Detect and prevent file tampering in multimedia files

A step by step guide using free and open-source software

By Doug Carner, CPP/CHS-III

Electronic files are vulnerable to tampering and corruption. Undetected, these changes can alter the meaning and value of critical evidence. By implementing a few simple steps, you can ensure that everyone is working from the exact same set of facts, and be able to prove if a file was altered prior to arriving into your care.

What you will learn:

- Version control through hash signatures
- How to access, interpret and alter video metadata
- Tamper detection using Video Error Level Analysis
- Using a hex editor to authenticate a file
- Using playback software to determine a file's origin
- Methods to embed or extract hidden parasite file data

What you should know:

- Familiarity with computer audio, video and image files
- Familiarity with installing PC based software from a website

INTRODUCTION

In your professional career, you are called upon to sort through a wide array of electronic file evidence, including discovery material originating from external sources beyond your control. These files are typically shared by disc or download using the honor system. If any file becomes accidentally or willfully corrupted, the altered version could go unnoticed. From there, the incorrect version could be shared with attorneys, the courts, clients, and various experts.

If the modification creates a new perceived reality, or an attempted falsity, such unwitting modification could compromise the entire case. If the modification inserts illegal content, you could become an unwitting distributor. Even if the only change is an unintentional reduction in quality, it can leave evidence open to interpretation. Such changes can be difficult or impossible to detect through a visual inspection. Fortunately, even a single bit change within a large file can be detected through a few simple software tools.

AUTHENTICATION

It can be challenging to trace changes to their source due to the number of people that have access to a given file. Achieving version control requires the ability to detect file modifications, even when they occur prior to your involvement. With this capability, recipients can determine if any files were altered prior to receipt. Although a full authentication analysis requires years of experience and sophisticated software, some basic tests can be quickly performed using free software and minimal guidance.

HASH VALUES

Upon receipt or creation of a file, you can quantify the current file version in a manner that can be replicated by others, so as to ensure that they are working with a file indistinguishable in every way from the version that you have. This is accomplished by calculating a unique identifying value for each file, and then providing a means for all other file recipients to do the same, so these values can be matched as proof of file authenticity.

A Hash value is an electronic identifier constructed solely from the file's contents and structure. The most common Hash is the 5th generation of the Message Digest algorithm, commonly known as MD5. There are dozens of free programs to calculate MD5 values and, regardless of the program used, the resulting MD5 value will always match for exact copies of the same file.

Two easy-to-use free MD5 programs are Digestit (*http://colonywest.us/digestit*) and Checksum (*http://corz.org/windows/software/checksum*). They support left click or drag-n-drop, and require less time to process than reading this paragraph. Other offerings, like Microsoft's™ File Checksum Integrity Verifier (*www.microsoft.com/en-us/download/details.aspx?id=11533*) are also free, but can be cumbersome to use.

As soon as you receive or generate an original electronic file or folder, use your preferred hash program to calculate its MD5 value (see Figure 1). You can then include a list of all the relevant MD5 values anytime you share those files. This list should be shared as read-only, and you can even create a hash value for the file listing the other Hash values.

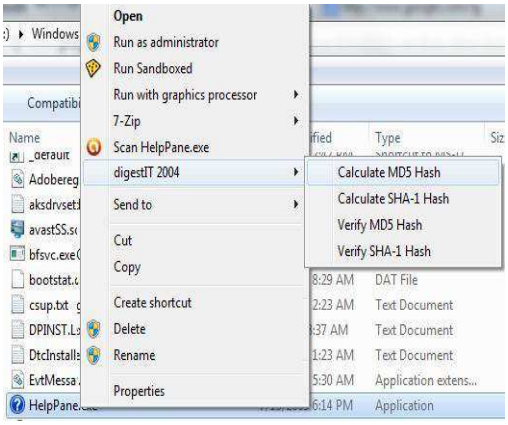

Figure 1. Left click to calculate a file's MD5

At any time, a file recipient can generate their own list of MD5 values and, if they match, be reasonably confident that their file versions are identical and indistinguishable from the file versions under your control. Any changes, even the simple act of opening and resaving a file without any content changes, will alter the calculated MD5 value. Copying, downloading and sharing a file will not alter a files metadata or MD5 value. Best of all, MD5 values work on media files, presentations, documents, DVDs, ZIP files and anything else that can be shared electronically.

MD5 hash values are the easiest method to provide version control of all your electronic files. The MD5 Hash value is so unique that you have significantly better odds of winning the Powerball lottery four times in a row, than altering a file to achieve the same MD5 value. However, because the MD5 algorithm is open-source, it has been reverse engineered and compromised under very controlled conditions. Although there are newer algorithms (example: MD6), MD5 remain the premiere balance of security and popular cross-platform support.

AUDIO AUTHENTICATION

Audacity and Audition are extremely common audio editing programs. Each program allows the user to isolate a given frequency for deeper analysis. Editing may be detected as disruptions in the cyclical pulse of a given frequency, shifts in the bit rate (see Figure 2), or shifts in the DC component of that signal. All of these tests require specialized training to interpret the significance of such anomalies.

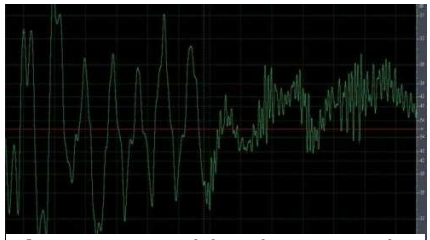

Figure 2. Sudden increase in audio bit rate

One solution is to open the file using any application that has a Spectrographic view, and then zoom in on areas of concern in search of data anomalies. If something looks suspicious, you will have cause for reaching out to a qualified expert.

IMAGE AUTHENTICATION

Most image viewing programs include a “Properties” or “Info” menu option to display a picture’s metadata. For images in the JPG format, additional metadata fields can be viewed using JpegSnoop (*http://sourceforge.net/projects/jpegsnoop*). JpegSnoop provides detailed information beyond the scope of this article (example: quantization tables) that requires training to interpret. However, JpegSnoop ends the report with summary information, including an interpretation if tampering is suspected (see Figure 3).

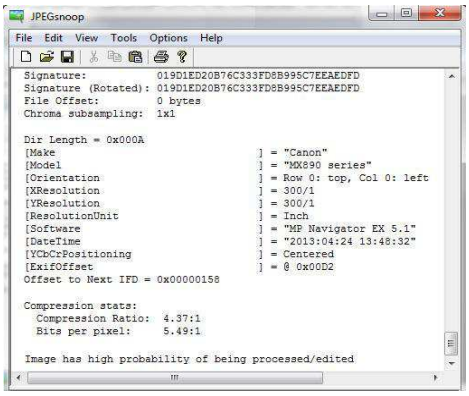

Figure 3. JpegSnoop’s detection results report

Figure 4. Analogexis can alter any Metadata field

This is especially important since JPG images are quite prone to tampering. It should also be noted that free open-source software, like Analogexis (*http://analogexif.sourceforge.net*), can untraceably alter every metadata field of JPG and TIFF images (see Figure 4). Similar programs exist for other image formats, including a few raw and proprietary file types.

Often times you can compare a file's metadata and/or header values to published tables (example: a Google search) to determine if that information matches the expected information for that file type. By comparing the file's header and/or metadata to known facts, you can quickly separate fact from fiction.

VIDEO AUTHENTICATION

Compression and decompression (CoDec) instructions guide a computer on how to store and reconstruct a compacted video. Modern video CoDecs conserve considerable file space by suppressing high-frequency data and merging color details in ways that are designed to be minimally perceptive to the human eye. Because these changes are lossy, they result in quantization data errors.

Each successive resave further deviates the video from the original version, but by decreasing amounts. If an object is added to, or relocated within, an already compressed video, it will be at the earlier stages of this process. When an altered video is resaved into a lossy format, the recent viewable manipulations will incur greater quantization changes than the remainder of the video content.

If you recompress a received video file, and then subtract that version from the file you received, the resulting video will only display the data errors between those versions. The resulting brightness directly relates the intensity of the data errors, with areas of greater contrast undergoing the greatest changes and resulting in greater brightness. If any pixels are disproportionally bright compared to other areas of comparable contrast, then those pixel locations denote suspected areas of tampering on the originally received video.

This test is called Video Error Level Analysis (VELA). VELA can be performed on any video using several video editing programs, including the popular open source VirtualDub program. VELA is based upon the concept of Error Level Analysis (*http://fotoforensics.com/*) which is used to authenticate single images. VELA’s advantage is that it exploits the motion vectors that are absent in single images. Step-by-step instructions and a guide to interpreting the results can be found at the author’s website (*http://ForensicProtection.com/VELA.html*).

DUBBED VIDEO TEST

Open any video with software that allows the user to advance frame-by-frame, and advance through the first fifty continuous frames that show objects or people in motion. If the frame numbers advance, but the viewed scene ever remains unchanged, then that video includes duplicate identical frames.

VLC (*http://sourceforge.net/projects/vlc*) is a popular open-source video player that can decode most formats without requiring the installation of other software. Once a video has been opened and paused in VLC, the "e" button on the computer keyboard will advance the video one frame at a time, making it easy to perform the dubbed video test.

Duplicate frames are a strong indicator that the video was recorded from a slower playing version. This is common for video DVD files which play at 29.97 frames per second for countries following the NTSC standard, like the USA. Duplicate frames almost always indicate that you do not have an accurate copy of the original recording.

REMOTE SCREEN CAPTURE VIDEO TEST

The more recent video CoDecs are based upon the h.264 standard which segments a video into a mosaic of small squares called "slices", typically only eight pixels (screen dots) wide and eight pixels tall. It is common for DVRs to skip slices during remote playback when the internet connection cannot keep up with the video's natural-speed playback data rate.

Using the software from the "Dubbed video test", if frame-by-frame playback depicts any frames with slices that are abnormally bright or dark for just that one frame, then the file likely originated from remote viewing capture software.

LOCAL SCREEN CAPTURE – CONVERSION VIDEO TEST

On-screen information is the last thing added to a video before it is saved onto, or played back by, a recording device. Use frame-by-frame playback to see if any video frame shows two different video frames or time stamps blended together. For example, a sequenced video is when one video input receives its signal from a cycling list of camera feeds.

If the video was converted at the incorrect speed (example: conversion software) or captured at a speed different than the video card's refresh rate (example: screen capturing software) the resulting video can create blended frames (see Figure 5). The resulting video is an inaccurate and lower quality representation of the original recording.

Figure 5. Two sequenced frames on either side of a blended frame

HIDDEN DATA

Every file has hidden data, even if the only purpose of the data is to denote the type of file it is or how it is to be played. Mining the hidden data can provide tremendous insight into the file's true origin and authenticity.

METADATA

It is well known that a computer's operating system can display a file's last Modified, last Accessed and first Created dates, collectively known as the MAC dates. In Windows, this information is accessed by right clicking on a file, and then left clicking on the Properties option. However, MAC information is saved by the computer's operating system and can become inaccurate from several causes including user error, virus, file transfer or free programs (see Figure 6) like FileDateCH (*http://www.nirsoft.net/utils/filedatech.html*). A more reliable date source is reading the metadata located within the target file.

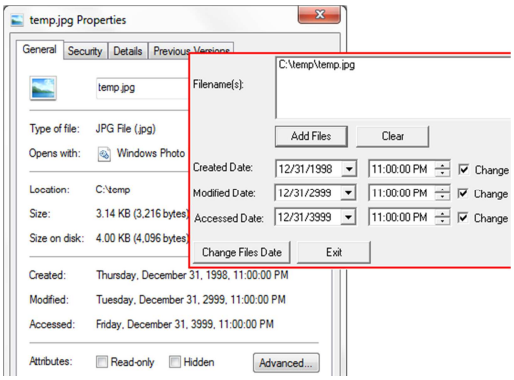

Figure 6. FileDateCH changing MAC dates

Some metadata can be read with the tools you already have. For example, with Windows you can right click on the file, left click on Properties, and then left click on the Details tab. You may also be able to access certain metadata by simply choosing the Properties option located under the File menu of whatever software you are using to open the file.

In the case of a media file, MediaInfo (*http://sourceforge.net/projects/mediainfo*) lets you review every metadata field of most audio and video files, even the hidden fields. If these details deviate from the known facts of the case, they become a strong indication of file tampering.

It is increasingly common for file metadata to include the Global Positioning System (GPS) coordinates denoting the actual location where the recording was created. You can enter these GPS values directly into a Google™ search box to translate them into an approximate street address. If the metadata includes a field labeled as "@xyz" these are the GPS coordinates in decimal format. The third GPS coordinate, the "z" value, is the above sea level altitude denoted in meters.

Depending on the GPS reception and capabilities of the recording unit, the GPS coordinates can pinpoint a specific room of a high-rise building. This accuracy results from advances in GPS technology that incorporates data from satellites and cell phone towers (aka Enhanced GPS). GPS data can make or break a case. For example, I recently had a case where a file's GPS coordinates matched the address of someone with video editing skills, instead of the location depicted in the video.

The file's metadata may include the date-time of the file's creation and who was the last person to access it. Metadata may also include details about the software or equipment used to capture the recording, including the user settings in effect at the time of file saving. If these details do not match the case facts, then you may have strong evidence of after-the-fact file tampering. For example, if a video's metadata lists the file as being in a Windows format, but the event was captured on an iPhone, then you can be confident that you are not looking at the original recording as saved by that iPhone. Metadata inconsistencies should always initiate an extremely thorough series of authentication tests.

HEX EDITORS

Hex editors allow the user to examine and modify any bit or byte of file data. Changing this inform will alter the file's Hash value, but reading this data can provide deep insight into any programs that previously affected the underlying file data. Most of the identifying information is located at the beginning (header) and end (footer) of a file (see Figure 7). One of many free hex editors, and my personal favorite is BeHexEditor (*http://sourceforge.net/projects/hexbox*).

```
 0: 46 56 45 52 04 00 00 00-2E 01 00 00 45 56 54 43  FVER........EVTC
10: AC 07 00 00 3C 44 43 5F-43 4F 4E 46 49 47 3E 3C  ....<DC_CONFIG><
20: 41 55 44 49 4F 3E 3C 43-4F 4D 50 52 45 53 53 45  AUDIO><COMPRESSE
30: 44 5F 52 41 54 45 3E 34-30 30 30 3C 2F 43 4F 4D  D_RATE>4000</COM
40: 50 52 45 53 53 45 44 5F-52 41 54 45 3E 3C 53 41  PRESSED_RATE><SA
50: 4D 50 4C 45 5F 52 41 54-45 3E 38 30 30 30 3C 2F  MPLE_RATE>8000</
60: 53 41 4D 50 4C 45 5F 52-41 54 45 3E 3C 2F 41 55  SAMPLE_RATE></AU
70: 44 49 4F 3E 3C 45 56 45-4E 54 3E 3C 41 55 44 49  DIO><EVENT><AUDI
80: 4F 5F 45 4E 41 42 4C 45-3E 31 3C 2F 41 55 44 49  O_ENABLE>1</AUDI
90: 4F 5F 45 4E 41 42 4C 45-3E 3C 46 52 4F 4E 54 5F  O_ENABLE><FRONT_
A0: 45 4E 41 42 4C 45 3E 31-3C 2F 46 52 4F 4E 54 5F  ENABLE>1</FRONT_
B0: 45 4E 41 42 4C 45 3E 3C-50 4F 53 54 54 52 49 47  ENABLE><POSTTRIG
```

Figure 7. File header as read by a hex editor

For example, the numeric value of each set of four characters (bytes), of the first 52 bytes of an AVI file, detail how the file is to be played (example: size, speed, audio type, offset, etc…). In this example, the value of bytes 32 through 35 denotes the videos width (in pixels) and the next four denote the height. The offset value allows the insertion of additional metadata which may include information about the creating software, creation time, geographic location, etc…

STEGANOGRAPHY

Steganography is the art of hiding parasite data (example: messages, passwords, illegal information, etc…) within a host file, typically as a text inside an image file. The hidden data cannot be detected using a visual inspection, metadata tools or hex editors. There are several Steganography programs and data hidden with one program is nearly invisible to the others. The most popular is the open-source application OpenPuff (*http://embeddedSW.net*).

Unless protected through Hash version control, Steganography would allow a file to contain illegal or personal information without the knowledge of the person distributing the file. As with any content change, adding or changing content with a Steganography program will change that file's MD5 Hash value.

BIBLIOGRAPHY

- Video codecs and decompressors, retrieved 2013 from *http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3232547/*
- Dr. Neal Krawetz, "A Picture's Worth...", 2007 Black Hat conference, Caesars Palace - Las Vegas, retrieved 08/01/07, *http://blackhat.com/html/bh-usa-07/bh-usa-07-speakers.html#Krawetz*

REFERENCES

- Photoshop CS3 for Forensic Professionals, George Reis, ISBN 978-0470114544
- Basic Television and Video Systems (6th edition), Bernard Grob, ISBN 978-0028004372
- How Video Works (2nd edition), Marcus Weise and Diana Weynand, ISBN 978-0240809335
- Best Practices for the Retrieval of Video Evidence from Digital CCTV Systems, *http://www.rcfl.gov/webinar/Video_Evidence_From_CCTV_system_flipbook.pdf*
- CCTV Networking & Digital Technology (2nd edition), Vlado Damjanovski, ISBN 978-0750678001

IN SUMMARY

It is an unavoidable reality that electronic files are vulnerable to tampering and corruption. Undetected, these changes can alter the meaning and value of critical evidence. From version control to authentication, the above steps can help ensure that everyone is working from an authentic and identical set of facts.

ABOUT THE AUTHOR

I am an audio - video enhancement and authentication expert, and am board certified by the American College of Forensic Examiners and the American Society for Industrial Security. Over the last twenty years I have processed evidence in over a thousand cases worldwide including the George Zimmerman - Trayvon Martin shooting, and the Mathew Clark beating. I have pioneered industry innovations and routinely donate my time to innocence projects, indigent clients and as an industry educator. I am the founder and lead analyst of Forensic Protection, a world class lab that has received both prosecution and defense praise for detailed work leading to an exceptionally high rate of pretrial case dismissal or settlement.

A complete CV is available at *http://ForensicProtection.com*