# **Hash functions**

## **Introduction**

A hash function is a function that can be used to map data of any size to a number of fixed size. The values returned by a hash function are called hash values, hash codes, digests, or hashes. The values are used to index a fixed-size table called a hash table. Using hash functions to index a hash table is called hashing. (Knuth, D. 1973, The Art of Computer Science, Vol. 3)

Hashes do not need to be able to be decrypted. If 2 pieces of data have the same hash then it is (probably) the same file.

A hash function takes an input as a key, which is associated with a method and used to map it to the data storage and retrieval application. The keys may be fixed length, like an integer, or variable length, like a name. In some cases, the key is the data itself. The output is a hash code used to index a hash table holding the data or records, or pointers to them.

A hash function has three requirements:

1. It must be fast, but it can’t be too fast or else it becomes too easy to crack.
2. It must have an avalanche effect. Changing one bit ANYWHERE in the file should change the entire Hash.
3. It must be able to avoid collisions. Think of an analogy with cables. If there is a collision with Hashes, then it’s like trying to plug 2 cables into the 1 socket.

## **MD5**

A Hash function is considered broken if it’s possible to create collisions deliberately. MD5 was previously the most widely used but now it is considered broken for this exact reason. People found out how to deliberately create collisions.

The problem with this is that if you are able to intercept a file and edit it and have the hash stay the exact same then you can put something malicious into the file and cause problems.

MD5 is so broken now that you can figure out what its hash’s represent by typing them into google. People used to store passwords this way and if anyone were to use MD5 for a password nowadays someone could crack it by just looking it up on Google.

**SHA1, 2 and 3**

People then began to move to SHA-1 which was created by the NSA. However now it’s thought that this might start to become broken as well as computers get faster and faster.

Some moved to SHA-2, which for the time being is secure.

SHA-3 is being checked by agencies. In a few years this will become the standard.

None of these should ever be used for storing passwords. They can become broken too easily. They should only be used for file transfer.

## **Hash tables**

Hash functions are used with Hash table to store and retrieve data items or data records. The hash function translates the key associated with each piece of data into a hash code which is used to index the hash table. When an item is added to the table, the hash code might index an empty slot (bucket), in which case the item is added to the table there. If the hash code indexes a full slot, some kind of collision resolution is required. (*Menezes, Alfred J.; van Oorschot, Paul C.; Vanstone, Scott A (1996).* [*Handbook of Applied Cryptography*](https://archive.org/details/handbookofapplie0000mene)*. CRC Press.*)

The new item may:

* not be added to the table
* replace the old item
* added to the table according to some other rules.

REFRENCES

Knuth, D. 1973, The Art of Computer Science, Vol. 3, Sorting and Searching, p.527. Addison-Wesley, Reading, MA., United States

<https://youtu.be/b4b8ktEV4Bg>

<https://youtu.be/DMtFhACPnTY>

*Menezes, Alfred J.; van Oorschot, Paul C.; Vanstone, Scott A (1996).* [*Handbook of Applied Cryptography*](https://archive.org/details/handbookofapplie0000mene)*. CRC Press.*