Hash Value Tool (Or "Digital Fingerprint") Increasingly Noted In Cases Involving Electronic Evidence

Tue, 02/19/2013 Editor's blog Printer-friendly version

Over the past few years, an increasing number of cases have discussed the role of "hash values" (mathematical algorithms) used to identify electronic images, records, files or other evidence; hash values (commonly referred to as "digital fingerprints") have unique identification capabilities that have a high degree of accuracy to confirm whether two records or files are a match or are dissimilar, such as in *United States v. Cartier*, 543 F.3d 442, 444 (8th Cir. 2008) (No. 07-3222) ("Every digital image or file has a hash value, which is a string of numbers and letters that serves to identify the image or file.") (footnote omitted)

As we have previously noted, "hash" values are an important tool used to identify and authenticate digital evidence. *See* Using "Hash" Values In Handling Electronic Evidence; *see also* Hash Values Used To Confirm Seized Video Clips And Images; Federal Judicial Center, *Managing Discovery of Electronic Information: A Pocket Guide for Judges*, at 24 (2007) ("'Hashing' is used to guarantee the authenticity of an original data set and can be used as a digital equivalent of the Bates stamp used in paper document production.") (quoted in *Lorraine v. Markel American Ins. Co.*, 241 F.R.D. 534, 546-47 & n.23 (D. Md. 2007) ("Hash values can be inserted into original electronic documents when they are created to provide them with distinctive characteristics that will permit their authentication under Rule 901(b)(4).")).

A recent review of cases referring to the use of hash values highlights the growing acceptance of this tool on forensic issues involving electronic evidence. As summarized below, hash values are commonly referred to as "digital fingerprints" or "digital DNA" and have been described as having more than a 99 percent level of accuracy to confirm two files or records match.

Using Hash Values

A "hash value" is an algorithm that can be used to confirm that two digital files or objects are either the same or different. As the Fourth Circuit recently summarized:

> A "hash value" is an alphanumeric string that serves to identify an individual digital file as a kind of "digital fingerprint." Although it may be possible for two digital files to have hash values that "collide," or overlap, it is unlikely that the values of two dissimilar images will do so. *United States v. Cartier*, 543 F.3d 442, 446 (8th Cir. 2008) (No. 07-3222). In the present case, the district court found that files with the same hash value have a 99.99 percent probability of being identical.

United States v. Wellman, 663 F.3d 224, 226 n.2 (4th Cir. 2011) (No. 10-4689) (identifying suspected child pornography by hash values), *cert denied*, 132 S.Ct. 1945, 182 L.Ed.2d 800 (2012); *see also United States v. Farlow*, 681 F.3d 15, 19 n.2 (1st Cir. 2012) (No. 11-1975) (defining hash value as "a short, unique set of numbers and letters produced by running the complex strings of data that make up a computer file through a mathematical algorithm"); *United States v. Henderson*, 595 F.3d 1198, 1199 n.2 (10th Cir. 2010) (No. 09-8015) ("A SHA value of a computer file is, so far as science can ascertain presently, unique. No two computer files with different content have ever had the same SHA value.") (quoting *United States v Klynsma*, No. CR 08-50145-RHB, 2009 WL 3147790, at *6 (D.S.D. Sept. 29, 2009)).

Generally, there are two common types of hash values that are used:

- **Secure Hash Algorithm Version 1 (or SHA-1):** As some cases have noted, "SHA-1 stands for Secure Hash Algorithm Version 1 — a digital fingerprint of a computer file. It is a 32-digit number that is calculated for a file and unique to it." *United States v. Glassgow*,

EXHIBIT A

682 F.3d 1107, 1110 n.2 (8th Cir. 2012) (No. 11-2611); *see also* ***United States v. Miknevich***, 638 F.3d 178, 181 n.1 (3rd Cir. 2011) (No. 09-3059) ("A SHA1 (or SHA-1) value is a mathematical algorithm that stands for Secured Hash Algorithim used to compute a condensed representation of a message or data file.").

- **Message-Digest Algorithm 5 (MD5):** "An MD5 hash value is a unique alphanumeric representation of the data, a sort of 'fingerprint' or 'digital DNA.'" ***United States v. Crist***, 627 F. Supp. 2d 575, 578, 585 (MDPA 2008) (No. 07-cr-211)). An MD5 also generates a unique, but shorter alphanumeric value than the SHA-1 for a particular file or object.

Digital Fingerprints and Digital DNA

Hash values have unique identification features. Recognizing this role, a number of cases refer to hash value determinations as "digital fingerprints," including the following cases:

- ***United States v. Chiaradio***, 684 F.3d 265, 271 (1st Cir. 2012) (No. 11-1290) (referring to hash values as "essentially, the digital fingerprint" used to compare files)
- ***United States v. Cunningham***, 694 F.3d 372, 376 n.3 (3rd Cir. 2012) (No. 10-4021) ("Each hash value 'is an alphanumeric string that serves to identify an individual digital file as a kind of "digital fingerprint."'") (quoting *Wellman*, 663 F.3d at n.2)
- ***United States v. Farlow***, 681 F.3d 15, 19 (1st Cir. 2012) (No. 11-1975) (defendant suggesting how investigators could "have employed a limited search" by "using the image's 'hash value' — a sort of digital fingerprint tied not only to a specific file but also to that file's precise location on a computer")
- ***United States v. Richardson***, 607 F.3d 357, 363 (4th Cir. 2010) (No. 09-4072) (describing how the AOL Image Detection and Filtering Program "recognizes and compares the digital 'fingerprint' (known as a 'hash value') of a given file attached to a subscriber's email with the digital 'fingerprint' of a file that AOL previously identified as containing an image depicting child pornography")
- *See also* ***United States v. Miknevich***, 638 F.3d 178, 181 n.1 (3rd Cir. 2011) (No. 09-3059) (noting how a SHA1 mathematical algorithm "can act like a fingerprint")

As another means of describing this identification role, some cases have also referred to hash values as a form of "digital DNA":

- ***United States v. Crist***, 627 F. Supp. 2d 575, 578, 585 (MDPA 2008) (No. 07-cr-211) ("An MD5 hash value is a unique alphanumeric representation of the data, a sort of 'fingerprint' or 'digital DNA.'") ("By subjecting the entire computer to a hash value analysis — every file, internet history, picture, and 'buddy list' became available for Government review" and the "examination constitutes a search.") (granting motion to suppress warrantless search of computer which ultimately had been provided to law enforcement after the defendant failed to pay his rent)
- *United States v Beatty*, No. 1:08–cr–51–SJM, 2009 WL 5220643, *1 n.5 (WDPA 2009) (in denying motion to suppress evidence seized from the defendant's computer, noting agent's affidavit described "the SHA1 'digital fingerprint' as "more unique to a data file than DNA is to the human body"), *aff'd*, **437 Fed.Appx. 185** (3rd Cir. 2011) (No. 10-3634)
- *United States v Wellman*, No. CRIM A 08CR00043, 2009 WL 37184 (SDWVA 2009) (noting investigator described "a hash value or algorithm is '[a] digital fingerprint or a DNA of a file'"), *aff'd*, **663 F.3d 224** (4th Cir. 2011), *cert denied*, 132 S.Ct. 1945, 182 L.Ed.2d 800 (2012)

Degree Of Accuracy

Many of the cases have noted the high degree of accuracy of hash values. In fact, few other evidence matches are as precise. Hash values have been said to be more precise than a match for DNA evidence. ***State v. Mahan***, 2011 Ohio 5154, n.2 (Court of Appeals, 8th Appellate Dist. Ohio 2011) (noting investigator testimony that "that SHA1 values are accurate in identifying a file to the 160th degree, which is 'better than DNA'").

The following cases involved evidence suggesting the accuracy of a hash value match exceeds 99 percent:

- ***United States v. Glassgow***, 682 F.3d 1107, 1110 n.2 (8th Cir. 2012) (No. 11-2611) (noting "there was a 99.9999% probability that exhibit 1 contained the same video clips that Glassgow possessed")

- ***State v. Mahan***, 2011 Ohio 5154, n.2 (Court of Appeals, 8th Appellate Dist. Ohio 2011) ("There is a certainty exceeding 99.99 percent that two or more files with the same SHA1 value are identical copies of the same file regardless of the file name.")

- *United States Nelson*, No. CR. 09-40130-01-KES (DSD July 12, 2010) ("When two files have the same hash value, there is a 99.99 percent chance that they are the same file.")

- *See also* ***United States v. Cartier***, 543 F.3d 442, 446 (8th Cir. 2008) (No. 07-3222) (in challenge to probable cause supporting search warrant, rejecting argument "that it is possible for two digital files to have hash values that collide or overlap")

Theoretically, it is possible for two different files to have the same hash value (referred to as a collision). But this theoretical possibility has yet to be demonstrated in the real world and is extremely unlikely. For the MD5 hash value, the likelihood is 1 in 340 billion billion billion billion. *See, e g* , Richard P. Salgado, **"Fourth Amendment Search And The Power Of The Hash,"** 119 HARV. L. REV. F. 38, 39 n.6 (2006) ("The range of values generated from commonly used hash algorithms is huge. For example, the prolific algorithm MD-5 can generate more than 340,000,000,000,000,000,000,000,000,000,000,000,000 (that's 340 billion, billion, billion, billion) possible values. The widely used SHA-1 algorithm generates a range of values over four billion times larger than that. Thus, although there is a finite number of possible hash values and an infinite number of possible data inputs, the odds of a collision are infinitesimally small."); *see generally* **Data Validation Using The MD5 Hash** ("There are actually 3.402 x 10^38 or 340 billion billion billion billion or a little more than 1/3 of a googol possibilities. When you consider that most people have never seen a million of anything the actual number becomes really difficult to conceptualize."); **HashCheck Shell Extension - FAQ** ("For 128-bit checksums (MD4, MD5), the probability [of a collision] is an unfathomably small 1 in 340 billion billion billion billion, and for SHA-1, it is even smaller.").

Generally, courts have rejected questions about the authentication or admissibility of evidence based on remote possibilities unless there is an articulable probability that the validity of the evidence should be doubted. *See, e g* , ***Cartier***, 543 F.3d at 446 (while theoretically "hash values could collide ," accepting government view "that no two dissimilar files will have the same hash value"); *see also* ***United States v. Safavian***, 435 F.Supp.2d 36, 41 (D.D.C. 2006) ("The possibility of alteration does not and cannot be the basis for excluding e-mails as unidentified or unauthenticated as a matter of course, any more than it can be the rationale for excluding paper documents (and copies of those documents).... Absent specific evidence showing alteration, however, the Court will not exclude any embedded e-mails because of the mere possibility that it can be done."), *rev'd on other grounds*, **528 F.3d 957** (D.C. Cir. 2008).

Identification Of Suspected Child Pornography Images

Once a hash value is obtained for a particular file, record or image, it can be used to confirm or locate other matches. In this manner, hash values are commonly used to identify suspected child pornography images. A known library of child pornography images can be used to determine whether suspected child pornography images are used or possessed. If a match in hash values between the known and suspected images is confirmed, law enforcement has used this information in support of a search warrant. *See, e g* , ***United States v. Brown***, 701 F.3d 120, 122 (4th Cir. 2012) (No. 11-5048) (hash values of downloaded files used to obtain search warrant in child pornography investigation); ***Cunningham***, 694 F.3d at 376 (hash values were used to identify child pornography images and used to show probable cause to seize the defendant's computer); ***Chiaradio***, 684 F.3d at 271 (hash values used in an "enhanced peer-to-peer software" to "compare the hash value (essentially, the digital fingerprint) of an available file with the hash values of confirmed videos and images of child pornography"; information was used to obtain a search warrant to seize the defendant's computer); *United States v Cartier*, 543 F.3d 442, 446 (8th Cir. 2008) (No. 07-3222) (hash values were used to identify child pornography images and used to show probable cause to seize the defendant's computer).

A few years ago, there were not many cases noting the application and use of hash values. As this review shows, the acceptance and use of this tool for electronic evidence has become more common and widely applied.

---

developments in the rules of evidence."

- **Inside Track - State Bar of Wisconsin** - *"The Federal Evidence Review blog is another place that attempts to consolidate jury instructions, both according to circuit and subject matter."*
- National Institute of Military Justice BLOG-CAAFLOG - ***"The ever excellent Federal Evidence Review** has this nice summary and analysis"* of White v. Illinois; Courtesy of ***FederalEvidence.com** **"here is their list** of potential significant evidence issues affecting criminal cases this coming year."*
- **Law Librarian Blog** - *"Check out the very good Federal Evidence Blog"*
- **Evidence Prof Blog** - *Notes that the Federal Evidence Review "nicely summarized" the implications of FRE 502*
- **Internet Legal Research Weekly / Inter Alia: an internet legal weblog** - Noted as a *Blawg of the Week (Sept. 9, 2008)*
- **Indiana Civil & Business Lawyer** - *The Federal Evidence Review Blog "actually is very helpful for those times I get into federal court"*
- **EDD: Issues, Law, and Solutions** - *Notes that the Federal Evidence Review assessed reasons for lack of FRE 502 waiver provision*
- **Montana Personal Injury & Civil Litigation Blog** - *On FRE 502 the "Federal Evidence Review has posted an outstanding analysis"*
- **Drug and Device Law: Interesting Stuff On The Web** - Notes the Federal Evidence Blog's coverage of FRE 502
- Hofstra School of Law, Deane Law Library, **Virtual Library Cat's Eye View** - - Noting that on newly adopted FRE 502, *"For the text of the rule, discussion/analysis and important links such as the "Statement of Congressional Intent", check out the Federal Evidence Review."*
- **Full List**

Share this post with colleagues on:

For More On Related Issues:

- Rule 901(a) - General Provision
- Rule 901(b)(4) - Distinctive Characteristics And The Like
- Rule 901 Requirement of Authentication or Identification
- Electronic Evidence
- Hash Values
- Internet Evidence

Comments

Post new comment

Subject:

Comment: *

- Web page addresses and e-mail addresses turn into links automatically.
- Lines and paragraphs break automatically.

More information about formatting options

Preview Save