Hiding in Text using Information Integrity Service

Dujan B. Taha *dujan_taha @uomosul.edu.iq*
Ahmed S. Nori *ahmed.s.nori @uomosul.edu.iq*
Yaseen H. Ismaiel *yaseen-hikmat @uomosul.edu.iq*
College of Computer Science and Mathematics
University of Mosul, Iraq

Received on:24/11/2008 **Accepted on:04/12/2008**

ABSTRACT

Modern computer networks make it possible to distribute documents quickly and economically. This is because of the decreasing cost of the equipment needed to ***copy***, ***print***, ***process the information***. The widespread adoption of electronic distribution of copyrighted material is accompanied by illicit copying and illicit distribution. This is why people used steganography. Steganography is the art of hiding transmitting data through apparently innocuous carriers in an effort to conceal the existence of the data.

The carrier can be anything used to transfer information, including for example: ***wood*** or ***state tablets***, ***hollow heels***, ***images under stamps***, ***tiny photographs***, or ***word arrangements***. Digital carriers include: ***disk space***, ***disk partitions***, ***text***, ***e-mail***, ***audio***, ***images***, and ***video***.

In this paper, we propose a new conceptual framework for text file steganography by using integrity tool. Hash function was used to produce a checksum.
Keywords: Steganography, Hash function.

الإخفاء في النص باستخدام ميزة تكامل البيانات

دجان بشير طه **أحمد سامي نوري** **ياسين حكمت اسماعيل**
كلية علوم الحاسبات والرياضيات / جامعة الموصل

تاريخ الاستلام: 2008/11/24 **تاريخ القبول: 2008/12/04**

الملخص

لقد مكنت شبكات الحاسوب الحديثة من توزيع الوثائق بصورة سريعة و اقتصادية . يعزى هذا لقلة كلفة الأجهزة اللازمة لنسخ وطباعة ومعالجة المعلومات . رافق هذا الانتشار السريع لتوزيع الوثائق الالكترونية النسخ غير المشروع والتوزيع غير المشروع . وهذا هو سبب استخدام الستيكانوكرافي من قبل الناس . الستيكانوكرافي هو فن اخفاء المعلومات المنقولة خلال وسط ناقل علني غير مؤذي كمحاولة لإخفاء وجوده . يمكن أن يكون الناقل أي شئ يستخدم لنقل المعلومات متضمنا كمثال خشب أو لوحة أقراص كعوب أحذية مجوفة ، صور مطبوعة ، صور صغيرة جدا أو ترتيبات كلمة . تتضمن النواقل الرقمية البريد الالكتروني ، الصوت و الفيديو ، النص ، الفراغات الموجودة على القرص المغناطيسي ، أجزاء القرص والصور .

في هذا البحث استخدمت هيكلية جديدة لإخفاء ملف نصي باستخدام ميزة سلامة البيانات (integrity) . تم استخدام الدالة الهاشية (hash function) للحصول على قيمة فحص المجموع.

الكلمات المفتاحية: الاخفاء، الدالة الهاشية.

1. المقدمة Introduction :

ستيكانوكرافي (حرفيا الكتابة المغطاة) هي إخفاء الرسائل السرية ضمن رسالة أخرى تبدو غير مؤذية أو ناقل (Carrier) يمكن أن يكون الناقل أي شيء يستخدم لنقل المعلومات ، متضمنا مثلا خشب أو لوحة أقراص ، كعوب أحذية مجوفة ، صور مطبوعة ، صور صغيرة جدا أو ترتيبات كلمة .تتضمن النواقل الرقمية البريد الالكتروني (E-MAIL) الصوت (Audio) ، والرسائل الفيديوية ، الفراغات الموجودة على القرص المغناطيسي ، أجزاء القرص المغناطيسي والصور .

الستيكانوكرافي تشبه التشفير لأنها وسائل تؤمن السرية . تؤمن الستيكانوكرافي ذلك من خلال إخفاء وجود الاتصال ، بينما يفعل التشفير ذلك من خلال ترميز الرسالة حتى لا يمكن فهمها . يمكن مقاطعة الرسالة المشفرة من قبل المتصنت لكن المتصنت قد لا يعرف حتى بوجود رسالة الستيكانوكرافي . إن هدف الستيكانوكرافي هو تجنب جلب الشكوك إلى تراسل الرسالة السرية . كشف الرسالة السرية هو هجوم ضد الستيكانوكرافي يعتمد على حقيقة أن إخفاء معلومات في وسط رقمي يحور الناقل ويقدم خصائص غير اعتيادية من خلال البيانات المتضمنة قد تكون المفتاح لمثل هذا الهجوم . الهجوم الناجح ضد العلامة المائية الرقمية ، من ناحية أخرى تسترجع علامة مائية غير مفيدة أو غير مقرؤة [1] .

واحدة من أوائل المستندات التي توصف الستيكانوكرافي هي من إلياذة هيرودتس . في اليونان قديما يكتب النص على ألواح مغطاة بالشمع . طريقة مبدعة أخرى هي بحلق راس الرسول وكتابة رسالة أو طبع صورة على راس الرسول . بعد السماح للشعر بالنمو فان الرسالة لايمكن كشفها إلا بعد حلق راس الرسول مرة أخرى . صيغة عامة أخرى في الكتابة المخفية هي من خلال استخدام الأحبار المخفية . استخدم مثل هذه الأحبار بنجاح خلال الحربين العالميتين الأولى والثانية . قد تحتوي رسالة بريئة على رسالة مختلفة جدا مكتوبة بين السطور . قديما في تقنية السيتكانوكرافي في الحرب العالمية الثانية كانت تتألف حصرا من الأحبار المخفية . المصادر العامة للأحبار المخفية هي الحليب ، الخل ، عصير الفواكه ، والإدرار (Urine) ، جميعها يغمق لونها عندما تسخن . مع تطور التقنية وسهولة فتح ترميز هذه الأحبار المخفية فقد تم تطوير أحبار أكثر تطورا والتي تتفاعل مع مواد كيمياوية متنوعة تم تطوير بعض الرسائل لتكون مشابهة للصور فيمكن إظهارها باستخدام عدد من المواد الكيمياوية في مختبر المعالجة [2] .

شفرات القيم اللاغية Null Ciphers (رسائل غير مشفرة وتعرف أيضا بالرموز المفتوحة) قد تم استخدامها أيضا حيث تخفى الرسالة الحقيقية في رسالة تظهر أنها بريئة . بسبب ظهور العديد من رسائل الرموز المفتوحة ، فان منقيات البريد اكتشفت الاتصالات المشكوك بها ، على كل حال الرسائل التي تظهر أنها بريئة سمح لها بالمرور . كمثال على رسالة مكتوبة على شفرة القيمة اللاغية والتي أرسلت حقيقة من قبل جاسوس ألماني في الحرب العالمية الثانية :

Apparently neutralu s protest is thoroughly discounted and ignored Isman hard hit.Blockade issue affects pretext for embargo on Byproducts,ejecting suets and vegetable oils.

بأخذ الحرف الثاني من كل كلمة فتظهر الرسالة المخفية التالية :

Pershhng sails from NYjune 1.

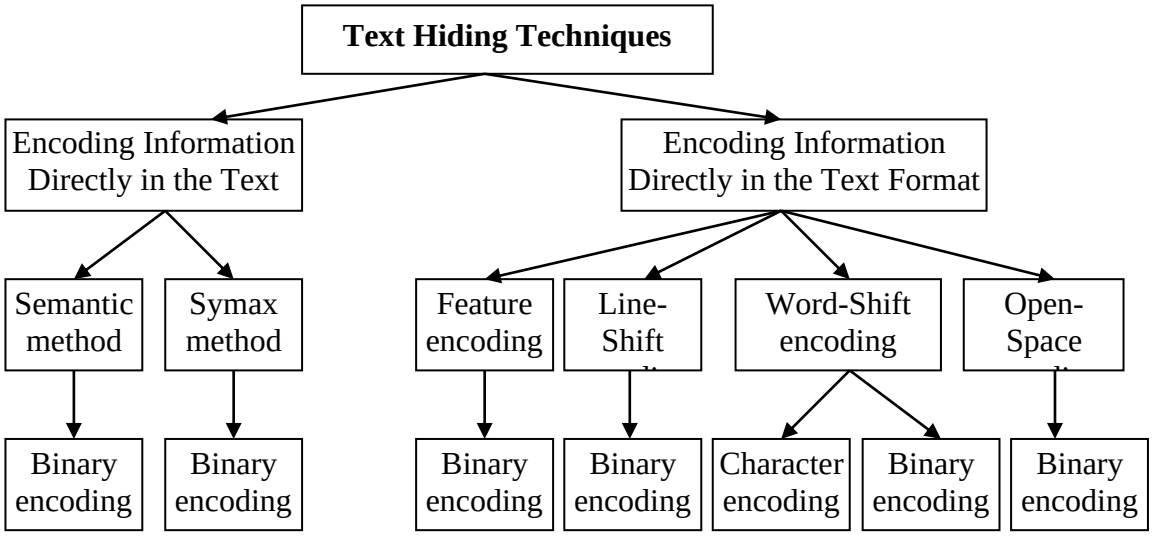
مثلما تطور كشف الرسائل فقد تم تطوير تقنيات جديدة تستطيع امرار معلومات أكثر وتكون حتى اقل ظهورا. مع طرق عديدة تم اكتشافها ومقاطعتها فقد اتخذ مكتب الرقابة الأمريكي إجراءات نهائية مثل منع تسليم الورد ، الذي يحتوي تواريخ تسليم أحجية الكلمات المتقاطعة وحتى بطاقات التقرير لأنها يمكن أن تحتوي على رسائل سرية .لقد ذهبت إجراءات المنع إلى مدى ابعد مثل إعادة صياغة الرسائل واستبدال الطوابع على ظروف الرسائل [1,3] .

تم في هذا البحث استخدام تقنية أخفاء نص في نص أخر باستخدام طريقة مقترحة تعتمد أسلوب أظهار الرسائل المتضمنة و الوسط الناقل (carrier) سوية . تختلف هذه الطريقة عن طرق أخفاء النص السابقة وطريقة الشفرة اللاغية المستخدمة سابقا . اعتمدت الطريقة المقترحة ميزة سلامة البيانات (integrity) في عملية التضمين والاسترجاع . استخدمت الدالة الهاشية (function hash) لتطبيق هذه الميزة .

2. تقنيات الإخفاء في النص :

يكون عادة من السهل جدا معرفة النسخ الأصلية للكتب من الكتب المستنسخة لان النوعية تكون مختلفة تماماً . تصبح هذه المعرفة أكثر صعوبة عندما نتعامل مع النسخ الالكترونية للنصوص . حيث تكون النسخ متطابقة ومن المستحيل معرفة ما إذا كانت النسخ أصلية أو مستنسخة . فلتضمين معلومات ، مثل العلامة المائية، داخل مستند تستطيع بكل سهولة تحوير بعض خصائص المستند. يمكن أن تكون هذه تشكيل النص أو خصائص الرموز نفسها . قد يظن الشخص بأن تحوير هذه الخصائص يصبح مرئي وواضح إلى أشخاص آخرين أو مهاجمين . إن المفتاح لهذه المشكلة هو أننا نحور المستند بطريقة تكون غير مرئية إلى عين الإنسان ورغم ذلك يكون بامكان الحاسوب أن يفتح رموزها [4] .

ترمز تقنيات الستيكانوكرافي المعلومات بطريقتين رئيسيتين وكما موضحة في الشكل[1] .



شكل (1) تصنيف تقنيات إخفاء النص

*** يمكن توضيح طرق ترميز المعلومات لتحقيق الستيكانوكرافي والموضحة بالشكل [1] بالنقاط التالية : [1,3]**

أ. طريقة النحو Syntax Method : تستخدم هذه الطريقة الترقين Punctuation والإلقاء Diction وتركيب النص بدون تغيير المعنى بصورة ملحوظة .

ب. طريقة دراسة معاني الكلمات Semantic Method : تتضمن هذه الطريقة تغيير الكلمات نفسها و انها تخصص قيمتين مترادفة رئيسية او ثانوية .

ج. طريقة الفراغ المفتوح Open Space Method : تدل هذه الطريقة على الترميز خلال معالجة الفراغ الأبيض (فراغ غير مستعمل) على الصفحة المطبوعة .

ء. ترميز إزاحة السطر Line-Shift Coding : في هذه الطريقة ، تزاح اسطر النص عموديا (تنقل إلى الأعلى أو الأسفل) حسب بتات الرسالة السرية ، بينما تبقى الأسطر الأخرى ثابتة من اجل غاية التزامن . في التنفيذ الأصلي ، ترسل المعلومات في كل سطر ثاني ، وذلك بتحريك السطر الثاني إما 1/300 من الانج إلى الأعلى أو الأسفل والذي لا تلاحظه العين البشرية. إذا نقل السطر إلى الأعلى، يرمز واحد (1) وإذا تحرك نحو الأسفل يرمز صفر (0) . لا يكتشف هذا التنقل من قبل العين البشرية ، لكنه يكشف من قبل الحاسوب عندما يقيس المسافة بين كل سطر من الأسطر .

هـ. ترميز إزاحة–الكلمة Word-Shift Coding : في هذه الطريقة ترمز الكلمات الرمزية في المستند من خلال إزاحة المواقع الأفقية أو العمودية للكلمات ضمن اسطر النص ، بينما يدام ظهور الفراغات الطبيعية . تكون هذه الطريقة مطبقة فقط على مستندات ذات فراغات متغيرة بين الكلمات المتجاورة . كنتيجة لهذه الفراغات المتغيرة ، يكون من الضروري أن تحتفظ بالصورة الأصلية ، أو على الأقل نعرف الفراغ بين كلمات المستند غير المرمز .

و. طريقة ترميز الصفة Feature Coding Method : في ترميز الصفة ، تحور صفة نص معينة ، أو لا تحور اعتمادا على كلمات الترميز. مثلا ، يستطيع شخص ترميز بتات في النص من خلال توسيع أو تقليص خطوط النهايات العمودية أو العليا للحروف مثل : b, d, h... الخ . بصورة عامة قبل الترميز ، تحدث عشوائية للصفة . هكذا ، أطوال نهاية الخط للرمز تطول أو تقصر عشوائيا ، بعد ذلك تحور مرة ثانية لترميز البيانات المحددة . يزيل هذا الاحتمالية للترميز المرئي ، كما أن أطوال نهاية الخط الأصلية ستكون غير معروفة . وهنا يحتاج المرء للترميز إلى الصورة الأصلية .

3. خوارزمية ملخص الرسالة MD5 :

تم تطوير خوارزمية ملخص الرسالة MD5 من قبل رون رايفست Ron Rivest وهو احد مطوري خوارزمية التشفير غير المتناظر RSA . كانت خوارزمية MD5 هي من أكثر خوارزميات الهاش الأمنية المستخدمة إلى قبل سنين قليلة حيث ظهر الاهتمام بتحليل الشفرة وخاصة هجوم القوة الوحشية (Brute-Force) . تأخذ الخوارزمية إدخال الرسالة ذات الطول المختلف وتنتج كإخراج ملخص رسالة ذو طول 128 بت. تتم عملية معالجة الإدخال على شكل كتل ، يكون حجم الكتلة الواحدة 512 بت [6] .

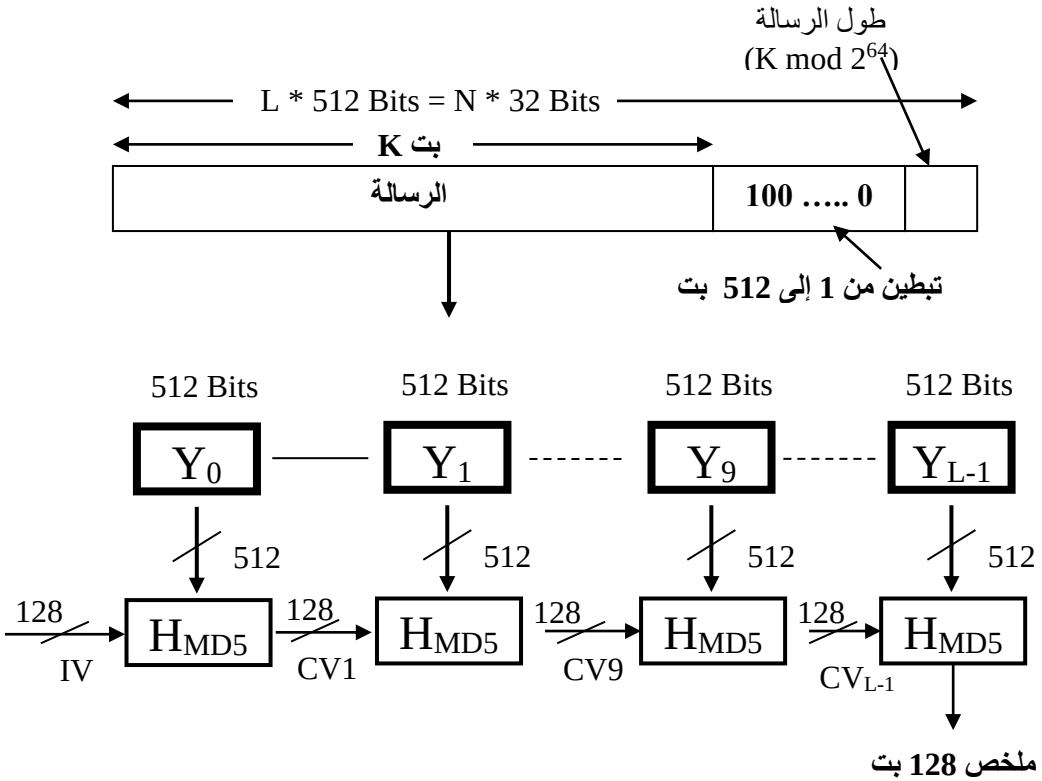
يوضح الشكل (2) المعالجة الكاملة لرسالة لإنتاج الملخص (digest) . يتبع هذا الهيكلة العامة .

تتألف المعالجة من الخطوات التالية :

1. إلحاق البتات المبطنة : يتم تبطين الرسالة حتى يكون طولها بالبتات محول إلى (الطول=448 mod 512) هكذا ، يكون طول الرسالة المبطنة هو 64 بت اقل من العدد الذي يكون مكرر إلى 512 بت . ويضاف التبطين دائماً، حتى إذا كانت الرسالة هي في الطول المطلوب . مثلاً ، إذا كان طول الرسالة هو 448 بت ، فإنها تبطن بـ 512 بت إلى طول 960 بت . وهكذا ، يكون عدد البتات المبطنة هو في مدى 1 إلى 512. يتكون التبطين من بت 1 منفردة متبوعة بالعدد الضروري من 0 بت .

2. طول التبطين : 64 بت تمثل الطول للرسالة الأصلية (قبل التبطين) يتم إضافتها إلى نتيجة الخطوة الأولى [1] (البايت الأقل أهمية أولا). إذا كان الطول الأصلي هو اكبر من 264 ، فيستعمل فقط 64 بت الأقل أهمية من الطول. هكذا، يحتوي الحقل على طول الرسالة الأصلية موديولو 264 .

تؤدي النتيجة للخطوات (1) و (2) إلى رسالة يكون طولها أعداد مضروبة إلى 512 بت . في الشكل (2) تمثل الرسالة الموسعة على شكل سلسلة من كتل 512 بت Y0' Y1' YL-1 حتى يكون الطول الكلي للرسالة (512*L) بت .



شكل(2) توليد ملخص رسالة باستخدام MD5

3. إنشاء المساحة الخزنية MD. مساحة خزنية طولها 128 بت تستعمل لخزن النتائج الوسطية والنهائية لدالة الهاش . يمكن تمثيل هذه المساحة الخزنية كمسجلات ذات طول 32 بت (D،C،B،A).

4. معالجة الرسالة على شكل كتل ذات حجم 512 بت (16 كلمة) . إن قلب الخوارزمية هو فعالية الكبس (Compression) التي تتألف من أربعة جولات من المعالجة . هذا الجزء مؤشر في الشكل (2) على شكل HMD5 .

5. بعد معالجة جميع L ذات 512 بت في الكتلة ، فأن الناتج من مرحلة Lth هي ملخص الرسالة ذو طول 128 بت .

تكمن قوة MD5 بأنه يمتلك خاصية أن كل بت في رمز الهاش هي دالة لكل بت في الإدخال. إن الإعادة المعقدة للوظائف الأساسية (I، H، G،F) هي جولات أربعة لها نفس الهيكلة ولكن كل واحدة تستخدم دالة منطقية أساسية مختلفة . تأخذ كل جولة كإدخال الكتلة الحالية (512 بت) المعالجة من قبل y وقيمة 128 بت في المساحة البينية وتحديث محتويات المساحة البينية . تنتج نتائج ممزوجة بصورة جيدة ، لذلك من غير الممكن اختيار رسالتين عشوائياً ، حتى وان أظهرتا تنظيم متشابه ، ويكون لهما نفس الرمز الهاشي .

4. الطريقة المقترحة :

قدمت هذه الطريقة مسار جديد في الستيكانوكرافي . حيث استخدمت النص كناقل لنص آخر سري باستخدام ميزة فحص المجموع (checksum) . تم إطلاق مصطلح الغربلة (Winnowing) على هذه الطريقة، حيث أنها تغربل النص المتضمن السري عن النص المزيف بالاعتماد على قيمة فحص المجموع الصحيحة . استخدمت الدالة الهاشية (hash function) لإنتاج قيمة فحص المجموع .

أ. خوارزمية التضمين :

*تقطيع النص السري إلى حزم متعددة طولها و عددها يعتمد على المستخدم .

*تطبيق خوارزمية MD5 على كل حزمة لإنتاج قيمة فحص المجموع .

*تقطيع النص المزيف (الناقل) إلى حزم متعددة أيضا بنفس عدد حزم النص السري .

*إعطاء قيم MD5 غير صحيحة لكل حزمة .

*دمج النص السري مع النص المزيف وإرسال الرسالة .

ب. خوارزمية الاسترجاع :

*أخذ الحزمة الأولى وحساب قيمة MD5 .

*اذا كانت القيمة صحيحة فالحزمة تابعة للنص السري وألا فهي للنص المزيف .

*تكرار الخطوات 1 و 2 لجميع الحزم .

<u>مثال تطبيقي :</u>

أفترض أن أحمد يريدان يرسل الرسالة النصية التالية الى علي :

"Hi Ali, I'II meet you at 5:00 PM., your friend Ahmed"

أولا يجب تقطيع الرسالة إلى حزم وإضافة قيمة فحص المجموع لكل حزمة . لو فرضنا بأن الرسالة ستقطع إلى أربع حزم مع إضافة قيمة 6 أرقام كناتج للدالة الهاشية .

(1, "Hi Ali", 498253)
(2, "I 'II meet you", 390024)
(3, "at 5:00 PM" , 759241)
(4, "Your friend Ahmed", 258133).

في الخطوة الثانية يتم اختيار الوسط الناقل لها وهو الرسالة المزيفة التي تبدو كالتالي :

"Hi Suzan, I ' II call you tomorrow, Your brother Ahmed"

تقطع الرسالة الناقلة إلى حزم أيضا وتعطى قيم خاطئة كناتج الدالة الهاشية . تضمن الرسالة الأصلية بها حيث تضاف كل حزمة مزيفة قبل أو بعد الحزم الأصلية . يظهر الشكل النهائي لها كالتالي :

(1, " Hi Ali", 498253)
(1, " Hi Suzan", 57801)
(2, " I' II call you", 533966)
(2, " I' II meet you", 390024)
(3, " at 5:00 PM", 759241)
(3, " tomorrow", 707224)
(4, " your brother Ahmed", 539421)
(4, " your friend Ahmed" 258133)

5 . المناقشة والاستنتاجات :

قدمت هذه الطريقة مسار جديد للستيكانوكرافي . تستخدم الطريقة المقترحة وسط نصي لتضمين رسالة نصية باستخدام الطريقة الدالة الهاشية للحصول على قيمة البصمة أو فحص المجموع و امتازت بما يلي :

نسبة التضمين (data rate) عالية جدا وتعتمد على حجم الغطاء (الوسط النصي الناقل) فكلما كبر حجم الغطاء كلما زادت كمية المعلومات المتضمنة أي ممكن تصل إلى نسبة 100% .

يمكن خزن الرسالة النصية بصيغ مختلفة دون أي تدمير مؤذي لها , لان عملية التضمين لا تعتمد على شكل الرسالة أو صيغتها .

إن هدف الكتابة المخفية هو تجنب جلب الشك إلى تراسل الرسالة المخفية . لذلك يبقى غير مكتشف . إذا أرتفع الشك يصبح الهدف فاشل . الهجمات والتحليل على النص تأخذ المراحل التالية :

الكشف Detection وهو عادة الخطوة الأولى وهي فقط معرفة أن احد الأشخاص يستخدم قناة اتصال مخفية . في الطريقة المقترحة الكشف سوف يكون صعب جدا لان الناقل هو نص واضح و مفهوم ولا يدعو للشك .

الاستخلاص Extraction وهي الخطوة التالية في تحليل النص المخفي والتي تفتح ترميز الملف المكتشف امعرفة ما مخفي بداخله. في الطريقة المقترحة عملية الاستخلاص غير ممكنة لان الطريقة تفترض بان خوارزمية فحص المجموع المستخدمة تبقى سرية بين الباعث والمستلم .

التشويش Confusion أو إعادة الكتابة Overwriting أو تضمين معلومات مزيفة Counterfit Information . وفي جميع الحالات سوف يتم كشف التلاعب عند المستلم لان مبدأ الإخفاء أصلا يعتمد على فكرة فحص المجموع أو الدالة الهاشية والتي من مميزاتها كشف أي تحوير أو تزييف في الرسالة .

التدمير Destruction هي الخطوة الأخيرة وهي الأسهل . إذا لم يستطع المهاجم إثبات نظريته بان هنالك رسالة مخفية في الغطاء فمعنى ذلك أن النظام نظريا يثبت بأنه أمين .

كتوصية لتطوير الطريقة ، استخدام الدالة الهاشية ذات المفتاح السري واعتماد سرية المفتاح كوسيلة للتضمين .

المصادر

[1] Al_hamami A.H,information hiding steganography and watermark , university bookshop , alsharaqa,2008 .

[2] doralhy e. denning “information warfare and security”, Addison_wesley , 1999.

[3] kazaenbeisser s.,and peticolas f.,”inforation hiding techniques for steganography and digital watermarking “ ,artech house, 2000.

[4] Neil F. jonson , zoran duric and Sushil Adjoin , “Information Hiding : Steganography and watermarking Attacks and Countermeasures” , Kluwer Acadmic publishers , 2001 .

[5] Preneel B. , “ the state of cryptographic hash function “ , proceeding , eurocryp’ab , 1996 , new york : springer – verlag .

[6] R. , Rivest , 1992 , “ the MD5 message Digest Algorithm “ , Internet Engineering task force , RFC 1321 .