## Cryptographic hash function

A hash function $h(\mathtt{m})$ is a message digest; in some sense, the message is condensed. Hash functions are routinely used to check integrity or for error detection of transmitted messages. Nick and Alex must agree on a hash function. If Nick is sending a message to Alex, he might create a hash of the message and transmit it along with the message. After receiving the message, Alex creates a hash message that he received using the hash function that he and Nick have agreed to use. The two hashes should be the same. If they are, Alex can assume that the message has not been altered intentionally or unintentionally during transmission.

Hash functions should accept messages of any length as input, produce a fixed-length output, and be fast.

A hash function that will be used for cryptographic purposes should have some other properties:

> A cryptographic hash function should be one-way. Knowing an output $h$ of the hash function it should computationally infeasible to find a message $\mathtt{m}$ which hashes to that output; i.e., for which $h(\mathtt{m}) = h$. (This property is called pre-image resistant.)
>
> A cryptographic hash function should also be second pre-image resistant – given a message $\mathtt{m}_1$, it should be computationally infeasible to find another message $\mathtt{m}_2$ with $\mathtt{m}_1 \neq \mathtt{m}_2$ having $h(\mathtt{m}_1)=h(\mathtt{m}_2)$.
>
> A cryptographic hash function should be strongly collision free. It should be computationally infeasible to find two different inputs that have the same hash; i.e., it should be computationally infeasible to find messages $\mathtt{m}_1 \neq \mathtt{m}_2$ having $h(\mathtt{m}_1)=h(\mathtt{m}_2)$. Of course, the number of inputs is much larger than the number of outputs; so, collisions will occur but collisions should be unlikely.

There are two widely used families of cryptographic hash functions – the MD family (MD = message digest) and the SHA family (SHA = secure hash algorithm). Rivest and RSA laboratories developed MD4 and now MD5. The original MD was never published; MD2 was the first of the family to appear, and it was followed by MD4. The NSA developed SHA-1 and SHA-2. Around February 2005, problems with SHA-1 became public.

Since the development of public-key cryptography, mathematics plays a significant, visible role in cryptography and cryptanalysis. The construction of hash functions, however, remains mostly non-mathematical and is probably more an art than a science. (In reality, the same can still be said about cryptography and cryptanalysis despite the scientific image the mathematical ideas impart.)

Hash functions allows authentication to occur without double encryption of the entire message.

Nick and Alex must agree on a hash function. Then Nick can (for security) send his message using Alex's public key. Also, he creates a hash of the plaintext and (for authentication) sends it

using his private key. Using his private key, Alex decrypts the ciphertext encrypted with his public key and creates a hash of the plaintext using the hash function that he and Nick have agreed to use. Alex also decrypts the ciphertext of the hash using Nick's public key. The two hashes should be the same. If they are, Alex can assume that the message is secure and that it came from Nick.

Message authentication codes (MAC) check both integrity and authenticity. MACs require the parties in the communication to agree on an algorithm and possess a secret key. The MAC algorithm uses the secret key and the message as input, and it outputs a message authentication code. Authorized receivers who possess the secret key can input the key and the message that they received and check their calculation of the MAC against the MAC transmitted with the message. If the two MACs agree, the receiver can be fairly sure of both the integrity and authenticity of the message. Cryptographic hash functions and block ciphers are often used to construct MAC algorithms.

Cipher suites typically contain key exchange algorithms, signature algorithms, and cryptographic hash functions.

## References

Susan Landau, "Find Me a Hash," *Notices of the American Mathematical Society*, 53(3), March 2006, 330 – 332.

Richard Spillman, *Classical and Contemporary Cryptology*, Prentice Hall, 2005. This text has a nice discussion of MD5 and SHA.