2000) فقد أعلن العلماء بأنهم استطاعوا فك رموز الجينات الإنسانية, وجاء هذا نتيجة جهد فريق مؤلف من مجموعة تضم (1000) باحث, وأكد هؤلاء الباحثون أن هذا العمل سيحدث ثورة في علم التشخيص ومعالجة الأمراض وأنّ ترجمة (3) مليار رمز من الحامض النووي (DNA) البشري يعدّ أحد معالم التاريخ العلمية العظيمة [11].

4. المعلوماتية الحيوية Bioinformatics

تُعرّف المعلوماتية الحيوية بأنها استخدام الحاسوب لمعالجة المعلومات الحيوية. وهو علم ركيزته الأساسية قواعد بيانات المعلومات الحيوية بمكوناتها الرئيسة (الجينات والبروتينات)، ويجمع عددا من العلوم الأخرى بهدف

الاستفادة من هذه المعلومات كعلوم الرياضيات، والحاسوب، والإحصاء، والطب، والكيمياء. ويمكن أن نلخص تعريف المعلوماتية الحيوية بأنها تطبيق التقانة الحاسوبية والمعلوماتية في إدارة المعلومات الحيوية. ومثال ذلك تحليل المعلومات الحيوية (الجينات والبروتينات) باستخدام الحاسوب والتقنيات الحاسوبية الحديثة[14,13] .

1.4. أهداف المعلوماتية الحيوية

للمعلوماتية الحيوية أهداف رئيسة ثلاثة هي[14,13]:

1. تطوير تقنيات جديدة وبناء خوارزميات تساعد على تحصيل المعلومات من مجموعة ضخمة من البيانات.
2. تحليل وتفسير الأنماط المختلفة من البيانات التي تتضمن سلاسل الأحماض الأمينية والقطع والبنى البروتينيّة.
3. تطوير وتنفيذ أدوات تساعد على إدارة فعّالة للأنماط المختلفة من المعلومات.

تتضمّن المعلوماتية الحيوية المعالجة البارعة، والتقصي، والتنقيب عن بياناتData Mining لمتتابعات DNA. إن تطوير تقنياتِ الخَزْن والتقصي عن متتابعات DNA أدّيا إلى التقدّمِ الكبير جدا في الجانب التطبيقيِ في علوم الحاسوب، خصوصاً في مجالات خوارزميات تقصي سلسلة حروف String Searching Algorithms وتَعَلّم الماكنة Machine Learning ونظرية قاعدة البيانات Database Theory. إن التقصي عن متتابعات DNA يُعنى بدراسة حدوثَ سلسلة حروف داخل سلسلة أكبر مِنْ الحروف، للتقصي عن متتابعات معيّنة مِن النكليوتيدات Nucleotide . وتستعمل سلاسل ماركوف لتشخيص الأشياء الشاذّة Anomaly ، وتصليح البيانات Repair Data، وتقييم سلامةَ البيانات Assess Data Integrity [15].

5. قواعد البيانات Data Base

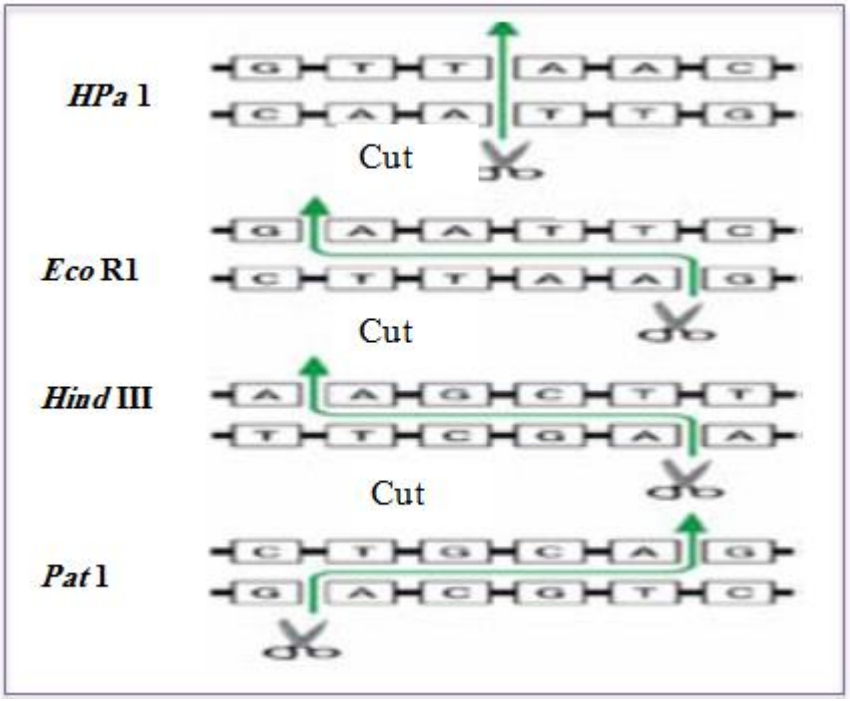
قواعد البيانات عبارة عن جداول او أرشفة الكترونية تستخدم لخزن وتنظيم البيانات بطريقة يمكن معها استرجاع المعلومات المطلوبة بسهولة باستخدام معايير بحث مختلفة .ان الهدف الرئيس من تطوير قواعد البيانات هو لتنظيم البيانات في مجموعة هيكلية من القيود لتسمح بعد ذلك من استرجاع المعلومات .كل قيد يسمى ايضا مدخل ، ويحوي عددا "من الحقول التي تحتوي على عناصر البيانات الحقيقية، على سبيل المثال حقول مثل الأسماء، أرقام الهاتف، العناوين والتواريخ .ولغرض استرجاع قيد معين من قواعد البيانات، فان المستخدم يدخل المعلومة المطلوبة وتسمى هذه العملية بالاستعلام. مثلا البحث اسم شخص معين بالاعتماد على رقم هاتفيه او أي معلومة من ضمن الحقول المدخلة ضمن قاعدة البيانات. من قواعد البيانات المتعلقة بالعمليات البايولوجية (الجزيئية) ما تم تصميمه من قبل الحكومة الامريكية وهي قاعدة البيانات (National Center for Biotechnology Information) NCBI

تحتوي هذا المركز على بنك الجينات(Gene Bank) التي تتضمن الالاف من تتابعات DNA و RNA والبروتينات. ان نظام استرداد للبيانات ضمن موقع NCBI هو Entrez حيث يمثل نقطة بداية البحث في قواعد NCBI ويكون من خلاله البحث سهلا ولكن محدود. يكون النظام مدخلا للبحث في قواعد الحوامض النووية مثل بنك الجينات وكذلك قواعد البروتينات فضلا عن وجود روابط تخص المراجع PubMed وروابط تخص التصنيف وغيرها.

يتم عمل تحديث على بنك الجينات كل 18 شهر حيث يضاف اليه التتابعات الجديدة وتحذف بعض التتابعات المكررة [16] .

6. انزيمات التقييد Restriction Enzymes

لاشك ان كان لكل حي طرق دفاعية مختلفة تحميه من غارات الاعداد وهجوم المعتدين، فالبكتريا هي احدى هذه الكائنات والتي لها اعداء كثيرة ومن اهم اعدائها الفايروسات المختلفة. وقد قامت بعض من البكتريا بإنتاج المقصات مهمتها تدمير الفايروسات. وتقوم هذه المقصات او ادوات القطع بقص الحامض النووي للفايروسات وبذلك يشل عمله ويبطل مفعوله. يبين الشكل(3) بعض اشكال الانزيمات القاطعة [17] .



الشكل(3) امثلة لبعض الانزيمات القاطعة

ان كل انزيم قاطع يعبر عن مقص خاص لقطع DNA في نقطع محددة، ويتعرف الانزيم القاطع على مكان القطع حسب تسلسل DNA للقطعة. فكل انزيم قاطع يقطع في تسلسل محدد. فمثلا انزيم القطع *Hpa*1 يقطع عندما يجد 6 من القواعد النيتروجينية s بهذا التسلسل GTTTAAC بينما الانزيم القاطع *Eco*R1 يقطع عندما يجد 6 من القواعد النيتروجينية بهذا التسلسل GAATTC. يوجد نوعين رئيسيين من المقصات، النوع الاول يقص شريط DNA بشكل مستقيم Blunt ends كما في *Hpa*1 والنوع الثاني يقص بشكل متعرج Staggered cuts وبالتالي يجعل طرفي DNA المقطوع قابلة في اللصق قطعة غريبة من DNA فيها كما في *Eco*R1، والجدول(1) يوضح تسلسل مواقع القطع Recognition sites لبعض الانزيمات القاطعة الاتية[17]،وبما ان القواعد النيتروجينية للنهايات الناتئة متكاملة فإنها يمكن ان ترتبط مع بعضها مرة اخرى عن طريق الاواصر الهيدروجينية المتكون بين القواعد المتكاملة حيث ترتبط النهايات اللاصقة اولا عن طريق الاواصر الهيدروجينية المتكونة من القواعد المتكاملة مما يسهل عملية لحم الكسور بواسط انزيم لايجيز الدنا DNA Ligase لتكوين جزيئة دنا هجينة صنعية [18] .

الجدول(1) تسلسل مواقع القطع Recognition sites لبعض الانزيمات القاطعة

الانزيم القاطع	المصدر	المقطع الذي يميزه القطع
BamHI	*Bacillus amyloliquefaciens H*	GGATCC
EcoRI	*Escherichia coli RY13*	GAATTC
HaeIII	*Haemophilus aegyptius*	GGCC
HindIII	*Haemophilus nfluenza Rd*	AAGCTT
HpaI	*Haemophilus parainfluenzae*	GTTAAC
HpaII	*Haemophilus parainfluenzae*	CCGG
MboI	*Moraxella bovis*	GATC
NotI	*Nocardia otitidis-caviarum*	GCGGCCGC
SfiI	*Streptomyces fimbriatus*	GGCCNNNNNGGCC

بالنظر لاكتشاف اعدادا كبيرة من انزيمات التقيد كان لابد من ايجاد نظام تسمية على اساسه كل الانزيمات المكتشفة. يعتمد نظام التسمية الحالي على الاقتراح الذي قدمه سميث وناثانس في 1973 مع اجراء بعض التغييرات البسيطة لتسهيل عملية التسمية ، تكون انزيمات التقييد حسب هذا النظام بأسماء مكونة من ثلاثة حروف تكتب بصورة مائلة او يوضع تحته خط. يدعى الانزيم التقييد المعزول من البكتيريا الهيموفلس بارا انفلونزا *Hin* بـ *influenza Hemophilus* [18] .

1.6. القطع المحددة Restriction Fragments

عندما يضاف الانزيم القاطع الى محلول به شريط من DNA فانه يقطعه الى عدة قطع وليس قطعتين فقط، وكل قطعة مقطوعة بالإنزيم تسمى قطعة محددة، ويختلف الطول هذه القطع حسب المسافة بين كل قطعة واخر. ولكن يجب ان تكون كل قطعة محددة لها نفس الحجم في كل نوع من الكائنات الحية، وقد انشا العلماء خريطة تسمى خريطة القطع المحددة لكثير من الكائنات الحية حيث تبين هذه الخريطة مكان القطع ومحلها وطولها والتتابع [17].

7. الخوارزمية المقترحة Suggested Method

1.7. خوارزمية التضمين Embedding Algorithm

المرسل: تتضمن خوارزمية بناء تتابع DNA الحاوية للرسالة السرية عدة خطوات :-

الخطوة الاولى: يكون التعامل مع موقع NCBI لقراءة سلسلة DNA بطول 6000 قاعدة (Seqn) او باي طول لجين يكون الاتفاق عليها مسبقاً، او بالإمكان توليدها باستخدام دالة التوليد العشوائي (Randseq) من برنامج الماتلاب اصدار 2012b .

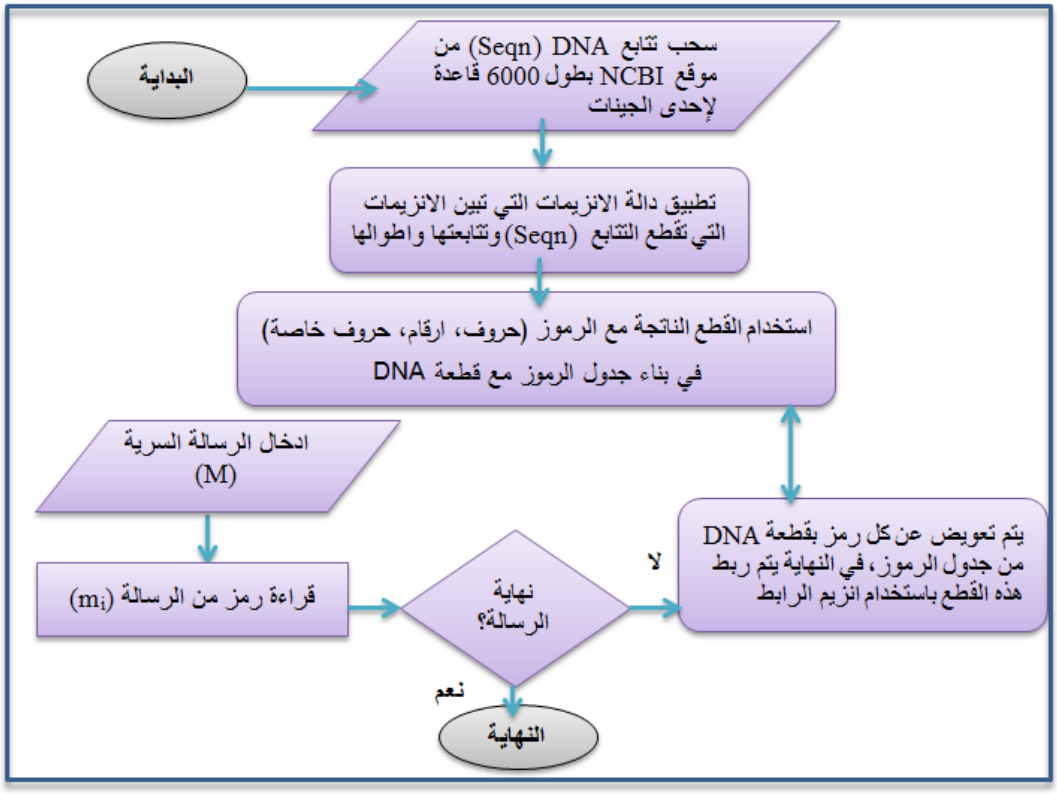
الخطوة الثانية: ادخال السلسلة (Seqn) الى دالة لإيجاد الانزيمات التي تقطع هذه السلسلة والقطع التي تقطعها واطوال كل قطعة واماكن القطع.

الخطوة الثالثة: استخدام القطع الناتجة (fragments) من الانزيمات لبناء جدول الرموز (symbol table) وذلك بإعطاء كل قطعة ناتجة الى رمز معين (حروف، ارقام، حروف خاصة،...).

الخطوة الرابعة: إدخال الرسالة السرية(M) (نصية، صوتية، صورية، فديوية).

الخطوة الخامسة: بالاعتماد على الجدول الذي تم بنائه في الخطوة الثالثة يتم تحويل كل رمز من الرسالة السرية M الى ما يقابها من قطع DNA.

الخطوة السادسة: يتم ربط القطع الناتجة مع بعضها بأنزيم الرابط وبالأخير انتاج سلسلة DNA (Seqsec) وارسالها الى الطرف المستلم. والشكل(4) يوضح المخطط الانسيابي لعملية بناء سلسلة DNA سري.



الشكل(4) المخطط الانسيابي لعملية بناء سلسلة DNA سري

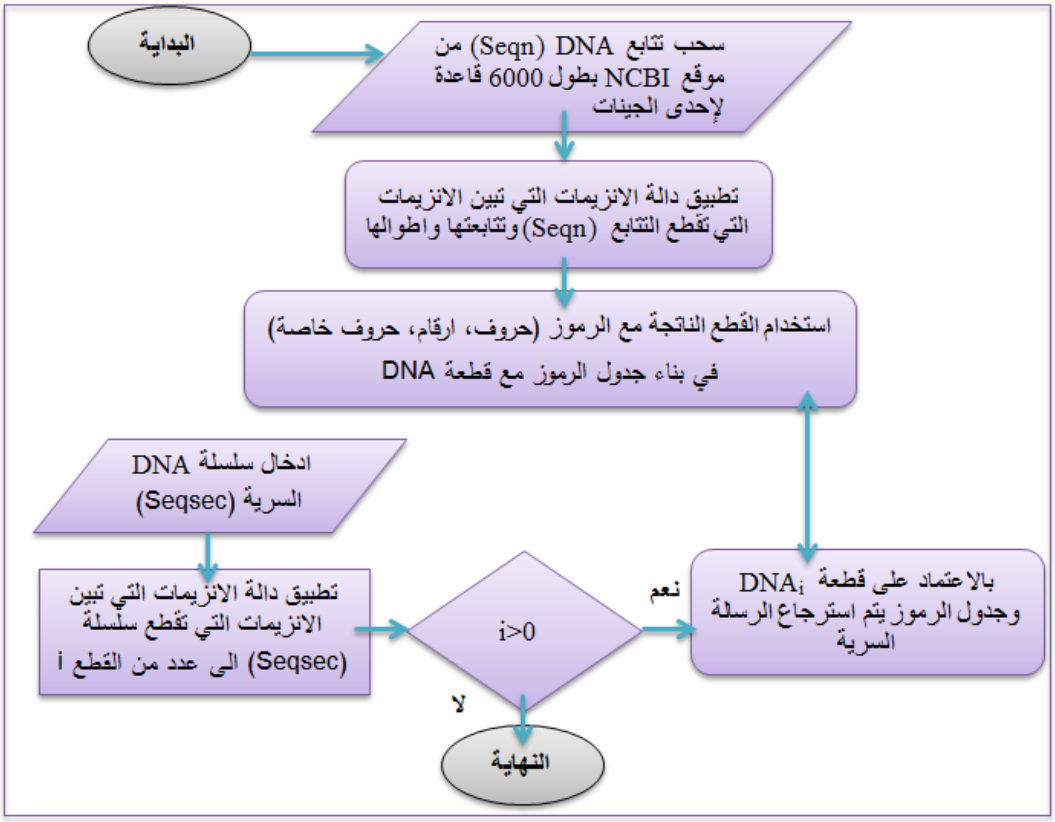
2.7. خوارزمية الاسترجاع Retrieving Algorithm

المستلم: تتضمن خوارزمية استرجاع الرسالة السرية من سلسلة DNA: بان يقوم المستلم بنفس خطوات الثلاثة الاولى التي تم التطرق اليها عند المرسل.

الخطوة الاولى: يستخدم المستلم الانزيمات القاطعة لقطع سلسلة المستلمة (Seqsec) الى قطع (Fragments).

الخطوة الثانية: قراءة القطعة الناتجة من الخطوة السابقة ومقارنتها مع جدول الرموز وبناء string من الرموز.

الخطوة الثالثة: حسب نوع الرسالة تعاد صيغتها ويتم واسترجاع الرسالة السرية. والشكل (5) يوضح المخطط الانسيابي لعملية استرجاع الرسالة السرية من سلسلة DNA.



الشكل(5) يوضح المخطط الانسيابي لعملية استرجاع الرسالة السرية من سلسلة DNA

8. الاستنتاجات Conclusions

1. ان استخدام الطرق الحديثة وخاصة الدمج بين تقنيات علوم الحاسوب وعلم البايولوجي الجزيئي (bioinformatics) اعطت اهمية عالية في اخفاء المعلومات.

2. ان الطريقة المقترحة اعطت سرية عالية للرسالة السرية المرسلة لعدة اسباب منها:

- ان استخدام الانزيمات القاطعة لترميز الرموز المختلفة اعطت احتمالية عالية جداً لكون وجود ما يقارب 1500 انزيم قاطع وبمختلف الاطوال من قطع DNA .
- كون اطوال قطع DNA ومحتوياتها مختلفة فعند بناء سلسلة DNA السرية الحاوية على الرسالة السرية يصعب استرجاع الرسالة منها من قبل المتطفل.

المصادر

[1] Shejul, A. A. and Kulkarni, U. L., (2011), "A Secure Skin Tone Based Steganography using Wavelet Transform", International Journal of Computer Theory and Engineering, Vol. 3, No. 1, pp. 16-22.

[2] Reddy, T. CH. S., Prasad, D., and Reddy, B. V., (2012), "Protection through Oblivion: An Image Based Steganography ", Int. J. Computer Technology & Applications, Vol. 3(1), pp. 420-424.

[3] القصاب، اسماء موفق،(2013)،"تقنية الكتابة المغطاة في الصور الكسورية باعتماد مجموعة جوليا"، رسالة ماجستير، قسم علوم الحاسوب، كلية علوم الحاسوب والرياضيات، جامعة الموصل، العراق.

[4] Tejinderpal S., And Simpel J., (2011)," Image Steganography Using Mixed Channel Replacement ", International Journal Of Engineering Sciences Issn: 2229-6913 Issue, Vol. 4, pp. 62 – 70.

[5] فاضل، سميرة عباس،(2013)،" تنفيذ تقنيات الكتابة المغطاة على منصة لينكس"، رسالة ماجستير، قسم علوم الحاسوب، كلية علوم الحاسوب والرياضيات، جامعة الموصل، العراق.

[6] Kriti S., and Pradeep K. S., (2010), " A Variant of LSB Steganography for Hiding Images in Audio", International Journal of Computer Applications, Vol. 11, No.6, pp. 12 – 16.

[7] M. Indra S. R., M. Purushotham R., K. and Subba R., (2012), " Different Medias of Steganography- An Emerging Field of Network Security" , International Journal of Computer Science and Information Technologies, Vol. 3 , No. 2 , pp. 3517-3522.

[8] النعيمي، سماح فخري عزيز،(2011)،" تحليل الإخفاء بالاعتماد على تقنيتي آلة المتجه الداعم ومميز فيشر الخطي في الصور الملونة"، رسالة ماجستير، قسم علوم الحاسوب، كلية علوم الحاسوب والرياضيات، جامعة الموصل، العراق.

[9] الغريري، شهد عبد الرحمن حسّو، (2003)، " تصميم نظام حماية هجين و تطبيقه على النصوص"، رسالة ماجستير، قسم علوم الحاسوب، كلية علوم الحاسوب والرياضيات، جامعة الموصل، العراق.

[10] قاسم، عمر صابر (2009)." تطبيق التقنيات الذكائية في المعلوماتية الحيوية "، أطروحة دكتوراه، كلية علوم الحاسوب والرياضيات، جامعة الموصل.

[11] سواحل، وجدي عبد الفتاح ,(2001)، "استخدام الهندسة الوراثية في التحقيق الجنائي : أساليب وتطبيقات", المركز القومي للبحوث , القاهرة.

[12] Calvino, M., Gomez, N. and Mingo, L.F.,(2007), "DNA Simulation of Genetic Algorithms: Fitness Computation", International Journal,Information Theories & Applications, Vol.14.

[13] زمزوم، فاطمة محمود حسن،(2011)،" استخدام النمذجة الماركوفية في المعلوماتية الحيوية"، رسالة ماجستير، قسم علوم الرياضيات، كلية علوم الحاسوب والرياضيات، جامعة الموصل، العراق.

[14] Shoemaker, J.S. and Lin, S.M. (2005). "Methods of Microarray Data Analysis IV", Springer Science + Business Media, Inc.

[15] الخياط ، باسل يونس ذنون وسليمان ، مثنى صبحي(2011) ،" التنبؤ عن الحالات المطرية في مدينة الموصل"، بحث مقدم إلى المؤتمر العلمي الثالث / لعلوم الرياضيات ، 27 – 28 نيسان 2011 ، جامعة الزرقاء ، الاردن.

[16] الخفاجي، زهرة محمود وابراهيم، علي عبد الحافظ ،(2012)،" المعلوماتية الحيوية "،معهد الهندسة الوراثية والتقنيات الحيوية، جامعة بغداد، العراق.

[17] عبيده، علي ابراهيم ومحمود، احمد عبدالفتاح،(2009)،" اساسيات التقنية الحيوية"، كلية الزراعة، سابا باشا، جامعة اسكندرية، مصر.

[18] البكري، غالب حمزة ، (1991) ، "مبادئ الهندسة الوراثية"، كلية العلوم ، جامعة البصرة، العراق.