Scientific Working Group on Digital Evidence

SWGDE Position on the Use of MD5 and SHA1 Hash Algorithms in Digital and Multimedia Forensics

The version of this document is in draft form and is being provided for comment by all interested parties for a minimum period of 60 days.

SWGDE encourages stakeholder participation in the preparation of documents. Suggestions for modifications are welcome and must be forwarded to the Secretary in writing at secretary@swgde.org. The following information is required as a part of the response:

a) Submitter's name
b) Affiliation (agency/organization)
c) Address
d) Telephone number and email address
e) Document title and version number
f) Change from (note document section number)
g) Change to (provide suggested text where appropriate; comments not including suggested text will not be considered)
h) Basis for change

Disclaimer:

As a condition to the use of this document and the information contained therein, the SWGDE requests notification by e-mail before or contemporaneous to the introduction of this document, or any portion thereof, as a marked exhibit offered for or moved into evidence in any judicial, administrative, legislative or adjudicatory hearing or other proceeding (including discovery proceedings) in the United States or any Foreign country. Such notification shall include: 1) The formal name of the proceeding, including docket number or similar identifier; 2) the name and location of the body conducting the hearing or proceeding; 3) subsequent to the use of this document in a formal proceeding please notify SWGDE as to its use and outcome; 4) the name, mailing address (if available) and contact information of the party offering or moving the document into evidence. Notifications should be sent to secretary@swgde.org.

It is the reader's responsibility to ensure they have the most current version of this document. It is recommended that previous versions be archived.

Redistribution Policy:

SWGDE grants permission for redistribution and use of all publicly posted documents created by SWGDE, provided that the following conditions are met:

1. Redistribution of documents or parts of documents must retain the SWGDE cover page containing the disclaimer.
2. Neither the name of SWGDE nor the names of contributors may be used to endorse or promote products derived from its documents.
3. Any reference or quote from a SWGDE document must include the version number (or create date) of the document and mention if the document is in a draft status.

Intellectual Property:
Unauthorized use of the SWGDE logo or documents without written permission from SWGDE is a violation of our intellectual property rights.

Individuals may not misstate and/or over represent duties and responsibilities of SWGDE work. This includes claiming oneself as a contributing member without actively participating in SWGDE meetings; claiming oneself as an officer of SWGDE without serving as such; claiming sole authorship of a document; use the SWGDE logo on any material and/or curriculum vitae.

Any mention of specific products within SWGDE documents is for informational purposes only; it does not imply a recommendation or endorsement by SWGDE.

SWGDE Position on the Use of MD5 and SHA1 Hash Algorithms in Digital and Multimedia Forensics

Table of Contents

1. Purpose

The purpose of this document is to explain that the use of the MD5 and SHA1 hash algorithms remains acceptable for certain functions in digital and multimedia forensic disciplines despite the algorithms having been shown to be inappropriate for broader cryptographic purposes.

2. Scope

This document addresses the use of the MD5 and SHA1 hash algorithms for integrity verification and file identification.

3. Summary

There are many types of hashes that are used for different purposes. This paper discusses four commonly-used hashing algorithms: MD5, SHA1, SHA2, and SHA3.

While SWGDE promotes the adoption of SHA2 and SHA3 by vendors and practitioners, the MD5 and SHA1 algorithms remain acceptable for integrity verification and file identification applications in digital forensics. Because of known limitations of the MD5 and SHA1 algorithms, only SHA2 and SHA3 are appropriate for digital signatures and other security applications.

4. Background

4.1 Hashing: General Background

Hash algorithms use complex mathematics to create a value that is typically represented as a string of hexadecimal characters (called a hash) based on a given set of data. If the data changes, so will the hash. When two different datasets produce the same hash, this is called a collision. There are two types of collisions: random collisions and deliberately engineered collisions. While there have been engineered collisions in some modern hash algorithms, no random collisions have yet been identified. (See Reference Section for additional information.)

Hashing serves a variety of functions. It is used for integrity verification, file identification, random number generation, creating a unique representation of a file to be digitally signed, and many other uses. Most public key digital signature applications require a hash as part of the process. In the field of digital forensics, hashing is primarily used for integrity verification and file identification.

4.2 Integrity Verification

The goal of integrity verification is to determine if data has changed since the hash value was calculated. This is a common use of hashes in digital and multimedia forensics. Integrity verification is important for maintaining a chain of custody. When a file is hashed, a “digital fingerprint” of a file is created, which is unique to the file. The change of even one bit in a file will cause the hash to change.

4.3 File Identification

The goal of file identification is to efficiently scan a digital image or other digital object for specific items. Because a hash is significantly smaller than the file it describes and it is not possible to re-create a file based on its hash, many organizations use hashes to identify files. The hashes are easier to distribute and protected information is not disclosed. Many common forensic tools use lists of hash values to identify known and notable files in an examination. For example, the National Institute of Standards and Technology (NIST) National Software Reference Library distributes hashes of known software. See www.nsrl.nist.gov.

4.4 What makes hash algorithms good or bad?

Hash algorithms are ranked according to their strength for various underlying properties, including collision resistance and preimage resistance.

- Collision resistance means that there are not two files with the same hash. Collisions can occur randomly or be engineered. If it is possible to create two files with the same hash, the algorithm is considered broken with respect to collision resistance.
- Preimage resistance is the inability to create a file with the same hash as a previously calculated hash.
- Other properties. There are other properties of hashing algorithms, such as the inability to reverse engineer a file from its hash and the computational efficiency of the algorithm.

4.5 Description of Commonly-Used Hash Algorithms

There are four common hash algorithm families in current use. (See Reference Section for additional information.)

1. **MD5:** MD5 is defined in IETF RFC 1321 and is the oldest among these hash algorithms. The first practical public attack on MD5 was published in 2004. This attack can be used to engineer hash collisions, involving the deliberate creation of two files with the same hash. It cannot be used to create a different file whose hash matches a pre-existing file's hash.
2. **SHA1:** SHA1 was first allowed for federal use in NIST publication FIPS 180 in 1995 and subsequently disallowed in 2010. The attack on MD5 raised the possibility that the same type of attack could be used on SHA1. The first successful SHA1 collision attack was published in 2017. However, like the MD5 attack, it cannot be used to create a different file whose hash matches a pre-existing file's hash.
3. **SHA2:** NIST added SHA2 to FIPS 180 in 2006 after concerns about SHA1 surfaced. It includes multiple versions, the most common being SHA256 and SHA512. There are no known significant attacks against SHA2.
4. **SHA3:** NIST added SHA3 to FIPS 180 in 2015. It uses a different algorithm than previous SHA versions. There are no known successful attacks against SHA3.

The numbers involved with random collisions and preimage attacks are hard to visualize. The following table provides a visual analogy:

	Random Collision Probability Numerical	**Random Collision Probability** Analogy
MD5	1 in 2^{64} (about 1 in 1.84×10^{19})	One drop out of all the water on Earth
SHA1	1 in 2^{80} (about 1 in 1.21×10^{24})	One drop out of all the water in the solar system
SHA2/3	Between 1 in 2^{112} and 1 in 2^{256} (between about 1 in 5.19×10^{33} and 1 in 1.16×10^{77})	One drop out of all the water in the Milky Way galaxy

5. Recommendations for the Appropriate Uses of MD5 and SHA1

Because MD5 and SHA1 have proven to be susceptible to engineered collisions, they should only be used for certain functions. It is still appropriate to use MD5 and SHA1 for the following situations:

- **Integrity Verification**
  It is appropriate to use both MD5 and SHA1 for integrity verification provided the hash is securely stored or recorded in examination documentation. This will prevent an individual from substituting a different file and its hash. This is true for all hash algorithms.
- **File Identification**
  Since there are no preimage attacks against MD5 and SHA1, it is appropriate to use both algorithms for file identification.

6. References:

[1] Internet Engineering Task Force, The MD5 Message-Digest Algorithm, April 1992, https://www.ietf.org/rfc/rfc1321.txt

[2] National Institute of Standards and Technology, NIST Policy on Hash Functions, August 5, 2015, https://csrc.nist.gov/Projects/Hash-Functions/NIST-Policy-on-Hash-Functions

[3] Google Security Blog, Announcing the first SHA1 collision, February 23, 2017, https://security.googleblog.com/2017/02/announcing-first-sha1-collision.html

[4] Xiaoyun Wang, Dengguo Feng, Xuejia Lai, and Hongbo Yu, Collisions for Hash Functions MD4, MD5, HAVAL-128 and RIPEMD, Cryptology ePrint Archive: Report 2004/199, https://eprint.iacr.org/2004/199

[5] National Institute of Standards and Technology, FIPS 180-1: Secure Hash Standard, April 17, 1995, https://csrc.nist.gov/publications/detail/fips/180/1/archive/1995-04-17

[6] National Institute of Standards and Technology, SP 800-107, Recommendations for Applications Using Approved Hash Algorithms, August 2012, https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-107r1.pdf

[7] National Institute of Standards and Technology, NIST Comments on Cryptanalytic Attacks on SHA-1, April 26, 2006, https://csrc.nist.gov/News/2006/NIST-Comments-on-Cryptanalytic-Attacks-on-SHA-1

[8] IOP Science, Journal of Physics: Conference Series, A comparative study of Message Digist 5 (MD5) and SHA256 algorithm, March 2018, http://iopscience.iop.org/article/10.1088/1742-6596/978/1/012116

[9] Sharma, Arvind & K Mittal, S. (2018). Comparative Analysis of Cryptograhpic Hash Functions, https://www.researchgate.net/publication/327664102_COMPARATIVE_ANALYSIS_OF_CRYPTOGRAPHIC_HASH_FUNCTIONS

[10] Preshing on Programming, Hash Collision Probabilities, May 04, 2011, http://preshing.com/20110504/hash-collision-probabilities/

SWGDE Position on the Use of MD5 and SHA1 Hash Algorithms in Digital and Multimedia Forensics

History

Revision	Issue Date	Section	History
1.0 DRAFT	2018-09-20	All	Initial draft created and voted by SWGDE for release as a Draft for Public Comment.
1.0 DRAFT	2018-11-20	--	Formatted and released as a Draft for Public Comment.