MD5 and the Effect of Collisions on Message Integrity

Joseph Patrick C. Galvez

American Military University

ISCC266: Cryptography Concepts

Mr. Joel Stewart

June 1,2025

MD5 and the Effect of Collisions on Message Integrity

In information security the need arises to verify if data has changed, for example if data is stored it may be difficult to verify if it has changed even in a small way if there is a large amount of it, in addition to this problem it may be insecure to store passwords as plaintext. This is where hashing comes in – hashing is an algorithm that creates a digest which is used to create a unique representation of the original data which changes drastically even to tiny changes. For a digest to be secure it must be near impossible for a different input to create the same digest, it must be practically impossible to extract the original data from the digest, slight changes must make a largely different output, and the last requirement is that hashes are fixed width meaning that the length of a hash will be the same regardless of the length of the data (Yaga et al., 2018, p.7).

The NIST defines a collision as “An event in which two different messages have the same

message digest” (Dang, 2012, p.3). The ability of a hash to remain secure is by ensuring that collisions do not occur, a quality known as collision resistance; collision resistance is calculated by finding the amount of computational power needed to find a collision for a hash function (Dang, 2012, p.6). If a collision occurs then the primary purpose of hashing, that of verifying data integrity, is defeated; this is why it is important that hashes remain unique, and the probability of collisions are maintained to be at a negligible level.

As was previously stated it is critical to maintain a low level of collision probability and to maintain security. Some methods are better than others, and one such method that is no longer used due to its ability to maintain security and a low probability of collisions is MD5. MD5 or Message Digest 5 is a hashing method that relies on the Merkle-Damgard Hash Function to create a digest by breaking down the plaintext and using two compression functions to create a message digest (Burr, 2005, slide 2).

It is further explained that MD5 is vulnerable to two types of attacks one being a collision attack and the other being a second preimage attack. A collision attack consists of using two hashes with the same digest and using the identical digest to attack the notion that the original data is the data meant to be verified, to defend against one must find collisions before signing. A signature second preimage attack uses a collision to give the attacker a digest that is usable as a signature for the data to be protected, this attack is possible even after signing since a collision can still be found afterwards. It is also stated in the presentation that MD5 was fully broken in 2004 (Burr, 2005, slide 11). The aforementioned reasons are why MD5 is not used and why the SHA families of hashing methods are preferred.

In conclusion, Hashing is the practice of using an algorithm to create a digest or a sequence that is unique to a set of data to verify the integrity of this data. Hashing has vulnerabilities called collisions that can be exploited by collision attacks and second preimage attacks that can either repudiate a signature or create a forged signature. And lastly, MD5 is considered to be too vulnerable to these attacks which is why the SHA family of algorithms are preferred by the NIST.

References

Yaga, D., Mell, P., Roby, N., & Scarfone, K. (2018). Blockchain technology overview (NIST Internal Report 8202). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.IR.8202>

Dang, Q. H. (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>

Burr, W. E. (2005, December). *NIST hash function standards: Status and plans* [Presentation]. Information Security and Privacy Advisory Board (ISPAB) Meeting. National Institute of Standards and Technology. <https://csrc.nist.gov/csrc/media/events/ispab-december-2005-meeting/documents/b_burr-dec2005-ispab.pdf>

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