Using Extended File Information (EXIF) File Headers in Digital Evidence Analysis

Special Agent Paul Alvarez
Air Force Office of Special Investigations,
Computer Investigations and Operations

Abstract

An obstacle in any Child Pornography (CP) investigation is the investigator's ability to determine whether the pictures in question have been altered. Because of the court ruling in Ashcroft v. Free Speech, many agents are asked on the stand if they can prove the pictures they recovered were altered in any way. If the picture doesn't match any known CP hashes, then it can be very difficult to prove they are untouched. One way an investigator may be able determine if a picture is authentic is through extraction of metadata. In the case of digital pictures, they may contain EXIF headers that can help the investigator to verify the authenticity of a picture.

Introduction

In the Supreme Court case of John D. Ashcroft, United States Attorney General v. The Free Speech Coalition, the court ruled that modified pictures (pictures that were edited, morphed or created to look like child pornography) were protected under the First Amendment of the Constitution [1]. The court felt that because a real child was not harmed the pictures are protected by the freedom of speech. This may create additional hurdles for investigators. With the use of hash databases compiled by law enforcement, it is possible to find previously identified CP on digital media. But how can we verify the remaining pictures? What if the person under investigation was taking pictures of a victim not previously identified by law enforcement? One way to help verify the authenticity of a picture is through the use of metadata. It must be stressed that any information gleaned from metadata cannot stand on its own. Metadata is not strictly bound to a file, but can provide useful information if used in the proper context. Therefore, an investigator must evaluate all available information before assessing a picture's authenticity. The subject's computer expertise and tools residing on the system should be taken into account when making any conclusions.

Metadata in Digital Photography

Metadata is, quite simply, data about data. For example, a Microsoft Word document's metadata may contain the author's name and the dates the document was created/modified. Metadata may contain useful information for an investigator. Specifically, digital camera pictures may contain an Extended File Information (EXIF) header, which saves information about the camera that took the picture. The EXIF format was created by the Japan Electronic Industry Development Association and is referenced as the preferred image format for digital cameras in ISO 12234-1 [2]. Many

digital camera manufacturers, such as Canon, Sony and Kodak implement the use of EXIF headers. This header is stored in an "application segment" of a JPEG file, or as privately defined tags in a TIFF file. This means that the resulting JPEG or TIFF is still in a standard format readable by applications that are ignorant of EXIF information [3]. Below is a typical EXIF header (in human readable format):

File name: 0805-153933.jpg
File size: 463023 bytes
File date: 2001:08:12 21:02:04
Camera make: Canon
Camera model: Canon PowerShot S100
Date/Time: 2001:08:05 15:39:33
Resolution: 1600 x 1200
Flash used: No
Focal length: 5.4mm (35mm equivalent: 36mm)
CCD Width: 5.23mm
Exposure time: 0.100 s (1/10)
Aperture: f/2.8
Focus Dist. : 1.18m
Metering Mode: center weight
Jpeg process: Baseline

Value in Retrieving EXIF Headers

By reviewing EXIF headers, some valuable information can be recovered. For example, the one above shows two dates. The first is the file creation date/time. The other is the date/time the picture was taken. The date/time the picture was taken will not change, even if the file is copied to another medium. The EXIF header also shows the camera make and model. The EXIF header is placed in the file by the camera that took the picture. If the file is modified with picture editing software, the EXIF header will be lost. Therefore, if a file contains EXIF information, then it is possible that the picture is unaltered. However, not all digital cameras use EXIF headers; pictures taken with such cameras do not have EXIF data.

Retrieval of EXIF Headers

Although it is possible to retrieve EXIF headers by looking at each picture in a disk editor, a considerable amount of time is required to translate the hex codes into human readable format. Fortunately, there are other ways. An open-source program called *jhead* allows the retrieval of EXIF headers from jpg files [4]. Only 88K in size, this small utility has some useful features:

- Outputs results in space delimited format for use in spreadsheets;
- Ignores files with no EXIF headers (with correct command line switch);
- Allows searches for specific camera models;
- Allows for recursive directory searches and wildcards;
- Runs on Windows, Linux and Mac OS-X..

To test the validity of the results returned by jhead, sample pictures with EXIF headers were downloaded from EXIF.org. MD5 hashes of all pictures were obtained before and after running jhead and all the hashes matched. After altering a picture in Microsoft Paint and running the picture through jhead, all EXIF header information was lost. Figure 1 shows a sample picture used to test the program and Figure 2 shows the results obtained from jhead.

Figure 1: Sample Picture Taken with Kodak EasyShare LS443

```
File name: esc_ekn026347_00002_obj0020.jpg
File size: 658027 bytes
File date: 2003:05:21 07:03:34
Camera make: EASTMAN KODAK COMPANY
Camera model: KODAK LS443 ZOOM DIGITAL CAMERA
Date/Time: 2002:01:06 10:20:22
Resolution: 2448 x 1632
Flash used: No
Focal length: 21.2mm
Exposure time: 0.011 s  (1/90)
Aperture: f/4.7
Exposure: program (auto)
Jpeg process: Baseline
```

Figure 2: jhead Output

Caveats

Adobe Photoshop 6.0 and higher attempt to preserve EXIF header data during editing. According to Kodak, Photoshop preserves EXIF data in files when written to file formats that support arbitrary extra data (PSD TIFF, EPS and PDF), and rewritten as EXIF markers when the file is saved as a JPEG [5]. Photoshop inserts version information into the comment, so it should be easy to spot edited files. Below is jhead output of a file edited by Adobe Photoshop 5.0 (note that this version of Photoshop removed all EXIF header information).

File name: baboon2.jpg
File size: 38113 bytes
File date: 2004:01:30 11:18:58
Resolution: 256 x 173
Jpeg process: Baseline
Comment: File written by Adobe Photoshop¿ 5.0

Figure 3: jhead Output of Picture Edited with Photoshop 5.0

jhead is just one of many EXIF metadata viewers. A list of some Windows and Mac EXIF viewers can be found at http://graphicssoft.about.com/cs/exifsoftware.

Conclusion

The problems faced by law enforcement personnel when analyzing Child Pornography are significant. The availability of CP hashes alleviates the problem somewhat, but other techniques are needed when encountering new pictures. Extracting EXIF headers from JPEG files may help investigators to distinguish untouched digital pictures from those that are altered to look like CP. Future versions of open-source and commercial forensic software should attempt to incorporate this functionality into their software. This would make it easier than extracting the pictures in question and analyzing them with a stand-alone tool.

References

[1] The American Center for Law and Justice (2002), Supreme Court Decision in Ashcroft v. Free Speech Coalition to Overturn Virtual Child Pornography Law http://www.aclj.org/resources/pornography/supctdec.asp

[2] Digital Photography Review (2003), Glossary of digital photography terms http://www.dpreview.com/learn/Glossary/Camera_System/EXIF_01.htm

[3] Digital Camera EXIF Library (2003) http://home.cfl.rr.com/genecash/digital_camera.html

[4] EXIF Jpeg camera setting parser and thumbnail remover (2003) http://www.sentex.net/~mwandel/jhead/

[5] Kodak Research and Development (2001), Metadata in Adobe Photoshop 6.0 http://www.kodak.com/US/en/corp/researchDevelopment/technologyFeatures/metadataAdobe.shtml

About the Author

Special Agent Paul Alvarez is a Computer Crime Investigator for the Air Force Office of Special Investigations. He received his Bachelor of Science degree in Computer and Information Science from Troy State University Montgomery, Alabama in 2001. He is currently pursuing a Master's degree in Network Security from Capitol College in Laurel, Maryland. Special Agent Alvarez is responsible for Air Force criminal and counterintelligence computer investigations in the Western Pacific area. Contact: paul_at_thealvarez.net