Metodología para análisis de archivos descargados de foros de hacking

Archivos: 7z, rar, xz, txt, csv

ReadMe: El presente documento describe una metodología práctica basada en buenas prácticas para el análisis estático de archivos obtenidos de foros de hacking. El proceso se desarrolla en un laboratorio aislado de Internet mediante una máquina virtual (VM) e incorpora herramientas especializadas como Detect It Easy (DIE) y Malcat para inspeccionar archivos con extensiones como .7z, .rar, .txt, .csv, entre otras, con el objetivo de identificar posibles indicadores de malware antes de interactuar con su contenido.

🕵️ El Rol del Threat Intelligence Researcher

El rol de un **Cyber Threat Intelligence Researcher** consiste en recopilar, analizar y transformar información del ciberespacio en inteligencia útil para apoyar la toma de decisiones. En este contexto, una de las fuentes de información son los foros de hacking, donde se publican filtraciones de datos que pueden afectar a empresas, organizaciones y organismos gubernamentales.

El proceso de investigación comprende dos etapas principales. La primera corresponde a la **fase de adquisición**, donde se recopilan los archivos publicados en estos foros. La segunda corresponde a la **fase de análisis**, cuyo objetivo es verificar que los archivos descargados no representen un riesgo antes de proceder con su extracción o revisión.

En esta segunda etapa resulta fundamental aplicar una metodología de análisis estático que permita identificar indicadores de comportamiento sospechoso, contenido ejecutable oculto o cualquier otra evidencia que pueda comprometer la seguridad del entorno de investigación.

Es importante considerar que la mayoría de los archivos obtenidos en este tipo de investigaciones corresponden a formatos como .txt, .csv, .7z o .rar. Los archivos .txt y .csv no son ejecutables por naturaleza y, por sí mismos, no ejecutan código. Por este motivo, el análisis debe centrarse inicialmente en verificar la verdadera naturaleza del archivo, validar su estructura e identificar la posible presencia de contenido ejecutable o elementos sospechosos antes de descomprimirlo o visualizar su contenido.

🔍 Metodología de Análisis Estático

FASE 1 — Preservación de Evidencia

En esta fase solo se realizará la tarea de creación de directorios de trabajo.

```
Case_01/
    -> Original/
    -> Extracted/
    -> Reports/
    -> Hashes/
```

```
C:\Users\malware\Documents\darkforums_leak\1M_ARGENTINA_EmailPass_UHQ_Combolist>tree
Folder PATH listing
Volume serial number is D27F-CE24
C:.
├───extracted
├───hashes
├───original
└───reports
```

FASE 2 — Hashing

Calcular los hashes (MD5, SHA1, SHA256) de nuestros archivos con el objetivo de identificar archivos duplicados, mantener cadena de custodia y posterior comparación. Utilizar la herramienta “Certutil” de línea de comandos de Windows para calcular el hash de los archivos.

[COMANDO]: `certutil -hashfile "Ruta\Al\Archivo.extension" [Algoritmo]`

```
C:\Users\malware\Documents\darkforums_leak\1M_ARGENTINA_EmailPass_UHQ_Combolist\hashes>dir
 Volume in drive C has no label.
 Volume Serial Number is D27F-CE24

 Directory of C:\Users\malware\Documents\darkforums_leak\1M_ARGENTINA_EmailPass_UHQ_Combolist\hashes

07/21/2026  03:24 PM    <DIR>          .
07/21/2026  03:24 PM    <DIR>          ..
07/21/2026  03:20 PM               141 md5_hashfile.txt
07/21/2026  03:24 PM               176 sha256_hashfile.txt
               2 File(s)            317 bytes
               2 Dir(s)  32,667,459,584 bytes free
```

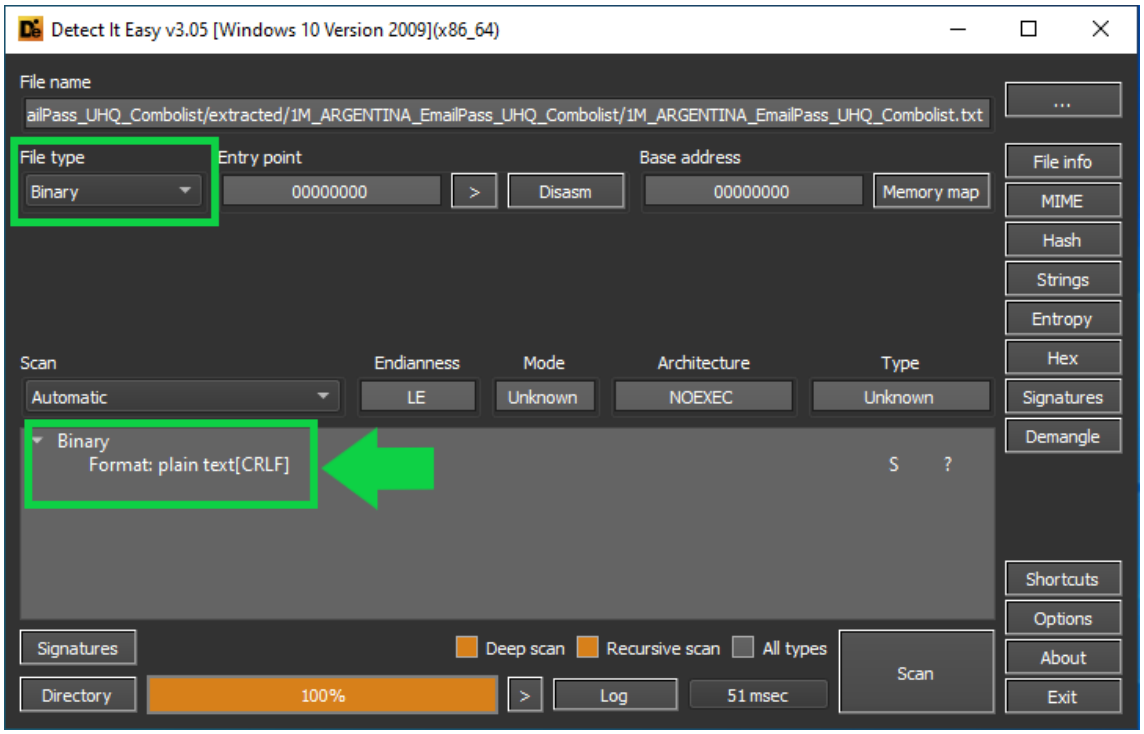
```
C:\Users\malware\Documents\darkforums_leak\1M_ARGENTINA_EmailPass_UHQ_Combolist\hashes>type md5_hashfile.txt
MD5 hash of 1M_ARGENTINA_EmailPass_UHQ_Combolist.7z:
1d083ef67001c25d5cc64f2ee991217b
CertUtil: -hashfile command completed successfully.

C:\Users\malware\Documents\darkforums_leak\1M_ARGENTINA_EmailPass_UHQ_Combolist\hashes>type sha256_hashfile.txt
SHA256 hash of 1M_ARGENTINA_EmailPass_UHQ_Combolist.7z:
4b517f1ef6e782c8622ed6dbeada064ff106be4faa821b942750d76c06259598
CertUtil: -hashfile command completed successfully.
```

FASE 3 — Identificación Real del Archivo

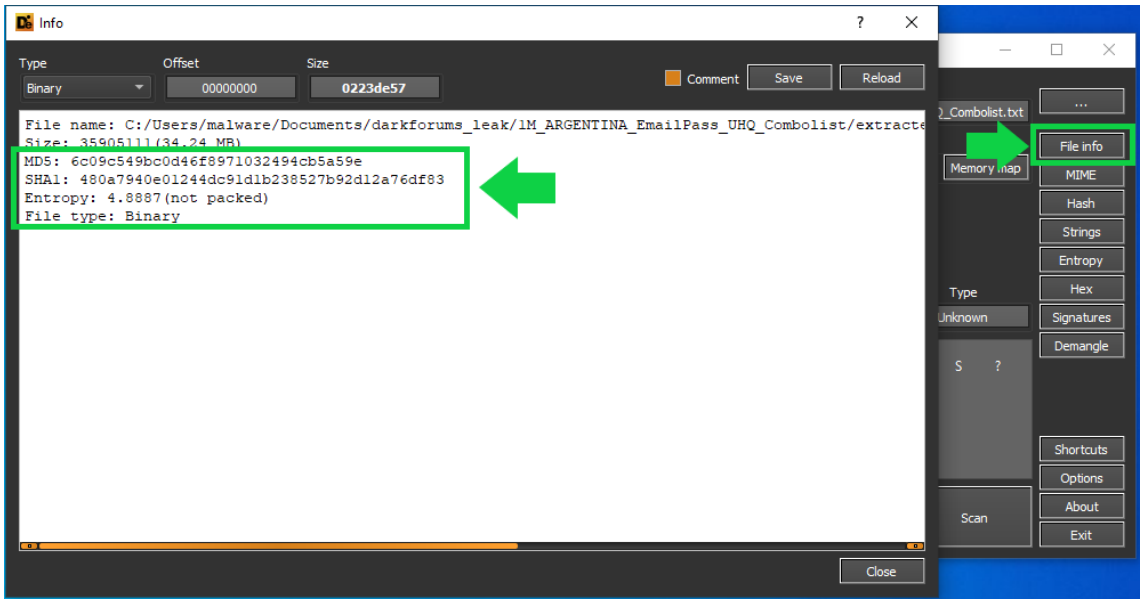
Identificar el tipo de archivo real mediante la visualización del valor del encabezado (`PE32, ELF, Mach-O, Archive, Text, Binary`). Si el tipo de archivo que estamos analizando es un TXT pero la herramienta Detect It Easy (DIE) nos arroja “`PE32 Executable`” entonces estamos ante un archivo ejecutable (red flag).

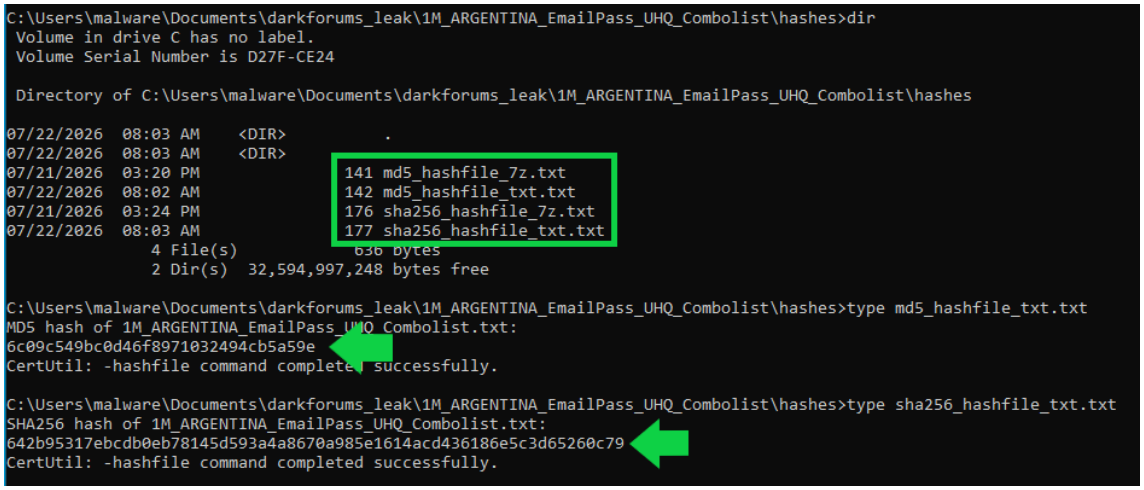
Herramienta:

- Detect It Easy



La herramienta DIE identifica que nuestro archivo es un binario de formato texto plano (plain text). También podemos consultar esta información con la opción “MIME”.



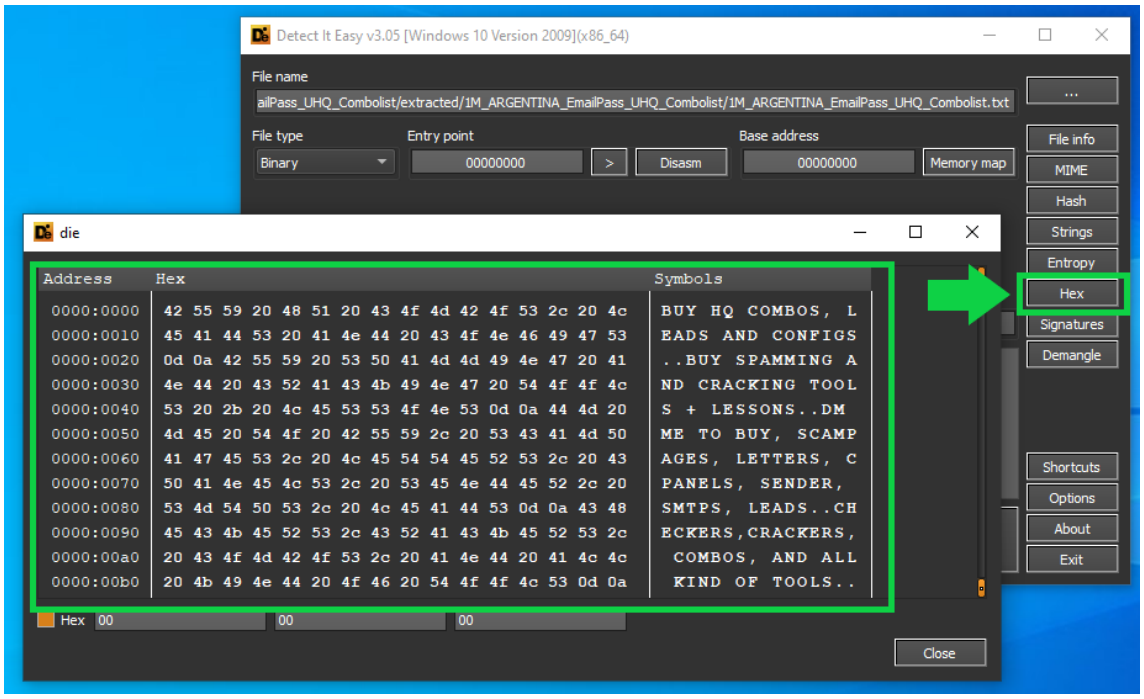


En la opción “File Info” y “Hash” podemos verificar los mismos hashes que generamos previamente en nuestro directorio “hashes”.

FASE 4 — Magic Bytes

Verificar el valor de la cabecera de los “Magic Numbers” o “Magic Bytes”. Este valor deberá coincidir con el tipo de extensión del archivo analizado (ej: TXT, CSV, 7Z, RAR, TAR).

Magic Numbers List: https://en.wikipedia.org/wiki/List_of_file_signatures



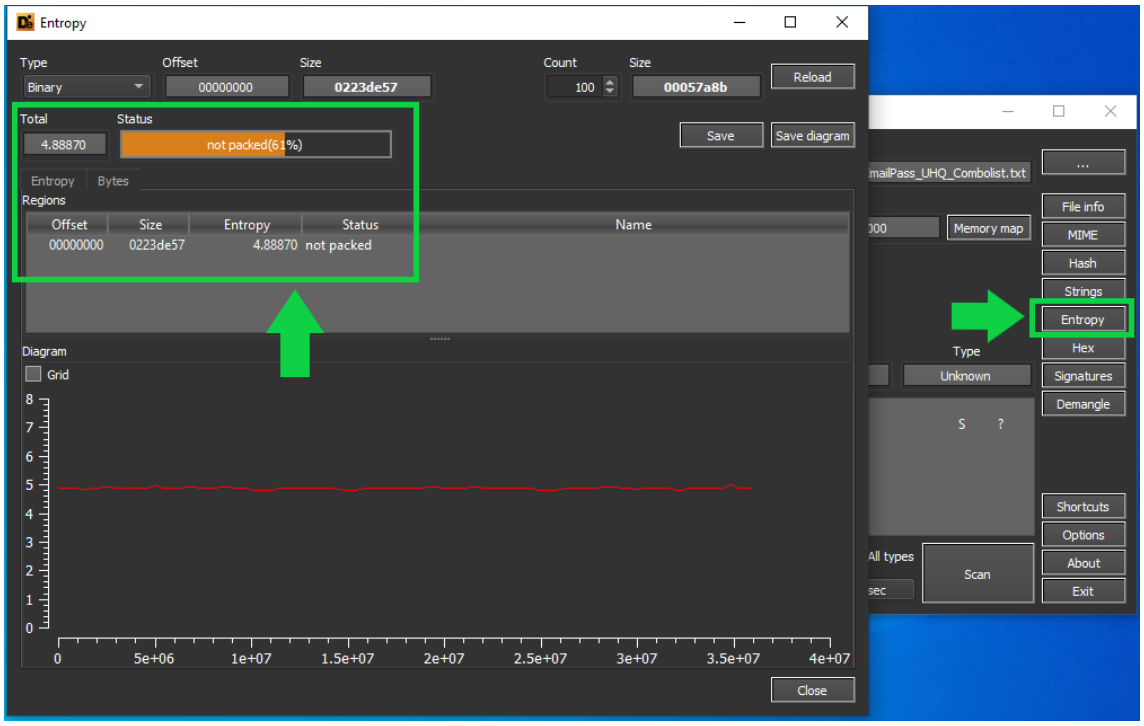
IMPORTANTE: los archivos de texto plano (.txt) no tienen un número mágico (Magic Number). Dado que los archivos de texto plano consisten únicamente en datos de caracteres sin procesar (como ASCII o UTF-8).

En el caso de la imagen se puede observar que nuestro archivo (.txt) no posee Magic Number, comienza directamente a mostrar el texto almacenado en el archivo de texto.

FASE 5 — Entropía

Analizar la entropía del archivo. Cualquier archivo con entropía muy alta merece una revisión adicional, ya que podría contener datos cifrados, comprimidos o binarios embebidos.

Normal	Comprimido	Cifrado/Packet
0-4	5-6	7+



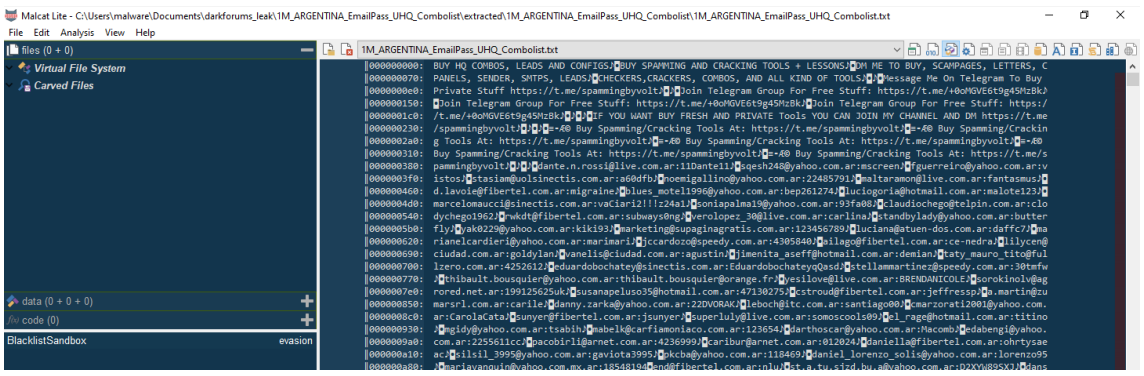
Nuestro archivo de texto plano (.txt) posee baja entropía (4.88) y está catalogado sin compresión (Not packet).

FASE 6 — Strings

Analizar el archivo mediante un visualizador de strings para identificar cadenas de texto sospechosas. En TXT y CSV normalmente deberían verse únicamente texto o datos estructurados. La presencia de comandos o cadenas propias de ejecutables puede justificar una inspección más profunda del contenido.

Si se identifican algunas de las siguientes cadenas el archivo es sospechoso:

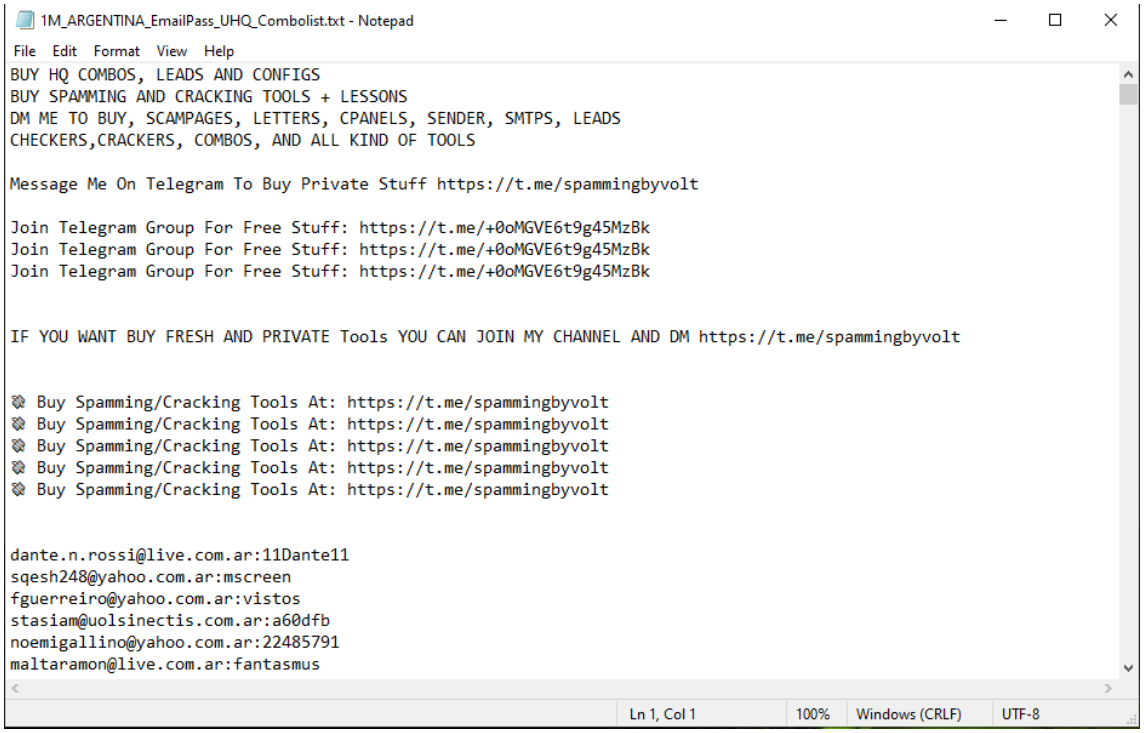
`http, https, .onion, powershell, cmd.exe, rundll32, regsvr32, wscript, cscript, certutil, curl, wget, Invoke-, FromBase64, MZ, This program cannot be run.`



FASE 7 — Inspección del Contenido

ADVERTENCIA: Solamente ejecutar este paso si efectivamente se verifico que el archivo no es malicioso.

Descomprimir el archivo y utilizar preferentemente un editor de texto plano para visualizar la información.



[📡] Contacto

[🔗] https://www.linkedin.com/in/david-padron-9a74aa323/

[🔗] https://github.com/FeathersMcgr4w

[🔗] https://davidpadron.info