

Herramientas para Recuperar Archivos Borrados

# RESTAURACIÓN

Gernot Krautberger, Fotolia

Los sistemas de ficheros modernos complican la tarea de recuperar ficheros por medio de técnicas forenses. Herramientas como Foremost y Scalpel identifican estructuras de datos y recuperan ficheros de una imagen del disco duro. **POR RALF SPENNEBERG**

Expertos en tecnologías de la información e investigadores tienen diversas razones por las que recuperar ficheros borrados. Ya sea porque un intruso haya borrado un registro para ocultar un ataque o porque un usuario haya borrado accidentalmente una colección de fotografías digitales con *rm -rf*, algún día puede que tengamos que enfrentarnos con la necesidad de recuperar datos borrados. En el pasado, los expertos en recuperaciones podían obtener fácilmente los ficheros perdidos, ya que las primeras generaciones de los sistemas de ficheros simplemente borraban la entrada en el directorio. La meta información que describe la localización física de los datos en el disco se preservaba, mientras que herramientas como “The Coroner’s Toolkit (TCT [1])” y “The Sleuth Kit (TSK [2])” podían recuperar la información necesaria para restaurar el fichero.

Actualmente, la mayoría de los sistemas de ficheros borran el conjunto completo de metainformación, dejando los bloques de datos. La operación de juntar estas piezas de forma correcta se denomina restauración de ficheros; los expertos forenses restauran los datos del disco y reconstruyen los ficheros a

partir de ellos. Cuanto más fragmentado esté el fichero, más complicada será la tarea.

Existen diversas herramientas de código abierto que automatizan el proceso de restauración: La lista está encabezada por Foremost [3] y sus derivados Scalpel [4], aunque hay otras como PhotoRec [5] y Ftimes [6]. PhotoRec no soporta la restauración genérica de cualquier tipo de ficheros y Ftime es tan complicado de utilizar que para la mayoría de los usuarios no les vale la pena.

Foremost y Scalpel no se interesan por el sistema de ficheros subyacente. Simplemente esperan que los bloques de datos de los ficheros residan secuencialmente en la imagen a investigar. Las herramientas encontrarán las imágenes gracias a los volcados de *dd*, a los volcados RAM o a los ficheros de intercambio. La restauración ayudará a identificar y reconstruir ficheros en sistemas de

ficheros corruptos, en los huecos al final de los clusters o incluso tras la instalación de un nuevo sistema operativo, siempre y cuando sigan existiendo los bloques de datos necesarios.

Por supuesto, ninguna de estas herramientas hacen milagros y no están diseñadas para obtener datos de discos duros dañados físicamente. Tampoco el proceso de restauración podrá acceder a los bloques de datos que hayan sido sobrescritos.

Como las herramientas de restauración no se basan en el sistema de ficheros, necesitan otras fuentes de información para descubrir dónde comienza y termina un fichero. Afortunadamente, muchos tipos de ficheros poseen estructuras conocidas. A menudo, todo lo que se necesita para identificar el tipo de fichero y su localización es la cabecera y el pie del fichero. El comando *file* de Linux también utiliza la información de la cabecera y el pie para identificar los tipos de los ficheros.

Los restauradores de ficheros investigan el disco duro completo, o la imagen del disco, para localizar las cabeceras y los pies. Luego restauran los bloques entre la cabecera y el pie, y almacenan los datos como un fichero nuevo.

## Listado 1: Configuración

```
01 gif y 155000000 \x47\x49\x46\x38\x37\x61 \x00\x3b
02 gif y 155000000 \x47\x49\x46\x38\x39\x61 \x00\x00\x3b
03 jpg y 200000000 \xff\xd8\xff\xe0\x00\x10 \xff\xd9
04 jpg y 200000000 \xff\xd8\xff\xe1\xff\xd9
05 jpg y 200000000 \xff\xd8 \xff\xd9
```

## Listado 2: Ejecución de Foremost

```

01 Foremost version 1.5.3 by Jesse Kornblum, Kris Kendall, and Nick Mikus          20 [...]
02 Audit File                    21 20: 00045015.zip 274 KB 23047680
03                               22 21: 00007982.png 6 KB 4086865 (1408 x 1800)
04 Foremost started at Sat Feb 9 18:36:29 2008 23 22: 00033012.png 69 KB 16902215 (1052 x 360)
05 Invocation: ./foremost -v -T -i 24 23: 00035391.png 19 KB 18120696 (879 x 499)
  ../dfrws-2006-challenge.raw    25 24: 00035431.png 72 KB 18140936 (1140 x 540)
06 Output directory:            26 *|
  /linux-magazin/foremost/foremost-1.5.3/output_Sat_Feb_9_18_36_29_2008
07 Configuration file:          27 Finish: Sat Feb 9 18:36:32 2008
  /linux-magazin/foremost/foremost-1.5.3/foremost.conf
08 Processing: ../dfrws-2006-challenge.raw 28
09 _____                    29 25 FILES EXTRACTED
10 File: ../dfrws-2006-challenge.raw 30
11 Start: Sat Feb 9 18:36:29 2008 31 jpg:= 11
12 Length: 47 MB (49999872 bytes) 32 htm:= 5
13                               33 ole:= 2
14 Num Name (bs=512) Size File Offset Comment 34 zip:= 3
15                               35 png:= 4
16 0: 00003868.jpg 280 KB 1980416          36 _____
17 1: 00008285.jpg 594 KB 4241920          37
18 2: 00011619.jpg 199 KB 5948928
19 3: 00012222.jpg 6 MB 6257664          38 Foremost finished at Sat Feb 9 18:36:32 2008

```

Algunos tipos de ficheros no poseen pies únicos. Los restauradores tendrán al menos que suponer dónde termina el fichero a partir de que conozcan dónde comienza la siguiente cabecera. Por supuesto, cualquier cantidad de datos no identificados podría residir entre el final del fichero y el comienzo de la siguiente cabecera.

Para evitar tener que recolectar datos innecesarios, los programas de restauración permiten que los usuarios establezcan el tamaño máximo de los ficheros. Desafortunadamente, las cabeceras y los pies son a menudo pequeños, lo que lleva a que se produzcan numerosos falsos positivos.

Los formatos de imágenes son una excepción. Por ejemplo, cada fichero JPEG comienza con la secuencia de bytes `0xFFD8`, seguida normalmente por `0xFFE00010`. Los restauradores de ficheros lo tienen en cuenta a la hora de identificar las imágenes JPEG. Sin embargo, si algunos bloques han sido sobrescritos o si el fichero está fragmentado, las herramientas sólo restaurarán una parte del fichero en el mejor de los casos (Figura 1).

## Foremost y Scalpel

Jesse Kornblum y Kris Kendall, pertenecientes a la Oficina de Investigaciones Especiales de la Fuerza Aérea de los Estados Unidos, desarrollaron Foremost en Marzo de 2001 como una herramienta para analizar y recuperar ficheros borrados. La herramienta de restauración Foremost está inspirada en un programa ante-

rior llamado CarvThis, creado en 1999 por el Laboratorio Forense de Defensa Informática, que sólo se utilizó internamente y no llegó al público en general. Foremost es ahora código abierto y Nick Mikus mantiene el código fuente tras mejorar el programa de forma significativa como parte de su doctorado.

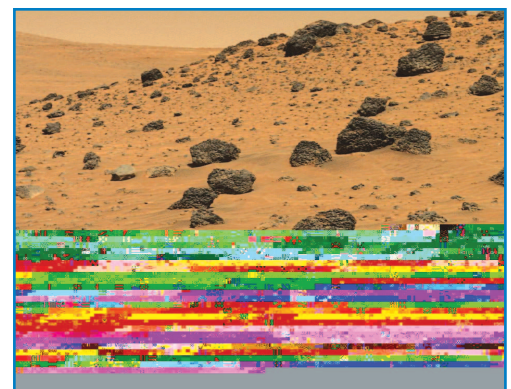
Golden G. Richard III desarrolló un programa separado llamado Scalpel basándose en Foremost 0.69. Durante bastante tiempo, Scalpel se ha considerado una herramienta avanzada. Algunas fuentes dicen que incluso los desarrolladores de Foremost recomiendan Scalpel [7]. Para ser más precisos, ambos proyectos se encuentran bajo un desarrollo activo. Aunque Scalpel era bastante mejor que su predecesor en 2005, con su habilidad de analizar imágenes unas 10 veces más rápido, Foremost lo ha alcanzado recientemente gracias a Nick Mikus, siendo superior en algunas tareas.

Tanto Foremost como Scalpel utilizan ficheros de configuración para especificar qué ficheros tienen que buscar (Listado 1). La primera columna designa el tipo de fichero y también especifica la extensión del fichero a añadir a cualquier otro que el programa encuentre. Los ficheros para los que las mayúsculas y las minúsculas son relevantes en la cabecera y en el pie tienen una *y* en la segunda columna; para el resto se pone *n*. La siguiente columna define el

tamaño máximo del fichero, seguido por la secuencia de bytes de la cabecera y la secuencia de bytes del pie, en el caso de existir. La cadena `\x` permite insertar bytes en notación hexadecimal; las otras posibilidades son `\s` para un espacio y `?` como comodín para cualquier carácter. Pueden seguir otras opciones al final.

## Buscador Rápido

Por su origen, Scalpel utiliza el mismo fichero de configuración que Foremost, aunque las dos herramientas funcionan internamente de forma diferente. Ambas encuentran más o menos los mismos ficheros, pero existen algu-



**Figura 1: Los restauradores de ficheros ignoran el sistema de ficheros y restauran las imágenes directamente desde los bloques de datos. En el caso de ficheros fragmentados, el restaurador devuelve una fotografía imperfecta, pero esta imagen podría ser suficiente para identificar el asunto.**

nas discrepancias en la identificación de éstos. Los expertos forenses recomiendan el uso de los dos programas.

A partir de la versión 0.9.1, Foremost utiliza una nueva técnica para identificar los ficheros ZIP, JPEG, Office y otros. Los formatos están implementados directamente en Foremost, de este modo el programa no requiere que en el fichero de configuración aparezca la información de la cabecera y el pie para el proceso de identificación. Foremost activa esta nueva función de detección si se añade el parámetro `-t` en la línea de comandos seguido de los tipos de ficheros requeridos:

```
foremost -T -t jpg,gif,pdf
-i fichero_imagen
```

Los formatos soportados se listan en la Tabla 1. Para activar todas estas funciones sólo hay que introducir `-t all`. La línea de comandos anterior también utiliza el parámetro `-T` para indicarle a Foremost que escriba cualquier fichero que encuentre a un directorio que tenga junto al nombre una marca de tiempo. De esta forma se organiza fácilmente la investigación forense, ya que cada vez que se ejecute se escribirán los resultados en un directorio nuevo.

### Requisitos de Espacio

La posibilidad de falsos positivos implica que la herramienta de restauración identi-

que una gran cantidad de datos, así que hay que asegurarse de que haya espacio libre en el sistema de ficheros de destino. El proceso de restauración no requiere necesariamente grandes cantidades de copias de archivos. Los sistemas de ficheros virtuales, como CarvFS [8], están diseñados para acceder a los datos directamente desde la imagen original. CarvFS, que está basado en FUSE (Sistema de Ficheros en el Espacio del Usuario), sólo espera que la herramienta de restauración proporcione una tabla que describa los ficheros que están disponibles y sus posiciones físicas. El sistema de ficheros CarvFS fue creado por el proyecto OCFA (Arquitectura Forense de Computación Abierta) de la policía holandesa (véase el artículo OCFA en este mismo número). Se utiliza en situaciones en las que el copiado de todos los ficheros a una localización aparte produce grandes cantidades de datos. En otros casos, sin embargo, el copiado de los datos es más eficiente que su acceso desde la imagen original.

En el Listado 2 se muestra una ejecución típica de Foremost sin activar las opciones internas. La imagen para este ejemplo fue proporcionada por el reto del DFRWS [9] (Taller de Investigación Forense Digital). DFRWS creó esta competición en 2006 para probar restauradores de ficheros y promover su desarrollo. Al final de la competición los organizadores publicaron una lista de los ficheros que contenía la imagen.

### PhotoRec

Si el sistema de ficheros no está completamente destruido, una alternativa importante a herramientas como Foremost y Scalpel la constituyen herramientas que evalúan el sistema de ficheros. La de recuperación PhotoRec [5] fue desarrollada por Christophe Grenier para rescatar fotografías de memorias Flash corruptas. PhotoRec también funciona en el caso de que la tabla de particiones esté dañada.

Una vez que PhotoRec haya identificado el sistema de ficheros, extrae una enorme variedad de tipos de ficheros. Además de ficheros fotográficos también restaura ficheros EXE y ZIP.

La herramienta soporta más de 180 tipos de ficheros. El programa se controla por medio de un menú en formato texto, con lo que se reduce el peligro por errores de los usuarios. Desafortunadamente, en la actualidad PhotoRec no puede analizar volcados RAM ni ficheros de intercambio.

### Fallos de Memoria

Los restauradores de ficheros ayudan a los investigadores forenses a recuperar ficheros borrados. Foremost y Scalpel ignoran el sistema de ficheros y pueden incluso restaurar datos de volcados RAM y ficheros de intercambio. Su velocidad es bastante impresionante.

Si el sistema de ficheros aún existe, una herramienta como PhotoRec también es útil para encontrar ficheros perdidos. ■

**Tabla 1: Opciones Internas de Foremost**

Formato	Comentario
<b>Imágenes</b>	
<b>JPG</b>	Formatos JFIF, Exif y RAW
<b>GIF</b>	Formato de Intercambio Gráfico
<b>PNG</b>	Gráficos Intercambiables a través de la Red
<b>BMP</b>	Fichero de mapas de bits de Windows
<b>Ejecutables</b>	
<b>EXE</b>	Windows PE, DLL y EXE
<b>Vídeo y Audio</b>	
<b>AVI</b>	Audio y Vídeo Entrelazado
<b>MPG</b>	Detecta todos los formatos MPEG que comiencen por 0x000001BA
<b>WMV</b>	Vídeo de Windows; WMA (Audio de Windows) en parte
<b>MOV</b>	Películas Quicktime
<b>Documentos</b>	
<b>PDF</b>	Formato de Documento Intercambiable
<b>OLE</b>	Objetos Enlazados y Embebidos; por ejemplo, PowerPoint, Word, Excel, Access, Starwriter
<b>DOC</b>	Sólo ficheros de Word
<b>HTM</b>	Lenguaje de Marcas de Hipertexto (sitios web)
<b>Formatos de Archivo</b>	
<b>ZIP</b>	ZIP, JAR, MS Office 2007, Open Office 2.0 (documentos XML comprimidos)
<b>RAR</b>	Archivo Roshal
<b>CPP</b>	Código fuente C; muchos falsos positivos.

### RECURSOS

- 1] The Coroner's Toolkit: <http://www.porcupine.org/forensics/tct.html>
- 2] The Sleuth Kit: <http://www.sleuthkit.org>
- 3] Foremost: <http://foremost.sf.net>
- 4] Scalpel: <http://www.digitalforensicssolutions.com/Scalpel/>
- 5] PhotoRec: <http://www.cgsecurity.org/wiki/PhotoRec>
- 6] FTimes: <http://ftimes.sourceforge.net/FTimes/>
- 7] Foremost en la Wiki Forensics: <http://www.forensicwiki.org/wiki/Foremost>
- 8] OCFA: <http://ocfa.sourceforge.net/libcarvpath/>
- 9] El reto de restauración DFRWS: <http://www.dfrws.org/2006/challenge/>