Generating Spanish stegotext for fun and profit.

Alfonso Muñoz Muñoz.
Departamento de Ingeniería y Arquitecturas Telemáticas. DIATEL
Universidad Politécnica de Madrid. E.U.I.T Telecomunicación
Carretera de Valencia Km.7 – 28031. Madrid. España
amunoz@diatel.upm.es

Resumen- La generación automática de estegotextos es una rama de investigación dentro de la esteganografía lingüística con un enorme potencial. Por desgracia, en la actualidad existen pocas investigaciones públicas sobre el potencial de este tipo de esteganografía en lengua española. En este artículo se demuestra la utilidad de utilizar una variante mejorada del algoritmo de imitado de Peter Wayner, así como se introduce el concepto de plantilla para mejorar posibles errores gramaticales en la creación automática de estegotextos en lenguaje natural. Este tipo de procedimientos facilitaría la creación de comunicaciones subliminales robustas al resistir ataques estadísticos, gramaticales y de coherencia. Se libera la herramienta Stelin que lleva a la práctica estos principios. http://stelin.sourceforge.net.

Palabras Clave- Esteganografía lingüística, estegotextos, generación automática, Stelin, redes sociales, Peter Wayner.

I. INTRODUCCION A LA ESTEGANOGRAFIA. TRABAJOS PREVIOS.

La esteganografía es la ciencia y el arte de ocultar una información dentro de otra, que haría la función de *tapadera o cubierta*, con la intención de que no se perciba ni siquiera la existencia de dicha información [1]. En teoría, sólo quienes conozcan cierta información acerca de esa ocultación (un secreto) estarían en condiciones de descubrirla. En criptografía no se oculta la existencia del mensaje sino que se hace ilegible para quien no esté al tanto de un determinado secreto (la clave). Por este motivo, los mensajes que se procuran ocultar usando técnicas esteganográficas, habitualmente, son previamente cifrados. La ocultación de mensajes usando procedimientos esteganográficos puede tener fines legítimos o ilegítimos, que pueden ser beneficiosos para proteger la privacidad de las comunicaciones o burlar censuras, o, por el contrario, ser vehículos para perpetrar actos criminales. Por estos motivos, en la presente década se está realizando una inversión importante en la detección de comunicaciones ocultas.

El estegoanálisis es la ciencia y el arte que permite detectar (ataques pasivos) o anular (ataques activos) esa información oculta. En general, detectar si un estegomedio tiene información oculta y cuantificarla. La extracción y recuperación de la información real enmascarada dependerá de las medidas de protección empleadas. Esta difícil tarea pertenece a la ciencia del criptoanálisis.

En la última década los avances en ocultación y detección han sido notorios, en muchos casos imitando diseños o características de procedimientos criptográficos. Es significativo como actualmente las herramientas esteganográficas que "pretenden" ser más seguras publican sus algoritmos y su robustez depende exclusivamente de una clave externa conocida por emisor-receptor [2].

La presente década ha visto como se han publicado numerosos procedimientos de ocultación para medios muy diversos [2], siendo el principal protagonista el contenido multimedia, precisamente en este campo es donde más ha avanzado la detección de comunicaciones. Numerosos algoritmos estadísticos (estegoanalíticos) se han publicado (chi-square, RS, simple-pair…), así como avances en estegoanálisis a ciegas[1]. Por otro lado, las técnicas de ocultación han ido evolucionado paralelamente. Un ejemplo significativo son las técnicas de modificación de imágenes (pixeles, índices a paleta de colores, DCTs, etc). Desde la inocente modificación de los bits de un pixel de forma secuencial a la modificación de los pixeles de manera aleatoria, la reducción del impacto en la imagen, la distribución multiportador, matrices de codificación, preselección de los pixeles a modificar mediante la utilización de los algoritmos estegoanalíticos publicados, etc. Sin duda, la clásica lucha entre "el gato y el ratón" [2].

Este artículo pretende por un "momento" olvidarse de todas esas investigaciones comentadas y enfocar el problema de la creación de canales subliminales utilizando un medio muy común, quizás el medio más común y antiguo de todos: el texto en lenguaje natural. Los motivos y las condiciones que se consideran son varias.

a) Seleccionar un medio (portador) muy común para ocultar información dificulta la tarea de un estegoanalista de separar "el grano de la paja". El texto está presente en todo, por ejemplo, las redes sociales son un buen ejemplo de ello.

b) Los procedimientos de ocultación en lenguaje natural deberían ser lo más automáticos posibles y no estar basados en oscuridad, su seguridad debería depender exclusivamente de una información secreta que poseyera emisor-receptor.

[1] Caracterización de un medio portador mediante la definición de características propias de cada medio consiguiendo diferenciar entre cubiertas originales y potenciales estegomedios. Para ello se utilizan clasificadores de todo tipo, SVM, Fisher, etc.

c) El mejor mensaje cifrado es aquel que no lo parece.

Cuando se utilizan procedimientos esteganográficos "clásicos" aplicados a contenido multimedia minimizar el efecto visual (o sonoro) es el primer paso. En la práctica los ataques estegoanalíticos van a la información que el usuario "no ve" y por tanto a simple vista no es capaz de evaluar la seguridad del estegomedio producido-modificado. La idea de ocultar información en texto en lenguaje natural es que cualquier persona sin necesidad de conocimientos matemáticos pueda evaluar, al menos en primera instancia, la calidad de la ocultación, sólo necesita saber leer. Es más, como se comentará, incluso la capacidad de escribir información en lenguaje natural manualmente podría hacer los estegotextos más robustos. La idea perseguida es clara, utilizar los conceptos de la ciencia de la esteganografía y la computación lingüística para generar estegotextos en lenguaje natural que permitan la transmisión de mensajes con una alta seguridad.

II. ESTEGANOGRAFIA LINGÜÍSTICA.

La utilidad de usar la información textual como un potencial estegomedio no es ni mucho menos nueva. A lo largo de los siglos se han documentado múltiples formas de ocultación en soportes y formas varias [3]: cartas, libros, telegramas, poesías, canciones, revistas, periódicos (por ejemplo, *newspaper code* en la época Victoriana o la verja de Cardano en el siglo XVI); o más recientemente en canales de mensajería instantánea (messenger, IRC), utilizando el "ruido" de las traducciones automáticas, etc [4]. Es común ver una clasificación clásica de estas técnicas en términos de *open codes y semagrams*, en terminología inglesa.

Los *open codes* genéricamente se refieren a textos de apariencia inocente, que ocultan información recuperable utilizando ciertas letras, palabras, frases del texto o comunicación (métodos basados en esto son: Cues, Null Ciphers, Jargon Code y Grilles), mientras que los *semagrams* son el conjunto de técnicas que consisten en la utilización (variación) de la estructura y formato de los elementos de un texto, modificaciones que aunque "visibles" no por ello son fáciles de detectar.

La ocultación de información en textos en lenguaje natural puede realizarse, al menos, de tres formas genéricas: ocultación basada en el orden, basada en la modificación de un texto existente y basada en la generación automática de un estegotexto.

Independientemente del mecanismo seleccionado la seguridad de los estegotextos debe ser analizada desde diferentes puntos de vista, considerando ataques lingüísticos (sintáctico, semántico y de coherencia) por parte de máquinas y analistas, así como ataques puramente estegoanalíticos y estadísticos (análisis de entropía, análisis de frecuencia de caracteres-palabras, ataques basados en conocimiento de cubierta original y cubierta modificada, etc). En la práctica, el lenguaje natural, como estegomedio, es muy poco redundante (ruidoso) en comparación con otros estegomedios como son las imágenes o los videos, lo cual hace más complicado la creación de algoritmos robustos de ocultación de información en lenguaje natural (información textual), lo que hace que las técnicas de ocultación requieran una gran cantidad de información textual para ocultar una cantidad de información no muy abultada. Este hecho puede hacer que estas técnicas solo sean interesantes en escenarios limitados.

Históricamente las técnicas más socorridas para ocultar información han sido las basadas en la modificación de textos existentes. Los problemas principales de estos mecanismos residen en su poca flexibilidad. La mayoría están basados en oscuridad (como puede ser la ocultación en una cierta posición de un texto, usando mayúsculas y minúsculas, etc) y sobre todo tienen el peligro de que un potencial estegoanalista pudiera conseguir el texto original sin modificar y realizar ataques clásicos de comparación estegotexto – texto original, lo que delataría fácilmente la presencia de información oculta. Ante este hecho, una solución posible es la realización de modificaciones sobre un estegotexto creado automáticamente a la medida (e incluso único por cada comunicación) o simplemente la ocultación de información en el propio estegotexto generado automáticamente. Las propuestas de generación automática de estegotextos en otras lenguas oscilan en procedimientos que imitan la gramática (sintaxis) y la imitación estadística de textos considerados como de entrenamiento.

En el presente artículo se profundiza en un procedimiento de generación automática de estegotextos en lengua española basada en imitación estadística (N-Gram) de uno o varios textos fuente de entrenamiento. El presente artículo muestra que es posible aunar conocimientos actuales de lingüística computacional con esteganografía lingüística para crear herramientas públicas de generación automática de estegotextos[1].

III. GENERACIÓN AUTOMÁTICA DE ESTEGOTEXTOS EN LENGUA ESPAÑOLA.

Peter Wayner en la década de los 90 publicó un procedimiento de generación automática de estegotextos (T) basado en el imitado estadístico de una o más fuentes de textos (S) que resulta interesante analizar [6]. La idea es sencilla:

"Cójase una función de imitado f que modifique un fichero A de forma que asuma las propiedades estadísticas de otro fichero B. Es decir, si p(t,A) es la probabilidad de que una cadena t suceda en A, entonces una función de imitado f, hace que la p(t,f(A)) sea aproximadamente p(t,B) para toda cadena t de tamaño menor que n".

La complejidad del modelo estadístico de imitado (análisis de frecuencia) depende precisamente del orden estadístico n (orden de complejidad del algoritmo). Según está idea, Wayner definió el siguiente algoritmo de imitado:

1. Constrúyase una lista de todas las diferentes combinaciones de n letras que ocurran en S (el/los texto/s de

[1] Si el lector desea conocer más acerca de otras líneas de investigación como las construcciones CFGs (Context-Free-Grammar) derivadas de los principios de la teoría de la gramática generativa propuesta por el lingüística A.Noam Chomsky en la década de los 60 puede ver a modo de introducción [5].

entrenamiento) y contabilícese el número de veces que ocurren en S.

2. Elegir una de ellas aleatoriamente que actuará de semilla inicial. Esto generará las primeras n letras de T (el estegotexto).

3. Repetir este punto hasta que se genere todo el estegotexto deseado:

a. Cójase las n-1 letras siguientes de T. Buscar en la tabla estadística (creada) todas las combinaciones de letras que comienzan con esas n-1 letras.
b. La última letra de esas combinaciones forma el conjunto de posibles elecciones para la siguiente letra que será añadida a T.
c. Elegir entre esas letras y usar la frecuencia de sus ocurrencias en S para "evaluar" cuál es la mejor elección.
d. Añadirla a T.

Por ejemplo, un primer orden de imitado genera caracteres aleatorios de acuerdo a la distribución estadística del texto de entrenamiento. En un segundo orden imita la distribución de parejas de caracteres de los textos S de entrenamiento, y así sucesivamente para órdenes mayores. El proceso de ocultación de información se realiza mediante la selección de las opciones de la próxima letra a mostrar. Wayner justificó como esto se podría hacer, entre otras opciones, utilizando un árbol de Huffman que basándose en las frecuencias de aparición de los caracteres (por ejemplo) les asignaría un código (código que se utilizará para ocultar una información). Si la selección de las ramas de este árbol (que imita la estadística a la fuente) es aleatoria el texto resultante imitará (o se aproximará) a la distribución estadística del texto fuente.

Se supone (por la información publicada [6]) que para lengua inglesa, dependiendo del texto y del orden (texto de al menos decenas de KB y orden mayor que 8), pueden obtenerse estegotextos con validez léxica y sintáctica. En lengua española, esta afirmación no puede mantenerse en general. El resultado depende mucho de la fuente de entrenamiento elegida. En la práctica, la ocultación de unas pocas decenas de octetos producirá estegotextos con algún error léxico, gramatical o de repetición de términos [7].

La idea de Peter Wayner podría ser mejorada (o eso se piensa) si se considera un nivel de atomicidad de entrenamiento diferente, por ejemplo, la utilización de la palabra en lugar del carácter, dado que se espera que los resultados mejoren, al menos sintácticamente (y se eviten errores léxicos). Otro nivel atómico sería posible pero esto condicionaría el factor de expansión y la capacidad de ocultación, por ejemplo un párrafo, un verso, etc.

IV. STELIN. UNA HERRAMIENTA PÚBLICA DE GENERACIÓN AUTOMÁTICA DE ESTEGOTEXTOS EN LENGUA ESPAÑOLA.

La herramienta libre Stelin, implementada en lenguaje JAVA, permite generar automáticamente estegotextos en lengua española basada en estos criterios y algunas mejoras adicionales (http://steling.sourceforge.net). El algoritmo implementado en Stelin para imitar una o más fuentes de texto de entrenamiento, considerando como nivel de atomicidad la palabra es el siguiente:

1. El proceso de generación se basa en el análisis de bloques de n palabras, extraídas del texto de entrenamiento, mediante una ventana deslizante que se desplaza una posición para cada nuevo bloque. Es decir, el primer bloque tendrá los términos de 0 a n-1, el segundo bloque de 1 a n, y así sucesivamente.

2. N define el orden de complejidad del algoritmo, lo que significa el número de palabras a considerar consecutivamente.

3. Las palabras se relacionan mediante nodos enlazados en los que se contabiliza el número de veces que se han repetido en el texto de entrenamiento. Según esto, existirá una tabla raíz que almacenará todas las "palabras diferentes" que existan en el texto fuente.

Basado en lo anterior, el algoritmo de generación de estegotextos funcionaría, en general, de la siguiente manera:

a) Se selecciona una "palabra" aleatoriamente de la tabla raíz (podría considerarse otro criterio con fines sintácticos, por ejemplo, hacer que el texto empezase por un artículo o mayúscula). De esta forma, para un mismo texto de entrenamiento se podrían obtener diferentes estegotextos.

b) Si esta palabra no tiene sucesores (no apunta a otro nodo), se elije otro término de la tabla raíz (paso a). Si el nodo sucesor solo tiene una palabra, esta palabra se añade al estegotexto (no es posible ocultar información en este caso) y se elige el siguiente nodo disponible. Si el nodo sucesor tiene varias palabras posibles entre las que elegir se elije aquella cuya rama del árbol de Huffman, generado de las posibles palabras (y sus frecuencias), coincida con la información a ocultar, y se elige el siguiente nodo disponible.

c) Si se llega al último nodo (orden n=8, por ejemplo, 8 palabras consecutivas) se elige la última palabra seleccionada para el estegotexto y se vuelve al paso b). Este proceso se repite hasta que se genere el estegotexto que oculta la información deseada.

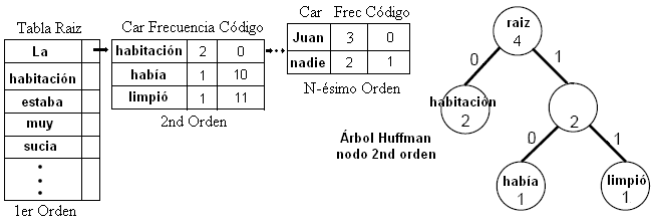

Fig. 1. Ejemplo del algoritmo implementado en Stelin. Atomicidad = Palabra.

d) El receptor necesita construir la tabla de frecuencias del texto de entrenamiento seleccionado para conocer los bits que ocultan cada palabra del estegotexto recibido.

Esta variante, genera estegotextos de mayor tamaño (a mayor nivel de atomicidad es más probable que los elementos no tengan tantos sucesores diferentes), pero es más fácil

obtener textos con validez léxica y sintáctica, e incluso, en ocasiones, con apariencia semántica.

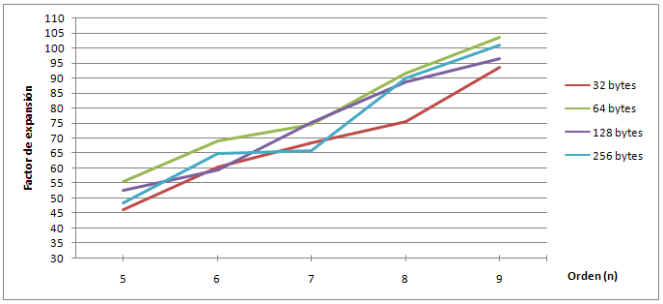

Fig. 2. Relación orden de complejidad y factor de expansión (mensaje a ocultar-estegotexto creado) para diferentes mensajes a ocultar de pequeño tamaño. Texto original: versión digital de Poesías Completas (290KB) de Antonio Machado (51.531 palabras).

[INICIO TEXTO] *viento... El viento traía perfume de rosas, dolor de campanas... Doblar de campanas lejanas, llorosas, suave de rosas aromado aliento ... ¿Dónde están los huertos floridos de la fuente sueñan... Sí, te conozco, tarde alegre y en el hogar campesino armó la envidia pelea. Casáronse los mayores; tuvo Alvargonzález nueras, que le trajeron cizaña, llenan la tierra maldita, tenaz a pico y a la tarde de abril que moría: ¿Al fin la sombra del sendero y el agua del mesón en el azul lejano. De tu morena gracia, de tu sombra a un hombre pensativo y a un agua de la peña? El hombre es por natura la bestia paradójica, un animal absurdo que necesita lógica. Creó de nada un mundo y, su obra terminada, "Ya estoy en el secreto —se dijo—, todo es la laguna insondable. Un buhonero, que cruzaba aquellas tierras errante, fue en Dauria acusado, preso y muerto en el aire ha abierto, y una mata de espliego castellano lleva en el pico a tu jardín diserto —mirto y laureles— desde el alto llano en donde el ojo alcanza su pleno mediodía (un diminuto bando de cuervos enronquece en busca de su peña denegrida, vuelve mi corazón a su faena, con néctares del campo el agua clara corriendo, mientras los dos asesinos tienen la maldición en sus campos. Ya el pueblo impío que juega al mus, de espaldas a la mar y la llanura, caminata o singladura, siempre larga, diéronle, para su prosa, viento recio, sal amarga, y el eco duerme, rodea; agua clara donde beben las rosas blancas, y ante el blanco lino que en los labios ... De tu mirar de sombra quiero llenar mi vaso. Para tu linda hermana arrancaré los montes sin nieve son de violeta. La tierra de mi corazón. Di, ¿por qué acequia escondida, agua, vienes hasta mi, manantial de nueva vida de donde nunca vivida* **[FIN TEXTO]**

Fig. 3. Ejemplo de estegotexto generado automáticamente en lengua española. Ocultación de 16 octetos (0xAA, 0x71, 0xF0, 0x0F, 0xAA, 0x71, 0x54, 0x72, 0xAA, 0x71, 0x28, 0xF5 ,0xAA, 0x77, 0x00, 0xAC). Texto Fuente obra "Poesías Completas" de Antonio Machado (290KB texto plano, versión digital). Orden de complejidad 10. Expansión 1:109. Modo Palabra.

Puede observarse en el ejemplo de la Fig.3 que los estegotextos generados no finalizan necesariamente con una estructura puramente sintáctica. Este problema puede solucionarse de diversas maneras teniendo en cuenta que después de la última palabra del estegotexto al receptor le da igual que información vaya (la herramienta detecta el fin de la ocultación), es decir, la frase puede finalizarse manualmente o por otro procedimiento, o si se desea se puede adjuntar este mensaje a otro texto.

En este proceso de ocultación la selección de los textos de entrenamiento y orden de complejidad son vitales para la generación de estegotextos de calidad. Diferentes tipos de textos podrían ser considerados como fuente para ocultar información (poemas, novelas, artículos periodísticos, código de programación, etc). Si el texto fuente es más grande es más probable que existan diferentes alternativas que sucedan a una palabra y por tanto la capacidad de ocultación sea mayor (piénsese en lengua española por ejemplo en la presencia de preposiciones y determinantes). Desde un punto de vista lingüístico deberían, al menos, evitarse o filtrase fragmentos de texto que claramente afecten a la coherencia en los estegotextos creados. Entre estos, índices, títulos, numeraciones (a), b), c), I, II, III), fechas, referencias, etc.

Por otro lado, a falta de una mejor formalización, las pruebas realizadas indican, que en general, un orden 7 o superior proporciona unos resultados léxicos y sintácticos razonables. Por las pruebas realizadas un orden mayor de 7 o 8 (depende del texto de entrenamiento) puede ser suficiente.

Considerando lo anterior no siempre la selección de un orden N+1 sería mejor (estadística y lingüísticamente) que un orden N o menor, además de que a mayor orden es más probable que el factor de expansión sea mayor y el estegotexto generado (más grande) sea más "atacable".

4.1 ATAQUES A MODELOS N-GRAM.

a. Problemas estadísticos del algoritmo de Peter Wayner y variantes.

Los estegotextos generados deben ser analizados no sólo lingüísticamente, sino además, con toda una serie de análisis estadísticos y estegoanalíticos. El ejemplo más clásico son los estudios basados en la estadística. En la práctica la aproximación estadística de la fuente de entrenamiento (texto) realizada por la idea de Wayner y variantes dependerá de varios factores, entre otros de la función de imitado utilizada.

La implementación actual de Stelin utiliza como función de imitado el algoritmo de Huffman (al igual que la idea original de Peter Wayner), por ser una buena elección, de esta manera, una posible codificación para 3 elementos (a,b,c) con probabilidades (0'80, 0'13, 0'07) sería (1, 01, 00). Si su función inversa es usada como función de imitado los caracteres aparecerían con frecuencia (0.5,0.25,0.25) lo cual dista de ser una aproximación estadística razonable. La explicación de este hecho es debido a la utilización de un árbol binario para la representación de los elementos, ya que su distribución estadística siempre será una potencia negativa de 2. Esta potencia depende de la distancia entre la raíz y la hoja correspondiente del árbol. En general, la existencia de muchos caracteres-palabras en el árbol hará que su profundidad sea mayor y la aproximación estadística a la fuente también, así como la utilización de un orden de complejidad mayor.

Conocido esto, el algoritmo implementado, a nivel de palabra en la herramienta Stelin, no tendría por qué cumplir las aproximaciones estadísticas a nivel de carácter justificadas en el algoritmo original de Peter Wayner. Lo cierto es que, por los estudios realizados, la variante implementada se

aproxima a la distribución reflejada en la fuente de entrenamiento para órdenes grandes, 7 o más. (Véase por ejemplo, Fig.4).

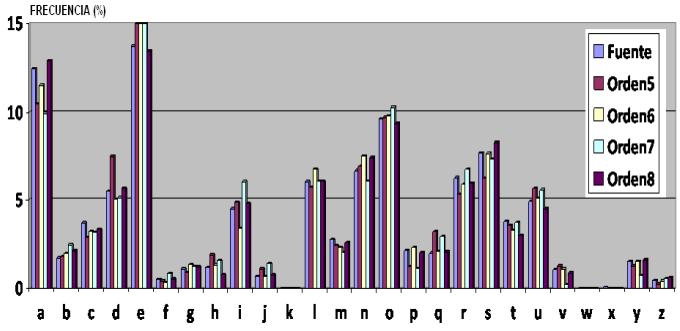

Fig. 4. Ejemplo de comparación de la distribución de frecuencias de caracteres de una fuente de texto de entrenamiento (los 13 primeros capítulos del Quijote con un total de 29.805 Palabras) y un estegotexto generados de ocultar 256 octetos (2048 bits) de información oculta a partir de dicha fuente (como nivel de atomicidad la palabra).

Palabra	Fuente	Estego	Palabra	Fuente	Estego
Y	3.52%	4.12 %	Se	1.10%	1.6%
En	3.61%	4.12 %	Un	1.33%	1.37%
El	3.87%	3.89%	Mi	0.6%	1.37%
Que	3.90%	3.66%	A	1.59%	1.37%
La	3.58%	3.66%	Yo	1.21%	1.14%
De	3.91%	2.98%	Con	0.74%	1.14%
Las	1.4%	1.83%	Me	0.74%	0.91%

Fig. 5. Ejemplo de comparación de frecuencias de las palabras más probables en un estegotexto que oculta 16 octetos mediante un orden 10. Texto Fuente: RIMAS (42 KB) de Gustavo Adolfo Bécquer. Factor expansión 1:144.

Palabra	Fuente	Estego	Palabra	Fuente	Estego
Que	3.9%	5.6%	Yo	1.21%	3.11%
La	3.58%	4.39%	el	3.87%	3.05%
En	3.61%	4.36%	se	1.10%	2.20%
Y	3.52%	3.87%	los	1.25%	1.6%
De	3.91%	3.44%	al	1.19%	1.56%
Del	1.10%	3.13%	par	0.04%	1.49%

Fig. 6. Ejemplo de comparación de frecuencias de las palabras más probables en un estegotexto que oculta 256 octetos mediante un orden 10. Texto Fuente: RIMAS (42 KB) de Gustavo Adolfo Bécquer. Factor expansión 1:166. Tamaño de estegotexto generado comparable al tamaño de la fuente.

b. Ataques estegoanáliticos y criptoanáliticos.

Adicionalmente a los estudios estadísticos, un ataque clásico estegoanalítico a analizar es la posibilidad de un ataque basado en cubierta conocida, es decir, estudiar qué sucedería si un atacante conociera el texto fuente de entrenamiento (que es secreto y compartido entre emisor y receptor), así como el orden de complejidad utilizado para generar un estegotexto concreto (este podría iterarse al ser en general bajo).

Centrémonos en la situación no deseada de que un atacante poseyera los textos de entrenamiento. Un ataque de este tipo permitiría reconstruir los árboles Huffman correspondiente y decodificar cada palabra del estegotexto a un código concreto (suponemos que el atacante conoce que texto es estegotexto). La información recuperada, en general binaria, podría ser analizada posteriormente de diferentes maneras. Un análisis clásico consiste en estudiar su entropía (ecuación de Shannon o aproximación de Shamir y Van Someren [8]) para delatar la existencia de una información cifrada (alta entropía). Este ataque podría ser dificultado de de las siguientes formas[1]:

1. La información a ocultar debería ser cifrada, para en el peor de los casos impedir la recuperación de la información original. La herramienta Stelin utiliza un cifrado basado en el algoritmo Rinjdael (clave 256bits - bloque 256bits) en modo contador, cuya seguridad fue analizada en [9].

2. Los ataques derivados de análisis de entropía y recuperación de información podrían dificultarse de diversas maneras. Actualmente la herramienta Stelin dificulta estos aplicando ideas clásicas de cifradores basados en reducción de redundancia [10]. Stelin utiliza un generador PRNG [9] (Rinjdael-256 en modo contador) que asigna a cada rama de cada árbol Huffman (ver Fig.1) una codificación aleatoria (0 o 1) en función de una clave (no de forma fija, por ejemplo, codificación 0 a la rama derecha y 1 a la izquierda), de modo que un atacante que conozca el texto de entrenamiento y el orden de complejidad tendría dificultades en asignar un código concreto a una palabra del estegotexto determinado. Esto dificultaría, entre otras cosas, extraer la información ocultada y aplicar análisis estadísticos a la misma (por ejemplo, análisis de entropía para revelar la presencia de información cifrada).

3. La herramienta Stelin implementa actualmente 2 mecanismos de ocultación de información. Un mecanismo tradicional de ocultación de cualquier información (cualquier fichero que se convierte a binario) y otro mecanismo que realiza una codificación de 6 bits por carácter a ocultar. Esta última opción está pensada para enmascarar información textual (mensajes cortos, urls, etc) con el menor número de bits con la intención de generar el estegotexto más pequeño posible. Aunque podría seleccionarse otro número de bits, este número es útil ($2^{6\ \text{bits/carácter a ocultar}} = 64$ caracteres) para ocultar mensajes o urls muy variadas. El alfabeto seleccionado es: ***abcdefghijklmnñopqrstuvwxyz*** (27), ***0-9*** (10), 25 caracteres especiales (**/;=@¿?%&|!_" '<>()*.[]espacio+-**), 1 código de mayúscula, 1 código de fin de mensaje. Se incluye un *código de mayúscula* de forma que si se elige esta opción de ocultación y se encuentra una letra mayúscula en la información a ocultar esta letra generará dos códigos (código de mayúscula + código de la letra), es decir, 12 bits a ocultar. En general, es conveniente en esta opción ocultar información sólo en minúsculas, no obstante, en ocasiones, como puede ser una url, es necesario la presencia de una o más mayúsculas.

La asignación de un código concreto de 6 bits a uno de los elementos del alfabeto (2^6 elementos) se realiza pseudoaleatoriamente en función del PRNG implementado y la clave criptográfica seleccionada por emisor-receptor.

[1] La seguridad de la herramienta se aposenta en una clave criptográfica compartida entre emisor y receptor.

c. Coherencia global del texto.

Sin duda, el mayor problema en la ocultación de información mediante esteganografía lingüística, ya sea en textos existen o generados, es conseguir que el estegotexto resultante presente coherencia global, es decir, que no sea una consecución de frases "más o menos pegadas". Este problema y la corrección de los pequeños errores introducidos en la generación de estegotextos se reduce notoriamente en la herramienta Stelin mediante el uso de plantillas y anotación manual.

V. MEJORANDO LAS TÉCNICAS DE ESTEGANOGRAFÍA LINGÜÍSTICA BASADA EN N-GRAM MEDIANTE USO DE PLANTILLAS.

La herramienta Stelin propuesta genera automáticamente estegotextos en lenguaje natural basada en la imitación estadística (N-Gram) de una o más fuentes de datos, en principio, conocidas exclusivamente por emisor y receptor (una misma fuente de datos puede dar lugar a diferentes estegotextos). Emisor y receptor solo necesitan mantener privado las fuentes de entrenamiento y una clave criptográfica, el resto del sistema es público. Es cierto que el orden de complejidad es en principio privado, pero en general este valor como muy alto estará en torno a 9, un poco por encima o un poco por debajo (dependerá de la calidad de los textos fuentes) con lo que podría considerarse como público.

El uso de un nivel de atomicidad igual a palabra permite obtener textos con mejor calidad léxica y gramatical. No obstante, dependiendo de la fuente de entrenamiento los estegotextos generados todavía podrían tener pequeños errores gramaticales, por ejemplo, signos de puntuación que se abren y no se cierran, etc, así como posibles problemas de coherencia global (que es el mayor problema en la ciencia de la esteganografía lingüística).

Una idea interesante, hasta que se consiga un algoritmo que solucione estos problemas, sería la posibilidad de editar los estegotextos generados después de su generación para corregir los problemas presentes. En general, este hecho tiene el problema fundamental de sincronización con el receptor, es decir, introducir cambios en el estegotexto generado implicaría que el receptor debiera conocerlos (lo cual no es muy interesante) o problemas en la recuperación de la información ocultada al desincronizarse la información esperada. Hasta lo que se tiene constancia no se han publicado soluciones en este sentido, si bien sí soluciones que permiten mejorar la calidad del estegotexto en el proceso de generación [5] pero no mejora posterior en el estegotexto generado sin perjuicio para el receptor. Esta idea puede ser usada con la herramienta Stelin con una serie de consideraciones y limitaciones.

Pensemos por un momento en la estructura en forma de árbol utilizada por la herramienta Stelin para ocultar una información. Véase el ejemplo de la Fig.7.

Una información se oculta mediante la selección de una palabra entre las disponibles en el siguiente nivel. Es decir, si se selecciona la palabra "La" la siguiente palabra podría ser "habitación" (bit 0), "había" (bit 10), "limpió" (bit 11) en función de los bits a ocultar. El receptor al recibir el estegotexto construiría la tabla de frecuencia al igual que el emisor e iría relacionando palabra por palabra del estegotexto recibido con la tabla generada, de esta forma, el receptor al recibir "La habitación", "La había", "La limpió" sabría que se ha ocultado un bit (0) o dos (10-11) respectivamente.

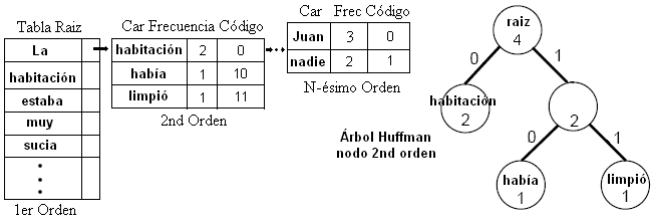

Fig. 7. Ejemplo de codificación de palabras y árbol Huffman.

En este proceso de desocultación y de sincronización podría actuarse de varias formas. Una consistiría en que la herramienta diera error si después de una palabra del "supuesto" estegotexto se encuentra otra palabra que no coincide con ninguna de las esperadas (si se piensa este hecho si el atacante tuviera información de las tablas de frecuencia esto le simplificaría el hecho de descartar mensajes sin información oculta). Otra variante consistiría en despreciar todas las palabras que se lean hasta que se encuentre una de las palabras posibles para el nivel de la tabla (lista enlazada) donde nos encontremos. Realmente lo importante es que el receptor no pierda la sincronización respecto a la tabla de frecuencias y al estegotexto recibido.

Este pequeño detalle no publicado, permite mejorar, en principio manualmente, a posteriori estegotextos generados por el emisor sin que ello afecte al receptor. La única condición es que el emisor puede utilizar cualquier palabra que no se encuentre en el nivel posterior para que el receptor no pierda sincronía. Es decir, en el ejemplo anterior entre las palabras que forman las parejas "La habitación", "La había", "La limpió", podría utilizarse cualquier palabra que no fuera "habitación, había o limpió". De esta forma el emisor puede corregir posibles errores gramaticales y mejorar la coherencia del texto sin necesidad que el receptor conozca esta información. Por ejemplo, si ocultamos una pequeña información de 126 bits, usando un orden 9 y fuente de entrenamiento las poesías completas de Antonio Machado obtendríamos entre los estegotextos posibles uno como el siguiente:

planeta por donde cruza errante la sombra de Caín criminal. ¡Gloria a Caín! Hoy sólo quedan lágrimas para llorar. No hay camino, sino estelas en la mar. ¡Fugitiva ilusión de ojos guerreros, que el polvo barre y la ceniza avienta. ¿Qué has hecho? La muerte no hay camino, se hace camino al andar. El que espera desespera, dice la mano viril que la blandiera, no por los salones de sal-si-puedes suena el rebato de la tarde en la arboleda! Mientras el corazón pesado. El agua en sombra pasaba tan melancólicamente, bajo los arcos del puente al ímpetu del río sus pétreos tajamares; la guerra nos devuelve los muertos milenarios de la tierra pamplonesa; encinas de Extremadura, a un ventanuco asoman, al declinar el sol, sobre el romero, tan disparatada! sobre el campanario. Es una tarde mustia y desabrida de un otoño sin camino, como el niño que en la mar te empuje por valles y barrancas, la tarde habrá caído sobre la tierra, y una

Fig. 8. Estegotexto que oculta 126 bits. Orden 9. Clave:Alfonso. Fuente de entrenamiento poesías completas de Antonio Machado.

Como puede observarse este estegotexto tiene una serie de pequeños errores gramaticales (dejamos a un lado problemas de coherencia global). Véase Fig.9.

planeta por donde cruza errante la sombra de Caín criminal. ¡Gloria a Caín! ***H****oy sólo quedan lágrimas para llorar. No hay camino, sino estelas en la mar.* ***¡****Fugitiva ilusión de ojos guerreros, que el polvo barre y la ceniza avienta. ¿Qué has hecho?* ***L****a muerte no hay camino, se hace camino al andar. El que espera desespera, dice la mano viril que la blandiera, no por los salones de sal-si-puedes suena el rebato de la tarde en la arboleda****!*** ***M****ientras el corazón pesado. El agua en sombra pasaba tan melancólicamente, bajo los arcos del puente al ímpetu del río sus pétreos tajamares; la guerra nos devuelve los muertos milenarios de la tierra pamplonesa; encinas de Extremadura, a un ventanuco asoman, al declinar el sol, sobre el romero, tan disparatada****!*** *sobre el campanario. Es una tarde mustia y desabrida de un otoño sin camino, como el niño que en la mar te empuje por valles y barrancas, la tarde habrá caído sobre la tierra, y una*

Fig. 9. Algunos de los errores gramaticales presentes en el estegotexto de la Fig.8.

Para solucionar estos problemas la herramienta Stelin genera una plantilla con las palabras posibles en cada nivel, de forma que el emisor pueda seleccionar que palabras añadir entre palabras del estegotexto, palabras que serán despreciadas por el receptor.

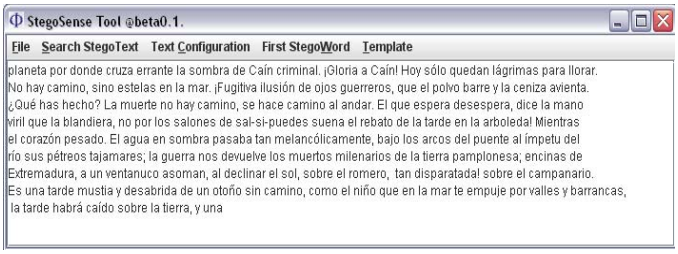

Fig 10. Herramienta StegoSense útil para la modificación manual de estegotextos generados por Stelin.

Fig. 11. Plantilla generada para un estegotexto concreto.

Por ejemplo, seleccionamos de Fig.9 la frase "*de la tarde en la arboleda! Mientras el corazón*". Veamos un trozo de la plantilla generada:

[WORD:en]
[muerta][flota][.][bella][roja][en][,][sobre][arrebolada][y]
[WORD:la][sus][la]
[WORD:arboleda] [arboleda]
[WORD:!] [!]
[WORD:Mientras] [Mientras]
[WORD:el] [el]
[WORD:corazón][querido][sueño][fondo][mar][temblor][semblante][tictac][vino][aire][sol][ataúd][silencio][blanquecino][maestro][solitario][blanco][mármol][fruto][encanto][hálito][patio][pretil][ambiente]…

Teniendo en cuenta esto editamos la frase (entre las múltiples opciones posibles elegimos la siguiente):

"tarde en la ***dulce*** *arboleda****, ¡qué sensación****!. Mientras el corazón"*

De esta forma tan sencilla puede mejorarse sustancialmente la calidad del estegotexto generado.

Para solucionar problemas derivados de signos de puntuación que se abren y no se cierran (o viceversa) u otros, la herramienta Stelin considera los signos de puntuación como palabras individuales para poder añadir información delante de ellos con la intención de compensar posibles errores que se produjeran.

La modificación de los estegotextos por este procedimiento requiere un poco de práctica. Por suerte en los textos en lenguaje natural en español (y en otras lenguas) las palabras tienen colocaciones determinadas, es decir, es más probable que ciertas palabras vayan detrás de otras y es más probable que existan más palabras detrás de unas que de otras. En general, existirían pocas palabras después de las cuales será más costoso elegir una palabra nueva (porque existen en el nivel de la tabla correspondiente muchas opciones) y muchas palabras después de las cuales existirán pocas opciones con lo que se tendrá más libertad para añadir palabras nuevas.

Por ejemplo, en textos en lenguaje natural palabras como de, la, que, el, en, y, a, los, etc, son más probables luego es más probable que existan más palabras unidas a estas. Si nos fijamos en el ejemplo anterior resultaría trivial añadir información antes del artículo "el" pero más difícil encontrar palabras después de esta palabra y antes de la palabra "corazón". Teniendo en cuenta estos principios esto puede ayudar para reducir el tiempo y las posiciones donde es mejor trabajar para corregir los posibles fallos gramaticales.

Según todo lo anterior es posible generar estegotextos en lengua española automáticamente y corregir los errores manualmente sin afectar al receptor, creando estegotextos de una calidad más que notable.

Actualmente la ocultación de información con la herramienta Stelin tiene una capacidad de ocultación aproximada de 1 PALABRA-1 BIT, esta capacidad de ocultación disminuye en función del número de palabras que se añadan mediante edición manual. Para ocultar informaciones por encima de la centena de bits los estegotextos generados serán de tamaño notorio (depende del tamaño y de la "calidad" de la fuente de entrenamiento) y por tanto la edición manual llevará un tiempo considerable. El interés de invertir más o menos tiempo en la calidad de los estegotextos generados dependerá de la importancia de la información intercambiada entre emisor y receptor.

Partiendo de todo lo anterior es posible otro funcionamiento de la herramienta aprovechándose de la característica de edición manual. Este procedimiento consistiría en la generación de estegotextos de "mala calidad"

pero que ocuparan poco espacio. Desarrollemos un poco más esta idea.

Dado que podemos modificar el estegotexto resultante podríamos usar un orden muy bajo (lo que produciría un estegotexto de mala calidad) para generar un estegotexto de poco tamaño y a continuación mejorarlo manualmente, además podríamos forzar un cambio de coherencia global. Véase Fig12-15.

La tarde se ha ido llegando las hojas de la fuente se oía tañer de una tierra. Nunca se cansa. Pasado habían el agua muda que enorme muro de la fuente. Yo no conozco el agrio zumo dorado de amor. El tren, abril galán. ¡Oh, dime si son mías. La tarde caía, que a mi

Fig 12. Orden 3, 126 bits ocultos, clave=alfonso, Poesías Completas A. Machado. Capacidad de ocultación 1 Palabra-2 Bit.

La tarde se ha ido llegando *y el viento contra* ***las hojas*** *esculpidas* ***de la fuente se oía tañer de*** *campanillas, melodía para* ***una tierra*** *entristecida. Esa fuente de la que fluye agua como nectar sin cesar.* ***Nunca se cansa. Pasado*** *10 años ya no* ***habían*** *enamorados que bebieran* ***el agua muda*** *de la fuente* ***que*** *en otra época, tras un* ***enorme muro****, retaba a demostrar su amor bebiendo libertad* ***de la fuente*** *prohibida.* ***Yo no conozco el agrio*** *muro y tampoco el* ***zumo dorado de amor*** *de esos jóvenes.* ***El*** *tren del amor ya pasó****,*** *mi* ***abril*** *cuando fui* ***galán*** *pasó como estrella fugaz en el cielo****. ¡Oh, dime si*** *puedo recuperarlo!. Mis paranoias* ***son mías*** *pero puede que haya esperanza.* ***La tarde caía, que*** *rápido...* ***a mi*** <u>*me pareció como un suspiro.*</u>

Fig 13. Ocultación de 126 bits. Modificamos estegotexto de Fig.12 creando con coherencia global. 1 PALABRA – 1 BIT

agua de la fuente de la primavera blanca entre los verdes hojas el campo verde que a tu sombra, ¿No es más aparente escisión del ser. En la fuente de la tarde, la fuente, en el campo verde, el agua, en el agua clara, casi con placidez de alma de la luna y de ceniza, estos limonares verdes. ¡Oh fe y la tierra de una tarde muerta. ¡Ay, lo otro inasequible." Su reflexión autoinspectiva. ¿Ya no le lleva a la mar, está más allá de los montes, y en la tierra. Y en aquella ausencia en esta paz con los ojos abiertos los balcones del viejo pueblo paseando solo, en la tierra de tu huerto, colmenar y campo y el poeta, y la luz y de la noche. ¡Qué importa que en los ojos me recuerdan un día. Como atento no es, y el campo. Juan lentamente avanza, sierra fría, y en los ojos de llama, el corazón del amor, en el alma. ¡Oh tierra ingrata y fuerte olor de un sueño. Larga es la fuente de la vida, sin luna, en la fuente de la tarde, dijeron tu pena, sé con qué se hicieron?, insiste en preguntar lo que se apaga o beso que no es, y a la vera del camino. Recio viento sopla, tienen la maldición en sus ojos? ¿Tu hermana es la canción que deje cenizas en la clara, casi de primavera, y el poeta

Fig 14. N=3. 570 bits ocultos, clave=perroblanco, Poesías Completas-A.Machado. [ST335]. TEXTO: **manifa 20:00 plaza cibeles policia id preparados 21:00 retirada sicroniza twitter/perroblanco** [94 caracteres]. 1 palabra- 2bit.

agua de la fuente de la primavera blanca entre los *álamos* ***verdes*** *y las* ***hojas*** *amarillas. Fue ese* ***el campo verde que*** *recordó tu cara,* ***a tu sombra*** *delgada regaló una rosa****,*** *creció y floreció.* ***¿No es más aparente*** *la* ***escisión*** *de tu belleza en una rosa que* ***del ser*** *triste en el cual te has convertido?****. En la fuente de la*** *arboleda ayer* ***tarde*** *recordé esa rosa****, la fuente*** *brotaba agua enrojecida****, en el campo verde*** *con álamos sedientos****,*** *donde* ***el agua*** *teñida fluía****, en el agua*** *poco* ***clara,*** *parecida a tu vida, vi tu rostro. Vi tu rostro* ***casi*** *completo* ***con*** *la* ***placidez de*** *mi* ***alma*** *o* ***de la luna y de ceniza*** *es mi recuerdo****,*** *de ceniza. Tu olor me recuerda a* ***estos limonares verdes*** *entre los árboles****. ¡Oh fe*** *bendita* ***y*** *valiente!. Fue en* ***la tierra de una*** *ciudad como esta donde una* ***tarde muerta*** *te declaré mi amor****. ¡Ay,*** *que pena!, me regalaste una sonrisa, l****o otro*** *un beso tuyo "fue* ***inasequible." Su reflexión autoinspectiva****, algo triste, la apartó de mi lado****. ¿Ya*** *marchó?, ya* ***no le lleva a la mar*** *su mirada perdida****, está más allá de los montes, y en la tierra*** *clavado grité su nombre****. Y en aquella ausencia****, en esta*** *soledad, busqué la* ***paz con los ojos abiertos*** *mirando a través de* ***los balcones del viejo pueblo*** *donde nacimos. Busqué tu aroma* ***paseando solo****, busqué* ***en la tierra de tu huerto*** *tu néctar,* ***colmenar y*** *hogar de tu dulce miel. Tu* ***campo y*** *tu huerto recuerdan nuestra infancia, como diría* ***el poeta, y*** *refleja* ***la luz*** *en la mañana* ***y de la noche*** *estrellada****. ¡Qué importa*** *todo esto ahora!* ***que*** *grite o* ***en*** *delirios* ***los ojos*** *llorando sangre* ***me recuerdan un día*** *tras otro tu aroma angelical****. Como atento*** *estoy a mis recuerdos* ***no es*** *cierto que****, y el*** *todopoderoso es testigo, el* ***campo*** *esté desamparado****. Juan*** *mi fiel jardinero* ***lentamente*** *planta rosas,* ***avanza*** *poco a poco, porque la* ***sierra fría*** *no entiende de recuerdos ni de anhelos****,*** *muerde almas* ***y en los ojos de*** *las tormentas se vislumbra la fiereza, la* ***llama*** *negativa****,*** *de lucha encarnecida. Mientras* ***el corazón*** *beba* ***del amor, en el alma*** *habrá esperanza****. ¡Oh tierra ingrata y*** *vacía, ayúdame a conquistar a mi amada!. Sé que* ***fuerte*** *es el* ***olor de un sueño****.* ***Larga es la fuente de la vida,*** *pero vacía* ***sin luna,*** *pensad* ***en la fuente*** *sin agua* ***de la tarde,*** *vacía. Me* ***dijeron****: ¿****tu pena*** *es compartida?****,*** *no estás solo ella también sufrió. ¿Sus recuerdos sabes y* ***sé con qué se hicieron?, insiste en preguntar*** *Juan. Yo le recordé* ***lo que se apaga*** *no entiende de excusas, ni su abrazo ausente ni su boca* ***o beso que no*** *fue, ni* ***es, y*** *además* ***a la vera del camino*** *seguiré esperándola. A pesar que no es fácil.* ***Recio viento sopla,*** *¿quizás* ***tienen la maldición en sus ojos*** *tus amantes por desear tu boca****?***

¿Tu hermana *también te buscó sin aliento?, ¿qué fue de tu familia?. ¿Donde estarás?, recito con mi guitarra,* ***es la canción que*** *grito al aire y a fuego vivo* ***deje cenizas en la clara*** *madrugada, madrugada de olvido. Te escribo sin* ***casi*** *esperanza* ***de*** *recuperarte, así como vuelva la* ***primavera,*** *primavera que vence al duro invierno. Volveré a escribirte* ***y el poeta*** *que llevo dentro florecerá de nuevo.*

Fig 15. Ejemplo de modificación de estegotexto Fig.14. Capacidad de ocultación 1 palabra-1 bit. Se fuerza un estegotexto con coherencia global y se busca un "estilo poético".

Como puede verse incluso generando estegotextos de mala calidad (orden bajo o como resultado de mala selección de la fuente de entrenamiento) el uso de plantillas puede ayudar no solo a corregir errores gramaticales sino incluso

proporcionar coherencia completa al texto. Como puede verse si el orden elegido es grande (7,8 o 9) la corrección de los textos será más rápida (al existir menos errores) si bien el estegotexto generado será mayor. Por el contrario para ordenes bajos (por ejemplo, N=3) el texto será más pequeño pero la corrección implicara más tiempo y posiblemente sea necesaria añadir más palabras.

Por las pruebas realizadas la herramienta Stelin tiene utilidad (sin excesivo coste en tiempo) para la ocultación de información de manera "muy segura" de una centena de bits (por ejemplo, 512/6 bits=85 caracteres a 1024/6=170 caracteres) en textos en lenguaje natural con seguridad estadística y lingüística.

VI. CONCLUSIONES.

La esteganografía lingüística es una ciencia que en los últimos años está despertando el interés de la comunidad científica por su enorme potencial. Existen multitud de problemas a solventar, uno de los principales consiste en que el lenguaje natural como medio de ocultación de información es poco redundante (ruidoso) en comparación con otros estegomedios de uso más común (imágenes, vídeo, etc), lo cual dificulta la ocultación de información de forma imperceptible, estadística y lingüísticamente. La aplicación actual de la esteganografía lingüística en lengua española es muy pobre, de hecho, existen pocos trabajos de interés que analicen algunos de sus aspectos con seriedad. Por este motivo, en el presente artículo profundiza en un procedimiento de generación automática de estegotextos aplicado a lengua española. Se incorpora la posibilidad de corregir manualmente los errores gramaticales e incluso mejorar sustancialmente la coherencia de los estegotextos generados mediante el uso de plantillas.

En la práctica se consiguen estegotextos de gran calidad pero es necesario invertir mucho tiempo (depende de la habilidad de cada emisor) para ocultar más de una centena de bits (intercambio de claves criptográficas, urls, mensajes de movilización, coordenadas, etc). Si bien es cierto, en todo momento la calidad del estegotexto puede ser modificada por el emisor y perfeccionarla. Si la información a ocultar es lo suficientemente importante puede que el tiempo invertido sea adecuado.

El emisor y el receptor deben compartir una clave criptográfica y uno o más textos fuentes de entrenamiento. Si por algún motivo los textos fuentes son comprometidos, pueden cambiarse sin ningún problema eligiendo nuevos textos con los que trabajar.

Queda mucho terreno por avanzar y no queda claro si es posible obtener procedimientos esteganográficos lingüísticos robustos, automatizados y que permitan una ocultación alta de información, mientras tanto Stelin puede ser una buena herramienta para conseguir estegotextos de alta calidad.

REFERENCIAS

[1] Carracedo, J.: Seguridad en Redes Telemáticas. Mc-Graw Hill InterAmericana de España. ISBN: 84-481-4157-1 (2004), páginas 123-131.

[2] Muñoz, A., Carracedo, J., Sánchez, S: Detection of distributed steganographic information in social networks. EATIS 2008. Euro American Conference on Telematics and Information Systems, September 10-12. ACM-DL Proceedings will have ISBN # 978-1-59593-988-3. Copyright © 2008.

[3] Kahn, D: The Codebreakers: The Comprehensive History of Secret Communication from Ancient Times to the Internet (Hardcover) Scribner 1996. ISBN-13: 978-0684831305.

[4] Bergmair, R.: A comprehensive Bibliography of Linguistic Steganography. SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents, *volume 6505, January 2007.*

[5] Chapman, M., Davida, G.I.: Hiding the hidden: A software system for concealing ciphertext in innocuous text. In: Han, Y., Quing, S. (eds.) ICICS 1997. LNCS, vol. 1334, pp. 11–14. Springer, Heidelberg (1997)

[6] Wayner,P.: Disappearing Cryptography, Second Edition – Information Hiding: Steganography and Watermarking. Morgan Kaufmann Series; 2 edition (May 2002). ISBN-13: 978-1558607699.

[7] Muñoz, A., Carracedo, Justo: Estegoanálisis aplicado a la generación automática de estegotextos en lengua española.Actas del V Congreso Iberoamericano de Seguridad Informática. CIBSI'09. Montevideo, Uruguay. ISBN: 978-9974-0-0593-8c.

[8] Shamir, A., Van Someren, N: Playing 'Hide and Seek' with Stored Keys. Lecture Notes in Computer Science. Springer Berlin. Volume 1648/1999. ISBN 978-3-540-66362-1.

[9] Muñoz, A., González, M: PRNG based on new HCI devices entropy sources. Wii ReMote study case. EuroAmerican Conference on Telematics and Information Systems. EATIS PRAGUE 3-5 June 2009.

[10] Hwang, M., A New Redundancy Reducing Cipher. Informatica, vol. 11, no. 4, pp. 435-440, Oct. 2000.