استخدام البرمجة الجينية GP في محاكاة خوارزمية بيركامب – ماسي

م.باسم سهر ياسين

قسم علوم الحاسوب – كلية شط العرب الجامعة

استخدام البرمجة الجينية GP في محاكاة خوارزمية بيركامب – ماسي

م. باسم سهر ياسين

USING THE GENETIC PROGRAMMING TO SIMULATE THE BERLEKAMP – MASSEY ALGORITHM

Basim Sahar Yaseen , M.Sc.in computer science

ABSTRACT

This research suggests a genetic program (GP) which is equivalent in its work to Berlekamp – Massey algorithm to find the LFSR equivalent for a given chains Hence, the genetic program deals with a population of generating structures in a random state as independent programs that generate digital random bits chains ,and it is possible for the given chain to be one of them or part of them, and these programs work independently. Each a program that represents a suggested structure (recorded and described with a specific length and with a good link equality)in the population of random programs, given a value for fitness function that represents the digital value for the extent of fulfilling the results of the final population , for the description of the given chain. The function of the subordinate program which simulate with the genetic algorithm is to find the population of the final initial values for each a genetic program. The aim of this research is to build and deal with the Berlekamp – Massey algorithm throughout the genetic programming(GP) by following a way improves the situation of work of this algorithm in order to overcome some problems that it may face for example the disavaliability of the bits of the given chain, in addition to the probability of being non- linear generating chain.

Keywords : GP , Berlekamp – Massey algorithm , LFSR , simulation.

استخدام البرمجة الجينية GP في محاكاة خوارزمية بيركامب – ماسي

- المجلد السابع
- العدد الثالث عشر
- تشرين الثاني 2014
- استلام البحث: 2013/5/26
- قبول النشر: 2013/10/2

م. باسم سهر ياسين

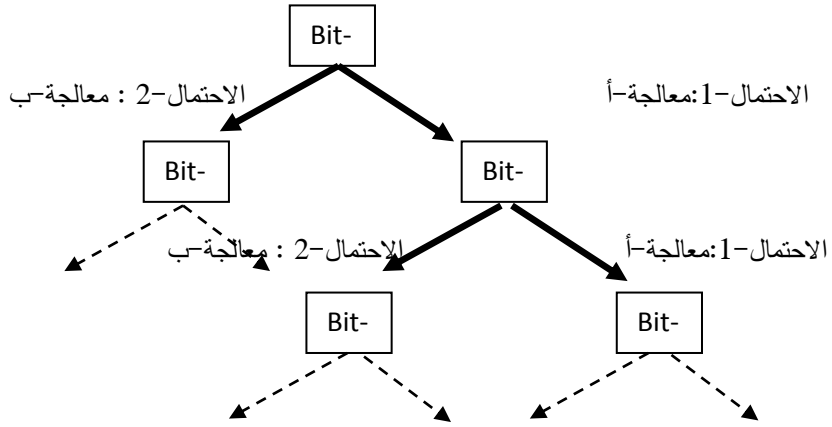
المستخلص

البحث يقترح برنامجا جينيا GP يكافئ في عمله خوارزمية بيركامب – ماسي لايجاد المكافئ الخطي لسلسلة معطاة, حيث يعالج البرنامج الجيني مجتمع لهياكل مولدة بصورة عشوائية كبرامج مستقلة تقوم بتوليد سلاسل ارقام ثنائية عشوائية, والتي يمكن ان تكون السلسلة المعطاة احداها او جزءا من احداها, وتعمل هذه البرامج بصورة مستقلة, يعطى كل برنامج يمثل هيكلا مقترحا (سجلا موصفا بطول معين وبمعادلة ربط جيد) في مجتمع البرامج العشوائية قيمة لدالة صلاحيته تمثل القيمة الرقمية لمدى تحقيق المجتمع النهائي لحلوله ,لمواصفات السلسلة ألمعطاة وتكون وظيفة البرنامج الفرعي الذي يحاكي الخوارزمية الجينية هو ايجاد مجتمع الحالات الابتدائية النهائي لكل برنامج جيني. الهدف من البحث بناء ومحاكاة خوارزمية بيركامب – ماسي من خلال البرمجة الجينية GP بطريقة تحسن من ظروف عمل هذه الخوارزمية, للتغلب على بعض المشاكل التي قد تواجهها مثل شحة ثنائيات السلسلة المعطاة (توفر سلسلة مخرجات منظومة مسجلات ازاحة بعدد قليل من الثنائيات), وكذلك احتمالية ان تكون السلسلة متولدة من نظام لا خطي .

الكلمات المفتاحية : البرمجة الجينية , خوارزمية بيركامب– ماسي , المكافئ الخطي LFSR, المحاكاة.

مقدمـــة :

تعمل خوارزمية بيركامب – ماسي [10][8] على ايجاد المكافئ الخطي [11][4][2] لسلسلة ارقام ثنائية عشوائية معطاة وبطول محدد, ويجب ان تتصف هذه السلسلة بمواصفات معينة, كأن تمتلك طولا لا يقل عن حد معين, وكذلك تكون ناتجة من نظام خطي [20][11][7], ومنذ كتابة الخوارزمية من العالم بيركامب في عام 1967م وتسخيرها في ايجاد المكافئ الخطي من العالم ماسي عام 1969م [8] والى حد الان تعد الخوارزمية طريقة كفوءة لمهاجمة نظام تشفير انسيابي خطي [4][2] من خلال مخرجاته, على الرغم من ظهور طرق اخرى بمتطلبات مختلفة كهجوم الارتباط السريع بالخوارزميتين A وB [7]والتعبير عن نظام التشفير بنظام معادلات وحله بالطرق الرياضية المعروفة كأختزال كاوس وغيرها من الطرق. وقد نفذت عدت برامج محاكاة لهذه الخوارزمية أستغلت فيها تقنيات بحث و تحليل حديثة كالخوارزمية الجينية [14] والشبكات العصبية [16][3] واساليب برمجية منوعة, الا ان هذه التقنيات عجزت عن احتواء المشاكل المرافقة للسلسلة المدخلة للخوارزمية, والمعبر عنها بعدم توفر المواصفات المطلوبة بالسلسلة, وقد حققت التقنيات المذكورة نسب نجاح متفاوتة في ايجاد المكافئ الخطي. ولو نظرنا الى الاسلوب الذي تعمل به خوارزمية بيركمب – ماسي لوجدناه يسلك اسلوب اخذ احتمالات كل ثنائية مدخلة وبنفس ورودها الى خطوات الخوارزمية لغرض بناء هيكل المكافئ الخطي وانشاء مكونات متعددة الحدود المعبرة عن ربطه, ويمكن ان نتصور الشجرة الثنائية في ادناه كتعبير عن سلوكية الخوارزمية:



الشكل–1 : الشجرة الثنائية لسلوكية خوارزمية بيركمب – ماسي .

هناك احتمالان لكل ثنائية مدخلة للخوارزمية, ومعالجتان لحساب الزيادة او عدم الزيادة في طول المكافئ الخط المقترح و اضافة الخانة الحالية للربط ام لا.

وتبقى خطوة ايجاد المكافئ الخطي دونا عن غيره من الهياكل, اهم خطوة في مهاجمة التشفير الانسيابي الذي يعتمد الهياكل الخطية وغير الخطية, كون ايجاده يسهل من تنفيذ خطوات المهاجمة الاخرى كتكوين وحل نظام المعادلات الخطية بطرق بسيطة. وفي هذا البحث تم اعتماد طريقة البرمجة الجينية لايجاد المكافئ الخطي لم تعتمد لهذا الغرض سابقا لتعطي مرونة اكثر في ايجاد افضل المواصفات للمكافئ الخطي, حيث ينقسم البحث الحالي علما: (فقرة مقدمة عن ماهي البرمجة الجينية والفروقات الاساسية بينها وبين الخوارزمية الجينية, مع ذكر اهم المصادر التي تقدم لهذه الفقرة بصورة وافية, الفقرة التي تليها هي سرد للايجابيات الناتجة عن استخدام البرمجة الجينية في

هكذا مجال (وهي نتائج العمل الحالي), ولغرض العمل على سلسلة حقيقية وهاجمتها تم تخصيص الفقرة (2) لبرمجة وانتاج ثنائيات من مولد تشفير انسيابي من الواقع العملي موضح مخططه في ضمن الفقرة نفسها, في الفقرة (3) تم شرح اجزاء وفقرات العمل في البرنامج الجيني المقترح.

2-1 **البرمجة الجينية** GP [14][13]

وهو اسلوب بحث متقدم ,يستخدم للبحث عن حلول موصفة في وسط وعينة كبيرة ومعقدة, وقد اثبت هذا الاسلوب كفاءته بالبحث وتفوق على الخوارزمية الجينية GA بنواحي عدة , وهو اسلوب مطور عنها , حيث تكون المجتمعات التي يعالجها عبارة عن هياكل وبرامج وخصائص تركيبية ...الخ , والتي قد تختلف من مجتمع لاخر, ومن كروموسوم لاخر بالمواصفات والتركيب والسلوك .

3-1 **ايجابيات استخدام البرمجة الجينية في ايجاد المكافئ الخطي**

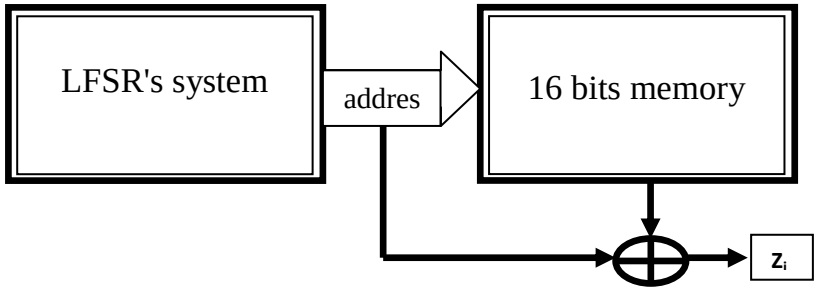
يمكن ان نجمل ايجابيات استخدام GP بدلا من اعتماد GA وحدها في المحاكاة بالنقاط التالية:

- تقوم GP بتوسيع فضاء الحلول المستخدم بالبحث, لمواجهة الزيادة بتعقيدية المشاكل المراد حلها,وبذلك يجعل احتمالية عدم التوصل لحل مقبول امرا مستبعدا ,وهذه الميزة يمكن ان تلخص بأن البحث عن الحل يتم في فضاءات متوازية بدلا من البحث في فضاء واحد كما في GA.
- لا يقتصر البحث على ايجاد الحالة الابتدائية للمكافئ الخطي, كما في GA بل يتعداها الى ايجاد مكافئ خطي بأفضل مواصفات من بين العديد من المكافئات التي تكون مجتمعات الحلول.

- وكما أظهرته نتائج البحث الحالي وأبرزته النقطة الاولى اعلاه, يمكن ان يتعامل برنامج GP مع سلاسل تعجز الخوارزمية الاصلية من التعامل معها لكونها سلاسل لا خطية او لكونها أطوالا دون المطلوب.

2- توفير سلسلة الارقام العشوائية الثنائية

لغرض ابراز الايجابيات الموضحة في ثنايا البحث الحالي عن استخدام البرمجة الجينية, فأننا سوف نعتمد كمدخلات على سلسلة ثنائيات عشوائية مولدة من نظام التشفير الانسيابي اللاخطي الذي يتكون من منظومة LFSR's تتكون من 4 مسجلات و ذاكرة تتكون من 16 خانة خزنية , تملأ جميع الهياكل المذكورة بصورة عشوائية, ويوضح المخطط-2 نموذج النظام المعتمد لإنتاج السلسلة المراد مهاجمتها بمكافئ خطي .



شكل -2:

نظام لا خطي معتمد لإنتاج سلسلة ثنائية عشوائية لغرض التحليل.

وبأعتماد القيم الابتدائية التالية :

initials="1100101,100110,00110010,11001" LFSR's

memory = "1101010100101101"

سوف تتولد سلسلة الارقام الثنائية العشوائية التالية : "101011101111" وهي سلسلة متكونة من 12 bits وهذه الكمية لا تلبي حاجة خوارزمية بيركمب – ماسي من الثنائيات , لذلك قد تفشل الخوارزمية في ايجاد افضل مواصفات للمكافئ الخطي وقد يكون الناتج يحوي نسبة من الخطأ , اما بالنسبة للمعالجة بالبرمجة الجينية فسوف يتم تجاوز هذه المشكلة .

1. البرنامج الجيني المقترح

كما هو مشهور عن البرمجة الجينية, فأنها تقوم بتنفيذ اجراءات ومعالجات الخوارزمية الجينية على مجتمع من البرامج وهياكل المعالجة لغرض الوصول الى افضلها [14] [12],وهذه بدورها تقوم بتنفيذ نفس الاجراءات على مجتمعاته المحلية وصولا الى افضل المواصفات لكل برنامج او هيكل معالجة, فيكون تنفيذ اجراءات الخوارزمية الجينية بالنسبة للبرامج متداخلا nested . وقد اعد البرنامج المقترح في بيئة البرمجة المستقلة VB.Net مستخدما مفاهيم المحاكاة العملية في محاكاة نظام رقمي ثنائي قائم ومستند الى اجراءات ودوال البرمجة والخوارزمية الجينية, ويرجع سبب استخدام VB.Net الى امكانية ربط البرنامج بسهولة ضمن اي بيئة برمجية تستند اليها شبكة بيانات معينة, وكذلك الى امكانية اجراء التطوير والصيانة من خلال تلك الشبكة

3 . 1 الربط الجيد للمكافئات الخطية

يتم انشاء ملف للربوطات الجيدة للمكافئات الخطية التي من المحتمل ان يتعامل معها البرنامج الجيني, واقصى حجم لمكافئ خطي تم تثبيته في مجتمعات البرنامج هوبطول 256 bits, والربط الجيد [11][2] :هو الربط الذي يعطي دورة عظمى لمحتويات المكافئ الخطي, وتكون متعددة الحدود المعبرة عن الربط عبارة عن

معادلة غير مختزلة irreducible ,لذلك فيجب اضافته كمحدد على توليد الكروموسومات في مجتمعات البرامج الرئيسية.

3 . 2 كيفية انشاء مجتمع البرامج

تمت مراعاة عدة نقاط مهمة عند انشاء المجتمع الابتدائي للبرامج (هياكل المكافئات الخطية) ,وهي:

- لتلافي مشكلة صغر طول السلسلة المتوفرة من النظام الاصلي, يتم اكمالها بثنائيات تولد عشوائيا ,لغرض الوصول الى الطول المطلوب والكافي لتنفيذ خطوات خوارزمية بيركمب – ماسي[8] ضمن البرنامج الجيني. فيكون الناتج ملفا من السلاسل التي تكون ثابتة بجزء منها ومتغيرة بالجزء المولد عشوائيا وجميعها بطول واحد ثابت.
- تستخدم خطوات الخوارزمية الاصلية ادناه في انتاج هيكل لمكافئات خطية ,تكون كروموسومات البرنامج الجيني الابتدائي ومعبر عنها بصيغة الثنائيات.

Input: A sequence $a_N = a_0, a_1, \cdots, a_{N-1}$.
Output: LFSR (f_{N-1}, l_{N-1})
STEPS:begin

1. Procedure LFSR(a_N)
2. Check if $a_i = 0, i = 0, 1, \cdots, k-1$ and $a_k = 1$,
   then set $f(x) = x_{k+1} + 1$
   $l = k + 1$
   $g(x) = 1$
   $a = k$
   $b = 0$

T(x) = 0

3. for n from k + 1 to N − 1 do

(a) Compute: $d = a_n + P_l - 1$

$i=0\ c_i = a_n - l + i$

(b) if d = 0 then b = b + 1

(c) if d 6= 0 and 2l > n then

f(x) = f(x) − xa−bg(x)

b = b + 1

(d) if d 6= 0 and 2l _ n then

T(x) = f(x)

f(x) = xb−af(x) − g(x)

l = n + 1 − l

g(x) = T(x)

a = b

b = n − l + 1

end

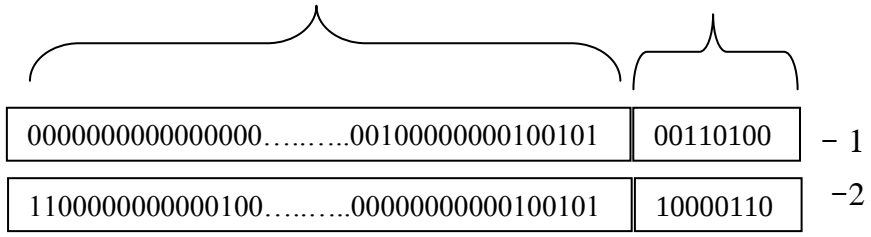
4. return (f(x), l)

5. end LFSR

- تحسب دالة الصلاحية لكل برنامج بمقياس معدل افضل مواصفات لمجتمع محلي متحقق.

فيكون شكل مجتمع البرامج (هياكل المكافئات الخطية) بالشكل:

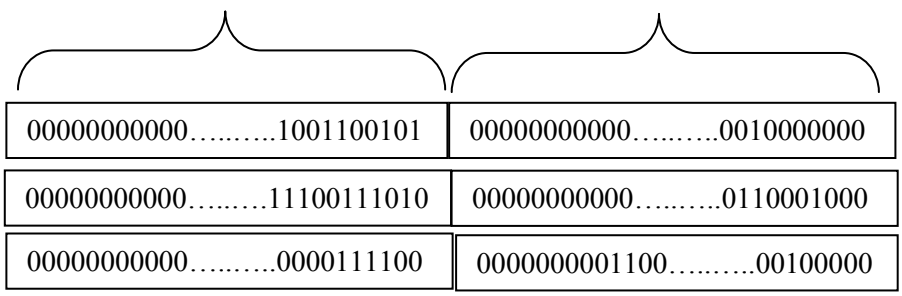
طول المكافئ الخطي معادلة ربط المكافئ الخطي



3 .3 مجتمعات الخوارزمية الجينية

من المجتمع الاولي للبرامج (المتولد من الفقرة 3. 2) الى المجتمعات التي سوف تليه, يتم انشاء مجتمعات الخوارزمية الجينية التي يتكون فيها مجتمع الكروموسومات بواسطة معالجة الكروموسوم الرئيس من البرنامج الجيني وانشاء كروموسومات محلية كلها بنفس طول المكافئ الخطي ولكن بمتعددات حدود وقيم ابتدائية initial له مختلفة , فمثلا لهيكل المكافئ الخطي الاول في المجتمع اعلاه (بطول (00110100) 52 bits) فيكون كل كروموسوم في المجتمع المحلي بالصورة:

القيم الابدائية للمكافئ الخطي　　معادلة الربط



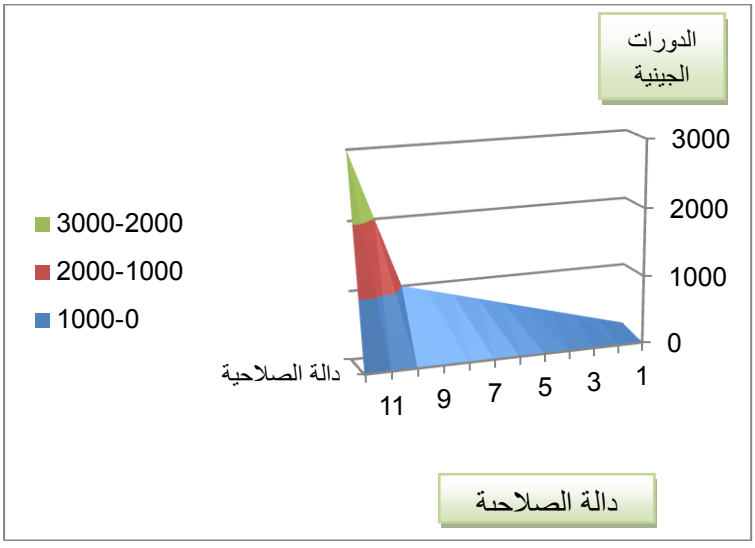
اما دالة الصلاحية فتحسب بمقدار عدد التطابقات للثنائيات الخارجة من مواصفات المكافئ الخطي الممثل بالكروموسوم, مع ثنائيات السلسلة الاصلية بدون الزيادات العشوائية, وبأي موقع من تلك السلسلة, مقسوما على العدد الكلي للثنائيات المكونة للسلسلة مضروبا بمئة. ويمكن ان نلاحظ مدى تطور دالة الصلاحية للمجتمع المحلي مع تصاعد عدد الدورات الجينية من خلال المخطط –3 ادناه , والمأخوذ من التنفيذ الاول للبرنامج الجيني المقترح. اما الجدول – 1 فيمثل تطور الوصول الى المكافئ الخطي بأفضل المواصفات الممكنة من خلال التنفيذ الاولي لمعالجات البرنامج الجيني على مجتمع هياكل المكافئات الممثلة بكروموسومات .وبمعاملات متغيرة تبدأ بقيم PC=0.73 و PM=0.03

جدول القيم لعدد الدورات الجينية وطول المكافئ الخطى الناتج من البرنامج الجينى

طول المكافئ	عدد الدورات الجينية
250	100
240	200
235	300
210	400
190	500
187	600
182	700
170	800

900	163
1000	103
2000	89
3000	63

جدول –1 : تطور طول المكافئ الخطي مقابل عدد الدورات الجينية



الشكل – 3

تطور دالة الصلاحية في المجتمعات المحلية قياسا بعدد الدورات الجينية

4. 3 النتائج العملية

عند الوصول لشرط التوقف المحدد داخل اجراءات البرنامج الجيني (Genetic cycle=3000) ,كانت النتائج تشير الى الكروموسوم رقم 37 كأفضل مواصفات تم الوصول لها ,حيث طول المكافئ الخطي هو (63) وبدالة ربط = 1+X1+X_{34}+X_{35}+X_{36}+X_{37}+X_{63}=p(x) وبدالة صلاحية هي (1.0) .

37	00000000000000000.....0000000000011111
1.0	01000000000000000000000000011110000000000000000000000000100000001

4 .المناقشة

بالرغم من أن البرنامج الجيني المقترح يستخدم خطوات الخوارزمية الاصلية في انشاء مجتمع البرامج , وذلك لتقليل الاعتماد على العشوائية في ايجادها, وجعلها تقتصر على تأثير السلاسل الجزئية المكملة للسلسلة الاصلية وتأثيرها في ايجاد مكافئات مختلفة ومتنوعة في المجتمع الابتدائي للبرامج, وكذلك لتوفر سبب مبرر آخر الا وهو ضمان انشاء مجتمع يمتلك المواصفات الرياضية ويتقيد بالحدود المفروضة على هيكلية المكافئات الخطية والتي توفرها خطوات الخوارزمية الاصلية, على الرغم من ذلك الا ان البرنامج في اجراءاته الجينية كالتزاوج والطفرة والاختيار وإنتاج المجتمعات المحسنة انتهج اسلوبا جديدا مغايرا للمألوف وحقق نتائج مشجعة, ولقد نجح الاسلوب الهجين بين الاثنين في الافادة من مزايا الطريقتين للتغلب على المشاكل المرافقة لمعالجة السلاسل العشوائية المشار لها في فقرات البحث. من الميزات الايجابية لإدخال GP في هذا المجال هو امكانية التعامل مع سلاسل مختلفة المنشأ (خطية ولا خطية) ,على عكس الكثير من التقنيات التي استخدمت سابقا والتي

تخصصت في التعامل مع الانظمة الخطية, وقد اختزل الاسلوب المقترح اعمال اضافية تترتب على ايجاد مكافئ خطي بمواصفات ليست الافضل دائما, فنجد ان باستخدام الخوارزمية الاصلية على سلسلة معطاة بطول معين يمكن ان نجد مكافئ خطي بحجم اكبر من المكافئ الخطي المنشأ من خلال GP ولنفس السلسلة, ويوضح المخطط-4 النتيجة هذه من خلال عدة تنفيذات للبرنامج الجيني ولخوارزمية بيركمب-ماسي على سلاسل ثنائية عشوائية.

5. الاستنتاجات

لم تشر ادبيات علم التشفير لاستخدام GP في ايجاد المكافئ الخطي, بل استخدمت وسائل وتقنيات عديدة لمحاكاة خوارزمية بيرلكامب – ماسي لايجاد المكافئ الخطي لسلسلة معطاة, مثل الخوارزمية الجينية والشبكات العصبية وانشاء نظام المعادلات للمكافئ وحلها بطريقة رياضية كأسلوب كاوس , وغيرها من الاساليب الا ان لكل طريقة مستخدمة محددات وعيوب وايجابيات , تتفوق طريقة البرمجة الجينية عليها كونها اسلوب مرن يتعامل مع اي عدد متوفر من اخراجات المنظومة المهاجمة وكذلك تنوع مواصفات المكافئات الخطية الناتجة من الطريقة وسرعة الوصول وكمقارنة بين البرنامج الجيني المقترح مع برامج لخوارزمية جينية وشبكة عصبية , كانت النتائج كالتالي:

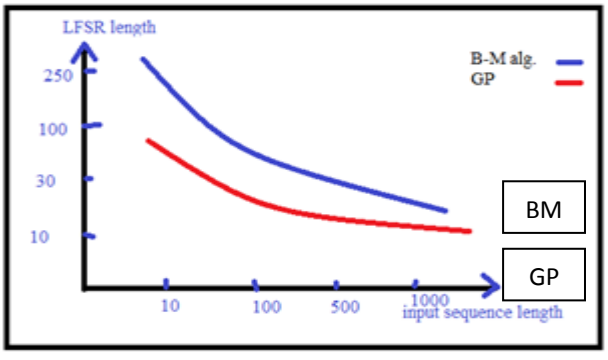
جدول-1

مقارنة بين اساليب غير تقليدية في محاكاة بيركمب-ماسي

البرنامج	السرعة للوصول للحل	الدقة في اختيار المواصفات	تنوع مكونات المجتمع النهائي	الوقت اللازم لايجاد المكافئ	الملاحظات
GP	سريع	دقيق	اعطاء مكونات متنوعة	ملائم نوعا ما	يتم العمل على مواصفات متنوعة لمكافئات خطية مختلفة لغرض اختيار افضلها والتي تحقق افضل التطابقات
GA	سريع ضمن المكافئ المحدد	دقيق	اعطاء مواصفات مكافئ واحد	ملائم نوعا ما	العمل على مواصفات واحدة لمكافئ خطي واحد لايجاد محتوياته التي تعطي تطابق مع السلسلة المعنية
ANN	البطئ في عملية التعلم	وجود نسبة خطأ	اعطاء مواصفات مكافئ واحد	باستثناء وقت التعلم يعتبر سريع	كذلك

الشكل 4

العلاقة بين طول المكافئ الخطي قياسا بطول السلسلة المتوفرة في طريقتي بيركمب – ماسي و GP



6 . المصادر

[1] Abraham Sinkov “Elementary Cryptanalysis”,Second Edition ,Mathematical Association of America,2009.

[2] Alfred J. Menezes,Paul C. van Oorschot, Scott A. Vanstone " HANDBOOK of

APPLIED CRYPTOGRAPHY",USA ,1996.

[3] Ben Krose , P.Van der smagt,"An Introduction to Neural Networks",Eighth Edition ,1996

[4] Bruce schneier ,"Applied Cryptography",2nd Edition.

[5] Christopher Swenson," Modern Cryptanalysis: Technique for Advanced Code Breaking ",University of Tulsa,2008.

[6] David Naccache,"Cryptography and Security : from Theory to Application",2012.

[7] E.R.Berlekamp," Algebraic Coding theory", McGraw-Hill, New York, 1968.

[8]F. G. Gustavson, "Analysis of the Berlekamp-Massey Linear Feedback

Shift-Register Synthesis Algorithm",USA ,1992.

[9] Henk C.A.van Tilborg," Fundamentals of Cryptology ",Eindhoven University of Technology, Netherlands.

[10] J.L. Massey, "Shift-Register Synthesis and BCH Decoding", IEEE Trans. on Inform. Theory vol. IT-15, January 1969.

[11] Jeffrey Hoffstein & Jill Pipher &Joseph H.Silverman,"An Introduction to Mathematical Cryptography", Springer ,USA,2008.

[12] Kumara Sastry & D. D. Johnson, David E. Goldberg, Pascal Bellon,"Genetic Programming for Multiscale Modeling", by Begell House, Inc.,2004 .

[13] LEE SPECTOR, ALAN ROBINSON," Genetic Programming and Evolvable Machines", Kluwer Academic Publishers. Manufactured in The Netherlands ,2002.

[14] Matthew Walker,"Introduction to Genetic Programming", October 7, 2001.

[15] Melanie Mitchell ,"An Introduction to Genetic Algorithm",Cambridge,Massachusetts,London,England,1999.

[16] Raul Rojas,"Neural Networks-A Systematic Introduction", Springer ,1996.

[17] Salil P.Vadhan "Theory of Cryptography",2009,ISBN:3540709355.

[18] Serge Vaudenay,"A Classical Introduction to Cryptography: Applications for Communications Security",2006.

[19] Thomas Johansson,"Stream Ciphers: Cryptanalytic Techniques", Department of Electrical and Information Technology ,Lund University Sweden,2007.

[20] Wade Trappe"Introduction to Cryptography with Coding Theory ",Prentice Hall,2010,ISBN:0130618144.

[21] William Stallings," Cryptography and Network Security:Principles and Practice",2010.