해싱과 해시 함수

컴퓨터공학부 김상원

https://youtu.be/RmkEzws7lic

해싱이란

해시 함수란

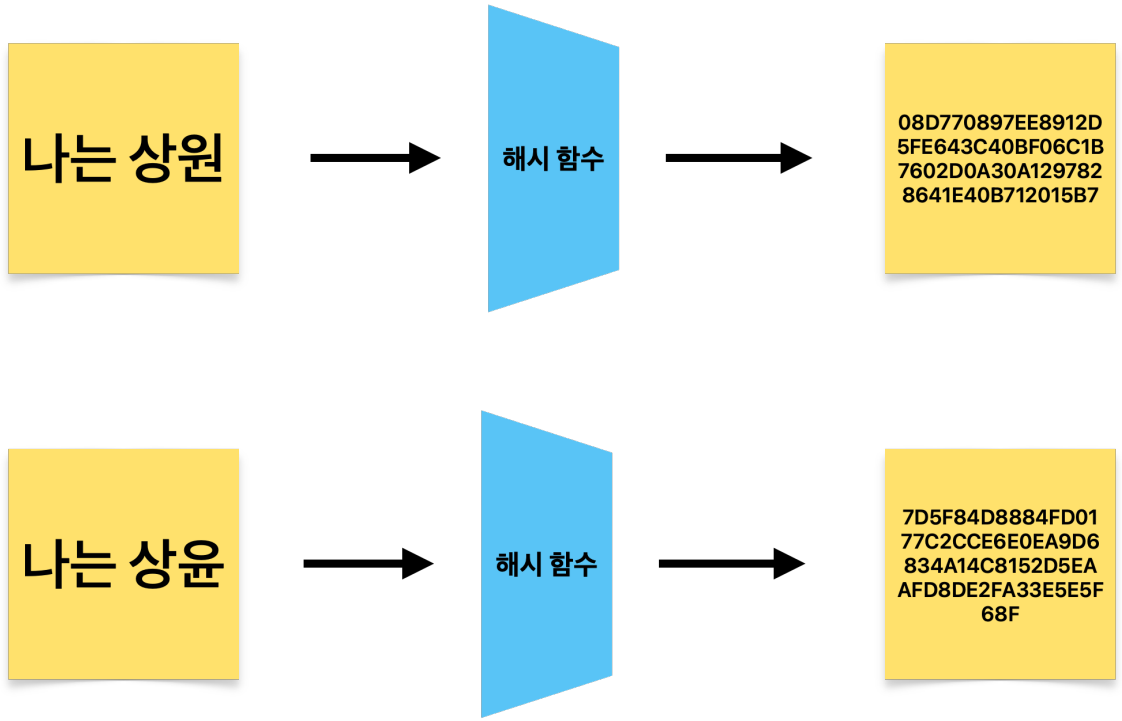
해시 알고리즘의 종류

해시 함수의 한계

해싱이란

키(Key) 값을 해시 함수(Hash Function)라는 수식에 대입시켜 계산한 후
나온 결과를 주소로 사용하여 바로 값(Value)에 접근하게 할 수 하는 방법이다.

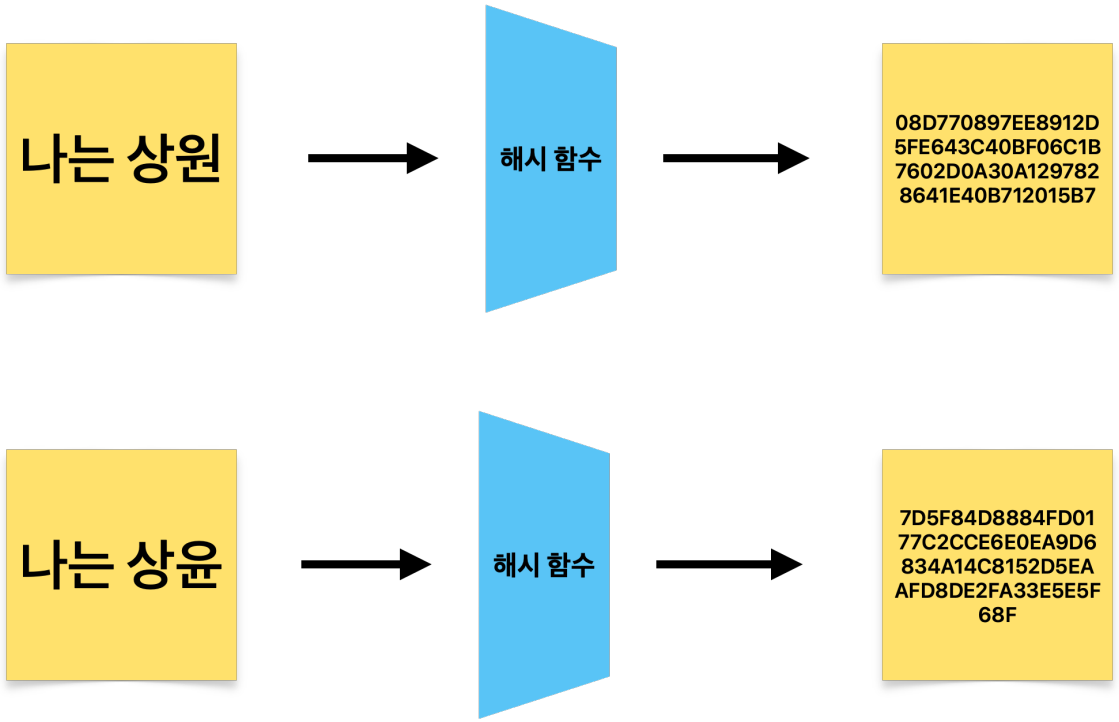
해시테이블 (hash table) : 키를 연산해서 얻은 값으로 직접 접근 가능한 구조
해싱 (hashing) : 해시 테이블을 이용한 탐색



해시 함수 입·출력 동작 원리

해시 함수란

- 해시 함수는 임의의 길이를 갖는 임의의 데이터에 대해 고정된 길이의 데이터로 매핑하는 함수를 말한다.
- 해시 함수를 적용하여 나온 고정된 길이의 값을 해시 값(Hash Value)이라고 한다.

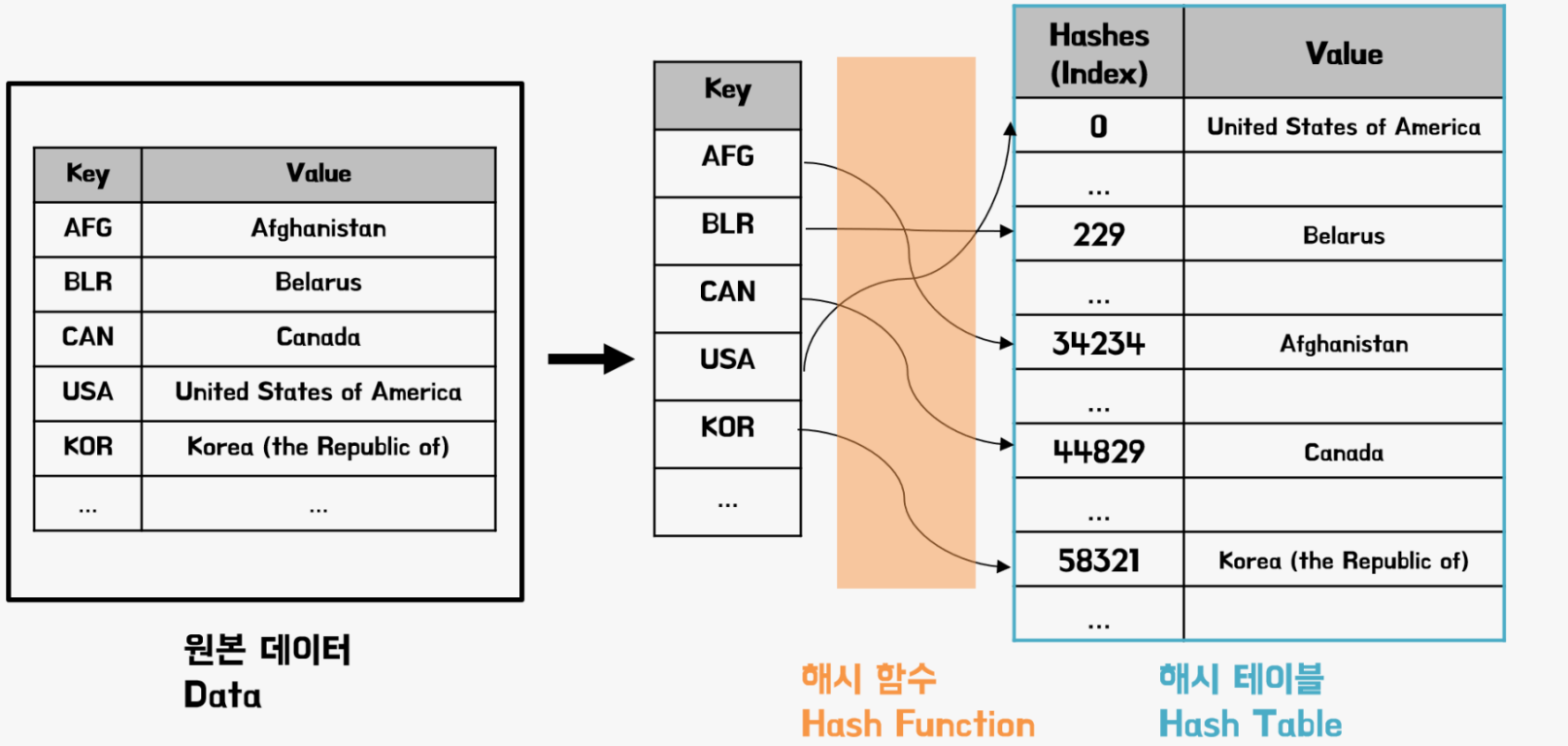


해시 함수 입·출력 동작 원리

해싱 활용

해시 테이블(Hash Table)

- Key, Value 형태인 데이터를 검색이 쉬운 형태로 저장하는 자료구조
- 임의의 길이를 가진 데이터를 고정된 크기로 인덱싱



방대한 자료에서
원하는 자료를 빠르게 검색

해시 충돌

- 해시 함수가 서로 다른 두 개의 입력 값에 대해 동일한 출력 값을 내는 상황을 의미한다. 해시 함수가 무한한 가짓수의 입력 값을 받아 유한한 가짓수의 출력 값을 생성하는 경우, 비둘기집 원리에 의해 해시 충돌은 항상 존재한다.
- 해시 충돌은 해시 함수를 이용한 자료구조나 알고리즘의 효율성을 떨어뜨리며, 따라서 해시 함수는 해시 충돌이 자주 발생하지 않도록 구성되어야 한다. 암호학적 해시 함수의 경우 해시 함수의 안전성을 깨뜨리는 충돌 공격이 가능할 수 있기 때문에 의도적인 해시 충돌을 만드는 것이 어렵도록 만들어야 한다.

암호화 해시 함수

해시함수의 일종으로, 해시 값으로부터 원래의 입력 값과의 관계를 찾기 어려운 성질을 가지는 경우를 의미한다. 암호화 해시 함수가 가져야 하는 성질은 다음과 같다.

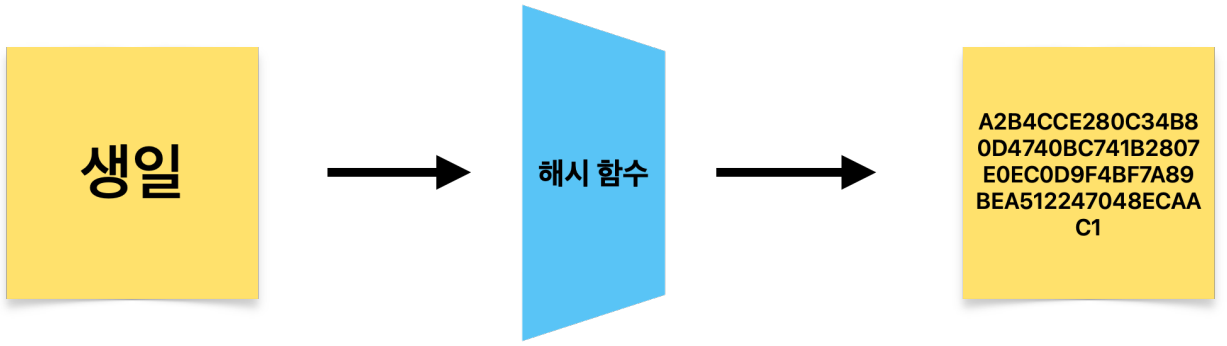
- 역상 저항성 : 주어진 해시 값에 대해, 그 해시 값을 생성하는 입력 값을 찾는 것이 계산상 어렵다. 즉, 제1 역상 공격에 대해 안전해야 한다. 이 성질은 일방향함수와 연관되어 있다.
- 제 2 역상 저항성 : 입력 값에 대해, 그 입력의 해시 값을 바꾸지 않으면서 입력을 변경하는 것이 계산상 어렵다. 제 2 역상 공격에 대해 안전해야 한다.
- 충돌 저항성 : 해시 충돌에 대해 안전해야 한다. 같은 해시 값을 생성하는 두 개의 입력 값을 찾는 것이 계산상 어려워야 한다.

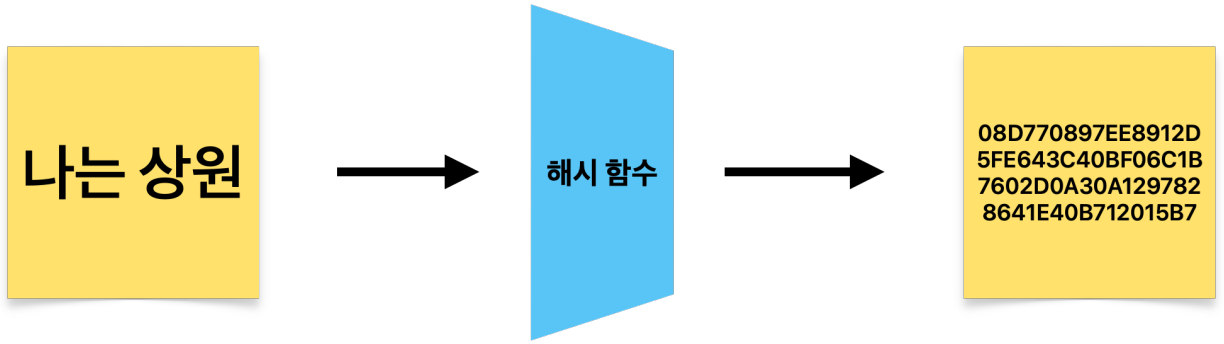
- **제 1 역상 공격**(first preimage attack): 해시값이 주어져 있을 때, 그 해시값을 출력하는 입력값을 찾는다.
- **제 2 역상 공격**(second preimage attack): 입력값이 주어져 있을 때, 그 입력과 같은 해시값을 출력하는 다른 입력값을 찾는다.

해시 함수의 활용

- 블록체인(SHA-256)
- 특징

단방향 변환
눈사태 효과
고정된 길이의 해시 값





해시 알고리즘의 종류

알고리즘	출력 비트 수	내부 상태 크기[c 1]	블록 크기	Length size	Word size	라운드 수	공격 가능성 (복잡도:최대 라운드 수)[c 2]		
							충돌	2차 역상	역상
GOST	256	256	256	256	32	256	2^{105}	2^{192}	2^{192}
HAVAL	256/224/192/160/128	256	1,024	64	32	160/128/96	가능		
MD2	128	384	128	-	32	864	$2^{63.3}$		2^{73}
MD4	128	128	512	64	32	48	3	2^{64}	$2^{78.4}$
MD5	128	128	512	64	32	64	$2^{20.96}$		$2^{123.4}$
PANAMA	256	8,736	256	-	32	-	가능		
RadioGatún	Up to 608/1,216 (19 words)	58 words	3 words	-	1-64	-	2^{352} 또는 2^{704}		
RIPEMD	128	128	512	64	32	48	2^{18}		
RIPEMD-128/256	128/256	128/256	512	64	32	64			
RIPEMD-160	160	160	512	64	32	80	2^{51}:48		
RIPEMD-320	320	320	512	64	32	80			
SHA-0	160	160	512	64	32	80	$2^{33.6}$		
SHA-1	160	160	512	64	32	80	2^{51}		
SHA-256/224	256/224	256	512	64	32	64	$2^{28.5}$:24		$2^{248.4}$:42
SHA-512/384	512/384	512	1,024	128	64	80	$2^{32.5}$:24		$2^{494.6}$:42
Tiger(2)-192/160/128	192/160/128	192	512	64	64	24	2^{62}:19		$2^{184.3}$
WHIRLPOOL	512	512	512	256	8	10	2^{120}:4.5		

가장 널리 사용되는 해시 함수에는 MD5와 SHA-1이 있으나, 이들은 안전하지 않다는 것이 알려져 있다.
미국 US-CERT에서는 2008년 MD5를 사용하지 말아야 한다고 발표했다.
NIST에서는 2008년 SHA-1의 사용을 중지하며 SHA-2를 사용할 것이라고 발표했다.

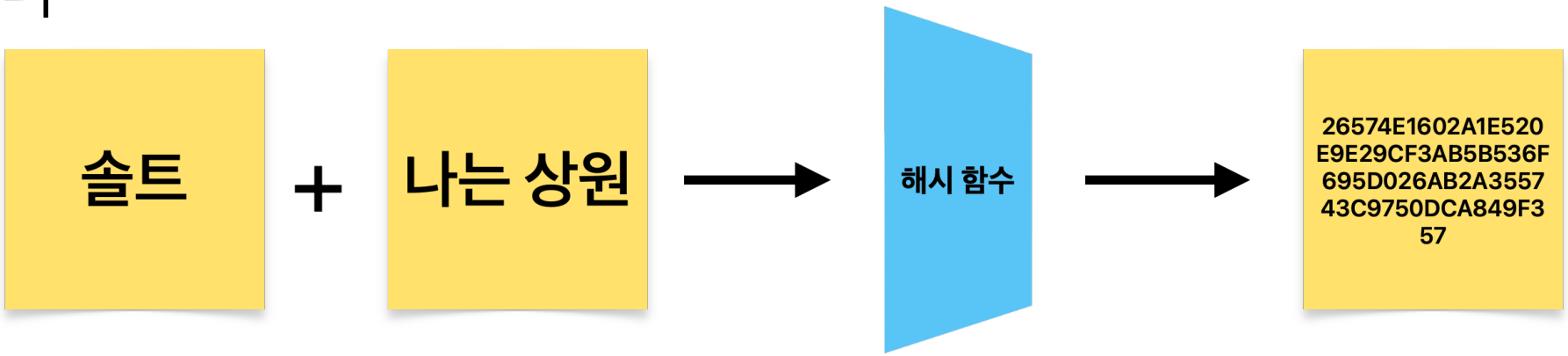
해시 알고리즘의 한계

레인보우 테이블(rainbow table)은 해시함수(MD-5, SHA-1, SHA-2 등)를 사용하여 만들어낼 수 있는 값들을 대량으로 저장한 표이다. 보통 해시함수를 이용하여 저장된 비밀번호로부터 원래의 비밀번호를 추출해 내는데 사용된다.

해시 알고리즘 보완

암호학에서 솔트(salt)는 데이터, 비밀번호, 통과암호를 해시 처리하는 단방향 함수의 추가 입력으로 사용되는 랜덤 데이터이다. 솔트는 스토리지에서 비밀번호를 보호하기 위해 사용된다. 역사적으로 비밀번호는 시스템에 평문으로 저장되지만 시간이 지남에 따라 추가적인 보호 방법이 개발되어 시스템으로부터 사용자의 비밀번호 읽기를 보호한다. 솔트는 이러한 방식의 하나이다.

솔트는 레인보우 테이블과 같은 미리 계산된 테이블을 사용하는 공격을 방어한다



Q & A