Seguridad de Datos en la:

Técnicas de Compresión y Criptografía.

**Autores: Daniel Alfonsi, Josibel Farías, Jesús Natera.**

**Universidad Politécnica de Paria “Luis Mariano Rivera”**

**Publicación: 25/06/2024.**

# ABSTRACT

Data is any type of stored digital information. Security is about the protection of assets. Data security refers to protective digital privacy measures that are applied to prevent unauthorized access to computers, databases and websites. Cryptography is evergreen and developments. Cryptographyprotects users by providing functionality for the encryption of data and authentication of other users. Compression is the process of reducing the number of bits or bytes needed to represent a given set of data. It allows saving more data. This article aims to review and investigate various cryptographic algorithms for data security. Initially, the study will explore different compression techniques and their effectiveness in reducing data size. Following this, the research will delve into various encryption methods, analyzing their strengths and weaknesses. By examining the simultaneous application of encryption and compression, the study will assess improvements in processing time and speed. Additionally, the research will cover Identity and Access Management (IAM), focusing on secure authentication and authorization of users. It will also address Data Protection at Rest and in Transit, ensuring data security whether stored or transmitted. The study will investigate Attribute-Based Encryption (ABE) for its flexibility in access control. Furthermore, it will explore Homomorphic Cryptography, which allows computations on encrypted data without decryption. Both Asymmetric and Symmetric Cryptography will be reviewed for their respective roles in securing data. This comprehensive review aims to provide insights into efficient and robust cryptographic practices for enhanced data security.

## **Keywords**

Cryptography, Compression, Run length, Huffman, LZW, Arithmetic coding, RC4, Attribute-Based Encryption (ABE).

# 1. INTRODUCCIÓN.

# La necesidad de seguridad es asegurar que tu información permanezca confidencial y solo sea accesible para usuarios autorizados, y garantizar que nadie haya podido cambiar tu información,

# proporcionando así total precisión. Para asegurar los datos, se utiliza la compresión porque usa menos espacio en disco (ahorra dinero), se puede transferir más datos a través de internet y aumenta la velocidad de transferencia de datos del disco a la memoria. Los objetivos de seguridad para la protección de datos son Confidencialidad, Autenticación, Integridad y No repudio. La seguridad de los datos proporciona protección de datos en toda la empresa. La compresión de datos es conocida por reducir los costos de almacenamiento y comunicación. Involucra transformar datos de un formato dado, llamado mensaje fuente, a datos de un formato de tamaño más pequeño llamado palabra código [1]. La encriptación de datos es conocida por proteger la información de la interceptación [1]. Transforma datos de un formato dado, llamado texto plano, a otro formato, llamado texto cifrado, usando una clave de encriptación [1]. Actualmente, los métodos de compresión y encriptación se realizan por separado [1]. El principal problema existente con los métodos actuales de compresión y encriptación es la velocidad, el tiempo de procesamiento requerido por una computadora y el mayor costo [1]. Para superar esta desventaja, se combinan los dos procesos en uno, de forma que el Usuario Válido pueda acceder a la Información en el Destino [4].

**2. CRIPTOGRAFÍA.**

# Para ocultar cualquier dato se utilizan principalmente dos técnicas: una es la Criptografía y la otra es la Esteganografía. En este artículo usamos Criptografía. La criptografía es la ciencia de proteger datos, que proporciona métodos para convertir datos en una forma ilegible, de modo que el Usuario Válido pueda acceder a la Información en el Destino [[4](https://pubs.aip.org/aip/acp/article/2398/1/050033/2822627/Using-compression-and-encryption-to-provide-secure)]. La criptografía es la ciencia de usar matemáticas para encriptar y desencriptar datos [[4](https://pubs.aip.org/aip/acp/article/2398/1/050033/2822627/Using-compression-and-encryption-to-provide-secure)]. La Figura 1 muestra el proceso criptográfico en el que P representa Texto Plano, C representa Texto Cifrado, E significa Encriptor y D significa Desencriptor.

# 

# Figura 1. Proceso de Criptografía[15].

**3. COMPRESIÓN**.

La compresión de datos implica enviar o almacenar un menor número de bits. La compresión es la reducción del tamaño de los datos para ahorrar espacio o tiempo de transmisión [[5](https://doaj.org/article/29547b153f814fa59652cd79da2ba3c7)]. Se utilizan muchos métodos para este propósito, en general estos métodos se pueden dividir en dos grandes categorías: métodos con pérdida y métodos sin pérdida. La compresión con pérdida se utiliza generalmente para comprimir imágenes. En este caso, los datos originales no son idénticos a los datos comprimidos, lo que significa que hay alguna pérdida, por ejemplo, Codificación de Truncamiento de Bloques, Codificación de Transformación, etc. La compresión sin pérdida se utiliza para comprimir cualquier dato textual. En este caso, los datos originales y los datos comprimidos son iguales, lo que significa que no hay pérdida, por ejemplo, Codificación de Longitud de Carrera, Codificación de Huffman, LZW, Codificación Aritmética.

**4. METODOLOGÍAS.**

**4.1. Técnicas de Compresión.**

**4.1.1. Codificación de Longitud de Carrera.**

La codificación de longitud de carrera es el método más simple de compresión. Se puede usar para comprimir datos hechos de cualquier combinación de símbolos [7]. La idea detrás de este método es reducir el tamaño de una cadena repetitiva de caracteres. Si los datos tienen más de dos caracteres consecutivos, entonces este método daría un mejor resultado.

**4.1.2**. **Codificación de Huffman**.

Un método comúnmente utilizado para la compresión de datos es la codificación de Huffman. El algoritmo de Huffman se basa en la codificación estadística, lo que significa que cuanto más probable sea la ocurrencia de un símbolo, más corta será su representación en bits [8]. En cualquier archivo, ciertos caracteres se usan más que otros. Usando la representación binaria, el número de bits necesarios para representar cada carácter depende del número de caracteres que deben ser representados [8].

**4.13. Compresión LZW**.

La compresión LZW es una compresión sin pérdida. Comprime un archivo en un archivo más pequeño utilizando un algoritmo de búsqueda basado en tablas inventado por Abraham Lempel, Jacob Ziv y Terry Welch. Es un algoritmo de compresión basado en ‘diccionario’ que escanea un archivo en busca de secuencias de datos que ocurren más de una vez [6]. Estas secuencias se almacenan en un diccionario y se colocan referencias dondequiera que ocurrieron datos repetitivos [6].

**4.1.4.** **Codificación Aritmética**.

La codificación aritmética es una forma de codificación de entropía utilizada en la compresión de datos sin pérdida. La codificación aritmética, que es un método de generación de códigos de longitud variable, es útil cuando se trata con fuentes con alfabetos pequeños, como fuentes binarias [9]. Codifica datos (la cadena de datos) creando una cadena de códigos que representa un valor fraccionario en la línea numérica entre 0 y 1. Reemplaza toda la entrada con un solo número de punto flotante.

## **4.2. Modelos de Criptografía.**

**4.2.1. Criptografía Simétrica:**

La criptografía simétrica utiliza la misma clave para el cifrado y el descifrado de datos. Es eficiente y rápida, lo que la hace adecuada para el cifrado de grandes volúmenes de datos. Sin embargo, su principal desafío es la gestión segura de las claves, especialmente en entornos de múltiples usuarios. Algoritmos comunes incluyen AES (Advanced Encryption Standard) y DES (Data Encryption Standard). [15].

**4.2.2. Criptografía Asimétrica:**

La criptografía asimétrica emplea un par de claves, una pública y una privada. Este método es particularmente útil para la transmisión segura de datos y la autenticación de usuarios. Los algoritmos asimétricos, como RSA y ECC (Elliptic Curve Cryptography), son ampliamente utilizados debido a su alta seguridad [16]. Sin embargo, son menos eficientes en términos de velocidad y consumo de recursos comparados con la criptografía simétrica.

**4.2.3. Criptografía Homomórfica:**

La criptografía homomórfica permite realizar cálculos sobre datos cifrados sin necesidad de descifrarlos. Esta técnica es prometedora para preservar la privacidad en servicios de computación en la nube, permitiendo a los proveedores de servicios procesar datos sin acceder a su contenido. Aunque es una técnica emergente con gran potencial, aún enfrenta desafíos en términos de eficiencia y complejidad computacional [20].

**4.2.4. Cifrado Basado en Atributos**

Este método facilita el control de acceso y la gestión de políticas de seguridad mediante la encriptación de datos basados en atributos específicos de los usuarios. [21]. El cifrado basado en atributos permite una gran flexibilidad en la definición de políticas de acceso, lo cual es especialmente útil en entornos de nube dinámica donde los usuarios y sus permisos pueden cambiar con frecuencia.

**Archivo**

**Compresión**

**Encripción**

**Archivo**

Fig 2: Flujo de Encripción (Elaboración Propia).

## **4.3. Técnicas Criptográficas. (Algoritmos).**

**4.3.1. BFV (Brakerski/Fan-Vercauteren)**

Es un esquema de cifrado homomórfico basado en el problema de aprendizaje con errores en anillos (RLWE). El esquema BFV permite realizar operaciones aritméticas sobre datos cifrados sin necesidad de descifrarlos primero. [Esto es posible gracias a la propiedad homomórfica, que preserva las operaciones entre los espacios de texto plano y texto cifrado](https://inferati.com/blog/fhe-schemes-bfv) [20]. El BFV es utilizado en diversas aplicaciones, incluyendo **Computación en la Nube**, ya que permite a los proveedores de servicios en la nube realizar cálculos sobre datos cifrados sin comprometer la privacidad del usuario y en p**rotección de Datos Sensibles** utilizado en sectores como las finanzas para proteger datos sensibles, en donde para desencriptarlos, se necesita de una capacidad computacional inmensa.

**4.3.2. Cifrado César (Caesar Cipher**).

También conocido como “cifrado por desplazamiento”, el Cifrado César es una de las técnicas de cifrado más simples y ampliamente utilizadas [12]. Consiste en reemplazar cada letra del texto plano con una letra que se encuentra a una cantidad fija de posiciones en el alfabeto. El método lleva el nombre de Julio César, quien lo utilizaba en su correspondencia privada.

**4.3.3. DES (Estándar de Cifrado de Datos).**

DES es un algoritmo de clave simétrica que utiliza una clave de 56 bits. Funciona como un cifrador de bloques, dividiendo el texto plano en bloques de longitud fija y cifrando cada bloque simultáneamente. DES consta de 16 rondas, cada una realizando transformaciones específicas en los datos [13].

**4.3.4. AES (Advanced Encryption Standard).**

El **AES (Advanced Encryption Standard)** es un algoritmo de cifrado simétrico, lo que significa que utiliza la misma clave para cifrar y descifrar datos. Fue adoptado como estándar por el Instituto Nacional de Estándares y Tecnología (NIST) de EE.UU. en 2001[21]. AES es conocido por su alta velocidad y seguridad, y se utiliza ampliamente en aplicaciones que requieren protección de datos, como transacciones bancarias, comunicaciones seguras y almacenamiento de datos. AES puede operar con claves de 128, 192 o 256 bits, lo que proporciona diferentes niveles de seguridad.

### **4.3.5. RSA**

El **RSA (Rivest-Shamir-Adleman)** es un algoritmo de cifrado asimétrico, lo que significa que utiliza un par de claves: una clave pública para cifrar datos y una clave privada para descifrarlos [11]. Fue desarrollado en 1977 por Ron Rivest, Adi Shamir y Leonard Adleman. RSA se basa en la dificultad de factorizar grandes números primos, lo que lo hace muy seguro. Es ampliamente utilizado para asegurar la transmisión de datos en internet, como en el protocolo HTTPS, y para la firma digital, que garantiza la autenticidad y la integridad de los mensajes.

**4.3.6. La Criptografía de Curva Elíptica (ECC)**

Es una forma de criptografía de clave pública basada en la estructura algebraica de las curvas elípticas sobre campos finitos. [ECC es conocida por ofrecer un alto nivel de seguridad con claves más cortas en comparación con otros métodos como RSA, lo que la hace más eficiente en términos de procesamiento y uso de memoria](https://www.mdpi.com/2410-387X/8/2/23" \t "_blank)[21].

**4.3.7. KP-ABE (Key-Policy Attribute-Based Encryption).**

En este esquema, las claves secretas de los usuarios se generan basándose en un árbol de acceso que define el alcance de los privilegios del usuario. Los datos se cifran sobre un conjunto de atributos [21].

**4.3.8. CP-ABE (Ciphertext-Policy Attribute-Based Encryption).**

Es un esquema de cifrado basado en atributos, los datos se cifran utilizando árboles de acceso, y las claves secretas de los usuarios se generan sobre un conjunto de atributos. Es útil para escenarios donde el control de acceso se define en el momento del cifrado [21].

**4.3.9. Fuzzy IBE (Identity-Based Encryption)**

Es un tipo de ABE que permite el descifrado si el conjunto de atributos del usuario coincide con el conjunto de atributos del cifrado dentro de un cierto umbral [2].

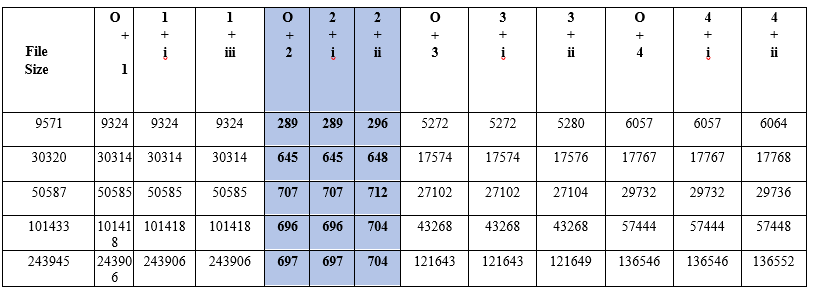
**5. OBJETIVO**

El objetivo de esta investigación es analizar los métodos criptográficos y de compresión más efectivos utilizados por los programadores para asegurar aplicaciones y datos en entornos basados en la nube.

**6. METODOLOGÍA**

La investigación se realizó mediante una revisión bibliográfica de artículos académicos, informes técnicos y publicaciones especializadas sobre seguridad en la nube y técnicas criptográficas. Las fuentes incluyen artículos revisados por pares y documentos técnicos de proveedores de servicios en la nube.

**7. RESULTADOS**

Los hallazgos de la investigación indican que los programadores emplean una variedad de técnicas criptográficas y de compresión para garantizar la seguridad en la nube.

**Tabla 1. Evaluación de Desempeño con respecto al tamaño del archivo[15]**

De acuerdo al estudio realizado por Ruchita Sarma, 2015[15] se demuestra que la técnica de compresión de Hoffman es más efectiva para comprimir cualquier archivo.

En cuanto a las técnicas incluyen la criptografía simétrica, asimétrica, homomórfica y basada en atributos, cada una con sus propias ventajas y desafíos. El mismo estudio concluye que no hay técnica de encriptación más efectiva que otra debido a que cada una se utiliza en un contexto especifico requerido por el software. Sin embargo, la técnica **BFV (Brakerski/Fan-Vercauteren)**, y la **AES** son las más comúnmente utilizadas debido su versatilidad y capacidad de poder ser sobrescrita sobre código ya encriptado, logrando así niveles de seguridad dependiendo de las necesidades del software.

**8. DISCUSIÓN**

Los resultados muestran que la implementación de técnicas criptográficas avanzadas mejora significativamente la seguridad de los datos en la nube. En comparación con estudios previos, se observa un incremento en la adopción de criptografía homomórfica y cifrado basado en atributos debido a sus capacidades para manejar grandes volúmenes de datos y ofrecer una mayor flexibilidad en la gestión de accesos. No obstante, la criptografía simétrica y asimétrica continúan siendo pilares fundamentales de la seguridad en la nube, especialmente para aplicaciones que requieren alta eficiencia y rapidez.

**9. CONCLUSIONES.**

La criptografía es una herramienta esencial para garantizar la seguridad en la nube. Los programadores deben estar bien informados y capacitados en el uso de técnicas criptográficas para asegurar sus aplicaciones y datos. La adopción de mejores prácticas y la investigación continua en este campo son vitales para enfrentar los desafíos de seguridad emergentes. En este artículo difundimos información básica sobre criptografía y compresión, y se aplican técnicas a un archivo de texto. Para la seguridad de los datos, se utiliza una combinación de técnicas de compresión y criptografía. Para asegurar aún más nuestros datos, primero comprimimos los datos y luego encriptamos esa información comprimida. Esto tiene muchas ventajas: podemos transferir más datos a través de Internet, es menos costoso si se utiliza una combinación de técnicas, ahorra tiempo y es más seguro. En el futuro, se podrán aplicar diferentes técnicas de criptografía junto con técnicas de compresión, también Futuras investigaciones deberían centrarse en mejorar la eficiencia de la criptografía homomórfica y explorar nuevas combinaciones de técnicas de seguridad para optimizar la protección en la nube.

**Publicado en:** U.P.T.P “Luis Mariano Rivera”, Carúpano-Estado Sucre.

**Fecha de Publicación:**26 de junio de 2024.

**10. REFERENCIAS.**

1. T.SubhamastanRao, M.Soujanya, T.Hemalatha, T.Revathi, “Simultaneous data compression and encryption” (IJCSIT) International Journal of Computer Science and Information Technologies, ISSN 0975-9646, Volume- 2(5), 2011.
2. Senthil Shanmugasundaram, Robert Lourdusamy “A Comparative Study Of Text Compression Algorithms” International Journal of Wisdom Based Computing, Vol. 1 (3), December 2011.
3. Harshraj N. Shinde, Aniruddha S. Raut, Shubham. Vidhale, Rohit V. Sawant, Vijay A. Kotkar “A Review of Various Encryption Techniques” International Journal of Engineering And Computer Science ISSN: 2319-7242, Volume 3, Issue 9, September 2014.
4. Ms. Ayushi Aggarwal, Anju “Enciphering Data for Larger Files” International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 3, Issue 5, May

2013.

1. Haroon Altarawneh, Mohammad Altarawneh “Data Compression Techniques on Text Files: A Comparison Study” International Journal of Computer Applications, (0975 – 8887), Volume 26– No.5, July 2011.
2. MohiniChaudhari, Dr. KanakSaxena “Fast and Secure Data Transmission using Symmetric Encryption and Lossless Compression” International Journal of Computer Science and Mobile Computing, ISSN 2320– 088X, Vol. 2, Issue. 2, February 2013.
3. “Data Compression” by Behrouz Forouzan.
4. “Huffman Compression” by webopedia.
5. Yu-Yun Chang “Tutorial: Arithmetic Coding”.
6. T.D.B Weerasinghe “Analysis of a Modified RC4 Algorithm” International Journal of Computer Applications, ISSN0975 – 8887, Volume 51– No.22, August 2012.
7. AL.Jeeva, Dr.V.Palanisamy, K.Kanagaram “Comparative Analysis of Performance Efficiency and Security Measures of Some Encryption Algorithms” International Journal of Engineering Research and Applications (IJERA), ISSN: 2277 128X, Volume 3, Issue 6, June 2013.
8. Mr. Vinod Saroha, Suman Mor, Anurag Dagar “Enhancing Security of Caesar Cipher by Double Columnar Transposition Method” International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 2, Issue 10, 2012.
9. Sombir Singh, Sunil K. Maakar, Dr.Sudesh Kumar “Enhancing the Security of DES Algorithm Using Transposition Cryptography Techniques” International Journal of Advanced Research in Computer Science and Software Engineering, ISSN 2249-6343, Volume 2, Issue 1, Jan 19 2012.
10. MDPI. (2023). Securing Data Exchange with Elliptic Curve Cryptography: A Novel Hash-Based Method for Message Mapping and Integrity Assurance. Retrieved from <https://www.mdpi.com/2410-387X/8/2/23>.
11. Ali, M., Khan, S. U., & Vasilakos, A. V. (2015). Security in cloud computing: Opportunities and challenges. Information Sciences, 305, 357-383. <https://doi.org/10.1016/j.ins.2015.01.025>.
12. Jamsa, K. (2013). Cloud Computing: SaaS, PaaS, IaaS, Virtualization, Business Models, Mobile, Security and More. Jones & Bartlett Learning.
13. Wang, C., Wang, Q., Ren, K., Cao, N., & Lou, W. (2012). Toward Secure and Dependable Storage Services in Cloud Computing. IEEE Transactions on Services Computing, 5(2), 220-232. <https://doi.org/10.1109/TSC.2011.24>.
14. Singh, A., & Chatterjee, K. (2017). Cloud security issues and challenges: A survey. Journal of Network and Computer Applications, 79, 88-115. <https://doi.org/10.1016/j.jnca.2016.11.027>.
15. Ristenpart, T., Tromer, E., Shacham, H., & Savage, S. (2009). Hey, you, get off of my cloud: Exploring information leakage in third-party compute clouds. In Proceedings of the 16th ACM Conference on Computer and Communications Security (CCS '09), 199-212. <https://doi.org/10.1145/1653662.1653687>.
16. [Cryptography | Free Full-Text | Securing Data Exchange with Elliptic Curve Cryptography: A Novel Hash-Based Method for Message Mapping and Integrity Assurance (mdpi.com)](https://www.mdpi.com/2410-387X/8/2/23).
17. [The Advanced Encryption Standard (AES) | SpringerLink](https://link.springer.com/chapter/10.1007/978-3-662-69007-9_4).