Text Encryption Using OTP Keys From Randomly Generated DNA Sequences

Dr. Kinda Abu Kassem*
Tayseer Salman**

(Received 12 / 5 / 2019. Accepted 29 / 8 / 2019)

□ ABSTRACT □

Cryptographic keys is one of the main issues when designing symmetric encryption algorithms. Several encryption algorithms have used Pseudo Random Number Generators (PRNG). The problem with PRNGs is that they are deterministic where usage of the randomly generated key will be repeated and it's occurrence depends on its length since the generation of the key is done based on mathematical mechanism or by using sources of randomness like thermal or chemical sources. The randomness property of DNA introduces a solution for the problem of key generation, which results from random succession of DNA nucleotides. Besides, DNA offers keys with lengths as required. Real DNA sequences, taken from public genetic databases exploited as a source for encryption keys. This paper demonstrates the use of OTP encryption key extracted from random generated DNA sequences as a substitute for DNA sequences taken from public genetic databases and performance evaluation of the proposed algorithm by applying the algorithm on text files of different sizes.

Keywords: symmetric encryption, One Time Pad, PRNG, public genetic databases

* **Associate Professor, Department of Computer Engineering and Automatic Control, Faculty of Mechanical and Electrical Engineering, Tishreen University, Lattakia, Syria.**
** **PhD student, Department of Computer Engineering and Automatic Control, Faculty of Mechanical and Electrical Engineering, Tishreen University, Lattakia, Syria.**

تشفير النصوص باستخدام مفاتيح OTP من تسلسلات DNA مولدة عشوائياً

الدكتورة كندة سليمان أبو قاسم *

تيســـير عزت ســـلمـان **

(تاريخ الإيداع 12 / 5 / 2019. قُبِل للنشر في 29/ 8 / 2019)

□ ملخّص □

يعتبر توليد مفاتيح التشفير أحد المسائل الأساسية في تصميم خوارزميات التشفير المتناظر. تم استخدام مولدات الأرقام العشوائية PRNG من قبل العديد من خوارزميات التشفير. تكمن مشكلة مولدات الأرقام العشوائية بأنها تتصف بالحتمية أي أن المفتاح المولد عشوائياً سوف يتكرر، وحدوث التكرار يحدده طول مفتاح التشفير، لان التوليد يتم بناءً على آلية رياضية أو يمكن استخدام مصادر للعشوائية مثل مصادر حرارية أو كيميائية. تقدم العشوائية الكبيرة DNA حلاً لمشكلة توليد المفاتيح والتي تنتج من التتالي العشوائي لأسس DNA أو النيكليوتيدات؛ كذلك يقدم مفاتيح بأطوال حسب المطلوب. تم استخدام تسلسلات DNA حقيقية مأخوذة من قواعد بيانات جينية عامة كمصدر للحصول على مفاتيح التشفير. تقدم هذه الورقة كيفية استخدام مفاتيح تشفير OTP مستخلصة من تسلسلات DNA تم توليدها عشوائياً وذلك كبديل عن تسلسلات DNA المأخوذة من قواعد البيانات الجينية العامة وتقييم الأداء لخوارزمية التشفير المستخدمة من خلال تطبيق الخوارزمية على ملفات نصية بحجوم مختلفة.

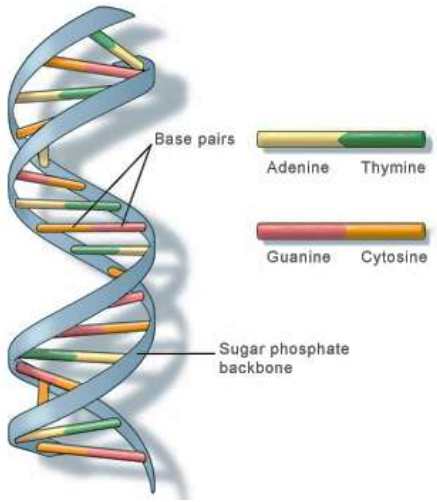
الكلمات المفتاحية: تشفير متناظر، One Time Pad، PRNG، قواعد البيانات الجينية العامة.

* **أستاذ مساعد – قسم الحاسبات والتحكم الآلي– كلية الهندسة الميكانيكية والكهربائية– جامعة تشرين– اللاذقية– سورية.**

** **طالب دكتوراة – قسم الحاسبات والتحكم الآلي– كلية الهندسة الميكانيكية والكهربائية– جامعة تشرين– اللاذقية– سورية.**

مقدمة:

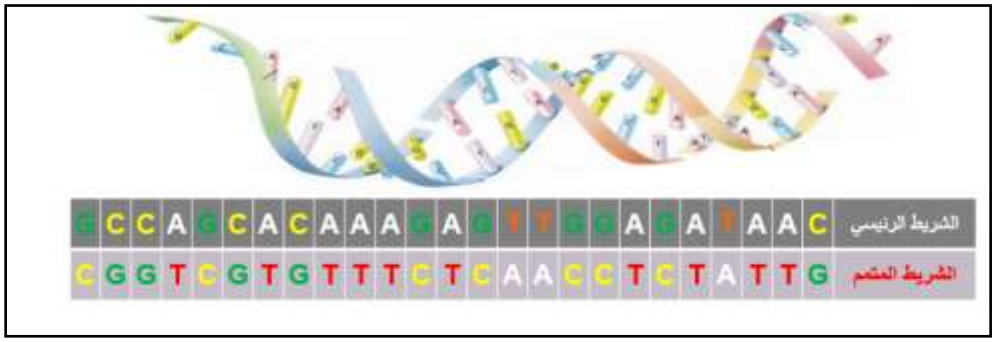
إن استخدام مفاتيح تشفير لمرة واحدة OTP أو التشفير المضمن [1]، يؤمن حماية البيانات بشكل غير قابل للكسر حيث أن مفتاح التشفير OTP يستخدم لمرة واحدة فقط خلال عملية التشفير وفك التشفير الواحدة بشرط أن يكون هذا المفتاح عشوائياً بشكل حقيقي TRNG [2]. يستخدم مولد الأرقام العشوائي الحقيقي مصدر غير حتمي لتوليد العشوائية. معظم مولدات الأرقام العشوائية الحقيقية تعمل من خلال قياس العمليات الطبيعية غير القابلة للتنبؤ مثل استخدام التقلبات العشوائية الناتجة عن دوران الأقراص الصلبة أو من خلال ضغطات المستخدم على لوحة المفاتيح [3]. أدى التطور الكبير في سَلسَلة الجينات سواء كانت بشرية أو حيوانية أو نباتية إلى ظهور مصدر آخر للعشوائية ومناسب لتطبيقات الحوسبة ويمكن ملاءمته لتشكيل مفاتيح التشفير واستخدامها مع خوارزميات التشفير التقليدية أو استنباط خوارزميات جديدة تم تسميتها بالتشفير المعتمد على DNA [4]. حيث تتصف تسلسلات DNA الطبيعية بعشوائية في تسلسل قواعدها الكيميائية والتي تسمى بالنيكليوتيدات وهي باختصار الثايمين T والأدنين A والغوانين G والسيتوزين C وهي عبارة عن سكريات خماسية تشكل الحمض النووي الريبي منقوص الاكسجين او DNA، كما هو موضح في الشكل(1) والذي يبين البنية ثلاثية الأبعاد لجزيء DNA وهي البنية التي وضعها كل من Watson وCrick عام 1953. نتج عن مشروع الجينوم البشري قاعدة بيانات جينية ضخمة من تسلسلات DNA وكذلك سلسلة جينات الأنواع الأخرى من بكتريا ونباتات وهذا أمن مصدرا كبيراً لتسلسلات DNA التي يمكن استخدامها أو استخدام أجزاء منها كمفاتيح للتشفير[5].



الشكل 1: البنية الفراغية لجزيء DNA وفقاً لتصور Watson وCrick

بعد عملية السَّلسَلَة للصبغي أو لجزيء DNA، يمكن تخزين التسلسلات الناتجة رقمياً من خلال القراءة الضوئية لهذه التسلسلات الحقيقية ومن ثم تحويلها الى الشكل الرقمي وحفظها في قواعد البيانات الجينية العامة حيث أنه من السهل تحويل تسلسلات DNA الى الصيغة الرقمية من خلال قواعد التحويل بين رموز أو حروف DNA والترميز الثنائي كما في الشكل(2).

تنتج تسلسلات DNA من عملية السلسلة وهي عملية تلي استخلاص DNA من خلايا كائن حي سواء حيواني أو نباتي وقراءة تتالي النيكليوتيدات على شريط DNA وتسجيلها ضمن ملف نصي غالباً كما في الشكل 2.



الشكل 2. تسلسلات *DNA*

عند تخزين DNA رقمياً يتم اعتباره مكونا من شريط واحد من النيكليوتيدات كون الشريط الآخر هو المتمم ويحوي نفس المعلومات الوراثية ويمكن استنتاج الشريط المتمم من أي تسلسل DNA من خلال قاعدة المتمم، الجدول (1).

الجدول 1: قاعدة المتمم

الرمز	الرمز المتمم
A	T
T	A
G	C
C	G

أما الترميز الثنائي لقواعد DNA (النيكليوتيدات) الأربعة A–T–C–G فيتم من خلال قواعد التحويل والتي يتم تطبيق إحداها عند تحويل تسلسلات DNA الى الصيغة الثنائية كما في الجدول (2) والشكل (3).

الجدول 2: قواعد ترميز حروف DNA ثنائياً

	1	2	3	4	5	6	7	8
A	00	00	01	01	10	10	11	11
T	11	11	10	10	01	01	00	00
G	01	10	00	11	00	11	01	10
C	10	01	11	00	11	00	10	01

تسلسل	A	T	A	G	A	G	T	C	A	G	A	....	C	T	A	G	A	C
المقابل	00	11	00	10	00	10	11	01	00	10	00	....	01	11	00	10	00	01

الشكل 3. ترميز تسلسل *DNA* ثنائياً

أهمية البحث وأهدافه:

تكمن أهمية البحث في اعتماد التشفير باستخدام تسلسلات DNA المولدة عشوائياً كمفاتيح سرية لمرة واحدة بدون الاعتماد على قواعد البيانات الجينية العامة، والهدف تحقيق تشفير للنصوص يتصف بالسرعة والأمان.

طرائق البحث ومواده:

- بيئة تطوير JetBrains PyCharm Community Edition 2018.2.4
- لغة Python 3.7
- قواعد البيانات الجينية العامة
- حاسب مكتبي HP Intel® Core™i3 -7100 CPU @ 3.9 GHz 4GB
- حاسب محمول Intel® Celeron® CPU 1.10GHz Dual Core
- حزمة اختبارات العشوائية NIST Randomness Tests

1- الدراسات المرجعية Literature review

هناك العديد من الدراسات المرجعية التي اعتمدت على تصميم خوارزميات للتشفير باستخدام DNA وحصلت على المفاتيح من تسلسلات DNA طبيعية من قواعد البيانات الجينية العامة [6]، [7]، [8]، [9]، [10]، [11]، [12]، [13].

- اعتبرت الدراسة [6] بأن قواعد البيانات الجينية العامة هي الحل لمشكلة تأمين مفاتيح التشفير المضمنة أو OTP باعتبار أن هذه التسلسلات تشكل مفاتيح عشوائية وحقيقية ويمكن الوصول اليها مجاناً وبأي وقت وبالتالي يمكن الاعتماد عليها كمصدر للتسلسلات العشوائية في تشفير النصوص.

قدمت الدراسة [7] طريقة لتشفير البيانات النصية من خلال استخدام تسلسلات DNA ذات أطوال كبيرة يتم الحصول عليها من قواعد البيانات الجينية العامة، هذه التسلسلات تستخدم كحامل للبيانات من خلال عملية فهرسة DNA. أما مفتاح التشفير فيتم توليده من خلال مولد عشوائي حيث يشفر النص المرمز ثنائياً مع المفتاح المولد عشوائياً للحصول على نص مشفر بالشكل الثنائي بعد ذلك يتم تحويله الى ترميز DNA ومن ثم البحث ضمن التسلسل الذي تم اخذه من قاعدة البيانات الجينية العامة لفهارس كل بايت من النص المشفر ويتم ارسال الفهارس الى المستقبل. تتميز هذه الطريقة بأنها لا تقوم بإرسال البيانات المشفرة وانما تقوم بإرسال فهارس البيانات على تسلسل DNA موجود لدى طرفي الارسال والاستقبال بالإضافة الى المفتاح. ربما تقلل هذه الطريقة من حجم البيانات المشفرة الواجب ارسالها لكن يجب ارسال تسلسل DNA اللازم لعملية الفهرسة بالإضافة الى مفتاح التشفير وفك التشفير.

استخدمت الدراسة [8] خوارزمية تشفير متناظر وعمليات المصفوفات وتقنية XOR واعتمدت على مفاتيح OTP والتهجين مع تقنية DNA لتقليل التعقيد الزمني وهو الموضوع الذي ركزت عليه الخوارزمية ربما تكمن سيئة هذه الطريقة بأن طول مفتاح OTP هو طول المفتاح المولد مضروباً بطول الرسالة بالشكل الثنائي مما ينتج نصاً مشفراً أطول 5 مرات من النص الأصلي على الأقل.

في الدراسة [9] تم استخدام التشفير المتناظر بالاعتماد على DNA حيث يتم توليد مفتاح التشفير عشوائياً وتشفير المفتاح نفسه قبل إرساله الى جهة المستقبل. أيضاً اعتماد تسلسلات DNA المولدة بناءً على النص الأصلي كحامل للمعلومات، حيث يتم تشفير التسلسل الناتج عن النص الأصلي مع المفتاح المولد عشوائياً. لم تقم هذه الدراسة بتقييم أداء الخوارزمية المستخدمة.

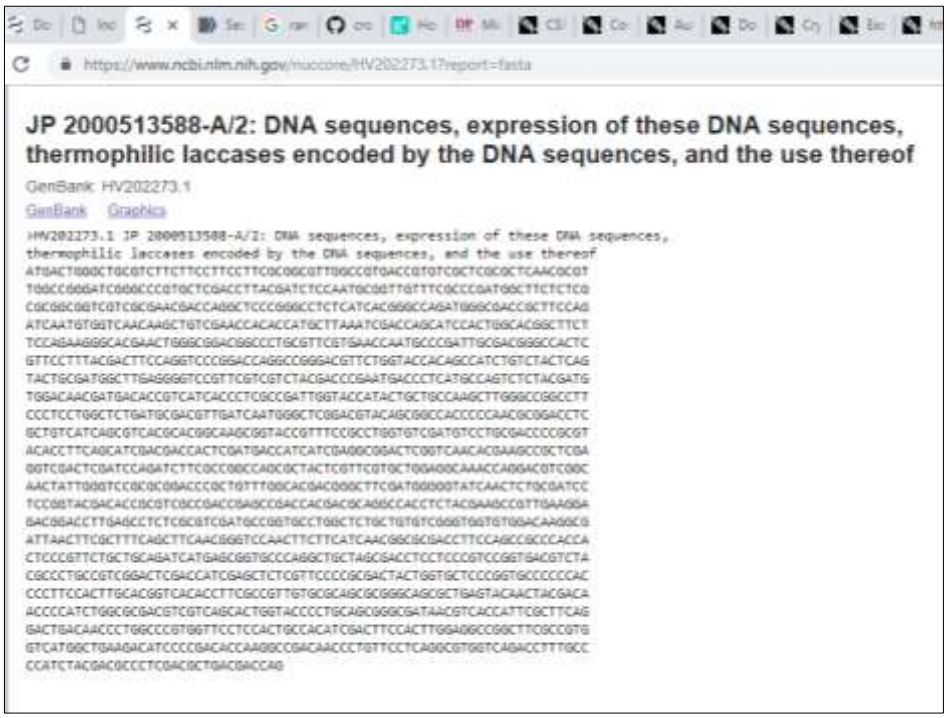
استخدمت الدراسة [10] مفاتيح OTP بالاعتماد على تسلسلات DNA من قاعدة بيانات جينية عامة وإجراء ترميز Huffman على النص الأصلي ثم اجراء تشفير للنص الناتج مع المفتاح حيث يقسم النص الأصلي المرمز الى عدد من القطع حجم كل منها يساوي حجم المفتاح، ثم إضافة بيانات مزيفة الى النص المرمز بعد ذلك يتم إجراء XOR بين النص المرمز والمفتاح OTP. وفي النهاية يتم ارسال النص المشفر على شكل تسلسل DNA. الملاحظ في هذا النوع من التشفير وحسب التحليل الوارد في هذه الدراسة نص بطول 40 حرف مع الفراغات ومفتاح OTP بطول 20 محرف DNA كان النص المشفر الناتج بطول 1137 محرف DNA وهذا يعود الى الرموز الوهمية التي تمت إضافتها الى النص المشفر. يعتبر النص المشفر أطول بمقدار 28 ضعف أكبر من النص الأصلي وهذا قد يكون غير عملياً عندما يزداد حجم النص الصريح ليصبح من مرتبة الكيلو بايت مما سيؤثر على عملية الارسال وفك التشفير بالأخص عندما تكون الموارد محدودة من ذاكرة ومعالجة وعرض حزمة لقناة الاتصال.

اعتمدت الدراسة [12] على فهرسة DNA حيث يتم اعتماد تسلسل من قاعدة بيانات جينية عامة كحامل للبيانات ومفتاح أيضاً مستخلص من قاعدة البيانات الجينية العامة. بعد أخذ النص الأصلي وتشفيره بالمفتاح OTP يتم أخذ موقع كل بايت على التسلسل الحامل وتخزين هذه المواقع ثم ارسال مواقع البايتات أو الفهارس الى طرف الاستقبال ليتم فك التشفير بعملية معاكسة. من مساوئ التشفير بفهرسة DNA هو وجوب توفر التسلسل الحامل والمفتاح لدى طرفي الارسال والاستقبال وهذا ما يزيد من عبء تأمين سرية نقل المفاتيح وتسلسلات الفهرسة الى طرف الاستقبال، أيضاً زيادة التعقيد الزمني من خلال عمليات البحث ضمن تسلسل كبير للحصول على مواقع البايت من المعلومات والذي يمكن أن يتكرر في العديد من المواقع على التسلسل. لكن بالمقابل تعتبر ميزة هذه الطريقة بأنه لا يتم إرسال المعلومات الفعلية وإنما مواقعها (فهارسها) على صبغي متفق عليه من قبل المرسل والمستقبل وهذا يعتبر أحد الصعوبات أمام محللي الشيفرة في حال عدم معرفتهم للصبغي المستخدم.

اعتمدت الدراسة [13] على نمذجة للمبدأ الأساسي في علم الاحياء الجزيئي Central Dogma of Molecular Biology والذي ينص على أن التعليمات الوراثية اللازمة لتشكيل البروتينات والتي بدورها تعتبر مسؤولة عن تشكيل صفات الكائن الحي تمر بعدة مراحل حيث تبدأ العملية بنسخ التعليمات الوراثية من على موقع معين من جزئ الـ DNA بعملية تسمى النسخ Transcription وينتج عن ذلك شريط RNA يسمى الناقل أو mRNA شريط mRNA مطابق لشريط DNA حيث أنه يحمل المعلومات الوراثية نفسها مع فارق أن القاعدة A تستبدل بالقاعدة U. تستخدم كل ثلاثة رموز من mRNA لإنتاج حمض أميني وتسلسل الاحماض الامينية الناتجة تشكل البروتينات وتسمى هذه العملية الترجمة Translation. في النهاية ينتج النص المشفر على شكل تسلسل من البروتينات. تحدث عملية معاكسة في طرف الاستقبال مع أن المبدأ الأساسي لعلم الاحياء الجزيئي غير عكوس في الطبيعة.

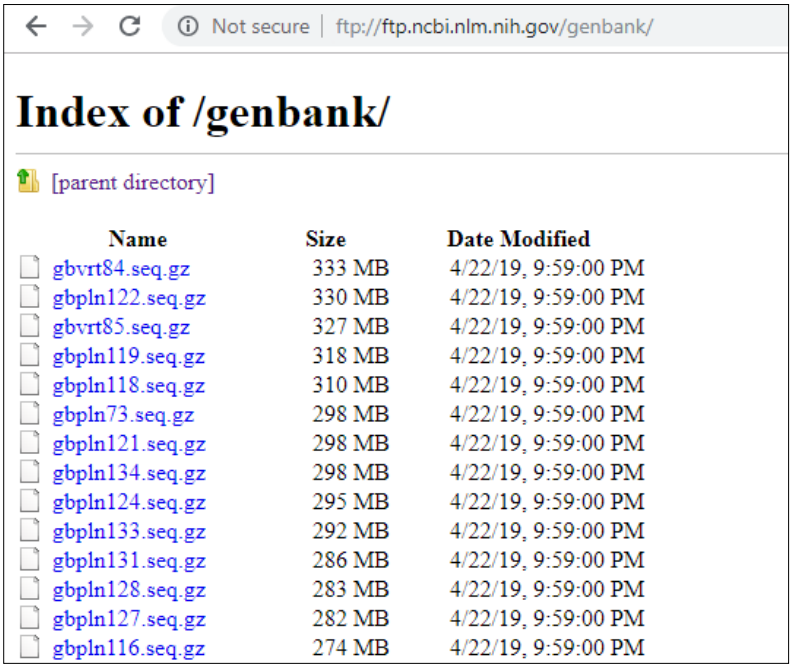
ملخص وملاحظات حول الدراسات المرجعية المعتمدة Summary of Literature Review

من خلال استعراض وتحليل الأوراق البحثية المعتمدة كدراسات مرجعية، تبيّن أنها اعتمدت على قواعد البيانات الجينية العامة للحصول على تسلسلات DNA المستخدمة كمفاتيح تشفير، لكن لم يتم ذكر الصعوبات التقنية المرافقة لذلك. من الجيد أن نعرف بأن التسلسلات الموجودة على مواقع قواعد البيانات الجينية العامة مثل الموقع الأكثر شهرة واستخداماً NCBI [14] هي عبارة عن ملفات مضغوطة تتراوح حجومها بين عدة كيلو بايت الى المئات من الغيغابايت تحوي ملفات نصية وبداخلها المعلومات الجينية التي هي تسلسلات DNA أو تسلسلات RNA أو تسلسلات بروتينات. يحوي كل ملف نصي على معلومات إضافية كما في الشكل (4) حول الكائن الذي تم أخذ التسلسلات منه ومعلومات عن جهاز السلسلة المستخدم الخ.



الشكل 4: عينة من تسلسل DNA المصدر https://www.ncbi.nlm.nih.gov/nuccore/HV202273.1?report=fasta

تشير معظم الدراسات التي تستخدم قواعد البيانات الجينية العامة الى خوارزمية بسيطة للحصول على التسلسلات حيث تتم عملية مزامنة بين المرسل والمستقبل على استخدام تسلسل أو صبغي معين بناءً على معرفة هذا التسلسل ضمن قاعدة البيانات الجينية العامة. لكن لم يتم التطرق في هذه الدراسات الى العملية أو سلسلة العمليات المرافقة لاختيار أحد التسلسلات ثم فك ضغط الملف الذي تم اختياره بناءً على المعرِّف ID وحذف المعلومات غير الضرورية وكم من الزمن والمعالجة اللازمة لذلك خصوصاً أن بعض الملفات بحجوم من مرتبة عشرات ومئات الميغابايت واحياناً بضع عشرات من الغيغا بايت كما في الشكل (5)؛ عدا المشاكل المتعلقة بتوفر خدمة قواعد البيانات الجينية العامة من عدمها.


Not secure | ftp://ftp.ncbi.nlm.nih.gov/genbank/

Index of /genbank/

[parent directory]

Name	Size	Date Modified
gbvrt84.seq.gz	333 MB	4/22/19, 9:59:00 PM
gbpln122.seq.gz	330 MB	4/22/19, 9:59:00 PM
gbvrt85.seq.gz	327 MB	4/22/19, 9:59:00 PM
gbpln119.seq.gz	318 MB	4/22/19, 9:59:00 PM
gbpln118.seq.gz	310 MB	4/22/19, 9:59:00 PM
gbpln73.seq.gz	298 MB	4/22/19, 9:59:00 PM
gbpln121.seq.gz	298 MB	4/22/19, 9:59:00 PM
gbpln134.seq.gz	298 MB	4/22/19, 9:59:00 PM
gbpln124.seq.gz	295 MB	4/22/19, 9:59:00 PM
gbpln133.seq.gz	292 MB	4/22/19, 9:59:00 PM
gbpln131.seq.gz	286 MB	4/22/19, 9:59:00 PM
gbpln128.seq.gz	283 MB	4/22/19, 9:59:00 PM
gbpln127.seq.gz	282 MB	4/22/19, 9:59:00 PM
gbpln116.seq.gz	274 MB	4/22/19, 9:59:00 PM

الشكل 5: عينة من الملفات المضغوطة للتسلسلات التي يوفرها موقع ncbi.nlm.nih.gov /genbank

الدراسة المقترحة Proposed Method

تهدف الدراسة المقترحة الى تصميم خوارزمية تشفير متناظر لتشفير النصوص بالاعتماد على التشفير باستخدام DNA. سيتم توليد التسلسلات المطلوبة بناءً على نسب محددة من كل من الرموز (A–T–C–G) بحيث يكون التسلسل الناتج عشوائيا؛ سيتم استخدام التسلسل الناتج كمفتاح OTP يستخدم لمرة واحدة لعملية تشفير وفك تشفير واحدة ثم يستبعد كي لا يتم استخدامه مرة ثانية. لزيادة الأمان؛ تستخدم قاعدة لتحويل التسلسل الناتج الى الترميز الثنائي وبشكل عشوائي حيث تستخدم قاعدة من بين ثمان قواعد وهذا يؤمن حماية من كشف المفتاح المستخدم.

التشفير Encryption

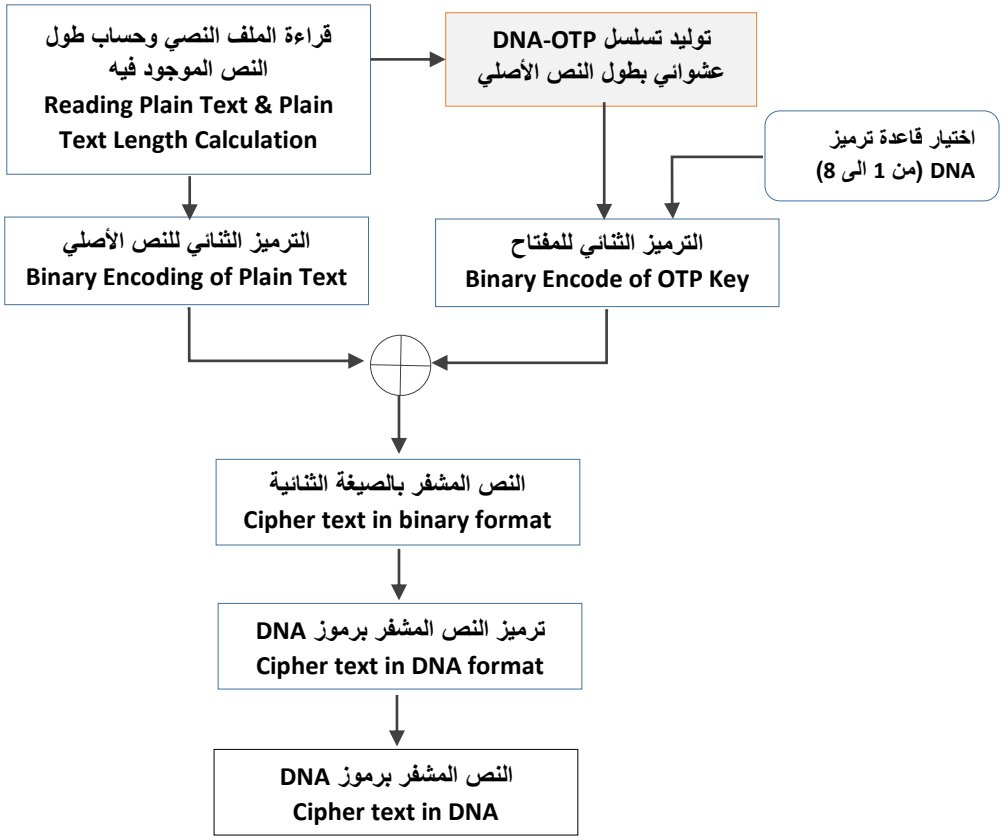
1- توليد تسلسل DNA بطول النص الأصلي حيث أن هذا التسلسل هو عبارة عن مفتاح التشفير OTP واختبار فيما إذا كان قد تم توليده سابقا ليستبعد.

2- ترميز النص الأصلي ثنائياً.

3- ترميز OTP المولَّد ثنائياً.

4- تشفير النص المرمز ثنائياً مع المفتاح المرمز ثنائياً من خلال إجراء عملية XOR عليهما لينتج النص المشفر.

5- ترميز النص المشفر برموز DNA وفق قاعدة التحويل التي تم اختيارها عشوائياً حسب الجدول (2). يوضح الشكل (4) المخطط الانسيابي لخوارزمية التشفير.



الشكل 4. مخطط خوارزمية التشفير

فك التشفير Decryption

يوضح الشكل (5) مخطط فك التشفير بعملية معاكسة للتشفير وفق الخطوات التالية:

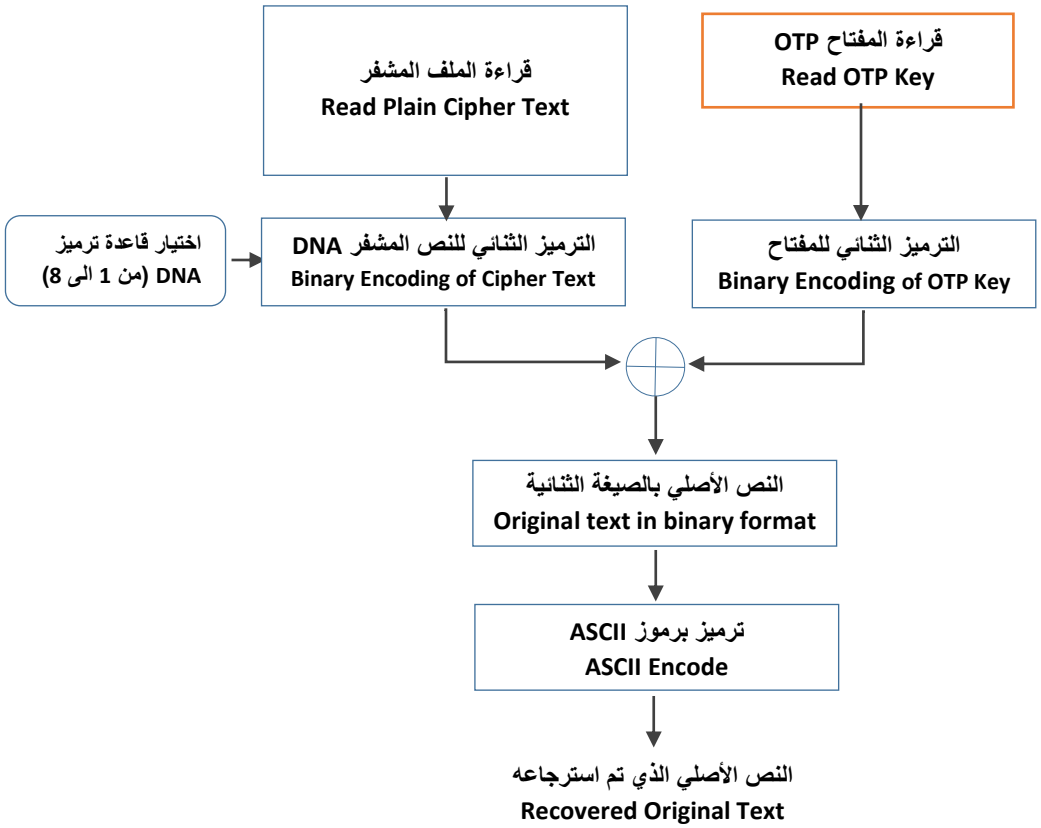
1- استقبال النص المشفر الذي هو عبارة عن تسلسل DNA.

2- ترميز هذه التسلسل ثنائياً وفق قاعدة الترميز التي تم اختيارها عشوائياً من الجدول (2).

3- ترميز المفتاح OTP الذي جرى تبادله مسبقاً مع طرف الاستقبال بالترميز الثنائي.

4- إجراء عملية XOR بين النص المشفر والمرمز ثنائياً والمفتاح OTP.

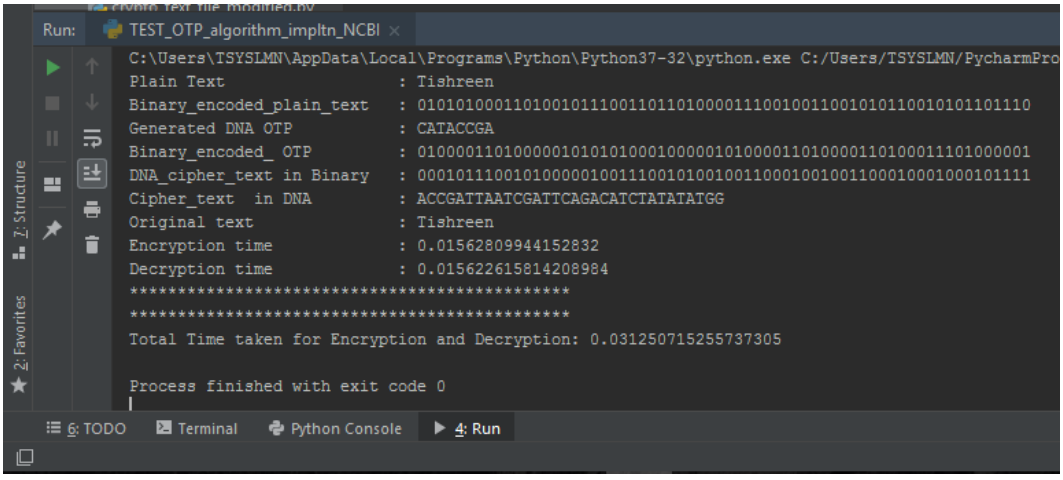
5- تحويل النص الثنائي الناتج والذي تم فك تشفيره الى النص الأصلي.



الشكل 4. مخطط خوارزمية فك التشفير

توضيح عمليتي التشفير وفك التشفير:

بفرض أن الملف الذي نود تشفيره يحتوي على العبارة Tishreen والتي تمثل النص الأصلي. تم أخذ لقطة شاشة لخرج البرنامج الذي قام بعملية التشفير وفك التشفير، الشكل (5).



الشكل 5: لقطة شاشة لخرج البرنامج الذي قام بعمليتي التشفير وفك التشفير

Plain Text (النص الأصلي) : **Tishreen**

Binary encoded plain Text (النص الأصلي المرمز ثنائياً):

01010100011010010111001101101000011100100110010101100101011101110

Generated DNA OTP is (مفتاح التشفير لمرة واحدة والذي تم توليده عشوائياً): **CATACCGA**

Binary Encoded OTP (مفتاح التشفير لمرة واحدة مرمز ثنائياً):

01000011010000010101010001000001010000110100001101000111010000001

Cipher text in Binary (النص المشفر بالترميز الثنائي):

00010111001010000010011100101001001100010010011000100010001011111

Cipher text in DNA النص المشفر مرمز برموز:

ACCGATTAATCGATTCAGACATCTATATATGG

Original text (النص الأصلي) : **Tishreen**

Encryption time (زمن التشفير) : 0.015627145767211914

Decryption time (زمن فك التشفير) : 0.01562356948852539

Total Time (الزمن الكلي للتشفير وفك التشفير): 0.031250715255737305

بالنظر إلى النتيجة السابقة والتي هي خرج البرنامج لعمليتي التشفير وفك التشفير لعبارة "Tishreen" والتي هي مؤلفة من ثمانية محارف، فكان طول المفتاح OTP أيضاً بثمانية محارف DNA وهي "**CATACCGA** " أما النص المشفر برموز DNA فكان " **ACCGATTAATCGATTCAGACATCTATATATGG** " وهو بطول 32 محرف. بعد الحصول على كل من المفتاح والنص برموز DNA ، تتم عملية ترميز ثنائي لكل من سلسلة المفتاح والنص الأصلي بهدف إجراء عملية XOR عليهما فينتج النص المشفر على شكل سلسلة ثنائية من الاصفار والواحدات بطول 64 بت وهذا الرقم نتج بسبب أن عملية الترميز الثنائي لكل من النص الأصلي والمفتاح تمت باستخدام 8bit لكل محرف مما أنتج سلسلتين من الاصفار والواحدات بطول 8*طول النص الأصلي = 64 لكن عند ترميز النص المشفر والذي هو بالصيغة الثنائية الى رموز DNA وفق قاعدة عشوائية تم اختيارها حسب الجدول (2) انخفض طول النص المشفر الى النصف عند تحويله الى رموز DNA حيث أن كل قاعدة DNA (نيكليوتيد) يتم ترميزها برمزين ثنائيين، ونفس القاعدة المستخدمة في ناحية التشفير يجب استخدامها في ناحية فك التشفير لضمان الحصول على النص الأصلي نفسه الذي قمنا بتشفيره.

تم إجراء اختبارات للخوارزمية المستخدمة والتي تم تطبيقها باستخدام لغة البرمجة Python 3.7 وبيئة التطوير Pycharm Community Edition 2018.2.4. أجريت الاختبارات على جهازي حاسب مختلفين بالعتاد الصلب، الأول جهاز حاسب محمول SMASUNG لديه المواصفات Intel® Celeron® CPU 1.10GHz Dual Core 2GB RAM والجهاز الثاني عبارة عن حاسب مكتبي HP Intel® Core™i3 −7100 CPU @ 3.9 GHz 4GB RAM. يوضح الجدول (3)، الاختبارات على ملفات نصية مختلفة الحجوم وقياس الزمن الوسطي لعمليتي التشفير وفك التشفير بالثانية:

الجدول 3: اختبار الخوارزمية المقترحة على جهازي حساب مختلفين بالعتاد الصلب من حيث زمن التشفير وفك التشفير بالثانية

م	حجم الملف	طول المفتاح DNA OTP	HP Core ™ I3 3.9 GHz		SAMSUNG Dual Core 1.10 GHz	
			زمن التشفير بالثانية	زمن فك التشفير بالثانية	زمن التشفير بالثانية	زمن فك التشفير بالثانية
.1	1KB	1024	0.01562237	0.01565933	0.12499570	0.06249690
.2	2KB	2048	0.03121113	0.03125429	0.23435354	0.12500143
.3	3KB	3072	0.04688191	0.03122234	0.45310831	0.15624594
.4	4KB	4096	0.04690194	0.04690504	0.51560878	0.21873474
.5	4.327KB	4423	0.04814553	0.04739274	0.32813119	0.21877360
.6	5KB	5120	0.05635711	0.05435456	0.28126454	0.18750882
.7	6.457KB	6601	0.07813382	0.06250381	0.46877908	0.32813620
.8	8KB	8192	0.09373450	0.06250286	0.84373545	0.42185544
.9	16KB	16384	0.17189168	0.17192602	1.42184591	0.89058256
0	32KB	32768	0.29690504	0.25005245	2.43742847	1.82806181
1	64KB	65536	0.59384322	0.48441839	4.62487435	3.62488174
2	68KB	69632	0.60944366	0.51564717	4.46898722	3.68768811
3	128KB	131072	1.15636706	0.96888327	8.42163276	7.04664897
4	256KB	262144	2.28148603	1.95333337	17.78070831	14.60892343
5	512KB	524288	4.59423470	3.87540888	39.31128239	29.82719612
6	1000KB	1024000	9.17287182	7.86017656	74.70079803	59.09191679
7	2000KB	2048000	18.8769946	17.1893215	159.83878302	118.54319167
8	4000KB	4096000	43.3006381	33.7948503	345.74074053	363.46760010

تقييم أداء الخوارزمية المقترحة

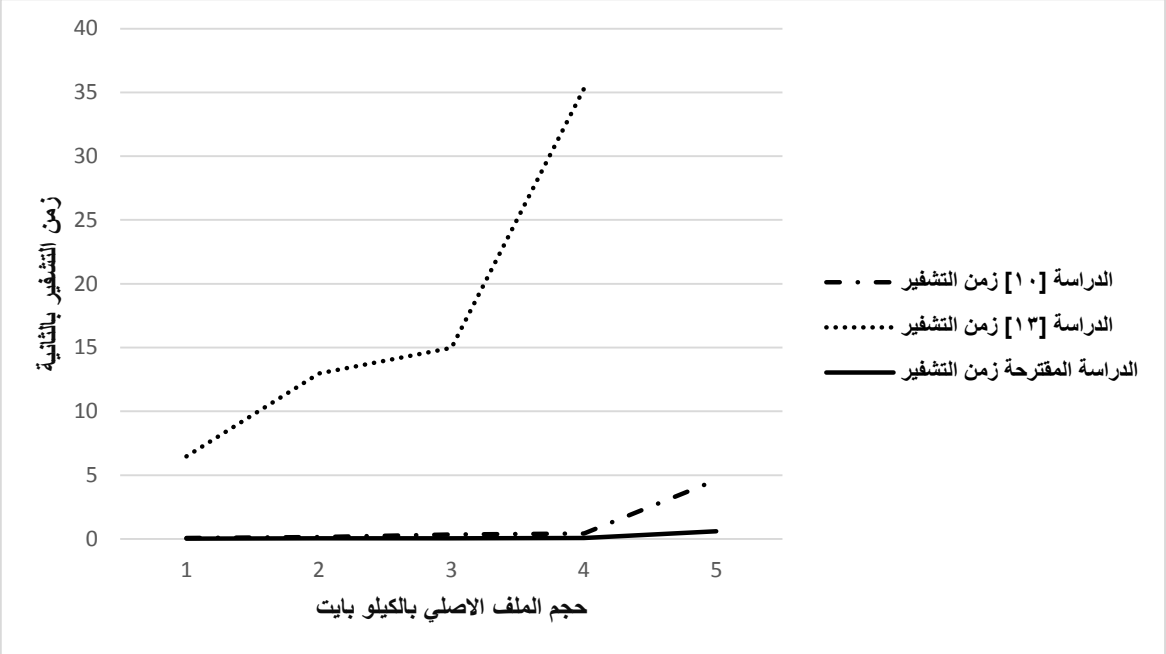
تقييم الأداء من ناحية سرعة التشفير وفك التشفير

بالمقارنة بين أداء الخوارزمية المقترحة والخوارزميات المعتمدة في الدراستين [10] و[13] حيث اكتفت الأولى بتشفير ثلاثة ملفات والثانية بتشفير ملفين من دون ذكر زمن فك التشفير في الدراسة [10]. تمت الإشارة في الجدول (3) وبخط ثخين إلى أزمنة التشفير التي توافق حجوم الملفات التي جرى اختبارها من قبل الخوارزمية المقترحة مع حجوم الملفات التي تم اختبار زمن تشفيرها في هاتين الدراستين حيث تم تلخيص هذه المقارنة في الجدول (4)، حيث تم اعتماد الزمن الذي تم قياسه على الحاسب المكتبي HP للدراسة المقترحة كون نفس المواصفات للعتاد الصلب

المستخدم مع الدراسة [13]. يبين الشكل (5) المخططات التي توضح العلاقة بين زمن التشفير وحجم الملف الأصلي وذلك كمقارنة بين الدراسة المقترحة والدراستين [10] و[13]

الجدول 4: مقارنة بين أزمنة التشفير للخوارزمية المقترحة مع أزمنة التشفير بالثانية للدراستين [10] و [13].

مقارنة بين أزمنة التشفير للخوارزمية المقترحة مع دراستين مرجعيتين سابقتين (الزمن بالثانية)						
حجم الملف [KB]	الدراسة [10]		الدراسة [13]		الدراسة المقترحة (على الحاسب HP)	
	زمن التشفير	زمن فك التشفير	زمن التشفير	زمن فك التشفير	زمن التشفير	زمن فك التشفير
2	0.06701827	--	--	--	0.03121113	0.03125429
4.327	--	--	12.958858	14.0100687	0.04814553	0.04739274
5	0.33630251	--	--	--	0.05635711	0.05435456
6.457	--	--	35.287274	36.7980848	0.07813382	0.06250381
68	4.62120795	--	--	--	0.60944366	0.51564717



الشكل 5 مخطط التشفير كعلاقة بين حجم الملف الأصلي وزمن التشفير كمقارنة بين الدراسة المقترحة والدراستين [10] و[13]

من خلال الجدول (4) والشكل (5)، نلاحظ أن زمن التشفير للملفات النصية كان أقل باستخدام الدراسة المقترحة بالمقارنة مع الدراستين [10] و[13] وهذا يدل على أن الخوارزمية المقترحة أسرع بمقدار 269 الى 451 مرة من

الدراسة [10] وأسرع بمقدار 2.14 مرة من الدراسة [13] حسب الجدول (4) وذلك يتوقف على حجم الملف النصي المراد تشفيره.

تجدر الإشارة الى أن الزمن الذي تم حسابه في عمليتي التشفير وفك التشفير لا يتضمن الزمن اللازم لتسليم مفتاح التشفير لطرف الاستقبال.

تقييم الخوارزمية من ناحية عشوائية المفاتيح التي تولدها

لدراسة عشوائية المفاتيح التي تولدها الخوارزمية تم تطبيق اختبارات NIST [15] على سلاسل ثنائية من هذه المفاتيح ومقارنة النتائج مع الخوارزمية [13] والتي اعتمدت نفس الاختبارات، تشمل اختبارات NIST لتقييم عشوائية تسلسل ثنائي على 15 اختبار احصائي، تعتمد هذه الاختبارات على القيمة الاحتمالية P–Value لتحديد عشوائية التسلسل، حيث تتراوح قيمة P–value ضمن المجال [1,0] وتعتبر القيمة الحدية 0.01 حيث إذا كانت القيمة المحسوبة P–value أقل من 0.01 يعتبر التسلسل غير عشوائي وإذا كانت القيمة أكبر من 0.01 يعتبر التسلسل عشوائياً ويمكن اعتماده. تم تطبيق هذه المعايير على المفاتيح المولدة من قبل الخوارزمية المقترحة ومقارنتها مع الدراسة [13] حيث اجتازت المفاتيح جميع هذه الاختبارات كما يوضح الجدول (5).

الجدول: 5 دراسة مقارنة لعشوائية المفاتيح المولدة باستخدام الخوارزمية المقترحة مع الدراسة [13]

#	الاختبار	P– value		اجتياز الاختبار
		الدراسة [13]	الخوارزمية المقترحة	
.1	Frequency	0.739918	0.783854	√
.2	Block Frequency	0.035174	0.043174	√
.3	Linear Complexity	0.350485	0.532142	√
.4	Longest Run of Ones in a	0.534146	0.634146	√
.5	Binary Matrix Rank Test	0.911413	0.945756	√
.6	Discrete Fourier Transform	0.122325	0.1314362	√
.7	Non Overlapping	0.213309	0.9999998	√
.8	Overlapping Template	0.350485	0.9999997	√
.9	Maurer's Test	0.350485	0.4853504	√
.10	Run Test	0.739918	0.742325	√
.11	Serial Test	0.739918	0.9999998	√
.12	Approximate Entropy	0.739918	0.783854	√
.13	Cumulative Sums [forward]	0.122325	0.9999998	√
.14	Cumulative Sums [invers]	---	0.9999998	√
.15	Random Excursions	---		√

نلاحظ من الجدول (5) القيم الناتجة من كلا الدراستين بالنسبة للاختبارات من 1 الى 13، لكن الدراسة [13] لم تقم بإجراء الاختبار 15 والذي يركز على العدد الكلي من المرات التي تتم فيها زيارة حالة محددة (بمعنى آخر حدوث هذه الحالة) في حالة سير عشوائي ضمن جمع تراكمي. الهدف من هذا الاختبار هو التحقق من الانحراف عن الرقم المتوقع للزيارات الى الحالات المختلفة للسير العشوائي. يتضمن هذا الاختبار سلسلة من ثمانية عشر اختباراً واستنتاجاً، اختبار واحد واستنتاج لكل حالة من الحالات: −9، −8،.....، −1 و+1،+2،.....،+8،+9 وكانت نتيجة تطبيق هذا الاختبار التي تم أخذه مباشرة من خرج البرنامج حيث تبين النتيجة اجتياز هذا الاختبار:

15. **Random Excursions Variant Test**: [('−9.0', −9.0, 3, 0.903478, **True**), ('−8.0', −8.0, 2, 1.0, **True**), ('−7.0', −7.0, 1, 0.889706, **True**), ('−6.0', −6.0, 2, 1.0, **True**), ('−5.0', −5.0, 4, 0.738882, **True**), ('−4.0', −4.0, 4, 0.7054569861112734, **True**), ('−3.0', −3.0, 4, 0.65472, **True**), ('−2.0', −2.0, 4, 0.563702, True), ('−1.0', −1.0, 3, 0.617075, **True**)]

الاستنتاجات والتوصيات:

الاستنتاجات:

تم استخدام التشفير المتناظر لتشفير الملفات النصية باللغة الإنكليزية بالاعتماد على مفهوم تشفير DNA والمفتاح OTP المولد عشوائياً من أربعة رموز A−T−C−G والتي تشكل أسس تسلسلات DNA حيث تم أخذ مفتاح الجلسة بطول النص الأصلي ومن أجل كل عملية تشفير يستخدم مفتاح مختلف ثم يستبعد وبذلك يتم ضمان أن المهاجم لن يستطيع الحصول على مفتاح التشفير خلال الهجوم العنيف brute force attack كذلك لن يتمكن من إجراء عملية تحليل إحصائي للنص المشفر الناتج بناءً على معلومات المفتاح. من خلال الاختبارات في الجدول (3) والشكل (5)، يبدو بشكل واضح أن أزمنة التشفير وفك التشفير هي أزمنة قليلة بالنسبة للملفات الصغيرة الحجم التي هي أصغر من 2 ميغابايت؛ لكن يزيد الزمن بشكل كبير عند تشفير ملفات بحجوم أكبر من 2 ميغابايت. وبالنسبة لفضاء المفاتيح المولدة عشوائياً؛ وباعتبار طول النص الأصلي n عندها سيكون عدد الاحتمالات الممكنة لمفاتيح OTP التي سيتم توليده عشوائياً هو 4^n باعتبار أن رموز DNA هي اربع وبالتالي سيكون أمام المهاجم 4^n محاولة لمعرفة المفتاح بالإضافة الى معرفة قاعدة الترميز المتبعة عند تحويل رموز DNA الى الترميز الثنائي في طرف التشفير وفك التشفير. تمت دراسة عشوائية المفاتيح التي يتم توليدها باستخدام الخوارزمية المقترحة والتي اجتازت اختبارات العشوائية القياسية مما يجعلها مناسبة للقيام بتشفير النصوص. تم تجريب هذه الخوارزمية على النصوص الإنكليزية والعربية والروسية والهندية حيث تم التشفير وفك التشفير بنجاح علماً أن جميع الدراسات المرجعية اكتفت بتشفير وفك تشفير النصوص الإنكليزية فقط.

عند استخدام مولدات تسلسلات DNA عشوائية يمكن الاستغناء عن قواعد البيانات الجينية العامة لتأمين التسلسلات والحصول على مفاتيح التشفير منها لأن ذلك يتطلب من كل من طرفي التشفير وفك التشفير البحث ضمن قاعدة البيانات الجينية العامة عن تسلسل معين من خلال المعرف الخاص به ثم فك ضغط الملف الذي هو بشكل عام ذو حجم كبير ثم استخلاص تسلسلات DNA من الملف الناتج وحذف المعلومات الإضافية لكل تسلسل، كل هذا يتطلب

زمناً إضافياً ومعالجة مما يؤدي إلى عبء إضافي. بالإضافة الى أنه لا يمكن التنبؤ متى تقوم المواقع التي تقدم بياناتها الجينية مجانا بإيقاف أو حجب الوصول العام الى مواردها.

التوصيات :

تم اعتماد تسلسلات DNA مولّدة عشوائيا من قبل مولد تسلسلات عشوائي حيث استخدمت هذه التسلسلات كمفاتيح تشفير للبيانات التي تم تحويلها أيضاً الى تسلسلات DNA بعد تحويلها الى الترميز الثنائي وفق قاعدة تحويل يتم اختيارها بشكل عشوائي وهذا بدورة يزيد من سرية التشفير. يمكن إجراء دراسات مستقبلية باعتماد توليد تسلسلات DNA المستخدمة في التشفير باستخدام تقنيات التعلم العميق من خلال نموذج يتم تدريبه على تسلسلات DNA حقيقية. تجدر الملاحظة إلى أنه عندما يكون النص المراد تشفيره قصيراً، بالأحرى أقل من 10 محارف عندها سيكون طول المفتاح أيضاً أقل من 10 محارف وللتعويض عن هذه الحالة والتي يمكن أن تضعف من أمان النص المشفر، يمكن إضافة تسلسل مزيف عشوائي الى النص المشفر وإزالته في طرف الاستقبال ويمكن إضافة مرحلة إضافية على النص المشفر الناتج لزيادة قوة التشفير من خلال استخدام قاعدة المتمم على هذا التشفير الناتج والذي هو عبارة عن تسلسل DNA أو من خلال إضافة رموز مزيفة الى النص المشفر.

المراجع:

[1] R. Daher," Information Security", Tishreen University Publications, 2015.

[2] M. Jakobsson, E. Shriver, B. Hillyer and A. and Juels, "A practical secure physical random bit generator," in *Proceedings of The Fifth ACM Conference on Computer and Communications Security*, 1998.

[3] K. A. Kassem, I. Chami and A. H. Kreaa, "Using IDEA and DESX Standards in Developing Security Applications," *Tishreen University Journal. Eng. Sciences Series*, vol. 1, 2008.

[4] G. A, L. T and R. J, "Dna-based cryptography.,," *DIMACS Series in Discrete Mathematics and Theoriti- cal Computer Science*, p. 233–249, 54 2000.

[5] A. Aich, A. Sen, S. R. Dash and S. Dehuri, "A Symmetric Key Cryptosystem Using DNA Sequence with OTP Key," *Springer India,* pp. 207 - 216, 2015.

[6] B. MONICA and O. TORNEA, "DNA secret writing Techniques," *8th IEEE International Conference in Communications*, pp. 451 - 456, 2010.

[7] Y. ZHANG, Y. ZHU, Z. WANG and S. RICHARD, "Index-Based Symmetric DNA Encryption Algorithm," in *4th International Congress on Image and Signal Processing IEEE*, 2011.

[8] A. TAUSIF, K. ABHISHEK and S. PAUL, "DNA Cryptography Based on Symmetric Key Exchange," *International Journal of Engineering and Technology (IJET),* vol. 7, no. 3, p. 938 – 950, 2015.

[9] A. ASISH, S. RANJAN, D. SATYA and D. SATCHIDANANDA, "A Symmetric Key Cryptosystem Using DNA Sequence with OTP Key," *Advances in Intelligent Systems and Computing 340, Springer India,* pp. 207-215, 2015.

[10] N. MOGALI, PRASHANTH and KAURA, "Encryption Algorithm Based on DNA Strand Technology," *IJCTA*, vol. 9, no. 33, pp. 49 - 59, 2016.

[11] Z. YUNPENG, L. XIN and S. MANHUI, "DNA based Random Key Generation and Management for OTP Encryption," *Biosystems, BIO*, no. 3753, 2017.

[12] S. KOLTE, N. K. KULHALLI, C. SHINDE and SAMRAT, "DNA Cryptography using Index-Based Symmetric DNA Encryption Algorithm," *International Journal of Engineering Research and Technology*, vol. 10, no. 1, pp. 810- 813, 2017.

[13] S. S. Nafea and P. D. M. K. Ibrahem, "Cryptographic Algorithm based on DNA and RNA Properties," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 7, p. 2278 – 1323, November 2018.

[14] Ncbi.nlm.nih.gov., "www.ncbi.nlm.nih.gov," National Center for Biotechnology Information, 2019. [Online]. Available: http://www.ncbi.nlm.nih.gov. [Accessed 20/02/ 2019].

[15] csrc.nist.gov., "https://csrc.nist.gov" , National Institute of Standards and Technology, 2019. [Online]. Avialable: https://csrc.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-22r1a.pdf.[Accessed 25/07/2019].