Hash Algorithms and the significance of bit length

Joseph Patrick C. Galvez

American Military University

ISCC266: Cryptography Concepts

Mr. Joel Stewart

June 11, 2025

Hash Algorithms and the significance of bit length

The idea of using a message digest to verify the integrity of data first arose in the 1970s, the most significant method being the development being that of the Merkle-Damgard Construction in 1979. Two such hash functions that use the Merkle-Damgard Construction are MD5 and SHA-1. Today MD5 is largely considered obsolete because it has been around for so long and it has been so thoroughly decoded that there are hashing calculators available online which can encrypt and decode MD5 hashes. On the other hand, SHA-1 was formally adopted by the NIST as an approved hashing function as a federal standard partially due to its longer bit length. In regard to bit length, it is only effective to increase bit length for a benefit to security only up to a certain point.

MD5 was created in 1992 by Ronald L. Rivest, it relies on a 128 bit key length and as previously stated it is based off of the Merkle-Damgard Construction. According to Rivest the MD5 algorithm utilizes five steps to perform its hashing functions. Step one consists of padding the message which means that it is extended to that it is 448 bits so that it is 64 bits short of 512 bits (Rivest, 1992, p.2). Next the message is appended so that the original message is exactly 64 bits which combines with the padded portion to make a 512 bit message block (Rivest, 1992, p.2). Next an MD buffer is initialized, the MD buffer consists of 4 32 bit variables that are predetermined by Rivest and are named: A, B, C, and D; the values are A: 01 23 45 67, B: 89 ab cd ef, C: fe dc ba 98, D: 76 54 32 10 (Rivest, 1992, p.3). The fourth step consists of dividing the message block into 16 sections each consisting of 32 bits and each section subject to 4 rounds of operations which uses non linear functions called F,G,H, and I. F applies to operations 1-16, G applies to 17-32, H applies to 33-48, and I applies to 49-64 these functions are nonlinear and perform a combination of XOR, AND, and NOT functions (Rivest, 1992, p.3 ). To further explain how the non linear functions work X,Y, and Z are derived from the buffer portion and the F,G,H, and I variables determine the order in which they appear relative to the logical operators in each equation, they also rotate to seem more random (Rivest, 1992. P.3). And lastly step 5 consists of taking the four 32 bit outputs, combining them with eachother, using little-endian output to present the least significant byte first and then creating a 128-hexadecimal string as a final output.

SHA-1 was created in 1995 by the NSA as a replacement for the original SHA algorithm as a version that features an extended output size. SHA-1 and MD5 have similarities in that they both use the Merkle-Damgard construction, they both use padding to create fixed length blocks with are both 512 bits during the creation of the digest, and they use little-endian format to make their outputs seem more random. However, the similarities end there. Step one is to pad the plaintext so that the fixed block length of 512 bits is achieved. The way this is done is determined through the formula L + 1 + k = 448 mod 512 where L is the original message, 1 is an extra bit, and k is the padding and is entirely made of zeroes (NIST, 2015, p.13). Afterwards the message is parsed into 512 bit blocks expressed as 16 32 bit words (NIST, 2015.p.14). Next initial hash values are determined for SHA-1 they are: H0⁰ = 67452301  
H1⁰ = EFCDAB89, H2⁰ = 98BADCFE, H3⁰ = 10325476, and H4⁰ = C3D2E1F0 (NIST, 2015, p.14). The blocks are then subjected to 80 rounds of processing with a nonlinear function and a left rotation in each round. The result of SHA-1 is a 160-bit message digest.

I stated in the introduction that it is only productive to increase the bit length of a hash and expect a reasonable gain in security up to a certain point. To be more specific if SHA-1 is compared to SHA-256 the difference in bit length is 96 bits comparing SHA-1’s 160 bits to SHA-256’s 256 bits; however, collision resistance only increased by 48 bits(Dang,2012,p.8). This is because of the birthday problem which asks how many people can be in a room before there’s a 50% chance one person shares a birthday with another person, intuitively it is not 183 but 23 which demonstrates how the probability of pairings grows disproportionately to the sample size. This is significant in hashing because the probability of a message digest having a collision grows disproportionately to the number of possible digests as a result of bit strength. As a result of this tendency it is often observed that doubling the bit strength only increases in an increase of collision resistance of 50%. It is however, notable that the increase of preimage resistance is linear.

References

Rivest, R. L. (1992). *The MD5 message-digest algorithm* (RFC 1321). Internet Engineering Task Force. <https://www.ietf.org/rfc/rfc1321.txt>

National Institute of Standards and Technology. (2015). *Secure hash standard (SHS)* (FIPS PUB 180-4). U.S. Department of Commerce. <https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.180-4.pdf>

Dang, Q. (2012). Recommendation for applications using approved hash algorithms (NIST Special Publication 800-107 Rev. 1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-107r1

*For additional information on APA Style formatting, please consult the* [*APA Style Manual, 7th*](https://apastyle.apa.org/style-grammar-guidelines) *Edition.*