Vulneración de Seguridad de Archivos Zip y Rar con Clave de Protección

Franklin Mamani Suca
Postgrado en Informática
Universidad Mayor de San Andrés
La Paz - Bolivia
cmphonk45@gmail.com

***Resumen*—El presente artículo mediante los análisis y el desarrollo práctico, describe la seguridad y vulnerabilidad de archivos Zip y Rar con clave de protección, la creación de los mencionados archivos es efectuada a la hora de comprimir y almacenar información; esta acción como una medida de seguridad, principalmente en conservar la confidencialidad de información o de archivos. Esta práctica es de habitual por los usuarios mediante el uso de programas de software para comprimir y descomprimir archivos, además añadiendo el cifrado de tal manera que se pueda compartir de una forma segura. Por otra parte, en medio de redes de Internet muestran la disponibilidad de métodos y programas software que facilitan a descomprimir archivos con clave de acceso, ya que esta contingencia se convierte en una medida necesaria en casos donde el usuario pierde la contraseña para descomprimir. Y esta acción irregular conlleva riesgos de la seguridad de datos, por lo tanto, se plantea el estudio. ¿Pero, el método de resguardar la confidencialidad de archivos mediante la compresión y con clave de acceso a estas, son realmente seguras para el usuario?**

Palabras clave—clave, cifrado, archivo, zip, rar, algoritmo

I. INTRODUCCIÓN

En la actualidad, el uso diario de ordenadores como una herramienta fundamental en ámbitos de uso personal, trabajo, negocios empresariales y entre otras, es de necesidad el acceso a programas software para ejecutar las tareas de interés. Y una de las necesidades es de un software que brinde la compresión y descompresión de archivos, el objetivo principalmente es lograr que un archivo ocupe menos espacio en la memoria almacenada y además compartir esta de una forma segura, especialmente cuando se trata de información confidencial para el usuario.

Asimismo, es importante destacar que ahora los programas para comprimir y descomprimir archivos van más allá de sus funcionalidades básicas, permitiendo no solamente comprimir y descomprimir en distintos formatos, sino también proteger la confidencialidad y la integridad de datos, también compartir los archivos a través de sistemas y plataformas muy diversas.

El uso de software WinRAR con la función principal de comprimir y descomprimir de archivos además el cifrado de múltiples archivos; con el objetivo prever la seguridad en las transferencias, esto gracias a los algoritmos especiales que se ejecutan dentro del ordenador, cumpliendo el objetivo de comprimir y luego extraerse de la misma sin afectar la integridad de la información [1], es una herramienta útil y conocida.

El método es comúnmente empleado por los usuarios en gestionar de manera eficaz los archivos y resguardar la privacidad de información las mismas, los programas de compresión brindan la posibilidad de establecer la contraseña de seguridad necesaria a la hora de extraer la información, de modo que nadie pueda descomprimirlo y abrirlo sin la contraseña establecida. Sin embargo, el olvido de la clave conlleva a la búsqueda de otras acciones como una medida de solución. Así mismo donde los archivos fuesen apoderados por las identidades son autorizas y estas busquen las formas de descomprimir archivos confidenciales.

A. Planteamiento del problema.

El uso ordinario de archivos Zip y Rar por los usuarios y la confianza brindada a estas técnicas, tal es el caso de comprimir archivos usando contraseñas débiles de acceso con el objetivo de resguardar la confidencialidad de archivos o datos, son riesgos de seguridad de archivos.

Por añadidura, según Grossman[2]recientemente descubrieron un fallo de seguridad en WinRAR que ha estado presente desde casi 15 años, la vulnerabilidad se encuentra en una librería para Windows llamada unacev2.dll que se usa solo para analizar archivos ACE1, para ello WinRAR usa librería de terceros para la función indicada y la misma no se ha actualizado desde año 2005.

Dada la posibilidad donde las identidades no autorizados logren la disposición de archivos confidenciales y estas a base de programas en Internet y así mismo los métodos para descomprimir sin la contraseña logren introducirse a los archivos vulnerando la confidencialidad de la información, en efecto el riesgo materializado y las consecuencias negativas hacia el usuario.

B. Objetivo general

Evaluar el método de compresión y descompresión de archivos Zip y Rar con clave de seguridad de acceso a archivos de información, estas generadas mediante programa software

Para referenciar este artículo (IEEE):

[N] F. Mamani, «Vulneración de Seguridad de Archivos Zip y Rar con Clave de Protección» *Revista PGI. Investigación, Ciencia y Tecnología en Informática*, nº 7, pp. 38-42, 2020.

[1] ACE: Es un formato de fichero de archivo de compresión de datos privativo desarrollado por e-merge GmbH.

WinRAR; realizando las pruebas en entorno de usabilidad y determinando el grado de seguridad y la vulnerabilidad.

C. Objetivos específicos

- Diagnosticar la estructura y la gestión de procesos ejecutadas por los programas de compresión y descompresión de archivos respecto a la seguridad de datos.
- Caracterizar los métodos de descompresión y las programas software que existan para el cometido.
- Realizar las pruebas de compresión de archivos con clave de seguridad y descompresión empleando los métodos hallados de esta manera denotando la vulnerabilidad existente.

II. MARCO TEÓRICO

Antes de entrar de lleno a las implicaciones que conlleva en el presente trabajo de investigación, es de suma importancia remarcar algunos conceptos dentro del marco teórico.

A. WinRar

Es un software de compresión de datos desarrollado por Yevgueni Roshal, es de tipo archivador de ficheros, programado en lenguaje de programación C++, el software es disponible en plataformas Windows, Mac OS X, FreeBSD, Linux. Pose un formato propio de compresión, el RAR (Roshal Archive) así mismo es compatible con otros formatos, tales como: ZIP (Zipx), CAB,7z, ACE, ARJ y entre otras. Otra de las funciones especiales es brindar el cifrado de archivos hasta AES-128 desde la versión 3.20 en adelante [2], para este estudio se tratará de los archivos de tipo Zip y Rar.

B. Archivador de ficheros

Un archivador de ficheros es un programa informático que combina el número de ficheros en un único fichero archivo, para un transporte o almacenamiento es benéfico. Muchos archivadores de ficheros pueden realizar compresión sin pérdida de datos para reducir el tamaño del fichero archivo. [1] Esta compresión se logra mediante una serie de algoritmos que permiten que los datos contenidos en un fichero ocupen menos tamaño sin que se pierda la información.

La mayoría de los archivadores básicos simplemente toman una lista de ficheros y concatenan su contenido secuencialmente en el archivo. Además, el archivo debe contener también alguna información sobre al menos los nombres y longitudes de los originales, para que sea posible una buena reconstrucción. Y estas también almacenan los metadatos de los ficheros que proporciona el sistema operativo, como marcas temporales, propietario y control de acceso.

C. Algoritmos de compresión sin perdida

Se denomina algoritmo de compresión sin pérdida a cualquier procedimiento de codificación que tenga como objetivo representar cierta cantidad de información utilizando u ocupando un espacio menor, siendo posible una reconstrucción exacta de los datos originales. [2] Es decir, la compresión sin pérdidas engloba a aquellas técnicas que garanticen generar un duplicado exacto del flujo de datos de entrada después de un ciclo de compresión / expansión. Por esta razón es utilizada para comprimir archivos que contienen datos que no pueden ser degradados o perdidos, como pueden ser documentos de texto, imágenes y sonido.

Se fundamenta en conceptos de la Teoría de la Información, como la Redundancia y Entropía de los datos (ver compresión de datos) y es generalmente implementada usando uno o dos tipos de modelos diferentes: el estático y aquel basado en diccionario [2].

D. Algoritmos de compresión con perdida

La compresión con pérdida solo es útil cuando la reconstrucción exacta no es indispensable para que la información tenga sentido. La información reconstruida es solo una aproximación de la información original. Suele restringirse a información analógica que ha sido digitalizada (imágenes, audio, video, etc.), donde la información puede ser "parecida" y, al mismo tiempo, ser subjetivamente la misma. Su mayor ventaja reside en las altas razones de compresión que ofrece en contraposición a un algoritmo de compresión sin pérdida. [3] La compresión con pérdida acepta una pérdida de datos para poder mejorar el factor de compresión. Se aplica generalmente al almacenamiento digital de datos analógicos como ficheros de gráficos y de sonidos. La gran ventaja de comprensión con pérdida es alcanzar una tasa de compresión más elevadas a costa de sufrir una pérdida de información sobre la imagen original.

E. Formato de compresión ZIP

El formato Zip fue diseñado originalmente por Phil Katz, fundador de PKWARE [2] . Katz liberó al público la documentación técnica del formato ZIP, y lanzó al mismo tiempo la primera versión de PKZIP en enero de 1989.

ZIP es un formato de fichero bastante simple, que comprime cada uno de los archivos de forma separada, comprime cada archivo independientemente del resto de archivos comprimidos permite recuperar cada uno de los ficheros sin tener que leer el resto, lo que aumenta el rendimiento [1]. La especificación de ZIP indica que cada archivo puede ser almacenado, o bien sin comprimir, o utilizando una amplia variedad de algoritmos de compresión. Sin embargo, en la práctica, ZIP se suele utilizar casi siempre con el algoritmo de Phil Katz.

En la actualidad, los archivos ZIP emplean la extensión de fichero ".zip" y tienen el tipo media MIME application/zip. Un archivo ZIP contiene uno o más archivos que están comprimidos o almacenados. Muchas aplicaciones software, distintas de PKZIP, están disponibles para crear, modificar o abrir archivos zip, destacan WinZip, WinRAR, PeaZip y 7-Zip.

F. Formato de compresión RAR

En informática, Rar (cuyas siglas significan Roshal Archive en inglés) es un formato de archivo propietario, con un algoritmo de compresión sin pérdida utilizado para la compresión de datos y archivado, desarrollado por el ingeniero de software ruso Eugene Roshal. [1] RAR utiliza un algoritmo de compresión basado en el LZSS que, a su vez, se basaba en el LZ77, de James Storer y Thomas Szymanski (1982). El tamaño del diccionario puede variar entre 64k y 1024 Mb.

El RAR es más lento que el Zip, pero posee una mayor tasa de compresión. Otra característica de Rar es que posee una mejor redundancia de datos que Zip.

[2] PKWARE Inc: Compañía de software empresarial.

Por otra parte, a diferencia del formato Zip, el formato Rar permite además de codificar los datos, codificar otras áreas sensibles del archivo: nombres de ficheros, tamaños, atributos, comentarios y otros bloques [2].

G. Seguridad en el cifrado en Zip y Rar

Hasta la fecha, los archivos RAR están codificados en AES-256 en modo CBC[3] para los archivos con formato RAR esta función disponible en la última versión 5.0 y con AES-128 en modo CBC para RAR 4.x. La función de derivación de la clave en archivos RAR 5.0 está basada en PBKDF2[4] usando HMAC-SHA256[5] [3]. Por otra parte, los archivos ZIP usan el cifrado AES-256 en modo CTR[6].

En efecto en su manual de usuario (WinRAR) para resguardar la información sugiere y recomienda el cifrado en el formato RAR según posee la mayor resistencia a los ataques de fuerza bruta. Así mismo sugiere el uso de contraseñas de al menos 8 caracteres, contraseñas mixtas de mayúsculas con minúsculas y caracteres especiales, evitar el uso de palabras comunes en la contraseña ya que la hacen más débil [3].

H. Métodos de descompresión de archivos de información con clave protegida

A continuación, se describe los métodos generalmente utilizados para descifrar claves o contraseñas y llegar al objetivo, así mismo estas técnicas son de base por los programas software de ayuda a descompresión con clave olvidado u otro:

a) Fuerza Bruta: El descifrado con fuerza bruta es una forma de probar los valores formados por todos los caracteres hasta que se haya confirmado la contraseña correcta. El método es primitivo pero efectivo a la hora de probar todas las combinaciones de contraseña, definitivamente puede encontrar la contraseña correcta.

TABLA I. TIEMPO ESTIMADO EN HALLAR LA CONTRASEÑA

Tipos de caracteres	Máx. tiempo de crack (aprox.)		
	Longitud		
	4 dígitos	6 dígitos	8 dígitos
Letras (sin distinción entre mayúsculas y minúsculas)	3 segundos	37 minutos	17 días
Letras (distinción entre mayúsculas y minúsculas) + números	2 minutos	5 días	50 años
Letras (distinción entre mayúsculas y minúsculas) + números + símbolo	9 minutos	54 días	1,000 años

[a.] Fuente: Rene E Laboratory

b) Diccionario: El descifrado de diccionario es una parte especial del descifrado de fuerza bruta, donde el diccionario posee la gran cantidad de posibles contraseñas de uso común basadas en la gran cantidad de idiomas humanos y estas disponibles en el internet. Así mismo estos posibles valores de contraseña se pueden obtener a base de una ingeniería social o estudio hacia la usuario víctima y estas incluirlas en el diccionario. Mediante un programa de software en un proceso de búsqueda de contraseña en el diccionario, reducen aún más la dificultad y el tiempo para descifrar y hallar la contraseña. Sin embargo, sin saber la longitud de la contraseña y los caracteres de la contraseña (letras, números, minúsculas y mayúsculas), si la contraseña es complicada, llevaría mucho tiempo (quizás 1,000 años) para descifrarla

c) Mascara: Otro de los métodos donde si conoce la longitud de la contraseña o el número de caracteres, se puede descifrar la contraseña por máscara y en menor tiempo posible a base de un programa de software.

III. MÉTODOS

Para la obtención de información, se empleó el método de observación y el uso de herramientas lógicas, computacionales, como el lenguaje de programación.

Según [4] sostiene que una investigación es científica, su razonamiento es la inferencia que hace uso de la noción de "fuerza inductiva", con la cual se obtiene conclusiones generales a partir de premisas particulares. Se distingue en cuatro pasos esenciales: la observación de los hechos para su registro; la clasificación y el estudio de estos hechos; la derivación inductiva que parte de los hechos y permite llegar a una generalización; y la contrastación.

Por otra parte, según Vara[5] afirma *"Los diseños explicativos se usan para determinar las causas de los fenómenos empresariales. Con estos diseños se puede explicar por qué ocurre, bajo qué condiciones se presenta, o por qué dos o más variables están correlacionadas. Las investigaciones explicativas no se contentan con descubrir qué es lo que causa un determinado hecho, sino también busca aclarar por qué lo causa"*.

Instrumentos de relevamiento de la información. A continuación, se describe el procedimiento empleado:

- Análisis y estudio de las características de RAR y ZIP respecto velocidad proceso, tipo de compresión y el tipo de cifrado de datos.
- Análisis y estudio de métodos de descompresión de archivos RAR y ZIP, protegidos con clave de acceso.
- Comprobar la efectividad en descompresión mediante programas disponibles en el internet a base de archivos de prueba modificando el tamaño y tipo archivos.

IV. RESULTADOS

En esta sección de denotará las formas de descompresión de archivos Zip y Rar estas previamente comprimidas y cifradas mediante el clave acceso respectivo, las pruebas se efectúa a base de los métodos de fuerza bruta y diccionario.

Así mismo, las contraseñas se plantearon de acuerdo el uso o creación de las mismas tales como caracteres mínimas y simples tales como: "54321", "12345abc", "password123" y otras generalmente usadas.

Forma 1: Descompresión de Archivo Zip o Rar mediante un programa escrita en lenguaje de programación Python.

[3] CBC: Cipher block chaining (cifrado por bloques).
[4] PBKDF2: Función de derivación clave con un costo computacional variable.
[5] SHA256: Función hash criptográfico.
[6] CTR: Counter Mode (cifrado de flujo aleatorio).

Fig. 1. El archivo "secreto.zip" se muestra con acceso restringido por contraseña (Creación Propia)

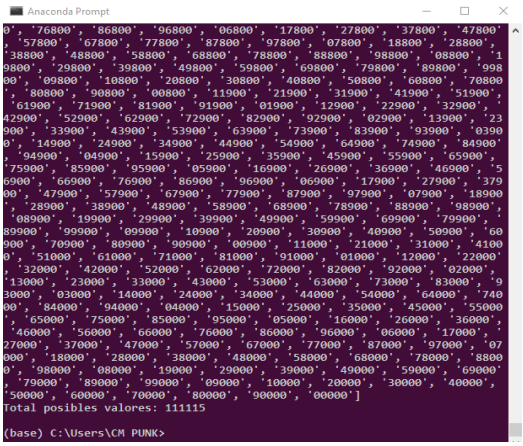

Fig. 2. El script desarrollado, para una contraseña de cinco caracteres de tipo numérico, el programa genera 111.115 de posibles combinaciones (Creación Propia)

```python
import zipfile

charlist='1234567890'
encontrado=[]
cont=0
for current in range(5):
    a= [i for i in charlist]
    for x in range(current):
        a= [y+i for i in charlist for y in a]
    encontrado=encontrado+a

#print(encontrado)
z=zipfile.ZipFile(r'C:\Users\CM PUNK\forsai\secreto.zip','r')
intentos=0

print("\nIniciando...")
for password in encontrado:
    try:
        intentos+=1
        z.setpassword(password.encode('ascii'))
        z.extractall(r'C:\Users\CM PUNK\forsai')
        print(f'Contraseña encontrado despues { intentos} intentos!, 
        break
    except:
        pass

print("Finalizado.")
```

Fig. 3. El programa ejecutado, según las instrucciones probar todos de valores hasta acertar la contraseña correcta y descomprimirlo el archivo víctima. (Creación Propia)

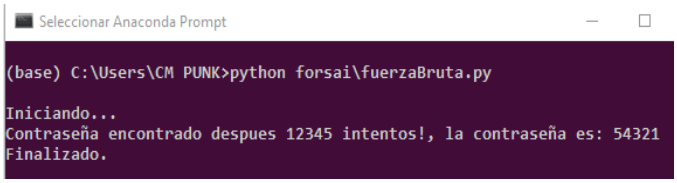

Fig. 4. Ejecución final de programa y el hallazgo de contraseña para descomprimir. (Creación Propia)

El primer método se evaluó con resultados esperados a través de un programa Python y demostrando la vulnerabilidad existente. Y esta una de las formas para hallar la contraseña y descomprimir el archivo protegido. Cabe destacar que la contraseña "54321" fue relativamente singular.

Forma 2: Descompresión de Archivo Zip por ataque de Diccionario a base de un programa escrita en Python

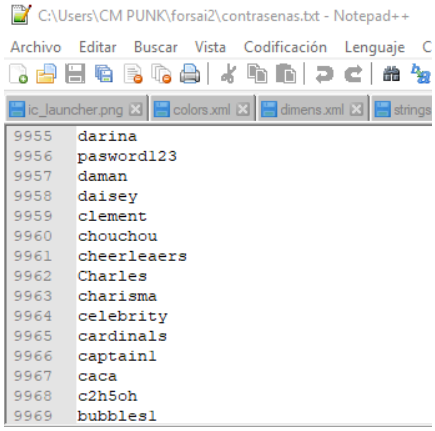

Fig. 5. Lista de contraseñas en el Diccionario y esta posee 9.969 de posibles valores (Creación Propia)

```python
#ATAQUE DE DICCIONARIO
import zipfile

posicion=0
#Abrir archivo median la libreria ZipFile
zip_file = zipfile.ZipFile(r'C:\Users\CM PUNK\forsai2\doc.nucleares.zip','
#Lectura de dicionario de contraseñas
with open(r'C:\Users\CM PUNK\forsai2\contrasenas.txt', 'r') as f:
    print("\nIniciando...")
    for line in f.readlines():
        password = line.strip('\n')
        try:
            posicion+=1
            zip_file.setpassword(password.encode('ascii'))
            zip_file.extractall(r'C:\Users\CM PUNK\forsai2')
            print(f'Contraseña encontrado en la posicion {posicion} y la c
            print("Finalizado.")
            break
        except:
            pass
```

Fig. 6. Ejecución de programa, según las instrucciones probar todos de valores existentes en el Diccionario hasta acertar la contraseña correcta y descomprimirlo el archivo víctima. (Creación Propia)

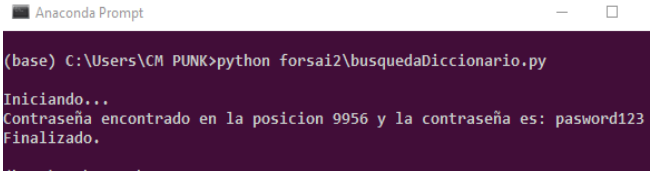

Fig. 7. Ejecución final de programa y el hallazgo de contraseña para descomprimir. (Creación Propia)

El segundo método se verifico la vulnerabilidad existente mediante el uso de diccionario, donde este diccionario posee la cantidad máxima de contraseñas posibles. Asimismo, denotar que en el Internet existen las grandes publicaciones de diccionario de contraseñas.

Forma 3: Descompresión de Archivo por Programas comerciales

Otra de las formas de descomprimir es a través de programas software disponibles en Internet, algunos de ellos con acceso gratis y de prueba.

Mediante una búsqueda de programas referente de nuestro interés, se encontraron la gran variedad de programas tales como: ARCHPR, Web Crack, Passfab, Passware. Kit Forennsic, Renee File Protector y entre otras.

Cabe destacar que, los programas mencionados se basan principalmente en los métodos de fuerza bruta, ataque de diccionario, mascara, texto sin formato y otras técnicas que ayudan a lograr el objetivo.

V. Discusión

De acuerdo a los resultados obtenidos, la forma de administrar archivos de información través de compresión de tipo Zip y Rar con clave de acceso, de puede determinarse seguras y efectivas según el usuario que use el programa, por ejemplo, crear una contraseña de mayor longitud de caracteres de letras, números y símbolos, ya que esta acción en un ataque dificultará a encontrar la contraseña correcta.

Asimismo, los creadores de software actualizan y publican las nuevas versiones de programas de compresión y descompresión con funciones especiales así mismo incluyendo los cifrado de archivos de manera continua, este a favor del usuario. Cabe destacar, los creadores de software recomiendan actualizar programas y así mismo el uso de contraseñas validas sin embargo el usuario siempre es el eslabón débil en su creación de contraseñas.

Por otra parte, WinRAR respecto a la vulnerabilidad en el uso de librería ACE, determinó eliminar por completo el uso de la librería mencionada y estas ya implementadas en las versiones posteriores a 5.70.

VI. Conclusión

En el presente trabajo de desarrollo practico y las pruebas de compresión y descompresión de archivos Zip y Rar, con restricción de acceso a identidades no autorizadas, se concluye la vulnerabilidad de confidencialidad existente, ya que existe métodos funcionales de descompresión de archivos de manera ilegal así mismo programas software para lograr el objetivo; sin embargo, esto será determinante por parte del usuario a la hora de crear una contraseña, en escenarios donde cumpla la forma de crear una contraseña con el uso de combinación de mayúsculas, minúsculas, número y símbolos; en beneficio la confidencialidad de archivos será a favor del usuario.

Finalmente, como recomendación es evitar el uso de contraseña débiles de caracteres simples tales como uso números simples, la secuencia inicial de abecedario y otras labras de común y también el uso de programas de software actualizado.

Referencias

[1] A. D. Catillo, «Aplicación de Programas Utilitarios en Sistemas Windows y Linux,» p. 56, 2015.

[2] N. Grossman, «Checkpoint,» 20 Febrero 2019. [En línea]. Available: https://research.checkpiont.com/2019/extracting-code-execution-from-winrar.

[3] A. Roshal, «RarLab,» 2020. [En línea]. Available: https://rarlab.com.

[4] WinRar5, «https://www.winrar.es/soporte/manual,» Marzo 2020. [En línea].

[5] K. Popper, «Revisión de su legado,» 2013. [En línea]. Available: https://www.academia.edu/27295692/Karl_R._Popper_Revisi%C3%B3n_de_su_legado.

[6] A. A. H. Vara, Siete pasos para una Tesis Exitosa, Lima-Perú: RenSol, 2012.

[7] C. Tori, Hacking Ético, Argentina, 2015.

[8] C. Mateo, Software Libre-Desarrollo de Aplicaciones Web, Barcelona: UOC, 2008.

[9] L. Javajan, «Javajan,» 26 Enero 2014. [En línea]. Available: http://www.guiadiseño.com/05_arquitectura.php.

[10] G. Hoglund y G. McGraw, Explotación de software, Addison-Wesley, 2004.

[11] A. J. Ayllon, *Criptografía y Seguridad en WhatsApp*, Barcelona, 2016.

[12] J. R. Aguirre, Criptografía Aplicada, Madrid: Creative Commons, 2019.

[13] «ReeneLaboratory,» ReeneLaboratory, [En línea]. Available: https://www.rennelaboratory.com.