HASH FUNCTIONS

 Version : 0.0.4 Update : 23/03/2022

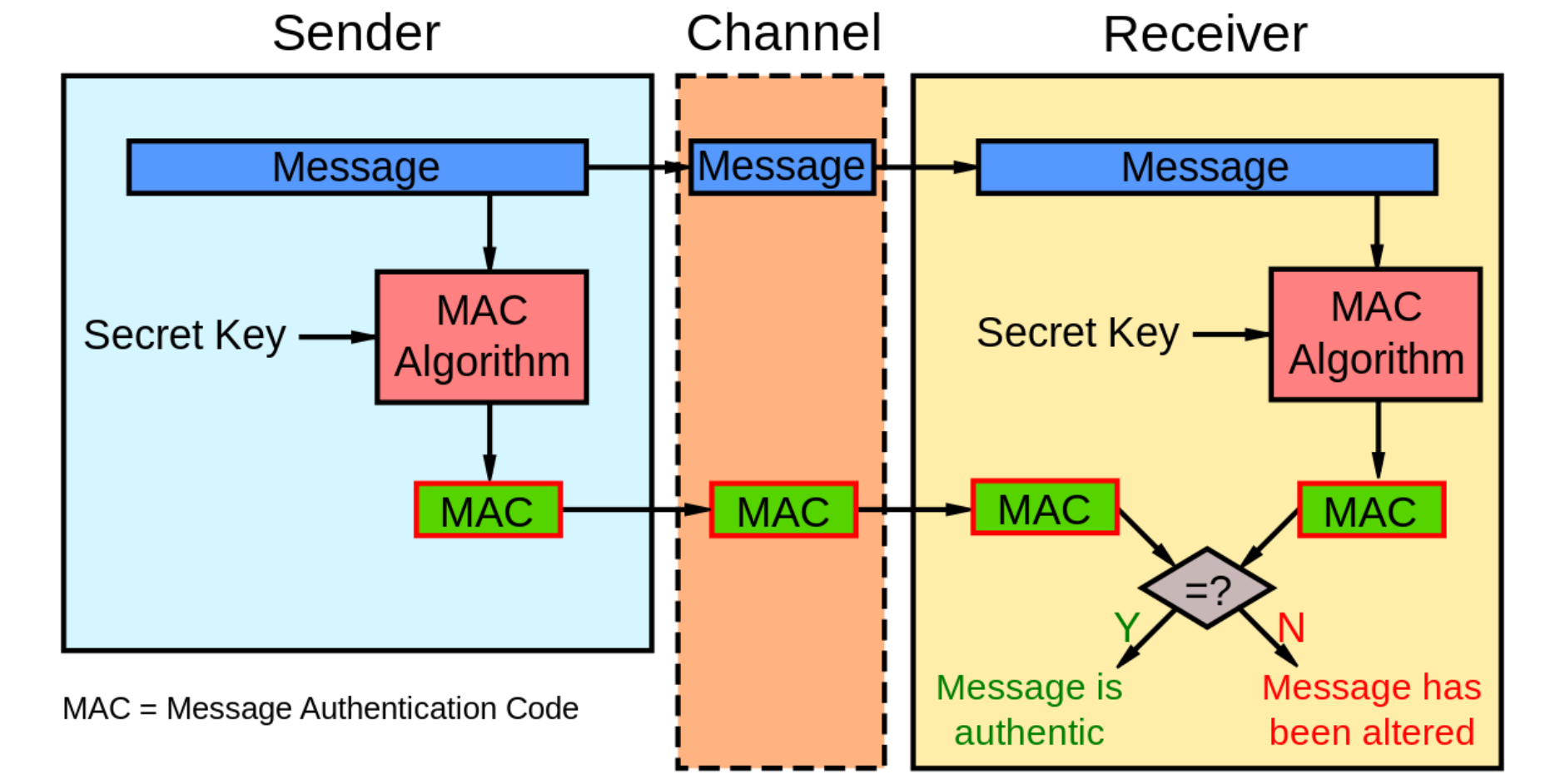
Introduction

- Hash function is a special function which, from data supplied as input, calculates a digital fingerprint used to quickly identify the initial data, just like a signature to identify a person. Hash functions are used in computing and in cryptography, in particular to quickly recognize files or passwords.
- This type of algorithm provides as a result a number whose size is fixed, whatever the size of the plain text input.
- A plaintext collision is when two different plaintext results in the same fingerprint.
- Checksum and CRC algorithms are not strong algorithms.
  - These two algorithms are not strong enough for cryptography, because the number of possible fingerprints is really too small (2 ^ 16 for the CHECKSUM algorithm or 2 ^ 32 for the CRC-32 algorithm to compare to 2 ^ 128 of the MD5 or even 2 ^ 160 of SHA-1).

Terms

- Message Authentication Code (MAC), is a short piece of information used to authenticate a message. In other words, to confirm that the message came from the stated sender (its authenticity) and hasn't been changed. The **MAC value protects** a message's **data integrity**, as well as its **authenticity**, by allowing verifiers (who also possess the secret key) to detect any changes to the message content.
- Hash-based Message Authentication Code (HMAC) is a specific type of message authentication code (MAC) involving a cryptographic hash function and a secret cryptographic key. HMAC can provide digital signatures using a shared secret instead of public key encryption.

Terms in image



Collision

- A hash function's “collision” is a pair of data distinct from its starting set, the checksums of which are identical.
- Collisions are generally considered undesirable but are generally impossible to avoid due to the size difference between the starting and ending sets of the function.
- This situation is considered rare, even impossible, depending on the quality level of the hash function. This is what makes it possible to consider that a file (or a password) corresponds to a unique signature. And therefore a given signature can only come from a single starting file (or password).
- A birthday attack is a type of cryptographic attack that exploits the mathematics behind the birthday problem in probability theory.

Collision

- Example:
  MD5 Collision Demo with
  these two blocks.
- Each of these blocks has
  MD5 hash :
  79054025255fb1a26e4bc422aef54eb4.

```
d131dd02c5e6eec4 693d9a0698aff95c 2fcab58712467eab 4004583eb8fb7f89
55ad340609f4b302 83e4888325 71415a 085125e8f7cdc99f d91dbdf280373c5b
d8823e3156348f5b ae6dacd436c919c6 dd53e2b487da03fd 02396306d248cda0
e99f33420f577ee8 ce54b67080a80d1e c69821bcb6a88393 96f9652b6ff72a70

d131dd02c5e6eec4 693d9a0698aff95c 2fcab50712467eab 4004583eb8fb7f89
55ad340609f4b302 83e4888325f1415a 085125e8f7cdc99f d91dbd7280373c5b
d8823e3156348f5b ae6dacd436c919c6 dd53e23487da03fd 02396306d248cda0
e99f33420f577ee8 ce54b6708028 0d1e c69821bcb6a88393 96f965ab6ff72a70
```

Length extension attack

- A length extension attack is a type of attack where an attacker can use Hash (message1) and the length of message1 to calculate Hash(message1 concatenate with message2) for an attacker-controlled message2, without needing to know the content of message1.
- Algorithms like MD5, SHA-1 and most of SHA-2 that are based on the Merkle–Damgård construction are susceptible to this kind of attack.
- When a Merkle–Damgård based hash is misused as a message authentication code with construction H(secret concatenate with message), and message and the length of secret is known, a length extension attack allows anyone to include extra information at the end of the message and produce a valid hash without knowing the secret.
- Since HMAC doesn't use this construction, HMAC hashes aren't prone to length extension attacks.

Message Digest

- The MD2, MD4, MD5 and MD6 algorithms are digital fingerprint calculation algorithms.
- If the MD5 algorithm is of significant historical interest, it is today considered to be outdated and absolutely unsuitable for any use in cryptography or in security.
- MD6 was developed to participate in the 2008 NIST hash function competition but was not selected.

Message Digest algorithm

- MD5 processes a variable-length message into a fixed-length output of 128 bits.
- The input message is broken up into chunks of 512-bit blocks (sixteen 32-bit words).
- The message is padded first a single bit “1”, is appended to the end of the message. This is followed by as many zeros as are required to bring the length.

Message Digest attacks

- John the ripper allows you to crack / reverse function MD5 by brute force.
- Rainbow tables can often crack in less than a second.
  - These tables use dictionaries established after several days, months or years of calculation.
  - These don't contain all of the possible MD5 keys, nor are they intended for brute force breaking (one hash has 128 bits, which represents about 4.10^38 of combinations).
  - These allow by examining the footprint to eliminate very large classes of combinations not to be tested.
  - The efficiency of rainbow tables decreases if the footprint is calculated with a "salt".

Secure Hash Algorithms

- The Secure Hash Algorithms are a family of cryptographic hash functions published by the National Institute of Standards and Technology (NIST) as a U.S. Federal Information Processing Standard (FIPS), including:
  - SHA-0: It was withdrawn shortly after publication due to an undisclosed "significant flaw" and replaced by the slightly revised version SHA-1.
  - SHA-1: Designed by the NSA to be part of the Digital Signature Algorithm. Cryptographic weaknesses were discovered in SHA-1, and the standard was no longer approved for most cryptographic uses after 2010.
  - SHA-2: Two similar hash functions with different block sizes SHA-256 and SHA-512. They differ in the word size; SHA-256 uses 32-byte words where SHA-512 uses 64-byte words. There are also truncated versions of each standard, known as SHA-224, SHA-384, SHA-512/224 and SHA-512/256.
  - SHA-3: A hash function formerly called Keccak, chosen in 2012 after a public competition among non-NSA designers. It supports the same hash lengths as SHA-2, and its internal structure differs significantly from the rest of the SHA family.

Secure Hash Algorithms

Algorithm and variant	Output size (bits)	Internal state size (bits)	Block size (bits)	Rounds	Security (in bits) against collision attacks	Capacity against length extension attacks	First published
MD5	128	128	512	64	**≤18 (collisions found)**	**0**	1992
SHA-0	160	160	512	80	**<34 (collisions found)**	**0**	1993
SHA-1	160	160	512	80	**<63 (collisions found)**	**0**	1995
SHA-2: - SHA-224 - SHA-256	 224 256	 256 256	 512 512	 64 64	 **112** **128**	 **32** **0**	 2004 2001
SHA-2 : - SHA-384 - SHA-512	 384 512	 512 512	 1024 1024	 80 80	 **192** **256**	 **128** **0**	 2001 2001
SHA-2 : - SHA-512/224 - SHA-512/256	 224 256	 512 512	 1024 1024	 80 80	 **112** **128**	 **288** **256**	 2012 2012
SHA-3: - SHA3-224 - SHA3-256 - SHA3-384 - SHA3-512	 224 256 384 512	 1600 1600 1600 1600	 1152 1088 832 576	 24 24 24 24	 **112** **128** **192** **256**	 **448** **512** **768** **1024**	 2015 2015 2015 2015