Merkle-Damgård Construction

Keane J. Moraes

1. History

The Merkle-Damgård Construction is a method of building one-way collision-resistant hash functions from one-way collision resistant compression functions. This construction was developed by Ralph Merkle for his doctoral thesis in 1979. In his thesis '*Secrecy, authentication, and public key systems*',[1]

2. Modern Usage

This construction is used in many modern cryptographic collision-resistant hash functions (CCRHFs) like SHA1, SHA2 and fairly recently SHA3. It was also used in MD5 which was a hash function developed by Ron Rivest at MIT, however, MD5 is no longer used since several collisions have been already found.

3. One-way Compression Functions

One-way Compression Functions are a class of functions that consume 2 values - a fixed-length input and a chaining value and they return a fixed length output. These one way compression functions differ from the traditional lossless or lossy data compression algortihms since those outputs can be reconstructed to give the input or some approximation of the input.
One-way compression functions (OWCFs) are often built from block ciphers where the message is the chaining value and the key is the fixed-length input.

3.1. Definition

One-way
A one-way function is some function that is easy to compute in one direction but hard to invert given some output of the function. These kinds of functions are widely used in cryptography. One famous example is the RSA Trapdoor Function. It is considered one-way for someone who does not have access to the private key and cannot efficiently factor the prime product. These one-way functions are also used in PRGs, digital signatures and authentication schemes.
In this case, the fixed-length output can be easily generated from some fixed-length input and a chaining value but given some output of an OWCF, it is extremely difficult (likely intractable) to derive the fixed-length input from it.

> **NOTE:**
> There is no way to prove that one-way functions exist. If one successfully does so, it would imply that the complexity classes NP and P are not equal.

[1] R.C. Merkle. Secrecy, authentication, and public key systems. Stanford. 1979

Compression
A compresssion function will take an input of fixed size N and output some value of fixed size M where $N > M$.
In an OWCF, the function mixes two fixed length inputs and produces a single fixed length output of the same size as one of the inputs.[2]. The fixed-length input could be of size 256 bits, the chaining value could be of size 128 bits and the output value could be of size 128 bits. This is essentially a compression from (128 + 256) bits to 128 bits
In OWCFs the compression is done is such a way that the full Avalanche Effect is acheived.[3]

3.2. Constructions

A One-way compression function should have the following properties :

1. **Easy to compute:** Given some fixed-length input and chaining message, computing a fixed-length output will be efficient.
2. **Primary preimage resistance:** Given an OWCF f and some output of f, say H, the probability of finding some m such that $f(m) = H$ should be negligible.
3. **Secondary preimage resistance:** Given an OWCF f and some input m_1 such that $f(m_1) = H$, the probability of finding some m_2 such that $f(m_2) = H$ should be negligible.
4. **Collision resistant:** Given an OWCF f, the probability of finding any two messages such that $f(m_1) = f(m_2)$ should be negligible.

Given these parameters, we can now look at some important constructions of OWCFs. In this section we will only look at constructions from block ciphers. Some key assumptions are :

- We will assume that all the constructions are built on 'ideal' block ciphers - An **ideal** block cipher $\mathbb{E} : \mathbb{K} \times \{0,1\}^n \rightarrow \{0,1\}^n$ is a collection of $|\mathbb{K}|$ random permutations of $n-$bit binary strings. (Using this assumptions we can state a theorem that the resultant compression function will have collisions only in $O(2^{n/2})$ time - which is the best case scenario. [4])
- The notation $\mathbb{E}(n, m)$ means - we are using a block cipher with n as the key and m as the message block. (Note that the key is **always** written first.)
- The notation H_i represents that i^{th} chained message digest and m_j represents the j^{th} message block.

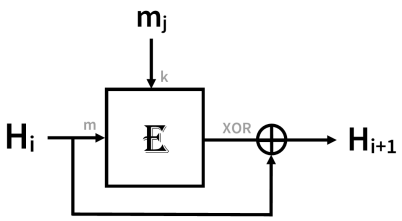
3.2.1. Single Block Length Compression

The constructions output the same number of bits as the fixed-size message input.

Davies-Meyer Construction
This construction for a compression function h is

$$h(H_{i-1}, m_j) = \mathbb{E}(m_j, H_{i-1}) \oplus H_{i-1}$$

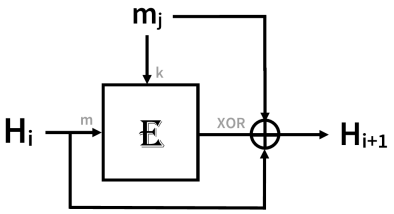
Note that the message block is the key and the i^{th} hash value is plaintext for the block cipher.



[2]From the Wikipedia article on One-way compression function
[3]Read on to find out what the Avanlanche Effect is in cryptography
[4]This is borrowed from Cryptography I (Coursera MOOC) by Dan Boneh of Stanford

Miyaguchi-Preneel
One variant for the construction of the compression function h is

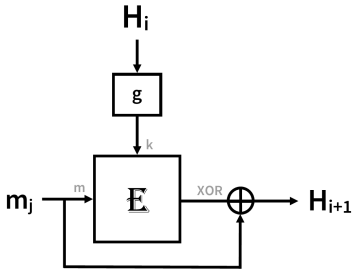
$$h(H_{i-1}, m) = \mathbb{E}(m_j, H_{i-1}) \oplus H_{i-1} \oplus m_j$$



This variant is used in the Whirlpool hash function. There are 12 other variants to the this construction. Some of the natural variants are not collision-resitant

Matas-Meyer Oseas
This construction for a compression function h is

$$h(H_{i-1}, m) = \mathbb{E}(H_{i-1}, m_j) \oplus m_j$$



This is considered the dual of Davies-Meyer since the message block is the plaintext and the i^{th} hash value is the key (the exact opposite of Davies Meyer). The g function is a 'normalization' function that corrects the size of the hash value to the correct key length.

3.3. Collision-Resistance

4. MD Construction

The Merkle-Damgård Construction is a domain-extension method that allows a one-way compression function to act on a message of arbitrary length.

4.1. Padding

4.2. Collision-Resistance Proof

The advantage of the MD Construction comes from the fact that if the compression function is collision-resistant then the whole hash function is collision resistant. The proof for this property is given below.[4]

Theorem 1. *If h is a collision-resistant compression function then H is a collision-resistant hash function.*

Proof. Take two messages from the message space $M, M' \in \mathbb{M}$ where $M \neq M'$. We will prove the contrapositive of the theorem - i.e. if a collision on H is found then h is not collision-resistant.

Suppose that $H(M) = H(M')$. Using this assumption we will build a collsion on h. For $H(M)$:

$$IV = H_0;\ H_1, \dots H_t;\ H_{t+1} = H(M) = h(H_t, m_t||PB)$$

For $H(M')$

$$IV = H'_0;\ H'_1, \dots H'_s;\ H_{s+1} = H(M) = h(H_s, m_s||PB')$$

Since $H(M) = H(M')$, therefore we get

$$h(H_s, m_s||PB') = h(H_t, m_t||PB)$$

Now suppose that □

5. Cryptographic Collision-Resistant Hash Functions

dhethe

5.1. Generic Birthday Attack

Let $\mathbb{H} : \mathbb{M} \to \{0,1\}^n$ be a hash function.
We can define an algorithm $\mathbb{ALG}$ to find a collision in $O(2^{n/2})$ time[4]. It is as follows :

1. Choose $2^{n/2}$ random messages in the message space $(\mathbb{M}) - (m_1, m_2, \dots m_{2^{n/2}})$ that are distinct.
2. For $i = 1, 2, \dots 2^{n/2}$ and compute $t_i := \mathbb{H}(m_i) \in \{0,1\}^n$
3. Look for collisions $t_i = t_j$ where $i \neq j$

This algorithm runs in $\approx 2^{n/2}$ time because it exploits the mathematics of the Birthday Paradox. Applying the result here essentially translates to - Out of 2^n binary strings of length n, one only needs $1.2 \times 2^{n/2}$ of them to have a non-negligible probability ($> \frac{1}{2}$) of getting a collision. A neat proof of the Birthday Paradox is given below.

Birthday Paradox
Let $r_1, r_2, \dots r_n \in \{1, \dots B\}$ be **independently generated, identically distributed** random variables over the set $\{1, \dots B\}$ where $B \in \mathbb{N}$

Theorem 2. *Given $n = 1.2 \times B^{1/2}$ then*

$$P\big[\ \exists i \neq j \mid r_i = r_j\big] > \frac{1}{2}$$

Proof. Note that

$$P\big[\ \exists i \neq j \mid r_i = r_j\big] = 1 - P\big[\ \forall\, i \neq j \mid r_i \neq r_j\big]$$

Therefore, our problem is reduced to finding the probability that **no collisions occur**. Now to choose some r_j, let $i = 1$ and $j = 2$.

- The probability that $r_1 \neq r_2$ is $\frac{B-1}{B}$.
- The probability that r_3 does not collide with r_1 and r_2 is $\frac{B-2}{B}$.
- The probability that r_4 does not collide with r_1, r_2 and r_3 is $\frac{B-3}{B}$.

Since the random variables are independent of each other therefore the probability that all of them provide no collisions is

$$P = \Big(\frac{B-1}{B}\Big)\Big(\frac{B-2}{B}\Big)\Big(\frac{B-3}{B}\Big)\cdots\Big(\frac{B-(n-1)}{B}\Big)$$

$$P = \prod_{i=1}^{n-1}\Big(\frac{B-i}{B}\Big)$$

$$\text{Thus } P\big[\ \exists i \neq j \mid r_i = r_j\big] = 1 - \prod_{i=1}^{n-1}\Big(\frac{B-i}{B}\Big)$$

Notice that $1 - x \leq e^{-x}$. This is because $e^x = 1 + x + \frac{x^2}{2!} + \ldots$, therefore $e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \ldots$. Hence $e^{-x} = 1 - x + \epsilon$ where $\epsilon > 0$.

$$\prod_{i=1}^{n-1}\Big(\frac{B-i}{B}\Big) \leq \prod_{i=1}^{n-1} e^{\frac{-i}{B}}$$

$$1 - \prod_{i=1}^{n-1}\Big(\frac{B-i}{B}\Big) \geq 1 - \prod_{i=1}^{n-1} e^{\frac{-i}{B}}$$

$$P\big[\ \exists i \neq j \mid r_i = r_j\big] \geq 1 - \prod_{i=1}^{n-1} e^{\frac{-i}{B}}$$

$$P \geq 1 - \Big[e^{\frac{-1}{B}} \cdot e^{\frac{-2}{B}} \cdots e^{\frac{-(n-1)}{B}}\Big]$$

$$P \geq 1 - e^{\frac{-1}{B}\sum_{i=1}^{n-1} i} \tag{1}$$

Note that using the Gaussian Summation Formula, we can bound the value of the sum

$$\sum_{i=1}^{n-1} i = \frac{(n-1)n}{2} \leq \frac{n^2}{2} \tag{2}$$

Now substitute (2) in (1),

$$P \geq 1 - e^{\frac{-n^2}{2B}} \tag{3}$$

Substiuting $n = 1.2 \times B^{1/2}$ in (3)

$$\begin{aligned} P &\geq 1 - e^{\frac{-n^2}{2B}} \\ &\geq 1 - e^{\frac{-0.72B}{B}} \\ &\geq 1 - e^{-0.72} \\ &\geq 1 - 0.486752 \\ &\geq 0.513247 > 0.5 \end{aligned}$$

$$P\big[\ \exists i \neq j \mid r_i = r_j\big] > \frac{1}{2}$$

□

NOTE:
This is called the Birthday Paradox since if $B = 365$ then $n \approx 23$ which means you only need 23 people in a room to get greater than 50% probability of having matching birthdays. However, this only really works if we assume that birthdays are randomlly scattered throughout the year. In reality, this is not the case since birthdays are not identically distributed random variables - there is some skew towards September.

We can alter the initial theorem to find the value of n such that $P > \frac{1}{2}$. A reformulation of the theorme would be

Theorem 3. *Find n such that for n random variables,*

$$P\big[\ \exists i \neq j \mid r_i = r_j\big] > \frac{1}{2}$$

Calculating this would result in $n = 1.2 \times B^{1/2}$

PBKDFs

6. Real-World Hash Functions

There are several well-known CCRHFs built via the Merkle-Damgård Construction. SHA1 and SHA2 have their own documentations but MD5 will be discussed below. SHA3 is omitted.

6.1. MD5

7. Further Reading