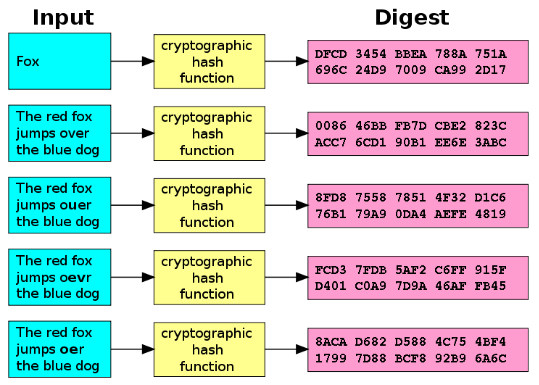
**MD5 Algorithm / Cryptographic Hashing:**

The MD5 message-digest algorithm is a widely used cryptographic hash function producing a 128-bit (16-byte) hash value, typically expressed in text format as a 32 digit hexadecimal number. MD5 has been utilized in a wide variety of cryptographic applications, and is also commonly used to verify data integrity.

MD5 stands for Message Digest algorithm 5, and was invented by celebrated US cryptographer Professor Ronald Rivest in 1991 to replace the old MD4 standard. MD5 is simply the name for a type of cryptographic hashing function Ron came up with, way back in ’91.

The idea behind cryptographic hashing is to take an arbitrary block of data and return a fixed-size “hash” value. It can be any data, of any size but the hash value will always be fixed.



Cryptographic hashing has a number of uses, and there are a vast number of algorithms (other than MD5) designed to do a similar job. One of the main uses for cryptographic hashing is for verifying the contents of a message or file after transfer.

If you’ve ever downloaded a particularly large file (Linux distributions, that sort of thing) you’ll probably have noticed the hash value that accompanies it. Once this file has been downloaded, you can use the hash to verify that the file you downloaded is in no way different to the file advertised.

The same method works for messages, with the hash verifying that the message received matches the message sent. On a very basic level, if you and a friend have a large file each and wish to verify they’re exactly the same without the hefty transfer, the hash code will do it for you.

Hashing algorithms also play a part in data or file identification. A good example for this is peer to peer file sharing networks, such as eDonkey2000. The system used a variant of the MD4 algorithm (below) which also combined file’s size into a hash to quickly point to files on the network.

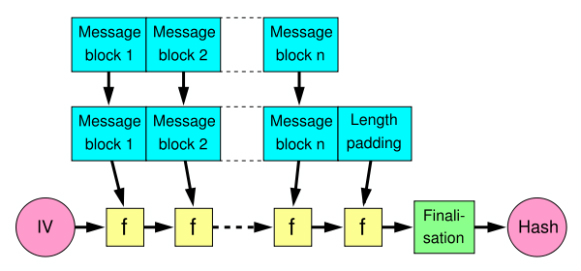
A signature example of this is in the ability to quickly find data in hash tables, a method commonly used by search engines.

Another use for hashes is in the storage of passwords. Storing passwords as clear text is a bad idea, for obvious reasons so instead they are converted to hash values. When a user inputs a password it is converted to a hash value, and checked against the known stored hash. As hashing is a one-way process, provided the algorithm is sound then there is theoretically little chance of the original password being deciphered from the hash.

Cryptographic hashing is also often used in the generation of passwords, and derivative passwords from a single phrase.

**Message Digest algorithm 5**

The MD5 function provides a 32 digit hexadecimal number. If we were to turn ‘makeuseof.com’ into an MD5 hash value then it would look like: 64399513b7d734ca90181b27a62134dc. It was built upon a method called the Merkle”“DamgÃ¥rd structure (below), which is used to build what are known as “collision-proof” hash functions.



No security is everything-proof, however and in 1996 potential flaws was found within the MD5 hashing algorithm. At the time these were not seen as fatal, and MD5 continued to be used. In 2004 a far more serious problem was discovered after a group of researchers described how to make two separate files share the same MD5 hash value. This was the first instance of a collision attack being used against the MD5 hashing algorithm. A collision attack attempts to find two arbitrary outputs which produce the same hash value – hence, a collision (two files existing with the same value).

Over the next few years attempts to find further security problems within MD5 took place, and in 2008 another research groups managed to use the collision attack method to fake SSL certificate validity. This could dupe users into thinking they are browsing securely, when they are not.

**Security:**

The security of the MD5 hash function is severely compromised. A collision attack exists that can find collisions within seconds on a computer with a 2.6 GHz Pentium 4 processor (complexity of 224.1). Further, there is also a chosen-prefix collision attack that can produce a collision for two inputs with specified prefixes within hours, using off-the-shelf computing hardware (complexity 239). The ability to find collisions has been greatly aided by the use of off-the-shelf GPUs. On an NVIDIA GeForce 8400GS graphics processor, 16–18 million hashes per second can be computed. An NVIDIA GeForce 8800 Ultra can calculate more than 200 million hashes per second.

These hash and collision attacks have been demonstrated in the public in various situations, including colliding document files and digital certificates. As of 2015 MD5 was demonstrated to be still quite widely used, most notably by security research and antivirus companies.

In April 2009, a preimage attack against MD5 was published that breaks MD5's preimage resistance. This attack is only theoretical, with a computational complexity of 2123.4 for full preimage.

**Conclusion:**

MD5 is one of many different methods of identifying, securing and verifying data. Cryptographic hashing is a vital chapter in the history of security, and keeping things hidden. As with many things designed with security in mind, someone’s gone and broken it.