Esteganografía Simulada para Análisis de Efectos sobre Portadores Imagen

Mg. Ing. Guillermo Sergio Navas[1], Mg. Ing. Gustavo Rodríguez Medina[2]

Gabinete de Computación / Fac. de Ingeniería / Univ. Nacional de San Juan [1,2]

Av. Libertador Gral. San Martin 1109 (oeste) – San Juan
0264 – 4211700 (Int. 435[1] / 285[1, 2])

snavas@unsj.edu.ar [1], grodriguez@unsj.edu.ar [2]

RESUMEN

La Esteganografía digital hace uso de variadas técnicas para ocultar información en un Portador, con el fin de que ella pase inadvertida para terceros [3]. Generalmente, los objetos portadores utilizados son elementos multimediales (imágenes, audio y video), mientras que el mensaje puede ser de cualquier tipo[1] [2].

La propuesta, expuesta en el presente artículo, se utiliza para implementar una herramienta software que es utilizada para el análisis de efectos esteganográficos sobre portadores imagen tipo BMP [1]. Ella surge como solución a la necesidad de analizar los efectos esteganográficos para poder evaluar los métodos y ponderar características deseables (perceptibilidad visual, detectabilidad y capacidad); teniendo como requisito que los mensajes deben ser de tamaños precisos y adecuados. Pero, disponer de "archivos mensajes" reales, con tamaños específicos para cada ensayo (son cientos) y evaluación, y aplicar esteganografía real, es una tarea excesivamente costosa y complicada.

Por ello, se planteó la *Esteganografía Simulada*, mediante la cual se genera computacionalmente un mensaje ficticio, de tamaño apropiado para cada uno de los ensayos, el mismo se inyecta en el portador, provocándole un efecto esteganográfico igualmente valido al que si se utilizara un mensaje real, pero el proceso resulta ser mucho más eficiente, rápido y con parámetros controlados por el operador.

La *Esteganografía Simulada*, expuesta en este artículo, es innovadora, no existen otros antecedentes al respecto.

Palabras clave: Esteganografía, Mensajes ficticios, Métodos LSB.

CONTEXTO

En este artículo se expone uno de los temas desarrollados en la tesis *"Exploración de efectos esteganográficos sobre portadores imagen de mapa de bits utilizando diferentes técnicas y algoritmos",* de la Maestría en Informática de la Univ. Nacional de la Matanza.

Cabe destacar que ese trabajo de Maestría a servido para la formación de un equipo de investigación en la temática, en la Facultad de Ingeniería de la Universidad Nacional de San Juan, y a partir del cual se han generado otras propuestas de Tesis de Posgrado.

1. INTRODUCCIÓN

La esteganografía trata el proceso de ocultar un objeto software de cualquier tipo (mensaje), en otro, denominado portador, con el fin de ser enviado desde un emisor hacia un receptor de

[1] Se hace referencia a un "archivo mensaje", de tamaño limitado a la capacidad del portador.

manera subrepticia, esto es, se lleva a cabo una transmisión encubierta [2].
Si bien pueden ser utilizados portadores de cualquier tipo, se prefiere los objetos multimediales, aprovechándose la limitación de los sentidos humanos: la vista y el oído no pueden detectar cambios sutiles en presentaciones visuales o de audio [2] [3].
Estos portadores resultan ser los preferidos en aplicaciones esteganográficas como elemento portador, dada la eficacia de las técnicas que aprovechan tales limitaciones humanas. En este tratado se utilizan portadores imagen Bitmap en formato BMP color de 24 bpp.
En la figura 1 se muestra un esquema general de un sistema de esteganografía.

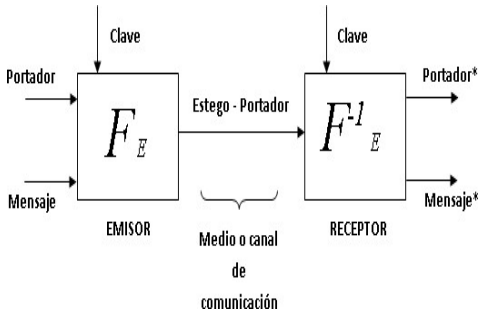

Fig. 1. Esquema de un sistema de Esteganografía.

Exploración de efectos esteganográficos

Existen diversas técnicas o métodos para ocultar un mensaje en un portador imagen. Entre las más difundidas están: la de Sustitución LSB 1 bit y la de inyección al final del portador[6][1]. En el desarrollo del trabajo de Tesis de Maes- tría [1] se han propuesto las variantes y combinaciones de la técnica de sustitución básica, como por ej. variar la cantidad de bits menos significativos a sustituir en el portador, utilizar diferentes canales de color RGB, alternar el uso entre ellos, usar dispersión en los pixels, etc, además de las posibles combinaciones entre variantes.
El objetivo central del trabajo de Tesis ha sido la exploración de los efectos provocados en portadores BMP por la aplicación de técnicas de sustitución, a fin de encontrar *las mejores* en cuanto a las características de *perceptibilidad* visual y de *capacidad*, eventualmente también de la característica de *detectabilidad*.
La implementación, y posterior prueba y valoración, de tal cantidad de variantes (más de mil preseleccionadas) puede llegar a ser bastante complicada y laboriosa si se la encara con un desarrollo de esteganografía real. Además, si para el caso que se requieran realizar varias pruebas sobre diversos portadores para determinar la perceptibilidad y la evaluación de la capacidad, esta última característica requiere encontrar un mensaje que se inserte exactamente en el 100% del portador, es decir que *"encaje"* en su totalidad exacta (byte a byte) en toda la imagen.

Simulación de esteganografía:

Por lo dicho en el punto previo, se ha tenido que pensar en soluciones más eficientes y senci- llas para realizar el estudio y exploración. Es así que se ha propuesto y desarrollado, con exce- lentes resultados, una forma de "provocar" los efectos bajo estudio sobre portadores imagen, pero con niveles de inserción controlados, ele- gidos por el usuario. De esta forma se puede, por ej., con total facilidad encontrar con certeza el valor de la capacidad que provee cada méto- do o técnica, cualquiera sea la variante utiliza- da. También generar, en escaso tiempo, efectos de multitud de técnicas sobre portadores testigo elegidos a fin de compararlas, valorarlas y pon- derar características de perceptibilidad, incluso analizar detectabilidad. Obviamente en función de las características halladas también se podrá evaluar la bondad de cada método.
Lo anterior se ha logrado implementando por software una suerte de *"Simulación de Esteganografía"*. Esto significa simular en un portador los efectos que puede provocar el ocultamiento de un *"mensaje verdadero",* pero usando un *"mensaje ficticio",* que es generado computacionalmente, por tanto factible de elegirlo en un tamaño muy preciso. El operador simplemente elige un método de sustitución a analizar

y el nivel de inserción, y la posterior generación del archivo estego es automática e inmediata.

Esteganografía real Vs. simulada:

La única diferencia entre la esteganografía real y la simulada es que, en la última, el mensaje oculto es ficticio, y por lo tanto no se requiere crearlo, leerlo, cargarlo, procesarlo ni recuperarlo, como sí se debería en esteganografía real. De hecho, para los fines perseguidos, no tiene ninguna relevancia recuperar el mensaje, sólo importa saber que sí se podrá recuperar en caso de una implementación análoga pero real y que los efectos provocados sobre el portador serán los mismos.
La sencillez de la lógica simulada frente a la esteganografía real es notable, dado que el mensaje se genera con lógica simple y en paralelo a la simulación.
La eficiencia de los algoritmos es mucho mayor, ya que no se lee ni se procesan archivos de mensajes, sino que se los genera usando rápidas lógicas booleanas.
La generación de estegoportadores es más simple y requiere muy escaso tiempo, lo cual agiliza notablemente los ensayos para análisis.
No se requiere buscar mensajes de largo adecuado para la prueba de cada portador y cada técnica (son cientos). El nivel de inserción lo elige el operador y el proceso se realiza automáticamente con un "mensaje" de largo exacto.
No se requiere la función inversa de recuperación del mensaje, sólo la certeza de que es factible lograrla sin complicaciones fuera de lo común.

Validez de la metodología por simulación:

Desde el punto de vista del diseño y la implementación, las *técnicas de sustitución simulada* y sus lógicas de generación, respetan exactamente los conceptos y reglas de los mismos métodos; es decir que la simulación esteganográfica se realiza de manera idéntica a como si fuera la real, por lo cual, desde ese punto de vista no hay nada que comparar ni comprobar, simulada o real usan idénticas técnicas. Lo único que cambia es el mensaje. La esteganografía simulada no requiere tener que operar con un mensaje real, evitando procesarlo, conocer su tamaño, la forma de lectura del archivo, acceso al disco, las lógicas de separación por canales, ocultamiento de parámetros de recuperación en el portador, etc.
En cuanto a los efectos provocados sobre el portador, la simulación de esteganografía podría llegar a diferir mínimamente de una aplicación real, la diferencia sería solamente debida a que se utiliza un mensaje ficticio y por lo tanto simulado; por lo cual, para lograr los efectos correctos se debe generar adecuadamente el mensaje, con validez comprobada. Se intenta que los efectos provocados sobre el portador sean lo más parecidos posible a los derivados de una esteganografía con mensaje real. Esto se logra realizando alteraciones, a nivel de bits, de carácter aleatorio con distribución uniforme.
Resulta necesario definir un elemento de medición, denominado *Nivel de Afectación*. Conceptualmente, es la relación porcentual entre la cantidad de Bytes utilizados en el proceso esteganográfico respecto a la cantidad de Bytes que efectivamente sufrieron alguna modificación. La expresión es la siguiente:

$$NivelAfectación = \frac{BytesUtilizados}{BytesModificados} \times 100$$

Se realizaron diversas y numerosas pruebas con LSB 1 real [6], utilizando variados tipos de mensajes u objetos a ocultar, tales como texto ASCII, imágenes BMP, archivos ZIP, ejecutables, etc. Como resultado, el nivel de afectación medio resultó del 50%, con una dispersión media en torno a ± 1,5%.
La conclusión de las pruebas con esteganografía real es que se alteran los bits LSB de los Bytes del portador en aproximadamente un 50% de los casos. Y alrededor de la mitad de los bytes utilizados no sufre ninguna alteración. Esto provoca finalmente en el portador deter- minado nivel de ruido.

Respecto a pruebas de validez de mensajes ficticios, se preparó para el análisis una muestra creada con un programa que genera números binarios aleatoriamente (pseudoaleatorios), y que cuenta la cantidad de unos y su promedio general. Se generaron 20 secuencias de 32 números cada una, en la suposición de que se está emulando 20 mensajes distintos de 4 caracteres cada uno. Se hicieron varias corridas del programa obteniéndose varias tablas de números binarios aleatorios.

Como resultado, se obtuvo que los binarios generados tienen una media que tiende al 50% de unos (y ceros) con una desviación bastante escasa, algo menor que la de mensajes reales. Por tanto se prueba que el algoritmo de simulación por generación aleatoria binaria es completamente válido, provocando efectos análogos a los que produce un mensaje real sobre un portador.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La línea de investigación corresponde a la temática *Esteganografía*, la que se enmarca en el área de Seguridad Informática[3]. En este sentido, los autores del presente artículo han elaborado hasta el momento dos tesis de Maestría, como así también la publicación de resultados en diferentes congresos [4][5][6].

Los resultados obtenidos desde el año 2006 a la fecha, aportan estudios y desarrollos innovadores en el área.

3. RESULTADOS OBTENIDOS/ESPERADOS

De las pruebas efectuadas, descriptas anteriormente, se desprende que si se generan los cambios en el bit menos significativo de un portador usando la función de generación aleatoria, los cambios obtenidos son casi iguales a una sustitución LSB 1 bits realizada con un mensaje real [6]. Y en consecuencia se debe obtener en el portador niveles de perturbación general en el canal y perceptibilidad de valores análogos a los conseguidos con mensajes verdaderos o reales.

Varias pruebas numéricas, como las descriptas, y otras mucho más extensas, dieron resultados satisfactorios, que indican que la simulación esteganográfica así realizada arroja resultados correctos, cualquiera sea el método de susitución.

Cabe destacar que oportunamente se desarrolló un software piloto, al que además de implementar algunas técnicas esteganográficas reales, se le dotó de la capacidad de aplicar esteganogra- fía simulada, para efectuar exploración de efec- tos. Ese software, cada vez que se hace una simulación, presenta un informe del *Nivel de Afectación* en los bytes del portador utilizado. Si con este software se aplica esteganografía real con la técnica LSB 1 bit, también se obtiene un reporte del *Nivel de afectación* (o de ruido). Por tanto el soft provee medios comparativos para conocer la bondad de la simulación.

En numerosas pruebas realizadas con le software piloto se ha comprobado que en el caso de mensajes ficticios la tendencia del *Nivel de Afectación* es muy marcada hacia el 50%, con una dispersión bastante baja, en general no mayor a ±1%, y en caso de mensajes reales es algo dependiente del tipo de mensaje, para un mismo portador, pero también con fuerte tendencia hacia un 50% aunque con algo más de dispersión, en torno a ±1,5%.

En cuanto a la evaluación del nivel de perceptibilidad de la simulación, comparado con la esteganografía real, es necesario *"poder ver"* los cambios LSB provocados. Esto se logra también con el software implementado, a través del filtrado de bits (los menos significativos del portador) y ajustes gráficos usando un editor de imágenes. La Fig. 2 muestra un caso, se observa la afectación de pixeles en una imagen totalmente blanca, los píxeles blancos son los no afectados.

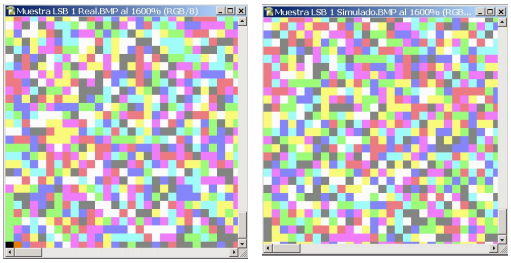

Fig. 2. Afectación de pixeles en un portador, con efecto real y simulado. Zoom 1600.

La validez y bondad de la simulación ha sido comparada numérica y visualmente. Por tanto es completamente válido realizar exploraciones de los efectos esteganográficos utilizando mensajes simulados, con lo cual se simplifica la implementación del software, y los ensayos para análisis resultan ser notablemente más sencillos, rápidos y precisos.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de investigación, encabezado por el autor principal, viene trabajando en la temática de Esteganografía desde 2005.

En 2006, defendió el trabajo de Tesis de Maestría titulado *"Exploración de efectos esteganográficos sobre portadores imagen de mapa de bits utilizando diferentes técnicas y algoritmos".*

Con el transcurso del tiempo, el anterior trabajo dio lugar a la propuesta, desarrollo y defensa de otra Tesis de Maestría titulada: *"Estudio, análisis, desarrollo y propuestas de algoritmos para la selección óptima de métodos de sustitu- ción en aplicaciones esteganográficas", 2015,* posibilitando de esta manera la formación de nuevos recursos humanos en el área.

5. BIBLIOGRAFÍA

[1] G. Sergio Navas. Exploración de efectos Esteganográficos sobre portadores imagen de mapa de bits utilizando diferentes técnicas y algoritmos. Tesis. Argentina. Univ. Nacional de la Matanza – Escuela de Posgrado. 2006.

[2] Abbas Cheddad, Joan Condell, Kevin Curran, Paul McKevitt. Digital image steganography: Survey and analysis of current methods [en línea]. El Sevier (Ed), Journal of Signal Processing, 90(3), 727-752, 2009. [Consultado Junio 2014]
Disponible en:
http://www.sciencedirect.com/science/article/pii/S0165168409003648

[3] David Frith, Steganography approaches, options, and implications [en línea]. Ed. Elsevier Ltd., Network Security, 2007. [Consultado marzo 2016.]
Disponible en:
http://www.sciencedirect.com/science/article/pii/S1353485807700715

[4] Navas, Sergio; Rodríguez, Gustavo; Eterovic, Jorge; "Aplicación del filtro de Canny a la esteganografía digital". XVI WICC, ISBN 978-950-34-1084-4. Ushuaia, 2014. Disponible en: http://sedici.unlp.edu.ar/handle/10915/40706

[5] Rodríguez, Gustavo; Navas, G. Sergio; Eterovic, Jorge; "Selección óptima de métodos de sustitución en aplicaciones esteganográficas". XVII WICC, ISBN 978-987-633-134-0. Salta, 2015. Disponible en :
http://sedici.unlp.edu.ar/handle/10915/45218

[6] Rodríguez, Gustavo; Navas, G. Sergio; "Esteganografía: Sustitución LSB 1 bit utilizando MatLab". XVIII WICC. ISBN 978-950-698-377-2. Entre Ríos, 2016. Disponible en :
http://sedici.unlp.edu.ar/handle/10915/52766.