مجلة نصف سنوية

العدد الثاني والثلاثون
السنة العشرون
١٩٩٨

مجلة الحاسبات الالكترونية

تكون جميع المراسلات بعنوان:
رئيس تحرير مجلة الحاسبات الالكترونية
وزارة التعليم العالي والبحث العلمي
ص.ب. ٣٢٣٦١ - السعدون
بغداد - العراق

تصدر عن - وزارة التعليم العالي والبحث العلمي - المركز القومي للحاسبات الالكترونية

هيئة تحرير المجلة

رقم الايداع في المكتبة الوطنية ٣٠٤-١٩٧٧

الاشتراك السنوي: داخل العراق: للدوائر والشركات: ١٠٠٠٠ دينار خارج العراق: للدوائر والشركات والاشخاص: ٢٥ دولار

طبعت في مكتب النسق للتنضيد والطباعة الالكترونية - بغداد - حي النضال - عمارة شكر النجار - قرب جمعية المهندسين - هاتف ٧١٧٦١٣٦

معلومات للراغبين في النشر

تهدف المجلة الى تعميق المعرفة بعلوم الحاسبات بين العاملين وفي أوسع القطاعات الجماهيرية وصولا الى جعلها مصدرا علميا متخصصا في هذا المجال باللغة العربية وسوف ينظر فقط في نشر المواد القيمة والمفيدة التي تتخذ أحد الاشكال التالية:

البحث، التقرير، المقالة، المواد المترجمة.

وتخضع جميع المواد للنشر للتقييم من قبل اختصاصيين مختارين وفق معايير خاصة لغرض ضمان رفيع من النوعية. كما ونشير الى ان المجلة سوف لاتنظر في نشر أي من المواد التي سبق نشرها او التي تنتظر النشر في مطبوع آخر وتمنح المجلة مكافأة تشجيعية للمواد التي تنشر، وندرج أدناه تعليمات خاصة بطريقة وضع وعرض وتقديم المواد للنشر، راجين مراعاتها تسهيلا المهمة.

١- مشروع المادة: تقدم المادة باللغة العربية بثلاثة نسخ مطبوعة على الة الطابعة على وجه واحد من الورق وبأسطر متباعدة، مع ترك مسافة كافية للهوامش.

٢- المحتوى: يكون ترتيب محتوى المادة على الوجه التالي:
- العنوان يليه أسم (أو أسماء) وعنوان (أو عناوين) الكاتب (أو الكتاب).
- ملخص لايزيد عن مائة كلمة يعطي فكرة موجزة عن الموضوع.
- فاتحة، صلب الموضوع، ثم خاتمة.
- قائمة بالمراجع.
- الرسوم والمخططات والصور الفوتوغرافية.

٣- المراجع: يشار الى المراجع بأرقام (مطبوعة على الطابعة) ويوصف كل مرجع في قائمة المراجع على النحو الاتي:
أسم كاتب المادة، عنوان المادة، المطبوع التي نشرت فيه، سنة النشر، الصفحة.

٤- الجداول: تكتب الجداول على الطابعة ويراعى ان تكون لها ارقام متسلسلة وعناوين واضحة وان لاتكون تكرار البيانات ظهرت في مكان آخر من المادة.

٥- الرسومات: تقدم الرسومات الاصلية مرسومة بوضوح (ويفضل ان تكون بالحبر الصيني على ورق شفاف)، ويراعى ان يكون لكل رسم عنوان وافي ودقيق، وان تشرح كل الرموز المستعملة فيه.

٦- الصور الفوتوغرافية: يفضل تقديم أكثر من نسخة من الصور الفوتوغرافية المستخدمة في المادة (ان وجدت).

٧- المعادلات: تترك مسافة قدرها سنتيمتر على الاقل بين المعادلة وما يسبقها ويليها من معادلات او كتابة، ويراعى تجنب استعمال الحروف والارقام التي قد يلتبس الامر بينها.

٨- برامج الحاسبات: اذا كانت الضرورة تقضي تضمين برنامجا فيراعى وضعها في ملحق الا اذا كان ذلك البرنامج او البرامج جزء رئيسي من المادة او موضوعها.

٩- الارقام العربية: تستخدم الارقام العربية الاصلية (3.2.1 الخ) في المواد المقدمة.

١٠- المصطلحات: يراعى استخدام المرادفات العربية للمصطلحات الاجنبية المستخدمة في كتابة المادة وللكاتب ان يضمن المصطلح الاجنبي بين قوسين في حالة اضطراره لاستعماله او عدم تمكنه من ايجاد بديل عربي.

١١- التقديم: ترسل المواد على العنوان التالي: رئيس تحرير مجلة الحاسبات الالكترونية، وزارة التعليم العالي والبحث العلمي، ص.ب ٣٢٦١ السعدون، بغداد-جمهورية العراق، تلكس-٢١٢١٦٣.

كلمة العدد

من منطلق التعاون بين هيئة تحرير مجلة الحاسبات والجمعية العراقية لعلوم الحاسبات لتحقيق الأهداف العلمية في نشر ما ينجزه الباحثين من بحوث ودراسات في مجال الحاسبات فقد تم الاتفاق مع الجمعية على نشر وقائع المؤتمرات العلمية والوطنية التي تعقدها الجمعية.

تصدر هيئة التحرير العدد الثاني والثلاثون لوقائع المؤتمر الوطني الخامس والمنعقد في تشرين الثاني لعام ١٩٩٧ دعما لمسيرة القطر العلمية والبحثية ولاسيما في هذا الحقل المميز.

مع فائق التقدير ....

هيئة التحرير

محتويات العدد

١- البحوث باللغة العربية

المحتويات	رقم الصفحة
- تطبيق محوسب لتصحيح بعض الأخطاء النحوية في الجملة العربية البسيطة محمد نعمان مراد	٥

٢- البحوث باللغة الانكليزية

Contents	Page No.
- An Approach for Breaking RSA Public Key Cipher System Dr. Ala'a. H. Al-Hamami	3
- Adaptive Ciphertext-Only Attack Using Genetic Programming With Indexed Memory Dr. W. A. K. Al-Hamdani, Dr. A. F. Abdul Kader, W. S. Awad	9
- Use of Genetic Algorithm (GAs) in The Cryptanalysis Nonlinear Stream Cipher (NLSC) Dr. W. A. K. Al-Hamdani, S. A. Al-Ageelee	15
- Infra Red Remote Pc Keyboard W. A. Jabbar	24
- Image Guider Ahmad S. Nori Laheeb M. Ibrahim Najla Badeaa	30
- Shorter Signature Verification Time With Improved Digital Signature Standard, DSS Hamza A. Al-Sewadi Khaldon I. Arif	38

تطبيق مُحوسب لتصحيح بعض الأخطاء النحوية في الجُملة العربية البسيطة

محمد نعمان مراد

خلاصة البحث

يهدف البحث الى اجراء تطبيق مُحوسَب لتصحيح بعض الأخطاء النحوية في الجمل العربية البسيطة، إذ يشمل التطبيق جانبي المعالجة اللغوية للجملة، الأول يتمثل في (تحليل) الجملة المُدخلة وإكتشاف الأخطاء النحوية وتصحيحها، أما الثاني فيتمثل في (توليد) الجملة الصحيحة.

ضُمنت قواعد نظرية تعالج المُطابقات بين الفعل والفاعل (المُسند والمُسند إليه)، إذ يلعب المعجم الدور الحاسم في عملية المُطابقات، فقد صُمم المعجم بحيث يتضمن الكلمة العربية مقترنة بسماتها النحوية وبعض السمات الدلالية التي تلبي تلك التطابق لإكتشاف الأخطاء النحوية في الجملة وتصحيحها على وفق أسس النظرية التحويلية التوليدية لجومسكي فهناك نوعان من التصحيح في عمل المنظومة، الأول يُصحح الجملة المدخلة وهي في بنيتها السطحية مباشرة الى جملة في البنية السطحية، أما النوع الآخر فيُصحح الجملة المدخلة بإيجاد بُنيتها العميقة، وبعدها تولد الجملة بِبنيتها السطحية.

استخدم نظام قواعد العبارة المحددة (DCG) لصياغة تمثيل القواعد النحوية العربية المُعرّفة في المنظومة، وذلك بإستخدام لغة البرمجة المنطقية (Prolog)، ونفذ التطبيق على حواسيب شخصية نوع (IBM/PC) والحواسيب المتوائمة معها، وتعمل المنظومة تحت مظلة نظام التشغيل (MS-DOS) إصدار 3.1 فما فوق.

١- المقدمة

تتطلب تطبيقات مجال معالجة اللغات الطبيعية (Natural Language Processing) القيام بالدور أو الهدف الذي نُفذت من أجله تلك التطبيقات، وهو بناء منظومات حاسوبية تُمكّن الحاسوب من فهم اللغة الطبيعية، وأن أمرا مثل هذا، بحد ذاته، ربما يكون صعبا، لأن قدرات بشرية وأحاسيس عديدة تولّد مفهوماً للكلام المنطوق مغاير لمفهومه المعبّر عنه. لذلك عُدّ هذا المجال فرعا أساسياً من فروع الذكاء الإصطناعي (Artificial Intelligence).

تحتاج منظومات معالجة اللغات الطبيعية حاسوبيا في تنفيذها وجعلها منظومات ذكية الى تلبية متطلبات المعالجة اللغوية المتمثلة في تحليل الجُمَل وتوليدها، هذا بدوره يتطلب توفير معطيات وقواعد (صرفية ونحوية ودلالية ومعجمية) تكون شاملة وكافية لتمثيلها التمثيل المناسب بما يُمكّن الحاسوب من تنفيذ مهامه بكفاءة وتحقيق الهدف من التطبيق.

اللغة العربية لها خصائصها (الصرفية والنحوية والدلالية والمعجمية) التي تتفرد بها وتميزها عن سواها من لغات البشر، وإن خضوعها الى التطبيق المُحوسب يتطلب صياغة رياضية تُعبّر بوضوح عن تلك الخصائص وبالشكل الذي تكون فيه تلك القواعد ملائمة للتطبيق المُحوسب.

مما يزيد من حدة التفاعل بين اللسانيات والحاسوب في مجال المعالجة النحوية المحوسبة إختلافهما الأساس من حيث الهدف، ففي حين يسعى المختصون في اللسانيات الى الصفاء النظري، والتغطية الكاملة للظواهر اللغوية المختلفة في شقيها التوليد والتحليل، يركز المختصون في الحاسوب، في معظم الأحيان، على تحقيق نتائج عملية بغض النظر عن مدى وجاهتها اللغوية الصرفية، وهم يستخدمون في ذلك أساليب ذات طابع إجرائي، أساسها التبسيط لا التمعيق، تتحاشى الدخول في متاهات الشذوذ والشرود اللغويين والتي يندر حدوثها، وذلك للمحافظة على كفاءة النظم المحوسبة". (علي، ١٩٨٨: ٣٨٩).

لقد نجم عن العمل البحثي في مجال معالجة اللغة

العربية حاسوبيا، خلال عقدين من الزمن، العديد من البحوث والتطبيقات في هذا المجال الحيوي. وبحثنا إذ يدخل في مجال معالجة اللغة العربية حاسوبيا، يختص بالدرجة الأساس في نحو اللغة العربية، فالتطبيق المحوسب مدار البحث يُحلل بعض الجمل العربية البسيطة ويتعرف على الأخطاء النحوية منها، ومن ثم يولّد الجمل الصحيحة. حاولنا أن تكون هناك شمولية في التطبيق البحثي على مستوى نحو اللغة العربية (تحليلا وتوليدا) مضافا اليه المعالجة الدلالية من خلال توفير بعض السمات الدلالية ضمن مفردات المعجم.

٢- نظرة الى النحو العربي

اللغة، عموما، لا تُستعمل في فراغ، بل هنالك أمران يحكمان الإستعمال اللغوي، أولهما السياق اللغوي نفسه الذي لا تأخذ المفردات معانيها بمعزل عنه (خرما، ١٩٧٨: ١٢٢). فعندما نبدأ جملة بـ:

أكل مصطفى ...

وقبل ان تتم الجملة، فإن السامع يتوقع في الحال ان تتم الجملة بإسم يدل على نوع من الطعام، ولكن المعنى الحقيقي لما تم النطق به من الكلام لا يتأتى إلا بإتمام الجملة، فإذا كان الكلام باللهجة المصرية، مثلا، يمكن أن تتم الجملة بالقول:

أكل مصطفى علقة

لقد خاب ظن السامع، وتغير فهمه لمعنى كلمة (أكل) تغييرا كبيرا، معناها (أصاب أو نزل بمصطفى سوء)، وعندما نتأمل الجُمل الآتية نرى كيف يكتسب الفعل (أكل) معان مختلفة لوقوعه في سياقات لغوية مختلفة:

(أيحبُ أحدكم أن يأكلَ لحمَ أخيه ميتا) (قرآن كريم)
أكل مصطفى طعامَهُ
أكل مصطفى مالَ اليتيم
أكل مصطفى أصابعَه ندما
أكل مصطفى ضربة على رأسه
أكلني جلدي أو رأسي
أكلت السكينُ اللحمَ
مصطفى يأكل عمرَه
مصطفى يأكل لحومَ الناس
أكل مصطفى علقة

هنالك إذن السياق اللغوي (Verbal Context) الذي يحدد معاني المفردات والذي بدونه لا يتم ذلك.

ولكن هنالك أيضا قرينة أخرى هي الموقف او المناسبة التي يقال فيها الكلام والتي اطلق عليها اللغويون العرب عبارة المقام فقالوا (لكل مقام مقال)، وهذا بالطبع يؤثر في معنى الجملة كلها تأثيرا كبيرا.

وعناصر هذا المقام عديدة: أولها المتكلم نفسه: هل هو ذكر أم أنثى؟ واحد أم إثنان أم جماعة أم جمهور؟ وما هو جنسيته ودينه وشكله الخارجي ونبرة صوته ومكانه الاجتماعي الى آخر هذه الصفات التي تميزه عن سواه. وهذا ينطبق على المستمع أيضا ويشمل إضافة الى ذلك علاقته بالمتكلم من حيث القرابة أو الصداقة أو المعرفة السطحية أو عدم المعرفة أو اللامبالاة أو العداوة؛ أو المركز الإجتماعي أو المالي أو السياسي ... الخ. ومن عناصر المقام أيضا موضوع الكلام، وفي أي جو يقال، وفي أي مكان وأي زمان؟ وكيف يقال، وما الداعي لقوله، وسوى ذلك من العناصر الكثيرة جدا التي يؤثر كل منها تأثيرا مباشرا على كيفية قول الكلام وعلى تركيبه وعلى معانيه وعلى الغرض من قوله (السامرائي، ١٩٨٩).

بينما يبحث علم النحو في علاقات المفردات بعضها ببعض في الجمل المختلفة، لابد من التنبيه بأن كلا من العلمين يرفد الآخر ويتصل به إتصالا وثيقاً لأن البنية الداخلية للكلمة تؤثر على علاقاتها مع الكلمات الأخرى في الجملة. فإذا إستعملنا فعلا مثل (قاتل) في بداية إحدى الجمل فإن المستمع يتوقع في الحال أن نتبع ذلك الفعل بفاعل يشير الى من قام بالمقاتلة وبمفعول به يشير الى من حصلت المقاتلة معه (خرما، ١٩٧٨). أي اننا نتوقع جملة كهذه.

قاتل الرجل عدوه

فإذا ماطرأ على الفعل (قاتل) تغيير داخلي (صرفي) بزيادة (التاء المفتوحة) في أوله، فأصبح (تقاتل) وأستُخدم الفعل في بداية إحدى الجمل، فإن تركيب الجملة (وهي ظاهرة نحوية) يتغير تبعا لذلك. فلا نعود نتوقع مفعولا به مثلا، بل نتوقع فاعلا فقط، كما نتوقع أن يكون الفاعل يشير الى المثنى أو الجمع، أي أن الجملة الناتجة تكون شبيهة (تقاتل الرجلان)، أو (تقاتل الرجال)، أو ان يكون الفاعل مفردا على أن تكمل الجملة بما يدل على اشتراك آخرين في العمل، كأن نقول:

تقاتل الرجل مع رفاقه

كما أن الصلة بين علمي النحو والدلالة واضحة،

فالفرق واضح في المعنى بين جملة (أخذتُ الكتاب منه)، على سبيل المثال، وجملة (أخذتُ الكتاب إليه). هذا الفرق نجم عن إستبدال حرف الجر (من) في الجملة الأولى بحرف جر آخر، هو (الى) في الجملة الثانية.

ولكن على الرغم من هذا الترابط الواضح بين أنظمة اللغة المختلفة، فأن على الدارس أن يعالج كلا منها على إنفراد في أحيان كثيرة تفاديا للتشابك الكبير القائم بينهما وتسهيلا للدراسة نفسها. وهذا ينطبق بوجه عام على معظم الدراسات اللغوية.

لذلك يعدُّ موقع النحو (Syntax) من اللغة، هو بمثابة القلب من جسم الإنسان، أما كلمة القواعد (grammar) فهي تشمل النحو بالإضافة الى الصرف كما تشمل النظام الصوتي ونظام المعاني أيضا. فهي بهذا إصطلاح شامل جدا لجميع القواعد التي لها علاقة بجميع وجوه اللغة المختلفة*

"من المعلوم أن علم (النحو) يُعنى أول ما يُعنى بالنظر في أواخر الكلم وما يعتريها من إعراب وبناء، كما يُعنى بأمور أخرى على جانب كبير من الأهمية، كالحذف والتقديم والتأخير وتفسير بعض التعبيرات غير أنه يولي العناية الأولى للإعراب". (السامرائي ١٩٨٩: ٥).

القواعد التحويلية (Transformational Rules) كما وصفها العالم اللغوي (جومسكي Chomsky) في نظريته الشهيرة (النظرية التحويلية التوليدية)، تعطي هذه القواعد وصفا للعلاقة بين البُنية العميقة (Deep Structure) والبنية السطحية (Surface Structure)، والعلاقة بين البُنيتين تشبه عملية كيميائية يعبَّر عنها بمعادلة، أحد طرفيها هو المدخلات قبل التفاعل، والطرف الأخر هو الناتج بعد التفاعل. فالبُنية العميقة تعطي المعنى الأساس للجملة، أما البنية السطحية فتتمثل في الجمل المستخدمة في الكلام أو الكتابة.

والقواعد التحويلية تستطيع أن تقدم تفسيرا مقنعا لقدرة المرء على أن ينتج وأن يفهم الجمل الجديدة وكيفية التمييز بين الجملة الصحيحة وغير الصحيحة، فضلا عن قدرتها وكفاءتها على تفسير تركيب الجمل المعقدة، إضافة الى تفسيرها بالحكم أن جملتين أو أكثر مترادفة في المعنى. لذلك كان لها وقع خاص في النحو (الخولي، ١٩٨٨ وعثمان، ١٩٩٠).

٣- الجملة في اللغة العربية

تتألف الجملة العربية من ركنيين أساسيين هما المُسنَد والمسنَدُ إليه، فالمُسنَد هو المتحدث به ويكون فعلا أو إسما، أما المُسنَد إليه فهو المتحدث عنه ولا يكون إلا إسما، وهذان الركنان هما عمدة الكلام وما عداهما فضلة أو قيد (السامرائي، ١٩٨٩).

يظهر تأليف الجملة العربية بصيغتين تبعا للمُسنَد، صيغة (فعل مع إسم)، وصيغة (إسم مع إسم)، وبالتعبير الإصطلاحي، صيغة (فعل وفاعل)، وصيغة (مبتدأ وخبر)، نحو (أقبل سعيد) و(سعيد مقبل)، وكل التعبيرات الأخرى إنما هي صيغ أخرى لهذين الأصلين.

فالصيغة الأساس للجمل التي مُسندها فعل أن يتقدم الفعل على المُسند إليه، كما في جملة (أقبل سعيد) ولا يتقدم الفاعل على الفعل، أو بتعبير أدق: لا يتقدم المُسند إليه على الفعل إلا لغرض يقتضيه المقام، أي أن البُنية الأساس للجملة العربية التي تحمل فعلا تكون بالشكل الأتي (Bakir, 1981):

(ف فا مف)وتعني (فعل فاعل مفعول).

والصيغة الأساس للجمل التي مُسندها إسم، أن يتقدم المُسند إليه على المسند، أو بتعبير آخر: أن يتقدم المبتدأ على الخبر، ولا يُقدَّم الخبر إلا لسبب يقتضيه المقام أو طبيعة الكلام.

والفرق بين هاتين الصيغتين-أي الجملة التي مُسندها فعل والجملة التي مُسندها فعل والجملة التي مُسندها فعل والجملة التي مُسندها إسم-أن الجملة التي مُسندها فعل إنما تدل على الحدوث، تقدم الفعل أو تأخر، والجملة التي مُسندها إسم تدل على الثبوت. تقول مثلا: (يجتهد زيد) و(زيد مجتهد)، و(يحفظ زيد) و(زيد حافظ)، و(يطلع سعيد) و(سعيدمطلع)، و(يتعلم سعيد) و(سعيد متعلم)، و(يجود مصعب) و(مصعب جواد). ففي هذه الأمثلة جميعها، يدل الفعل على التجدد والحدوث، والاسم يدل على الثبوت (السامرائي، ١٩٨٩). وهذا بدوره يتيح سعة في التعبير للمتكلم في التقديم والتأخير، إذ ان الكلمة تحمل معها مركزها في الجملة بعلامتها الإعرابية، فالجملة الآتية مثلا يمكن صوغها في عدة صور (بُنى سطحية) مع بقاء المعنى العام واحدا (بُنية عميقة):

أعطى محمد خالداً كتاباً
محمد أعطى خالداً كتاباً
كتاباً أعطى محمد خالداً
كتاباً خالداً أعطى محمد
أعطى خالداً كتاباً محمد
أعطى خالداً محمد كتاباً

وسواها من الصور الأخرى دون أن يحصل لبس بين المعطي والآخذ، فالمعطي في كل هذه الجمل هو (محمد)، والآخذ (خالد) وهو معلوم من حركة الأثنين؛ فالرفع يشير الى الفاعل والنصب الى المفعول. في حين هناك تقييد ولا يمكن التعبير عن مثل هذا الحدث في اللغات المبنية إلا بصورة واحدة ضيقة لا تتعداها، فهذه الجملة يقابلها في الانكليزية:

Mohammad gave Khalid a book

ولا يمكن صياغة صيغة ثانية لها، إلا من خلال تغيير أساس في الجملة، أو تغيير في المعنى، في حين أمكن تكوين سبع صور في العربية لهذا التعبير. إذ إن الإعراب يعطي المتكلم حرية وسعة بعكس البناء. ولا يقتصر الأمر على ذلك للمفرد، بل يشمل المثنى والجمع المذكر، فيُرفع المثنى بالألف، ويُنصب ويُجر بالياء ....، أما الجمع المذكر فإنه يرفع بالواو ويُنصب ويُجر بالياء.

٤- عن النحو المُحَوسَب

يُعدُّ النحو (Syntax) أحد المنظومات الفرعية الأساسية المكونة للمنظومة اللغوية التي تشمل أيضا على منظومات الصرف (Morphlogy)، والدلالة (Semantics)، والمعجم (Lexicon)، والصوتيات (Phonolgy)، والمقاميات (Pragmatics).

ويختص النحو في دراسة بنية الجملة دون معناها، وذلك من حيث تركيب عناصرها أو مكوناتها والعلاقات البنائية المتمثلة بالوظيفة التي تربط هذه العناصر. فالجملة ليست تعاقبا لكلمات، بل هي هيكلية ترتبط عناصرها من خلال قواعد محكومة بضوابط وقيود.

إن معالجة منظومة النحو حاسوبيا موضوع متعدد الجوانب، وذو تفاصيل دقيقة ويصب فيه الكثير من النظريات النحوية الحديثة وأساليب الذكاء الاصطناعي المتطورة (Shapiro, 1990).

وتمثل معالجة النحو حاسوبيا صلب اللسانيات الحاسوبية، وتواجه معالجة النحو العربي حاسوبيا مشكلات وصعوبات عديدة ومتداخلة ناجمة من خصوصية اللغة العربية، يمكن تلخيص تلك الصعوبات بالنقاط الآتية (علي، ١٩٨٨):

أ- غياب صياغة شكلية (formal) ورياضية للنحو العربي.
ب- إسقاط علامات التشكيل في معظم النصوص العربية.
جـ- تعدد حالات اللبس النحوي وتداخلها الشديد.
د- المشاكل الناجمة عن المرونة النحوية للعربية.
هـ- حدة ظاهرة الحذف النحوية.
و- قصور المعجم العربي، نحوياً ودلالياً.
ز- تعدد العلامات الإعرابية وحالات الجواز والتفضيل.
ح- عدم توفر الإحصائيات النحوية.

ويتكون المعالج النحوي المحوسب من مكونات رئيسة هي:

أ- المعجم، متضمنا المعطيات النحوية والدلالية للمفردات.
ب- قاعدة المعرفة النحوية، وتشمل قواعد النحو، وقيود الإنتقاء الدلالي التي تضمن توافق الأفعال مع عناصر إسنادها، والأسماء مع مكملاتها وملحقاتها، ...
جـ- روتينات برنامج المعالجة، وهي عبارة عن سلسلة من الإجراءات البرمجية التي تتعامل مع المعجم، وقاعدة المعارف النحوية.
د- برنامج التحكم، وهو الذي يحدد التسلسل الذي يتم به تنفيذ الروتينات المختلفة بحيث يمكن تحليل الجملة حاسوبيا في أقصر وقت ممكن، وبأقل موارد ممكنة.

لقد أسهم تطور النظريات اللغوية الى تطور وتنوع الستراتيجيات الحاسوبية في معالجة اللغات الطبيعية حاسوبيا، إذ تعتمد هذه الستراتيجيات على أنظمة قواعد صيغت بلغات برمجية مثل لغة البرمجة المنطقية (Prolog)، اللغة الأكثر تطبيقا في تطبيقات الذكاء الإصطناعي وأبحاثه. ومن بين تلك أنظمة القواعد نجد (Shapiro, 1990):

– نظام قواعد العبارة المحدد
(Definite Clause Grammar) (DCG)
– نظام قواعد الشبكات الإنتقالية المعززة
(Augmented Transition Neworks (ATN)

هناك تطبيقات عديدة للمعالج النحوي، من بينها

القدرة على تصحيح الأخطاء النحوية حاسوبيا، وهذا يختلف عن التصحيح المُحوسب للأخطاء الإملائية الذي يُجريه المعالج الصرفي المُحوسب. وعدم كفاية التصحيح الإملائي لاكتشاف جميع الأخطاء التي تَرِد في النصوص، خاصة بالنسبة للغة كالعربية تتميز بحدة التآخي النحوي الذي يربط بين كلمات الجملة. فالتصحيح الإملائي يعمل على مستوى الكلمة، لذا فلا يمكنه إكتشاف أخطاء مثل: (المواطنون السلبيين بجعلوا مصيرهم الى واضعون القوانين)، ففي هذا المثال هناك ثلاثة أخطاء نحوية (عدم مطابقة الصفة مع الموصوف، وعدم مطابقة الفعل والفاعل، وعدم حذف نون المضاف) وهي أنواع من الأخطاء التي لا يمكن للمصحح الإملائي أن يكتشفها، إذ لا يهمه سوى صحة الكلمات المنفردة (علي، ١٩٨٨).

يكتشف المصحح النحوي المُحوسب الخلل النحوي، ويتطلب ذلك إضافة إمكانيات جديدة على المعالج النحوي ليمكنه التعامل مع حالات الخطأ المختلفة، وربما تطلب ذلك وضع مجموعة من القواعد لوصف الحالات الشاذة للخطأ النحوي، إن وضع هذه القواعد يكون على أساس السلامة النحوية، إذ أن معظم الأخطاء النحوية ما هي إلا حيودا طفيفا عن النمط السليم للجُمل، وعلى الباحثين اللغويين والحاسوبيين محاولة وضع نظام قواعد للحيود النحوي في العربية المشكولة وغير المشكولة (علي، ١٩٨٨).

من السهل الحكم على الصحة النحوية للجملة العربية ككل، إذ سيفشل المُحلل المُحوسب في الوصول الى إعراب صحيح لها، المشكلة هي في تحديد موضع الخلل النحوي، وفي هذا الصدد، على المصحح النحوي المُحوسب أن يفترض نوع الخطأ الذي سبب فشل المحلل الإعرابي، ليقوم بناء على ذلك بتعطيل بعض إمكانات النظام المحوسب كمحاولة لتمرير الجُمل الخاطئة، علاوة على ذلك، يجب على المصحح النحوي المُحوسب الإحتفاظ بتحليل دقيق لخطوات عملها أثناء تحليل الجملة، يمكن تحديد موضع الخطأ على ضوئها.

٥/ تصميم منظومة حاسوبية لتصحيح الأخطاء النحوية في الجمل العربية

ضُمنت المنظومة كما مبين في المخطط رقم (١) من أجزاء أساسية، تُستهل بالمُدخلات، وهي عبارة عن جملة عربية بسيطة. بعدها يأتي دور تحليل الجملة، وفيه تقارن القواعد المؤلفة للجملة مع القواعد المعرفة في المنظومة، إذ وُضع تصميم خاص بالقواعد المُعرفة في المنظومة تشمل على قواعد تكشف الخطأ ومحاولة تصحيحه. أما بالنسبة للأخطاء التي يمكن للمنظومة إكتشافها، فهي تلك الأخطاء التي تأتي بين المُسند والمُسند إليه، سواء كانا فعل وفاعل، أو مبتدأ وخبر عندما يكون الخبر جملة فعلية. فقد ناقشنا التطابق بين الفعل والفاعل من حيث:

- العدد (الإفراد والتثنية والجمع)
- الجنس (المفرد والمذكر)
- الحالات الإعرابية (الرفع والنصب) بالنسبة للمثنى وللجمع المذكر السالم.

التطابق –آنف الذكر– يخص الناحية النحوية، وقد ناقشنا فضلاً عنه بعض التطابقات الدلالية بين المُسند والمُسند إليه في الجملة من خلال تضمين المعجم بعض السمات الدلالية التي تحقق ذلك. ففي حالة عدم وجود خطأ نحوي في الجملة المدخلة تُرسل رسالة للمستفيد عبر واجهة المستفيد. أما في حالة وجود خطأ نحوي، فهناك جزء يقوم بتصحيح الجملة.

لقد تناول التصحيح مدار البحث، مفهومين، الأول التصحيح بأسلوب البنية العميقة، أما الأسلوب الآخر هو الأسلوب المباشر. فالأسلوب الأول يستخدم قواعد تحويلية لإيجاد البنية العميقة لجملة العربية المدخلة، ومن ثم تصحيحها على وفق القواعد العربية التي تولّد الجملة الصحيحة. وعلى سبيل المثال، فأن تصحيح جملة مثل (لعب الولدين الكرة).

لعب	الولد	ين	الكرة	(بُنية سطحية)
1	2	3	4	
فعل ماضي	اسم منصوب مثنى			↓
لعب	الولد	// ان	الكرة	(بُنية سطحية)
1	2		4	
فعل ماضي	اسمُ مرفوع مثنى			

إذ إن الخطأ كما هو واضح فيها، إن حالة الفاعل

هي النصب، بينما يجب أن تكون الرفع، والتصحيح المبين يكون نفسه في النوعين من التصحيح المباشر أو بإستخدام القواعد التحويلية، وهو إجراء تبديل لاحقة الاسم التي تدل على النصب وهي (الياء والنون) وتبديلها بمثيلتها الدالة على الرفع وهي (الواو والنون)، وذلك لأن بُنية الجملة هو (ف فا مف) وهو ما يمثل البنية الأساس للجملة العربية.

وكذلك الحال بالنسبة لجملة مثل (لعبا الولدان الكرة)

لعب	// ا	الولدان	الكرة	(بُنية سطحية)
1	2	3	4	
فعل ماضي	ضمير	اسم مرفوع مثنى		

↓

لعب	∅	الولدان	الكرة	(بُنية سطحية)
1		3	4	
فعل ماضي		اسم مرفوع مثنى		

إذ يعني الرمز // عملية دمج أو إلصاق، اما الرمز ∅ فيعني عملية حذف. وفي مثالنا السابق حُذف (حرف الألف) الملصق بالفعل (لعب) الدالة على المثنى.

أما جملة مثل (الولدان لعبا الكرة)، فالتصحيح النحوي اللغوي يأخذ منحيين:

بنية سطحية   الولدان لعب الكرة

↓

بنية سطحية   الولدان لعبا الكرة

أما المنحنى الاخر، فهو على وفق التحويلية بإيجاد البُنية العميقة من البُنية السطحية، أي إرجاع الجملة الى بُنيتها الأساسية (ف فا مف)، فتكون بذلك الجملة الصحيحة هي (لعب الولدان الكرة).

بُنية سطحية   الولدان لعب الكرة

↓

قواعد تحويلية بُنية عميقة (إبدال موقعي المُسند والمُسند إليه)

↓

بُنية سطحية   لعب الولدان الكرة

وفي جملة (لعبت الولد الكرة) يكون التصحيح هنا بالشكل الأتي:

لعب	// ت	الولد	الكرة	(بُنية سطحية)
1	2	3	4	

↓

لعب	∅	الولد	الكرة	(بُنية سطحية)
1		3	4	

إذ يعتمد التصحيح هنا على السمات الدلالية المعرفة ضمن مفردات المعجم، ففي هذه الحالة (التاء) الملحقة بالفعل (لعب) تدل على التأنيث، بينما الفاعل هو مذكر، لذلك حذفت التاء من الفعل.

٦- تنفيذ المنظومة

بعد أن أجرينا تصميم المنظومة، كُتبت برامج حاسوبية لتنفيذها، وذلك بإستخدام لغة البرمجة المنطقية (Prolog)، ونُفذت المنظومة على حواسيب نوع (IBM/PC) والحواسيب المتوائمة معها.

وقد مُثلت قاعدة المعرفة المتمثلة بالمعجم والقواعد اللغوية بإستخدام نظام قواعد العبارة المحددة (DCG)لتحليل الجُمل العربية المُدخلة وتوليد الجُمل الصحيحة.

تُعرف القواعد اللغوية لأنواع الجمل العربية التي ستجري لها المعالجة الحاسوبية في المنظومة مكونات تلك الجمل بلغة البرمجة (Prolog) كما يأتي:

```
domains
 sentence = sentence_S_V_O(subj, verb,
    obj);
 sentence_S_V(subj, verb, obj);
 sentence_V_S_O(subj, verb, obj);
 verb = verb(string)
 subj = noun(string); pron(string)
 obj = noun(string); empty()
```

اما تعريف المحمولات الأساسية فهو بالشكل الآتي:

```
predicates
sentence(strlist, sentence)
v_sentence(strlist, sentence)
n_(strlist, sentence)
```

القاعدة analyze تتعرف أولا على الجملة فيما إذا كانت جملة أسمية أو فعلية:

```
analyze(Sentence):
n_sentence(Sentence);
v_sentence(Sentence).
```

إذ تحلل القاعدة n_sentence الجملة الأسمية، والقاعدة v_sentence تحلل الجمل الفعلية، وتصححان الخطأ بأسلوبي البنية العميقة بالقاعدة correct_S_Deep_S، والأسلوب المباشر فيتم عن طريق القاعدة correct_S_to_S وكما مبين في الأجزاء الآتية:

```
n_sentence(Sentence, sent_S_V_O
   (Subject, Verb, Object)):-
  for_ subj(Sentence, Rest, Subject),
  for_verb(Rest0, Sent_Rest1, Verb),
  for_obj(Sent_Rest_Rest2, Object),
  dote(Sent_Rest2),
  correct_S_Deep_S(OldSent, NewSent).
v_sentence(Sentence, sent_V_S_O (Verb,
  Subject, Object)):-
  for_subj(Sentence, Rest, Verb),
  for_verb(Rest0, Rest1, Subject),
  for_obj(Sent_Rest1, Sent_Rest2,
  Object),
  dote(Sent_Rest2),
  correct_S_to_S(OldSent, NewSent).
v_verb([Verb|Rest], Rest, verb(Verb)):-
   verb_type(Verb, _).
v_subj([Subj|Rest], Rest, subject(Subj)):-
   subj_nominal(Subj).
v-obj([Obj|Rest], Rest, object(Obj)):-
   noun(Obj).
```

نفذت المنظومة على مجموعة من الجمل وفيما يأتي بعض الامثلة:

- جملة (اللاعبون يلعبان الكرة) تصحح الى (اللاعبان يلعبان الكرة) أو (اللاعبون يلعبون الكرة)
- جملة (اللاعبان لعب الكرة) تصحح الى (اللاعبان لعبا الكرة) أو (لعب اللاعبان الكرة)
- جملة (يلعب اللاعبة الكرة) تصحح الى (تلعب اللاعبة الكرة)

٧- الإستنتاجات

بعد إجراء تجربتنا البحثية في بناء منظومة حاسوبية لتصحيح بعض الأخطاء النحوية وتطبيقها على بعض الجُمل العربية، وذلك من خلال إجراء التطابقات النحوية والدلالية بين المُسند والمُسند إليه (الفعل والفاعل) في الجملة العربية، إذ تضمنت المنظومة نوعين من تصحيح الأخطاء: الأول يُعنى بالتصحيح المباشر، وذلك بتحليل البُنية السطحية للجملة المدخلة، وتوليد بُنية سطحية للجملة الصحيحة، أما النوع الأخر فيُعنى بتحليل البُنية السطحية المُدخلة وإيجاد بُنيتها العميقة، وتصحيح الأخطاء النحوية على وفق البُنية العميقة ومن ثم تحويلها الى بُنية سطحية تمثل الجملة بصورتها النحوية الصحيحة.

توصلنا الى حقيقة مهمة، مفادها أن بُنية الجملة العربية المتمثلة بالصيغة (فا ف مف) لا تقل أهمية عن البُنية الأساس للجملة العربية المتمثلة بالصيغة (ف فا مف) من ناحية التطبيق المحوسب، وذلك لسبب مهم جدا، وهو أن جملة مثل (نحن نلعب الكرة)، على سبيل المثال، لا يمكن إلا أن تكون لها هذه الصورة في التعبير اللغوي، أي الصورة التي يظهر فيها المُسند وهو (الضمير نحن) قبل الفعل، بينما لا تكون الجملة مقبولة عندما تكون بالشكل (نلعب نحن الكرة)، بل الأصح القول (نلعب الكرة) بدون ذكر الضمير (نحن) الذي يدل على (الجمع المخاطب)، إذ ان (النون) في بداية الفعل إنما تدل على تلك الدلالة. فالضمير المنفصل عندما يأتي في بداية الجملة قبل الفعل يجب أن يذكر، لذلك يجب أن

يكون هناك وصفا شاملا للمفردات اللغوية وبما يتناسب وموقعها في الجملة، الوصف الذي نعنيه هو إبراز دور السمات النحوية والدلالية معجميا.

إضافة الى ذلك فإن تصحيح الجملة من خلال البنية العميقة يمكن أن يولد صيغة واحدة دائماً لمجموعة من الجمل في البنى السطحية، مما يؤدي الى سلامة لغوية من ناحية البُنية الأساسية للجملة العربية، لكن هذا يؤدي الى أن يكون هناك تمثيل لقواعد تحويلية عديدة في المنظومة الحاسوبية، بلا شك، تؤدي بشكل او بأخر، الى تأخير في عملية البحث (search) عن القاعدة النحوية المطابقة ضمن مجموعة القواعد المعرفة في المنظومة، لا سيما إذا كانت تلك القواعد كثيرة ومفردات المعجم هائلة.

إذا كنا وضعنا الأستنتاجات -آنفة الذكر- في مقدمة الإستنتاجات المهمة في تطبيقنا البحثي، فإننا لا يمكن أن ننسى الدور الذي لعبته الصياغة الشكلية والرياضية للقواعد اللغوية، سواء كانت للطريقة المباشرة، أو تلك التي أعتمدت القواعد التحويلية والبُنية العميقة لتصحيح الأخطاء النحوية، هذا الدور الذي سهل في صياغة القواعد على شكل تمثيل معرفة حاسوبية.

المصادر

١- حمودي، زهير شاكر (١٩٩٢). تصميم نظام لتصحيح بعض الأخطاء النحوية. الجامعة التكنولوجية-قسم علم الحاسبات (رسالة ماجستير).

٢- خرما، نايف (١٩٧٨). أضواء على الدراسات اللغوية المعاصرة. سلسلة عالم المعرفة: المجلس الوطني للثقافة والفنون والأداب، الكويت.

٣- الخولي، محمد علي (١٩٨١). قواعد تحويلية للغة العربية: دار المريخ للنشر، السعودية - الرياض.

٤- السامرائي، فاضل صالح (١٩٨٩). معاني النحو - الجزء الأول: بيت الحكمة للطباعة والترجمة والنشر - جامعة بغداد.

٥- الشبشيني، هشام وأيمن النجار (١٩٨٩). محللان نحويان للجمل العربية عن طريق الحاسب الآلي. وقائع أبحاث المؤتمر الثاني حول اللغويات الحسابية العربية: معهد الكويت للأبحاث العلمية.

٦- فاخوري، عادل (١٩٨٨). اللسانية التوليدية التحويلية: دار الطليعة، بيروت.

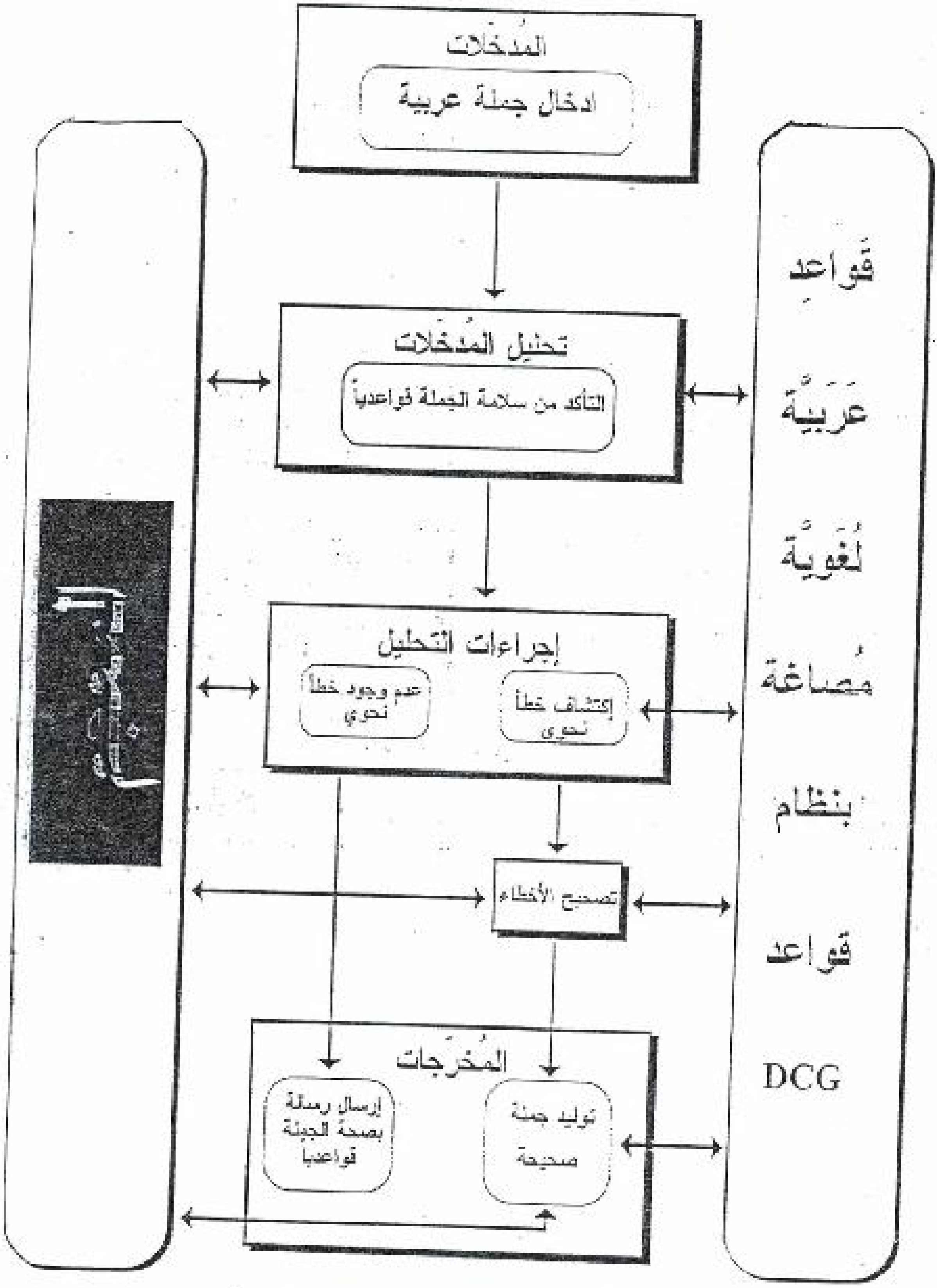
٧- الفهري، عبد القادر الفاسي (١٩٨٩). تحليل الجملة والمفردات العربية - ملامح التداخل والإشتقاق الواقعية. وقائع أبحاث المؤتمر الثاني حول اللغويات الحسابية العربية: معهد الكويت للأبحاث العلمية.

٨- عثمان، أكرم محمد (١٩٩٠). قواعد تحويلية للغة العربية لبناء نظم قواعد المعرفة. جامعة بغداد - كلية العلوم (رسالة ماجستير).

٩- علي، نبيل (١٩٨٨). اللغة العربية والحاسوب: مؤسسة تعريب للنشر - تركي العريض، الكويت.

10- Bakir, M.J., (1980). Aspects of clause structure in Arabic. A study in Word order variation in litrary Arabic. Ph.D. Thesis, Indiana University.

11- Shpiro S.C., (1990), Encyclopidia of Artifitial Intelligence. New York:Wiely-Interscince.



مخطط رقم (١): الهيكل العام لمنظومة تصحيح الأخطاء النحوية.

Algorithm	Signing time (sec)	Verification time (sec)
original DSA	43.83	72.39
McCurley	44.26	71.1
proposed	43.94	41.52

Table (1): Computation time comparison of the three algorithms

Appendix

A proof that v = r for the proposed DSA

The following steps prove that v = r for the proposed DSA:

$$v = (y^{u2} \bmod p) \bmod q$$
$$= (y^{(u1.s) \bmod q} \bmod p) \bmod q$$
$$= (y^{(((H(m)+r) \bmod q).s) \bmod q} \bmod p) \bmod q$$
$$= (y^{((((H(m)+r) \bmod q).s) \bmod q) \bmod q} \bmod p) \bmod q$$

{by lemman 1 of appendix A in Ref.[6]}

$$= (y^{(((H(m)+r).s) \bmod q} \bmod p) \bmod q$$
$$= (g^{x.(((H(m)+r).s) \bmod q} \bmod p) \bmod q$$
$$= (g^{(x(((H(m)+r).s) \bmod q) \bmod q} \bmod p) \bmod q$$
$$= (g^{((H(m)x+xr).s) \bmod q} \bmod p) \bmod q$$

let $w = (H(m)\,x + xr)^{-1}$

$$= (g^{((H(m)x+xr)(k.w) \bmod q) \bmod q} \bmod p) \bmod q$$
$$= (g^{((((H(m)x+xr).w) \bmod q)(k \bmod q)) \bmod q} \bmod p) \bmod q$$
$$= (g^{(k \bmod q)} \bmod p) \bmod q$$

Since $0 < k < q \rightarrow k \bmod q = k$

$$= (g^{k} \bmod p) \bmod q$$
$$= r$$

receives r', s' and m' computes u1, u2 and v as follows.

$$u_1 = (H(m') + r') \bmod q$$
$$u_2 = (u_1 . s') \bmod q \qquad \dots (4)$$
$$v = (y^{u2} (\bmod p)) \bmod q$$

then he verifies the validity of $r' = v$

It is anticipated that this method is much faster than McCurly verification method since it includes one modular addition, one multiplication and one exponentiations as compared with McCurly verification method which consists of three modular multiplications and two modular exponentiations. Such improvement in signature validation speed is highly welcomed in most applications. For a proof that $v = r'$, see the appendix.

IV- Results and Conclusion

A comparative study of the computation time required for signing and verifying an experimental message using the three algorithms, i.e., the original DSA of NIST, McCurly improvement on DSA and the proposed algorithms is given in Table (1) below. It shows clearly the anticipated improvement in the verification time for the proposed algorithm as compared with the other two. It should be pointed out that, similar computational algorithms for multiplication and exponentiation are used on the same computer for the three above methods in order to make sure that the noticed decrease in the verification time is totally due to the algorithm itself.

However, the drawback of this work is that the solution of equation (3) might be a bit less difficult especially for small values of x & k than McCurley equation, but this factorization problem gets more difficult as the value of x & k increases.

The attacks on this problems comes in two approaches, either deriving the secret key c from rearranging equations or getting k from rearranging equation (1) then find x with the help of k, i.e., discrete algorithm first and then factorization, the second approach is achieved by forging a signature pair for a message, then by trial and error choosing r and try to find s which is even more difficult than discrete logarithm problem.

Acknowledgment

The authors would like to thank the staff of the computer institute and bureau, Basrah, Iraq, for their help in typing out the manuscript of the paper.

References

1- National Institute of Standards and Technology, "A proposed federal information processing standard for digital signature (DSS)" Federal Register, Vol. 56, No. 169, Aug. 30, 1991.

2- NIST, "The digital signature standard", communication of ACM, Vol. 35, No.7, July 1992.

3- Rivest, R. L., Hellman M. E. And Anderson I. C. "Responses to NIST's proposal" communication of ACM, vol. 35, No.7, July 1992.

4- K. S. McCurely "An open comment letter from the Sandia National Laboratory on the DSA of the NIST", Nov.7, 1991.

5- Yen S. M. And Laih C. S. "Improved signature Algorithm" IEEE Trans. On computers" Vol. 44, No. 5 May 1995.

6- K. I Arif "Some improvement on the digital signature standard" M.Sc. Thesis, College of Science, Basrah University, Iraq.

verify the signature. A message digest or a condensed version of the data is obtained using a hash function. This digest is signed and sent to the intended recipient together with the message. The receiver must use the same hash function with the senders public key to verify the signature.

DSA Parameters[2]

p: a prime modulus where $2^{511} < p < 2^{512}$

q: a prime devisor of (p-1), where $2^{159} < q < 2^{160}$

g $= h^{(p-1)/q} \bmod p > 1$, where h is random integer $0 < h < p$.

x: an integer secret key, such that $0 < x < q$

y = $g^x \bmod p$, public key where $0 < y < p$

m: the message to be signed and transmitted

H: a random way hash function.

K: a random integer, $0 < k < q$.

The integers p, q and g are public and can be common for a group of users, while x and k are secret. Where k must be changed each time a message is signed.

Message Signature and Verification

For a message m, the signer computes

$r = (g^k \bmod p) \bmod q$

$s = (k^{-1} (H(m) + xr)) \bmod q$ ... (1)

r and s contain the signature of the message m, which are transmitted along with the message m to the receiver.

At the receiver side, getting m', s' and r' the signature is verified by computing

$w = s'^{-1} \bmod q$,

$u_1 = w\, H(m') \bmod q$ and ... (2)

$u_2 = (r'w) \bmod q$,

then verifying the validity of the following equation:

$v \overset{?}{=} (g^{u1} \cdot g^{u2}) \bmod p) \bmod q$.

The signature is verified if $v = r'$ which makes the receiver confident of the originator of the message.

II. Developments of DSA

Two major improvements where reported on the digital signature algorithm suggested by McCurley[4] and Yen and Laih[5]. Yen & Laih suggested an improvement which benefits from computing the modular inverse of the secret key (i.e., $x^{-1} \bmod q$) in advance and use it for each signature generation, while McCureley and Yen & Laih suggested independently a method resulted in the elimination of the computation of the $(s^{-1}) \bmod q$ in the verification part of the original DSA version. These suggestions proved to give shorter verification time. The modified algorithms presented by both are outlined and compared with each other and with original DSA in[6].

III- The Proposed Algorithm

The message m is signed by computing the following:

$r = (g^k \bmod p) \bmod q$

$s = (kx\,(H(m) + r)^{-1}) \bmod q$ ... (3)

The pair of numbers, r and s, constitute the signature using the user public key y and secret key x.

To verify the signature, the recipient

Shorter Signature Verification Time With Improved Digital Signature Standard, DSS

Hamza A. Al-Sewadi
Department of Computer Engineering, College of Engineering,
Basrah University

&

Khaldon I. Arif
Department of Computer Science, College of Science,
Basrah University

Key words: Digital signature, Digital Signature Algorithm (DSA), Public key cryptography, Ciphering, Modular arithmetic.

Abstract

An improved version of the digital signature algorithm (DSA) of the National Institute of Standard and Technology (NIST) is developed. The modification has led to considerable improvement in the signature verification speed. The results are compared with those of original DSA and McCurley's suggested algorithms.

The security of the improved version is somewhat affected as it is changing to be a factorization problem together with discrete algorithm rather than discrete algorithm alone. The difficulty of factorization increases as the parameters increase in value.

I- Introduction

The digital signature standards proposed by National Institute of Standard and Technology NIST, comprise of Digital Signature Algorithm (DSA). It is a pair of large digital numbers, appropriate for applications require digital rather than written signatures[1,2]. Many different responses to DSA of NIST outlining the pro and against aspects of this algorithm were reported, such as those given by Rivest, Hellman and Anderson[3].

However, digital signature is produced on a computer following a set of rules and parameters enabling it to be used to verify the identity of originator and the integrity of the data. The system includes signature generation and verification, the signature generation is achieved by using private key, while verification is achieved by using public key which corresponds to the private key.

Each user processes his own pair of keys, one of them is public, which is published in a directory containing all users public keys and the other is private and only known by the originator. Signature can only be generated by the owner of the secret key while any one can

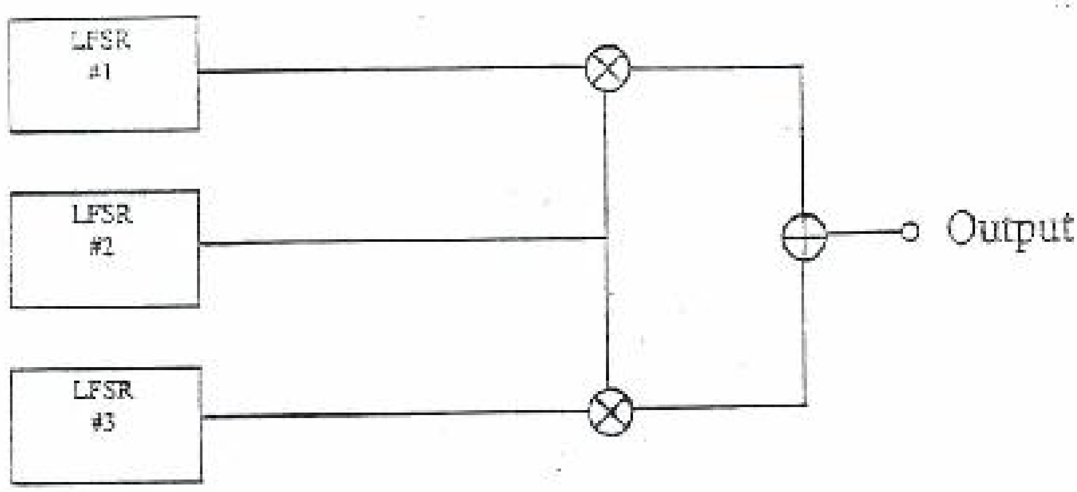


Fig.(2): Geffe generator.

Appendix (A)

```
/*-----------------------------------------------------------------------*\
            Display procedure for the BMP image file format
/*-----------------------------------------------------------------------*\
          void display-bmp (char * pic-name, int start)
  {       int              x = 0, Y = 0, left  = 0, right = 0;
          long int         i = 0;
          char             value =";
          FILE             *fp;
  if ((fp = fopen (pic-name, "rb")) = = NULL)
  Printf ("Cannot open File!");
  if (fseek (fp, start, SEEK-SET) ! = 0)
  Printf ("Error on file format !");
  i = x = 0;    y = 349;
  do {
  fread (& value, 1, 1, fp);            left = right = value
  left = (left & 240) >> 4;             right = right & 15; i ++;
  putpixel (x, Y, left);                putpixel (x+1, Y, right);
  x + = 2;                              if (x > 639) {x = 0; Y-;}
  } while (i < = 112000);
  fclose (fp);
  }
/*-----------------------------------------------------------------------*\
```

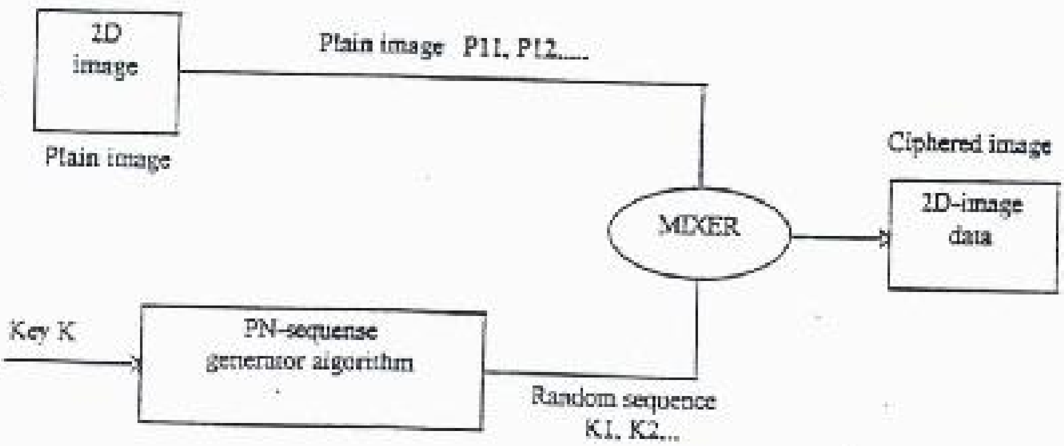


Fig (1): Block diagram of pseudo-random number cipher.

pictures.

- In this project we studied different type of already executable programs to display pictures like (SHOWPIC. EXE, ST. EXE), we distinguish between these already executable program and (SHOWPUT. EXE, SHOWPCX. EXE, SHOWBMP. EXE) program, see Table (1) which illustrate a comparison on size between these already executable program and our executable program.
  Table (1) illustrates the advantage of SHOWPUT. EXE, SHOWPCX. EXE, and SHOWBMP. EXE to help the user to display pictures without any warred about the size of memory.
- Design of an image encryption program (incrypt/ decrypt) technique (PROTEPIC. EXE) that can be used to encrypt/ decrypt picture under extension (PUT, BMP, PCX).
- Development for an easy and fast way to display and protect images with Turbo C 2.0

Executable Program	Size (Kbyte)
SHOWPIC. EXE	10037
ST. EXE	100838
SHOWPUT. EXE	7880
SHOWBMP. EXE	8972
SHOWPCX. EXE	9066

Table (1): Size of executable program.

References

- Copyright (C), 1989, Image 72, Promoting Enterprise Co. LTD.
- Dutton, Gail, 1996, Choosing a graphics file format, PC Novice, Vol.4, pp. 34-35.
- Ezzell, Ben, 1989, Graphics programming in Turbo C 2.0, Addison-Wesley Publishing Comp.
- Geffe, P, R., 1973, How to protect data with ciphers that are really hard to break, Electronics, Jan.
- Graef, Gerald L., 1989, Graphics formats, Byte, Vol.9, pp, 305-310.
- Hearn, D. And Baker M., 1986, Computer graphics, Prentice-Hall International Inc.
- Hearn, D. And Baker M., 1997, Computer graphics C version, Prentice-Hall International Inc.
- IBM, Corporation (C), 1990, STORY Board Live Help, U. S. Edition.
- John, R. Rankin, 1989, Computer graphics software construction, Prentice-Hall.
- Komura M. F. & Shiraishi T., 1986, An Encryption algorithm for digital image data, Trans. Inst. Electron. & Commun. Eng., Vol.J69b, pp. 1385-1392.
- Nelson, J., 1987, Advance graphics in C, Obsorn McGraw-Hill.
- Pavlidis, T., 1982, Algorithms for graphics and image processing, Computer Science Press Inc.

are required shift registers informations and output is a file is considered as a key file which will be XOR-ed with plain-image to produce the ciphered image. The most important user defined variables required for this program are as follows:

- Length of each shift register, Len.
- Initial value of each shift register, S.
- Output stage NO. Of each shift register, OUT.
- Length of sequence, Lenofseq.
- The feedback operations, OPR.
- The coefficients values, COF.
- SEQ1, SEQ2, SEQ3 which are arrays to store the sequence.
- Key-Byte, which is key value corresponding to one pixel value of image data.

Syntax

PROTPIC <File name> [/E, /D]

PROTPIC: executable program name.

<File name>: name of file you want to protect it.

(ext>: extension of file (PUT, BMP, PCX).

/E: encrypt.

/D: decrypt.

Algorithm

PURPOSE: Is used to generate key sequence generated from Geffe's generator, and save the output sequence into file.

BEGIN

FOR (i = 0; i < 3; i ++)/ * loop for 3 LFSR */

{

Ask the user to enter stages information which are: len, s, outs, and lenofseq.

Ask the user to enter feedback information which are: opr, and cof.

Compute the sum of stages data? * generate sequence from LFSR */

Rotate the data

Save the generated sequence into a file.

}

/* Apply Geffe's algorithm */

Set the key-byte ((seq1 and seq2) or (not seq2 and seq 3)).

Computer key-byte value for corresponding to each image data point.

Save the final generated sequence in a file seq-key

Close all open files

End of Pseudo-Random-Number

Conclusion

In this project we have attempted to give the user tools to display under extension (PUT, BMP, PCX) which are draw or scanned by using (Story board live, Image 72. Page scanner) packages into any system or programs. The major of our work focus on the execution time efficiencies although storage requirement have been taken into consideration.

In our project we constructs a tool to project the images by using encryption techniques. The project has demonstrated that (Pseudo Random Number Cipher Technique) can be an effective, efficient image encryption technique for an encrypt/decrypt images.

The main issues revealed by this work are listed below:

- You need a less space and time to display pictures under extension (PUT, BMP, PCX) into any system or program, this operation can be easily performed on

GET picture names.
OPEN the file, skip the header to the proper address.
INITIALIZE variables I, X, Y to zeros, and Y to maximum number.
ASSIGN the name location to the variables L & R.
PROCESS left-nibble, mask it with value 240, shifted right
PROCESS right-nibble mask it with value 15.
PLOT the point.
UPDATE position for X, Y.
CLOSE file.

Protection Picture (PROTPIC. EXE)

This project designed to given you the tool to protect pictures by create PROTPIC. EXE executable program written in program language (Turbo C 2.0), pictures is coded by using (Pseudo-Random-Number Cipher Technique).

3-1 Pseudo Random Number Cipher Technique

This technique is used:

1) Stream Cipher

Stream cipher algorithm are often needed for high data-rate security applications. Stream cipher can operate on allocate units as small as a bit or a character, a fact that has greatly contributed to their popularity. In general stream cipher consists of two main parts:
- Pseudo-Random Sequence Algorithm.
- Mixer.

The key is fed to an efficient algorithm which use the key to generate an infinite sequence (the algorithm is usually referred to as the key stream generator). Sequence generated by a good stream cipher is called a Pseudo-Random sequence.

(PN sequence). This sequence generated then is mixed with the plain data by using the mixer to generate the cipher data.

Key stream can be generated in a number of different ways, but nearly all of these methods employ shift registers, Shift register is a cascade connection of binary memory elements, which are controlled in such a way that the binary contents stored in the elements may be shifted along the register. (Komura, 1986, Geffe, 1973).

2) Encrypt 2D-Image Using Geffe's Generator

2D-Image data can be encrypted by using stream-cipher, see Figure (1). Based on the good properties of the key sequence we use an efficient NLFSR (non-linear feedback shift register) proposed by GEEFE, with shift register of length (23, 19, 7) to generate key stream of size (112000 bytes). This key stream will be XOR-ed with the image-data to produce the cipher image. (Komura, 1986, Geffe, 1973).

3) Decrypt 2D-Image Using Geffe's Generator

To reconstruct cipher image, apply the decipher algorithm by making XOR between the ciphered image and the same key stream generated before, see Figure (2).

3-2 PROTPIC. EXE Program

This program is designed to generate a file of random number sequence. Its inputs extension.

OPEN the file, READ information from header part assign the values to the variables of COLO, COHI, ROLO, ROHI.

CALCULATE the real values to the rows & Columns.

READ information byte by byte then, check it bit by bit for each.

If bit = 0 THEN PLOTE a WHITE pixel ELSE PLOTE a BLACK pixel.

CLOSE file.

The executable program SHOWPUT. EXE, written by using (Turbo C 2.0).

2-2 SHOWBMP program

SHOWBMP program used to display pictures rescanned under page scanner. Bit Maps (also called pixel or raster graphics) are the most common type of graphics file format in the PC world. Bit maps break the graphic into a grid, with a light value assigned to each block (or pixel) or the grid bit maps excel at recording complex and subtle images such as photographs and computer screen displays. (Graef, 1989, Dutton, 1996).

Syntax

SHOWBMP <File name>. BMP.

SHOWBMP: executable program name, write in small or capital letters.

<File name>: name of file you want to display it under extension BMP.

Algorithm

X - > X-axis.

Y - > Y-axis.

I- > total number of pixels.

GET picture names.

OPEN the file, skip the header to the proper address.

INITIALIZE variables I, X, Y to zeros, and Y to maximum number.

ASSIGN the name location to the variables L & R.

PROCESS left-nibble, mask it with value 240, shifted right

PROCESS right-nibble mask it with value 15.

PLOT the point.

UPDATE position for X, Y.

CLOSE file.

The executable program SHOWBMP. EXE, written by using (Turbo C 2.0), see appendix (A).

2-3 SHOWPCX program

SHOWPCX program used to display pictures rescanned under page scanner. The PCX bit-mapped format was designed for PC-based paint programs such as Windows paintbrush application and is one of the oldest graphics file formats, making it one of the most widely supported. (Graef, 1989, Dutton, 1996).

Syntax

SHOWPCX <File name>. PCX

SHOWPCX: executable program name, write in small or capital letters.

<File name>: name of file you want to display it under extension PCX.

Algorithm

X - > X-axis.

Y - > Y-axis.

I- > total number of pixels.

input-output technology for computers involving the creation, manipulation, and display of pictures with the aid of a computer, computer graphics represents the most recent development in improving the efficiency of communications between human beings and computer (Copyright(c), 1989, IBM, 1990), but by using these packages we can not display pictures under any programs because these is no commands used to do this operation, also if there is command used to display picture, this command have a big size (i.e., SHOWPIC. EXE used to display pictures draw by using Story Board Live package have a big size = 10037 Kbyte, if we want to use it in Foxpro package, the picture can not display and we see this message "TOO PIG TO FIT IN MEMORY").

This project is designed to given you the tools to display pictures under extension (PUT, BMP, PCX) into any packages or programs by create three executable programs (SHOWPUT. EXE, SHOWBMP. EXE, SHOWPCX. EXE).

By using this project, you may want to protect any pictures to prevent any one to see it by create executable program (PROTPIC. EXE).

SHOWPUT. EXE, SHOWBMP. EXE, SHOWPCX. EXE, and PROTPIC. EXE are written using Turbo C 2.0, because of Graphics and Turbo C are ideal partners. Turbo C's fast, device-independent graphics routines allow programmers to create high-quality graphics with minimal fuss and a speed that is addictive. As a result, Turbo C is the environment of choices for graphics programmers. Besides, that C is a structured high level language. It is very close to English language so it is very easy to read, write and understand. DOS and BIOS are full of powerful service that can be called from written Turbo C. Turbo C provides exceptionally rich set of graphics routines such as draw the forms and fit it ... etc. That can make graphics programming much easier. (Hearn, 1997, Ezzell, 1989).

Display Executable Programs

2-1 SHOWPUT program

By using SHOWPUT program we can display pictures rescanned by using Image 72 package under any programs or system. The Image 72 * PUT file format structure is a language standard. This allows other software to call the image (picture) file, where the PUT file format structure is as follow: (Copyright(c), 1989).

For example: 32 x 10 block

Horizontal pixels +	Vertical lines +	data
(Word)	(Word)	Pixels
32	10	0ffh, 0ffh, 0ffh, 0ffh
		.
		.
		10 ROWS
		.

* PUT image total length: (int ((width +7)/8) x height) + 4 bytes

Syntax

SHOWPUT <file name>. PUT

SHOWPUT: executable program name, write in small or capital letters.

<File name>: name of file you want to display it under extension PUT.

Algorithm

GET the name of the file of put

Image Guider

Ahmad S. Nori, Laheeb M. Ibrahim, Najla Badeaa
Department of Computer Science, College of Science,
Mosul University

Abstract

SHOWPUT. EXE, SHOWBMP. EXE, SHOWPCX. EXE and PROTPIC. EXE are an executable program written in (Turbo C 2.0) and tested successfully on IBM personal computer or compatibles to display pictures under extension (PUT, BMP, PCX) which are drawing or rescanned by using (Story board live, Image 72, page scanner) packages, into any system or programs, and to protect pictures by coding it to prevent any human to see it, if these pictures are importance and security.

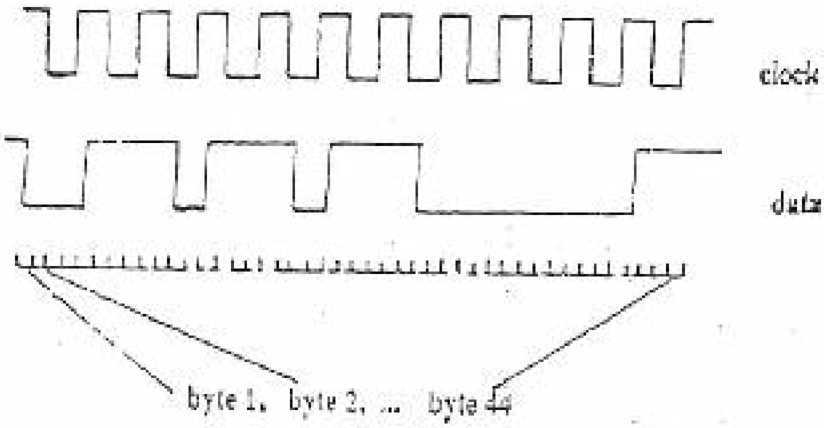
Introduction

The kinds of tasks we use computer to perform are changing. At one time computers were used primarily for accounting but now they are used for everything you can image. Analytical software depends on graphics to reveal every thing for flaws in metals to weather patterns to the properties of geologic formations. Graphics systems are used in graphic arts, medical scientific, robotics, security, and quality control applications Drafting, business planning and design applications are flourishing. (Nelson, 1987, John, 1989, Hearn, 1986).

Years ago, computer graphics were considered unnecessary impractical, or too expensive, Not so today. Hardware and software are beginning to be fast enough and powerful enough to make graphics not only feasible but essential. The emergence of desktop publishing integrates graphics with text in modern computing. Soon database applications will routinely include images as well as text. (Hearn, 1997).

Image processing is a topic of rapidly growing importance in the computer field. It has always been one of the most visually spectacular branches of computer technology, producing graphics (images) whose appearance and motion make them quit unlike any other form of computer output (Pavlidis, 1982, Ezzell, 1989).

(Story Board Live, Image 72, page scanner) are a modern visual picture-style



(80, 00, 00, 80, 0C, 40, 40, 0C,80,00,
00,80,0C,40,40,0C,80,00,00,80,0C,40
,40,0C,80,00,00,80,80,00,00,80,
80,00,00,80,80,00,00,80,80,00,00,0C)

Fig.(3): An example of scan code press signals

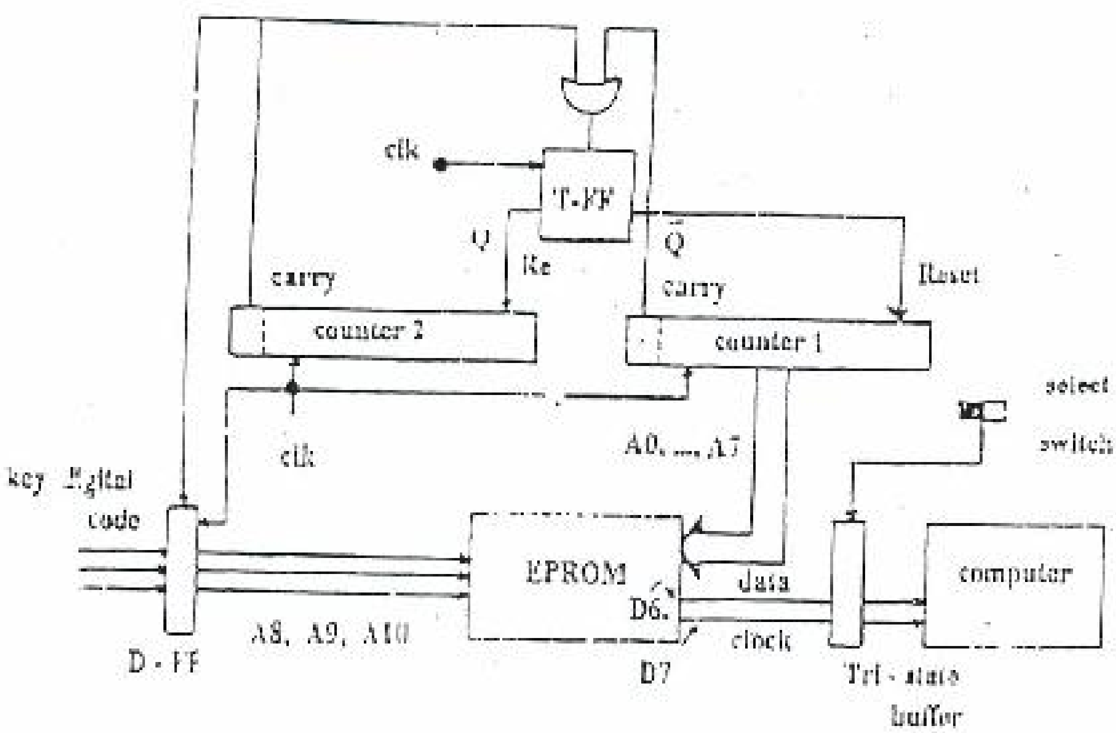
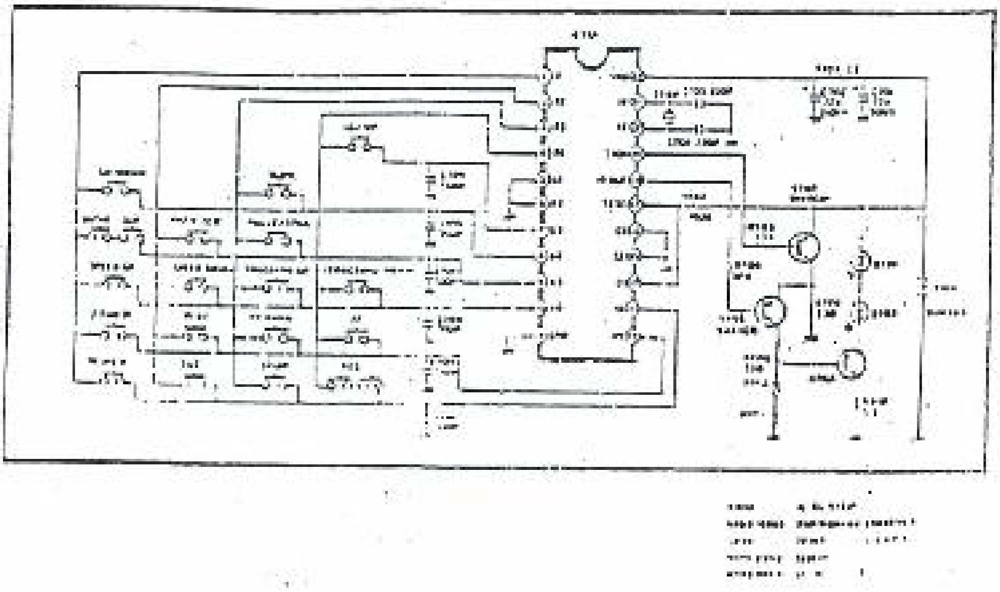
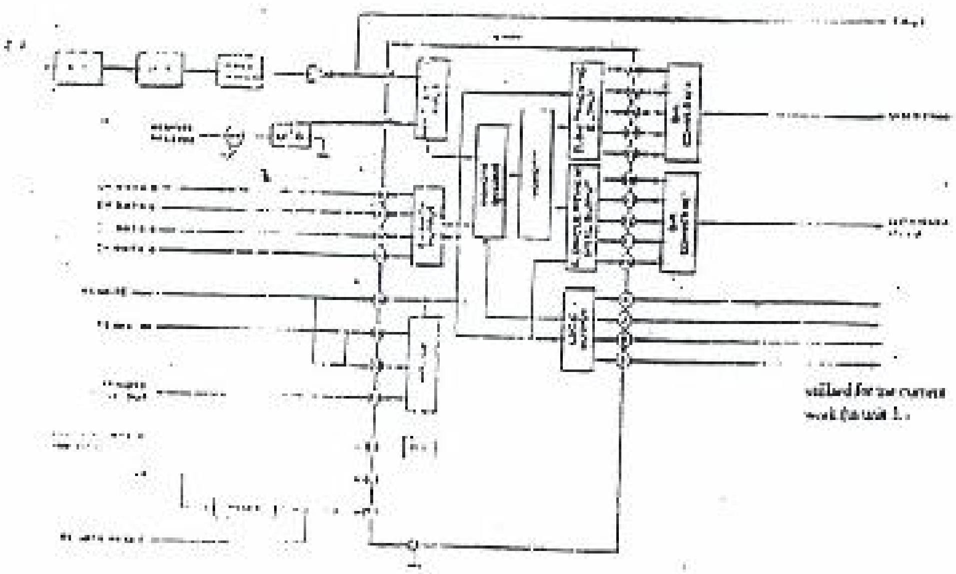


Fig.(4): Scan code generator and switching unit



a) Remote wireless remote controller circuit



b) Remote control receive block dingrium

Fig.(2): Toshiba TV remote transmitter receiver and shift registers.

The design aspect is drawn compatibly to IBM PC systems in both hardware and software objectives. Neither additional circuitries nor additional software drives are needed to mount this peripheral to the PC. Port 60h and INT 09h routine are still valid for peripheral configuration.

References

1- R. Nagarajan and W.A. Jabar, "Computer-aided testing of a DC motor", Aided Engineering Journal, June 1990.

2- W.A. Jabar, A.F. Marhoon, H.L. Saadon and H.A. Al Attar, "Microcomputer based data acquisition system for laser beam scattering pattern analysis", Basrah J. Science, Vol. 4, 1996.

3- W. A. Jabar, "Microcomputer based panel control for driving DC motor Net", Basrah J. Science, Vol. 12, 1994.

4- MIC 956 Microprocessor application module, feedback instruments Ltd, UK, 1981.

5- Labpack IBM PC application module Scientific solution press. 1985.

6- Toshiba TV model V-64 manual, Toshiba corporation.

7- Peter Abel, "IBM PC Assembly language and programming", Prentice Hall International Editions, 1987.

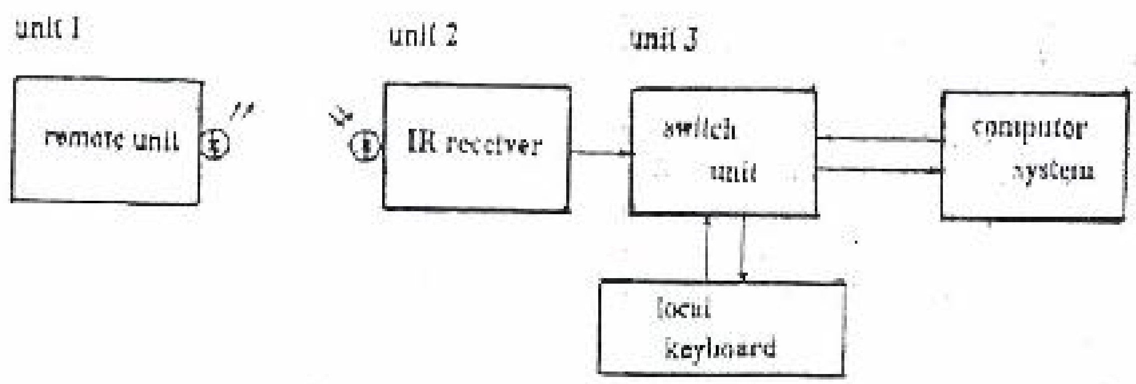


Fig.(1): Peripheral units

stored. An incremented counter (counter 1) is used to scan any segment in steps to generate the related data of a selected digital code. To avoid conflicts of an interrupter appearance of a digital code, the digital code is controlled by a D-FF to be triggered at a required intervals assuring the completion of a code and mounting a new one. To integrate this function another counter (counter 2) is used to allow for a sufficient time periods between mounting a scan code and another.

Practically, specific functions are sought emergently being driven in their remote mode. Therefore, a practical keyboard of 64 k bit for AT system is designed to have 8 segmented memory. The first segment (address = 00) is reserved for resting state that has been filled with FFH. While the rest 7 segments consists of the required scan codes.

To draw the whole scheme into PC system compatibility, the output is interfaced to the same port received to link traditional keyboard via a tri-state buffer. A switch is mounted to this peripheral in unit 3 to control the two modes of operation. By enabling and disabling this tri-state buffer, the remote keyboard is enabled or disabled in response.

3- The Application Scopes

Reasonable experimental investigation lies the usefulness of the presented peripheral to be realized by employing it in the following application scopes:

1- Control Schemes

As it is mentioned before that computers are utilized as main controllers in diverse controlling systems. In such systems, parts of these industrial plants may require adjustment processes or remote ordering. These processes practically invokes for free linking to simplify the communication of the operator to the main controller (computer to drive speed, position and other setting values of the different parameters of the control scheme.

2- Educational Laboratories

This peripheral may assist laboratory supervisor in teaching computer students who group around the computers. In these laboratories a supervisor often needed to guide his students in a procedural ordering. By the presented peripheral the supervisor assures an error free logged data to be fed to the distributed system by single keyboard. In other words, this peripheral could be regarded as an electronic pointing stick for local displays of the different groups.

3- Wireless Joystick

The growing up of computer games may employ this peripheral to add interactive features for driving game activities.

Conclusions

An infra red remote keyboard is designed and described in this paper. The design utilizes available successful schemes of infra red transmitter receiver that operate out of the wide band noise interference of computers.

Microprogramming methodology characteristics of this design. Passive network is chosen to avoid the complexity and cost in design when the structure is built on the bases of processor systems essential to control the programmable units of USART function of receiving a digital to be interpreted into a transmission serial synchronised signals of data and clock.

The design allows the implementation of either local keyboard (traditional) or a remote keyboard (the proposed keyboard) when the mode selection switch is set to the desired operation that directs the signals of either of these keyboards to the same port via the available connector reserved for traditional keyboard.

2- Infra Red Remote Keyboard Peripheral

Three units constitute the presented peripheral structure, Fig.(1). These units are:

1- Remote unit (keyboard and infra red transmitter).
2- Infra red receiver unit and,
3- Scan code generator and select switching unit.

The first two units are widely used with different instruments like Video and TV sets. These units, in spite of their modalities of structure and their operating frequencies furnish in their resultant function a digital output that decodes the operated switch present work is given in Fig.(2)[6]. The present paper will investigate mainly the third unit as forms the essential of the peripheral.

Scan Code Generation and Select Switching Unit

The main objectives attributed to this unit are to catch a switch code from unit 2 and response serial two signals set of data and clock is to be generated to the computer.

The characteristic specifications of these synchronised signals of data and clock are presented in many technical references that thoroughly discusses computer hardware view[7], and for the sake of design them, these specifications will not be discussed in detail here. However, to show design principles, an examples of a scan code in its serial form is adopted. Figure(3) gives the signals that are generated from traditional keyboard to decode a "q" character in its press details.

These characteristics in fact interpret the design performance into a low level of computer design in control unit. A code is to be interpreted into a series of control signals that is what really in instruction decoding in the control unit. This fact motivates the design towards micro-programming methodology to realize the interpretation process.

Serial signals of a scan code (data and clock) are multiplexed into a consecutive cells of memory storing unit. In our previous example these signals are interpreted into 44 code, Fig.(3). It is to be noted that the depress code is not important to be derived and the press code is quit enough to satisfy data transmission.

Then, the main task of the scan code generator on this bed is to associate a digital code with this 44 consecutive codes that are to be transmitted on after another. A suitable structure presented for this function is given in Fig.(4).

In this figure, the digital code of unit 2 is utilized to contribute the addressing structure in order to organise the memory into 8 segments. In each segment, the multiplexed form of serial scan code is

Infra Red Remote Pc Keyboard

W. A. Jabbar
Computer science Department, College of Science
Basrah University, Basrah, Iraq

Abstract

An infra red remote keyboard for PC is presented. The design aspect of this keyboard is based on a passive architecture that multiplexes the main signals of data and clock of keyboard seen code in to the bit mapping of a set of consecutive memory cells of an EPROM. The design utilizes a wireless infra red transmitter-receiver units that are widely implemented in instruments of commercial Video and TV sets.

1- Introduction

Developments of computer systems fascinate system designers to draw computers into a wide range of application fields. One of these important fields denote Computer Aided Engineering. In this field, computers are implemented as main controllers for different functional systems[1,2,3]. Computers through these applications invoke emergently to be provided with suitable peripherals and interfacing circuitries to integrate their performance in these engineering roles.

In this context, various laboratory peripherals were adopted to enhance computer set up in conducting control experimentation's[4,5]. The present work describes a remote PC keyboard peripheral. This peripheral utilizes infra red transmitter receiver wireless link for data transmission due to its successful features of minimizing the filtering process efforts in compressing the wide band electromagnetic noise problem. In addition, very successful schemes of these links appear in most of producer catalogues that caller digital output signals convenient for the interfacing management.

The peripheral scheme is designed around a simple passive addressing structure of a Read Only Memory drive. This results in cost reduction in design when it is compared with similar structure of the peripheral that is based on a processor system to achieve the main

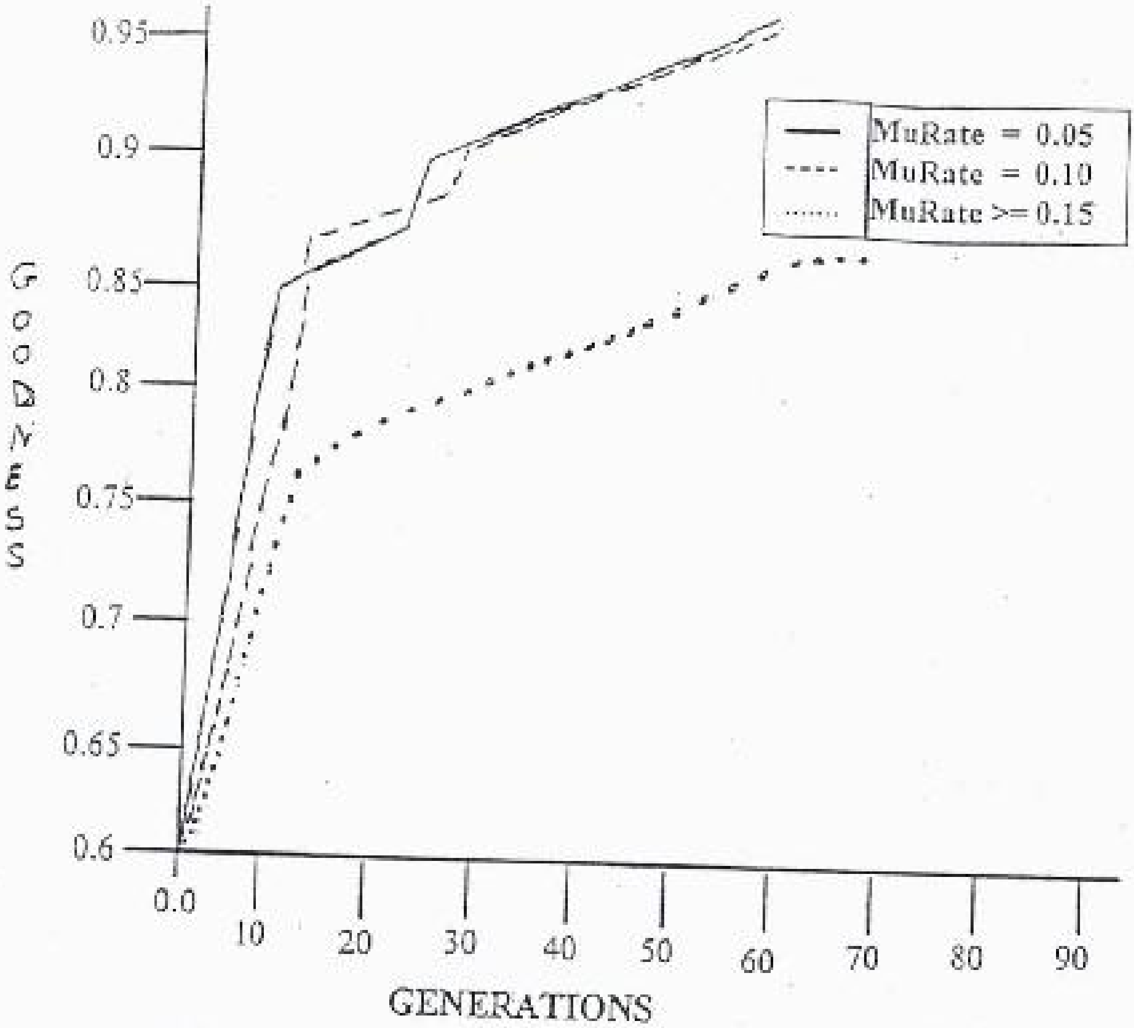


Fig (4): 10 Keys population size with a three mutation rates

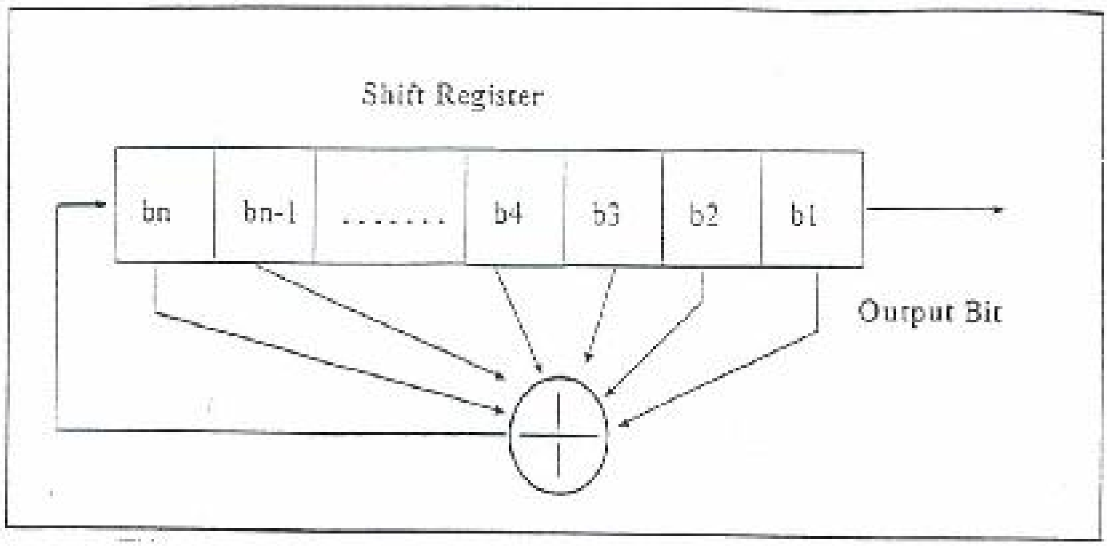


Fig (2): Linear feedback shift register

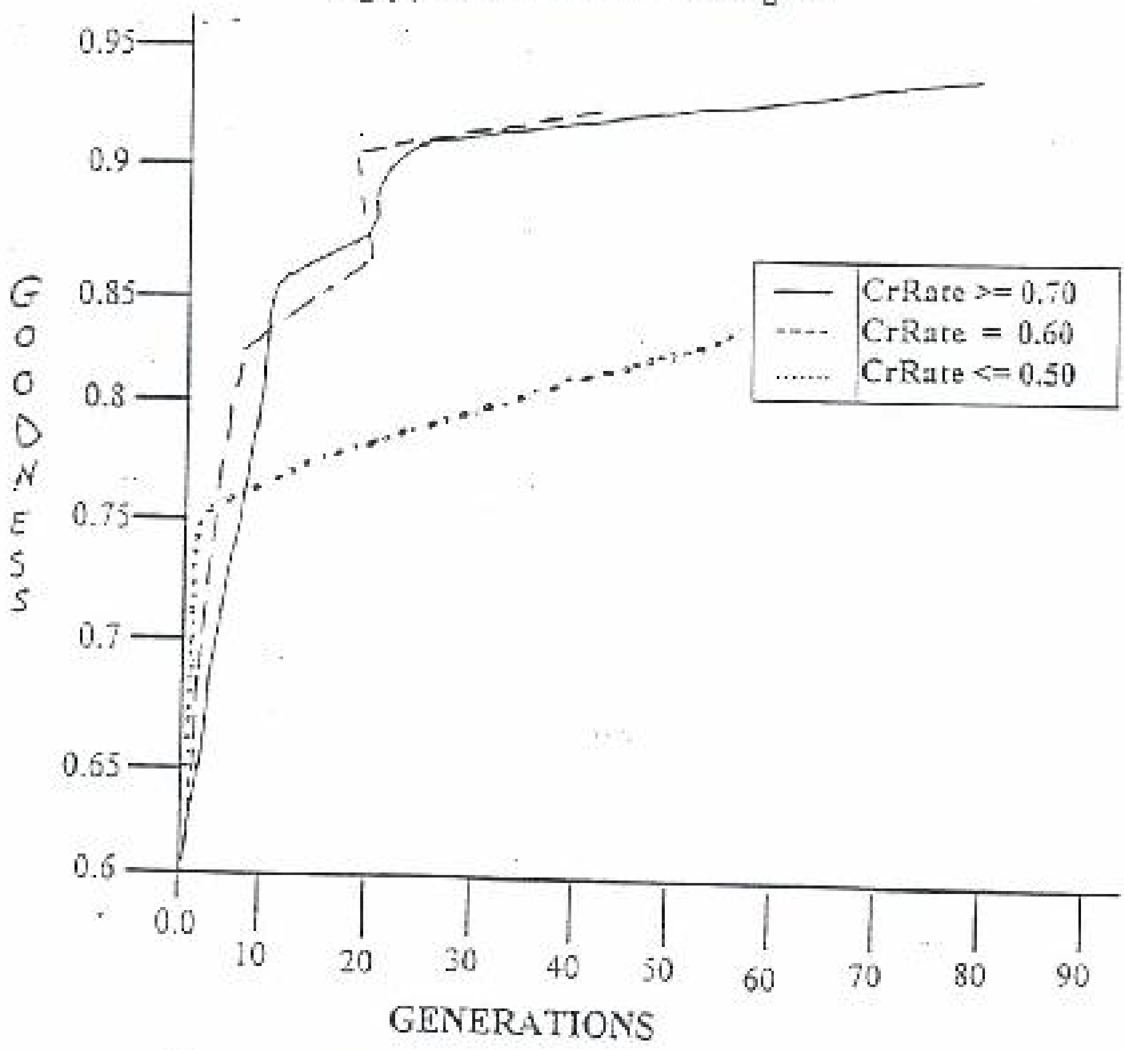


Fig (3): 10 Keys population size with a three crossover rates

Appendix 1

Some Primitive Polynomials Mod 2

1, 1	53, 6, 2, 1, 1	113, 9, 1	151, 67, 1
2, 1, 1	59, 7, 4, 2, 1	113, 15, 1	151, 70, 1
3, 1, 1	59, 6, 5, 4, 3, 1, 1	113, 30, 1	157, 6, 5, 2, 1
5, 2, 1	61, 5, 2, 1, 1	127, 1, 1	163, 7, 6, 3, 1
7, 1, 1	67, 5, 2, 1, 1	127, 7, 1	167, 6, 1
7, 3, 1	71, 5, 3, 1, 1	127, 63, 1	521, 32, 1
11, 2, 1	71, 6, 1	131, 8, 3, 2, 1	521, 48, 1
13, 4, 3, 1, 1	73, 25, 1	137, 21, 1	521, 158, 1
17, 3, 1	73, 4, 3, 2, 1	139, 8, 5, 3, 1	521, 168, 1
17, 5, 1	79, 9, 1	149, 10, 9, 7, 1	607, 105, 1
17, 6, 1	79, 4, 3, 2, 1	151, 3, 1	607, 147, 1
19, 5, 2, 1, 1	83, 7, 4, 2, 1	151, 9, 1	607, 273, 1
23, 5, 1	89, 38, 1	151, 15, 1	1279, 216, 1
29, 2, 1	89, 51, 1	151, 31, 1	1279, 418, 1
31, 3, 1	89, 6, 5, 3, 1	151, 39, 1	2281, 915, 1
31, 6, 1	97, 6, 1	151, 43, 1	2281, 1029, 1
31, 7, 1	101, 7, 6, 1, 1	151, 46, 1	3217, 67, 1
31, 13, 1	103, 9, 1	151, 51, 1	3217, 576, 1
37, 6, 4, 1, 1	107, 9, 7, 4, 1	151, 63, 1	4423, 271, 1
37, 5, 4, 3, 2, 1	109, 5, 4, 2, 1	151, 66, 1	9689, 84, 1
43, 6, 4, 3, 1			
47, 5, 1			

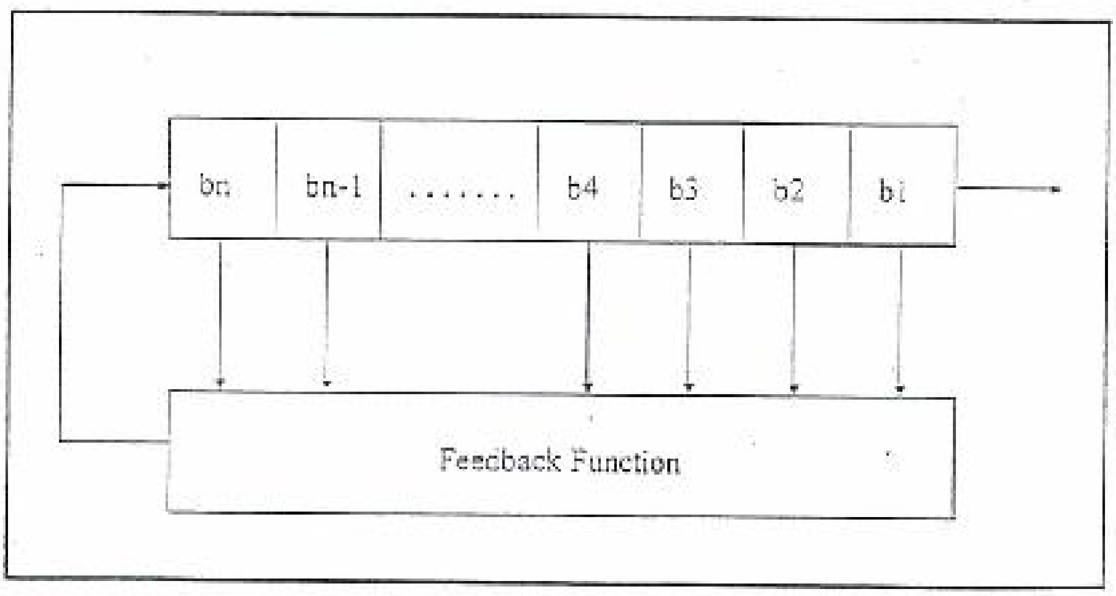


Fig (1): Feedback shift register

6- Conclusion

In this paper, we have argued that genetic algorithms are a valuable tool in the cryptanalysis of certain classes of cipher, and have shown that nonlinear stream ciphers can be broken using such GA. A class of pn-generators consisting of n subgenerators and a combining function f has been investigated. It has been pointed out that a weakness of these generators may be the statistical dependence between a single subgenerator sequence and the keystream.

These are several open avenues for further research. Variation on the crossover and mutation procedures may significantly affect the behaviour of the algorithm. Different fitness function might be used.

References

1- V. S. Pless "Encryption schemes for computer confidentiality" IEE Trans. Comput., Vol. C-25, pp. 1133-1136, Nov. 1977.

2- P. R. Geffee "How to protect data with ciphers that are realy hard to break" Electronic, pp. 99-101 Jan. 4. 1977.

3- J. O. Bruer "On nonlinear combination of linear shift register sequence" in Proc. IEEE Int. Symp. Inform. Theory les arcs, France, Jan. 21-25 1982.

4- Dr. W. A. K. Al-Hamdani & S. A. Al-Ageelee "Correlation attack using genetic algorithm for nonlinear stream cipher systems "Computer Magazine, No. 31 1997, National Computer Center (NCC) Baghdad, Iraq.

5- H. Beker & F. Piper "Cipher systems, the protection of communications" 1982.

6- O. Staffelbach "Correlation attacks on stream cipher" Gretag Aktiegenesellschaft, Althardstr. 70, Ch-8105 Regensdorf, Switzerland.

7- T. Siegenthaler "Decrypting a class of stream cipher using ciphertexts only" 1985.

8- B. Schneier "Applied cryptography" 2nd edition 1997.

9- T. Siegenthaler "Correlation immunity of nonlinear combining functions for cryptography application" 1984.

10- J. J. Grenfenstette "Optimization of control parameters of genetic algorithm" IEEE computer society press 1992.

11- D. E. Goldberg "Genetic algorithm search, optimization and machine learning" 1989.

12- S. A. Abbass "Design a package to build and perform stream cipher systems" 1992.

The initial population P(0) can be chosen heuristically or at random. The operation "evaluate strings in p(t)" refers to the assignment of a figure of merit to each of the population's strings. The strings of population P(t+1) are chosen from P(t) by a randomized selection procedure that ensures that the expected number of times a string chosen is approximately proportional to that strings performance relative to the rest of the population.

The most important recombinational operators we have used in our system are crossover and mutation operators.

Under the crossover operator, two strings in the new population exchange portions of their internal representation. For example, if the string are represented as binary strings, crossover can be implemented by choosing a point at random, called the crossover point and exchanging the segments to the right of this point[10].

Mutation operator is a secondary search operator which increase the variability of the population. After selection, each bit position of each string in the new population undergoes a random change with a probability equal to mutation rate. For a problem over a binary alphabet, the original allele is exchanged for its complement[11].

5- Results

The simulation of the algorithm was programmed in Pascal. It was applied to ciphertexts created using a nonlinear stream cipher systems whose combining functions are:

1- Multiplications (And)[5].
2- Or[5].
3- J-K Flip flop[5].
4- Pless system[1].
5- Geffe system[2].
6- Police[12].
7- Multiplexing[5].
8- Bruer system[3].

The system proved highly successful in finding the primitive feedback polynomials (PFPs) and the initial state which are used by the above nonlinear stream cipher systems. Variant shift registers lengths and primitive feedback polynomials were used.

Increasing the size of the population can reduce the number of generations required to find the correct setting, but whether you search with 10 strings over 100 generations or 20 strings over 50 generations, they both involve looking at about 100 strings.

The effect of the crossover rate and the mutation rate were explored. Figures 3 and 4 show the result of three attacks to the same nonlinear stream cipher. Different populations size were taken e.g. 4, 6, 8, 10, 12, 14, 16, 18 strings (keys). Three different crossover rates (CrRate > = 0.70, CrRate = 0.60, and CrRate < = 50) and mutation rates (MuRate = 0.05, MuRate = 0.10, and MuRate > = 0.15) were selected.

The highest crossover rate the more quickly new strings are introduced into the population. On the other hand, low crossover rates, the search may stagnate due to the lower exploration.

The highest mutation rates clearly hampered the search effort, on the other hand, low mutation rates seemed to be better results.

Correlation attacks and variations such as fast correlation attacks have been successfully applied to a number of LFSR-based key stream generators.

There are other general attacks against key stream generators. The linear consistency test attempts to identify some subset of the encryption key using matrix techniques.

There is also the meet in the middle consistency attack. The linear syndrome algorithm which relies on being able to write a fragment of the output sequence as linear equation. There is the best affine approximation attack and the derived sequence attack[8].

4- Application of the New Algorithm to Key Search

The purpose of the new genetic algorithm is to find the driving part subsystem i.e., the primitive feedback polynomials (PFPs) and the initial states which are used by the attacked generators.

4-1 The Complete Algorithm

In the following algorithm a primitive feedback polynomials file which contains all primitive feedback polynomials (PFPs) for all shift registers length should be available.

We have used the division method to find the contents of the feedback file (for each shift registers length there are a finite number of primitive feedback polynomials). The steps of the algorithm are:

1- A primitive feedback polynomial (no.t) which corresponds to the length of the attacked shift register is assigned.

2- A random population of strings (keys) is generated.

3- A fitness value for each string in the population is determined.

4- A biased random selection of parents is conducted.

5- The crossover operation is applied.

6- The mutation process is applied to the children.

7- A fitness value for each string in the new generation is determined.

This process will stop after a fixed number of generation or a specified value will met. If the fixed number of generation (MaxGen) is met and the specified value (Threshold) is not, the process returns to step number one of the algorithm and a new primitive feedback polynomial (no. T + 1) which corresponds to the length of the attacked shift register from the feedback file is assigned.

This algorithm is repeated until the EOF feedback file or solution is found. The above algorithm can be written as:

```
Max Value: = 0;
While Not Eof Feedback File And Max Value
< Threshold Do
Begin
t: = 0
Search Feedback File For Shift Register's length
And Choose Primitive feedback Polynomial No.t;
Initialize P(t);
Evaluate Strings In P(t);
While No. Of Gen. <Max Gen Or Max Value
< Threshold Do
Begin
t: = t + 1;
Select P(t) From P(t-1);
Recombine String In P(t);
Evaluate Strings In P(t);
End;
End;
```

n	&(n)	n	&(n)
1	1	13	630
2	1	14	756
3	2	15	1800
4	2	16	2048
5	6	17	7710
6	6	18	8064
7	18	19	27594
8	16	20	24000
9	48	21	84672
10	60	22	120032
11	176	23	356960
12	144	24	276480

Table (1): The number of primitive polynomials with degree at most 24.

In general, there is no easy way to generate primitive polynomial mod 2 for a given degree. The easiest way is to choose a random polynomial and test it whether it is primitive. Appendix 1 lists some primitive polynomials mod 2 of varying prime degree[8].

For example the listing (97, 6, 1) means that the following polynomial is primitive modulo 2.

X ^ 97 + X ^ 6 + 1

The first number is the length of LFSR. All the numbers specify the tap sequence. This listing (97, 6, 1) means that if you take a 97-bit shift register and generate the new bit by XORing the ninety seventh and sixth bits together the resultant LFSR will be maximal length; it will cycle through 2 ^ 97-1 values before repeating.

Primitive trinomials are fastest in software, because only two bits of the shift register have to be XORed generate each new bit.

Actually, all the feedback polynomials list in appendix 1 are sparse, meaning the they only have a few coefficients.

Sparseness is always a source of weakness, sometimes enough to break the algorithm. It is far better to use dense primitive polynomials, those with a lot of coefficients[8].

Generating dense primitive polynomials modulo 2 is not easy. In general, to generate primitive polynomials of degree K you need to know the factorization of 2 ^ k -1.

3- Some Methods for Analyzing Stream Cipher

Analyzing stream cipher is often easier than analyzing block cipher. One important metric used to analyze LFSR-based generators is linear complexity.

This is defined as the length, n, of the shortest LFSR that can mimic the generator output. Simple algorithm called Berlekamp-Massy algorithm, can generate this LFSRs after examining only 2n bits of the key stream[8].

Cryptographers try to get high linear complexity by combining the output of several output sequences in some nonlinear manner[7].

The danger here is that one or more of the initial output sequences can be corrected with the combined key stream and attacked using linear algebra.

Thomas Siegethaler has shown that correlation immunity can be precisely defined, and that there is a tradeoff between correlation immunity and linear complexity[9].

50) may only be possible if the correct LFSR-phase could be found faster than by an exhaustive search.

Al-Hamdani and Al-Ageelee have applied a new approach to cryptoanalysis based on the application of a directed random search algorithm called a genetic algorithm[4]. They used genetic technique to reduce the number of trials which are needed to find the correct initial setting of the attacked generators assuming that the feedback polynomial is known.

This paper presents a complete genetic algorithm to find the primitive feedback polynomials (PEPs) and the initial setting of a nonlinear stream cipher systems.

2- Linear Feedback Shift Register

Shift register sequences are used in both cryptography and coding theory. Stream ciphers based on shift registers have been the workhorse of military cryptography since the beginnings of electronics.

A feedback shift register is made up of two parts: a shift register and a feedback function see Fig. (1).

The shift register is a sequence of bits. Each time a bit is needed, all of the bits in the shift register are shifted one bit to the right. The new left-most bit is computed as a function of the other bits in the register. The output of the shift register is one bit, often the least significant bit[8].

The period of a shift register is the length of the output sequence before it starts repeating. The simplest kind of feedback shift register is a linear feedback shift register, or LFSRs see Fig (2).

The feedback function is simply the XOR or certain bits in the register; the list of these bits is a tap sequence. Cryptographers like to analyze sequence to convice themselves that they are random enough to be secure. LFSRs are the most common type of shift registers used in cryptography.

An n-bit LFSR can be in one of (2^n)-1 internal states. This means that it can, in theory, generate (2^n)-1 bit-long pseudo-random sequence before repeating[8].

Only LFSRs with certain tap sequences with cycle through all (2^n)-1 internal states; these are the maximal-period LFSRs.

In order for a particular LFSR to be maximal-period LFSR, the polynomial formed from a tap sequence plus the constant 1 must be a primitive polynomial mod 2. The degree of the polynomial is the length of the shift register. A primitive polynomial of degree n is an irreducible polynomial that divides (X ^ (2 ^ n)-1) +1.

To determine the number of primitive polynomial of a given degree we need to introduce the Euler function[5]. For any positive integer m the Euler function E(m) is the number of positive integer which are less than or equal to m that coprime to it.

For any given positive integer n the number of primitive polynomial of degree n over GF (2), which we denote by & (n), is given by the equation:

$$\&(n) = \frac{E((2^\wedge n) - 1)}{n}$$

Table (1) gives the number of primitive polynomials over GF (2) of degree n for 1 < = n < = 24[5].

Use of Genetic Algorithm (GAs) in The Cryptanalysis Nonlinear Stream Cipher (NLSC)

Dr. W. A. K. Al-Hamdani & S. A. Al-Ageelee
Department of computer Science, University of Technology,
Baghdad-Iraq

Abstract

Pseudonoise sequences generated by linear feedback shift registers (LFSRs) with some nonlinear combining function have been proposed as running key generators in stream ciphers[1,2,3].

We consider the use of genetic algorithm (GAs) as powerful tools in the breaking of cryptographic systems.

We have shown in our previous paper[4] that GAs can greatly facilitate cryptanalysis by efficiently searching large key spaces and demonstrated their use with nonlinear stream cipher systems.

This paper presented a complete genetic algorithm which can be used to reduce the number of trials which are needed to determine the primitive feedback polynomials and the initial states of nonlinear stream ciphers.

A well known systems are taken for the case of study: Pless system[1], Geffe system[2], Bruer system[3], J-K, OR, Multiplying, Multiplexing[5] and police systems[12].

Index term

Linear feedback Shift Registers (LFSRs), Primitive Feedback Polynomials (PFPs), Euler Function, Sparse Feedback, Dense Feedback, Linear Complexity, Tap Sequences.

1- Introduction

In the analysis of certain stream ciphers it is convenient to divide a running key generator into a driving part and a combining part. The driving subsystem is responsible for providing sequences with large periods and good statistical properties. It is often implemented as a set of linear feedback shift registers (LFSRs) whose output sequences are then fed into the (nonlinear) combining subsystem in order to produce the key stream[6].

For certain generators of this type, e.g., for the generators of Geffe, Bruer, or pless, there is a statistical dependence between the generator output and output of some internal shift registers.

The cryptanalytic significance of this fact was first recognized by Blaser and Heinzmann and was investigated by Siegenthaler[7]. In siegenthaler's analysis the generator output sequence Z is viewed as a perturbation of the appropriate internal LFSR-sequence X by a symmetric memoryless noise source with prob (0) = P.

Then if P (correlation-probability) # 0.5 the unknown sequence X can be found by correlating all candidates for X with the given sequence Z; a candidate is accepted if its correlation to Z exceeds a suitable threshold.

Such correlation attack can significantly reduce a brute force attack since the LFSRs can be attacked individually (divide and conquer). Attack on long LFSR (length > =

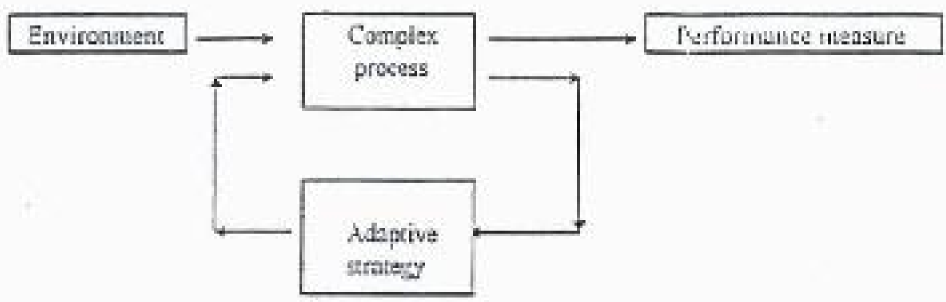


Fig.(1): Adaptive system model.

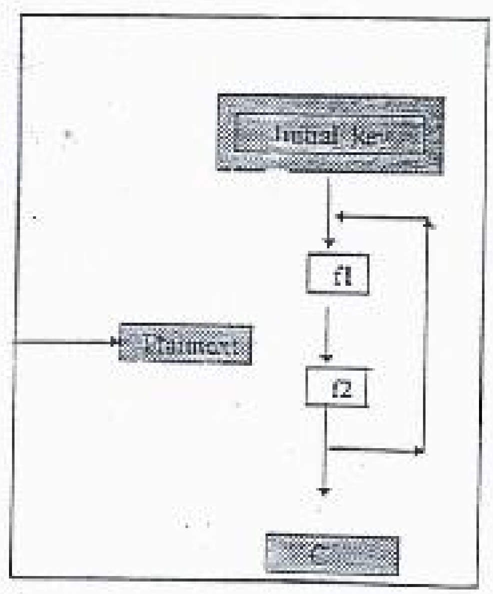


Fig.(2): Ciphertext feedback

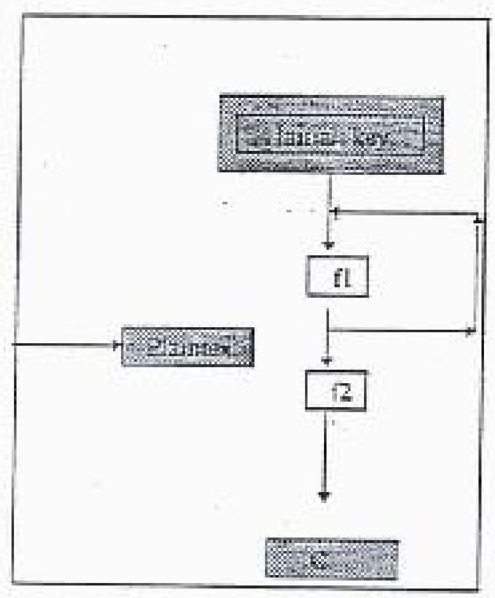


Fig.(3): Key feedback

Example (2)

S = 1

Generation number	Fittest program	Fit	Decrypted text
0	- * r a_8 0 a_0 8 + ^ 17 a_0 w a_8 0	0.3250	solejakwfycfootwvahgiesuzgeec xoeofeymmsuucjadgwfgkiodetg geewrcaivedagiggetvwve
1	- * r a_8 0 a_0 8 + ^ 17 a_0 w a_8 0	0.3250	sofejakwfycfootwvahgiesuzgeec xoeofeymmsuucjadgwfgkiodeg geewrcaivedagiggetvwve
2	+ 18 a_8	0.3824	thenegotiationsforsettlementoet hestrikeatanimpasseronmend weincreaseouroffe

Example (3)

S = 2

Generation number	Fittest program	Fit	Decrypted text
0	* * 14 a_0 * $a_1 a_1$ - r 41 w a_1 1	0.3547	honedrlarytoibmcaxizpdrdajqhrf otszhrznrehllghntefnjapotlftrdny ttrvdfelsnuzhrlfvjmbepkle
1	* * 14 a_0 * $a_1 a_1$ - r 41 w a_1 1	0.3547	honedrlarytitibmcsxizpdrdajqlrf otszhrznrehllghatefnjapotlftrdny ttrvdfelsnuzhrlfvjmbepkle
2	* * 14 a_0 * $a_1 a_1$ - r 41 w a_1 1	0.3547	honedrlarytitibmzsxizpdrdajqhrf otszhrtznrehllghntefnjapotlftrdny ttrvdfelsnuzhrlfvjmbepkle
3	a_0 - r 41 w a_1 1	0.3688	fojelrqabygisixmnsgitpircaqqor bohsmhetmihetlcghnketnnawool sthdiyrthrwdrcwsjumhulvwmrc qkie
4	a_0 - r 41 w a_1 1	0.3688	fojelrqabygisixmnsgitpircaqqor bohsmhetmihetlcghnketnnawool sthdiyrthrwdrcwsjumhulvwmrc qkie
5	a_0 - r 41 w a_1 1	0.3688	fojelrqabygisixmnsgitpircaqqor bohsmhetmihetlcghnketnnawool sthdiyrthrwdrcwsjumhulvwmrc qkie
6	- r 3 0 w a_0 0 - r 41 w a_1 1	0.3915	codebreakingisthemostimportan tformofsecretintelligenceinthewo rldtodayitproducesmuchmoreau dmae

6- Conclusion

This paper has presented a new class of attacking cipher systems called adaptive ciphertext-only attack, which determines the cipher systems for a given ciphertext. It uses GP methodology, and makes use of language statistical characteristics as a measurement for program fitness. The proposed method uses the technique of indexed memory in order to break the systems having a feedback. A number of examples have been presented which show the success of this method in breaking CSS, further work will examine more complex systems.

References

1- Andy Singleton, "Genetic programming with C++), Byte, Feb. 1994, pp. 171-176.

2- Beker Henry and piper Fred, "Cipher system", Northwood, 1982.

3- Brady J. M., "The theory of computer science", Chapman and Hall, 1977.

4- DE Jong K., "Adaptive system design: A genetic approach", IEEE Trans. On Sys., Man, and Cyp., Vol. SMC-10, No.9, Sept. 1980, pp. 566-574.

5- Denning D. E. R., "Cryptography and data security", Addison-Wesley, 1982.

6- Goldberg David E., "Genetic algorithms in search, optimization, and machine learning", Addison-Wesley, 1989.

7- John Holland, "Adaptation in natural and artificial systems", Univ. Of Michigan press, 1975.

8- Koza John R., "Genetic programming: On the programming of computer by means of natural selection", MIT press, 1992.

9- Mathews Robert A. J., "The use of genetic algorithms in cryptanalysis", Cryptologia, April 1993, pp. 187-201.

10- Schneir B., "Applied cryptography", John Wiley & Sons, 1996.

11- Spillman Richard and others, "Use of a genetic algorithm in the cryptanalysis of simple substitution ciphers", Cryptologia, Jan. 1993, pp. 31-44.

12- Teller A., "Turing completeness in the language of genetic programming with indexed memory", IEEE WCCI 1994, pp- 136-141.

notation, and it is evaluated recursively. For example + * 25 a_0 represents the expression $(2*5)+a_0$, and the interpretation of it is:

For i = 1 to size (chiphertext) do

$p_i = ((2 * 5) + c_i) \bmod 26$

where p_i is the ith plaintext letter and c_i is the ith ciphertext letter. When the decryption algorithm contain a feedback such as key feedback, this could be expressed as in the following example:

+ a_0 w + r 0 1 1 1

and the interpretation of it is:

$m_1 = 0$; $m_1 = m_1 + 1$

for i = 1 to size (ciphertext) do

begin

$p_i = (m_1 + c_i) \bmod 26$,

$m_1 = m_1 + 1$

end

where m_1 is the value of location 1 in the memory, in which the key is stored, and updated.

In each generation fitness values are assigned to programs. This value is a measurement of the text satisfiability of the desired language characteristics, and it is used to control the application of the operations that modify the structure in our population. The fitness value is calculated by using each program to decrypt the given ciphertext, and then compute the following functions:

$$\text{dif}_1 = \sum_{i=1}^{26} |sf_i - ff_i|$$

$$\text{dif}_2 = \sum_{i=1}^{26^2} |dsf_i - dff_i|$$

$$\text{dif}_3 = \sum_{i=1}^{Q} tf_i$$

$$\text{dif} = \text{dif}_1 + \text{dif}_2 + \text{dif}_3$$

$$\text{fit} = \frac{1}{1+\text{dif}}$$

where

sf_i : Standard relative letter frequency.

ff_i : Measured relative letter frequency.

dsf_i : Standard relative bi-gram frequency.

dff_i : Measured relative bi-gram frequency.

tf_i : Measured relative tri-gram frequency which never appear in English.

The initial population contains a number of programs generated randomly, such that, the generated programs are of different sizes and shape. So, we use the way described in[8] to generate the initial population. If the block size S is known, all the programs are generated for the block size, otherwise, the programs are generated for different block size which are determined randomly.

5- Examples

This section presents a number of examples which shows the success of the adoptive method to determine the cipher system for a given ciphertext.

Example (1)

S = 1

Generation number	Fittest program	Fit	Decrypted text
0	- - a_0 a_0 + a_0 1	0.3348	qpcbldgmmiuardjbwterbgmhp vtkfrojfiuikjrakfrxcidasepdoihin quldcgbjeldgvihmu
1	* - a_0 17 + 24 r 13	0.3378	stghcfcwtrvaoirfzhmperhcwbtm pydruzhdiyarlydrlagafiqetfuobav soxfgchzefbakabrwo
2	* - a_0 17 + 24 r 13	0.3378	stghcfowtrvaoirfzhmperhcwbtm pydruzhdiyarlydrlagafiqetfuobav soxfgchzefiaskabrwo
3	- r 14 w a_0 3	0.3788	iamaphdstudentinthecomputersc iencedepartmentoftechnologyand ihaveready curpep

a) Evaluate the fitness of each individual.
b) Create a new population by applying the following operations:
   i) Copy existing individual to the new population.
   ii) Create two new individuals (chromosomes) by genetically recombining randomly chosen sub strings from two existing strings.

3- The best individual that appeared in any generation is designated as the result of the GA for the run.

GP is a new programming methodology, the goal of it is to get computer to solve problems without being explicity programmed, thus the space of computer programs is the place to look[8]. Gives a good illustration about these applications. In this paper, we shall show a new application of GP which is cryptanalysis.

GP is used for evolving functions that perform well on assigned task. These evolved functions are represented in GP as S-expressions consisting of non-terminals (atomic functions) and terminals (variables and constants).

There are many problems that traditional GP cannot solve, due to the theoretical limitations of its paradigm. Thus, a new technique has been added which is the indexed memory, and it has been proved that GP with the technique of indexed memory is Turing complete[12]. This means that GP with indexed memory can be used evolve any algorithm. So, Read (to get a value from the memory) and Write (to store a value into the memory) are added as a new non-terminals, and each GP function is given access to its own array, indexed over integer numbers.

4- The New Attacking Method

Our method uses GP to determine a cipher system for a given ciphertext. The major steps in preparing to use GP to solve a problem involve:

1- Determining the function and terminal sets, F and T.

2- Determining the representation scheme, and

3- Determining the fitness measure.

The set of terminals includes all possible values of the keys, i.e., 1..26, and the variables which correspond to input letters (ciphertext letters), thus

$T = \{1, ..., 26, a_0, ..., a_{S-1}\}$

where S is the block size, for example, S = 1 for direct standard cipher system, and the value of S could be 2 for Hill system.

As shown, the basic functions are addition (+) and multiplication (*), so F = {+, *, r, w, ^} where (^) is the power function, r and w are Read and Write functions respectively. R and w functions are necessary in the case where the feedback is used. The general form of the Read function is r f_1 f_2 where the value of f_2 is a location in the indexed memory, and the value of f_1 is stored initially in that location. The value returned from this function is the value of location f_2. Also, the general form of Write function is w f_1 f_2 where the value of f_1 is stored in the memory at location determined by f_2, and the returned value is the value of f_1.

The chromosomes in GP are programs, each program is a string of characters which is represented using prefix polish

Ciphertext-Only Attack, in which only ciphertext is known, and the interceptor can find decryption algorithm, key, and hence plaintext.

The need for an adaptive solution to a problem arises in a wide variety of contexts. Typically, the inherent complexity of a problem or the uncertainly surrounding it prevents one from specifying an acceptable priori solution. Instead, an attempt is made to solve the problem adaptively as shown in Fig. (1)[4]. Finding a cipher system for any given ciphertext is a complex process, and there is no acceptable solution to solve this problem. So, we shall show how the adaptive method can solve the problem of cipher system determination for any given ciphertext.

In the new attacking method, the structure under adaptation is a set of programs, and the adaptive strategy used is GA. Our method can be used to break any cipher system, but here only conventional substitution cipher systems (CSS) are considered as an example.

2- Cipher System Structure

Any computable function f is regarded as constructed object, that is, it has to be built from some components say $F_1, \ldots, f_m$ which are also computable functions. Of course, it is normally the case that the components f_i themselves have to be built. But, these must be some functions that are not decomposable which are called basic functions, the set of basic functions is finite[3].

The encryption and decryption algorithms are computable functions, composed of a number of basic functions. For CSS, the set of basic functions is {+, -, /, *}. Since:

$a - b = (a + (-b)) \bmod 26$

$a / b = (a * (b^{-1})) \bmod 26$

This set could be reduced to {+, *}, where $-b$ and b^{-1} are additive and multiplicative inverse elements. The basic functions are combined according to the composition strategy such that:

If f, g, and x are computable functions, then a new computable function h can be constructed from these functions using composition strategy as follows:

$h = f(g(x))$

Also, cipher systems may contain a feedback which can be one of the following two types as shown in Fig.(2) and (3):

1- Ciphertext feedback.

2- Key feedback.

3- Genetic Programming With Index Memory

Genetic programming (GP) is an application of GA which is a simple tactic for computer learning that is inspired by natural evolution[1]. GAs were first suggested by John Holland in the early seventies[7]. Over the last 20 years it have been used to solve a wide range of search, optimization, and machine learning problems[6]. The steps of a simple GA can be summarized as follows:

1- Randomly create an initial population of individuals.
2- Iterative performed the following sub steps on the population until the termination criterion has been satisfied:

"A method
d public key
No.2, Vol.21

otograph an
, Advance in
ence series,

Adaptive Ciphertext-Only Attack Using Genetic Programming With Indexed Memory

Dr. W. A. K. Al-Hamdani, Dr. A. F. Abdul Kader, W. S. Awad
Department of Computer Science, University of Technology
Baghdad, Iraq

Abstract

There are a number of methods and tools to attack different cipher systems. A general solution for the problem of determining a cipher system for any given ciphertext is not known, So, in this paper, an adaptive method is presented to solve this problem. The proposed method uses genetic programming with indexed memory, where the structure under adaptation is a set of programs which presented decryption algorithms.

Key words: *Genetic algorithm, Genetic programming, Indexed memory, Cipher system, Cryptanalysis, Ciphertext-only attack.*

1- Introduction

Cipher systems are systems which are used for encrypting plaintext to produce ciphertexts, and vice versa. The purpose of such systems is the protection of information from the unauthorized persons. The set of steps which are taken by an encipherer are called encryption algorithm which depends on a key, and the reverse algorithm is called decryption algorithm which uses the same key or a new key derived from the previous key.

There are a number of attacking methods to break cipher systems which can be classified into ciphertext-only attack, known-plaintext attack, chosen-plaintext attack, chosen-ciphertext attack, chosen-key attack, and adaptive chosen-plaintext attack[10]. In ciphertext-only attack, an interceptor known ciphertext only, and any part of plaintext is not known. One of the methods used for attacking knowing ciphertext only is by searching the key space. It is clear that the cipher systems of small key space can be easily broken using the method, also the efficiency of the search process can be increased using genetic algorithm (GA)[9,11]. Another method used in ciphertext-only attack is the letters frequencies analysis which succeeded in breaking a number of conventional cipher systems.

In this paper, a new class of attacking is presented which is called **Adaptive** simplified database for strong the factorisation of N, by extracting the values of p and q, it is possible to compute the secret key value (D).

References

1- Rivest, Shamir, A, and Adleman, L "A method for obtaining digital signatures and public key correct value of the private key. No.2, Vol.21 February 1978.

2- Seberry J and Pieprzyk J, "Cryptograph an introduction to commuter security", Advance in computer advance in computer science series, 1989.

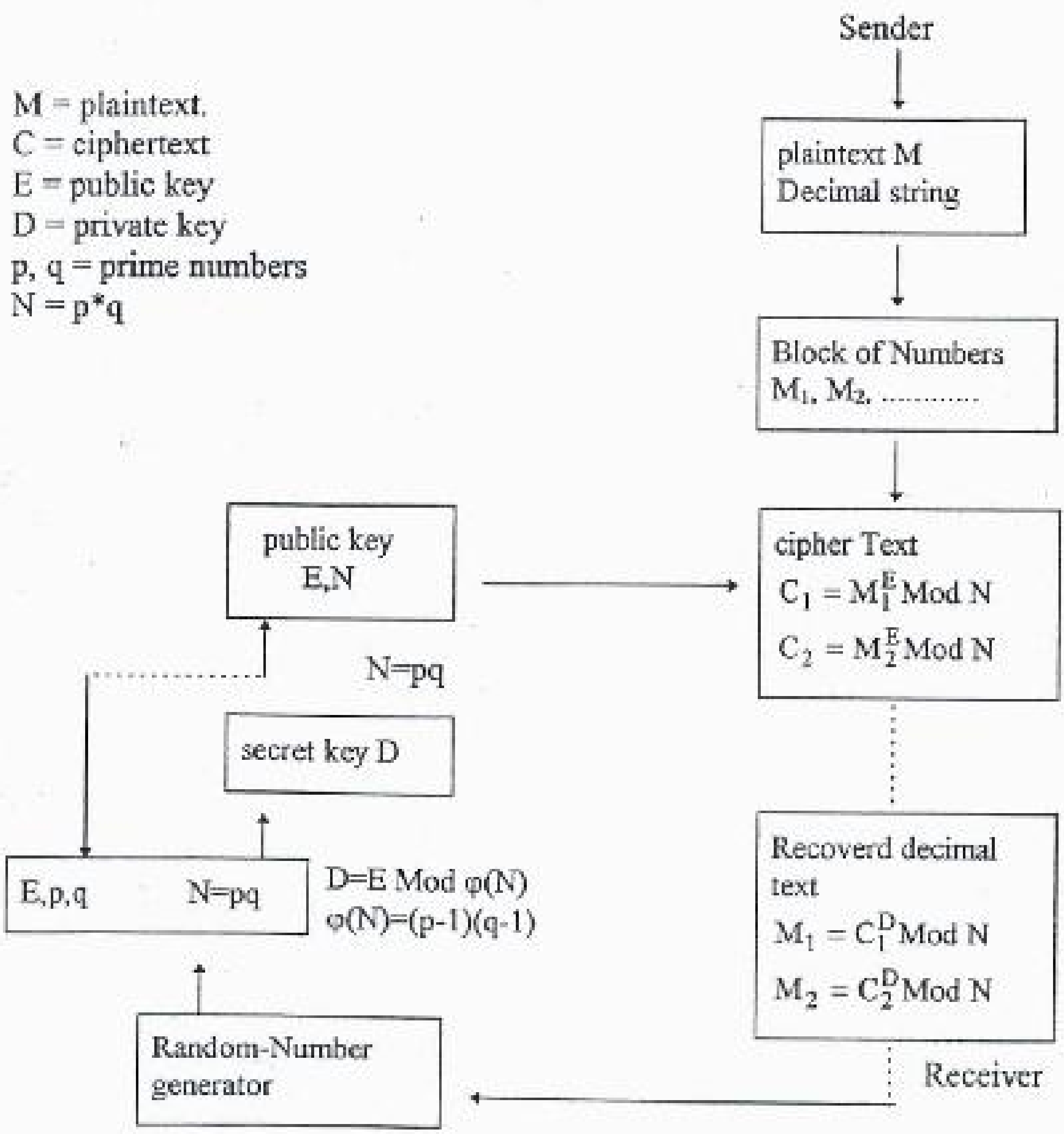


Figure (1): The concept of working of RSA system in details.

11	33	3	55	5	77	7	121	11	143	13

←→

Value of N = 143
Value of q = 13
and value of p = 11

All the (q) values will be tested for the primarality before it stored into the data base.

This design will make it possible to store all the possible values in the data base without the need for a very large database.

F- It is possible to avoid the creation of the database if the value of n is small and this could be done by using the following algorithm.

Enter odd Number say N
Read N
Given set of prime numbers
Divide n by each one of this element
Until no fraction is remain.
Check p, q if it is prime them stop else regect.
Until one of the set element become
> = N then either p = 1 & q = N or reject N.

5- Implementation

If we have the public key E = 5, N=51 so the public key (E, N)=(5, 51).

To beak the RSA cipher we should find the private key (D, N). We will follow the following steps:

A- We calculate the square root of 51 which is equal to nearly 7.

B- We create the dynamic database starting from the first prime number which is less than 7 and its value equal to one digit with excluding the even value of N and the non prime value of q:

3	9	3	15	5	21	7	33	11	39	13	51	17	57	19

↑

C- We access the data base to find the factorial of N (where N = 51) and in this case q= 17 and p = 3.

D- We find the value of $\phi(N)$.

$$\phi(N) = (p-1) * (q-1)$$
$$= (3-1) * (17-1)$$
$$= 32$$

E- We find the value of the private key D from the following equation:

$$D = \frac{GCD\ (\phi(N) * \phi(N) + 1}{E}$$

$$= \frac{GCD\ ((p-1),\ (q-1)) * \phi(N) + 1}{E}$$

To find GCD ((p-1), (q-1)):
GCD (2, 16) =2

$$\therefore D = \frac{(2 * 32) + 1}{5} = \frac{65}{5} = 13$$

The private key is = (13, 51).

F- To make sure that the value of d is correct, it must satisfy the following equation:

E * D (Mod $\phi(N)$) = 1
5 * 13 (Mod 32)
65 mod 32 = 1

We succeed 1 = 1 in finding the correct value of the private key.

6- Discussion

With a public key encryption system, each user would have a key that did not to be kept secret. The public nature of the key would not inhibit the secrecy of the system. The public key transformation is essentially a one-way encryption with a secret (private) way to decrypt.

This paper introduces an idea for breaking RSA cipher system by using a

$$D = \frac{GCD(\phi(N) * \phi(N) + 1)}{E}$$

$$= \frac{GCD((p-1), (q-1)) * \phi(N) + 1}{E}$$

To make sure that the value of D is correct, we can satisfy the following equation:

$ED \ (Mod\ \phi(N)) = 1$

As we discussed above the public key is (E,N) and its value known to the public. The problem is how to find p, q values. We know that:

$N = p * q$

So theoretically it is possible to build a database containing all the values of p and q, in practice, it is very difficult to store N values specially when the value of p or q is consist of 200 digits.

For this reason, we suggest our new technique to create a dynamic database with minimum size but it meets our requirement in finding the values of p and q by using the value of N.

To create the dynamic database we must follow the following steps:

A- $N = p * q$ or $N = q * p$

For that we choose either p * q or q * p to store it in the database and in this case we cut the size of the database to the half.

EX: p = 5 and q = 7

N = 35 we choose only 5 * 7 and not 5 * 7 and 7 * 5.

B- We select the odd numbers only for the values of p and q.

C- We apply the primarility test for the selected p and q values.

D- Create a relational database so the design of the relation as follows:

PKEY #	NVAL #	qVAL

occurs several times

Where:

PKEY # = Value of p and at the same time equals to the value of N. It is a key.

NVAL # = It refers to the value of and it represents a sub key.

qVAL # = Value of the q.

To explain this design let us follow the following example:

3	6	2	q	3	15	5

PKY # NVAL # QVAL

because N = p * q

PKEY = N = 3 * 1 = 3

PKEY = N = p at the same time.

The first occur which is 6 and 2

NVAL = qVAL * PKEY

6 = 2 * 3

It means N = 6, q = 2, p = 3

E- The creation of the dynamic relational database will be at the same time of the analysis and that means it is not necessary to prepare the database permanently. The creation of the database will be for the possible numbers combinations that lead to the value to p & q.

For example if the value of N is equal to 143. By taking the square root of 143 which is equal to nearly 12. This means that the value of p or q is consists from two digits.

We take the first prime number which is equal to two digits. So the creation of the data base is starts from two digits as the following.

3-2 Computing ϕ (n) Without Factoring N

If a cryptanalyst could compute ϕ (n) than he could break the system by computing D as the multiplicative inverse of E moulo ϕ (n).

This approach is no easier than factoring N since it enables the cryptanalyst to easily factor N using ϕ (N). This approach to factoring N has not turned out to be practical.

How can N be factored using $\phi(n)$? First, (p + q) is obtained from N and $\phi(n)$ = N - (p + q) +1. Then (p - q) is the square root of (p + q) -4N. Finally, q is half the difference of (p + q) and (p - q).

3-3 Determining D Without Factoring N or Computing ϕ (N)

Of course, D should chosen from a large enough set so that a direct search for it is unfeasible.

Computing D is no easier for a cryptanalyst than factoring N, Since once D is known N, could be factored easily. This approach to factoring has also not turned out to be fruitful.

A knowledge of D enables N to be factored as follows. Once a cryptanalyst knows D he can calculate E * D-1, which is multiple of ϕ (N). Therefore, if N is large a cryptanalyst should not be able to determined D any easier than he can factor N.

3-4 Computing D in Some Other Way

Although this problem of "computing the roots modulo N without factoring N" is not a well known difficult problem like factoring. We feel that it is computationally intractable.

4- <u>The New Approach</u>

With the RSA algorithm, there are two keys, D and E that work in pairs for decryption and encryption respectively.

A plaintext message M is encrypted to be a ciphertext C by:

$C = M^E \bmod N$

The plaintext is recovered by:

$M = C^D \bmod N$

The encryption key consists of integers (E, N), and the decryption key is (D, N). The starting point in finding keys for this algorithm is to select a value for N.

The value of N should be quite large, a product of two primers p and q. Next a relatively prime to (p-1)*(q-1) means that E has no factors in common with (p-1)*(q-1).

Finally it is possible to find D such that:

E*D mod (p-1)*(q-1) = 1

The cryptanalysis of the RSA encryption is not difficult but it required a large amount calculations. We want to find ϕ(N) (the number of positive integers less than N that are relatively prime E to ϕ(N) (Euler totient function). ϕ(N) can be found by using the primes number tables that are initially proved for this task. Then we can find the secret key by use method:

let Z = 1

Repeat

D = (Z * ϕ(N) +1)/E

Z = Z+1

Until D is integer number

End

It is possible to calculate D from the following equation:

day. The ingenious approach is certainly feasible. Some of the algorithms are based on known "hard problems". But, the cryptanalyst dose not necessarily have to solve the underlying problem to break the encryption of a single message.

Second, estimates of breakability are based on current technology. An enormous advance in the technology of computers has occurred within the last fifty years. Things that were infeasible in 1940s become possible in the mid 1950s, and every succeeding decade has brought greater improvement. Operating characteristics of computers, such as numbers of operations per second and numbers of bits stored, have regularly increased by an order of magnitude every few years. It is risky to pronounce an algorithm secure because it cannot be broken with current technology.

The analyst can do any or all of three different things:

1- Attempt to break a single message.

2- Attempt to recognize patterns in encrypted message in order to be able to break subsequent ones by applying a straight forward decryption algorithm.

3- Attempt to find general weaknesses in an encryption algorithm, without necessary having intercepted any message.

2- Public Key System (RSA)

Figure (1) illustrates the concept of working of RSA system in details.

How should you choose your encryption keys, if you want to use RSA method?

You first compute N as the product of two primes p and q;

$N = p * q$

These primes are very large, "random" primes. Although you will make N public, the factors p and q will be effectively hidden from everyone else due to the enormous difficulty of factoring N. This also hides the way D can be derived from E.

You then pick the integer D to be a large, random integer which is relatively prime to (p-1) * (q-1). That is, check that D satisfies:

$GCD(D,(p-1) * (q-1)) = 1$

"GCD" means "greatest common divisor".

The integer E is finally computed from p, q, and D to be the "multiplicative inverse" of D, Modulo (p-1) * (q-1).

Thus we have:

$E * D = 1 \pmod{(p-1) * (q-1)}$

note that:

$\phi(n) = (p-1) * (q-1)$

So: $E * D = 1 \pmod{\phi(N)}$

3- Breaking RSA

In the following section we consider ways a cryptanalyst might try to determine the secret decryption key from the publicly revealed encryption key.

3-1 Factoring N

Factoring N would enable an enemy cryptanalyst to break the RSA method. The factors of N enable him to compute $\phi(n)$ and thus D. Fortunately, factoring a number seems to be much more difficult than determining whether it is prime or composite.

An Approach for Breaking RSA Public Key Cipher System

Dr. Ala'a. H. Al-Hamami
Head of Computer Sciences Department
Al-Rafidain University College

Abstract

The only known cryptosystem which can be adapted for the authentication and secrecy at the same time is the Rivest, Shamir and Adleman (RSA) scheme(1). To encrypt a message M with RSA method, using a public encryption key (E, N). To decrypt the ciphertext is by using the decryption key which is a pair of positive integers (D, N). Each user makes his encryption key public, and keeps the corresponding decryption key private.

A cryptanalysts chore is to break an encryption; this means that the cryptanalyst will attempt to deduce the meaning of a ciphertext message, or determine a decrypting algorithm that matches an encrypting algorithm.

There are three basic methods of attack;

a- ciphertext-only attack,
b- known-plaintext attack,
c- and chosen-plaintext attack.

In this research a new approach is introduced for breaking RSA scheme and this by extracting the factoring of the public (N) to find the prime numbers (p, q) values. This could be done through the creation of a dynamic database. By using different techniques, it is possible to minimize the database size for possible implementation and efficiency. This could be done ignoring the even value of N and the no prime value of q.

Key words: *Public key, RSA, Database, Attacks, Crypto Analysis.*

1- Introduction

Cryptology, in general, splits into two subdivisions: cryptography and cryptanalysis. The cryptographer seeks to find methods to ensure the secrecy of messages, while the cryptanalyst seek to undo the former's work by breaking a cipher or by forging coded signals that will accepted as authentic. The objective of this research is to implement an approach for breaking the RSA public key system.

An encryption algorithm may be breakable, meaning that given enough time and data to an analyst could determine the algorithm. However, practically is also an issue. A particular cipher scheme may have an inverse deciphering scheme that requires 10^{30} operations. On a current technology computer performing on the order of 10^{10} operations per second, this decipherment would require 10^{20} seconds, or roughly 10^{12} years. In this case, although we know that theoretically a deciphering algorithm exists, the deciphering algorithm can be ignored as infeasible using current technology.

Note two things about the breakability of encryption algorithms. First, the cryptoanalyst cannot be expected to try just the hard, long way. In the example above the obvious decryption might require 10^{30} machine operation, but a more ingenious approach might required 10^{15} operations. At the speed of 10^{10} operations per second, 10^{15} operations take slightly more than one

Contents

1- English Section:

Contents	Page No.
- An Approach for Breaking RSA Public Key Cipher System Dr. Ala'a. H. Al-Hamami	3
- Adaptive Ciphertext-Only Attack Using Genetic Programming With Indexed Memory Dr. W. A. K. Al-Hamdani, Dr. A. F. Abdul Kader, W. S. Awad	9
- Use of Genetic Algorithm (GAs) in The Cryptanalysis Nonlinear Stream Cipher (NLSC) Dr. W. A. K. Al-Hamdani, S. A. Al-Ageelee	15
- Infra Red Remote Pc Keyboard W. A. Jabbar	24
- Image Guider Ahmad S. Nori Laheeb M. Ibrahim Najla Badeaa	30
- Shorter Signature Verification Time With Improved Digital Signature Standard, DSS Hamza A. Al-Sewadi Khaldon I. Arif	38

2- Arabic Section:

رقم الصفحة	المحتويات
٥	– تطبيق مُحوسَب لتصحيح بعض الأخطاء النحوية في الجملة العربية البسيطة محمد نعمان مراد

ELECTRONIC COMPUTERS

Electronic Computers Restricted Scientific Journal

Vol. No. 32	Issued By The National Computer Center Biannually

Correspondence:
Chief of Editorial Board,
Journal of Electronic Computers,
Ministry of Higher Education & Scientific Research
P.O.Box 3261, Sadoon-Baghdad-IRAQ

Annual Membership:
10000 I.D. for Gov. Establishments & individuals - inside Iraq.
25 U.S.D. for Gov. Establishments & individuals - outside Iraq.

Ministry of Higher Education & Scientific Research

National Computer Center

ELECTRONIC

COMPUTERS

Electronic Computers Restricted Scientific Journal

Vol. No. Thirty Second 1819H - 1998A

Electronic Computers	Issued By The National Computer Center Biannually