Revista de la Facultad de Contaduría y Ciencias Administrativas

OCULTAMIENTO DE INFORMACIÓN CONFIDENCIAL EN IMÁGENES BMP Y AUDIO WAV MEDIANTE EL MÉTODO LSB

LSB METHOD FOR HIDING SENSITIVE INFORMATION IN BMP AND WAV FILES

PEDRO CHAVEZ LUGO[1], GUSTAVO ALFONSO GUTIERREZ CARREON[1], SALVADOR ANTELMO CASANOVA VALENCIA[1].

UNIVERSIDAD MICHOACANA DE SAN NICOLAS DE HIDALGO [1]

México

Recibido el 3 de Agosto de 2020; Aceptado el 20 de Diciembre de 2020; Disponible en Internet el 31 de Diciembre de 2020.

E-mail de Contacto: pedro.chavez@umich.mx
© Universidad Michoacana de San Nicolás de Hidalgo (México)
Vol. 5, Nº 10, Pág. 92-. ISSN: 2448-6051

Av. Gral. Francisco J. Múgica S/N
Edificio AII C.P. 58030
Ciudad Universitaria
Morelia, Michoacán, México.
Tel. y Fax (443) 3-16-74-11
Email: rfcca@umich.mx
Web: http://rfcca.umich.mx.

Resumen— En la actualidad debido al uso constante de las TIC, las personas y organizaciones hacen y harán uso de herramientas de seguridad para proteger su información confidencial. En el caso de que la información confidencial sea expuesta, podría implicar pérdidas monetarias e incluso la afectación a la imagen y reputación del propietario, persona u organización. En este trabajo, se presenta la especificación para modificar e incluir el nombre, extensión, longitud y contenido de un archivo confidencial, dentro de los datos de una imagen en formato BMP y un audio en formato WAV, mediante el método de incrustación en bit menos significativo - LSB. El método utilizado solo modifica el bit menos significativo de cada byte de dato, por lo cual, la modificación de los datos de una imagen en formato BMP no debe generar cambios notables que sean perceptibles al ojo humano. Para el caso de un audio en formato WAV, la modificación de los datos debe ser imperceptible al oído humano. La idea principal de este trabajo es fomentar la cultura de la seguridad de la información, mostrando las características, ventajas y desventajas del método LSB, con el fin de poder hacer un uso correcto del mismo.

Palabras Clave— *información confidencial; ocultamiento; WAV; BMP; LSB; cultura de la seguridad de la información*

Abstract— Nowadays the constant use of computers, mobile devices and Internet, people and organizations use security tools to protect their sensitive information. When the sensitive information is exposed, it could entail monetary losses and even damage to the image and reputation of the owner. This work shows a specification to modify and include name, extension, length and content of a sensitive file, within the data of BMP and WAV files using the least significant bit embedding method - LSB. The method used only modifies the least significant bit of each data byte, therefore, the modification of the data of an image in BMP format should not generate noticeable changes that are perceptible to the human eye. In the case of audio in WAV format, the modification of the data must be imperceptible to the human ear. The main idea of this work is to promote the culture of information security, showing the characteristics, advantages and disadvantages of the LSB method, in order to make correct use of it.

Keywords— *sensitive information; hiding; Wav and Bmp files; security culture*

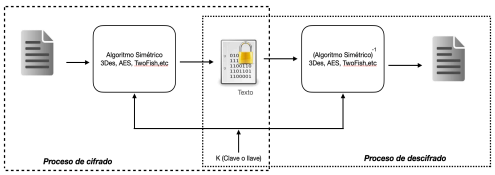
JEL CODE— *M00, L00, L96*

INTRODUCCIÓN

La seguridad de la información es un aspecto imprescindible que toda organización o persona debe procurar. Existe una gran cantidad de evidencia que demuestra la afectación que se tiene cuando la información confidencial es comprometida. Un ejemplo, es el ciberataque sufrido por Equifax en el año 2017, en el cual se comprometieron los datos de 143 millones de estadounidenses (El País, 2017). Otro ejemplo, es el caso del ciberataque a PEMEX ocurrido en el mes de noviembre de 2019 (Forbes, 2019; El Economista, 2019). Este ciberataque se realizó con un ransomware, el cual tiene la finalidad de ingresar sin autorización a un sistema informático, gracias a una vulnerabilidad y realizar diversas tareas, como pueden ser, el robo de información, afectar un servicio, monitorear actividades, etc. No solo las organizaciones han sido afectadas, las personas también han sido víctimas, perdiendo parte o toda la información relacionada con sus actividades diarias (Expansión, 2017). Se estima que en el año 2017 la expansión del ransomware WannaCry, provocó pérdidas de aproximadamente 4,000 millones de dólares. Sobran los ejemplos y las razones para proteger la información confidencial. La cultura de la seguridad de la información debe estar presente en todas las personas, es de vital importancia fomentar y educar con la finalidad de salvaguardar los activos informáticos más importantes para las personas y organizaciones.

Tradicionalmente, la criptografía es una de las técnicas mayormente empleadas para proteger a la información confidencial. Su función es transformar la información para que no se pueda interpretar de forma directa. En la Figura 1, se muestra el proceso de cifrado y descifrado simétrico. Un documento se cifra mediante un algoritmo y una clave, el algoritmo puede hacer la transformación combinando la transposición y la substitución (Douglas, R. Stinson 2006; Aumasson, J. P. 2017). El proceso de descifrado se realiza con la aplicación inversa del algoritmo de cifrado y la clave utilizada. Un aspecto importante a comentar del algoritmo de cifrado simétrico son los modos de operación, los cuales pueden implicar el éxito o fracaso del cifrado resultante. Este aspecto no es abordado en este trabajo.

Figura 1. *Proceso de cifrado y descifrado simétrico.*

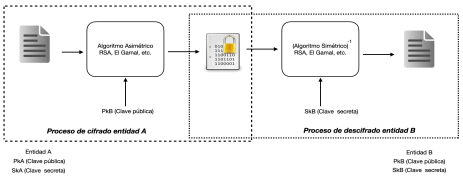


Fuente: *Elaboración propia.*

La criptografía asimétrica es otra opción para proteger la información, cada participante tendrá un par de claves, una secreta y una pública (Aumasson, J. P. 2017; Ferguson, N., Schneier, B., et al., 2010).

La clave pública se obtiene mediante la aplicación de operaciones matemáticas sobre la clave secreta. Las operaciones matemáticas permiten su aplicación en un solo sentido, de tal forma que es matemáticamente imposible obtener la clave secreta a partir de la clave pública. En la Figura 2, se muestra el proceso de cifrado y descifrado asimétrico. Todos los aspectos relacionados con la criptografía asimétrica como lo son la generación de claves, longitudes de clave, etc., no son abordados en este trabajo.

Figura 2. *Proceso de cifrado y descifrado asimétrico.*



Fuente: *Elaboración propia.*

La esteganografía, es otra técnica utilizada para proteger la información. A diferencia de la criptografía, la cual transforma la información, en la esteganografía se intenta ocultar la existencia de la información (Cox, I., Miller, M., Bloom, J., Fridrich, J., et al., 2007; Carracedo Gallardo, J. 2004)). Un ejemplo tan sencillo del uso de la esteganografía es el uso de la tinta invisible sobre un papel, que al momento de aplicarle algún fluido revelara la presencia de la información.

En la actualidad los procesos de la esteganografía se realizan sobre archivos de imágenes (Tao, J., Li, S., Zhang, X., & Wang, Z. 2018), audio y video (Sadek, M. M., Khalifa, A. S., & Mostafa, M. G. 2015; Liu, Y., Liu, S., Wang, Y., Zhao, H., & Liu, S. 2019; ussain, M., Wahab, A. W. A., Idris, Y. I. B., Ho, A. T., & Jung, K. H. 2018), de tal forma que al incrustar información en un archivo de imagen, la modificación no sea perceptible al ojo humano o mientras que al incrustar información en un archivo de audio, la modificación no sea perceptible al oído humano (Pradhan, A., Sahu, A. K., Swain, G., & Sekhar, K. R. 2016). La Figura 3, muestra el objetivo de la esteganografía sobre imagen y audio.

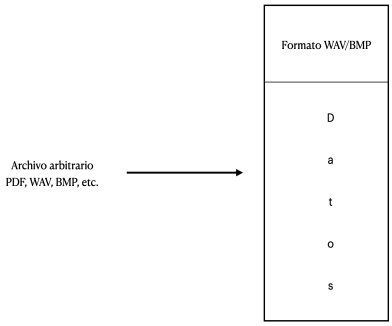
Figura 3. *Esteganografía sobre imagen y audio.*

Fuente: *Elaboración propia.*

MÉTODO DE INCRUSTACION EN BIT MENOS SIGNIFICATIVO - LSB

En el proceso de incrustación de información en archivos de imagen y audio, mediante el método de incrustación en bit menos significativo- LSB[1], es necesario tener presente que tales archivos son la representación de un formato. Tales formatos tienen características comunes, como lo son, una firma que los identifica, el tamaño y en el caso de imágenes, la información sobre la representación de colores, entre otros aspectos. En el caso de audio, se tiene información sobre el número de canales, audio estéreo o mono, etc. En ambos casos imagen o audio, después de las características del formato se continúa con los datos. La Figura 4, muestra la idea general de incrustar información en un archivo de audio en formato WAV[2] o imagen en formato BMP[3].

Figura 4. *Incrustación de información en los datos de una imagen o un audio.*



Fuente: *Elaboración propia.*

Los formatos BMP y WAV, fueron seleccionados por utilizar algoritmos sencillos de compresión y por tal razón, se obtienen archivos de gran tamaño. A diferencia, por ejemplo, de un formato JPEG[4] que utiliza un algoritmo de compresión complejo, el cual ofrece reducciones de tamaño considerable.

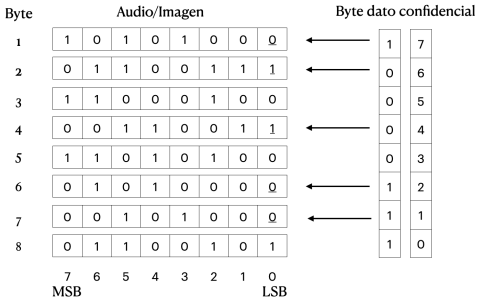
DESCRIPCIÓN DEL MÉTODO LSB

El método LSB, corresponde a la modificación del bit menos significativo, el cual corresponde a un método simple y sencillo de fácil implementación en código de programación. La Figura 5, muestra el

[1] Del Inglés Least Significant Bit.
[2] Del Inglés Wave.
[3] Del Inglés Windows BitMap.
[4] Del Inglés Joint Photographic Expert Group.

proceso de inserción de 1 Byte de dato confidencial en 8 Bytes de datos de una imagen BMP o audio WAV. Los bits menos significativos de los Bytes de dato 1, 2, 4, 6 y 7 de la imagen o audio, son modificados por los bits 7, 6, 4, 2 y 1 del byte de dato confidencial. Por el contrario, los bits menos significativos de los bytes de dato 3, 5 y 8, no son cambiados, ya que se tiene el mismo valor que los bits 5, 3 y 0 del byte de dato confidencial. Los bits menos significativos que son cambiados en los datos de la imagen o audio se encuentran subrayados.

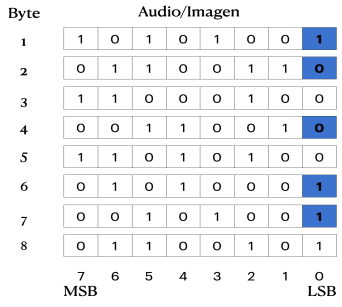
Figura 5. *Método LSB.*



Fuente: *Elaboración propia*

La Figura 6, muestra el resultado de la inserción del byte 10000111 en los bytes de dato de un audio o imagen, marcado de color azul los bits menos significativos que fueron modificados.

Figura 6. *Resultado de la aplicación del Método LSB.*



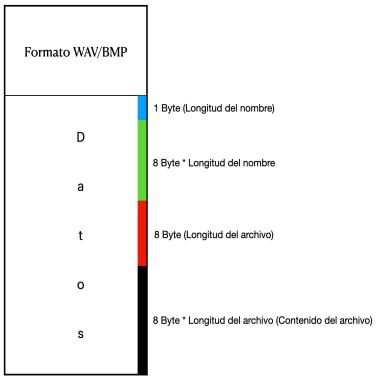
Fuente: *Elaboración propia.*

APLICACIÓN DEL MÉTODO LSB

En el proceso de aplicación del método LSB, se consideró incrustar la longitud del nombre del archivo, el nombre del archivo, la longitud del archivo y el contenido del archivo. En la Figura 7, se muestra la estructura de aplicación del proceso de incrustación. El color azul representa un byte incrustado que corresponde a la longitud del nombre, el color verde, corresponde al resultado de la multiplicación de 8 bytes por la longitud del nombre para incrustar los caracteres del nombre del archivo (máximo 256 caracteres). El color rojo, representa la incrustación de la longitud del archivo en 8 bytes (tamaño máximo 1.844674407370955e19 bytes) y finalmente el color negro, representa la incrustación del contenido del archivo, correspondiente a la información confidencial.

La relación de incrustación es 1 a 8, es decir, por cada byte de dato confidencial es necesario 8 bytes de datos de la imagen o audio. Previo a la incrustación es necesario verificar si es posible realizar el proceso en la imagen o audio. La incrustación será posible si la longitud de los datos de la imagen o audio cumple la relación por cada byte de dato confidencial es necesario 8 bytes de dato. Debido a la simplicidad del método LSB basada en la estructura de aplicación, es necesario combinarla con técnicas criptográficas para salvaguardar a la información confidencial.
Figura 7.

Figura 7. *Aplicación del Método LSB.*



Fuente: *Elaboración propia.*

MÉTODO LSB EN IMÁGENES BMP

Para la incrustación de información en una imagen en formato BMP, se utilizó la fotografía de un felino, tal como se puede observar en la Figura 8. El archivo de imagen tiene una longitud de 72,000,122 bytes y el archivo de datos confidenciales con una longitud de 212,989 bytes y corresponde a un contenido en formato PDF. En el proceso de incrustación, fue necesario cambiar 853,110 bits menos significativos de la imagen, correspondiente a menos del .2 % del contenido y el 99.852% no fue modificado. En este caso, se puede observar, una incrustación de información confidencial que ocupa una proporción

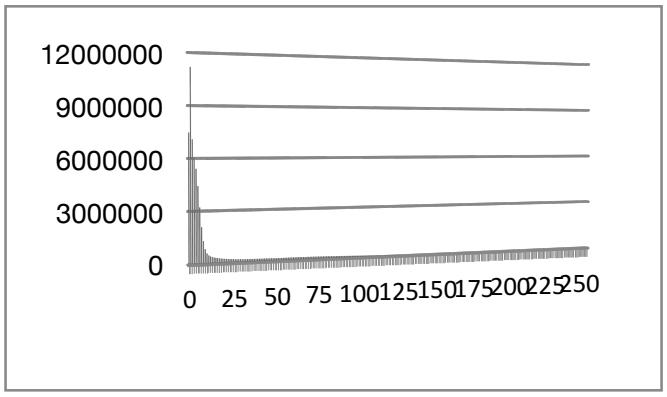
mínima de los datos de la imagen. Los cambios realizados a la imagen, no son perceptibles a la vista, las razones de esto, es que la información confidencial incrustada es muy pequeña y los bits cambiados no implican una alteración considerable. Pensando en términos del color negro, la modificación del bit menos significativo producirá un negro más claro, pero no tan claro para ser notable.

Figura 8. *Aplicación del Método LSB en una imagen BMP.*

Fuente: *Elaboración propia.*

La Gráfica 1, muestra el histograma de la imagen original en formato BMP, como se puede observar, existe un dato predominante con mayor conteo. A partir del dato 25 se observa que el conteo tiende a ser muy cercano a cero.

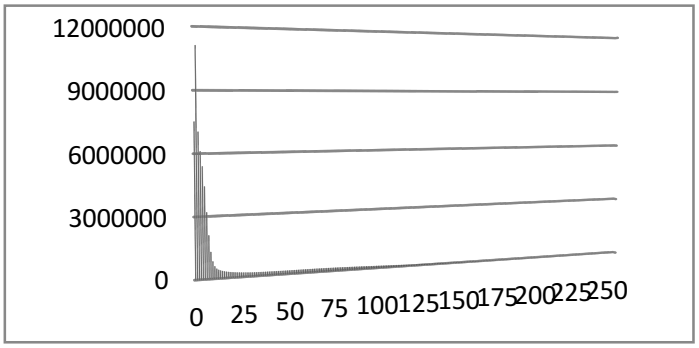
Gráfica 1. *Histograma de la imagen original.*



Fuente: *Elaboración propia*

La Gráfica 2, muestra el histograma de la imagen resultante de la incrustación de datos y a simple vista, se observa que la distribución es casi la misma de la Gráfica 1.

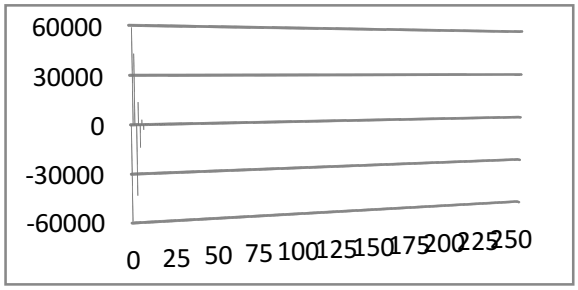
Gráfica 2. *Histograma de la imagen con datos incrustados.*



Fuente: *Elaboración propia.*

La Gráfica 3, muestra la diferencia entre los datos de la imagen y el resultado de la incrustación. En este caso es posible notar la diferencia de los histogramas de las Gráficas 1 y 2. Tres datos tienen mayor conteo en la incrustación y el resto tiene mayor presencia en la imagen original. Aproximadamente 80 datos tienen el mismo conteo en la imagen original y en la imagen resultante de la incrustación.

Gráfica 3. *Diferencia de datos (imagen original e imagen con datos incrustados).*



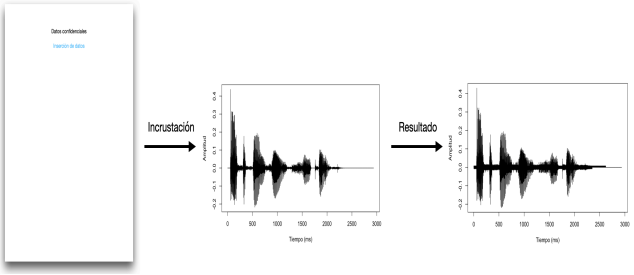
Fuente: *Elaboración propia.*

MÉTODO LSB EN AUDIO WAV

Para la incrustación de información en un audio en formato WAV, se utilizó un audio de duración de casi 3 segundos. En la Figura 9, se muestra la incrustación de información confidencial en un audio en formato WAV. El archivo de audio tiene una longitud de 94,226 bytes y el archivo de datos confidenciales con una longitud de 10,448 bytes y corresponde a un contenido en formato PDF. En el proceso de incrustación fue necesario cambiar 40,457 bits menos significativos del audio, correspondiente

al 5% del contenido y el 95 % restante no fue modificado.

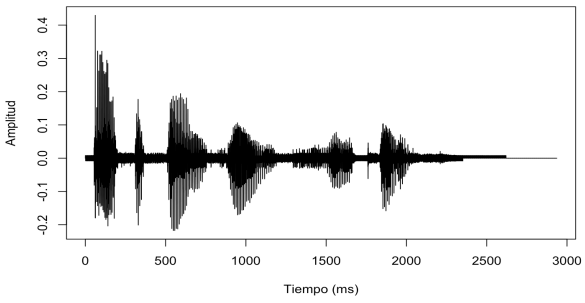
Figura 8. *Aplicación del Método LSB en un audio WAV.*



Fuente: *Elaboración propia.*

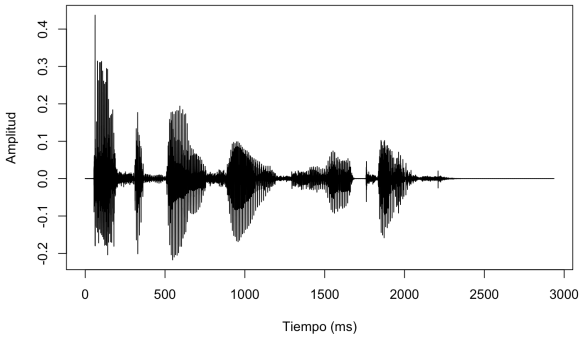
En la Gráfica 4 y 5, se muestran las señales del audio original y el audio resultante de la incrustación de datos confidenciales. Se pueden observar incrementos de amplitud en el audio resultante, lo cual se traduce en aumento de volumen. En este caso existe una percepción de ruido en el audio resultante de la incrustación comparado con el audio original. En este caso, existe una distinción una distinción entre ambos audios.

Gráfica 4. *Señal de audio original.*



Fuente: *Elaboración propia.*

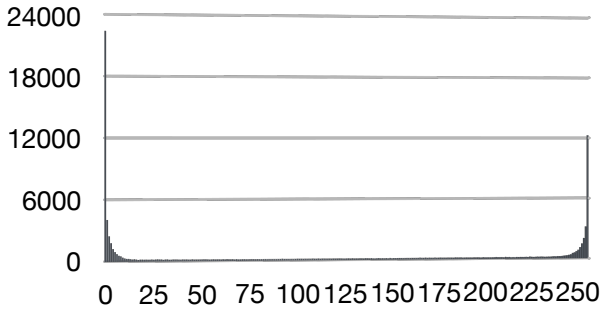
Gráfica 5. *Señal de audio con datos confidenciales incrustados.*



Fuente: *Elaboración propia.*

La Gráfica 6, muestra el histograma del audio original en formato WAV, como se puede observar, existe un dato predominante con el mayor conteo de casi 24,000. El resto de los datos tiene un conteo inferior a 12,000.

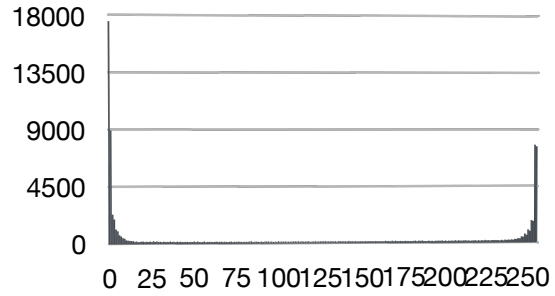
Gráfica 6. *Histograma del audio original.*



Fuente: *Elaboración propia.*

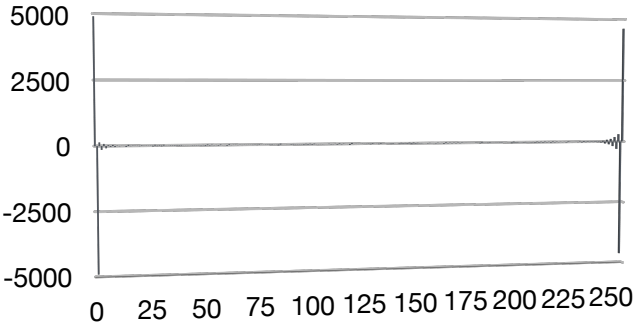
La Gráfica 7, muestra el histograma del audio con los datos confidenciales incrustados, como se puede observar, existe un dato predominante con mayor conteo, cercano a 18,000, mientras que el dato de mayor conteo ronda los 24,000. Esto implica un cambio significativo en los datos. La Gráfica 8, muestra la diferencia entre los datos del audio y el resultado de la incrustación. En este caso es posible notar la diferencia de los histogramas de las Gráficas 6 y 7. Dos datos tienen mayor conteo en la incrustación y también dos datos tienen mayor conteo en el audio original.

Gráfica 7. *Histograma del audio con los datos confidenciales incrustados.*



Fuente: *Elaboración propia.*

Gráfica 8. *Diferencia de datos (audio original y audio con datos incrustados).*



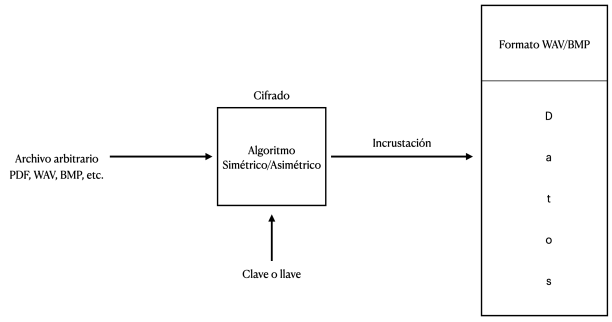
Fuente: *Elaboración propia.*

MAYOR SEGURIDAD CON EL MÉTODO LSB

La aplicación del método LSB para incrustar información confidencial en imágenes en formato BMP y audio WAV, es sencillo y fácil de codificar en un lenguaje de programación. Debido a su sencillez, también es fácil la extracción de los datos de una manera sencilla al conocer la especificación de incrustación. Para obtener un aumento de seguridad en los datos incrustados, se recomienda una etapa previa a la incrustación, correspondiente a la aplicación de la criptografía simétrica o asimétrica. De esta forma, los datos confidenciales serán cifrados con algún algoritmo simétrico y/o asimétrico e incrustados en una imagen o un audio.

En la Figura 9, se muestra la idea de combinar la criptografía y la esteganografía, en la búsqueda de un fortalecimiento de la seguridad de la información confidencial. Esta combinación dará como resultado una mejor estrategia en la protección de información confidencial, ya que con la criptografía se transformarán los datos confidenciales y con la esteganografía se intentará ocultar la existencia de los mismos. Existen algunas herramientas de código abierto que permiten aplicar esta estrategia de combinación, tal es el caso de Steghide (Stefan Hetzl 2003). Steghide, es un programa con licencia de GPL[5], codificado en C++, que permite hacer la incrustación sobre ciertos formatos de audio y video.

Figura 9. *Combinación de criptografía simétrica/asimétrica con esteganografía.*



Fuente: *Elaboración propia.*

CONCLUSIONES

Es imprescindible para todas las personas tener los conocimientos básicos sobre la cultura de la seguridad de la información. No importa la profesión o las actividades laborales que se realicen, es necesario contar con los principios básicos de la seguridad de la información. Conocer tales principios permitirá tomar las mejores decisiones de la selección y uso de herramientas de seguridad para proteger de la mejor manera la información confidencial. Para el caso de la esteganografía mediante el método LSB, saber que la modificación se hace a nivel de bit sobre el bit menos significativo, que numéricamente solo hace un decremento de una unidad. También es necesario tener presente en la aplicación de este método, la necesidad de combinarlo con la criptografía simétrica o asimétrica o incluso con técnicas de la criptografía clásica (métodos de transposición y substitución polialfabética), para aumentar el nivel de protección de la información confidencial.

Una de las aplicaciones para esta propuesta, es la protección de los documentos para la firma electrónica otorgados por el Servicio de Administración Tributaria. Ya que, al ser comprometidos, sería posible que sin autorización alguien emita facturas electrónicas, generando

[5] Del Inglés General Public License.

problemas al contribuyente en la aclaración de tales emisiones.

REFERENCIAS

El País, (2017, September 8). Un ciberataque masivo roba los datos de 143 millones de estadounidenses. EL PAÍS. https://elpais.com/tecnologia/2017/09/08/actualidad/1504856601_125518.html

Forbes, (2019, November 27). Las consecuencias de un ciberataque: caso Pemex. Forbes México. https://www.forbes.com.mx/las-consecuencias-de-un-ciberataque-caso-pemex/

El Economista (2019, November 12). El ciberataque a Pemex fue una acción dirigida: ESET. El Economista. https://www.eleconomista.com.mx/tecnologia/El-ciberataque-a-Pemex-fue-una-accion-dirigida-ESET-20191112-0089.html

Expansión. (2017, May 17). WannaCry costará al mundo 4,000 millones de dólares. https://expansion.mx/tecnologia/2017/05/16/wannacry-costara-al-mundo-4-000-millones-de-dolares

Douglas, R. Stinson (2006). Cryptography Theory and Practice Third Edition.

Aumasson, J. P. (2017). Serious cryptography: a practical introduction to modern encryption. No Starch Press.

Ferguson, N., Schneier, B., & Kohno, T. (2010). Cryptography engineering. Design Princi.

Cox, I., Miller, M., Bloom, J., Fridrich, J., & Kalker, T. (2007). Digital watermarking and steganography. Morgan kaufmann.

Carracedo Gallardo, J. (2004). Seguridad en redes telemáticas.

Stefan Hetzl . (2003, October 3). Steghide. Steghide. http://steghide.sourceforge.net

The GNU Privacy Guard. The GNU Privacy Guard. https://gnupg.org

Sadek, M. M., Khalifa, A. S., & Mostafa, M. G. (2015). Video steganography: a comprehensive review. Multimedia tools and applications, 74(17), 7063-7094.

Liu, Y., Liu, S., Wang, Y., Zhao, H., & Liu, S. (2019). Video steganography: A review. Neurocomputing, 335, 238-250.

Hussain, M., Wahab, A. W. A., Idris, Y. I. B., Ho, A. T., & Jung, K. H. (2018). Image steganography in spatial domain: A survey. Signal Processing: Image Communication, 65, 46-66.

Tao, J., Li, S., Zhang, X., & Wang, Z. (2018). Towards robust image steganography. IEEE Transactions on Circuits and Systems for Video Technology, 29(2), 594-600.

Pradhan, A., Sahu, A. K., Swain, G., & Sekhar, K. R. (2016, May). Performance evaluation parameters of image steganography techniques. In 2016 International Conference on Research Advances in Integrated Navigation Systems (RAINS) (pp. 1-8). IEEE.