Encryption & Hiding Information in Internet Files HTML & XML

Dujan B. Taha dujan_taha @uomosul.edu.iq
Ahmed S. Nori ahmed.s.nori @uomosul.edu.iq
Najla B. Ibraheem Najla.dabagh @uomousl.edu.iq
College of Computer Science and Mathematics
University of Mosul

Received on:13/07/2009 **Accepted on:03/11/2009**

ABSTRACT

In order to achieve communication security, cryptography and information hiding in different media are used.

In this work, a system for hiding text in Internet files namely, HTML and XML has been built. Two proposed algorithms have been designed and implemented to embed and extract secret information from these files.

Hiding in HTML files was done by first encrypting the message using Linear Feedback Shift Register (LFSR) and embed the encryption key into the HTML tags. Then, the encrypted secret message was embedded into an image in the HTML page.

Hiding in XML files was achieved using non linear feed back shift register to encrypt the secret message. The resultant encryption key was embedded inside XML definition file namely, Document Type Definition (DTD) file which is invisible to the user. The encrypted message was embedded inside the XML component of the file.

Experimental results demonstrated that the proposed algorithms are secure and efficient. The image carrying the secret information is identical (by Human Visual System HVS) to the original image as well as the ability to embed a lot of information inside the files. Visual C++ was used to access Internet files whereas Matlab Version 7 was used to implement the used encryption methods and graphical user interface.

Keywords- HTML, XTML, LFSR, DTD, HVS.

تشفير وإخفاء المعلومات في ملفات الإنترنت HTML و XML

دجان بشير طه **أحمد سامي نوري** **نجلاء بديع إبراهيم**
كلية علوم الحاسوب والرياضيات / جامعة الموصل

الملخص

ادت الحاجة لتحقيق امنية الاتصالات إلى استخدام طرائق مختلفة لهذا الغرض كالتشفير واخفاء المعلومات في وسائط مختلفة. تم في هذا العمل بناء نظام لإخفاء المعلومات في ملفات الانترنت HTML و(Hyper Text Markup Language) XML و (Extensible Markup Language) حيث تم تصميم وتنفيذ خوارزميتين مقترحتين لتضمين معلومات سرية واسترجاعها من الملفات المذكورة.

تتلخص خوارزمية الاخفاء في ملفات HTML بتشفير الرسالة السرية اولا باستخدام التشفير الانسيابي الخطي (LFSR_Linear Feedback Shift Register) وتضمين مفتاح التشفير في وسوم ملفات HTML اما الرسالة السرية فيتم تضمينها داخل صورة في ملف HTML باستخدام اسلوب التضمين

في الخليتين الثنائيتين الاقل اهمية . اما في خوارزمية الاخفاء في ملفات XML فقد تم اعتماد التشفير الانسيابي اللاخطي (NLFSR) لغرض تشفير الرسالة السرية وخزن مفتاح التشفير داخل الملف التعريفي (DTD_Document Type Definition) الذي يكون غير مرئي للمستخدم بينما تم خزن الرسالة المشفرة في مكون XML للملف.

اثبتت النتائج العملية كفاءة وامنية الخوارزميتين من ناحية ان المعلومات المخفية لم تحدث أي تغيير او تشوه على الصورة الحاملة لها بالاضافة إلى امكانية تضمين معلومات كبيرة الحجم داخل الملفات المذكورة. تم استخدام لغة (Visual C++) لغرض التعامل مع ملفات الانترنت بينما تم استخدام (Matlab) بإصداره السابع لغرض تنفيذ طرق التشفير المستخدمة وواجهة المستخدم الرسومية.

الكلمات المفتاحية: HTML، XML، LFSR، DTD.

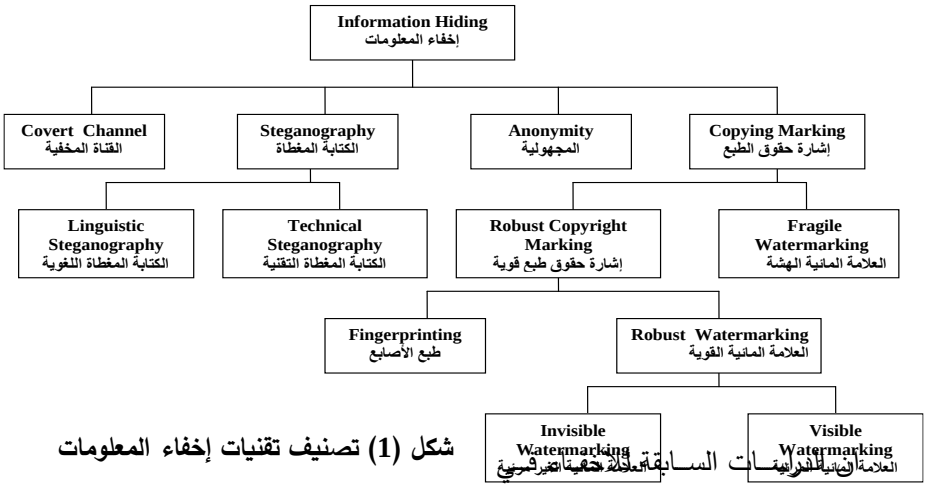
1. مقدمة:

أدى النمو المتزايد لتطبيقات الوسائط المتعددة على شبكات الاتصال إلى زيادة الحاجة إلى توفير طرائق كفوءة تعمل على حماية البيانات والملكية الخاصة بالفرد لذلك كان لابد من ظهور وسائل تعمل على توفير امن لهذه الوسائط لحمايتها من السراق والمتطفلين من العبث بها وتحريفها أو سرقتها ونشر المعلومات الحساسة منها. ومن هنا ظهرت الحاجة إلى توفير وسائل أمنية للبيانات، ومن هذه الوسائل علم التشفير Cryptography الذي يعني بتوفير حماية لخزن البيانات ونقلها عن طريق استخدام مفتاح سري، لذلك كان التشفير وما يزال الطريقة الناجحة لحماية البيانات المخزونة والمرسلة عبر الشبكة، ولكن مع ازدياد شبكات التناقل وشبكة المعلومات العالمية Internet أصبح من الصعب المحافظة على هذه البيانات، ولاسيما إنها تكون دائما في متناول الجميع عبر شبكة الانترنت في صيغة غير واضحة تبعث على الشك والاهتمام لدى المتطفل والسارق لفتح هذا التشفير أو تدمير المعلومات المرسلة.

إن إخفاء المعلومات (Information Hiding) يعني تضمين معلومات في معلومات أخرى ظاهرها لا يدعو إلى الشك ولا يلفت الانتباه، وتكون غير مدركة من قبل المتطفلين والمهاجمين، ولذلك لن تكون المعلومات مشاعة لمستخدمي الشبكة، بل يبقى محتواها حكراً على الجهات ذات العلاقة، والتي تكون على دراية بكيفية استخراج هذا المحتوى.

إن علمي الإخفاء والتشفير يستخدمان مع اختلافهما في تحقيق امنية البيانات، ففي التشفير يمكن لأي طرف إن يكتشف إن ثمة طرفين يتصلان بطريقة مشفرة، إما تقنية إخفاء المعلومات فتُخفى أصلا وجود الاتصال فلا يمكن لأحد إن يلاحظ وجود طرفين يتبادلان الرسائل عبر قنوات الاتصال. يوضح الشكل(1) تصنيف تقنيات إخفاء المعلومات [6]. تناول البحث اسلوب الكتابة المغطاة (Steganography) في تصميم وتنفيذ خوارزميات الاخفاء المقترحة.

منذ ان بدأت تقنيات الأنترنت تتطور بصورة مذهلة، فأن كمية المعلومات المتناقلة الكترونيا قد ازدادت بصورة كبيرة ومع زيادة عدد التطبيقات التي لاتستخدم النص الواضح فقط وانما البيانات المهيأة (formatted) مثل ملفات HTML و XML ازدادت اهمية تحقيق الامنية في مثل هذه الملفات. وبسبب ان المستندات المكتوبة باستخدام هذه الملفات تتوزع بصورة كبيرة على الويب لذلك فأن كشف الاتصالات التي تستخدم هذه الملفات يكون صعب مع الأخذ بنظر الاعتبار الصلاحية التقنية في مراقبة النظام كذلك، ومقارنة مع طرق الاخفاء الخاصة بالصور والصوت فأن هناك طرق قليلة لأخفاء المعلومات في النص ولاتوجد دراسة تقريبا على طرق اخفاء المعلومات في المستندات المهيأة HTML و XML.[10]



شكل (1) تصنيف تقنيات إخفاء المعلومات

ان الدراسات السابقة لاخفاء المعلومات في ملفات HTML كانت فكرتها الرئيسية الاستفادة من وجود الفراغ الابيض داخل نص صفحة الويب بحيث يُضَمَّن رمز واحد من البيانات السرية لكل فراغ ابيض. تلون البيانات السرية بنفس لون خلفية صفحة الويب HTML بعد ذلك يتم ادخال البيانات السرية الملونة داخل الفراغات البيضاء في نص صفحة الويب الاصلية.

اما في مجال الاخفاء في ملفات XML فقد اعتمدت التقنيات السابقة للاخفاء في هذه الملفات على تمثيل العناصر الفارغة، الفراغات البيضاء في العلامات، التسلسل الظاهر للعناصر، وظهور تسلسل الصفات [4][10]. اعتمدت جميع التقنيات اعلاه على الاخفاء في ملف XML فقط ولم تعتمد على مكونات XML الاخرى (CSS, XSL, DTD) كغطاء للمعلومات السرية.

ان الخوارزميتين المقترحتين في البحث اضافت للتقنيات الموجودة في الدراسات السابقة بعض الميزات المهمة والضرورية للاخفاء في ملفات الانترنت ففي خوارزمية الاخفاء في ملف HTML تكون البيانات المخفية باستخدام هذه الخوارزمية غير مرئية عند التصفح ولا في ملف

النص الاصلي اضافة الى ان حجم الملف بعد الاخفاء لايتغير. وبالنسبة لخوارزمية الاخفاء في ملف XML فقد اعتمدت على استخدام احد مكونات ملف XML الاساسية وهو ملف DTD لتضمين مفتاح التشفير فيه كون هذا الملف يكون غير مرئي للمستخدم. كما ان الطريقة المستخدمة للاخفاء في مكون XML تتيح اخفاء كميات كبيرة من البيانات. وبالاضافة لما ذكر اعلاه فقد تم اجراء عملية تشفير للبيانات السرية قبل التضمين باستخدام طريقة تشفير معتمدة ذات امنية عالية.

2. ملفات HTML و XML :

1.2 ملفات HTML:

تعالج لغة HTML النص الواضح بالإضافة إلى البيانات المهيأة المكتوبة، تعتبر هذه اللغة لغة النشر القياسية إلى الشبكة العنكبوتية العالمية (World Wide Web). تقدم لغة HTML وسائل لنشر المستندات مباشرا (online) مع العناوين والنص والجداول والقوائم والصور.....الخ. كذلك توفر إمكانية استرجاع المعلومات بشكل مباشر من خلال وصلات النص ألتشعبي (الملف النصي) بالإضافة إلى تصميم نماذج لمعالجة المعاملات بخدمات أخرى لاستخدامها في البحث عن المعلومات، عمل حجز، شراء مواد .... الخ. يتكون ملف HTML من جزئين، الجزء الأول يمثل الترويسة (Header) التي تحتوي على معلومات حول الصفحة كالعنوان، والجزء الثاني يمثل جسم الصفحة (body page) الذي يحتوي على مكونات الصفحة، المكونات تكون عبارة عن مجموعة من الوسوم بالإضافة إلى الخصائص التي يمكن إضافتها إلى داخل كل وسم من اجل التحكم بالشكل العام لتأثير الوسم. وتطلق كلمة خاصية على التعابير التي تضاف إلى الوسم، من اجل تحديد الكيفية أو الشكل الذي تعمل به هذه الوسوم. وبعبارة أخرى فأن الوسم يقوم بأخبار برنامج التصفح عن العمل الذي يجب القيام به أما الخاصية فتحدد الكيفية التي سيتم بها أداء هذا العمل [1][8].

2.2 ملفات XML:

تعتبرلغة XML (Extensible Markup Language) لغة حديثة تكونت في بداية سنة 2000، وهي لغة مشابهة للغة HTML ببعض صفاتها حيث لا تحتوي على جمل التحكم والدوران، وعند الحاجة لاستخدام هذه الجمل يجب تضمين شفرات من لغات JavaScript, CGI, JAVA كما في لغة HTML [3]. تستخدم بصورة واسعة في تبادل البيانات بين صفحات الويب وأنواع معينة من الملفات النصية (Text Files)، وهذه اللغة تستخدم العناصر

(Elements) أو ما يعرف بالوسوم (Tags) ليتم بواسطتها تعيين بداية ونهاية مقاطع معينة من النص لكي تعطي معنى ما أو لكي يتم التعامل معها بشكل ما.

هناك تشابه بين لغة HTML و لغة XML ففي كليهما يبدأ كل عنصر (Element) أو وسم (Tag) بالرمز '<' وينتهي بالرمز '>' (واللغتان تتعاملان مع الوسوم بنفس الطريقة المبينة سابقا)، وكلاهما أيضا يحتوي على الاختلاف ما بين كل امر فتح وسم معين وأمر إغلاقه، إذ إن أمر الإغلاق لأي وسم يجب أن يحتوي على الرمز '/' قبل بداية الاسم.

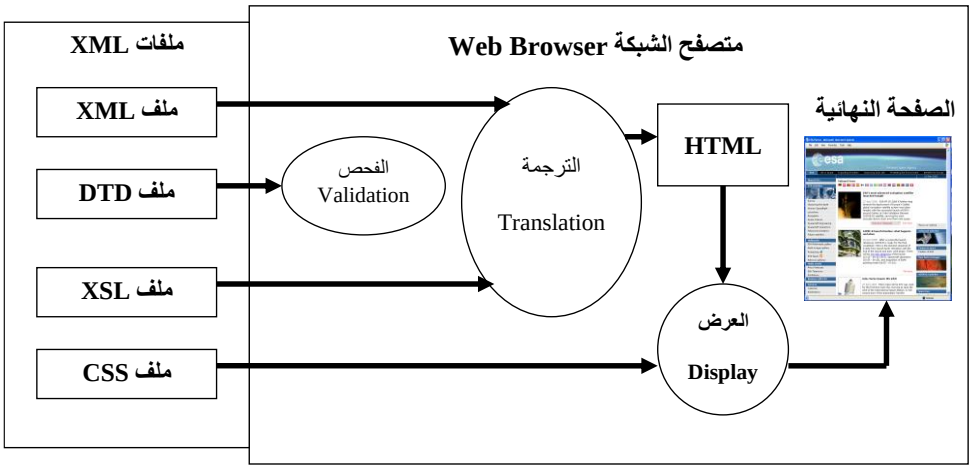
إن صفحات الشبكة (Web Pages) المكتوبة بلغة XML قد تحتوي على عدد غير محدد من الوسوم بشرط أن لا تتداخل هذه الوسوم فيما بينها. تتألف صفحة XML من[2][3]:

أ- ملف XML مكتوب فيه الهيكل المطلوب (مشابه لملف (source) الموجود في HTML)، وهذا العنصر يعتبر أساسيا.

ب- ملف تعريف DTD لكي يتم بواسطته فحص نوعية الوسوم المستخدمة وطريقة تعريفها وخصائصها، وهذا العنصر يعتبر أساسيا ايضا.

ج- ملف XSL (Extensible Style Language) الذي يتحكم بالعمل الحقيقي للوسوم (مثل عمل ربط (Link) بين موقعين أو عرض صورة معينة)، وهذا العنصر يعتبر اختياريا.

د- ملف CSS (Cascading Style Sheet) يقوم بعمل تنسيق للصفحة قبل عرضها (مثل تكبير وتصغير الخط أو تلوين الخط بلون معين)، وهذا العنصر يعتبر ايضا من العناصر الاختيارية. ويوضح الشكل(2) كيفية ربط العناصر الأربعة السابقة لتكوين صفحة XML، حيث ان كل الوسوم يتم تعريفها في ملف DTD ثم يتم استخدامها في الملف النصي لصفحة XML، وعند تشغيل صفحة XML سيقوم ملف XSL بتنفيذ عمل كل وسم، اما ملف CSS فإنه سيستخدم لتنسيق شكل الصفحة الناتجة.



شكل (2) كيفية ربط العناصر الاربعة لصفحة XML

1.2.2 تعريف الصفحة كصفحة XML :

يجب أن تحتوي كل صفحة XML على ما يعرف بالبادئة (Prolog) وهي عبارة عن جملة معينة تقوم بتعريف الصفحة على أنها صفحة XML وهي تكون بالشكل التالي :

<?xml version="1.0" encoding="UTF-8"?>

هذه الجملة يجب أن توضع في بداية كل صفحة (يجب أن لا تسبق بأي شيء حتى لو كان فراغا واحدا)، يجب أن تحتوي كل جملة بادئة على نوع النسخة (version) أما باقي الصفات فهي اختيارية (مثل صفة (encoding) المذكورة في المثال اعلاه) [2][9].

2.2.2 ملف تعريف نوع الصفحة DTD :

إن تصميم لغة XML يتيح للمستخدم ان يقوم بتعريف الوسوم المستخدمة وعمل كل منها، أي ان المستخدم سيقوم بتصميم لغة ثانوية تمكنه من التعامل مع مجال معين. يتعرف الحاسوب على معنى وعمل كل وسم عن طريق ملف تعريف نوع الصفحة (Document Type Definition) أو ما يسمى (DTD)، وهو عبارة عن ملف تحريري (Text File) يحتوي على هيكل صفحة XML إضافة إلى أسماء كل الوسوم والصفات التابعة لهذه الوسوم وترتيبها، بواسطة هذا الملف سيتعرف متصفح الشبكة على كافة الوسوم الموجودة في صفحة XML إضافة إلى التعرف على الهيكل المطلوب في تلك الصفحة، حيث يتم تعريف الوسم باستخدام الأمر (ELEMENT!) وتعريف كل الصفات لوسم معين عن طريق الأمر (ATTLIST!) [2][3].

يتم الربط بين ملف التعريف وصفحة XML بكتابة ملف التعريف كملف منفصل عن الصفحة الأصلية ثم يتم عمل وصلة (Link) ما بينهما باستخدام الجملة التالية (يكون موقعها بعد جملة البادئة مباشرة) :

<!DOCTYPE Student SYSTEM "DTD_Name.dtd">

حيث يمثل (Student) الوسم الرئيسي (Main Element) والذي من شروط اللغة أن يتكرر مرة واحدة فقط وان يحتوي على كافة الوسوم الأخرى، وبهذا فإن متصفح الشبكة (Internet Explorer) سيأخذ التعاريف من ملف التعريف ثم يتعامل مع الوسوم الموجودة في صفحة XML بشكل اعتيادي .

3.2 الفرق بين ملفات HTML وXML :

ان الفرق الرئيسي بين لغتي HTML و XML هو ان الاولى تستخدم وسوما جاهزة لتعريف عمل معين، أي أن كل الوسوم الموجودة قد تم تعريفها سابقا من قبل المصممين لكي تقوم بعمل معين (يتم خزن هذه الوسوم ومعانيها وعملها في متصفح الشبكة (Internet Explorer))، وهذا يسبب عدة مشاكل منها إن على المستخدم أن يحفظ ويتذكر هذه الأوامر المخزونة لكي يستطيع تصميم صفحة انترنت معينة، ومن المشاكل أيضا هو تحديد عدد هذه الوسوم (حيث يبلغ عددها حاليا ما يقارب 270 إلى 280 وسم) مما يحد من إمكانيات اللغة ويزيد من صعوبة تطويرها، لذلك تم تصميم لغة XML بطريقة ثورية تتيح للمستخدم أن يقوم بتعريف الوسوم التي سيقوم باستخدامها وكذلك تعريف العمل الذي ستقوم به هذه الوسوم، أي أن لغة XML هي لغة تقوم بتصميم لغات ثانوية لكي يتمكن المستخدم (مهما كان مجال عمله) من تصميم لغة تتيح له التعامل مع هذا المجال. بالإضافة إلى الفرق الرئيسي المذكور اعلاه هنالك اختلافات اخرى بين اللغتين يتم تلخيصها بما يلي[2][3][5][9] :

- غلق الوسوم (Closing Tags)

توفر لغة HTML استخدام وسوم مفردة ومزدوجة، عند كتابة الوسم المفرد لا نحتاج إلى إن نكتب الوسم الأيمن الذي يمثل وسم الإغلاق وهذا لا يسبب مشكلة في إثناء التنفيذ أو عرض الصفحة، بينما هنالك وسوم أخرى تتطلب وسم للنهاية. في لغة XML تكون المرونة اقل مما هي عليه في لغة HTML حيث تظهر رسالة خطا في حالة كتابة المصدر (Source) بصورة مخالفة لهيكل ملف XML.

- تداخل العناصر (Element Nesting)

على الرغم من أنها غير موصى باستخدامها، يمكن ان تتداخل وسوم لغة HTML بشكل مناسب وتؤدي وظيفتها بشكل مناسب أيضا حيث من الممكن إن نغلق الوسم الذي كتب اولا قبل الوسم الذي بعده، إما في ملف XML إذا بدأ وسم معين قبل وسم اخر, فيجب ان ينتهي الوسم الذي بدأ اولا اخيرا، إي إن الوسم الذي يمثل الأب (Parent) يجب إن ينتهي بعد الوسم الذي يعتبر الابن (Child) لذلك الأب.

- مواقع الصفات (Location of Attributes)

تعتبر لغة HTML أكثر مرونة من لغة XML حيث إن كتابة الخصائص داخل الوسم ليس من الضروري إن تتبع تسلسل معين، أما في ملفات XML فهناك ترتيب معين يجب إن نتبعه

لكتابة الخصائص داخل الوسم (ترتيب هذه الخصائص داخل الوسم يتعلق بالتعريف الخاص بهم في ملف التعريف (DTD) المرافق لملف XML).

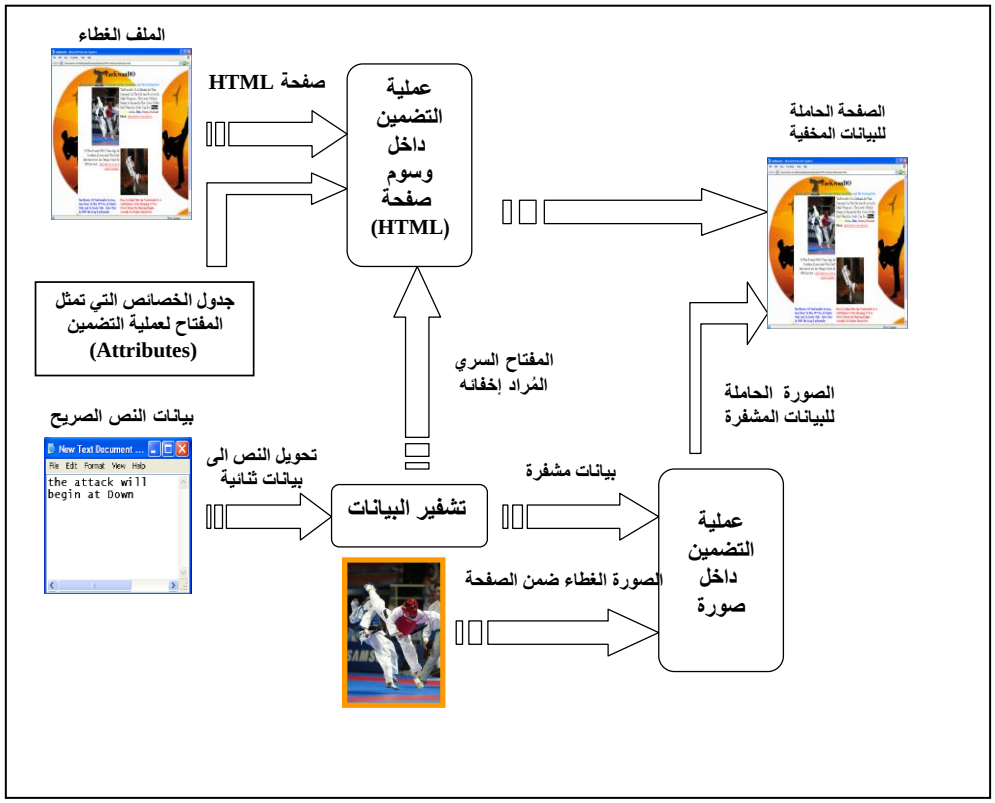
- حالة الحروف (Case Sensitivity)

تعتبر لغة XML حساسة من ناحية تقبلها لحالة الحروف صغيرة أو كبيرة بخلاف لغة HTML.

3. تصميم وتنفيذ خوارزميات الاخفاء المقترحة :

1.3 الاخفاء في ملف HTML

إن عملية حشر بيانات وبأي صيغة كانت داخل الملف النصي الخاص بملف HTML تكون ظاهرة ومرئية إما عن طريق عرض صفحة الويب أو في الملف النصي الأصل (Source)، لذلك سوف نتناول الطريقة المقترحة للإخفاء داخل صفحة الويب بحيث تكون هذه البيانات غير مرئية عند التصفح أو في الملف النصي الأصل، تتلخص عملية الإخفاء أولاً بتشفير الرسالة السرية باستخدام التشفير الانسيابي الخطي ثم أخفاء مفتاح التشفير داخل وسوم صفحة HTML وأخيراً تضمين الرسالة السرية المشفرة في صورة داخل ملف HTML. الشكل(3) يوضح عملية الإخفاء. وتتناول الفقرات التالية شرح مراحل عملية الإخفاء بالتفصيل.



شكل (3) عملية التضمين في صفحة HTML

1.1.3 تشفير الرسالة السرية :

تم استخدام التشفير الانسيابي الخطي لتشفير الرسالة السرية [7]. يتم أولاً إدخال المفتاح السري الذي يمثل مراحل (LFSR) والروابط فيما بينها، ثم إدخال بيانات النص الصريح (بعد تحويلها إلى سلسلة ثنائية).

بعد توليد مفتاح التشفير الانسيابي يتم عمل XORing بين سلسلة النص الصريح والسلسلة المولدة، فيكون الناتج سلسلة النص المشفر التي سيتم إخفائها داخل صورة ضمن ملف HTML.

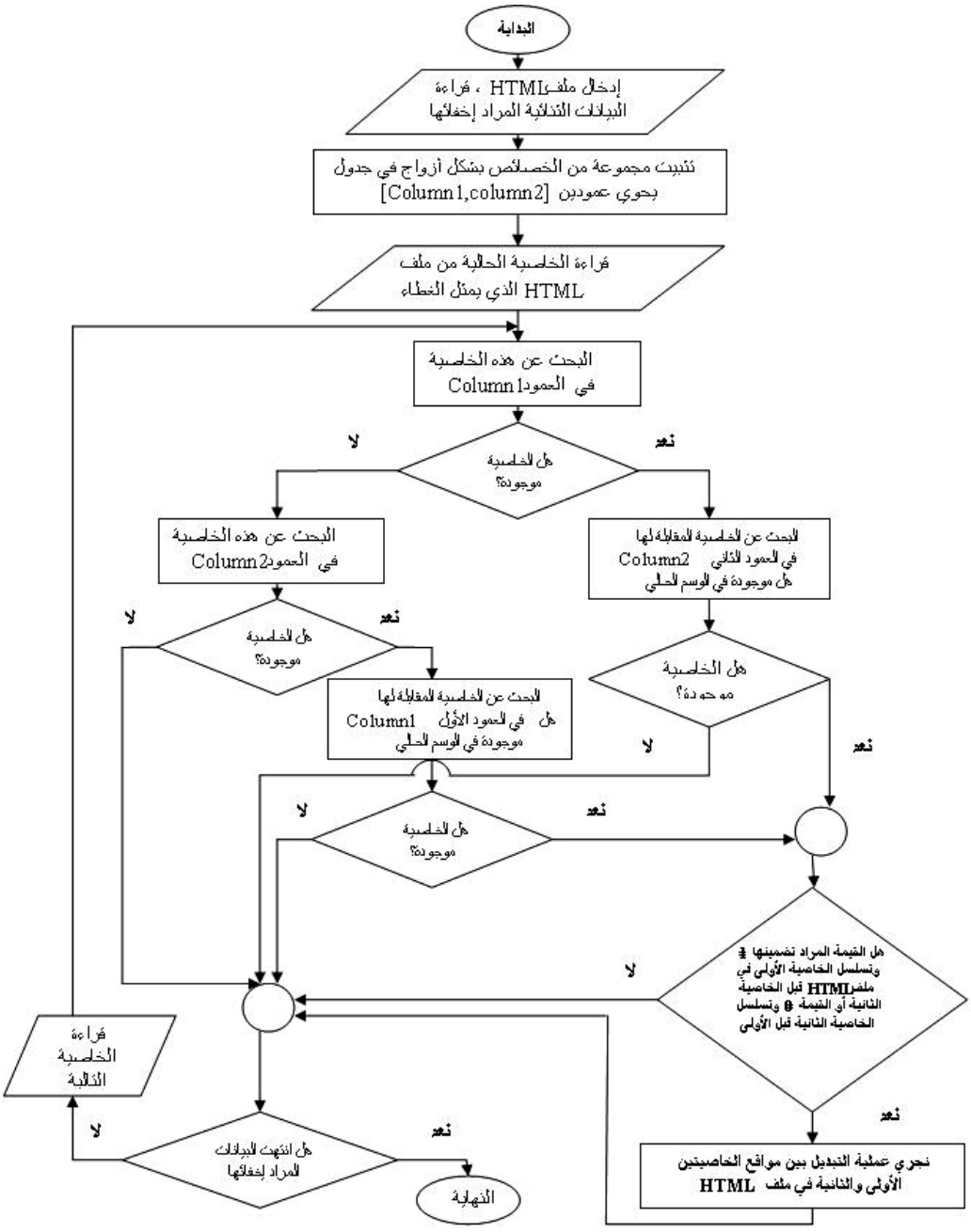
2.1.3 تضمين مفتاح التشفير داخل وسوم صفحة HTML :

كما ذكرنا سابقاً، يتكون الهيكل الخاص بملف HTML من مجموعة من الوسوم التي تتضمن الخصائص (Attributes). إن تسلسل الخصائص لكل وسم لا يؤثر على عرض

المعلومات على الصفحة لذلك من الممكن إن نستفيد من تغيير تسلسل الخصائص لكل وسم في عملية التضمين. في عملية التضمين يجب إن يكون هناك قاعدة بيانات أو جدول يحتوي على تسلسل مجموعة من الخصائص المزدوجة .

يتم قبل عملية التضمين حساب طول السلسلة المراد تضمينها ثم يتم تحويل هذا الطول إلى النظام الثنائي ممثل بكتلة ثمانية واحدة (8-bit) وتضاف إلى بداية السلسلة التي سيتم إخفائها. تعتبر هذه الخطوة ضرورية في عملية الاسترجاع. يتم في عملية التضمين استخدام الجدول الذي يحتوي على الخصائص المزدوجة. يتم أولا قراءة جميع الوسوم ثم يتم البحث عن كل خاصية للوسم الحالي في الجدول إذا كانت الخاصية موجودة ضمن الحقل الأول للجدول، ثم البحث عن الخاصية المقابلة لها في الحقل الثاني هل موجودة في الوسم الحالي إذا كانت موجودة ننظر إلى تسلسل الخاصيتين حيث إذا كانت الخاصية التي هي من الحقل الأول قبل الخاصية من الحقل الثاني وكانت القيمة المطلوب تضمينها (0) سوف يبقى نفس التسلسل ضمن الوسم إما إذا كان المطلوب تضمين القيمة(1) في هذه الحالة سوف نغير تسلسل الخاصيتين في ملف HTML، حيث تكتب الخاصية الموجودة في الحقل الثاني قبل الخاصية الموجودة بالحقل الأول في الملف.

إما إذا كانت الخاصية ليست موجودة في الحقل الأول سوف يتم البحث عنها في الحقل الثاني من الجدول، إذا كانت الخاصية موجودة ضمن الحقل الثاني للجدول، يتم البحث عن الخاصية المقابلة لها في الحقل الأول هل موجودة في الوسم الحالي إذا كانت موجودة ننظر إلى تسلسل الخاصيتين حيث إذا كانت الخاصية التي هي من الحقل الثاني قبل الخاصية من الحقل الأول وكانت القيمة المطلوب تضمينها (1) سوف يبقى نفس التسلسل ضمن الوسم إما إذا كان المطلوب تضمين القيمة (0) في هذه الحالة سوف نغير تسلسل الخاصيتين في ملف HTML، يتم تكرار هذه العملية إلى الانتهاء من تضمين جميع القيم. ويمثل الشكل (4) المخطط الانسيابي لعملية التضمين في وسوم ملف HTML.



شكل (4) مخطط انسيابي لعملية التضمين في وسوم ملف HTML

3.1.3 تضمين الرسالة المشفرة في صورة داخل صفحة HTML :

تم استخدام الصور الملونة من نوع BMP كغطاء لخزن البيانات السرية المراد إخفاؤها وذلك عن طريق تحويل الصورة إلى صيغة YCbCr (هذه الصيغة ستحول الصور الملونة إلى ثلاث طبقات هي Y وهي تمثل شدة إضاءة الصورة، Cb التي تمثل التلوينة الزرقاء و Cr التي تمثل التلوينة الحمراء) ثم تغيير نقاط الصورة في طبقة Y عن طريق استخدام الخليتين الثنائيتين الأقل أهمية (Two Least Significant Bits) من كل خلية ثمانية (Byte) (تم اختيار الصور من نوع BMP لكونها الصيغة القياسية للنوافذ Windows حيث يكون الخزن فيها بصيغة غير معتمدة على نوع أجهزة العرض ,كما إن الصور من هذا النوع لا تحتوي على إي نوع من أنواع المعالجة مما يوفر مساحة اكبر للإخفاء).

تم اخذ أول (8) نقاط صورية وخزن طول النص المشفر في الخليتين الثنائيتين الأقل أهمية من كل نقطة ثم اخذ باقي النقاط واستخدام نفس الطريقة لخزن النص المشفر فيها .

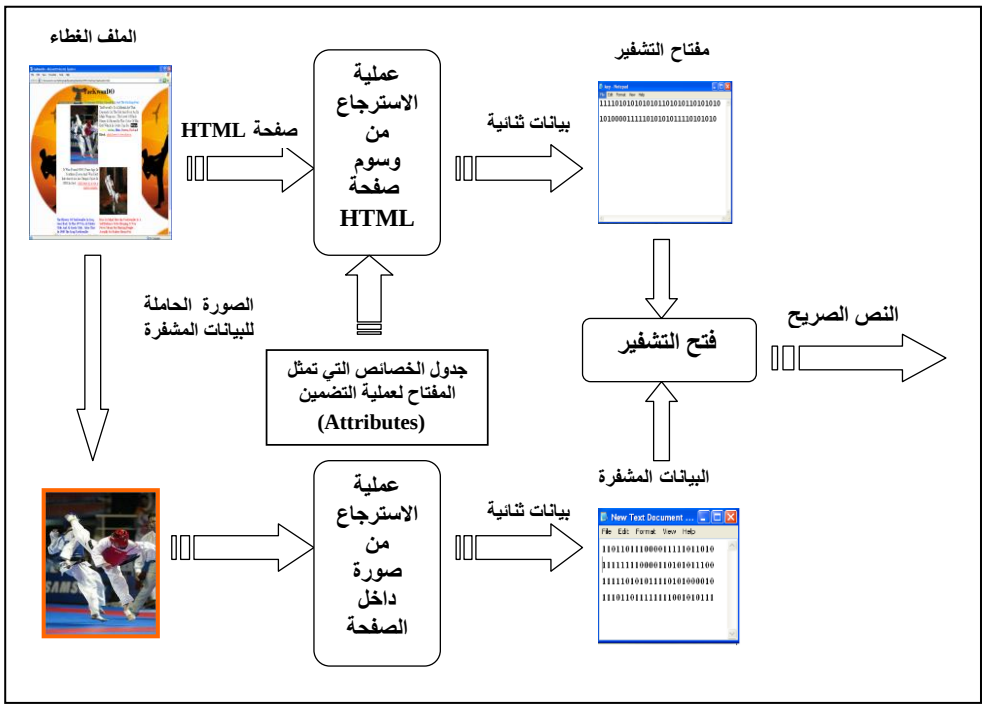
2.3 استرجاع المعلومات المخفية من ملف HTML :

تتلخص العملية أولاً باسترجاع مفتاح التشفير من صفحة HTML الذي يمثل مفتاح فك الشفرة، ثم استرجاع البيانات المشفرة من صورة داخل صفحة HTML، وأخيرا القيام بعملية فك الشفرة للبيانات باستخدام التشفير الانسيابي الخطي. والشكل (5) يوضح عملية الاسترجاع، وتتناول الفقرات التالية توضيح عملية الاسترجاع بالتفصيل.

1.2.3 استرجاع مفتاح التشفير من داخل وسوم صفحة HTML :

إن عملية الاسترجاع مشابهة لعملية التضمين ولكن بدون إجراء تبديل بين مواقع الخصائص حيث تتلخص عملية الاسترجاع بالبحث وبمسح ملف HTML وفحص التسلسل لتحديد القيمة "0" أو "1".

في البداية نقوم باسترجاع أول كتلة ثمانية (8–bit) من البيانات المخفية ونحولها إلى عدد صحيح الذي يمثل طول السلسلة التي سيتم استرجاعها. ثم القيام بقراءة جميع خصائص الوسم الحالي ونبحث عن كل خاصية للوسم في الجدول إذا كانت الخاصية موجودة ضمن الحقل الأول للجدول، يتم البحث عن الخاصية المقابلة لها في الحقل الثاني هل موجودة في الوسم نفسه إذا كانت موجودة ننظر إلى تسلسل الخاصيتين حيث إذا كانت الخاصية التي هي من الحقل الأول قبل الخاصية من الحقل الثاني تكون القيمة المسترجعة (0) وإلا فأن القيمة المسترجعة تكون (1).

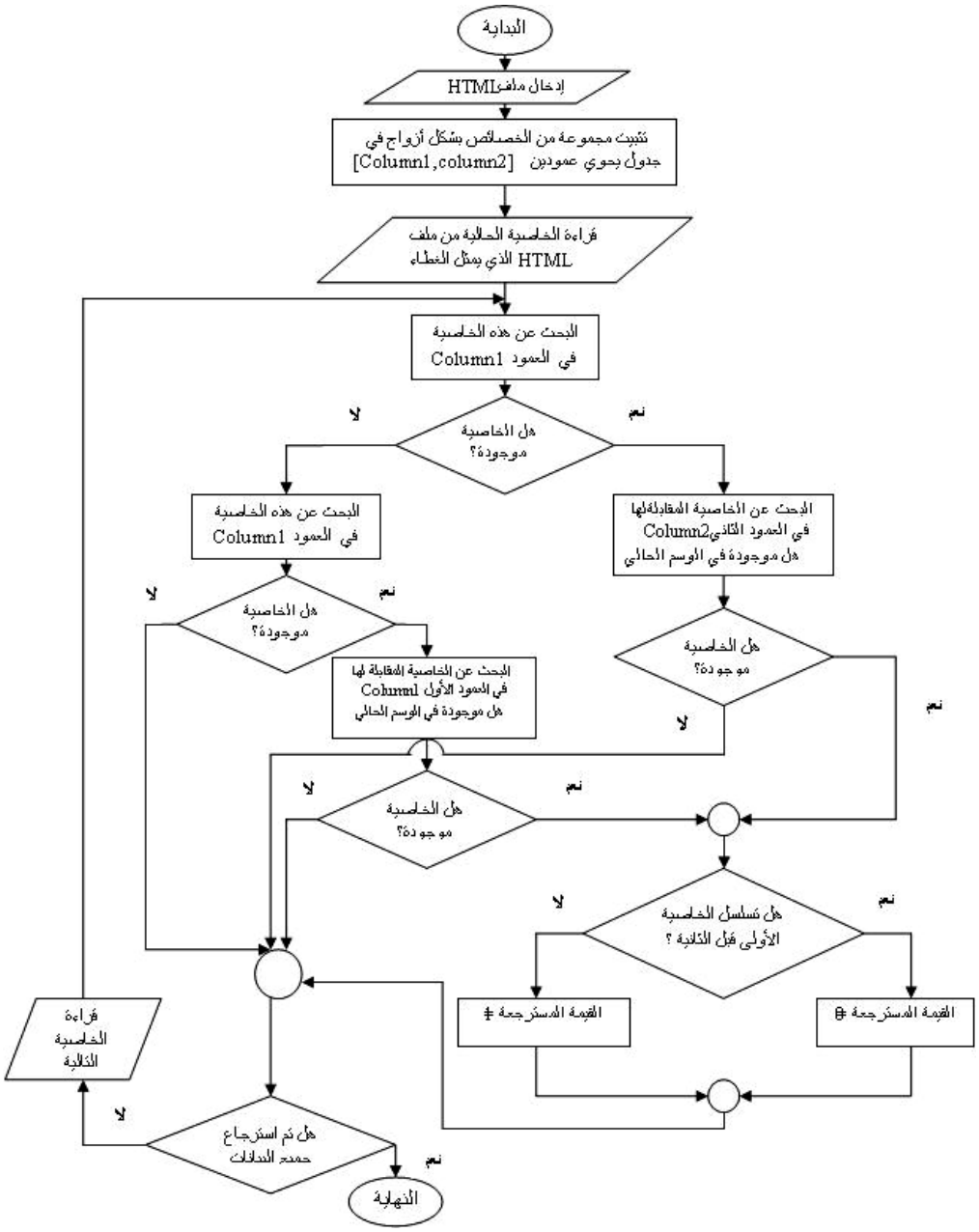


شكل (5) عملية الاسترجاع من ملف HTML

إما إذا كانت الخاصية ليست موجودة في الحقل الأول سوف يتم البحث عنها في الحقل الثاني من الجدول، إذا كانت الخاصية موجودة ضمن الحقل الثاني للجدول، يتم البحث عن الخاصية المقابلة له في الحقل الأول هل موجودة في الوسم الحالي إذا كانت موجودة ننظر إلى تسلسل الخاصيتين حيث إذا كانت الخاصية التي هي من الحقل الثاني قبل الخاصية من الحقل الأول في ملف HTML تكون القيمة المسترجعة (1) وإلا فأن القيمة المسترجعة تكون (0)، يتم تكرار هذه الخطوات إلى أن نصل إلى الطول الذي تم استرجاعه في البداية. ويمثل الشكل (6) المخطط الانسيابي لعملية الاسترجاع من ملف HTML.

2.2.3 استرجاع الرسالة المشفرة من الصورة :

في الاسترجاع سيتم اخذ قيمة الخليتين الثنائيتين الأقل اهمية من أول (8) نقاط صورية لمعرفة طول النص المشفر الذي تم إرساله ثم يتم الاعتماد على هذا الطول في استرجاع البيانات المخفية من باقي نقاط الصورة أيضا عن طريق اخذ قيم الخليتين الثائيتين الأقل اهمية من كل نقطة صورية. بعد استرجاع البيانات المشفرة ومفتاح التشفير يتم القيام بعملية فك الشفرة وذلك باستخدام نفس الخطوات التي تمت في عملية التشفير للحصول على النص الصريح.

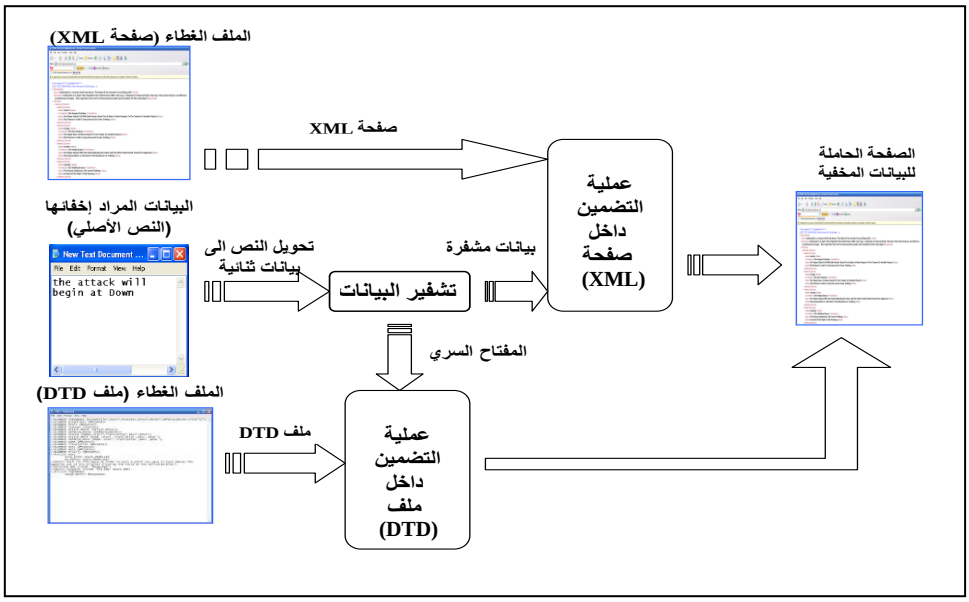


شكل (6) المخطط الانسيابي لعملية الاسترجاع من ملف HTML

3.3 الإخفاء داخل ملف XML :

إن الملفين الذين يكونان صفحة XML (ملف التعريف DTD والملف النصي الخاص بصفحة XML (Source)) يتعاملان مع البيانات بشكل مختلف ففي ملف التعريف DTD لا يكون أي تغيير في البيانات (على شرط إن لا يؤثر التغيير على الهيكل الخاص بالصفحة) مرئيا للمستخدم فمثلا إذا تم استخدام (*) أو (+) لتمثل عدد مرات تكرار الوسم فإن هذا لن يؤثر على صفحة XML التي يتم التعامل معها. إما الملف النصي فإنه عند العرض سيقوم بتحويل الصفحة إلى هيكلية محددة (Predefined Structure) تم وضعها من قبل مصممي اللغة، هذه الهيكلية ستلغي الكثير من العوامل الغير مرغوب بها عند العرض مثل الفراغات (White Space) فمثلا إذا تم وضع عشر فراغات في نهاية سطر معين فسيتم إلغائها عند العرض، كذلك سيتم إلغاء الأسطر الفارغة وإعادة ترتيب الوسوم بشكل هرمي يوضح علاقة كل وسم بالذي قبله.

تتلخص عملية الاخفاء اولا بتشفير الرسالة السرية باستخدام التشفير الانسيابي اللاخطي وتضمين هذه البيانات المشفرة في صفحة XML ثم اخذ المفتاح السري الخاص بعملية التشفير وتضمينه في ملف DTD. ويوضح الشكل(7) عملية الإخفاء في صفحة XML. فيما يلي سنشرح كيفية تشفير البيانات السرية ثم إخفائها والمفتاح السري في صفحة XML وملف DTD.



شكل (7) يوضح عملية الإخفاء في صفحة XML

1.3.3 تشفير الرسالة السرية باستخدام التشفير الانسيابي اللاخطي (NLFSR):

تم استخدام التشفير الانسيابي بنوعه اللاخطي لتشفير البيانات السرية قبل تضمينها [7]حيث تم اعتماد خوارزمية جيف وهي احدى خوارزميات التشفير الانسيابي اللاخطي لهذا الغرض.

2.3.3 إخفاء مفتاح التشفير داخل ملف التعريف DTD :

إن ملف التعريف يحوي عدة رموز يتم الاستفادة منها في تعيين عدد مرات تكرار الوسم وهي كالتالي:

1- الرمز '?' ومعناه إن الوسم سيكون غير موجود أو انه سيكون موجودا لمرة واحدة فقط.

2- الرمز '*' ومعناه إن الوسم سيكون غير موجود أو سيكون موجودا لمرة واحدة أو أكثر.

3- الرمز '+' ومعناه إن الوسم يجب إن يكون موجودا على الأقل لمرة واحدة.

والطريقة التي تم استخدامها للإخفاء هي تعيين وسم معين على انه وسم المفتاح (Key Element) ويتم هذا التعيين عن طريق الرمز '?' فيتم البحث في ملف التعريف عن الوسم المفتاح وبعد العثور عليه يتم البحث عنه في صفحة XML فإذا لم يتم العثور عليه في صفحة XML فالتعامل مع الرمزين ' + ' و' * ' سيكون كالتالي :

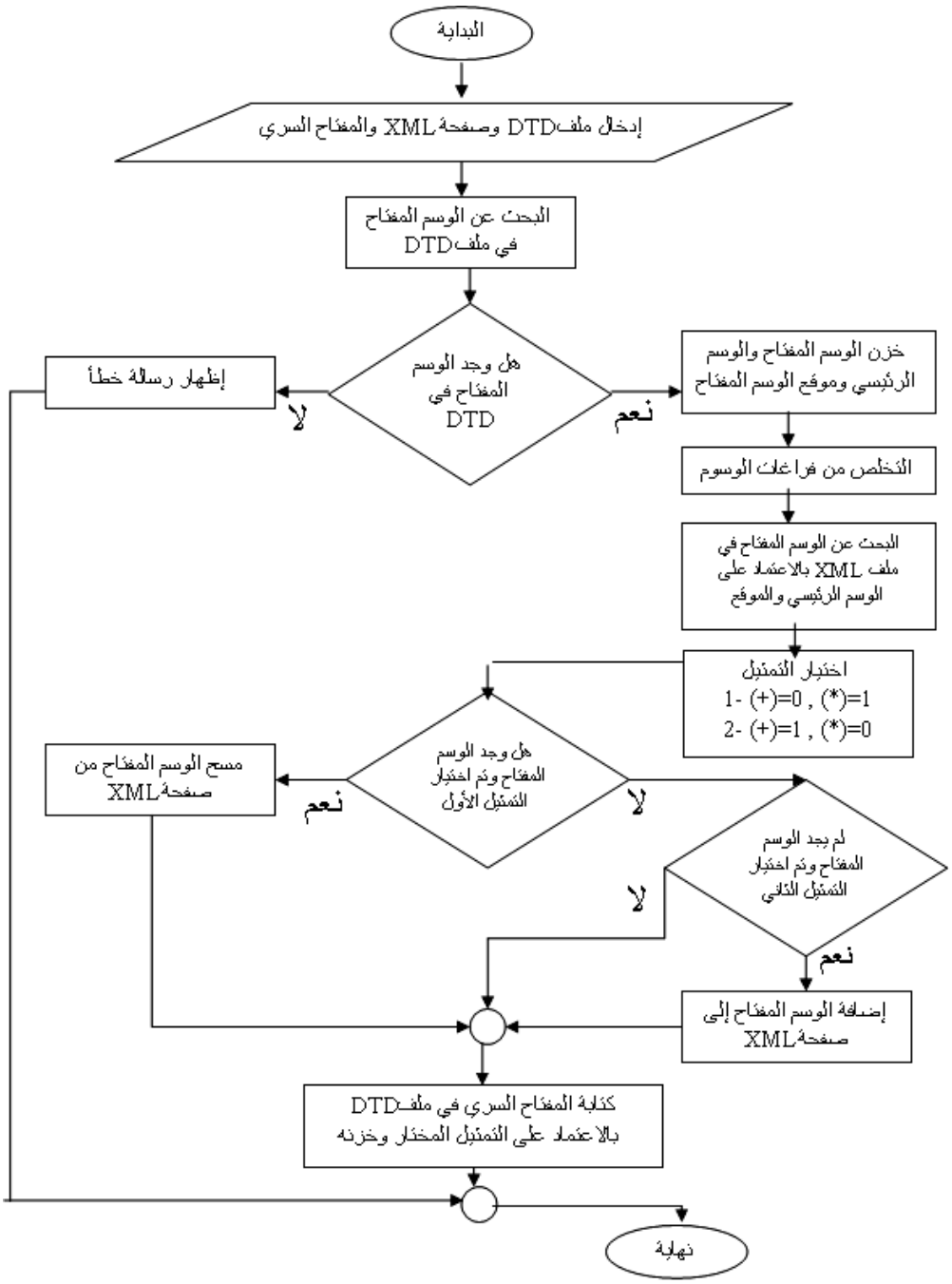
' + ' = 0 , **' * ' = 1**

اما اذا عثر عليه فإن التعامل معها على أساس :

' + ' = 1 , **' * ' = 0**

وبالاعتماد على ما تم ذكره فإن المفتاح السري سيتم خزنه واسترجاعه عن طريق قراءة ملف التعريف ثم تحويل '+' و '*' إلى (0) أو (1) حسب وجود أو عدم وجود الوسم المفتاح (عند الإخفاء بهذه الطريقة يتم اختيار التمثيل المطلوب من قبل الرمزين (+,*) وعلى أساسه يتم إضافة أو حذف الوسم المفتاح من صفحة XML).

لغرض اخفاء مفتاح التشفير، تم اخذ ملف تعريف الصفحة والبحث عن الوسم المفتاح فيه، وبعد العثور عليه يؤخذ هو والوسم الرئيس له إضافة إلى موقعه في الوسم الرئيس (لكي يتم تحديده بشكل دقيق وللتغلب على مشكلة استخدام كلمات مشابهة للوسم المفتاح) ثم يتم البحث عنه في صفحة XML ويتم اختيار نوع التمثيل، وعلى أساس هذا الاختيار سيتم إضافة أو مسح الوسم المفتاح من صفحة XML ثم يتم تضمين البيانات السرية حسب نوع التمثيل المختار. الشكل (8) يمثل المخطط الانسيابي لخوارزمية إخفاء مفتاح التشفير في ملف DTD .



شكل (8) المخطط الانسيابي لإخفاء المفتاح السري في ملف DTD

3.3.3 إخفاء البيانات السرية داخل صفحة XML :

تم إخفاء بيانات الرسالة المشفرة بالاعتماد على الوسوم نفسها، حيث إن لغة XML تتقبل وضع فراغ واحد قبل الرمز '>' وتتقبل وضع فراغين احدهما قبل الرمز '<' والآخر بعده، أي أن كل وسم من الممكن أن يحتوي على ستة أجزاء من الرسالة المشفرة وكالتالي :

1 <tag_name 2 > 3 Tag_Content 4 </tag_name 5 > 6

حيث إن الأرقام (1 إلى 6) تمثل موقع الفراغات التي سيتم التضمين فيها، وباعتبار إن (1) ممثل بفراغ وان (0) ممثل بغياب الفراغ من الممكن خزن كم هائل من البيانات في صفحات XML.

بالاعتماد على طريقة فراغات الوسوم فإن الرسالة المشفرة سيتم تضمينها في صفحة XML عن طريق اخذ كل خلية ثنائية منها ثم تضمينها بعد البحث عن الرمزين (< , >)، فإذا كانت الخلية الثنائية (1) سيتم إضافة فراغ إلى الصفحة في الموقع المحدد (إذا لم يكن هناك فراغ)، وإذا كانت (0) لن يتم إضافة أي فراغ (إذا كان هناك فراغ فيتم حذفه). ويمثل الشكل(9) المخطط الانسيابي لعملية الاخفاء.

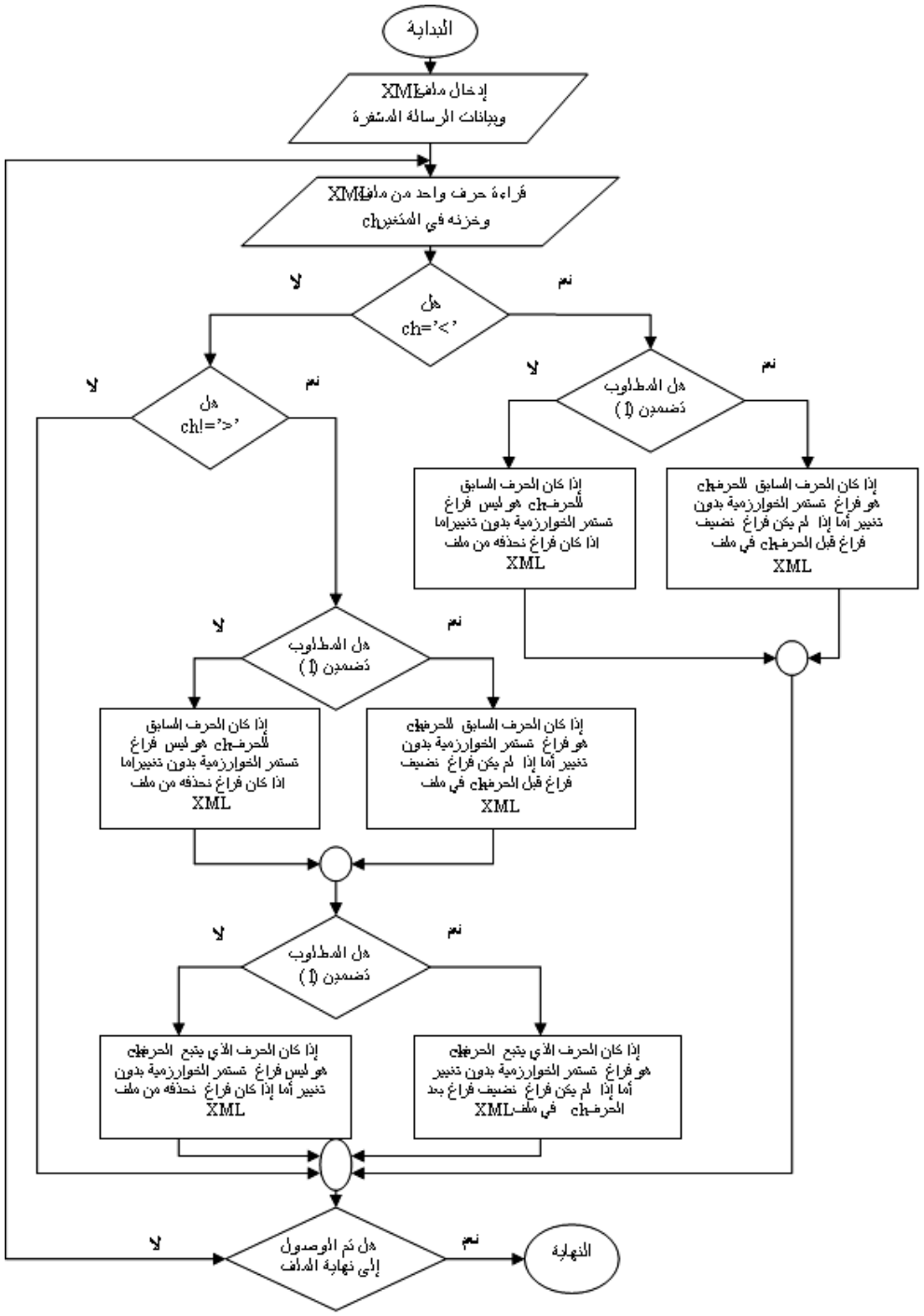
4.3 عملية الاسترجاع من صفحة XML :

في هذه المرحلة يتم أولا استرجاع مفتاح التشفير من ملف التعريف DTD، ثم استرجاع البيانات المشفرة من صفحة XML وبعد ذلك يتم تطبيق نفس خوارزمية التشفير لفك الشفرة للحصول على النص الصريح. ويوضح الشكل (10) عملية الاسترجاع من صفحة XML.

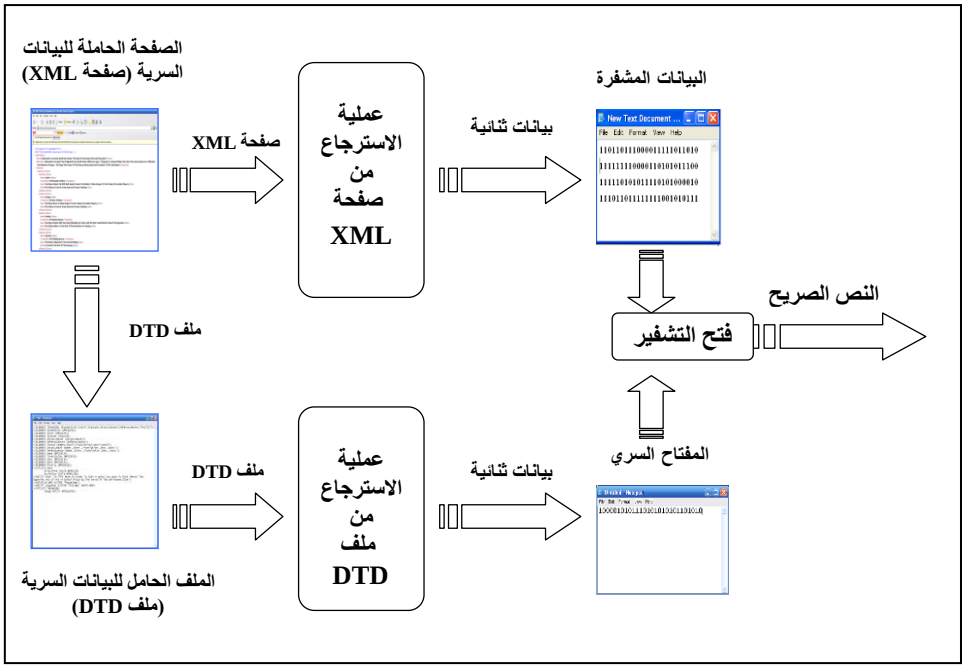
في الفقرات التالية سيتم شرح كيفية استرجاع مفتاح وبيانات سرية في ملفات صفحة XML إضافة إلى طريقة تشفير هذه البيانات.

1.4.3 استرجاع مفتاح التشفير من ملف DTD :

يتم اخذ ملف تعريف الصفحة والبحث عن الوسم المفتاح فيه، وبعد العثور عليه سيتم أخذه واخذ الوسم الرئيس له إضافة إلى موقعه في الوسم الرئيس (لكي يتم تحديده بشكل دقيق وللتغلب على مشكلة استخدام كلمات مشابهة للوسم المفتاح) ثم يتم البحث عنه في صفحة XML، فإن وجد الوسم المفتاح في صفحة XML يتم استرجاع البيانات من ملف DTD على أساس إن (1) ممثل بعلامة (+) وان (0) ممثل بعلامة (*). إما إذا لم يوجد الوسم المفتاح فسيتم اعتبار إن (1) يمثل بعلامة (*) وان (0) يمثل بعلامة (+) وعلى أساس التمثيل المختار سيتم استرجاع المفتاح السري (مفتاح التشفير). والشكل (11) يمثل المخطط الانسيابي لخوارزمية استرجاع مفتاح التشفير من ملف DTD.



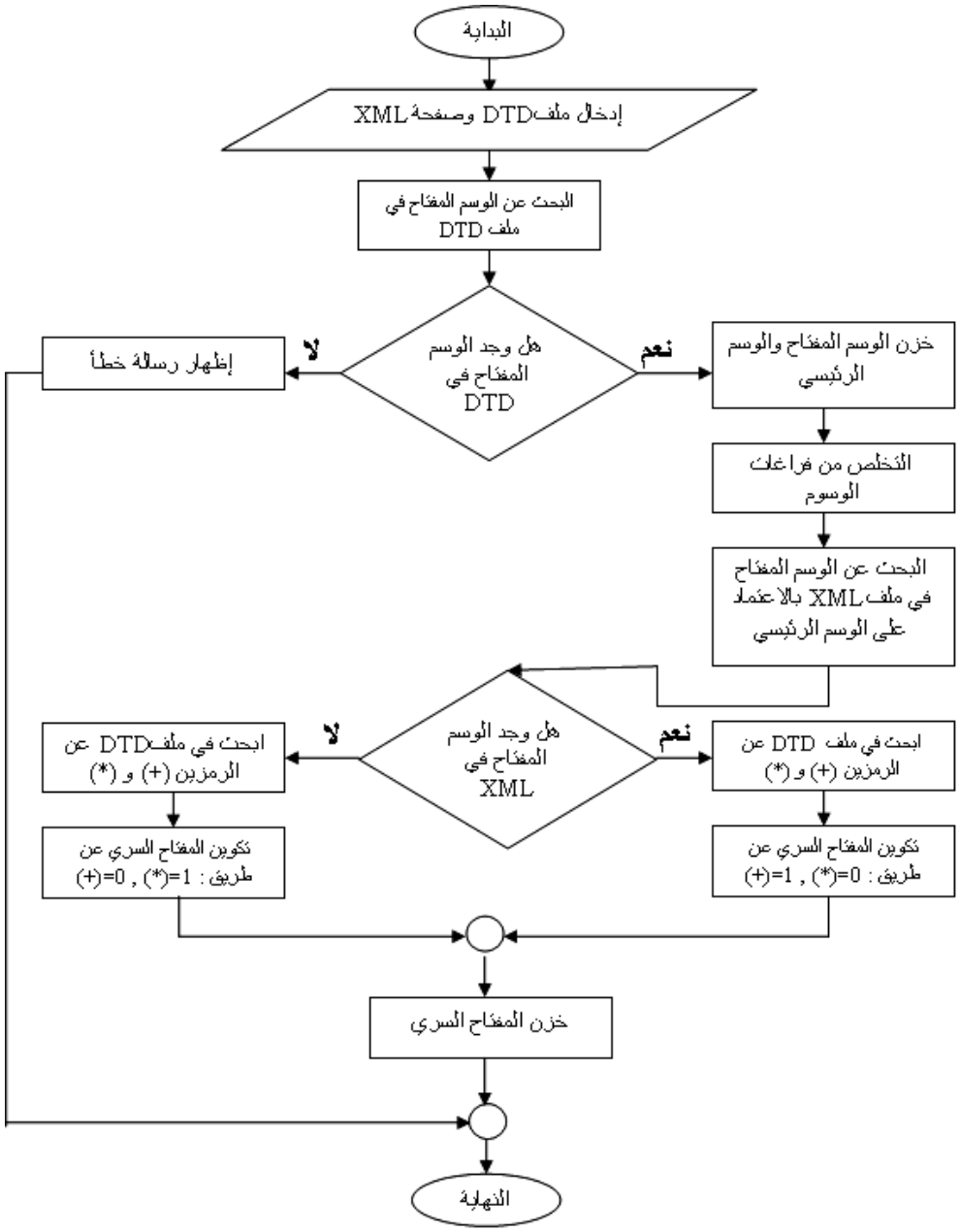
شكل (9) المخطط الانسيابي لعملية التضمين في صفحة XML



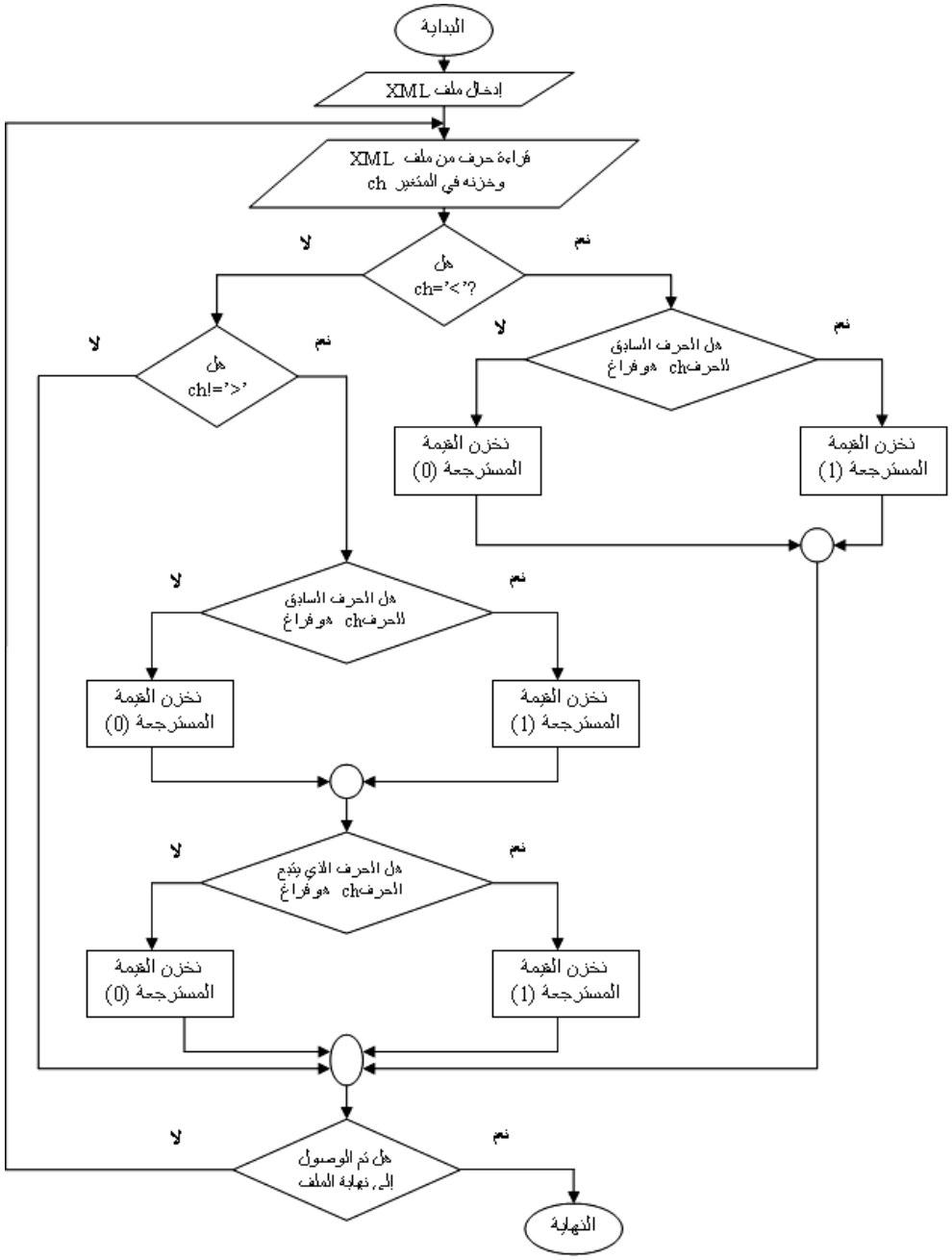
شكل (10) عملية الاسترجاع من صفحة XML

2.4.3 استرجاع الرسالة المشفرة من صفحة XML :

في الاسترجاع يتم البحث عن الرمزين (< , >) فعند العثور على (>) يتم البحث عن وجود فراغ قبله فإذا وجد يتم استرجاع (1) وإذا لم يوجد يتم استرجاع (0)، ونفس المعالجة تتم مع (<) مع فرق انه يتم البحث عن وجود فراغ قبله وكذلك بعده. ويمثل الشكل(12) المخطط الانسيابي لعملية الاسترجاع.



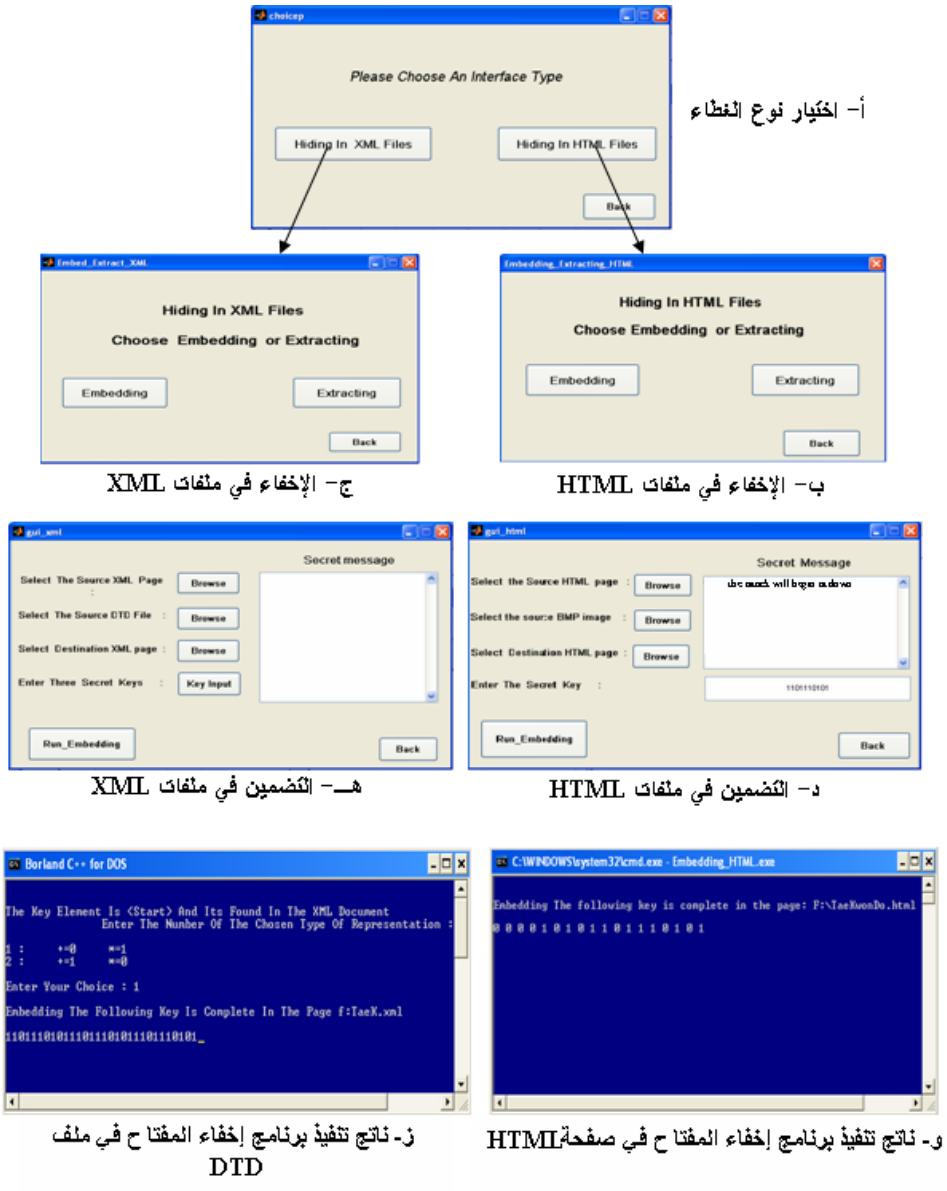
شكل (11) المخطط الانسيابي لاسترجاع مفتاح التشفير من ملف DTD



شكل (12) المخطط الانسيابي لعملية الاسترجاع من فراغات الوسوم

4. النتائج

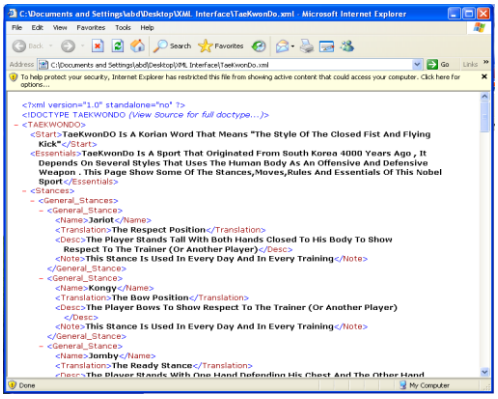
تم تصميم وتنفيذ نظام الاخفاء باستخدام واجهة المستخدم الصورية GUI المدعومة من قبل Matlab لتنفيذ خوارزميات الاخفاء المقترحة. يمثل الشكل (13) الواجهات المستخدمة في النظام.



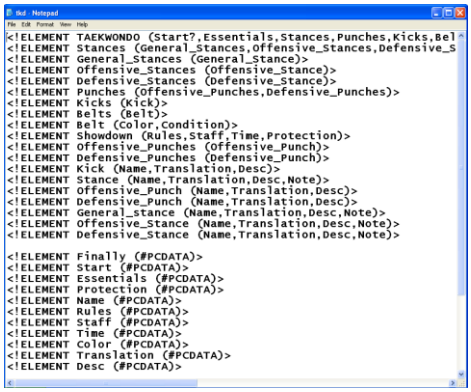
شكل (13) الواجهات المستخدمة في النظام

حـ – صفحة**HTML** الحاملة للمفتاح السري

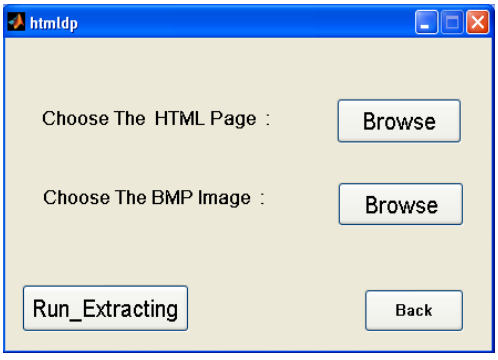
ط– الصورة الحاملة للرسالة المشفرة



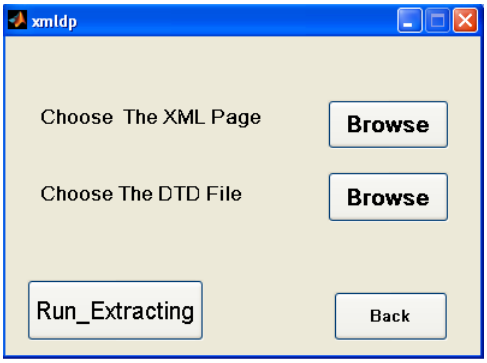
ي– صفحة **XML** الحاملة للرسالة المشفرة



ك– ملف DTD الحامل للمفتاح السري

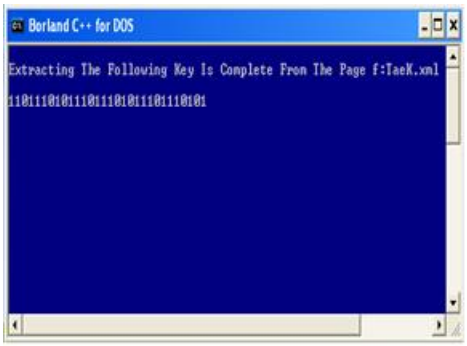


ل– الاسترجاع من ملفات HTML

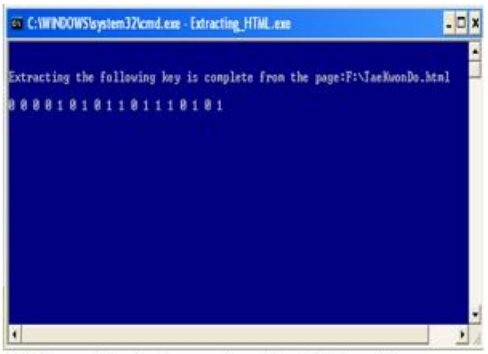


م– الاسترجاع من ملفات XML

شكل (13) الواجهات المستخدمة في النظام (تكملة)



س- ناتج تنفيذ برنامج استرجاع المفتاح ملف DTD



ن- ناتج تنفيذ برنامج استرجاع المفتاح صفحة HTML

ت - الرسالة المسترجعة

شكل (13) الواجهات المستخدمة في النظام (تكملة)

5 . الاستنتاجات والتوصيات :

- أثبتت النتائج العملية كفاءة الخوارزميات المقترحة من ناحية ان المعلومات المخفية لم تحدث أي تغيير او تشوه على الملفات الغطاء المستخدمة حيث تم قياس مدى صلاحية ووضوحية الصورة الناتجة بعد التضمين فكانت قيمة Peak Signal to Noise Ratio (PSNR) تساوي 41.63 dB.
- تتميز خوارزميات الإخفاء المستخدمة بالأمنية العالية حيث انه من الصعب جدا الاستدلال على وجود أي معلومات مخفية وكشفها.
- وفرت الخوارزميات المقترحة امكانية اخفاء كمية كبيرة من البيانات في الصورة الموجودة ضمن صفحة HTML او في الصيغة النصية للبيانات في ملف XML.
- يتميز نظام الاخفاء المستخدم بالمرونة العالية وسهولة الاستخدام.

فيما يخص الاعمال المستقبلية المقترحة للإخفاء في ملفات XML فنوصي بالاتي :

- ان يتم اسخدام ملف CSS لتنسيق شكل الصفحة والكتابة والألوان وان يتم إخفاء بيانات سرية في ملف CSS.
- أن يتم استخدام ملفات XSL, CSS, DTD وXML ولغة Java Script لتحسين صفحة XML عن طريق إضافة الوصلات التشعبية وعرض الصور والفيديوات وتصميم وتنفيذ عمل كل وسم.

المصادر

[1] Cole E., Krutz R., Conley J., (2005), "Network Security", Wiley Publishing, Inc.

[2] Deitel, H. M, (2001), "XML How To Program", Prentice Hall.

[3] Holzner, Steven, (2001), "Inside XML", New Riders Publishing.

[4] John, Corinna, (2008) Steganography 13- Hiding Binary Data in HTML Documents, The Code Project.

[5] Pence, James. H, (2001), "How To Do Every Thing With HTML", McGraw Hills Companies.

[6] Petitcolas F., Anderson R., Kuhn M., (1999) , " Information Hiding A Survey", Proceedings of the IEEE, special issue on protection of multimedia content.

[7] Stallings W., (1999), "Cryptography and Network Security", Prentice Hall Inc.

[8] The Official Web Site Of XML And HTML, "http://www.w3c.com".

[9] Wahlin, Dan, (2002), "XML For ASP .NET Developers", Sams Publishing.

[10] الحمامي، علاء حسين و الحمامي، محمد علاء، (2008)، "إخفاء المعلومات : الكتابة المخفية والعلامة المائية"، إثراء للنشر والتوزيع.

[11] العيسى، سميح يوسف، (2007)، "تنفيذ وبرمجة واجهات المستخدم الرسومية GUI في Matlab"، شعاع للنشر والعلوم .