ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA

TÉCNICAS, MECANISMOS DE SEGURIDAD Y ENCRIPTACIÓN DE LA INFORMACIÓN Y DESARROLLO DE APLICACIONES

ESTUDIO DE MECANISMOS DE DETECCIÓN DE MENSAJES OCULTOS UTILIZANDO MODELOS ESTADÍSTICOS

TRABAJO DE INTEGRACIÓN CURRICULAR PRESENTADO COMO REQUISITO PARA LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN TECNOLOGÍAS DE LA INFORMACIÓN

GUILLERMO ENRIQUE NARANJO VITERI

guillermo.naranjo@epn.edu.ec

DIRECTOR: MSc. WILLAMS FERNANDO FLORES CIFUENTES

fernando.flores@epn.edu.ec

DMQ, febrero 2022

CERTIFICACIONES

Yo, Guillermo Enrique Naranjo Viteri, declaro que el trabajo de integración curricular aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

GUILLERMO ENRIQUE NARANJO VITERI

Certifico que el presente trabajo de integración curricular fue desarrollado por Guillermo Enrique Naranjo Viteri, bajo mi supervisión.

MSc. WILLAMS FERNANDO FLORES CIFUENTES

DECLARACIÓN DE AUTORÍA

A través de la presente declaración, afirmamos que el trabajo de integración curricular aquí descrito, así como el (los) producto(s) resultante(s) del mismo, son públicos y estarán a disposición de la comunidad a través del repositorio institucional de la Escuela Politécnica Nacional; sin embargo, la titularidad de los derechos patrimoniales nos corresponde a los autores que hemos contribuido en el desarrollo del presente trabajo; observando para el efecto las disposiciones establecidas por el órgano competente en propiedad intelectual, la normativa interna y demás normas.

GUILLERMO ENRIQUE NARANJO VITERI

MSc. WILLAMS FERNANDO FLORES CIFUENTES

DEDICATORIA

Para mis padres, hermano y abuelita, sin ellos nada de esto fuera posible.

Para mis abuelitos, Guillermo y Enrique, donde quiera que estén.

Para Naty, lo más hermoso de mi vida.

AGRADECIMIENTO

A mi familia, en especial a mis padres, por sus sacrificios a lo largo de este tiempo.

A mis amigos, Jonathan, Santiago, Daniel y Viviana, por estar conmigo en todos estos años de universidad.

Al Ing. Fernando Flores, por hacerme parte del desarrollo de este proyecto.

ÍNDICE DE CONTENIDO

RESUMEN

El presente trabajo de integración curricular se propone estudiar los métodos para detectar mensajes ocultos en archivos multimedia haciendo uso de modelos estadísticos.

En el primer capítulo se realizará un estudio de los fundamentos de la ocultación de información que incluye una breve revisión de criptografía y marcación de agua. Además, se describe ampliamente el concepto de esteganografía, su evolución hasta la actualidad, los términos básicos para entender un proceso esteganográfico, clasificación según el archivo a usar para ocultar la información, las diferentes técnicas esteganográficas y por último se realiza una breve descripción de las características principales que afectan a la esteganografía.

En el segundo capítulo se presenta la situación actual del estegoanálisis, se describen los problemas comunes al realizar un proceso de recuperar un mensaje oculto desde el punto de vista probabilístico como lo es la heurística. Además, se describen los modelos estadísticos de primer orden usados para detectar mensajes ocultos cuando se ha utilizado la técnica del Bit Menos Significativo (LSB) y los modelos estadísticos de orden superior que se usan en el estegoanálisis, en este último caso se da una breve introducción a las Máquinas de Vectores de Soporte (SVM) que son usadas en conjunto a los modelos estadísticos para detectar mensajes ocultos.

En el tercer y último capítulo se presentan los resultados del estudio, que incluye la factibilidad del uso de los modelos estadísticos descritos anteriormente. Por último, se presentan las respectivas conclusiones y recomendaciones que deja el desarrollo de este trabajo de integración curricular

PALABRAS CLAVE: esteganografía, estegoanálisis, modelo, estadística, mensaje oculto.

ABSTRACT

This curricular integration work pretends to study the methods to detect hidden messages in multimedia files using statistical models.

In the first chapter, a study of the fundamentals of information hiding will be carried out, including a brief review of cryptography and watermarking. In addition, the concept of steganography is widely described, its evolution to the present, the basic terms to understand a steganographic process, classification according to the file used to hide the information, the different steganographic techniques and finally a brief description of the main characteristics that may affect steganography.

In the second chapter, the current situation of steganalysis is presented, the common problems are described when using a process of recovering a hidden message from the probabilistic point of view, such as heuristics. In addition, the first order statistical models used to detect hidden messages when the Least Significant Bit (LSB) technique has been used and the higher order statistical models used in steganalysis are described, in this last model we give a brief overview to Support Vector Machines (SVM) that are used in conjunction with statistical models to detect hidden messages.

The third and last chapter presents the results of the study, which includes the feasibility of using the statistical models described above. Finally, the respective conclusions and recommendations are presented left by the development of the present curricular integration work.

KEYWORDS: steganography, steganalysis, model, statistics, hidden message.

1 INTRODUCCIÓN

En la actualidad la comunicación es una necesidad básica y todo el mundo quiere seguridad de sus datos al momento de realizar este proceso. En nuestra vida diaria, utilizamos muchas vías como Internet o el teléfono para transferir y compartir información, pero son seguras hasta un cierto nivel. Para compartir la información de forma oculta se podrían utilizar dos técnicas. Estos mecanismos son la criptografía y la esteganografía.

En criptografía, el mensaje se modifica de forma cifrada con la ayuda de una clave de cifrado que solo conocen el remitente y el receptor. Nadie puede acceder al mensaje sin utilizar la clave de cifrado. Sin embargo, la transmisión de un mensaje cifrado puede despertar fácilmente las sospechas del atacante y, por lo tanto, puede ser interceptado, atacado o descifrado violentamente. Para superar las deficiencias de las técnicas criptográficas, las técnicas de esteganografía han sido desarrolladas.

La esteganografía es la ciencia de comunicar de tal manera que oculta la existencia de la comunicación. Así, la esteganografía oculta la existencia de datos para que nadie pueda detectar su presencia. En esteganografía el proceso de ocultar información dentro de cualquier contenido multimedia como imagen, audio, video se conoce como "Incrustación". Para aumentar la confidencialidad de la comunicación de datos, se pueden combinar ambas técnicas.

Las herramientas de esteganografía (ocultación de datos) están fácilmente disponibles para todo el público y existe la necesidad de diseñar herramientas de estegoanálisis que detecten la presencia de mensajes ocultos. Si bien la investigación en esteganografía está muy avanzada, la de estegoanálisis no ha sido muy tomada en cuenta. La razón principal de esto es que, en toda su generalidad, el estegoanálisis es un problema algo difícil de resolver; por ejemplo, se desconocen los datos de la cubierta original, se desconoce la tasa de ocultación (si los datos están ocultos) y la cantidad de esquemas de esteganografía es grande. Desafortunadamente, incluso los enfoques existentes más prometedores y el marco de aprendizaje supervisado tienen inconvenientes que limitan su uso práctico.

1.1 OBJETIVO GENERAL

El objetivo general de este trabajo es el estudiar los mecanismos de detección de mensajes ocultos utilizando modelos estadísticos.

1.2 OBJETIVOS ESPECÍFICOS

Los objetivos específicos de este trabajo son:

- Definir los conceptos fundamentales de la esteganografía y estegoanálisis.
- Realizar el estudio de diferentes mecanismos usados para ocultar información en archivos multimedia.
- Describir los modelos estadísticos usados en la detección de información oculta.
- Discutir los resultados que se pueden obtener a partir de implementar los métodos de detección de información oculta.

1.3 ALCANCE

El trabajo de integración curricular realizará el estudio de los mecanismos de detección de mensajes ocultos que usan modelos estadísticos. Para ello se describirán los fundamentos de ocultación de la información como lo es la esteganografía

Primero se describirán los fundamentos sobre ocultación de la información, esto incluye dar una amplia descripción sobre esteganografía, que incluye sus términos básicos, clasificación y características.

Se presentarán dos mecanismos de esteganografía para archivos multimedia (audio, imágenes) que son de los más conocidos en la actualidad para comprender como se realiza la ocultación de la información.

Se presentará la situación actual del estegoanálisis y los principales problemas que tiene para ser implementado, se va a dar énfasis en describir problemas probabilísticos como son la heurística. Una vez planteados estos inconvenientes, se estudiará los modelos estadísticos empleados para la detección de mensajes ocultos y se propondrá aquellos que sean los más desarrollados en la actualidad.

Por último, en base al estudio desarrollado se dará a conocer la factibilidad de usar los modelos estudiados y se dará a conocer alternativas que pueden servir para futuros trabajos en el área.

Es de suma importancia mencionar que este trabajo no cuenta con un producto final demostrable (implementación), debido a que se realizará específicamente el estudio de modelos para detección de mensajes ocultos que ya han sido desarrollados previamente.

1.4 MARCO TEÓRICO

1.4.1 ASEGURAMIENTO DE LA INFORMACIÓN

1.4.1.1 Descripción

Existen varias formas de ocultar información. Todas ofrecen algo de sigilo, pero no todos los mecanismos son tan poderosos como los demás. Algunos mecanismos proporcionan un mimetismo increíble con la ayuda de los usuarios y otros son en gran parte automáticos. Algunas técnicas se pueden combinar con otras para proporcionar múltiples capas de seguridad. Todos ellos explotan un poco de aleatoriedad, incertidumbre, etc. A continuación, la figura 1.1 presenta las técnicas más comunes en la actualidad para asegurar la información. [1]

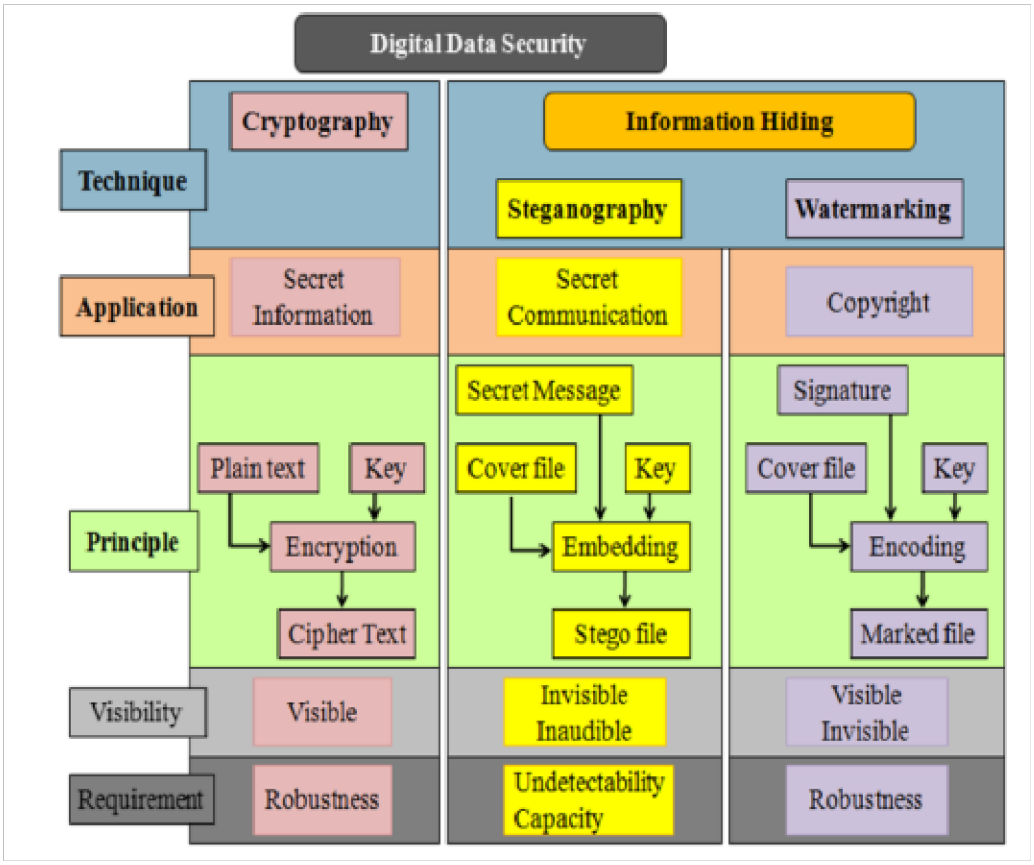

Figura 1.1. Métodos para asegurar la información digital

1.4.2 CRIPTOGRAFÍA

1.4.2.1 Definición

Es el proceso de cifrar la información al momento de la comunicación para hacerla ilegible para los observadores sin claves. Esta tecnología a veces se denomina codificación de datos. Existen dos tipos de criptografía: simétrica y asimétrica; en la simétrica la llave es compartida por ambos extremos de la comunicación y en el cifrado asimétrico existe una

llave pública que se usa para cifrar y una llave privada usada para descifrar la información. [2] La figura 1.2 demuestra como son los dos tipos de criptografía.

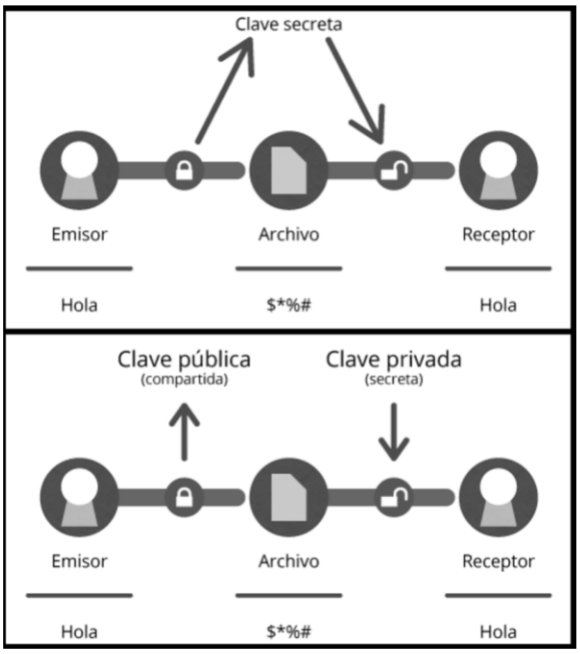

Figura 1.2. Diferencia entre cifrado simétrico y asimétrico

1.4.3 MARCACIÓN DE AGUA

1.4.3.1 Definición [3]

Una marca de agua es un patrón de bits insertado en un medio digital que puede identificar al creador o usuarios autorizados. Las marcas de agua digitales, a diferencia de las marcas de agua visibles impresas tradicionales, están diseñadas para ser invisibles para los espectadores. Los bits incrustados en una imagen están dispersos por todas partes para evitar su identificación o modificación. Por lo tanto, una marca de agua digital debe ser lo suficientemente robusta para sobrevivir a la detección, compresión y otras operaciones que podrían aplicarse a un documento.

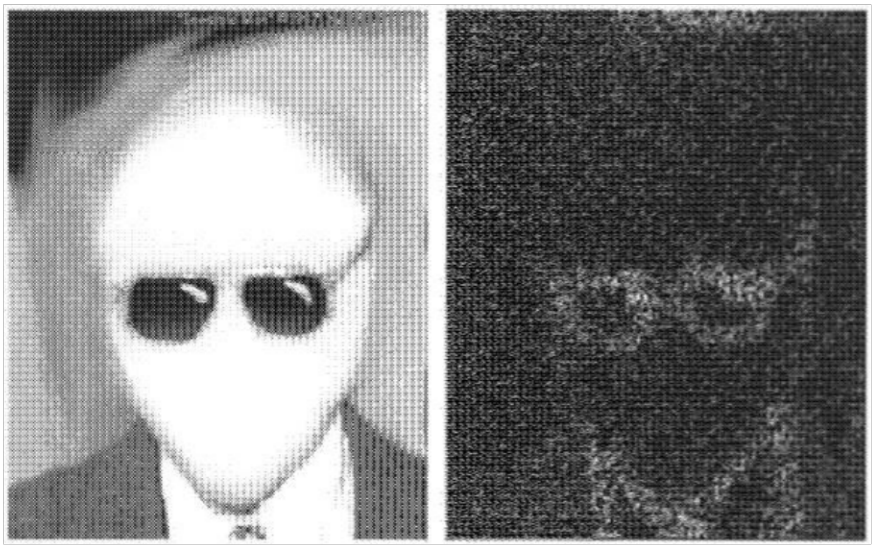

Figura 1.3. Ejemplo de imagen comercial con marca de agua

1.4.4 ESTEGANOGRAFÍA

1.4.4.1 Definición

La esteganografía digital tiene como objetivo ocultar información digital en canales encubiertos para que un usuario pueda ocultar y evitar la detección del mensaje oculto. Para la transmisión segura de datos confidenciales, archivos multimedia como audio, video e imágenes se utilizan como fuentes de cobertura para ocultar los datos. [4] Esto se hace principalmente porque los archivos multimedia son relativamente más grandes que el mensaje secreto a ocultar.

1.4.4.2 Historia [5] [6] [7]

La palabra esteganografía está compuesta por las palabras griegas steganos, que significa "cubierto" y graphia, que significa "escritura". En otras palabras, la esteganografía es el arte de la comunicación encubierta donde existe un mensaje secreto.

En la antigua Grecia algunos mensajeros solían afeitarse la cabeza, luego escribían un mensaje en su cabeza. Una vez que el mensaje había sido escrito, se dejaba que el cabello volviera a crecer. Después de que el cabello creciera, el mensajero era enviado a entregar el mensaje, el destinatario afeitaría el cabello del mensajero para ver el mensaje secreto.

Otro método utilizado en Grecia era que alguien quitaba la cera de una tableta que estaba cubierta con cera, escribía un mensaje debajo de la cera y luego volvía a aplicar la cera. El destinatario del mensaje simplemente quitaría la cera de la tableta para ver el mensaje.

Ocultar mensajes en un texto escrito se llama esteganografía lingüística o acrósticos. Los acrósticos eran un método esteganográfico antiguo muy popular. Para incrustar una firma

única en su trabajo y este sea plagiado, algunos poetas codificaron mensajes secretos como letras iniciales de oraciones o las iniciales de su nombre en un poema. Uno de los ejemplos más conocidos es Amorosa visione de Giovanni Boccaccio. Boccaccio codificó tres sonetos (más de 1500 letras) en las letras iniciales del primer verso de cada terceto de otros poemas.

La rejilla de Cardano fue popularizada en los años 1500, donde por la época del Renacimiento muchos de los mensajes tenían que ser enviados en secreto. Las letras del mensaje secreto forman un patrón aleatorio dentro del texto al que se puede acceder simplemente colocando una máscara (rejilla) sobre el mismo. La máscara desempeña el papel de una clave esteganográfica secreta que debe compartirse entre las partes que se comunican (esteganografía de llave pública).

Francis Bacon describió un precursor de los esquemas esteganográficos modernos, se dio cuenta de que al usar fuentes en cursiva o normales (alfabetos), se podía codificar la representación binaria de las letras en sus obras. Por ejemplo, Cinco letras del archivo de cubierta podrían contener cinco bits y, por lo tanto, una letra del alfabeto. La inconsistencia de la tipografía del siglo XVI hizo que este método pasara relativamente desapercibido.

También se han descrito métodos para ocultar datos en documentos de texto desplazando ligeramente las líneas de texto hacia arriba o hacia abajo en 1/300 de pulgada. Resulta que cambios tan pequeños no son perceptibles visualmente, pero son lo suficientemente robustos como para sobrevivir a que sean copiados. De esta forma, el mensaje podría extraerse incluso de documentos impresos o fotocopiados.

Durante los años 1800, se propuso una técnica muy ingeniosa que se utilizó en varias guerras de los siglos XIX y XX. La idea era encoger el mensaje tanto que comience a parecerse a motas de suciedad, micropuntos, que no tuvieran significado alguno al verse a simple vista, pero aún se puede leer con gran aumento. Los obstáculos tecnológicos para el uso de esta idea en la práctica fueron superados por el fotógrafo francés Dragon, quien desarrolló una tecnología para reducir el texto a dimensiones microscópicas. Dichos objetos pequeños podrían ocultarse fácilmente en las fosas nasales, los oídos o debajo de las uñas. En la Primera Guerra Mundial, los alemanes usaron esos micropuntos escondidos en las esquinas de las postales abiertas. Los micropuntos modernos del siglo XX podían contener hasta una página de texto e incluso contener fotografías.

Recientemente se propuso una versión moderna del concepto de micropunto para ocultar información en el ADN con el fin de marcar material genético importante.

Durante la Segunda Guerra Mundial, se usó tinta invisible para escribir información en pedazos de papel, de modo que el papel pareciera para la persona promedio como simples pedazos de papel en blanco. Se utilizaban líquidos como orina, leche, vinagre y jugos de frutas, ya que al calentar cada una de estas sustancias se oscurecen y se vuelven visibles al ojo humano.

En la actualidad con el avance diario de la tecnología, la comunicación electrónica es muy susceptible a las escuchas e intervenciones maliciosas por lo que los temas de seguridad y privacidad son más relevantes hoy que nunca. Las soluciones tradicionales se basan en la criptografía que es un campo maduro y bien desarrollado con fundamentos matemáticos rigurosos. El enfoque criptográfico de la privacidad es hacer que la información intercambiada sea ilegible para aquellos que no tienen la clave de descifrado correcta. Cuando se intercepta un mensaje encriptado a pesar de que el contenido del mensaje está protegido el hecho de que los sujetos se comunican en secreto es obvio. En algunas situaciones puede ser importante evitar llamar la atención y en su lugar incrustar datos confidenciales en otros objetos para que el hecho de que se envíe información secreta no sea ovio en primer lugar. Por tal motivo se desarrollan los mecanismos de esteganografía moderna aplicada a enviar mensajes ocultos en archivo multimedia.

1.4.4.3 Aplicaciones [8]

La esteganografía puede ser una solución que permita enviar noticias e información sin censura y sin temor a que los mensajes sean interceptados y rastreados hasta nosotros.

También es posible utilizar simplemente la esteganografía para almacenar información sobre una ubicación. Por ejemplo, varias fuentes de información como nuestra información bancaria privada, algunos secretos militares, pueden almacenarse en una fuente de cobertura. Cuando se nos requiere mostrar la información secreta en nuestra fuente de cobertura, podemos revelar fácilmente nuestros datos bancarios y será imposible probar la existencia de secretos militares en su interior.

La esteganografía también se puede utilizar para implementar marcas de agua. Aunque el concepto de marca de agua no es necesariamente esteganografía, existen varias técnicas esteganográficas que se utilizan para almacenar marcas de agua en los datos. La principal diferencia está en la intención, mientras que el propósito de la esteganografía es ocultar información, la marca de agua simplemente amplía la fuente de la portada con información adicional. Dado que las personas no aceptarán cambios notables en imágenes, archivos de audio o video debido a una marca de agua, se pueden usar métodos esteganográficos para ocultar esto.

El comercio electrónico permite un uso interesante de la esteganografía. En las transacciones de comercio electrónico actuales, la mayoría de los usuarios están protegidos por un nombre de usuario y una contraseña, sin ningún método real para verificar que el usuario es el titular real de la tarjeta. El escaneo biométrico de huellas dactilares, combinado con identificaciones de sesión únicas integradas en las imágenes de huellas dactilares a través de la esteganografía, permiten una opción muy segura para abrir la verificación de transacciones de comercio electrónico.

Junto con los métodos de comunicación existentes, la esteganografía se puede utilizar para realizar intercambios ocultos. Los gobiernos están interesados en dos tipos de comunicaciones ocultas: las que respaldan la seguridad nacional y las que no. La esteganografía digital proporciona un gran potencial para ambos tipos. Las empresas pueden tener preocupaciones similares con respecto a los secretos comerciales o la información de nuevos productos.

El transporte de datos confidenciales es otro uso clave de la esteganografía. Un problema potencial con la criptografía es que los espías saben que tienen un mensaje encriptado cuando lo ven. La esteganografía permite el transporte de datos confidenciales más allá de los espías sin que ellos sepan que los han pasado. La idea de utilizar la esteganografía en el transporte de datos se puede aplicar prácticamente a cualquier método de transporte de datos, desde correo electrónico hasta imágenes en sitios web de Internet.

1.4.4.4 Terminología [9]

1.4.4.4.1 Mensaje

Son los datos o mensaje que se quieren ocultar en el proceso de comunicación.

1.4.4.4.2 Archivo de cubierta

Son los medios o archivos que van a llevar incrustados el mensaje oculto.

1.4.4.4.3 Algoritmo de incrustación

Puede ser una función o modelo que toma como parámetros de entrada el archivo de cubierta y el mensaje para producir el estego-documento.

1.4.4.4.4 Estego-documento

Es una versión del archivo de cubierta modificado que contiene al mensaje oculto.

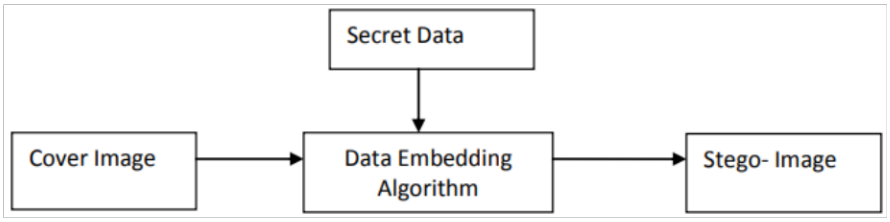

Figura 1.4. Diagrama de bloques para esteganografía en imágenes

1.4.4.5 Clasificación de la esteganografía [10]

Se pueden tomar dos enfoques a la hora de clasificar a la esteganografía.

El primer enfoque es sobre el uso de una llave al momento de usar el algoritmo de incrustación, similar a la criptografía se puede tener un intercambio de llaves al momento de realizar la comunicación. [11] La figura 5 describe a brevedad este enfoque.

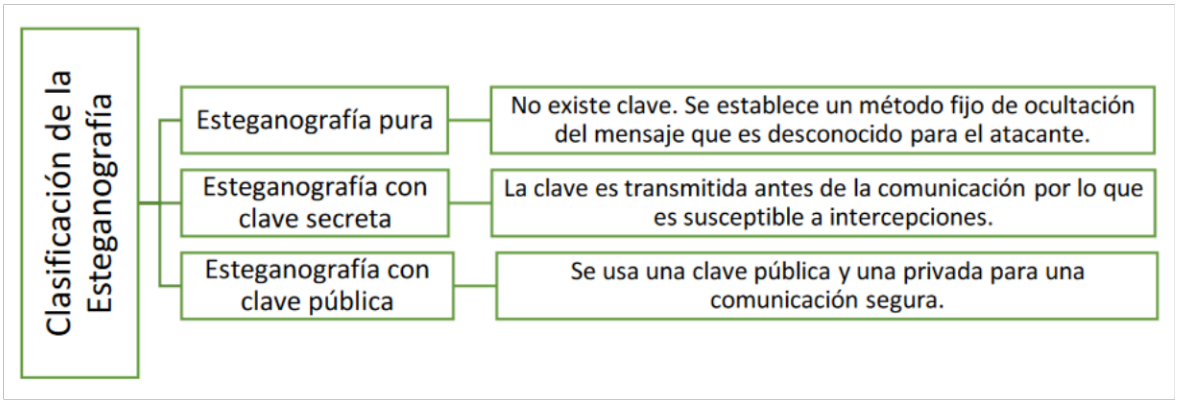

Figura 1.5. Clasificación de la esteganografía según el tipo de llave

El segundo enfoque es sobre el tipo de archivo usado como cubierta, dependiendo de este se tienen ciertas aplicaciones, algoritmos específicos, etc. Este último enfoque será detallado a continuación.

1.4.4.5.1 Esteganografía de texto

Consiste en ocultar información dentro de los archivos de texto. En este método, los datos secretos se ocultan detrás de cada enésima letra de cada palabra del mensaje de texto. Hay varios métodos disponibles para ocultar datos en archivos de texto.

1.4.4.5.2 Esteganografía de imágenes

Ocultar los datos tomando como archivo de cubierta una imagen. En la esteganografía de imágenes, se utilizan intensidades de píxeles para ocultar los datos. En la esteganografía digital, las imágenes son una fuente de cobertura ampliamente utilizada porque hay una gran cantidad de bits presentes en la representación digital de una imagen.

1.4.4.5.3 Esteganografía de audio

Implica ocultar datos en archivos de audio. Este método oculta los datos en archivos de sonido WAV, MP3, etc.

1.4.4.5.4 Esteganografía de video

Es una técnica para ocultar cualquier tipo de archivos o datos en formato de video digital. En este caso, el vídeo (combinación de imágenes) se utiliza como soporte para ocultar los datos. Generalmente, la transformada de coseno discreta (DCT) altera los valores que se usa para ocultar los datos en cada una de las imágenes en el video, lo cual es imperceptible para el ojo humano. H.264, Mp4, MPEG, AVI son los formatos utilizados por la esteganografía de video.

1.4.4.5.5 Esteganografía de protocolos de red

Implica ocultar la información tomando protocolos de red como: TCP, UDP, ICMP, IP, etc., como objeto de cubierta. En el modelo de red de capa OSI existen canales encubiertos donde se puede utilizar la esteganografía.

1.4.4.6 Técnicas de esteganografía [9] [10]

1.4.4.6.1 Métodos de dominio espacial

En este método, los datos secretos se incrustan directamente en la intensidad de los píxeles. Significa que algunos valores de píxeles de la imagen se cambian directamente durante la ocultación de datos. Las técnicas de dominio espacial se clasifican en las siguientes categorías:

- Bit menos significativo (LSB)
- Diferenciación de valor de píxel (PVD)
- Método de incrustación de datos basado en bordes (EBE)
- Método de incrustación de píxeles aleatorios (RPE)
- Método de mapeo de píxeles a datos ocultos
- Método de etiquetado o conectividad
- Basado en la intensidad de píxeles

1.4.4.6.2 Espectro ensanchado

En este método, los datos secretos se distribuyen en un ancho de banda de frecuencia amplio. La relación de señal a ruido en cada banda de frecuencia debe ser tan pequeña que resulte difícil detectar la presencia de datos. Incluso si se eliminan partes de los datos de varias bandas, todavía habrá suficiente información en otras bandas para recuperar los

datos. Por lo tanto, es difícil eliminar los datos por completo sin destruir por completo la cubierta. Es una técnica muy robusta que se usa principalmente en la comunicación militar.

1.4.4.6.3 Técnicas estadísticas

En esta técnica se incrusta el mensaje cambiando varias propiedades de la cubierta. Implica dividir la cubierta en bloques y luego incrustar un bit de mensaje en cada bloque. El bloque de cubierta se modifica solo cuando el tamaño del bit de mensaje es uno; de lo contrario, no se requiere modificación.

1.4.4.6.4 Dominio de la frecuencia

En esta técnica el mensaje secreto está incrustado en el dominio de frecuencia o transformación de la cubierta. Esta es una forma más compleja de ocultar un mensaje en una imagen. Se utilizan diferentes algoritmos y transformaciones en la imagen para ocultar el mensaje en ella. Las técnicas de dominio de transformación se clasifican ampliamente como:

- Técnica de transformación discreta de Fourier (DFT)
- Técnica de transformación de coseno discreta (DCT)
- Técnica de transformación de ondas discretas (DWT)
- Método sin pérdidas o reversible (DCT)
- Incrustación en bits de coeficiente

1.4.4.6.5 Técnicas de distorsión

En esta técnica, el mensaje secreto se almacena distorsionando la señal. El codificador aplica una secuencia de modificación a la cubierta. El decodificador mide las diferencias entre la cubierta original y la cubierta distorsionada para detectar la secuencia de modificaciones y consecuentemente recuperar el mensaje secreto.

1.4.4.7 Características de la esteganografía [9] [10]

1.4.4.7.1 Robustez

Se refiere a la capacidad de los datos incrustados de permanecer intactos si el estego-documento sufre transformaciones, como filtrado lineal y no lineal, nitidez o desenfoque, adición de ruido aleatorio, rotaciones y escalado, recorte o diezmado, compresión con pérdida.

Si bien la eliminación de datos secretos puede no ser un problema tan grave como su detección, la robustez es una propiedad deseable cuando el canal de comunicación está distorsionado por errores aleatorios (ruido del canal) o por interferencia sistemática con el objetivo de evitar el uso de esteganografía (ataques). Las métricas típicas de la robustez de los algoritmos esteganográficos se expresan en clases de distorsión, como ruido aditivo

o transformación geométrica. Dentro de cada clase, la cantidad de distorsión puede especificarse aún de mejor manera con medidas de distorsión específicas, como puede ser los parámetros de la fuente de ruido o genéricas, por ejemplo, la relación señal ruido máxima (PSNR). Cabe señalar que la robustez no ha recibido mucha atención hasta ahora en la investigación de esteganografía; sin embargo, la esteganografía robusta es un componente relevante para la construcción de tecnologías resistentes a la censura seguras y efectivas.

1.4.4.7.2 Imperceptibilidad

Significa invisibilidad de un algoritmo esteganográfico. Porque es el primer y más importante requisito, ya que la fuerza de la esteganografía reside en su capacidad para pasar desapercibida para el ojo humano.

1.4.4.7.3 Capacidad de ocultación

Se refiere a la cantidad de información secreta que se puede ocultar en la fuente de la cubierta. Las marcas de agua generalmente incorporan solo una pequeña cantidad de información de derechos de autor, mientras que la esteganografía se centra en la comunicación oculta y, por lo tanto, tiene suficiente capacidad de integración.

Esta capacidad se puede especificar en términos absolutos (bits) para una cubierta determinada, o en relación con el número de bits necesarios para almacenar el estego-documento resultante. La capacidad depende del algoritmo de incrustación y también puede depender de las propiedades de la cubierta. Por ejemplo, el reemplazo del bit menos significativo (LSB), algoritmo que se describirá más adelante, con un bit por píxel en una imagen en escala de grises de ocho bits sin comprimir logra una capacidad neta del 12,5%, o un poco menos si se tiene en cuenta que cada imagen se almacena con información de encabezado que no se usa para incrustar información. Esto también se puede representar como 1 bpp (bits por píxel), donde la información sobre las profundidades de bits reales de cada píxel debe conocerse a partir del contexto. Se debe tomar en cuenta que no todos los mensajes tienen la longitud máxima, por lo que los bits por píxel también se usan como una medida del uso de la capacidad o la tasa de incrustación. Para mejor representación se define una métrica p (proporción) para la longitud del mensaje secreto en relación con la longitud máxima del mensaje secreto de una cubierta. La tasa de incrustación p no tiene unidad y se define en el rango $0 \leq p \leq 1$. Por lo tanto, para una función de incrustación que incrusta un bit por símbolo de cobertura, sin embargo, encontrar medidas significativas para la capacidad y la tasa de incorporación no siempre es tan fácil, ya que algunos sistemas de esteganografía se integran los datos en cubiertas con formatos comprimidos.

En tales casos, es muy difícil ponerse de acuerdo sobre la mejor manera para realizar el cálculo de la capacidad porque el tamaño de la cubierta (por ejemplo, en bytes o en píxeles para imágenes) no es una buena medida de la cantidad de información en una cubierta. Por lo tanto, se necesitan medidas de capacidad específicas para formatos de compresión particulares de datos de cobertura. Por ejemplo, F5, un algoritmo esteganográfico para imágenes comprimidas en JPEG, se incrusta al disminuir el tamaño del archivo de forma casi monótona con la cantidad de bits incrustados [4].

Aunque contradictorio al principio, a la vista, esto funciona al reducir la calidad de imagen de la imagen comprimida con pérdida aún más por debajo del nivel de distorsión que ocurriría sin contenido esteganográfico. Como resultado, se ha propuesto bpc (bits por coeficiente DCT distinto de cero) como una métrica de capacidad en imágenes JPEG.

Es intuitivamente claro, y se puede demostrar teóricamente que los mensajes secretos más largos, requieren más cambios de incrustación y, por lo tanto, son estadísticamente más detectables que los más pequeños. Por lo tanto, la capacidad y la tasa de incrustación están relacionadas con la seguridad del algoritmo.

1.4.4.7.4 Relación señal ruido (SNR)

Es la relación entre la potencia de la señal y la potencia del ruido. Compara el nivel de una señal deseada con el nivel de ruido de fondo.

1.4.4.7.5 Error medio cuadrado (MSE)

Se define como la diferencia cuadrática media entre una imagen de referencia y una imagen distorsionada. Cuanto más pequeño sea el MSE, más eficiente será la técnica de esteganografía de imágenes. El MSE se calcula píxel a píxel sumando las diferencias cuadradas de todos los píxeles y dividiendo por el recuento total de píxeles.

1.4.5 IMÁGENES DIGITALES

Para el propósito del estudio de los métodos esteganográficos y los modelos de estegoanalisis, la mayoría de estos han sido desarrollados con pruebas usando imágenes como archivo de cubierta. Por ello se va a brindar una serie de conceptos para que sea más sencillo el entendimiento de estas técnicas esteganográficas.

1.4.5.1 Representación de imágenes [12]

En la mayoría de las pantallas de computadora, la imagen de la pantalla se compone de unidades discretas llamadas píxeles. Cada píxel ocupa una pequeña región rectangular en la pantalla y muestra un color a la vez. Los píxeles están dispuestos de manera que forman

una matriz bidimensional. Los objetos se dibujan en la pantalla ajustando el color de los píxeles individuales.

A lo largo de los años, la cantidad de píxeles que se muestran en los monitores de PC ha aumentado drásticamente.

1.4.5.2 Gráficas de vector y bitmap [12]

Así como los dispositivos de visualización tienen dos métodos generales de operación, los formatos de archivo de gráficos se pueden dividir en dos clases generales, vectoriales y de mapa de bits. Los formatos de gráficos vectoriales utilizan una serie de comandos de dibujo para representar una imagen. Un metarchivo de Windows es un formato de gráficos vectoriales de uso común.

Los formatos de gráficos vectoriales no se limitan a los dispositivos de salida, como los trazadores, que crean imágenes a través de comandos de dibujo. Los monitores de computadora y las impresoras láser suelen tener un software que convierte los comandos vectoriales en píxeles.

Hay dos inconvenientes principales con los gráficos vectoriales. En primer lugar, no son adecuados para reproducir fotografías o pinturas. Una pintura como requeriría decenas de miles de comandos de dibujo; simplemente determinar qué comandos usar para representar la pintura sería una tarea monumental. En segundo lugar, las imágenes complejas tardan mucho en mostrarse. En la mayoría de los sistemas de visualización, cada objeto vectorial debe convertirse en una imagen de píxeles.

Dichos formatos representan imágenes como matrices bidimensionales donde cada elemento de la matriz representa un color que se mostrará en una ubicación específica. Cuando se muestra en la pantalla de una computadora, cada elemento generalmente se asigna a un solo píxel de pantalla. Si los píxeles están lo suficientemente cerca en el dispositivo de visualización, se vuelve difícil para el ojo humano detectar la estructura de matriz que compone la imagen.

La mayor ventaja de las imágenes de mapa de bits es su calidad. A medida que la cantidad de espacio en disco y memoria ha aumentado junto con la velocidad de los procesadores, también se ha expandido el uso de imágenes de mapa de bits. Uno de los ejemplos más visibles de esto se encuentra en la industria de los juegos de computadora. Actualmente, incluso los juegos que requieren un alto rendimiento, como los simuladores de vuelo y los first person shooters, utilizan gráficos de mapa de bits.

Un gran inconveniente de las imágenes de mapa de bits es la cantidad de datos necesarios para almacenarlas. Otro inconveniente de las imágenes de mapa de bits es que dependen del tamaño y no son adecuadas para una edición extensa. Con formatos vectoriales, es fácil para los programas de dibujo agregar, eliminar y modificar elementos individuales. También es sencillo realizar transformaciones como perspectiva, ampliación y escala en una imagen vectorial.

Con imágenes de mapa de bits, incluso cambiar el tamaño causa problemas. Reducirlos requiere tirar información; agrandarlos produce efectos de bloqueo. Existen técnicas de suavizado para mejorar la apariencia de las imágenes redimensionadas.

1.4.5.3 Modelos de colores [12]

Hay muchas maneras de representar los colores numéricamente. Un sistema para representar colores se llama modelo de color. Los modelos de color generalmente están diseñados para aprovechar un tipo particular de dispositivo de visualización.

En la mayoría de los monitores a color hay tres transmisores (rojo, verde y azul) o emisores de luz para cada píxel. El ajuste de la intensidad de los transmisores individuales controla el color del píxel. Cuando los tres transmisores están en su intensidad mínima, el píxel aparece negro. En su máxima intensidad, el píxel aparece blanco. Si el fósforo rojo es el único activo, el píxel aparece rojo. Cuando los transmisores rojo y verde están encendidos, se combinan para producir tonos de amarillo y cuando los tres transmisores están a máxima intensidad, el píxel aparece en blanco.

El modelo de color más común utilizado en aplicaciones informáticas se conoce como RGB (Rojo-Verde-Azul). El modelo RGB imita el funcionamiento de las pantallas de ordenador. En RGB, los colores se componen de tres valores de componentes que representan las intensidades relativas de rojo, verde y azul. La gama de colores que se puede representar mediante un modelo de color se conoce como espacio de color.

En las discusiones matemáticas sobre el color, los valores de los componentes a menudo se representan como números reales normalizados en el rango de 0,0 a 1,0. En la programación y los formatos de imagen, casi siempre se utilizan valores de componentes enteros sin signo. El rango de valores para un componente de color está determinado por la precisión de la muestra, que es el número de bits utilizados para representar un componente.

1.4.5.4 Paleta de colores versus color verdadero [4] [12]

Por ejemplo, si el dispositivo de salida usa el modelo de color RGB para mostrar imágenes y que cada componente se representa usando 8 bits, su representación de colores está en el rango 0-255. Esta es la representación de color usada por la mayoría de los sistemas de computadoras personales. Tal sistema puede producir 16,777,216 colores distintos.

Hay sistemas informáticos que utilizan más bits para representar el color, por ejemplo, la escala de grises de 12 bits que se utiliza con frecuencia para las imágenes médicas. Algunos formatos de imagen admiten datos con más de 8 bits por componente (12 para JPEG, 16 para PNG). Para entender de mejor manera se va a asumir que se está trabajando en un sistema que utiliza 8 bits por componente.

Normalmente se utilizan dos métodos para asignar uno de los colores posibles a un píxel. El más simple es almacenar el valor de color para cada píxel en los datos comprimidos. Para imágenes con 24 bits por píxel, cada píxel tiene un valor de color de 3 bytes asociado. Las imágenes que usan 24 bits o más se denominan color verdadero porque en el rango de colores que puede mostrar un monitor, 24 bits por píxel es el límite de diferencias de color que un ser humano puede distinguir.

El problema con los gráficos de 24 bits es que, si bien un sistema puede mostrar 16 777 216 colores diferentes, es posible que no pueda hacerlo simultáneamente. Es posible que las computadoras más antiguas ni siquiera tengan una tarjeta de video capaz de usar un modo de visualización de 24 bits. Es posible que las computadoras más nuevas no tengan suficiente memoria de video para operar en modo de 24 bits con resoluciones de pantalla más altas. Una pantalla en una computadora personal configurada con una resolución de 1024 × 768 píxeles requeriría 2 359 296 (1024 × 768 × 3 = 2,25 MB) de memoria de video para mostrar imágenes de 24 bits. Si la computadora tuviera solo 2 MB de memoria de video, no podría mostrar imágenes de 24 bits a esta resolución, pero podría hacerlo a una resolución más baja de 800 x 600 (800 × 600 × 3 = 1,4 MB).

La solución ideada para representar colores antes de los días de pantallas capaces de 24 bits por píxel fue definir una paleta de colores que selecciona un subconjunto de los colores posibles. Conceptualmente, la paleta es una matriz unidimensional de elementos de 3 bytes que especifican el color. En lugar de especificar directamente el color, cada valor de píxel es un índice de la paleta de colores. El tamaño más común para una paleta es de 256 entradas, donde cada valor de píxel consta de 8 bits. La mayoría de las computadoras de hoy pueden mostrar gráficos de 8 bits en todas sus resoluciones de pantalla, pero las computadoras muy antiguas estaban limitadas a tamaños de paleta aún más pequeños.

1.4.5.5 Compresión [4] [12]

Dado que las imágenes de mapa de bits en color normalmente requieren más de un megabyte de almacenamiento, la mayoría de los formatos de archivo de imagen incorporan técnicas de compresión. Las técnicas de compresión aprovechan los patrones dentro de los datos de la imagen para encontrar una representación equivalente que ocupe menos espacio. Los datos completamente aleatorios no se pueden comprimir. La eficacia de una técnica de compresión depende del tipo de datos.

La mayoría de los formatos de archivo de imagen utilizan lo que se conoce como compresión sin pérdidas. Con esto se quiere decir que, si toma una imagen, se la comprime usando una técnica sin pérdidas y se la expande de nuevo, la imagen resultante es idéntica bit a bit a la original.

Algunos métodos de compresión (especialmente JPEG) tienen pérdidas. Usando la secuencia de compresión descrita anteriormente, la compresión con pérdida produce una imagen que es cercana a la original pero no una coincidencia exacta. Es decir, un píxel con un valor de color RGB de (128 243 118) en una imagen comprimida puede producir (127 243 119) cuando se expande. En la compresión de imágenes, las técnicas con pérdida aprovechan el hecho de que al ojo le resulta difícil distinguir entre colores casi idénticos.

La razón para usar la compresión con pérdida es que generalmente proporciona una compresión significativamente mayor que los métodos sin pérdida. En muchas situaciones, se aceptan pequeñas pérdidas de datos a cambio de una mayor compresión.

1.4.6 FORMATOS MULTIMEDIA PARA CUBIERTAS

1.4.6.1 JPEG [4] [6]

Dado que los ojos humanos son mucho menos sensibles a los cambios en la crominancia que en la luminancia, las señales de crominancia a menudo se representan con menos bits sin introducir una distorsión visible en la imagen. Este hecho se utiliza en el formato de compresión JPEG y también se usa para señales de TV, donde se asigna un ancho de banda más pequeño a las señales de crominancia y un ancho de banda más amplio se usa para la luminancia. Los formatos de imagen digital que utilizan el modelo YC_rC_b incluyen IIF, TIFF, JFIF, JPEG y MPEG (JPEG con movimiento).

Las personas perciben las imágenes naturales como una colección de segmentos llenos de textura en lugar de matrices de píxeles. En particular, las pruebas en sujetos humanos mostraron que nuestro sistema visual es bastante insensible a los pequeños cambios de color o al ruido de alta frecuencia espacial. Por lo tanto, es muy ineficaz almacenar imágenes naturales como matrices rectangulares de colores. Los ingenieros que trabajan

en la compresión de datos hace tiempo que se dieron cuenta de este hecho y propusieron varios formatos de imagen mucho más eficientes que funcionan al transformar la imagen en un dominio diferente donde se puede representar en una forma escasa y fácilmente comprimible. Dichos formatos suelen tener pérdidas, lo que significa que la conversión de formato introduce cierta pérdida de percepción que es imperceptible en condiciones de visualización normales. Los ahorros sustanciales en espacio de almacenamiento justifican la ligera pérdida de fidelidad. Las dos transformadas más utilizadas en la actualidad son la transformada de coseno discreta (DCT) y la transformada de wavelet discreta (DWT). El DCT está en el corazón del formato JPEG, mientras que el DWT se usa en JPEG2000. En este trabajo se presenta solo el formato JPEG, ya que la esteganografía JPEG2000 no está desarrollada a profundidad actualmente.

JPEG significa el Grupo Conjunto de Expertos Fotográficos que finalizó el estándar en 1992. La compresión JPEG consta de cinco pasos.

- **Transformación de color**
  El color se transforma del modelo RGB al modelo YC_rC_b. Aunque este paso no es necesario (JPEG puede funcionar directamente con la representación RGB), normalmente se usa porque permite relaciones de compresión más altas con la misma radio de fidelidad.
- **División en bloques y submuestreo**
  La señal de luminancia Y se divide en bloques de 8 × 8. Las señales de crominancia C_r y C_b se pueden submuestrear antes de dividirlas en bloques.
- **Transformada DCT**
  Las señales YC_rC_b de cada bloque se transforman del dominio espacial al dominio de frecuencia con la DCT. La DCT se puede considerar como un cambio de base que representa matrices de 8 × 8.
- **Cuantización**
  Los coeficientes de transformación resultantes se cuantifican dividiéndolos por un valor entero y se redondean al entero más cercano. Las señales de luminancia y crominancia pueden utilizar diferentes tablas de cuantificación. Los valores más grandes de los pasos de cuantificación producen una relación de compresión más alta, pero introducen más distorsión a la imagen.
- **Codificación y compresión sin pérdidas**
  Los coeficientes DCT cuantificados se organizan en orden de zig-zag, se codifican con bits y luego se comprimen sin pérdidas utilizando códigos de línea de Huffman o codificación aritmética. El flujo de bits resultante se antepone con un encabezado y se

almacena con la extensión jpg o jpeg. Para aplicaciones en esteganografía, no es necesario entender a profundidad este paso de codificación.

Para ver una imagen JPEG, primero se debe obtener la representación del dominio espacial del archivo JPEG, lo que se logra esencialmente invirtiendo los cinco pasos anteriores. El flujo de bits JPEG se analiza primero, luego se descomprime y luego se forma la matriz bidimensional de coeficientes DCT cuantificados. A continuación, los coeficientes de cada bloque se multiplican por los pasos de cuantificación y se aplica la DCT inversa para producir los valores de píxel sin procesar. Finalmente, los valores se redondean a números enteros de un cierto rango dinámico, generalmente el conjunto {0, ... , 255}. Si bien la compresión sin pérdida y la DCT son procesos reversibles, la cuantificación es un paso irreversible y, en general, la imagen descomprimida no será idéntica a la imagen original antes de la compresión.

A continuación, se muestra en la figura 1.6 un ejemplo de diagrama de bloques para el proceso de compresión y representación de una imagen en escala de grises.

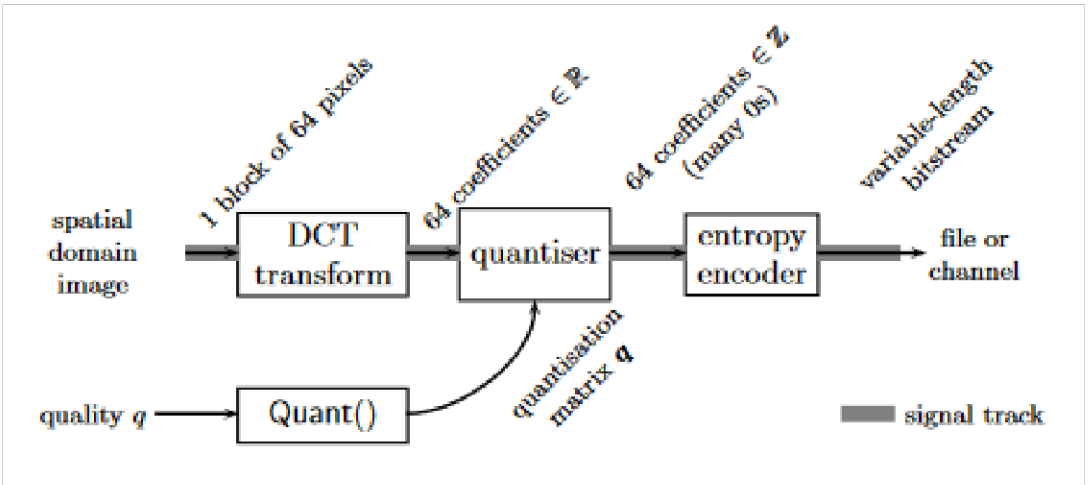

Figura 1.6. Diagrama de bloques para compresión de imagen JPEG en escala de grises [4]

1.4.6.2 MP3 [4]

El Grupo de expertos en imágenes en movimiento (MPEG) se formó en 1988 para producir estándares para representaciones codificadas de audio y video digital. El popular formato de archivo MP3 para señales de audio comprimidas con pérdida se especifica en el estándar ISO/MPEG 1 Audio Layer-3.

El estándar MP3 combina varias técnicas para maximizar el equilibrio entre la calidad de audio percibida y el volumen de almacenamiento. Su principal diferencia con muchos métodos de compresión anteriores y menos eficientes es su diseño como un enfoque de

dos pistas. La primera pista transmite la información de audio, que primero se pasa a un banco de filtros y se descompone en 32 sub-bandas de frecuencia igualmente espaciadas.

Estos componentes se transforman por separado al dominio de la frecuencia con una transformación de coseno discreta modulada (MDCT). Una operación de cuantificación posterior reduce la precisión de los coeficientes de MDCT. Hay que tomar en cuenta que los factores de cuantificación se denominan factores de escala en la terminología de MP3. A diferencia de la compresión JPEG, estos factores no son constantes en todo el flujo.

Finalmente, la codificación de entropía sin pérdidas de los coeficientes cuantificados asegura una representación compacta de los datos de audio MP3. La segunda pista es una pista de control.

Además, comenzando nuevamente desde la señal de entrada de modulación de código de pulso (PCM), se utiliza una FFT (transformada directa de Fourier) de 1024 puntos para alimentar el espectro de frecuencia de una ventana corta en el tiempo como entrada a un modelo psicoacústico. Este modelo emula las particularidades de la percepción auditiva humana, mide y valora la distorsión y deriva funciones de enmascaramiento para que la señal de entrada cancele las frecuencias inaudibles.

El modelo controla la elección de los tipos de bloque y los factores de escala específicos de la banda de frecuencia en la primera pista. En general, el enfoque de dos vías encuentra de forma adaptativa un compromiso óptimo entre la reducción de datos y la degradación audible para una señal de entrada dada.

En cuanto al formato de datos subyacente, un flujo de MP3 consta de una serie de fotogramas. Las etiquetas de sincronización separan los cuadros de audio MP3 de otra información que comparte el mismo flujo de transmisión o almacenamiento (por ejemplo, cuadros de video). En la siguiente figura se representa el proceso de compresión para el formato MP3.

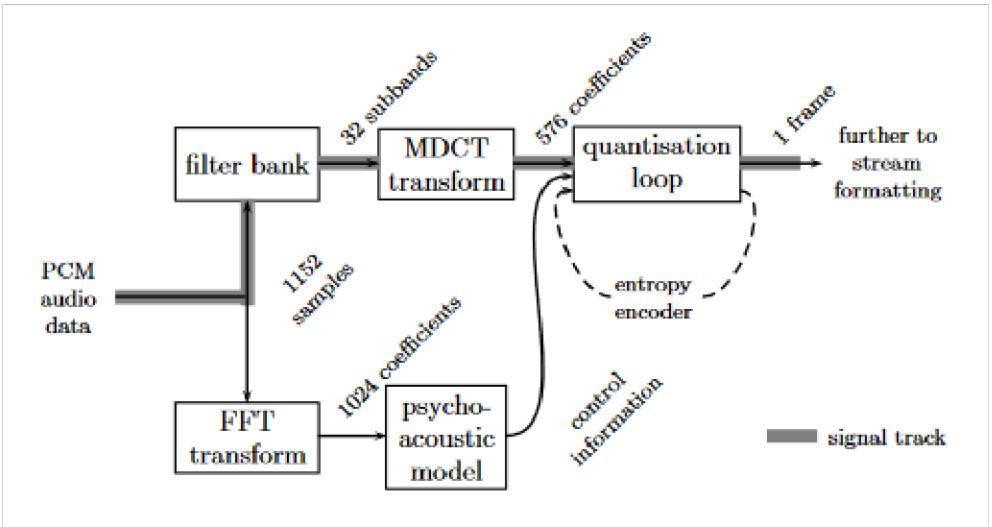

Figura 1.7. Representación de la compresión del formato MP3

Para una tasa de bits determinada, todos los fotogramas MP3 tienen un tamaño comprimido fijo y representan una cantidad fija de 1152 muestras PCM. Por lo general, un cuadro MP3 contiene 32 bits de información de encabezado, una suma de verificación de redundancia cíclica (CRC) opcional de 16 bits y dos gránulos de datos de audio comprimidos. Cada gránulo contiene uno o dos bloques, para señales mono y estéreo, respectivamente.

Ambos gránulos en un marco pueden compartir (parte de) la información del factor de escala para ahorrar espacio de almacenamiento. Dado que el tamaño de bloque real depende de la cantidad de información que se requiere para describir la señal de entrada, los tamaños de bloque y gránulo pueden variar entre fotogramas. Para equilibrar eficientemente los tamaños de gránulos flotantes en marcos de tamaños fijos, el estándar MP3 introduce el llamado mecanismo de depósito. Los marcos que no utilizan su capacidad total se llenan (parcialmente) con datos de bloque de marcos posteriores. Este método garantiza que las secciones locales altamente dinámicas en el flujo de entrada se puedan almacenar con una precisión superior a la media, mientras que las secciones menos exigentes asignan un espacio inferior a la media. Sin embargo, el alcance del uso del depósito está limitado para disminuir las interdependencias entre marcos más distantes y para facilitar la resincronización en posiciones arbitrarias en una corriente.

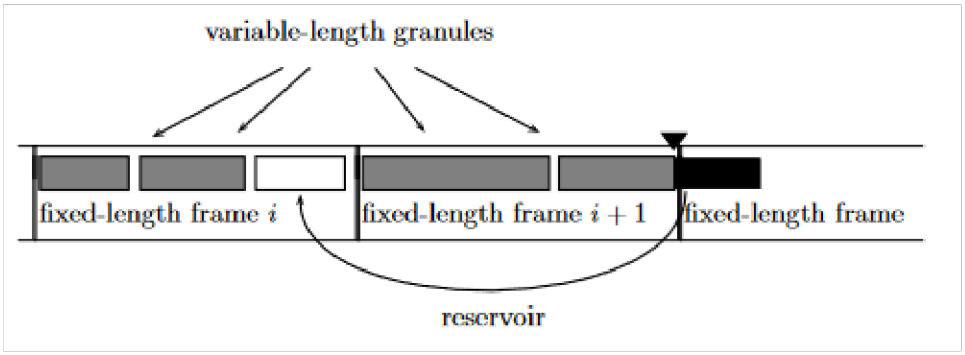

Figura 1.8. Representación de los gránulos para mono y audio estéreo presente en MP3

2 METODOLOGÍA

2.1 TÉCNICAS ESTEGANOGRÁFICAS

Para entender los modelos estadísticos que se usan al momento de recuperar un mensaje oculto, se debe realizar un análisis de las técnicas esteganográficas que se usan para ocultar información en archivos multimedia como imágenes y audio. Este análisis nos va a ayudar a comprender como se esconde la información y los problemas que se puede tener al momento de recuperar los mensajes.

2.1.1 BIT MENOS SIGNIFICATIVO (LSB) [13]

Al hablar de LSB se debe tomar en cuenta que fue uno de los primeros métodos utilizados para ocultar información en el dominio espacial; de igual manera, es de los mecanismos más sencillos en aplicar.

El método LSB funciona mejor en archivos de imagen que tienen una alta resolución y usan una variedad de colores, y en archivos de audio que tienen muchos sonidos diferentes y tienen una tasa de bits alta.

LSB generalmente no aumenta el tamaño del archivo, pero dependiendo del tamaño de la información enmascarada en el archivo, el archivo puede distorsionarse significativamente. Este método permite una alta incorporación de datos y es relativamente fácil de implementar o combinar con otras técnicas de enmascaramiento

Tradicionalmente, se basaba en incrustar cada parte del mensaje en la parte menos significativa del archivo de cubierta de forma determinista. Por ejemplo, para audio muestreado a 16 kHz, se enmascararán 16 kbps de datos.

Sin embargo, esta técnica se caracteriza por una baja resistencia a las interferencias, lo que reduce su rendimiento de seguridad, ya que se vuelve vulnerable incluso ante un simple ataque. Es muy probable que la filtración, la amplificación, la adición de ruido y la compresión con pérdida del estego-audio pueda destruir los datos. Además, dado que los datos están incrustados de una manera muy determinista, un atacante puede descubrir fácilmente el mensaje simplemente eliminando todo el plano LSB.

Se ha aplicado una estrategia LSB simple para incorporar un mensaje de voz en una comunicación inalámbrica. Si bien este método logra la imperceptibilidad a una alta tasa de incrustación, la seguridad y la solidez de los datos ocultos se ven fácilmente comprometidas. En un intento por aumentar la capacidad de ocultación y minimizar el error en el estego-audio, se adoptó un método de reemplazo de error mínimo al incorporar cuatro

bits por muestra. Luego, el error de integración se difunde en las siguientes cuatro muestras.

Para mejorar la solidez del método LSB frente a la distorsión y la adición de ruido, se ha aumentado la cantidad de incrustación a la cuarta, sexta y a la octava posición sin afectar la transparencia perceptible de la señal de estego-audio. Solo los bits en la sexta posición de cada muestra de 16 bits de la señal de cubierta original se reemplazan con bits del mensaje. Para minimizar el error de incrustación, los otros bits se pueden voltear para tener una nueva muestra más cercana a la original.

Por ejemplo, si el valor de la muestra original era 4, que se representa en binario como 0100, y el bit que se ocultará en la cuarta posición LSB es 1, en lugar de tener el valor 12=”1100” generado por el algoritmo LSB convencional, el algoritmo propuesto produce una muestra que tiene un valor de 3=”0011”, que está mucho más cerca del valor de la muestra original (es decir, 4). Por otro lado, se ha desplazado la incrustación de LSB a la octava posición y ha evitado ocultarse en periodos de silencio o cerca de puntos de silencio de la señal de cubierta. La aparición de instancias de incrustación en el octavo bit aumentará ligeramente la solidez de este método en comparación con los métodos LSB convencionales. Sin embargo, la capacidad de ocultamiento disminuye ya que algunas de las muestras deben permanecer inalteradas para preservar la calidad de percepción de audio de la señal. Además, la facilidad de recuperación de mensajes ocultos sigue siendo uno de los principales inconvenientes del LSB y sus variantes, si los bits ocultos en la sexta u octava posición se revelan maliciosamente fuera de la señal de estego-audio.

A continuación, en la tabla 2.1 se presentan las mejoras que se tiene al momento de aplicar LSB y sus métodos derivados.

Tabla 2.1. Características del método LSB y sus mejoras

	LSB Convencional	**Variantes LSB**	**Intervalos de Silencio**
Imperceptibilidad	X	X	X
Mejora contra ruido		X	X
Compresión			X

2.1.2 TRANSFORMADA DISCRETA DE FOURIER BIDIMENSIONAL (2D DFT) [14]

Los procesos esteganográficos también se pueden realizar en el dominio de la frecuencia. Por ejemplo, la imagen se transforma primero a su dominio de frecuencia. A diferencia del

dominio espacial, donde los cambios se realizan directamente en los valores de los píxeles, en el dominio de la frecuencia, los valores de los píxeles se transforman en coeficientes de frecuencia, mediante una técnica de transformación que pueden ser: Coseno Fourier, Wavelet, etc. Cualquiera que sea el procesamiento que se va a realizar, se lleva a cabo en el dominio de la frecuencia, y la imagen resultante se somete a una transformada inversa para obtener la imagen requerida.

La transformada discreta de Fourier (DFT) tiene la facilidad de reconstruir la imagen original sin pérdida de información y no presenta inconvenientes al cambiar de dominio. La transformada puede ser unidimensional o bidireccional, para el estudio de la esteganografía una imagen se considera una matriz por lo que se utiliza la transformada bidimensional.

Tomando en cuenta una imagen de cubierta con matriz N x N se debe crear máscaras (subimágenes de matrices más pequeñas), a todas estas máscaras se les aplica una DFT directa descrita en la ecuación 2.1, para transformar del dominio espacial al dominio de la frecuencia y obtener los coeficientes que van a ser usados para esconder la información.

$$F(u,v) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) e^{-j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)} \quad (2.1)$$

$$Donde\ u = 0\ a\ M-1$$

$$v = 0\ a\ N-1$$

Una vez en el dominio de la frecuencia se puede usar cualquier método para ocultar información en los coeficientes (el mismo método LSB anteriormente descrito puede ser usado).

Por último, se aplica una DFT inversa a cada máscara para obtener en el dominio espacial los pixeles de la imagen con información adicional oculta.

2.1.3 HERRAMIENTAS PARA ESTEGANOGRAFÍA [15]

Hoy en día existen disponibles para todas las personas varios programas de esteganografía gratuitos o comerciales. Debido al avance en el desarrollo de software la mayoría de las herramientas son similares en cuanto a funcionalidad y facilidad de uso.

La mayoría de las herramientas de steganografía que están disponibles en la actualidad brindan la opción de proteger los datos con una contraseña, cifrar los datos o comprimir los datos para mayor seguridad.

Aunque hay cientos de herramientas de steganografía; a continuación, se describirán algunos de estos programas que se basan en técnicas esteganográficas de inserción y sustitución.

2.1.3.1 S-Tools [15] [16]

Es un programa gratuito con una interfaz de arrastrar y soltar (drag and drop) que se ejecuta en la mayoría de las versiones de Windows. Puede ocultar datos en archivos de imagen GIF o bmp o en archivos de sonido wav. También puede aplicar técnicas de cifrado usando algoritmos como: IDEA, DES, Triple-DES y MDC. Comprimir archivos también es una opción.

S-Tools ofrece la posibilidad de ocultar varios mensajes secretos en un archivo de cubierta. Para todos los formatos de archivo, oculta los datos en los tres bits menos significativos de cada byte de datos.

Para los archivos de sonido, los datos se colocan directamente en los tres bits menos significativos del archivo, esto funciona con archivos de audio en formato wav de 8 o 16 bits.

Para ocultar datos en imágenes de 8 bits, S-Tools modifica la imagen para usar solo una paleta de 32 colores en lugar de 256. Los 32 colores se duplican 8 veces (32 × 8 = 256), para llenar la tabla de colores con entradas duplicadas. S-Tools puede usar las entradas duplicadas para almacenar el mensaje secreto en los tres LSB para cada entrada RGB. Dado que cada color de la imagen modificada se puede representar de ocho formas diferentes, la información se puede ocultar en cualquiera de las representaciones que se repiten.

2.1.3.2 Hide and Seek [15]

Es un programa gratuito que oculta datos en archivos de imagen GIF utilizando el bit menos significativo de cada byte de datos para codificar caracteres. A continuación, utiliza la dispersión para distribuir los datos (y, por tanto, la degradación de la calidad de la imagen) un poco por todo el GIF de forma pseudoaleatoria. Un algoritmo pseudoaleatorio parece bastante aleatorio, pero en realidad es predecible para permitir la extracción de datos.

Este método es fundamentalmente el mismo que el método de 8 bits utilizado por S-Tools. La única diferencia es que Hide and Seek reduce la tabla de colores a 128 colores y crea dos duplicados. Hide and Seek solo funciona con codificación de color de 8 bits.

El ruido se nota cuando se usa esta herramienta con archivos cubierta más grandes, pero los archivos más pequeños permanecen prácticamente intactos. Como recomendación se

tiene que el mensaje a ocultar también debe ser corto (no más de 19Kbytes) porque cada carácter necesita 8 píxeles para ocultarse.

Hide and Seek en realidad consta de dos ejecutables, uno para ocultar y otro para extraer el mensaje. Ambos programas se ejecutan desde una línea de comandos (CLI) y el usuario pasa los argumentos para los nombres de archivo.

Debido a que el método para ocultar el mensaje es relativamente débil y fácil de descifrar, también se puede proporcionar una contraseña que encripte el mensaje y lo haga más difícil de descifrar.

2.1.3.3 J-Steg [15] [16]

J-Steg oculta los datos en imágenes JPEG. Ocultar datos en imágenes JPEG es bastante diferente de todas las otras técnicas descritas anteriormente, ya que JPEG utiliza un algoritmo de compresión con pérdida para almacenar datos de imagen. Es por eso que las imágenes JPEG se usan en el Internet debido a que estes ocupan menos espacio.

Un algoritmo de codificación o compresión de datos con pérdida es aquel que pierde, o desecha deliberadamente, los datos de entrada durante el proceso de codificación para obtener una mejor relación de compresión. Cuando se utiliza la compresión con pérdida, los mensajes almacenados en los datos de la imagen se dañarán. Debido a esto, pensaría que este formato de archivo no podría usarse para ocultar datos.

Para superar este inconveniente, en lugar de almacenar mensajes en los datos de imagen, J-Steg utiliza los coeficientes de compresión para almacenar datos. Las imágenes JPEG utilizan un esquema de compresión de transformada discreta de coseno (DCT). Los datos comprimidos se almacenan como números enteros y la compresión implica cálculos extensos que se redondean al final. Al realizar estas transformaciones, los mensajes pueden integrarse en los coeficientes DCT y de esta manera los mensajes ocultos son bastante difíciles de detectar.

2.1.3.4 EZ Stego [15] [16]

EZ Stego es un software que está escrito en el lenguaje de programación Java. Este programa oculta datos usando los bits menos significativos de las imágenes GIF. Lo hace almacenando los datos en la tabla de colores, pero la tabla no se modifica como ocurre con S-Tools y Hide and Seek. EZ Stego ordena la tabla de colores de la imagen GIF por su equivalencia en colores RGB, en lugar de reducir el color de la imagen y duplicar las entradas de color. Esta clasificación se realiza de tal manera que los colores similares

aparecen uno al lado del otro en la tabla de colores. Esto ayuda a garantizar que la imagen de salida (estego-imagen) no se degrade gravemente.

2.1.3.5 Image Hide [15]

Image Hide es un programa de esteganografía que tiene una GUI sencilla para el usuario. Este programa puede ocultar información en una variedad de formatos, lo hace de manera similar a las otras técnicas sobre las que ha leído en este capítulo, reemplazando los bits menos significativos de los píxeles individuales de una imagen. Sin embargo, hace la incrustación sobre la marcha, lo que hace que este programa sea único.

Dado que Image Hide usa una técnica de sustitución, este no aumenta el tamaño del archivo; por lo tanto, existen límites en la cantidad de datos que puede ocultar.

2.1.3.6 Digital Picture Envelope [15]

Digital Picture Envelope conocido por sus sigles en inglés, DPE, es una herramienta para steganografía que oculta información en un formato de archivo bmp. El método para ocultar información con esta herramienta es similar a las anteriores descritas, lo que hace que esta técnica sea única es que puede ocultar grandes cantidades de datos en una imagen y hacerlo sin cambiar el tamaño del archivo.

El programa viene con dos opciones: una para incrustar la información y otra para extraerla. Digital Picture Envelope se basa en un algoritmo de incrustación llamado Segmentación Compleja del Plano de Bits (BCPS), que fue desarrollado por Eiji Kawaguchi de Japón. DPE puede usar imágenes bmp de 8 bits, pero para ocultar la cantidad máxima de datos, lo usaría con imágenes de varios colores (24 bits). Los datos están incrustados en los planos de bits (similares a los píxeles) de las imágenes, que es lo que le permite ocultar una cantidad tan grande de datos. Este método también se llama steganografía de gran capacidad.

2.1.3.7 Camouflage [15]

Camouflage es una herramienta de inserción de datos relativamente simple que funciona en una variedad de formatos. Oculta los datos al final de un archivo, colocándolos después del marcador de fin de archivo. La aplicación ignora esta información. Debido a que Camouflage utiliza una técnica de inserción, puede ocultar grandes cantidades de datos, pero esos datos son bastante fáciles de detectar. Lo que hace que este programa sea único es que para usarlo no se ejecuta una aplicación especial para ocultar y extraer información. En el menú contextual del programa se tiene dos opciones adicionales, una para esconder la información (camuflar) y otra para extraer el mensaje oculto (descamuflar).

2.1.3.8 Gif Shuffle [15]

Gif Shuffle es un programa que oculta datos en imágenes GIF. Lo hace manipulando la tabla de colores de la imagen. El orden de la tabla de colores no importa para las imágenes GIF porque en una imagen GIF típica, la tabla no está ordenada. Gif Shuffle toma una imagen y ordena la tabla de colores, organizándola por colores similares. Esto se hace usando los valores RGB (rojo, verde y azul). Luego se realiza una operación de modificación y cada parte del mensaje encubierto se oculta en la tabla de colores.

Gif Shuffle es una herramienta de línea de comandos (CLI) que puede ser utilizada para ocultar y extraer un mensaje. Debido a que esta técnica manipula el mapa de colores, existe un límite en la cantidad de datos que se pueden ocultar, por lo que el programa viene con una opción que puede ser ingresada como comando que le muestra al usuario la cantidad de datos que se pueden ocultar en un archivo.

Para extraer un mensaje, escriba el nombre del programa, seguido del nombre del archivo que tiene datos ocultos y aparecerá el mensaje.

2.2 ESTEGOANÁLISIS

2.2.1 DEFINICIÓN [4] [17]

El estegoanálisis es la parte contraria de la esteganografía y se encarga de detectar información oculta en un determinado grupo de objetos (estego-documentos), para lo cual se utiliza diferentes técnicas del campo de la computación como análisis de imágenes y señales, inteligencia artificial y reconocimiento de patrones, entre otras.

No hay un método único y formal para crear un estego-documento por lo que el estegoanálisis requiere de múltiples técnicas de diferentes áreas para poder reconocer aquellos objetos que pudieran estar escondiendo información. Esto hace que el problema de la detección de mensajes sea desafiante y complejo por las implicaciones prácticas que conlleva, partiendo de la experimentación, hallar el método adecuado para poder encontrar las características que permiten la detección de los posibles estego-documentos. Seguido de esto es necesario escoger un mecanismo de clasificación que en base a la información obtenida del archivo pueda clasificar de manera correcta estos objetos y determine si contienen información oculta o no.

Lo que se intenta explicar es que para que exista un sistema esteganográfico operable, la incrustación y la extracción del mensaje oculto son computacionalmente fáciles, mientras que la detección confiable requiere muchos más recursos.

2.2.2 TIPOS DE DETECTORES [18]

Dentro del estegoanálisis se definen dos tipos de detectores que son usados para determinar si un estego-documento contiene o no información oculta, estos pueden ser dedicados o específicos y universales.

Cualquiera de los dos tipos de detectores tiene dos enfoques: activo y pasivo, el primer enfoque consiste en modificar todos los elementos que viajan en el canal de comunicación, de manera que no afecte su uso legítimo; por ejemplo, la visualización de una imagen, pero cuando se requiera recuperar la información, hace que esto no sea posible. Esto supondría una solución final a uno de los principales problemas de la detección de la esteganografía como lo es la imperceptibilidad, pero su implementación es solamente teórica. Estos enfoques serán discutidos a continuación.

2.2.2.1 Guardían del canal de comunicación

En la esteganografía, el canal físico utilizado para la comunicación generalmente se asume libre de ruido, ya que esto puede garantizarse mediante la corrección de errores y los protocolos estándar de Internet. En cambio, las propiedades del canal son definidas por el guardián. El guardián se considera parte del canal porque puede o no interferir con la comunicación. Se toma en cuenta dos tipos de guardianes principales: el pasivo y el activo, existe otro tipo de guardián el cual se define por sus intenciones en el canal de comunicación (malicioso).

2.2.2.1.1 Guardián pasivo [19]

El guardián se llama pasivo si no puede modificar el contenido enviado por el extremo A antes de que el extremo B lo reciba (es decir, el guardián solo puede evitar o permitir la entrega del mensaje del extremo A). En este escenario, el guardián prueba cada comunicación del extremo A para detectar la presencia de un mensaje encubierto. Si la prueba del guardián es negativa, la comunicación se transmite al extremo B. De lo contrario, está bloqueado. Este es el escenario comúnmente más asumido del porque la mayoría de los algoritmos esteganográficos no están diseñados para ser robustos.

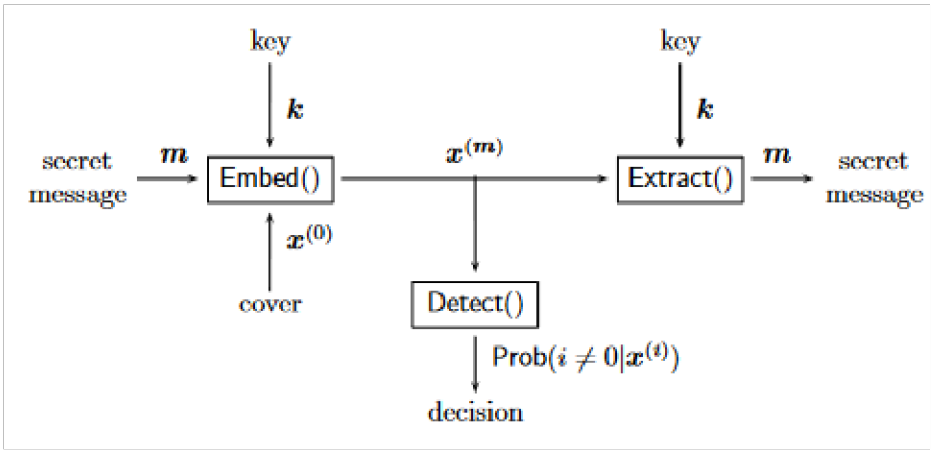

Figura 2.1. Diagrama de bloques para esteganografía con guardián pasivo [4]

2.2.2.1.2 Guardián activo [19]

El guardián se considera activo si modifica intencionalmente el contenido enviado por el extremo A antes de que el extremo B lo reciba. En este escenario, es posible que la directora no esté completamente segura de su programa de estegoanálisis. Por lo tanto, aunque sus pruebas sean negativas, el guardián puede alterar el contenido, con la esperanza de que la modificación destruya cualquier mensaje esteganográfico que pueda estar presente. Si el algoritmo esteganográfico asume un guardián pasivo, entonces hay una buena posibilidad de que las alteraciones al contenido degraden severamente o eliminen el mensaje oculto. Los tipos de modificación que podría aplicar un guardián activo incluyen la recompresión con pérdida de imágenes y clips de audio, el uso de filtros pasa bajos y otros procedimientos que degradan ligeramente el contenido.

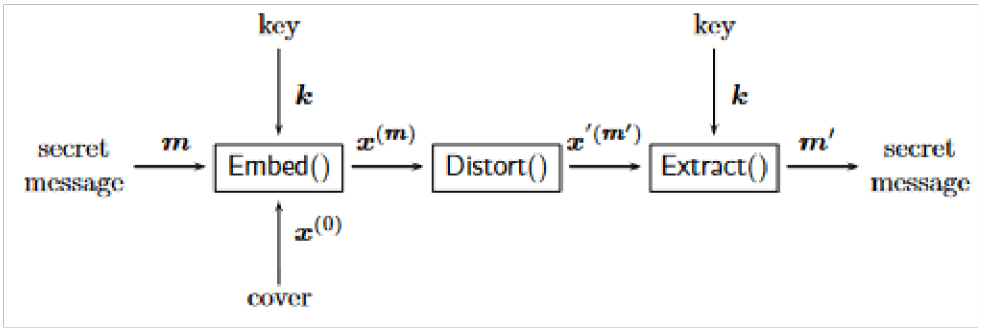

Figura 2.2. Diagrama de bloques de esteganografía con guardián activo [4]

2.2.2.1.3 Guardián malicioso [19]

El guardián se llama malicioso si sus acciones se basan en los detalles del esquema esteganográfico y tienen como objetivo atrapar a los prisioneros comunicándose en secreto. Esto puede incluir que el guardián intente hacerse pasar por el extremo A o el extremo B o que los engañe. Un guardián malicioso generalmente se considera en la

esteganografía de clave pública. En este escenario, se conoce la clave steganográfica y cualquiera puede extraer el mensaje secreto. Sin embargo, el mensaje se cifra mediante un sistema criptográfico de clave pública. Solo aquellos que poseen la clave privada del extremo B pueden descifrar el mensaje del extremo A. Aunque se conoce la clave steganográfica, es difícil distinguir entre un mensaje cifrado y una secuencia de bits aleatoria extraída de un archivo de cubierta. Sin embargo, dado que el guardián también conoce la clave steganográfica, tiene más opciones para atacar el sistema steganográfico.

2.2.2.2 Detector Específico

Este tipo de detector sirve para un tipo de técnica esteganográfica en particular, su ventaja es el de un porcentaje detección mayor comparado con otros tipos de detectores, pero con la desventaja de que debe de saber a priori que tipo de técnica fue utilizada en el objeto a analizar.

2.2.2.3 Detector Universal

Son aquellos que detectan más de dos tipos de técnicas esteganográficas, su desventaja es que tienden a tener una tasa de detección menor que la de los detectores dedicados, pero no requieren saber que técnica esteganográficas fue empleada en el objeto a analizar.

2.2.3 PROBLEMAS DEL ESTEGOANÁLISIS

Los métodos de estegoanálisis existentes se basan en la heurística y dado un método de esteganografía en específico, no existe un enfoque sistemático para diseñar un método de estegoanálisis.

Cada método de estegoanálisis tiene algunos parámetros que determinan el rendimiento del método. Idealmente, los parámetros de prueba deben elegirse únicamente teniendo en cuenta la base de la cubierta para cumplir con el rendimiento objetivo. Estos métodos se encuentran muy poco desarrollados.

Debido a la poca base teórica del tema, no se sabe cómo se comparan las pruebas de estegoanálisis actuales con pruebas en casos reales para determinar a esos métodos desarrollados como óptimos.

Aún quedan por entender muchas cuestiones fundamentales en el estegoanálisis. Un enfoque prometedor para desarrollar un marco sistemático para el estegoanálisis y que puede servir como solución a ciertos problemas es la teoría de la prueba de hipótesis.

2.2.3.1 Plausibilidad heurística [4] [18]

Los mensajes intercambiados entre los extremos de una comunicación deben pasar desapercibidos; es decir, indistinguibles de cubiertas plausibles, esto significa que se

pueda sospechar que la cubierta lleve información oculta. Sin embargo, la plausibilidad es principalmente un criterio empírico, subjetivo y difícil de capturar con métodos formales. Un enfoque común es definir la plausibilidad en un sentido probabilístico, es decir, determinar que los objetos que probablemente contienen mensajes ocultos son plausibles por definición. A este ajuste se lo conoce como plausibilidad heurística que quiere decir que para determinar las posibles cubiertas con mensajes ocultos se basarán en investigaciones anteriores, así como en métodos no tan rigurosos. Esto es importante porque la mayoría literatura sobre el tema confía implícitamente en esta heurística.

El objetivo de este trabajo no es resolver todas las consecuencias de la plausibilidad heurística, por lo que solo se dará una breve descripción del impacto de este concepto en el estegoanálisis. La plausibilidad heurística implica que un mensaje, objeto u archivo se considera más o menos probable de tener información oculta con respecto a una función de probabilidad universal.

En la práctica, los extremos de comunicación o quien está a cargo del proceso (un guardián de la comunicación) no poseen pleno conocimiento de esta función, porque eso correspondería a tener acceso a un oráculo global (un adivino con todas las respuestas), y es una cuestión filosófica si tal función puede existir en absoluto. En cambio, es razonable suponer que todas las partes tienen conceptos privados de qué mensajes creen que son plausibles o no. Entonces se puede asumir que existe cierta superposición entre los conceptos privados de ambos extremos de la comunicación y el guardián; de lo contrario, el proceso de comunicación sería imposible.

Pero esto puede dar lugar a malentendidos. Por ejemplo, en la comunicación esteganográfica, cada uno de los participantes de la comunicación forman expectativas racionales sobre lo que el otro podría considerar plausible o no y, por lo tanto, pueden desviarse tanto de su propia noción de plausibilidad como de la función de probabilidad universal imaginaria.

Por lo tanto, los guardianes realistas deben formarse expectativas sobre lo que es plausible para los extremos, teniendo en cuenta todos los conocimientos previos disponibles. Por supuesto, cada una de estas expectativas se puede formular como funciones de probabilidad condicional, pero esos son solo modelos para reemplazar la falta de comprensión de las otras partes. Dada esta explicación, se puede identificar al menos tres enunciados que acompañan a la plausibilidad heurística y que son común a saber:

- Noción de plausibilidad universal en lugar de específica por el contexto.
- Simplificación del razonamiento por funciones de probabilidad.

- Mala noción de plausibilidad de las otras partes, debido a conjeturas subjetivas dentro de todo el contexto.

2.2.3.2 Archivos de cubierta heterogéneos [4]

Para hablar sobre los problemas que causa tener varias fuentes de información, se propone que el modelo de estegoanálisis sea representado como una hipótesis por la denotación P_0.

Una de las razones por las que P_0 es tan difícil de modelar, es el hecho de que los formatos típicos de los archivos de cubierta son tan generales que se pueden almacenar y transmitir información dentro de una una gran cantidad de fuentes diferentes. Cada uno de los procesos de generación de un estego-documento de varias fuentes está gobernado por mecanismos físicos específicos, lo que da como resultado diferentes características estadísticas de las representaciones digitales. Incluso es difícil evaluar la cantidad de variación entre las fuentes de cubierta, ya que la unidad de análisis para estimar las propiedades estadísticas sería la fuente y no el estego-documento individual. Los conjuntos de datos de estego-documentos comunes consisten en muchas cubiertas de una sola fuente (homogénea), mientras que el muestreo controlado de una gran cantidad de fuentes heterogéneas es mucho más costoso.

En consecuencia, muchos esfuerzos para refinar los modelos de incrustación de información dan como resultado modelos más específicos para fuentes particulares, pero lo que realmente se necesita es un enfoque unificado para tratar los archivos de cubierta heterogéneos en general.

Hay que tener en cuenta que la variación en una variable aleatoria es demasiado grande para ignorar el error de aproximar sus realizaciones mediante parámetros fijos de la función de distribución (por ejemplo, momentos, si corresponde). La multimodalidad (es decir, máximos locales de la función de distribución) o la dispersión no despreciable (por ejemplo, colas pesadas) son causas comunes y signos de heterogeneidad. Por supuesto, el criterio de lo que es aceptable depende de la aplicación a desarrollar, en el caso del estegoanálisis, está relacionado con el rendimiento de seguridad o detección alcanzable.

En el contexto de este trabajo, se argumenta que la heterogeneidad se puede abordar mejor con modelos de cubierta condicionales, que combinan varios modelos de cubierta más específicos para subconjuntos homogéneos de fuentes de cobertura heterogéneas, de manera jerárquica.

La idea está motivada por la observación de que medir dependencias incondicionales en cubiertas empíricas de alta dimensión a menudo no es práctico, pero las versiones condicionales de P_0 son más manejables. Después de presentar este caso, cabe recordar que el principio básico de las distribuciones mixtas son una forma elegante y estadísticamente bien fundamentada de lidiar con la heterogeneidad.

El objetivo de este trabajo sigue siendo estudiar un marco conceptual general, que se pueda adaptar a aplicaciones específicas de estegoanálisis.

2.2.4 MÉTODOS DE ESTEGOANÁLISIS [9]

2.2.4.1 Ataques visuales [16]

La mayoría de los programas y mecanismo esteganográficos incorporan bits de mensaje de forma secuencial o de forma pseudoaleatoria. En la mayoría de los casos, los bits del mensaje se eligen de forma no adaptativa, independientemente del contenido de la imagen. Si la imagen contiene áreas conectadas de color uniforme o áreas con el color saturado entre 0 o 255, se puede buscar bits sospechosos mediante una simple inspección visual después de preprocesar la estego-imagen. Aunque los bits sospechosos no se pueden ver fácilmente, lo mejor es trazar un plano de bits (por ejemplo, el plano LSB) e inspeccionar solo el plano de bits. Este ataque es especialmente aplicable a las imágenes para incrustar LSB en los índices de la paleta.

Si al mismo tiempo, el mensaje se incrusta secuencialmente, se puede tener un argumento convincente para la presencia de mensajes esteganográficos en una imagen. Sin embargo, como se puede leer en otras literaturas, puede ser imposible distinguir las imágenes con ruido o las imágenes con mucha incrustación de información de las estego-imágenes utilizando esta técnica.

Aunque los ataques visuales son simples, son difíciles de automatizar y su confiabilidad es muy cuestionable.

2.2.4.2 Análisis de histogramas [9] [16]

Estudios previos en el campo del estegoanálisis introdujeron un poderoso ataque estadístico que se puede aplicar a cualquier técnica esteganográfica en la que un conjunto fijo de pares de valores (PoV) se invierten entre sí para incrustar bits de mensaje. Por ejemplo, los PoV pueden estar formados por valores de píxeles, coeficientes DCT cuantificados o índices de bits que difieren en el LSB. Antes de incrustar, en la imagen de cubierta, los dos valores de cada par se distribuyen de manera desigual. Después de la incrustación del mensaje, las ocurrencias de los valores en cada par tenderán a volverse iguales (esto depende de la longitud del mensaje a incrustar). Dado que cambiar un valor

por otro no cambia la suma de ocurrencias de ambos colores en la imagen, se puede usar este hecho para diseñar una prueba estadística de chi-cuadrado.

Se puede probar la importancia estadística del hecho de que las ocurrencias de ambos valores en cada par son iguales. Si, además de eso, el algoritmo esteganográfico incrusta bits de mensaje secuencialmente en píxeles, índices o coeficientes posteriores comenzando en la esquina superior izquierda, se observará un cambio abrupto en la evidencia estadística cuando se encuentra el final del mensaje.

Para esto se calcula la probabilidad de que los coeficientes en frecuencia sean iguales a los coeficientes de ocurrencia en la imagen de cubierta (k).

$$p = 1 - \frac{1}{2^{\frac{k-1}{2}}\Gamma(\frac{k-1}{2})} \int_0^{x_{k-1}^2} e^{\frac{x}{2}} x^{\frac{k-1}{2}-1} \tag{2.2}$$

Para un mensaje incrustado secuencialmente, se puede escanear la imagen en el mismo orden en que se ha incrustado el mensaje y evaluar el valor p para el conjunto de todos los píxeles ya visitados. El valor de p al principio estará cerca de 1 y luego de repente bajará a 0 cuando se llega al final del mensaje. Permanecerá en cero hasta que se llegue a la esquina inferior derecha.

Por lo tanto, esta prueba nos permite no solo determinar con una probabilidad muy alta que un mensaje ha sido incrustado, sino también determinar el tamaño del mensaje secreto.

Si los píxeles que llevan el mensaje en la imagen se seleccionan al azar en lugar de secuencialmente, esta prueba se vuelve menos efectiva a menos que la mayoría de los píxeles se hayan utilizado para incrustar (tamaño del mensaje comparado con la cantidad de píxeles en la imagen). Si esta técnica se aplica a diferentes áreas más pequeñas de la imagen, el valor p fluctuará con un grado decreciente a medida que aumenta la longitud del mensaje. Esto se debe a que un mensaje difundido aleatoriamente se debe a que el azar estará más concentrado en algunas áreas que en otras. Al cuantificar esta observación, no se podría, en principio, detectar incluso mensajes dispersos al azar y estimar su longitud. Desafortunadamente, no existe ningún análisis estadístico adicional de esta observación ni ningún otro detalle que respalde esta afirmación.

Finalmente, se puede decir que cualquier técnica de estegoanálisis basada en el análisis de recuentos de muestras (histograma) será fácil de eludir. Existen estudios sobre el desarrollo de una técnica de incrustación de JPEG que conservará los recuentos originales de muestras en sus puntos de vista, por lo tanto, evitará la detección de mensajes utilizando este tipo de análisis.

2.2.4.3 Métodos estadísticos [4] [7]

Existen propiedades estadísticas para la percepción humana de imágenes naturales. La longitud de onda de la luz genera estímulos visuales. Cuando la longitud de onda cambia dinámicamente, el color percibido varía de rojo, naranja, amarillo, verde y azul a violeta, lo que se conoce como el espectro de luz visible. Las sensibilidades visuales humanas están estrechamente relacionadas con el color; por ejemplo, los humanos son más sensibles al verde que al rojo y al azul. Se puede dividir la composición de un color en tres componentes: brillo, matiz y saturación. La adaptación del ojo a un tono particular distorsiona la percepción de otros tonos; por ejemplo, el gris visto después del verde, o visto contra un fondo verde, se ve diferente y no como debería de ser.

Los investigadores en esteganálisis estadístico se han esforzado en mejorar sus métodos de detección debido a que en la actualidad no se presenta mayor variación de las estadísticas al usar modelos de primer orden para detectar mensajes ocultos, esto debido a que los algoritmos de esteganografía se han venido preparando para evitar la detección. El cifrado de mensajes ocultos también dificulta la detección porque los datos cifrados generalmente tienen un alto grado de aleatoriedad.

Además de la detección, la recuperación de mensajes ocultos parece ser un problema complicado, ya que adicionalmente se requiere el conocimiento del algoritmo de criptografía o la clave de cifrado.

La inserción de LSB en una imagen basada en una paleta producirá una gran cantidad de colores duplicados. Algunos colores idénticos (o casi idénticos) pueden aparecer dos veces en la paleta. Algunas técnicas esteganográficas que reorganizan el orden de la paleta de colores para ocultar mensajes pueden provocar cambios estructurales, lo que generará que el algoritmo esteganográfico pueda ser reconocido. Los colores similares tienden a estar agrupados y cambiar el LSB de un píxel de color no degrada la percepción visual general.

Existen algunos métodos para estoanálisis, incluido el método singular-regular (RS), el método de pares de valores, el método de análisis de pares y el método de pares de muestras, que aprovechan los cambios estadísticos al modelar cambios en ciertos valores como una función del porcentaje de píxeles con esteganografía incrustada.

Las técnicas esteganográficas generalmente modifican las propiedades estadísticas del archivo de cubierta; un mensaje oculto más largo alterará más a la cubierta que uno más corto. El análisis estadístico se utiliza para detectar mensajes ocultos, particularmente en esteganografía a ciegas, donde no se conoce el algoritmo de incrustación. El análisis

estadístico de las imágenes puede determinar si sus propiedades se desvían de la norma esperada; esto se puede realizar de varias maneras, como realizar pruebas de medias, varianzas y chi-cuadrado.

Los métodos estadísticos que comienzan con recuentos de muestras, en su mayoría, descuidan una gran cantidad de información muy importante: la ubicación de los píxeles en la estego-imagen. Es intuitivamente claro que al utilizar las correlaciones espaciales en la estego-imagen, uno debería poder construir una detección mucho más confiable y precisa. Sin embargo, no es fácil descubrir y cuantificar la débil relación entre algunos componentes pseudoaleatorios presentes en la imagen y la imagen misma. Una vez que esta relación se cuantifica usando una medida, se podría estudiar cómo se presentan los cambios con la incrustación de mensajes. La relación derivada puede servir como base para las técnicas estegoanalíticas.

Las estadísticas de primer orden son todas las medidas que describen datos independientemente de las dependencias entre las muestras. En otras palabras, primer orden significa invariancia ante permutaciones arbitrarias de muestras. Esta propiedad se aplica a los histogramas y todas las cantidades que se pueden derivar de ellos (como las estadísticas de momentos).

Por el contrario, las estadísticas de orden superior consideran la relación entre muestras o su posición en el vector de datos. Incluyen, por ejemplo, medidas de correlación entre píxeles adyacentes en una imagen.

2.2.4.3.1 Modelos estadísticos de primer orden [15] [17]

Si bien la atención de los investigadores se centró en gran medida en el estudio de los detectores de orden superior, ya que se creía que modelos esteganográficos como el algoritmo MB1 [23] era seguro para ataques de detección con modelos estadísticos de primer orden, se tiene avances en el estudio de estos detectores.

Para describir la detección de mensajes usando un método estadísitco de primer orden se va a suponer que se tiene una imagen de cubierta con M×N píxeles y con valores de píxeles de un conjunto P y que previamente ha sido usado el algoritmo de LSB para incrustar un mensaje.

Para una imagen en escala de grises de 8 bits, $P=\{0, \ldots, 255\}$. Se captura las correlaciones espaciales utilizando una función de discriminación f que asigna un número real $f(x_1, \ldots, x_n) \in R$ a un grupo de píxeles $G=(x_1, \ldots, x_n)$. En este caso, se usa la función f definida en la

ecuación 2.3 la cual mide la suavidad de G y donde más ruidoso es el grupo G, mayor es el valor de la función de discriminación.

$$f(x_1, x_2, \dots, x_n) = \sum_{i=1}^{n-1} |x_{i+1} - x_i| \tag{2.3}$$

La incrustación de LSB aumenta el ruido en la imagen, por lo tanto, se espera que el valor de f aumente después de la incrustación de LSB. El proceso de incrustación de LSB se puede describir convenientemente utilizando una función de inversión F_1: $0 \leftrightarrow 1$, $2 \leftrightarrow 3$, ..., $254 \leftrightarrow 255$. Usar LSB en un nivel gris x es lo mismo que aplicar la inversión de F a x. También se define un concepto llamado LSB desplazado descrito en 2.4, que invierte F_{-1} como $-1 \leftrightarrow 0$, $1 \leftrightarrow 2$, $3 \leftrightarrow 4$, ..., $253 \leftrightarrow 254$, $255 \leftrightarrow 256$, finalmente, para completar se define también F_0 como identidad de permutación $F(x)=x$, $\forall x \in P$.

$$F_{-1}(x) = F_1(x+1) - 1 \; \forall x \tag{2.4}$$

La función de discriminación f y la operación de inversión F definen tres tipos de grupos de píxeles: R, S y U dependiendo de cómo cambia el valor de la función de discriminación. En general, es posible que se desee aplicar diferentes operaciones de inversión a diferentes píxeles en el grupo G.

En imágenes típicas (imágenes con varios colores), invertir el grupo G conducirá con mayor frecuencia a un aumento en la función de discriminación f en lugar de una disminución. Así, el número total de grupos regulares será mayor que el número total de grupos singulares. Si se denota el número relativo de grupos regulares para una máscara M no negativa como R_M (en porcentajes de todos los grupos) y sea S_M el número relativo de grupos singulares. Se tiene que:

$$R_M + S_M \leq 1 \qquad R_{-M} + S_{-M} \leq 1 \tag{2.5}$$

La hipótesis para detectar un posible mensaje oculto de este método esteganalítico es que, para imágenes de cubierta típicas, el valor de R_M es aproximadamente igual al de R_{-M}, y lo mismo debería ser cierto para S_M y S_{-M}.

$$R_M \cong R_{-M} \qquad S_M \cong S_{-M} \tag{2.6}$$

Se puede justificar esta hipótesis de forma heurística inspeccionando la ecuación 2.4. Usar la operación de inversión F_{-1} es lo mismo que aplicar F_1 a una imagen cuyos colores se han desplazado en uno (LSB). Debido a que la función de discriminación f captura la suavidad, agregar el valor de 1 a todos los píxeles no debería influir en las estadísticas de los grupos regulares (R) y singulares (S) de manera significativa. De hecho, se tiene amplia evidencia experimental de que la ecuación 2.6 es muy precisa para imágenes tomadas con

una cámara digital tanto para formatos JPEG (presenta compresión) como formatos sin comprimir. También funciona bien para imágenes procesadas con operaciones comunes de procesamiento de imágenes y para fotografías escaneadas.

La relación en la ecuación 2.6, sin embargo, se no se cumple después de aleatorizar el plano LSB (por ejemplo, debido a que se incrustan más bits usando mejoras del algoritmo de esteganografía LSB).

La aleatorización del plano LSB hace que la diferencia entre R_M y S_M sea cero a medida que aumenta la longitud del mensaje incrustado. Después de cambiar el LSB del 50% de los píxeles (que es lo que sucedería después de incrustar un bit de mensaje aleatorio en cada píxel), se obtiene que $R_M \cong S_M$ que es lo que se esperaría. Pero el plano LSB tiene el efecto opuesto en R_{-M} y S_{-M}, su diferencia aumenta con la longitud del mensaje incrustado.

Como consecuencia de esto, se han desarrollado otros modelos de detección mejorados donde se analizan gráficas que contienen el porcentaje de píxeles modificados y el número de grupos singulares y regulares presentes en la imagen.

2.2.4.3.2 Modelos estadísticos de orden superior [19] [20]

El enfoque adoptado aquí se basa en la construcción de modelos estadísticos de orden superior para imágenes digitales y en la búsqueda de desviaciones de dichos modelos. Con estos modelos y a través del uso de una gran cantidad de imágenes digitales, se puede demostrar que existen fuertes regularidades estadísticas de orden superior dentro de una descomposición tipo wavelet. La incrustación de un mensaje altera significativamente estas estadísticas y, por lo tanto, el mensaje oculto se vuelve detectable.

Para entender estos mecanismos se va a realizar la descomposición de imágenes utilizando funciones básicas que se localizan en posición espacial, orientación y escala; por ejemplo, wavelets ha demostrado ser extremadamente útil en una variedad de aplicaciones como: la compresión de imágenes, codificación de imágenes, eliminación de ruido y síntesis de texturas. Una razón para que sean tan populares es que tales descomposiciones exhiben regularidades estadísticas que pueden explotarse. A continuación, se describe una descomposición de este tipo y un conjunto de estadísticas recopiladas a partir de esta descomposición.

Esta descomposición se basa en filtros de espejo en cuadratura separables (QMF). Esta descomposición divide el espacio de frecuencias en múltiples escalas y orientaciones. Esto se logra mediante la aplicación de filtros pasa bajos y pasa altos separables a lo largo de los ejes de la imagen, generando una sub-banda vertical, horizontal, diagonal y de pasa

bajos. Las escalas posteriores se crean filtrando recursivamente la sub-banda de pasa bajos. Las sub-bandas verticales, horizontales y diagonales en la escala i = 1, ..., n se indican como $V_i(x, y)$, $H_i(x, y)$ y $D_i(x, y)$ respectivamente. En la figura 2.3 se muestra una descomposición en tres niveles de una imagen.

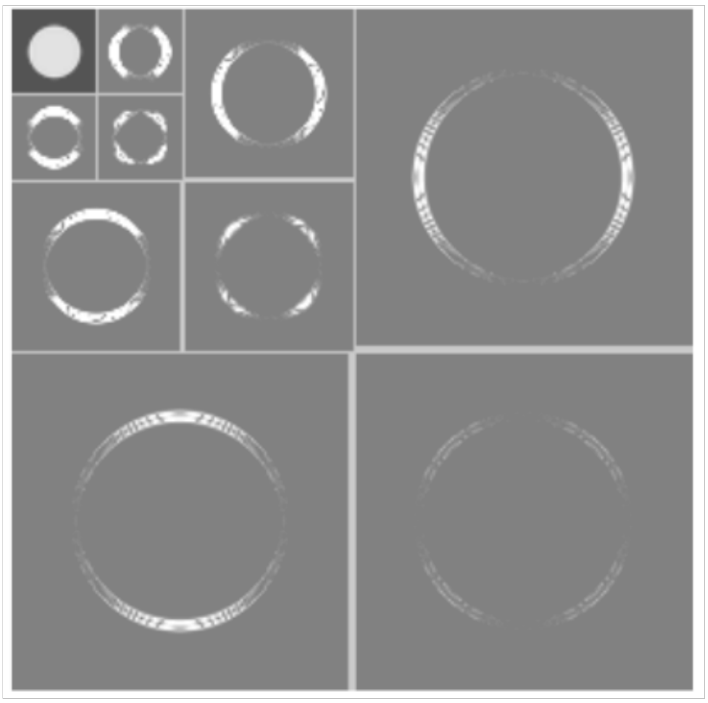

Figura 2.3. Sub-bandas presentes en la descomposición de una imagen

Dada esta descomposición de imágenes, el modelo estadístico se compone de la media, la varianza y la asimetría de los coeficientes de sub-banda en cada orientación y en las escalas i = 1, ..., n − 1. Estas estadísticas caracterizan las distribuciones de coeficientes básicos, el segundo conjunto de estadísticas se basa en los errores de un predictor lineal óptimo de la magnitud del coeficiente. Dichos coeficientes de sub-banda están correlacionados con sus vecinos de manera espacial, en la orientación y su escala.

Para entender de mejor manera este modelo, se considerará primero una banda vertical, $V_i(x, y)$, a escala i y un predictor lineal para la magnitud de estos coeficientes en un subconjunto de todos los vecinos posibles, que se representa por:

$$V_i(x,y) = w_1 V_i(x-1,y) + w_2 V_i(x+1,y)$$

$$+ w_3 V_i(x,y-1) + w_4 V_i(x,y+1)$$

$$+ w_5 V_{i+1}\left(\frac{x}{2},\frac{y}{2}\right) + w_6 D_i(x,y)$$

$$+w_7 D_{i+1}\left(\frac{x}{2},\frac{y}{2}\right) \tag{2.7}$$

Donde w_k son los valores de peso escalares. Esta relación lineal también se puede expresar de manera mejor en forma matricial como:

$$V = Qw \tag{2.8}$$

En la ecuación 2.8 se tiene que el vector columna w = $(w_1, \ldots, w_7)^T$ y el vector V contiene las magnitudes de los coeficientes de $V_i(x, y)$ de igual manera en un vector columna, y las columnas de la matriz Q contienen las magnitudes de los coeficientes vecinos como se especifica en la ecuación 2.6 también representados en vectores columna. Los coeficientes se determinan minimizando la función de error cuadrático como se presenta a continuación en la ecuación 2.9.

$$E(w) = [V - Qw]^2 \tag{2.9}$$

Para minimizar esta función de error, se utiliza la derivada respecto a w, como se representa en la ecuación 2.10. Si se resuelve esta ecuación y se iguala el resultado a 0 se puede obtener el error logarítmico del predictor lineal que viene dado en la ecuación 2.11.

$$\frac{dE(w)}{dw} = 2Q^T[V - Qw] \tag{2.10}$$

$$E = \log_2(V) - \log_2(|Qw|) \tag{2.11}$$

Es a partir de este error que se recopilan estadísticas adicionales como la media, la varianza y la asimetría. Este proceso se repite para cada sub-banda vertical a escalas i = 1, ..., n − 1, donde en cada escala se estima un nuevo predictor lineal. De igual manera todo este proceso se vuelve a realizar para el otro tipo de sub-bandas horizontales y diagonales. Se puede asumir que el predictor lineal para las sub-bandas horizontales y diagonales están representados por las siguientes ecuaciones.

$$\begin{aligned} H_i(x,y) &= w_1 H_i(x-1,y) + w_2 H_i(x+1,y) \\ &+ w_3 H_i(x,y-1) + w_4 H_i(x,y+1) \\ &+ w_5 H_{i+1}\left(\frac{x}{2},\frac{y}{2}\right) + w_6 D_i(x,y) \\ &+ w_7 D_{i+1}\left(\frac{x}{2},\frac{y}{2}\right) \end{aligned} \tag{2.12}$$

$$D_i(x,y) = w_1 D_i(x-1,y) + w_2 D_i(x+1,y)$$

$$+w_3 D_i(x, y-1) + w_4 D_i(x, y+1)$$

$$+w_5 D_{i+1}\left(\frac{x}{2}, \frac{y}{2}\right) + w_6 H_i(x, y)$$

$$+w_7 V_i(x, y) \tag{2.13}$$

La misma métrica de error que se presenta en la ecuación 2.11 puede ser usada tanto para los cálculos de las sub-bandas verticales como para las sub-bandas horizontales y diagonales. En este caso de ejemplo se obtienen se obtienen un total de 12(n−1) estadísticas de error.

La combinación de estas estadísticas con las estadísticas de coeficiente 12(n−1) produce un total de 24(n−1) estadísticas que forman un vector de características que se utiliza para discriminar entre imágenes que contienen mensajes ocultos y aquellas que no.

Al momento de combinar y correlacionar unas estadísticas con otras es donde se forman los modelos de orden superior.

2.2.4.4 Clasificadores

A partir de las estadísticas medidas de un conjunto de imágenes de entrenamiento con y sin mensajes ocultos, el objetivo es determinar si una imagen de prueba contiene un mensaje. Existen varios tipos de clasificadores; por ejemplo, se puede utilizar un análisis discriminante lineal de Fisher (FLD) [24].

Para el motivo de este estudo se van a presentar a las Máquinas de Soporte de Vectores como clasificadores más sencillos y flexibles. En complejidad creciente, existen tres clases de SVM.

El primer caso es una SVM lineal separable, la cual es matemáticamente la más sencilla. El segundo caso es una SVM lineal no separable, la cual se enfrenta a situaciones en las que no se puede encontrar una solución en el caso anterior y es más similar a un FLD.

El tercer caso una SVM no lineal que ofrece el esquema de clasificación más flexible y que para este método de presenta la mejor precisión de clasificación.

La siguiente figura describe de manera simplificada las dos clases de SVM lineales.

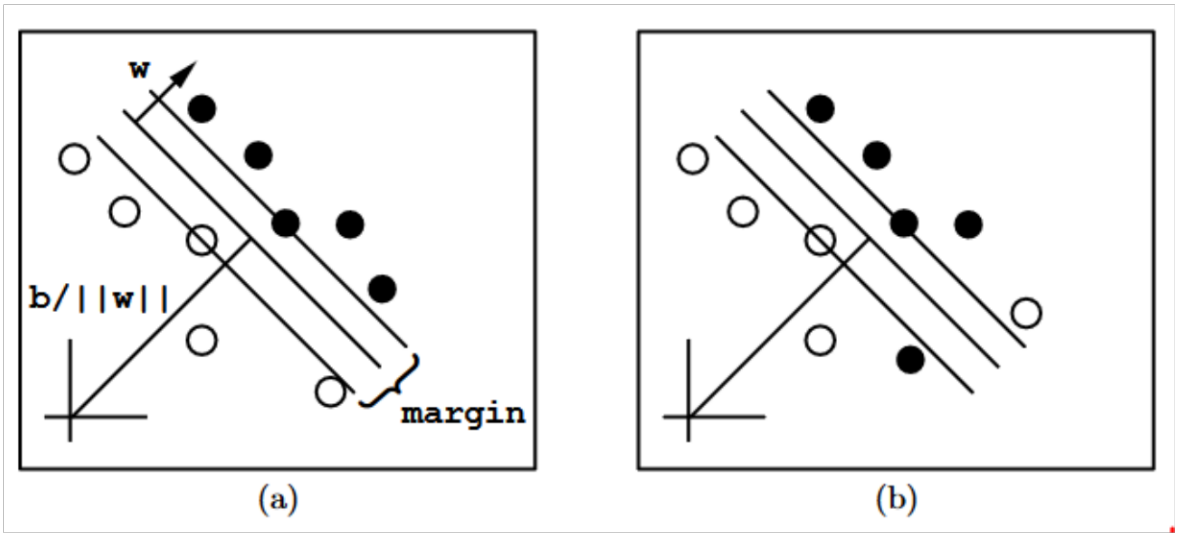

Figura 2.4. Representación de SVM lineales, en (a) SVM separable y en (b) SVM no separable

2.2.4.4.1 Máquinas de soporte de vectores (SVM) [26]

Los primeros algoritmos de aprendizaje automático tenían como objetivo aprender representaciones de funciones simples. Por lo tanto, el objetivo del aprendizaje era generar una hipótesis que realizara la clasificación correcta de los datos de entrenamiento y se diseñaron algoritmos de aprendizaje temprano para encontrar un ajuste preciso a los datos. La capacidad de una hipótesis para clasificar correctamente los datos que no están en el conjunto de entrenamiento se conoce como su generalización.

SVM es una técnica útil para la clasificación de datos, aunque se considera que las Redes Neuronales son más fáciles de usar que esto, sin embargo, en ocasiones se obtienen resultados insatisfactorios debido a que estas pueden llegar a generalizar de manera muy fácil. Una tarea de clasificación generalmente implica entrenamiento y prueba de datos que consisten en algunos conjuntos de datos. Cada instancia del conjunto de entrenamiento contiene un valor objetivo y varios atributos. El objetivo de SVM es producir un modelo que prediga el valor objetivo de las instancias de datos en el conjunto de prueba que reciben solo los atributos.

La clasificación en SVM es un ejemplo de aprendizaje supervisado. Los datos conocidos ayudan a indicar si el sistema está funcionando correctamente o no. Esta información apunta a una respuesta deseada, validando la precisión del sistema, o se puede utilizar para ayudar al sistema a aprender a actuar correctamente. Un paso en la clasificación de SVM implica la identificación de cuáles están íntimamente conectados con las clases conocidas. Esto se llama selección de características o extracción de características. La

selección de características y la clasificación SVM juntas tienen un uso incluso cuando la predicción de muestras desconocidas no es necesaria.

2.2.4.4.2 Redes Neuronales (NN) [24] [25]

Las NN son básicamente modelos computacionales paralelos masivos que imitan la función del cerebro humano. Una ANN consta de una gran cantidad de procesadores simples vinculados por conexiones ponderadas. La salida de cada nodo depende solo de la información que está disponible localmente en el nodo, ya sea almacenada internamente o que llega a través de las conexiones ponderadas.

Cada unidad recibe entradas de muchos otros nodos y transmite su salida a otros nodos. Por sí mismo, un solo elemento de procesamiento no es muy poderoso; genera una salida escalar con un solo valor numérico, que es una función no lineal simple de sus entradas.

Como se observa en la figura se presenta una entrada a la red neuronal y se establece una respuesta deseada u objetivo correspondiente en la salida para tener un enfoque de aprendizaje supervisado.

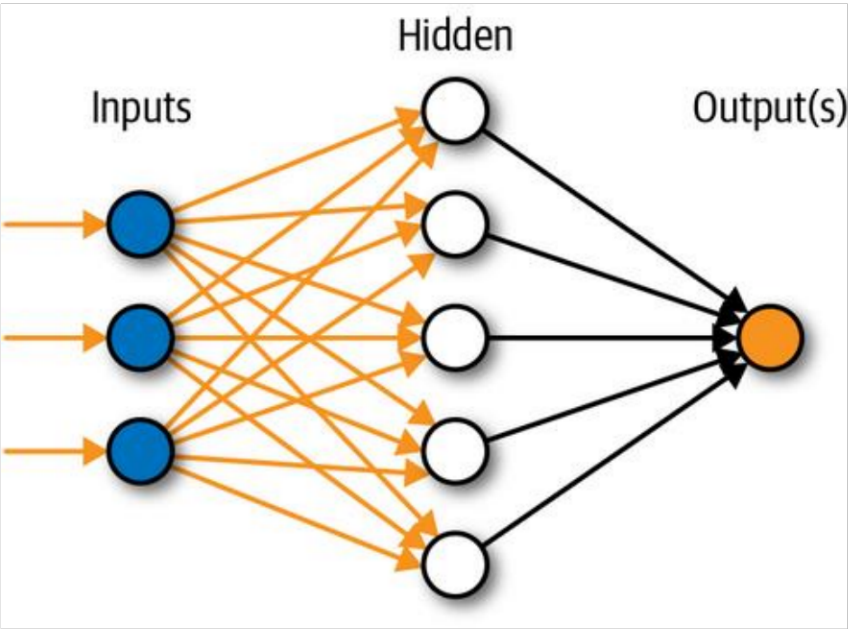

Figura 2.5. Representación básica del modelo de red neuronal

En cada nodo se obtiene un error que viene de la diferencia entre la respuesta deseada y la salida del sistema. Esta información de error se retroalimenta al sistema y ajusta los parámetros del sistema de manera sistemática (la regla de aprendizaje). El proceso se repite hasta que el rendimiento es aceptable. Está claro a partir de esta descripción que el rendimiento depende en gran medida de los datos. Si uno no tiene datos que cubran una parte significativa de las condiciones operativas o si son ruidosos, entonces la tecnología

de redes neuronales probablemente no sea la solución adecuada. Por otro lado, si hay muchos datos y el problema no se comprende bien para derivar un modelo aproximado, entonces la tecnología de redes neuronales es una buena opción. Este procedimiento de operación debe contrastarse con el diseño de ingeniería tradicional, hecho de especificaciones exhaustivas de subsistemas y protocolos de intercomunicación. En las redes neuronales artificiales, el diseñador elige la topología de la red, la función de rendimiento, la regla de aprendizaje y el criterio para detener la fase de entrenamiento, pero el sistema ajusta automáticamente los parámetros. Por lo tanto, es difícil incluir información a priori en el diseño, y cuando el sistema no funciona correctamente, también es difícil refinar la solución de forma incremental. Pero las soluciones basadas en ANN son extremadamente eficientes en términos de tiempo y recursos de desarrollo, y en muchos problemas difíciles, las redes neuronales artificiales brindan un rendimiento que es difícil de igualar con otras tecnologías. En la actualidad, las redes neuronales artificiales se están convirtiendo en la tecnología preferida para muchas aplicaciones, como el reconocimiento de patrones, la predicción, la identificación de sistemas y el control.

Al crear un modelo funcional de la neurona biológica, hay tres componentes básicos de importancia.

Primero, las sinapsis de la neurona se modelan como pesos. La fuerza de la conexión entre una entrada y una neurona se nota por el valor del peso. Los valores de peso negativos reflejan conexiones inhibitorias, mientras que los valores positivos designan conexiones excitatorias. Los siguientes dos componentes modelan la actividad real dentro de la célula neuronal. Un sumador suma todas las entradas modificadas por sus respectivos pesos. Esta actividad se conoce como combinación lineal. Finalmente, una función de activación controla la amplitud de la salida de la neurona. Un rango aceptable de salida suele estar entre 0 y 1, o -1 y 1.

El aprendizaje y la generalización son quizás los temas más importantes en la investigación de redes neuronales. El aprendizaje es la capacidad de aproximar el comportamiento subyacente de forma adaptativa a partir de los datos de entrenamiento, mientras que la generalización es la capacidad de predecir mucho más allá de los datos de entrenamiento. La potente capacidad de ajuste de datos o de aproximación de funciones de las redes neuronales también las hace susceptibles al problema de sobreajuste. El síntoma de un modelo de sobreajuste es que se ajusta muy bien a la muestra de entrenamiento, pero tiene poca capacidad de generalización cuando se usa con fines de predicción. La generalización es una característica más deseable y crítica porque el uso más común de un clasificador es hacer una buena predicción sobre objetos nuevos o desconocidos. Una

serie de problemas prácticos de diseño de redes relacionados con el aprendizaje y la generalización incluyen el tamaño de la red, el tamaño de la muestra, la selección del modelo y la selección de funciones. En la mayoría de los estudios se aborda estos problemas de aprendizaje y generalización dentro de un marco bayesiano general.

En general, un modelo simple o inflexible, como un clasificador lineal, puede no tener el poder de aprender lo suficiente sobre la relación subyacente y, por lo tanto, no ajustar los datos. Por otro lado, los modelos flexibles complejos, como las redes neuronales, tienden a sobreajustar los datos y hacen que el modelo sea inestable al extrapolar. Está claro que tanto el ajuste insuficiente como el sobreajustado afectarán la capacidad de generalización de un modelo. Por lo tanto, un modelo debe construirse de tal manera que solo el patrón sistemático subyacente de la población sea aprendido y representado por el modelo.

Los fenómenos de subajuste y sobreajuste en muchos procedimientos de modelado de datos pueden analizarse bien a través de la conocida descomposición de sesgo más varianza del error de predicción.

3 RESULTADOS, CONCLUSIONES Y RECOMENDACIONES

3.1 RESULTADOS

La mayoría de los algoritmos de incrustación actuales, como la incrustación de LSB en el dominio espacial o la modificación de los valores de los coeficientes DCT cuantificados en el dominio de la frecuencia (para imágenes JPEG) asumen de manera implícita que las modificaciones pequeñas y aleatorias de los valores del supuesto estego-documento no son necesariamente detectables porque esto solo representa el ruido que comúnmente se presenta en las imágenes al momento de la comunicación. Desafortunadamente, esta creencia bastante heurística es a menudo falsa. De hecho, los mecanismos de detección más complejos descritos en este trabajo sugieren que cualquier técnica esteganográfica de reemplazo de bits, como la incrustación de LSB en el dominio espacial o de frecuencia, o sus derivados, probablemente nunca puedan conducir a métodos más seguros para esconder información porque es difícil encontrar una regla general que identifique componentes pseudoaleatorios en imágenes (ruido) que puedan sustituirse por mensajes.

Los mensajes se pueden incrustar en imágenes digitales de manera imperceptible para el ojo humano, sin embargo, estos métodos de incrustación pueden alterar significativamente las estadísticas de una imagen. Para detectar la presencia de mensajes ocultos se ha estudiado un modelo estadístico de orden superior extraídos que se basa en la descomposición de una imagen en multiescalas. Este modelo incluye estadísticas de coeficientes básicos, así como estadísticas de error de un predictor lineal óptimo de la magnitud del coeficiente. Estas estadísticas de orden superior parecen capturar ciertas propiedades de las imágenes y lo principal es que estas estadísticas se modifican significativamente cuando se incrusta un mensaje en una imagen. Esto permite detectar, con un grado razonable de precisión bajo ciertas condiciones de prueba, la presencia de mensajes ocultos en imágenes digitales. Para evitar la detección, por supuesto, solo se necesita insertar un mensaje lo suficientemente pequeño que no altere significativamente las estadísticas de la imagen, pero esto nos lleva nuevamente al punto anterior donde si el mensaje incrustado es sumamente pequeño este puede ser confundido por los sistemas de detección como ruido en la estego-imagen.

Existen varios enfoques que se pueden explorar para mejorar la precisión de la detección. La elección particular del modelo estadístico es algo que puede llegar a ser importante en el método de detección, por lo que sería beneficioso elegir un conjunto de estadísticas que optimicen las tasas de detección.

La comparación indiscriminada de estadísticas usando como muestra una gran cantidad de imágenes podría reemplazarse con un análisis basado en clases, donde, por ejemplo, las escenas interiores y exteriores se comparan por separado.

Por último, aunque solo se estudió estos métodos de detección en imágenes (JPEG) como archivo de cubierta, no hay una razón alguna por la que estos modelos descritos aquí no funcionen para señales de audio o secuencias de video, formatos de archivo de imagen arbitrarios u otros algoritmos de ocultación de información. Cabe recalcar que estos modelos pueden servir como base para otros archivos de cubierta, esto es porque las estadísticas en cada tipo de cubierta son únicas y se deberían realizar los ajustes correspondientes a cada modelo estudiado.

Una ventaja del modelo de orden superior estudiado en el presente trabajo es que no se vuelva tan vulnerable a los contraataques para evitar la detección que coinciden con las distribuciones estadísticas de primer orden, como son: la intensidad de píxeles o los coeficientes de las transformadas (DCT). Sin embargo, es posible que se desarrollen algoritmos de incrustación más desarrollados que puedan frustrar el esquema de detección aquí estudiado. El desarrollo nuevas técnicas esteganográficas conducirá a su vez a mejores esquemas de detección, y así sucesivamente, formando un ciclo donde se desarrollan métodos esteganográficos más eficaces, es decir, robustos e imperceptibles; pero a su vez se desarrollan más métodos de detección que puedan descifrar estos algoritmos y determinar que existe información oculta en un objeto.

3.2 CONCLUSIONES

En este trabajo de integración curricular se trató en el marco teórico los conceptos y definiciones básicas correspondientes a la esteganografía, donde se presentó las maneras en que se pueden realizar las técnicas esteganográficas. En este mismo capítulo se dio a conocer las características que afectan a que un algoritmo esteganográfico sea mejor, como son su robustez, imperceptibilidad, el tamaño del mensaje incrustado, etc.

En el capítulo de Metodología se estudió como se desarrolla el estegoanálisis en la actualidad, dando principal enfoque al problema que se presenta cuando se desarrollan métodos de detección de manera general, es decir, sin conocer con anterioridad el archivo de cubierta que fue usado para incrustar información.

Existen varios tipos de ataques a los métodos de esteganografía. Estos van desde la simple inspección (analizar el LSB) que sirve para analizar diferentes tipos de archivos de cubierta, utilizar los histogramas de los archivos de cubierta y compararlos con los histogramas de los estego-documentos, hasta el desarrollo de diferentes modelos estadísticos que usan propiedades presentes en la descomposición de los estego-documentos para determinar mediante funciones errores en dichas estadísticas y así determinar si existe un mensaje incrustado.

Para el estegoanálisis es difícil decir que existe un método válido para aplicar en todos los casos que existe un mensaje oculto, por lo que las pruebas que se aplican a los métodos previamente estudiados son desarrolladas bajo ciertas condiciones y son difíciles de adaptar a casos reales y generalizados. A diferencia de la esteganografía donde sus métodos se encuentran bastante extendidos y por eso existen programas computacionales (software) que aplica estos algoritmos de una manera sencilla.

El continuo estudio de la esteganografía ha hecho que se desarrollen nuevos algoritmos cada vez más eficaces, esto quiere decir que son robustos e indetectables lo que ocasiona a la persona a cargo del estegoanálisis que los método de detección desarrollados con modelos estadísticos de primer orden queden casi obsoletos debido a que los mensajes incrustados causan pocos cambios en las estadísticas de la cubierta y estos puedan pasar desapercibidos, esto llevo al uso de modelos estadísticos de modelos superior, los cuales correlaciona más parámetros presentes en las estadísticas de los archivos de cubierta para que los mensajes puedan ser detectados.

Otro método desarrollado para la detección de mensajes cuando se usa esteganografía es el del uso de clasificadores para determinar si un archivo dado contiene información oculta o no. Esto incluye usar tecnologías como máquinas de soporte de vectores (SVM) o redes

neuronales artificiales (NN), las cuales usan conceptos de aprendizaje automático para la detección de mensajes ocultos; un problema de estas tecnologías es la generalización que se puede obtener al usarlas, es decir, que por el aprendizaje que estas tienen lleguen a equivocarse y determinar que algunos estego-documentos no contienen información oculta pese a que ello es totalmente falso.

Los clasificadores son de gran ayuda cuando se quiere detectar mensajes ocultos dado una basta cantidad de estego-documentos donde los mensajes han sido ocultados haciendo uso de una gran variedad de algoritmos de esteganografía por lo que se hace imposible usar determinado método de detección debido a la variedad de técnicas usadas.

3.3 RECOMENDACIONES

Para entender a mayor profundidad no solo los métodos estudiados en este trabajo, sino seguir con el estudio de la esteganografía y el estegoanálisis es fundamental tener un amplio conocimiento en álgebra lineal y estadística; pues en todos los métodos de detección y algoritmos de incrustación se utilizan funciones, distribuciones probabilísticas o transformaciones lineales.

Como se menciona en la mayoría de los métodos de esteganografía estos se vuelven de mejorer características cuando son usados en archivos multimedia de cubierta grandes; por ejemplo, los mensajes ocultos son más indetectables cuando son incrustados en imágenes que contienen gran cantidad o niveles de colores o en audios que tienen una gran tasa de btis.

En contraposición con el punto anterior es recomendable que los sistemas donde se van a procesar los estego-documentos deben ser de gran capacidad computacional, debido a que el mensaje incrustado puede ser muy grande y se necesite realizar un gran número de operaciones en el estego-documento. A esto se debe adicionar que dichas operaciones son descomposiciones o transformaciones a nivel estadístico que son complejas.

4 REFERENCIAS BIBLIOGRÁFICAS

[1] F. Djebbar, B. Ayad, K. A. Meraim, y H. Hamam, “Comparative study of digital audio steganography techniques”, *Eurasip J. Audio, Speech, Music Process.*, vol. 2012, núm. 1, pp. 1–16, 2012, doi: 10.1186/1687-4722-2012-25.

[2] J. C. Mendoza T., “Demostración de Cifrado Simétrico y Asimétrico”, *Ingenius*, pp. 46–53, 2008, [En línea]. Disponible en: https://dspace.ups.edu.ec/bitstream/123456789/8185/1/Demostración de cifrado simétrico y asimétrico.pdf.

[3] N. F. Johnson, Z. Duric, S. Jajodia, y N. Memon, *Information Hiding: Steganography and Watermarking—Attacks and Countermeasures*, vol. 10, núm. 3. 2001.

[4] R. Cramer y V. D. Gligor, *Advanced Statistical Steganalysis*. 2010.

[5] A. Mangarae, “Steganography FAQ”, *Most*, 2006.

[6] J. Fridrich, *Steganography in Digital Media*. 2009.

[7] F. Y. Shih, *Digital watermarking and steganography: Fundamentals and techniques*. 2017.

[8] M. Mudassar, A. Kumar, y K. Pooja, “Steganography-A Data Hiding Technique”, *Int. J. Comput. Appl.*, vol. 9, núm. 7, pp. 975–8887, 2010.

[9] Sheelu y B. Ahuja, “An Overview of Steganography”, vol. 11, núm. July, pp. 15–19, 2007.

[10] J. Kour y D. Verma, “Steganography Techniques: A Review”, *Int. J. Emerg. Res. Manag. &Technology*, vol. 3, núm. 5, pp. 132–135, 2014.

[11] G. E. ONOFRE CONCHA, “DESARROLLO Y ANÁLISIS DE UNA TÉCNICA ESTEGANOGRÁFICA EN ZONAS RUIDOSAS DE LA IMAGEN MEDIANTE TRANSFORMACIONES DE COLOR REVERSIBLES”, 2016.

[12] M. John, *Compressed Image File Formats: Jpeg, png, gif, xbm, bmp.* 1999.

[13] N. F. Johnson y G. Mason, “Exploring Steganography Seeing the Unseen”, *IEEE Trans. Image Process.*, p. 26, 1998.

[14] J. K. Mandal, *Reversible Steganography and Authentication via Transform Encoding*, vol. 901. 2020.

[15] E. Cole, *Hiding in Plain Sight : Steganography and the Art of Covert Communication*. 2003.

[16] A. Westfeld y A. Pfitzmann, “Attacks on Steganographic Systems Breaking the Steganographic Utilities EzStego, Jsteg, Steganos, and S-Tools—and Some Lessons Learned”, *Inf. Hiding*, vol. 1768, pp. 61–76, 2000.

[17] J. C. L. HERNÁNDEZ, “ESTEGOANÁLISIS DE IMÁGENES DIGITALES USANDO TÉCNICA DE RECONOCIMIENTO DE PATRONES”, 2009.

[18] J. de J. Serrano-Pérez, M. Salinas-Rosales, y N. Cruz-Cortés, “Sistema inmune artificial para estegoanálisis de imágenes JPEG”, *Res. Comput. Sci.*, vol. 114, núm. 1, pp. 187–199, 2016, doi: 10.13053/rcs-114-1-15.

[19] and T. K. Ingemar J. Cox, Matthew L. Miller, Jeffrey A. Bloom, Jessica Fridrich, *Digital Watermarking and Steganography*. 2008.

[20] O. Dabeer, K. Sullivan, U. Madhow, y S. Chandrasekharan, “Detection of Hiding in the Least Significant Bit”, *Inf. Sci. (Ny).*, pp. 1–4, 2003.

[21] H. Farid, “Detecting Steganographic Messages in Digital Images”, *Electron. Imaging 2004*, vol. 2, núm. 6, pp. 1–4, 2013, [En línea]. Disponible en: http://www.tjprc.org/view-archives.php?year=2014&id=15&jtype=2&page=7%0Ahttp://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=837665%0Ahttp://ijacsa.thesai.org/%0Ahttps://repo.zenk-security.com/Cryptographie . Algorithmes . Steganographie/EN.

[22] A. Razaghpanah *et al.*, “Apps, Trackers, Privacy, and Regulators: A Global Study of the Mobile Tracking Ecosystem”, feb. 2018, doi: 10.14722/ndss.2018.23353.

[23] D. Hu, L. Wang, X. Jiang, T. Zhu, y Y. Yue, “Detecting the MB1 with higher-order statistics”, *Proc. - 2008 Int. Conf. Comput. Intell. Secur. CIS 2008*, vol. 2, pp. 330–333, 2008, doi: 10.1109/cis.2008.35.

[24] H. Farid, “DETECTING HIDDEN MESSAGES USING HIGHER-ORDER STATISTICAL MODELS”, *Computer (Long. Beach. Calif).*, pp. 2–5, 2000.

[25] S. Lyu y H. Farid, “Detecting hidden messages using higher-order statistics and support vector machines”, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2578, pp. 340–354, 2003, doi: 10.1007/3-540-36415-3_22.

[26] V. Jakkula, “Tutorial on Support Vector Machine (SVM)”, *Sch. EECS, Washingt. State Univ.*, pp. 1–13, 2011, [En línea]. Disponible en: http://www.ccs.neu.edu/course/cs5100f11/resources/jakkula.pdf.

[27] A. D. Dongare, R. R. Kharde, y A. D. Kachare, “Introduction to Artificial Neural Network (ANN)”, *Int. J. Eng. Innov. Technol.*, vol. 2, núm. 1, pp. 189–194, 2012, [En línea]. Disponible en: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1082.1323&rep=rep1&type=pdf.

[28] G. P. Zhang, “Neural networks for classification: A survey”, *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 30, núm. 4, pp. 451–462, 2000, doi: 10.1109/5326.897072.