

The Foremost Open Source Forensic Tool

Ray Strubinger

(Reprinted from SysAdmin Magazine)

Several open source forensic tools have been created for the Unix platform; most notably, The Coroner's Toolkit (TCT) by Dan Farmer and Wietse Venema, as well as The @stake Sleuth Kit (TASK) and its browser-based front end, Autopsy, both created by Brian Carrier. While TCT and TASK/Autopsy have been written about frequently, there is another very useful application known as "foremost" that has received relatively little attention outside of a few computer forensics mailing lists.

Introduction to the Foremost Forensic Tool

Foremost is a free forensics tool created for the Linux platform and developed by Special Agents Kris Kendall and Jesse Kornblum of the U.S. Air Force Office of Special Investigations. (In accordance with 17 USC 105, this tool is not afforded any copyright protection because it is a work of the U.S. government.) The tool was inspired by, and designed to imitate the functionality of, the DOS program CarvThis, written by the Defense Computer Forensics Lab. Foremost enables forensic examiners to automatically recover files or partial files from a bit image (or the media itself) based on file header and footer types specified in a user-defined configuration file.

Foremost works by reading into memory a portion of the media or media image under examination. This portion is searched for a file header contained within the foremost configuration file. If a matching header is found, foremost writes the header and the data following the header to a file. Foremost will append data to the file until the footer is found (which signifies the end of the file) or until the file size limit listed in the configuration file is reached. Foremost will read additional data into memory from the media or media image if necessary to find the file footer or to satisfy the file size limit set within the configuration file. Using a file size limit serves as a means to stop foremost from adding data to a recovered file if the appropriate file footer is not found.

The foremost configuration file also allows the forensic examiner to customize the types of files that will be recovered and enables the use of wildcards for pattern matching. Customizing the configuration file and sources for file headers and footers will be covered later in the article.

As the capacity of storage media grows larger and larger, automated processing tools will become more necessary to supplement hand processing of forensic images. Although a hex editor can be used to recover files from a forensic image (and sometimes may be the best choice), the use of a hex editor can be tedious if a great many files must be recovered. Foremost is useful as a recovery tool when dealing with unfragmented files, because the tool in its current form will simply read data until a condition is met to cause it to stop extracting data. Newly formatted disks or partitions that have data written to them would be unfragmented until files are deleted and new files are written. (This type of situation is often found in a document or image archive.)

Foremost can be used to perform a cursory examination of media or media images for headers specified in the configuration file. When using this mode, foremost will only search the media; it will not extract any files it encounters. This technique is useful in situations where many drives are involved to determine whether a drive should be given a thorough examination. Foremost can recover files from images made by the Unix utility dd (as shown later in this article), as well as images created by commercial applications such as EnCase and Safeback.

Obtaining and Compiling Foremost

Foremost is available from SourceForge at:

<http://sourceforge.net/projects/foremost/>

The Foremost Open Source Forensic Tool

Ray Strubinger

(Reprinted from SysAdmin Magazine)

The source of the current version, 0.64, is less than 100,000 bytes in size. The program is written in C, and compilation is straightforward on a Linux system. Once you have obtained the archive, the following commands can be used to build the application:

```
$: tar xvfz foremost-0.64.tar.gz
$: cd foremost-0.64
$: make
```

You can install the tool to /usr/local/bin with the command:

```
$: make install
```

The tool can be uninstalled from the system with the command:

```
$: make uninstall
```

I have successfully compiled and executed foremost on several versions of Red Hat Linux (7.x-9.x). The documentation included with foremost indicates that some systems using older versions of glibc may produce harmless warnings that can be ignored regarding fseeko and ftello.

Using Foremost

Foremost opens bit images or devices in read-only mode, which is important for maintaining the forensic integrity and the validity of the investigation -- especially when the unfortunate situation arises whereby a forensic image of the media cannot be obtained. (Generally, forensic examiners only want to use the original media to produce a forensically sound image -- they do not want to conduct an examination on the original media to avoid the risk of contaminating the original in some way.)

If you are not a computer programmer or do not want to examine the foremost source code, a simple way to verify that foremost does not alter the image is to compute an MD5 checksum of the forensic image before and after running the tool on the image. If the checksums match, then the forensic image cannot have been altered. The image is read into memory in 10-MB chunks and any matching files are obtained from the image in memory, if possible, in order to increase processing speed by limiting the reads from the disk. Large files (those greater than 2 GB in size) are supported under Linux. As an additional safety precaution, foremost will only extract data to an empty directory.

Foremost requires a configuration file that contains, at a minimum, the header, the amount of data to extract for a file type (referred to as "size" in the configuration file), a "y" or "n" denoting the case sensitivity of the header and footer, and the file extension. Header and footer values can be specified in either hexadecimal or octal values and are decoded before use. A maximum of 100 file types may be specified in a single configuration file. Lines in the configuration file that begin with a pound sign (#) are treated as comments. The pound sign can be used at the beginning of a line containing a file type to prevent foremost from searching for that file type. An excerpt from a configuration file is shown in Figure 1.

In the example in Figure 1, the default foremost.conf configuration file is used. I encourage you to examine the configuration file and modify it to best suit the needs of your investigation. You may want to create custom configuration files that only contain files of one type, such as documents, graphical images, or binaries.

A great Web resource for finding file specifications is Wotsit's Format at:

The Foremost Open Source Forensic Tool

Ray Strubinger

(Reprinted from SysAdmin Magazine)

<http://www.wotsit.org>

Wotsit's Format contains information on hundreds of file formats and makes a nice complement to the foremost application by bringing together many file formats (which contain header and footer information) that can be used to populate the foremost configuration file. Users of the Wotsit site are also encouraged to submit information on file formats that are not listed on the site. A hex editor may also be used to obtain header and footer information from a given file type by opening the file and copying short hex strings that occur at the beginning and end of the file. These extracted strings would then be copied into the foremost configuration file. If you choose to use a hex editor, keep in mind that a machine utilizing little-endian architecture may require you to reverse the byte order of the header and footer to get an accurate representation.

Running foremost without any options will produce an error message, along with a brief explanation of the function of the program and the parameters that may be used with the application as seen in Figure 2. Running foremost with the -h option will also produce the list of parameters.

If the examiner creates an image directory, then copies the forensic image to be analyzed and the configuration file to that directory, the command **foremost -v -c foremost.conf ext2-home.dd** will produce output similar to that shown in Figure 3 (the output has been edited to save space). With the options shown in the command-line below, foremost will generate verbose output and use the foremost.conf configuration file found in the current directory against the image named ext2-home.dd. Since the output directory was not specified with the -o option, foremost defaults to using a directory called foremost-output for any files that it locates and recovers. This output directory will be created in the current working directory. As foremost recovers files, it assigns them numerical names starting at 00000000 and adds the relevant extension. Figure 3 shows output from a foremost run where the verbose (-v) option has been used.

While foremost executes, it produces an audit.txt file containing the command line options used, path information, and the name of the forensic image. The name foremost is assigned to the recovered file, then offset into the image where the file was found, and the file length is displayed. The column labeled "Interior" within the audit file denotes whether the recovered file was found off the start of a sector. An excerpt from the audit file produced on the above run is shown in Figure 4.

To recover entire files (assuming that the entire file is still intact), you should pay attention to the file sizes specified in the foremost.conf file. In this example, the audit.txt file shows that two recovered files reached the maximum size allowed by the configuration file. The default maximum size for a PDF file is 100000 bytes, and the default for an html file is 50000 bytes. Adjusting the file sizes in the foremost.conf file to support the recovery of larger PDF and html files and using the information in the "Found at Byte" column with the -s option, makes it easy to skip through the image and make another recovery attempt on the PDF and html file.

Conclusion

Although foremost has been used in various investigations, I am unaware at the time of writing of any court cases where the tool has been judicially validated or utilized in the gathering of submitted evidence. Bug reports are taken seriously by the tool's creators, due to the fact that it could be used to produce evidence. I have used foremost to recover data from forensic images from several file system types including ext2/3, Linux swap, UFS, JFS, NTFS, and various flavors of FAT including FAT 12, 16, and 32. I have also used foremost to extract data from single-disk elements in a multi-disk RAID-5 array and from partial bit images. Although no single tool can serve all needs because of the

The Foremost Open Source Forensic Tool

Ray Strubinger

(Reprinted from SysAdmin Magazine)

variations found in each situation, overall I have found foremost to be a useful addition to the growing collection of open source forensics tools.

References

Foremost Web Site -- <http://sourceforge.net/projects/foremost/>

Wotsit's Format (file specifications) -- <http://www.wotsit.org>