Sistema de cifrado para seguridad de archivos con detección de voz y cifrado con colores RGB

Trabajo Terminal No. 2021-A031

*Alumnos: De La Cruz Flores Daniel, *Leyva Benítez Alexis*
Directores: López Rojas Ariel, Durán Camarillo Edmundo René
**e-mail: alexisleyvacetis17@gmail.com*

Resumen – La implementación de sistemas que preserven la información en formato digital de las personas se ha convertido en una necesidad con el paso de los años debido a los altos incrementos en los robos de esta índole en recientes fechas. Se propone desarrollar un sistema de cifrado y descifrado a través de la voz del usuario, misma que será captada por el micrófono del equipo donde se esté utilizando el sistema, en conjunto con otro factor de validación formado por una combinación en formato RGB elegida por el usuario, esto con el fin de preservar la seguridad principalmente de información sensible. Dicho proceso se realizará utilizando el algoritmo DES (Data Encryption Standard), el cual utilizará la información obtenida del procesamiento de la voz y la combinación RGB (Red, Green, Blue) del color que haya elegido el usuario.

Palabras clave – Cifrado y descifrado de información, validación de información, detección procesamiento de voz, colores RGB.

1. Introducción

Con el paso del tiempo y las nuevas condiciones en las que actualmente se vive, debido a la contingencia, las tareas y actividades de la vida cotidiana, ya sean sociales o profesionales, han pasado a ser desarrolladas en una nueva modalidad donde se requieren herramientas y recursos, en su mayoría, tecnológicos, de manera que las personas puedan mantenerse en constante comunicación.

Debido a que la comunicación entre las personas se desarrolla de manera indirecta, o bien, haciendo uso de las herramientas tecnológicas necesarias para poder comunicarse y realizar las tareas que satisfagan cada una de las necesidades de las personas, la seguridad de la información, por supuesto, está más limitada, debido a que la información, al viajar por la red, tiende a estar más vulnerable a que algún tercero pueda acceder a ella y usarla de manera inadecuada.

Atendiendo a estos problemas, a lo largo de los años, se han implementado nuevas tecnologías, que hacen que cada vez sea más segura la transmisión de información a través de la red, sin embargo, no solo depende de lo seguro que pueda ser el medio por el que se transmite la información, sino que, en general, la información más sensible suele guardarse en aplicaciones o sitios seguros que requieran de un acceso restringido (login), por lo que cualquier persona que sepa estos datos, puede acceder a la información.

La implementación de nuevos sistemas que permitan establecer una mayor seguridad en la información de los archivos y/o documentos, de forma que el acceso a esta información sea única y exclusivamente para una persona, haciendo uso de una característica, la cual es exclusiva y diferente para todas las personas: la voz, es una alternativa que puede resultar viable para poder realizar la identificación de usuarios, sin embargo, se debe tomar en cuenta los ataques que el sistema puede tener por usuarios no autorizados, que pudiesen a la información de otros mediante diversos medios, por ejemplo: grabaciones y/o reproducciones de audio. Por lo tanto, es recomendable tomar en cuenta una nueva alternativa que pueda ofrecer mayores privilegios de seguridad, como puede ser la validación por algún factor, que, a su vez, pueda ser utilizado como herramienta de implementación dentro del algoritmo de cifrado de la información, por ejemplo: combinación de colores, de las cuales existen aproximadamente 16 581 375 combinaciones que representan a cada uno de los colores en formato RGB [1], por lo tanto, su descifrado sería mucho más complicado.

2. Objetivo

Implementar un prototipo de sistema de cifrado de archivos de un tamaño de hasta 10 MB, utilizando en conjunto, la voz del propietario de dichos documentos, misma que se obtendrá a través del micrófono que viene contenido de fábrica en el equipo para trabajar con esta información y poder realizar el proceso correspondiente para verificar la identidad del usuario, aunado a esto, se utilizará una combinación en formato RGB que represente algún color en específico, como una llave que permita cifrar y descifrar dicha información, dicha llave será generada a partir de la combinación RGB del color que el usuario seleccione en una paleta de colores, de forma que del color seleccionado se genere su combinación RGB y esta sea utilizada como factor de validación y clave para el algoritmo DES, y así poder continuar con el proceso de cifrado o descifrado.

Objetivos específicos:

- Diseño e implementación de la base de datos para la gestión de archivos de hasta 10 MB durante el proceso de cifrado.
- Programación del sistema para procesamiento de información a través de combinación de colores RGB.
- Desarrollo del tratamiento de la voz con uso de un DSP.
- Programación del sistema para cifrado de archivos y procesamiento de voz con uso del algoritmo de cifrado DES.
- Desarrollo e implementación de una aplicación de escritorio para el usuario.

3. Justificación

De acuerdo con una cifra arrojada por Kaspersky, sólo un tercio de los latinoamericanos almacena información sensible en la nube, específicamente hablando de México el 57% preferiría no guardar información en la nube por considerarlo peligroso. [2]

Como ya informó IT Digital Security, INCIBE-CERT emitió en 2018 un total de 228 avisos de vulnerabilidades relacionadas con los sistemas de control industrial, cifra que constata un aumento del 28% con respecto a 2018. Por grado de severidad, el 33% de los avisos correspondieron a vulnerabilidades críticas; el 43% fueron de nivel alto; el 22% de nivel medio y el 2% restante de nivel bajo. Hubo una serie de vulnerabilidades más frecuentes. En este sentido, como muestra el gráfico, primaron las relacionadas con la obtención de información sensible se mantienen en la primera. [3]

El robo de información sensible o los virus para inutilizar sistemas son las ciber amenazas que mayor daño generan en las organizaciones, en el año 2019 causaron pérdidas económicas que ascienden en promedio a los 7.7 millones de dólares anuales en México.[4]

La situación en cuanto a robos de esta índole ha ido en aumento conforme al paso del tiempo, cada vez más personas desconfían de las plataformas para almacenar información personal. Si bien estas plataformas son seguras nunca existirá un sistema que sea invulnerable, aunado a esto, las plataformas de almacenamiento tienen un costo mensual o anual por cierta capacidad de almacenamiento.

Este proyecto brindará seguridad a las personas, una manera en la que se pueda asegurar la información es a través de su voz tanto para cifrar su información como para descifrarla.

La voz como un validador es recomendable puesto que a pesar de que el sonido son ondas y esto significaría que podrían existir voces idénticas, el sonido una vez que sale de la laringe sufre una serie de modificaciones resonanciales nasobuco-faríngeas. Estas modificaciones consisten en el aumento de la frecuencia de ciertos sonidos y la desvalorización de otros. Esto depende de muchos factores que dan lugar al timbre y la calidad vocal.

Esto significa que hay distintas calidades de sonidos que las personas emiten, constituyendo un aspecto de la singularidad individual, lo que hace particular y singular a la voz del individuo, dependiendo sus características de tres elementos: altura, intensidad y timbre. Estas características o cualidades de la voz humana, una vez que sale de los resonadores, son modificadas y moldeadas por los articuladores, transformándose en sonidos del habla: fonemas, sílabas, palabras.[5]

Aunado a esto, se implementará un cifrado a través del Pantone de colores RGB en busca de aumentar la seguridad del sistema.

4. Productos o Resultados esperados.

El sistema estará compuesto por un conjunto de módulos muy importante que realizarán una determinada tarea, el primero consiste en un gestor de archivos donde el usuario pueda cargar archivos que desee enviar, o bien, visualizar aquellos que ha recibido, dichos archivos deberán tener una tamaño de hasta 10 MB, el segundo módulo será un sensor detector de voz que procesará el espectro para poder trabajar con la información que éste emita, además de que mediante una paleta de colores con su respectiva combinación RGB cada uno ellos, se seleccione el color que permita continuar con el proceso de cifrado. Una vez detectado el espectro de la voz y la respectiva combinación RGB, se trabajará con el tercer módulo, el cual consiste en analizar el espectro para poder trabajar con él, de manera que se generen llaves, y a su vez, se cifre la información que contienen los archivos mediante un algoritmo en específico. Con el archivo ya cifrado, el cuarto módulo consistirá en un control de almacenamiento, donde se guardará el archivo cifrado en el equipo del usuario. Una vez guardado el archivo cifrado, se pasará al proceso de la detección del espectro de la voz y la respectiva combinación RGB que validé los argumentos para el descifrado eficiente del archivo (quinto módulo), esto para poder obtener la información necesaria para intentar descifrar el documento (sexto módulo) y así, visualizarla para poder descargar el archivo que se está utilizando (séptimo módulo).

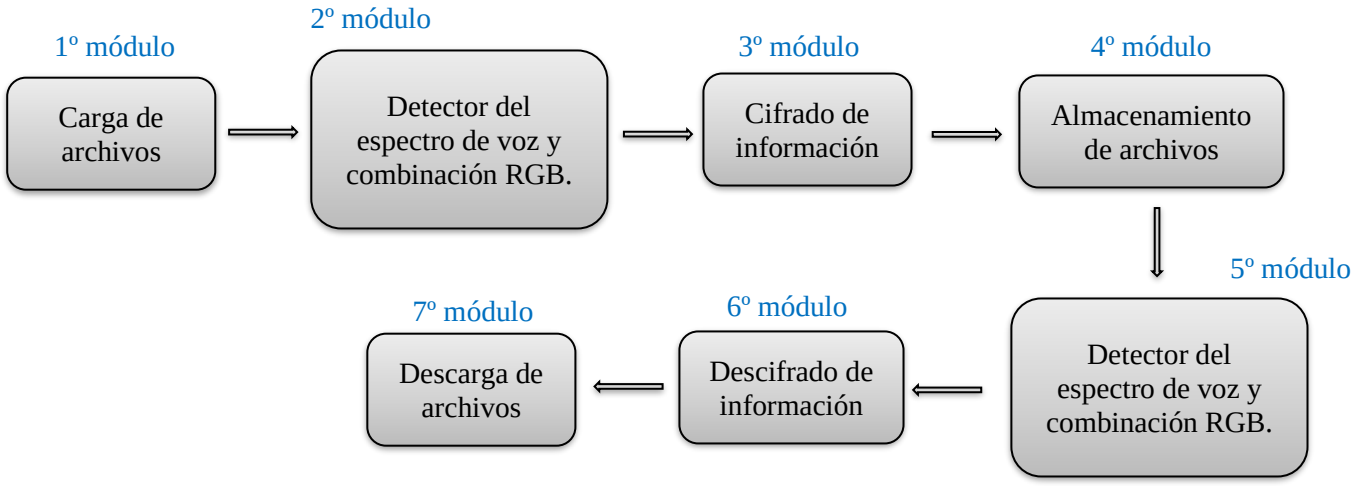

Figura 1. Diagrama de bloques del Software.

1. Software que permite cifrar y descifrar archivos.
2. Detector del espectro de voz de los usuarios y procesamiento de la información adquirida de éste.
3. Detección de combinaciones RGB para validación e identificación de entradas de información y posterior cifrado o descifrado de información.
4. Implementación del algoritmo de cifrado DES para poder proteger el contenido de los archivos.
5. Aplicación que permita asegurar documentos (cifrar y descifrar) para la protección de información sensible.
6. Documentación técnica del sistema.
7. Manual de usuario.

5. Metodología.

Para el desarrollo del sistema de software, se implementarán las estrategias de la metodología Scrum, la cual está enfocada y es aplicada especialmente para el desarrollo de sistemas de software [6], y destaca como una de las más utilizadas en el desarrollo de sistemas de software [7]. El proceso partirá de un módulo de Backlog, donde se definirán los requisitos del sistema con base en cada una de sus funcionalidades. Una vez planteado lo anterior, se pasa a una etapa de clasificación y ordenamiento de estos requerimientos (Sprint Backlog), para poder comenzar con la etapa de desarrollo, donde habrá un conjunto de actividades especiales para cada factor dentro del desarrollo, hasta que se pueda obtener el resultado esperado en el tiempo que se estimó ser obtenido. La especificación y orden de cada etapa se muestra en la siguiente figura.

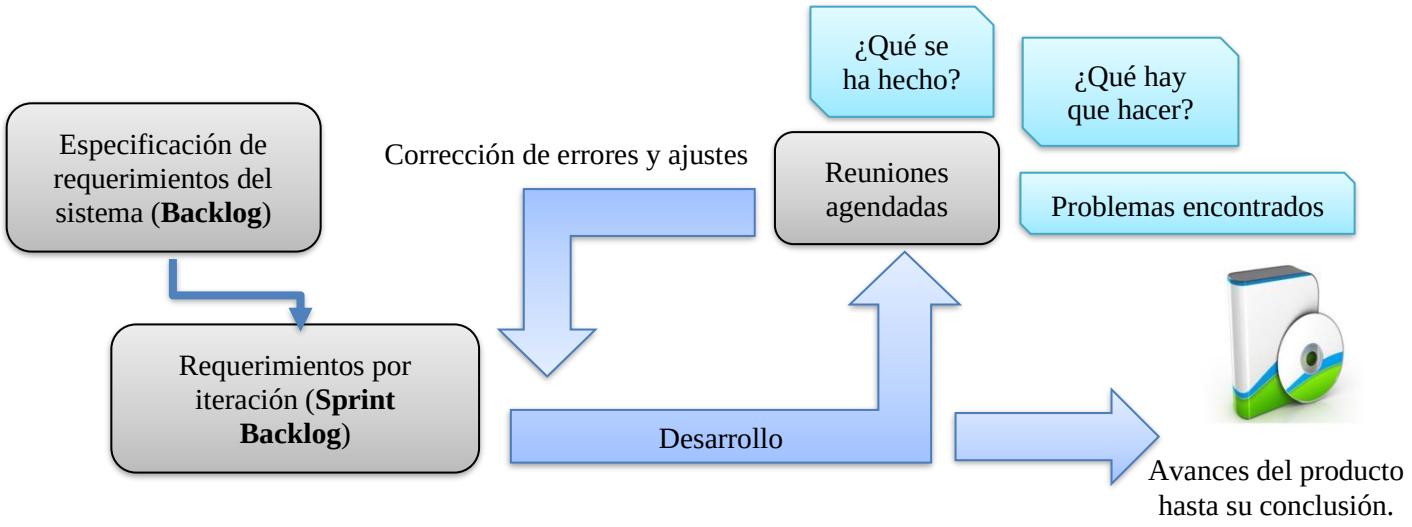

Figura 1. Metodología Scrum.

El proceso inicia con la especificación de los requerimientos del sistema de software, con el fin de que una vez planteados, se puedan plantear las estrategias pertinentes a implementar para satisfacer cada uno de ellos. Posteriormente, con base en cada uno de los requerimientos planteados, se procede a analizar cada uno de ellos, de manera que puedan ser separados y asignados a un subconjunto, donde habrá requisitos con características en común, que puedan ser atendidos bajo ciertas condiciones en común, lo cual dará una mejor distribución de trabajo, de manera que el proceso sea más manejable. Ahora, después de toda la planificación, se procede a implementar cada una de las estrategias planteadas en la etapa de desarrollo, donde se harán sesiones o reuniones con el fin de observar los avances, si hay o no complicaciones, o qué hay que cambiar en el proceso para hacer más eficiente el mismo. Este proceso es realizado de manera cíclica durante el tiempo estimado de desarrollo, tomando en cuenta que en determinado tiempo hay que mostrar avances del producto para analizar su funcionamiento en ese instante, hasta poder obtener el resultado final. Para lograrlo contemplamos realizar 6 sprints, 3 para la primera parte del proyecto y 3 para la segunda parte, realizándolos con una frecuencia de 20 días de manera que se pueda distribuir de manera uniforme en la duración del proyecto. Estos sprints tendrán los siguientes propósitos y características:

Sprint 1. Diseño de la base de datos para almacenamiento de información y análisis sobre la implementación y uso de recursos para la identificación y tratamiento de voz mediante el uso del DSP.
Sprint 2. Creación de base de datos y análisis de alternativas viables para la selección del proceso de recepción de combinaciones RGB.
Sprint 3. Elección acerca del algoritmo más eficiente a implementar, con base en la información disponible para cifrar y descifrar, para el posterior desarrollo del software.
Sprint 4. Observación de resultados ante la implementación del algoritmo de cifrado y descifrado aplicado a los archivos del software.
Sprint 5. Observación de resultados ante la corrección de errores e implementaciones faltantes sobre el proyecto.
Sprint 6. Preparación de presentación del producto final.

6. Cronograma.

Nombre del alumno: De La Cruz Flores Daniel

Actividad	**Ago**	**Sep**	**Oct**	**Nov**	**Dic**	**Feb**	**Mar**	**Abr**	**May**	**Jun**
Investigación de métodos para captar audio con dispositivo DSP.										
Investigación del funcionamiento e implementación de cifrados.										
Evaluación de TT I										
Implementación de la base de datos										
Implementación de algoritmo para validar entradas de información con pantone de colores										
Implementación de algoritmo DES para cifrado de archivos										
Desarrollo de aplicación de prueba										
Pruebas conjuntas del sistema.										
Generación de manual de usuario.										
Generación del reporte técnico										
Evaluación de TT II										

Nombre del alumno: Leyva Benítez Alexis

Actividad	**Ago**	**Sep**	**Oct**	**Nov**	**Dic**	**Feb**	**Mar**	**Abr**	**May**	**Jun**
Investigación de funcionamiento y métodos para tratamiento de audio.										
Investigación de algoritmo para validación de información a través de pantone de colores.										
Evaluación de TT I										
Configuración de dispositivo DSP										
Implementación de algoritmo para captación de audio										
Implementación de algoritmo para tratamiento de audio										
Desarrollo de aplicación de prueba										
Pruebas conjuntas al sistema										
Generación de manual de usuario.										
Generación del reporte técnico										
Evaluación de TT II										

7. Referencias.

[1] Fotonostra.com. 2021. *Modelo de colores RGB, CMYK y sRGB*. [online] Available at: <https://www.fotonostra.com/grafico/rgb.htm> [Accessed 2 April 2021].
[2] eSemanal - Noticias del Canal. 2021. "Un tercio de los latinoamericanos almacena información sensible y fotos íntimas en la nube". [online] Available: https://esemanal.mx/2021/02/un-tercio-de-los-latinoamericanos-almacena-informacion-sensible-y-fotos-intimas-en-la-nube/ [Accessed: 2 March 2021].
[3] Group, I., 2021. "INCIBE-CERT hace balance de los ataques a los sistemas industriales de 2018". [online]. Available: https://www.ituser.es/seguridad/2019/02/incibecert-hace-balance-de-los-ataques-a-los-sistemas-industriales-de-2018/. [Accessed: 2 March 2021].
[4] México, 2021. "*En México el robo de información provoca pérdidas anuales por 7.7 mdd"*. [online] Milenio.com. Available: https://www.milenio.com/tecnologia/mexico-robo-informacion-provoca-perdidas-anuales-7-7-mdd [Accessed 3 March 2021].
[5] Castañeda, P. F. (s. f.). *"Calidad de Voz y sus Diferencias en las Personas".* [online] Biblioteca Central Pedro Zulen. Available: https://sisbib.unmsm.edu.pe/bibvirtual/libros/linguistica/leng_niño/cal_voz_diferen_pers.htm
[6] Abellán, E., 2021. *Metodología Scrum: qué es y cómo funciona*. [online] Wearemarketing.com. Available at: https://www.wearemarketing.com/es/blog/metodologia-scrum-que-es-y-como-funciona.html#:~:text=Scrum%20es%20una%20metodolog%C3%ADa%20de,en%20iteraciones%20cortas%20de%20tiempo.&text=Esto%20permite%20al%20cliente%2C%20junto,obtener%20ventas%20(Sales%20enablement). [Accessed 17 March 2021].
[7] Maldonado, M., 2018. *Las mejores metodologías ágiles para la creación de software*. [online] DIGITAL55. Available at: <https://www.digital55.com/desarrollo-tecnologia/mejores-metodologias-agiles-creacion-software/> [Accessed 5 April 2021].
[8] Ramírez Ramírez Hugo Alberto, Villalba Valdez Jorge Antonio. "Criptosistema de seguridad simétrico mediante el algoritmo DES y reconocimiento de voz", Trabajo Terminal, Escuela Superior de Cómputo - IPN, Ciudad de México, 2017.

8. Alumnos y directores.

De La Cruz Flores Daniel.- Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta: 2015080305 , Tel. 5534418339 , email cruzfloresdaniel@gmail.com

Firma:_____

Leyva Benítez Alexis.- Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta: 2018630791, Tel. 5543781634 , email alexisleyvacetis17@gmail.com

Firma:_____

López Rojas Ariel.- Docente en ESCOM, Ext 52032, email: arilopez@ipn.mx

Firma:_____

Durán Camarillo Edmundo René.- M. en C. del CINVESTAV-IPN en 1994, Ingeniero en Electrónica del Instituto Tecnológico de Orizaba, Veracruz en 1993, Docente en ESCOM. Áreas de Interés: Sistemas Neurodifusos y Redes Neuronales y Artificiales, Ext 52037, email: eduranc@ipn.mx

Firma:_____

Buenas noches estimados profesores, el motivo de este mensaje es para presentar, a nombre de su servidor Alexis Leyva Benítez y mi compañero Daniel De La Cruz Flores, el Protocolo final para nuestro Trabajo Terminal, con el fin poder compartirlo con ustedes, y a su vez, obtener el acuse de recibo para poder realizar la inscripción del Trabajo Terminal. De antemano les agradecemos mucho su apoyo y quedamos en espera de su respuesta.

A

ALR ARIEL-ESCOM-IPN 3/6/2021
para alexis.leyva.cetis17, Edmundo

Muy buenas noches,

Acuso de recibido y de Vo.Bo.

Saludos
Ariel López Rojas

Mostrar texto citado

E

Edmundo Rene Duran Camarillo 3/6/2021
para ALR, alexis.leyva.cetis17@gmail.com

Hola Alexis

Acuso de recibido y de Vo.Bo.

Gracias y buenas noches.

Saludos.
Edmundo René