07-003

DISTRIBUTED STORAGE SYSTEM TO SAFEGUARD AND STORE ENCRYPTED INFORMATION APPLYING CHAOS

Jiménez Rodríguez, Maricela [(1)]; Aguilar Santiago, Jorge [(2)]; González Novoa, María Guadalupe [(2)]; Flores Jiménez, Ariadna Berenice [(2)]; Gómez Rodríguez, Horacio [(3)]

[(1)] Universidad de Guadalajara, [(2)] Centro Universitario de la Ciénega, [(3)] Centro Universitario de los Altos

Currently, there is a requirement to implement different techniques in order to safeguard data from ill-intentioned users that can employ software tools to intercept any information traversing through the Internet or that are stored within a server without any form of encryption. Thus, this paper proposes a network or distributed storage system with a point-to-multipoint or mesh topology, in which there are several mobile devices, PC, and single-board computers (SBC), and storage media such as clouds, raids, flash drives, among others. On the network, one or more origin nodes collect information from photographic cameras, fingerprint readers, sensors, audio recorders, or any type of digital file; afterward, the developed procedure is employed to apply diffusion and confusion techniques to encrypt the data. Next, the cryptogram is segmented chaotically, generating fragments of different sizes and content, which can be transmitted and stored in a distributed manner or redundantly among other nodes on the net. The system provides data security and integrity, in that it allows the recovery of 100% of the encrypted information on employing the correct encryption keys.

Keywords: chaos; encryption; distributed storage

SISTEMA DE ALMACENAMIENTO DISTRIBUIDO PARA SALVAGUARDAR Y ALMACENAR INFORMACIÓN CODIFICADA APLICANDO CAOS

Hoy en día es necesario implementar diferentes técnicas para salvaguardar la información de usuarios malintencionados que pueden implementar algún software para tomar los datos que viajan a través de la red o que se encuentran almacenados en un dispositivo sin ningún tipo de cifrado. Por tal razón, en esta investigación se propone una red o sistema de almacenamiento distribuido con una topología punto a multipunto o malla, donde se pueden conectar diferentes dispositivos móviles, PCs, computadoras de placa reducida (SBC) y medios de almacenamiento tales como: cloud, raid, memorias, entre otros. En la red uno o más nodos origen recaban información de cámaras fotográficas, lectores de huella, sensores, sonidos, o cualquier tipo de archivo digital; posteriormente se utiliza el procedimiento desarrollado que aplicar las técnicas de difusión y confusión con la finalidad de cifrar y posteriormente segmentar de forma caótica la información, generando fragmentos del criptograma muy diferentes en tamaño y contenido, los cuales se pueden transmitir y almacenar tanto de forma distribuida como redundante en otros nodos conectados en la red. El sistema proporciona seguridad e integridad a la información, ya que permite que con las claves utilizadas para cifrar se puedan recuperar el 100%.

Palabras clave: caos; cifrado; almacenamiento distribuido

1. INTRODUCCIÓN

Las redes sociales y el Internet son parte de la vida cotidiana de las personas, frecuentemente los usuarios cuentan con varios dispositivos que les permiten trasladarse de un lugar a otro para trabajar ya sea en la oficina, su hogar, incluso en algún espacio de esparcimiento; lo que genera la necesidad de acceder a sus archivos actualizados desde cualquier equipo, por tal razón, hoy en día es muy frecuente utilizar medios de almacenamiento en la nube tales, como: Google Drive, Dropbox, icloud, etc. o incluso pueden tener su nube personalizada. En ocasiones los usuarios almacenan información muy importante, que desean proteger de algún ataque, por lo tanto, es necesario establecer medidas de seguridad o cifrar la información para evitar que un atacante pueda acceder a ella y hacer mal uso. La mejor solución a esta problemática es utilizar un sistema para cifrar con la finalidad de resguardar la información y que sólo las personas que cuenten con las claves de cifrado puedan tener acceso a ella, además la seguridad se incrementa si se almacena la información de forma segura en varios dispositivos. Los sistemas caóticos han sido ampliamente utilizados en criptografía debido a sus propiedades como alta sensibilidad a las condiciones iniciales y parámetros del sistema, además permiten cifrar aplicando sus órbitas en técnicas de difusión para mezclar la información y confusión para cambiar cada elemento de la información por otro, estos métodos son necesarios para proporcionar una criptografía más robusta (Pareek et al., 2005). Jiri Fridrich implementó un esquema de cifrado simétrico con una clave de longitud y tamaño de bloque variable implementado mapas caóticos bidimensionales (Fridrich, 1998). Ramzi Guesmi and M. A. Ben Farah, propusieron un algoritmo para cifrar imágenes médicas donde utilizan Secure Hash Algorithm (SHA-2) y mapas caóticos como: el Logístico, Tent y Sine (Guesmi et al., 2021). También los mapas de Sine y Logístico fueron implementados para aplicar esteganografía y cifrado con la finalidad de codificar los rostros identificados en una fotografía (Flores Siordia et al., 2018). Ping et al. desarrollaron un sistema para cifrar imágenes que se basa en autómatas celulares y caos, donde aplican difusión y confusión con un mapa caótico bidimensional 2D LASM (Ping et al., 2018). Mohamed Amin et al. propusieron un esquema de cifrado caótico por bloques, que usa el mapa de Tent tomando como entrada una imagen de 256 bits y genera una de 256 con claves de sesión de 256 bits (Amin et al., 2010). También los sistemas continuos como el oscilador de Rössler se usó para desarrollar un sistema de comunicación punto a punto para transmitir información cifrada mediante sincronización caótica (Rodríguez et al., 2016). Desarrollaron un algoritmo para cifrar imágenes con Rössler hiper-caótico que trabaja con matrices comprimidas (Huang et al., 2021). Para subsanar la problemática de seguridad al guardar en un solo dispositivo remoto la información y que esta pueda ser vulnerada por un usuario mal intencionado, se propone un sistema que implementa el modelo matemático caótico del oscilador de Rössler y mapa logístico para cifrar y segmentar la información de forma variable, posteriormente se puede almacenar cada segmento en su respectivo dispositivo remoto, con la finalidad de que sólo puedan acceder los usuarios autorizados para re-ensamblar y decodificar la información; también es posible almacenar la información de forma redundante. Está técnica es una alternativa de seguridad para evitar que las empresas que proporcionan almacenamiento en la nube, puedan acceder a la información de sus usuarios, ya que los segmentos codificados se almacenan en diferentes nubes o medios de almacenamiento.

El resto de este artículo se organiza de la siguiente forma: en la Sección 2, se explican los modelos matemáticos caóticos que se implementaron, también se expone el sistema de almacenamiento distribuido, además de los métodos usados para cifrar y

segmentar la información. En la Sección 3, se presentan los histogramas y diagramas de correlación para exhibir la robustez del sistema; por último, en la Sección 4, se presentan las conclusiones donde se remarcan los puntos importantes y logros de esta investigación.

2. METODOLOGÍA

2.1 OSCILADOR DE RÔSSLER

Es un sistema de ecuaciones continuas muy sencillo, el cual se utilizó en reacciones químicas, fue desarrollado por Otto Rössler; y está dado por las siguientes 3 ecuaciones:

$$\begin{aligned} \frac{dx}{dt} &= -(y+z), \\ \frac{dy}{dt} &= (x+ay), \\ \frac{dz}{dtt} &= c+z(x+b) \end{aligned} \tag{1}$$

Donde $\mathrm{x}, \mathrm{y}, \mathrm{z}$ son las variables de estado, mientras que a, b y c son los parámetros (Ibrahim et al., 2018), (Aguilar Santiago et al., 2020).

2.2 MAPA LOGÍSTICO

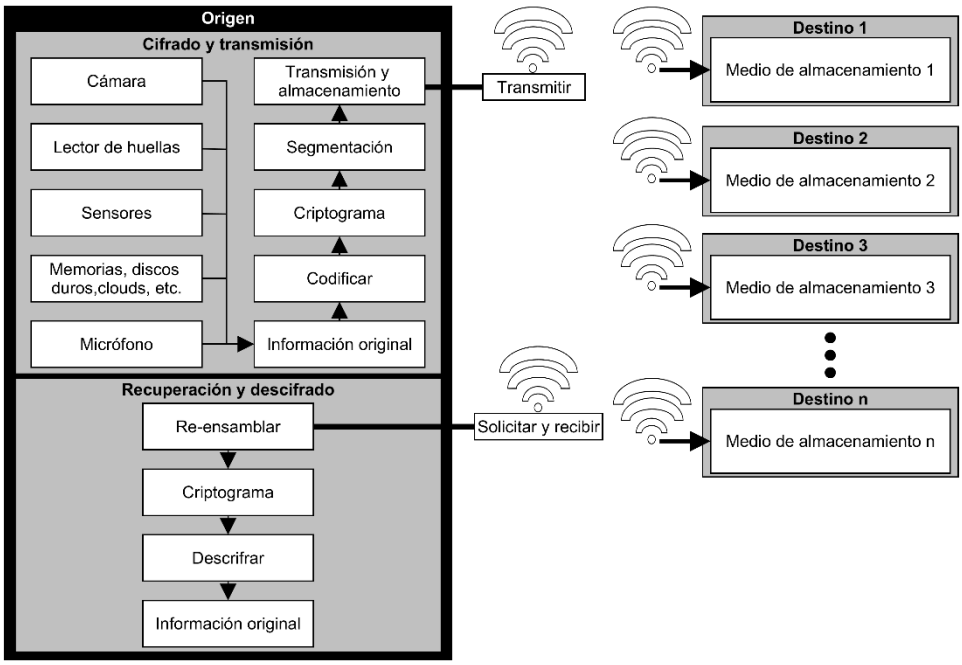
Es un modelo matemático caótico sencillo muy utilizado, cuenta con una variable x que puede tomar valores $0 \leq x \leq 1$ y un parámetro de control $1 < b < 4$, este modelo tiene la capacidad de producir un comportamiento caótico cuando $3.569 < b < 4$ (Alawida et al., 2022), su definición matemática es:

$$x_{n+1} = b \cdot x_n \cdot (1 - x_n) \tag{2}$$

2.3 SISTEMA DE ALMACENAMIENTO DISTRIBUIDO

La Figura 1, muestra el diagrama de la red de almacenamiento distribuido, donde el dispositivo origen puede recopilar información de sensores, lectores de huellas, cámaras, micrófonos o medios de almacenamiento, posteriormente cifra la información mediante un criptosistema caótico propuesto por (Aguilar Santiago et al., 2020), el cual se adaptó y además se le adicionó una nueva técnica que permite dar un segundo cifrado y segmentación a los archivos; el método propuesto considera el número de dispositivos remotos donde se desea almacenar la información cifrada para generar segmentos de criptogramas variables, los cuales transmite a su respectivo dispositivo destino. Para recuperar la información a su estado original se debe contar con las llaves de cifrado, los cripto-segmentos y el sistema para decodificar.

Figura 1. Proceso completo para cifrar, segmentar y almacenar de forma distribuida.



El sistema de cifrado y segmentación se puede emplear en dispositivos como: SBCs, PCs, tablets, celulares, en medios de almacenamiento en la nube como: Google Drive, Dropbox, icloud, personalizadas, etc.

2.3.1 Algoritmo para cifrar y segmentar

El algoritmo propuesto por (Aguilar Santiago et al., 2020*),* utiliza el modelo matemático del oscilador de Rössler (Ecuación 1), para generar órbitas caóticas que implementa al aplicar 2 técnicas de difusión que se encargan de mezclar de forma caótica la información de los rostros y 2 métodos de confusión que reemplazan cada uno de los elementos que conforman el rostro por otro; al implementar las técnicas de difusión y confusión se incrementa considerablemente la seguridad del cifrado. En este proyecto el criptosistema (Aguilar Santiago et al., 2020*)*, se adaptó para utilizar solamente una de sus técnicas de difusión y otra de confusión, con la finalidad de que el cifrado fuera más rápido y necesitará menos recursos de hardware. Además, incrementa la seguridad al agregar otra técnica donde aplica el mapa caótico Logístico (Ecuación 2), el cual es un modelo matemático muy simple que ocupa pocos recursos, este se utiliza para generar una órbita que permite crear cripto-segmentos de acuerdo al número de dispositivos remotos donde se desea almacenar la información de forma segura. Enseguida se explica de forma detallada el procedimiento, en este caso se utilizan 3 dispositivos destino, pero puede ser cualquier número de equipos remotos:

Segmentar
Llaves:
$x_0\ y\ b$: condición inicial y parámetros del modelo matemático Logístico (Ecuación 2).
$Iter$: número de veces que se resuelve la (Ecuación 2), antes de comenzar a cifrar; esto aumenta la seguridad, ya que es difícil para un atacante averiguar desde que iteración se inició el cifrado.

Paso 1. Resolver la (Ecuación 2) con las llaves x_0 y b, el número de veces indicado en la llave $Iter$, lo cual genera una órbita caótica.

$$log = [0.21348, 0.43452, 0.56741, \dots, 0.90349]$$

Paso 2. Reemplazar el valor de la llave x_0, por el último dato generado en log.

$$x_0 = 0.90349$$

Paso 3. Asignar a la variable $Ndisp$ el número de dispositivos remotos donde se almacenarán los cripto-segmentos (en este caso 3).

$$Ndisp = 3$$

Paso 4. Crear un archivo $archDestino\#$ para cada dispositivo destino; el cual almacenará los cripto-segmentos (donde # representa el número del dispositivo al que pertenece el archivo).

$$archDestino1, archDestino2 \; y \; archDestino3$$

Paso 5. Extraer los bytes del $criptograma$ (Figura 3(b)) y almacenarlos en un vector llamado $bytesArchivo$.

$$bytesArchivo = extraer\ bytes\ (criptograma)$$

El criptograma es la imagen de Mandril (Figura 3a) cifrada previamente con el sistema propuesto por (Aguilar Santiago et al., 2020).

Paso 6. Resolver la Ecuación 2 $Ndisp$ veces, utilizando como valor inicial $x_0 = 0.90349$ y b, cada valor caótico de x_n se asigna a un dispositivo.

$$Disp1 = 0.34564$$
$$Disp2 = 0.76854$$
$$Disp3 = 0.67237$$

Paso 7. Verificar

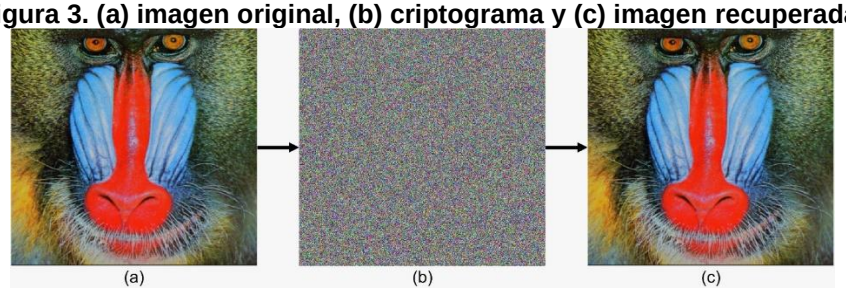
Si: $Disp1 > Disp2$ && $Disp1 > Disp3$
Entonces: almacenar un valor de $bytesArchivo$ en $archDestino1$
Si: $Disp2 > Disp1$ && $Disp2 > Disp3$
Entonces: almacenar un valor de $bytesArchivo$ en $archDestino2$
Si: $Disp3 > Disp1$ && $Disp3 > Disp2$
Entonces: almacenar un valor de $bytesArchivo$ en $archDestino3$

Paso 8. Repetir los pasos 6 y 7 asignando a x_0 el último valor de la órbita asignado a $Disp3$, hasta acomodar todos los valores de $bytesArchivo$ en los diferentes $archDestino\#$.
Paso 9. Enviar cada $archDestino\#$ a su respectivo destino, donde se guarda hasta que sea solicitado nuevamente para recuperar la información original.

3. RESULTADOS

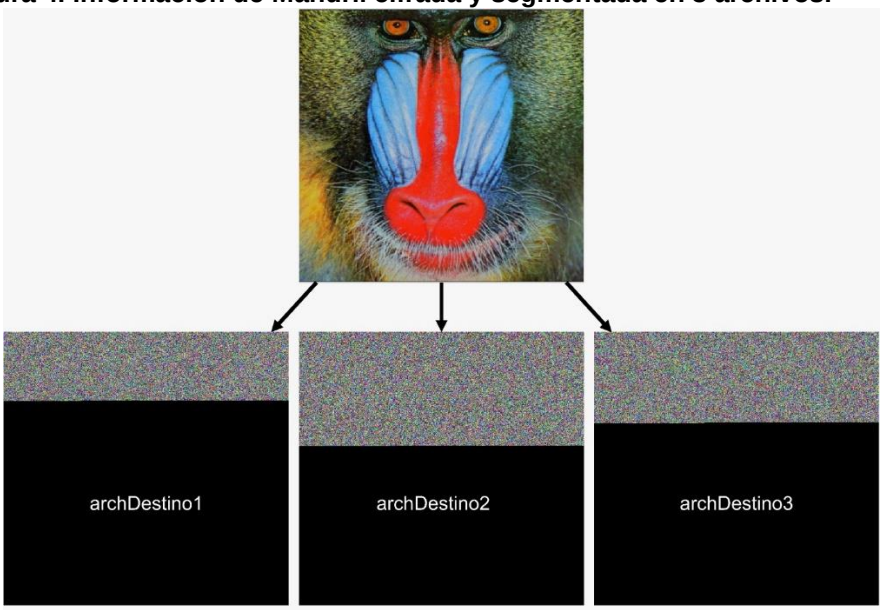
Se aplicó difusión, confusión con el algoritmo de cifrado y segmentación de la Sección 2.3.1 para codificar la imagen original de Mandril que se muestra en la Figura 3(a); el criptograma resultante se presenta en la Figura 3(b) y por último la Figura 3(c) presenta la imagen re-ensamblada y descifrada, con la cual se puede observar que no existe pérdida de información a simple vista.

Figura 3. (a) imagen original, (b) criptograma y (c) imagen recuperada.



La Figura 4, muestra los 3 cripto-segmentos generados a partir de la imagen de Mandril Figura 3(a); se puede observar que son de tamaño variable y presentan desorden o entropía. Lo anterior se debe a que se almacenan los valores de forma caótica en los diferentes archivos.

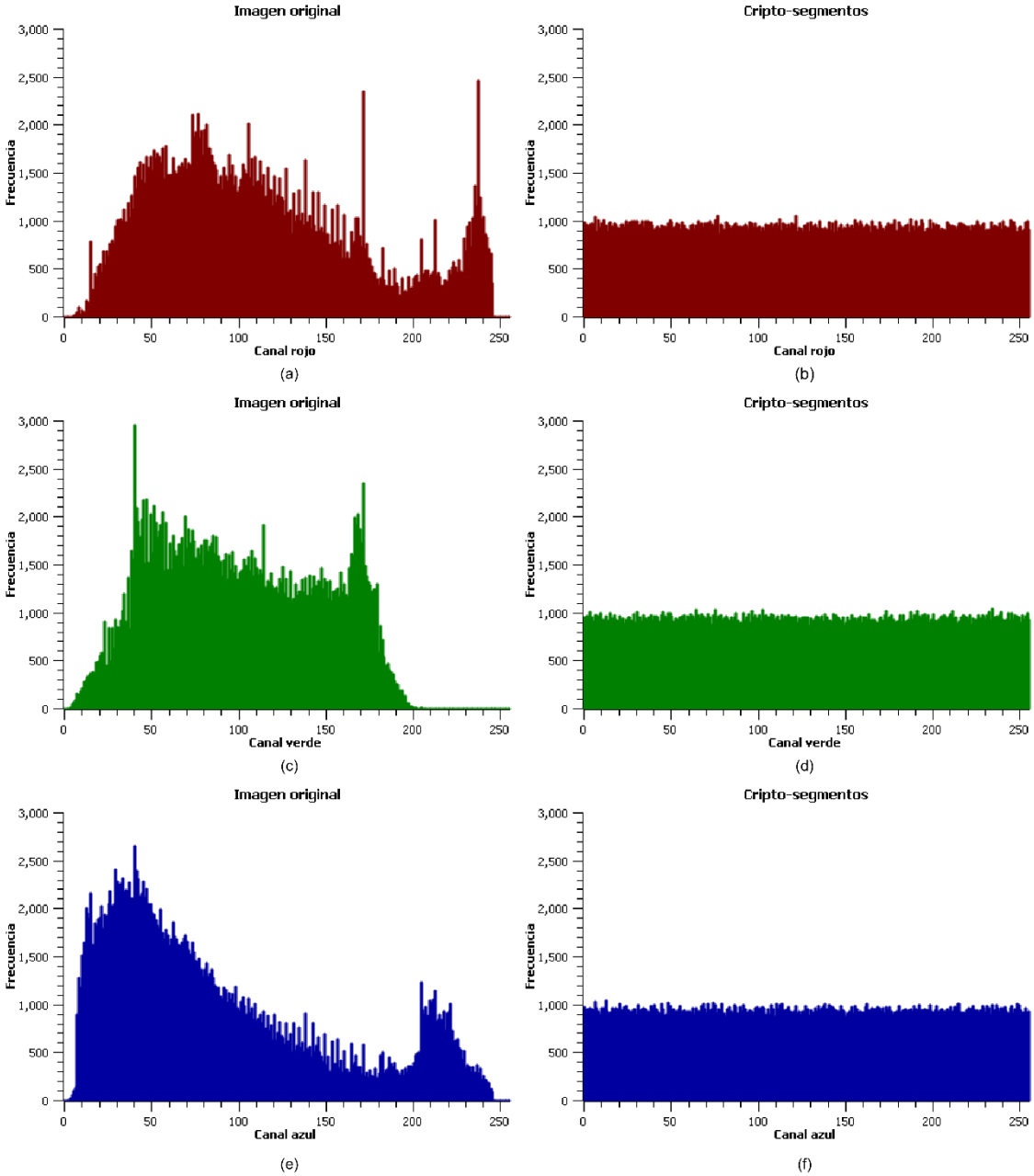
Figura 4. Información de Mandril cifrada y segmentada en 3 archivos.



3.1 HISTOGRAMAS

Permiten representar la distribución de los subpixeles que conforman las imágenes, el eje horizontal tiene un rango entre 0 y 255 de acuerdo a la intensidad de los colores y en el eje vertical se exhiben las frecuencias o número de veces que se repite el subpixel. Las Figuras 5 a, c y e, muestran los histogramas de la Figura 3 (a), correspondiente a la imagen original del Mandril; mientras que las Figuras 5 b, d y f exhiben los histogramas de los criptogramas, aplicando las técnicas de difusión y confusión de (Aguilar Santiago et al., 2020*),* y posteriormente se generan los cripto-segmentos. Al comparar los histogramas de acuerdo al color, se puede observar que son muy diferentes; los cripto-segmentos presentan los números de frecuencias muy homogéneos, por lo tanto, es más difícil para un atacante tratar de encontrar la relación que existe entre la imagen original y el criptograma segmentado.

Figura 5. (a)(c) y (e) histogramas rojo, verde y azul correspondientes a la imagen original. (b)(d) y (f) histogramas rojo, verde y azul del criptograma segmentado.



3.2 DIAGRAMAS DE CORRELACIÓN

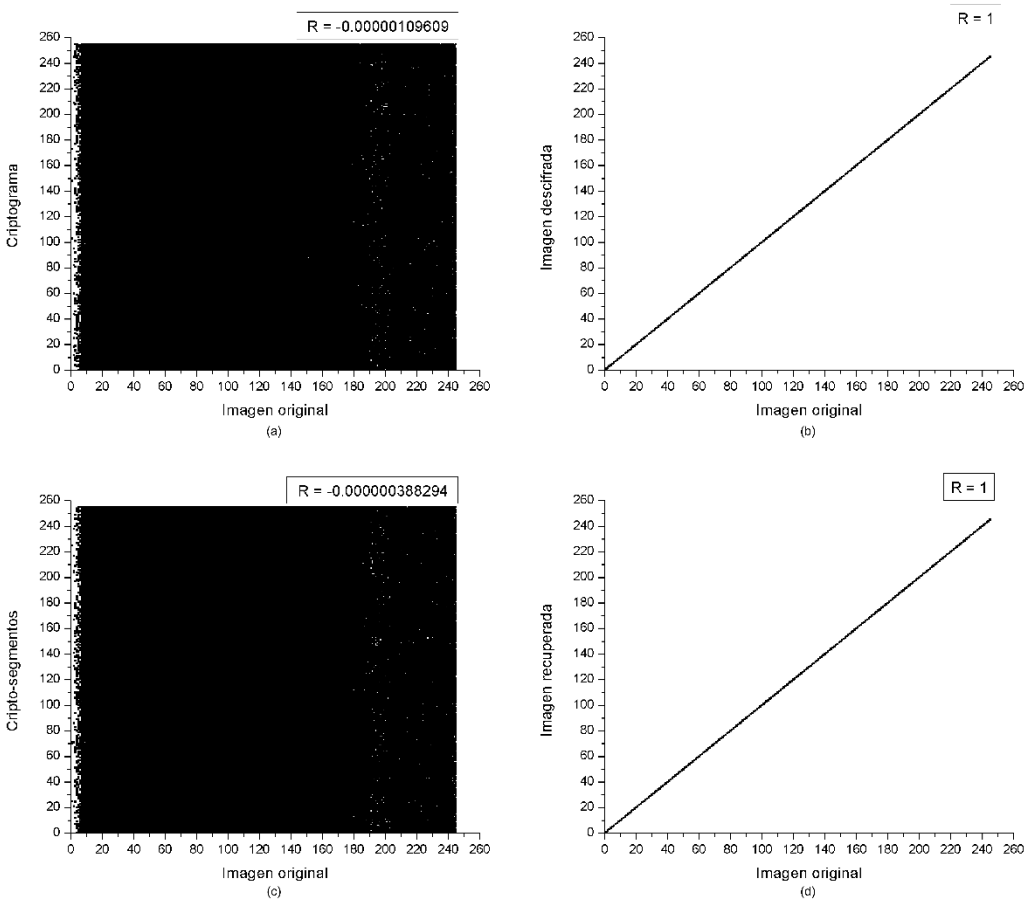
Se utilizan para demostrar la relación lineal que existen entre 2 variables, y arroja como resultado el coeficiente de correlación R, el cual puede tener un valor en el rango de -1 y 1; entre más se acerca a 0 indica mayor diferencia entre las variables; en cambio cuando es muy cercano a 1 son muy parecidas o si es igual a 1 demuestra que ambas variables son idénticas.

La Figura 6(a) exhibe el diagrama de correlación de la imagen original vs el criptograma generado al aplicar las técnicas de difusión y confusión (Aguilar Santiago et al., 2020); presenta un coeficiente de correlación $R = 0.00000109609$ cercano a 0, indicando que son muy diferentes y comprueba que el sistema proporciona robustez al momento de codificar. La Figura 6(b) presenta el diagrama de correlación entre la

imagen original y la descifrada con un coeficiente $R = 1$, demostrando que se puede recuperar el 100% de la información y que se mantiene la integridad de los datos.

La Figura 6(c) representa la asociación lineal entre la imagen original y los cripto-segmentos generados a partir de la imagen original de Mandril preparada para los diferentes dispositivos; el coeficiente $R = 0.000000388294$, se demuestra que también es muy cercano a 0, por lo tanto, son muy diferentes proporcionando mayor seguridad. La Figura 6(d) representa el cálculo del coeficiente de correlación entre la imagen original y la recuperada tras unir y descifrar los segmentos dando $R = 1$, concluyendo que no hay pérdida de información al volver a unir los segmentos.

Figura 6. Diagramas de correlación: (a) imagen original vs criptograma, (b) imagen original vs descifrada, (c) cripto-segmentos vs imagen original y (d) imagen recuperada vs original.



4. CONCLUSIONES

El sistema de cifrado y almacenamiento distribuido, se puede utilizar para codificar y segmentar cualquier tipo de información; genera cripto-segmentos muy diferentes con cualquier pequeño cambio que se realice a las llaves; permite decodificar los datos garantizando su integridad al descifrar sin pérdidas. Como los sistemas remotos tienen solo un cripto-segmento, no pueden descifrar la información sin los demás. El algoritmo de cifrado previo considera la propia información para codificar, por lo tanto, si se realiza un pequeño cambio en el archivo original, se genera un criptograma muy distinto. La técnica de segmentación propuesta también codifica al mezclar la información, pero se recomienda aplicar el cifrado previo para garantizar mayor seguridad. También se puede

realizar almacenamiento distribuido y redundante en dos o más dispositivos, lo cual lo hace más tolerante a fallas, porque si algún equipo tiene un error, se puede recuperar la información del dispositivo de respaldo.

5. REFERENCIAS

Aguilar Santiago, J., Flores Siordia, O., Guillen Bonilla, J. T., Estrada Gutiérrez, J. C., González Novoa, M. G., & Jiménez Rodríguez, M. (2020). Chaotic Cryptosystem for Selective Encryption of Faces in Photographs. *Security and Communication Networks*, *2020*, 1–22. doi: 10.1155/2020/8848356

Alawida, M., Teh, J. S., Mehmood, A., Shoufan, A., & Alshoura, W. H. (2022). A chaos-based block cipher based on an enhanced logistic map and simultaneous confusion-diffusion operations. *Journal of King Saud University - Computer and Information Sciences*, *34*(10), 8136–8151. doi: 10.1016/j.jksuci.2022.07.025

Amin, M., Faragallah, O. S., & Abd El-Latif, A. A. (2010). A chaotic block cipher algorithm for image cryptosystems. *Communications in Nonlinear Science and Numerical Simulation*, *15*(11), 3484–3497. doi: 10.1016/j.cnsns.2009.12.025

Flores Siordia, O., Gutiérrez, J. C. E., Leyferman, C. E. P., Santiago, J. A., & Rodríguez, M. J. (2018). System to Safeguard the Identity of Persons in Photographs through Cryptography and Steganography Techniques Using Chaos. *Security and Communication Networks*, *2018*, 1–16. doi: 10.1155/2018/4853134

Fridrich, J. (1998). Symmetric Ciphers Based on Two-Dimensional Chaotic Maps. *International Journal of Bifurcation and Chaos*, *08*(06), 1259–1284. doi: 10.1142/S021812749800098X

Guesmi, R., & Farah, M. A. B. (2021). A new efficient medical image cipher based on hybrid chaotic map and DNA code. *Multimedia Tools and Applications*, *80*(2), 1925–1944. doi: 10.1007/s11042-020-09672-1

Huang, W., Jiang, D., An, Y., Liu, L., & Wang, X. (2021). A Novel Double-Image Encryption Algorithm Based on Rossler Hyperchaotic System and Compressive Sensing. *IEEE Access*, *9*, 41704–41716. doi: 10.1109/ACCESS.2021.3065453

Ibrahim, K. M., Jamal, R. K., & Ali, F. H. (2018). Chaotic behaviour of the Rossler model and its analysis by using bifurcations of limit cycles and chaotic attractors. *Journal of Physics: Conference Series*, *1003*, 012099. doi: 10.1088/1742-6596/1003/1/012099

Pareek, N. K., Patidar, V., & Sud, K. K. (2005). Cryptography using multiple one-dimensional chaotic maps. *Communications in Nonlinear Science and Numerical Simulation*, *10*(7), 715–723. doi: 10.1016/j.cnsns.2004.03.006

Ping, P., Wu, J., Mao, Y., Xu, F., & Fan, J. (2018). Design of image cipher using life-like cellular automata and chaotic map. *Signal Processing*, *150*, 233–247. doi: 10.1016/j.sigpro.2018.04.018

Rodríguez, M. J., González-Novoa, M. G., Estrada-Gutiérrez, J. C., Acosta-Lúa, C., & Flores-Siordia, O. (2016). Secure point-to-point communication using chaos. *DYNA*, *83*(197), 180. doi: 10.15446/dyna.v83n197.53506

Comunicación alineada con los objetivos de Desarrollo Sostenible