해시(hash)란 단방향 암호화 기법으로 해시함수(해시 알고리즘)를 이용하여 "고정된 길이의 암호화된 문자열"로 바꿔버리는 것을 의미합니다.

해시함수(hash function)는 임의의 길이의 데이터를 고정된 길이의 데이터로 매핑하는 함수입니다. 이 때 ①매핑 전 원래 데이터의 값을 키(key), ③매핑 후 데이터의 값을 해시값(hash value), ②매핑하는 과정을 해싱(hashing)이라고 합니다.

SHA256 Hash Generator Online

Generate the SHA256 hash of any string.

passwordsgenerator.net

위의 주소에로 들어가시면 hash값을 만들어주는 사이트에 접속하실 수 있는데요. 입력값이 바뀔 때마다 hash값이 바뀌는 것을 볼 수 있습니다.

해시 알고리즘은 종류가 다양하고 알고리즘마다 hash 길이가 다릅니다. 해시알고리즘은 모두에게 공개되어있습니다. 즉, 해커에게도 공개되어있다는 말이기도 하죠. 그래서 이미 보안이 뚫린 해시 함수가 존재하는데요. MD5, SHA-1, HAS-180은 사용해선 안됩니다. SHA-256, SHA-512 등을 사용하기를 권고하고 있습니다(SHA-512가 더 좋다고 합니다).

해시값이란?

해시값이란 파일의 데이터를 해시 함수(Hash Funtion) 또는 해시 알고리즘(Hash Algorithm)을 통해 계산하여 산출된 값으로 이러한 값은 해당 파일의 고유한 값이 됩니다. 해시값은 고유한 값이기 때문에 서로 다른 두 파일의 해시값이 같다는 것은 두 파일이 완벽하게 100% 일치하는 동일한 파일임을 의미합니다.

아주 쉽게 이야기해서 100 이라는 값을 가진 파일이 있고 해시 함수가 곱하기 2 라고 하였을 때 이 둘을 계산하면 해시값은 200 이 되죠. 또 다른 파일을 같은 해시 함수로 계산해 보았더니 마찬가지로 해시값이 200 이 나왔습니다. 그럼 해당 파일도 100 이라는 값을 가진 파일이라는 의미가 되죠. 결론을 내리면 첫 번째 파일과 두 번째 파일 모두 100 이라는 동일한 데이터를 가진 완전히 똑같은 파일이라고 볼 수 있는 겁니다.

> 파일 1 해시값 200, 파일 2 해시값 200 -> 파일 1 과 파일 2 는 완전히 똑같은 동일한 파일
> 파일 1 해시값 200, 파일 2 해시값 300 -> 파일 1 과 파일 2 는 전혀 다른 파일

이러한 해시값은 주로 파일을 배포하는 측에서 다운로드 받은 파일이 온전한(또는 변조되지 않은) 파일인지를 검사해보라는 목적으로 파일과 함께 제공해주는 것이 보통입니다.

해시함수의 특성 1 : 충돌이 거의 발생하지 않는다

정해진 길이의 결과를 출력하는 해시함수의 특성상 무한대의 입력을 사용할 수 있다면 충돌되는 값을 찾는 것은 100% 가능합니다. 그래서 충돌이 없다(Collision free)라고 말하지 않는 것입니다. 대신 *충돌되는 경우를 찾기가 힘들다*라고 말하는 이유죠.
실제로 충돌이 되는 값을 찾기위해서는 정말, 정말, 정말..... 오랜 시간이 걸립니다.
인류가 지금까지 만들었던 모든 컴퓨터를 다 동원해서 우주의 시작부터 지금까지 시도를 해도 못 찾을 거라고 합니다. 그건 2^256이라는 숫자가 얼마나 큰 숫자인지 보면 좀 더 이해가 됩니다.

해시함수의 특성 2 - 원본 내용을 알 수 없다.

해시함수의 특징은 이 해시결과를 보고서는 원본 메시지가 무엇인지 알아내는 것이 거의 불가능하다는 것입니다.
바로 위에서 함수는 입력 데이터가 다르면 해시 결과가 다르다고 말했습니다. 같은 해시결과를 만드는 값은 반드시 존재하지만 실제로 그 값을 찾는 것은 불가능에 가깝다고 볼 수 있습니다. 다시 말하면 해시 결과를 알려줘도 그 결과를 만드는 입력 데이터를 찾는 일은 불가능하다는 거죠.

해시함수의 특성 3 - 퍼즐 게임을 만들 수 있다

해시 퍼즐 게임이 비트코인 채굴자(마이너)들이 푸는 문제와 같은 종류입니다.
흔히 암호를 푼다고 하는데요.
알고보면 암호를 푼다기 보다는 랜덤하게 값을 입력해서 원하는 해시가 나오는지 확인하는 작업을 반복하는 겁니다. 모래밭에서 금조각 찾기라고 할 수 있는데요. 금속 탐지기나 심지어는 사금 채취하듯 물에 흘려보내면서 찾는 등 어떤 기술 도입을 전혀 할 수 없다는거죠.