해싱(hashing)

https://youtu.be/spNap62g6xA

IT융합공학부 송경주

한성대학교 HANSUNG UNIVERSITY

CryptoCraft LAB
https://crypto.modoo.at

Contents

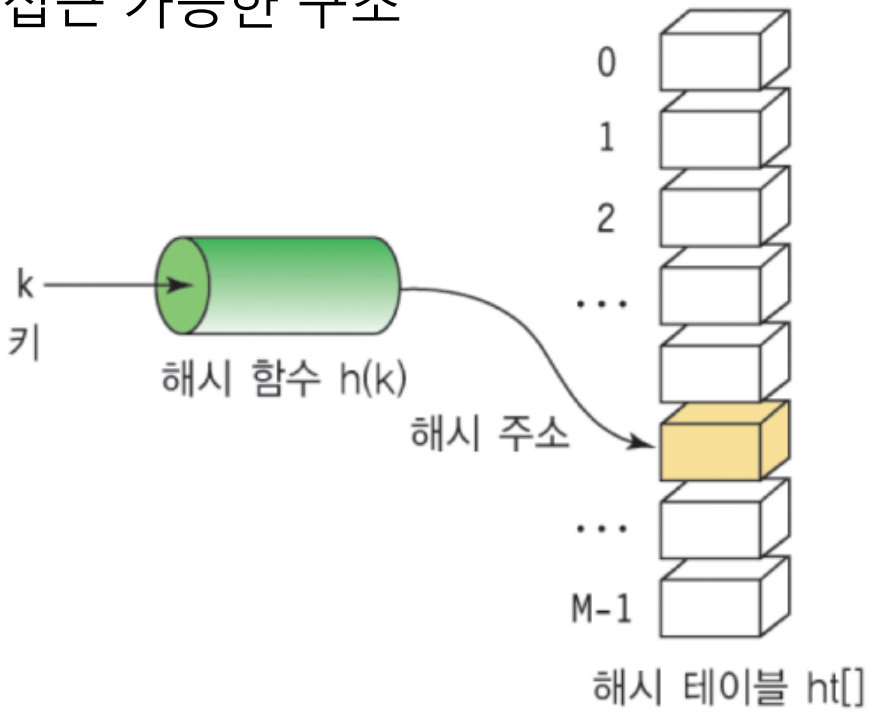
CryptoCraft LAB

해싱이란?

키 값에 산술적인 연산을 적용하여 항목이 저장되어 있는 테이블의 주소를 계산하여 항목에 접근하는 방식.

해시테이블 (hash table) : 키를 연산해서 얻은 값으로 직접 접근 가능한 구조

해싱 (hashing) : 해시 테이블을 이용한 탐색

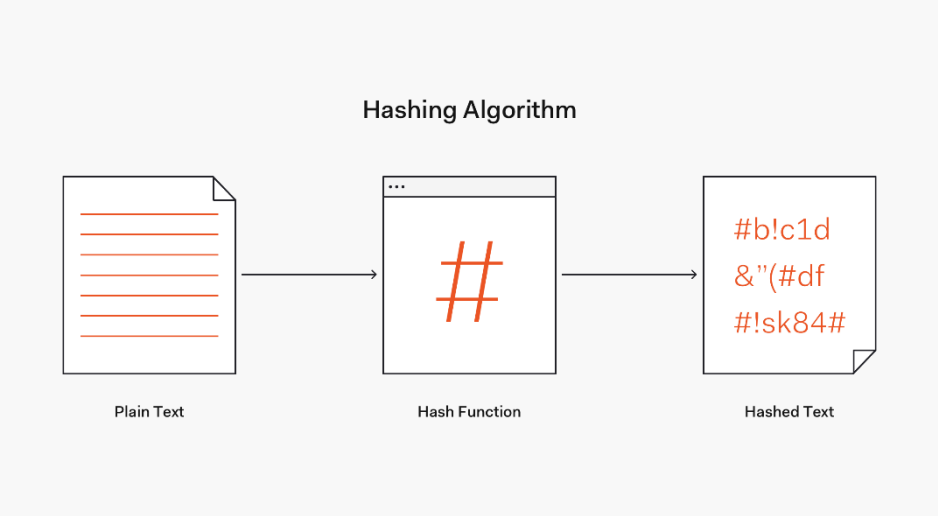


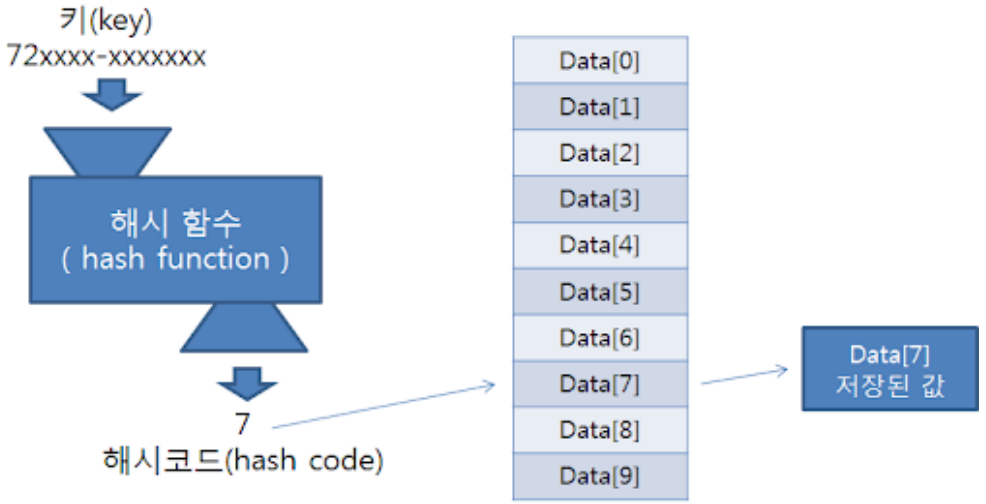
해싱의 구조

해싱에서의 자료 저장방식 : 배열

→특정 항목의 탐색 키만을 가지고 다른 요소에 접근하지 않고 바로 원하는 항목이 저장되어 있는 인덱스를 결정하여 값을 찾음.

해시함수 : 키를 입력으로 받아 해시주소를 생성





해싱의 구조

해시테이블 : M개의 버켓(bucket)으로 이루어짐.

해시함수는 $0 \leq h(k) \leq (M-1)$ 사이의 값을 출력하여 버켓주소를 찾고 값을 제공해야함.

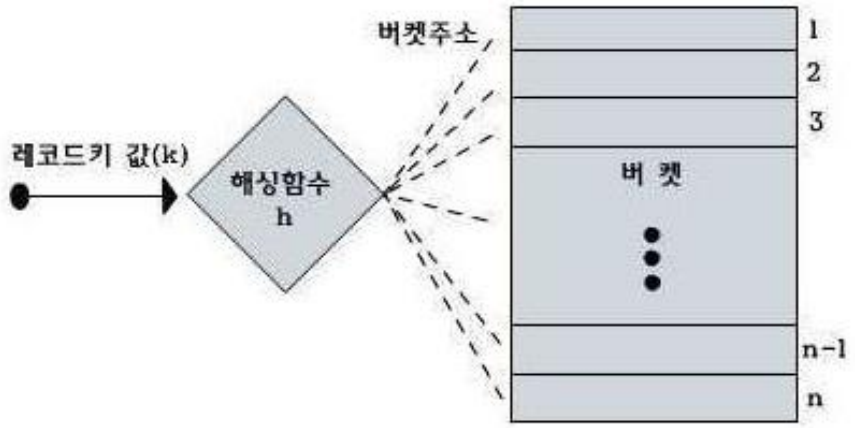
하지만 k의 값보다 버켓의 수가 훨씬 적으므로

다른 키가 해시함수에 의해 같은 해시 주소로 mapping 되는 경우가 발생함

→ h(k1) = h(k2) : 충돌(collision)

충돌이 버켓이 담을 수 있는 양을 넘게되면 오버플로(overflow) 발생

버킷당 담을 수 있는 항목이 1개라면 충돌이 즉 오버플로가 된다.



해시 함수

해시함수 : 특정 키를 입력으로 받아 해시주소를 생성.

좋은 해시함수의 조건

1. 적은 충돌.
2. 해시 함수 값이 해시 테이블 주소 영역 내에서 고르게 분포.
3. 빠른계산.

제산 함수

나머지 연산(mod)를 사용하여 탐색 키를 해시 테이블의 크기(M)로 나눈 나머지를 해시주소로 사용하는 방법.

$$h(k) = k \bmod M$$

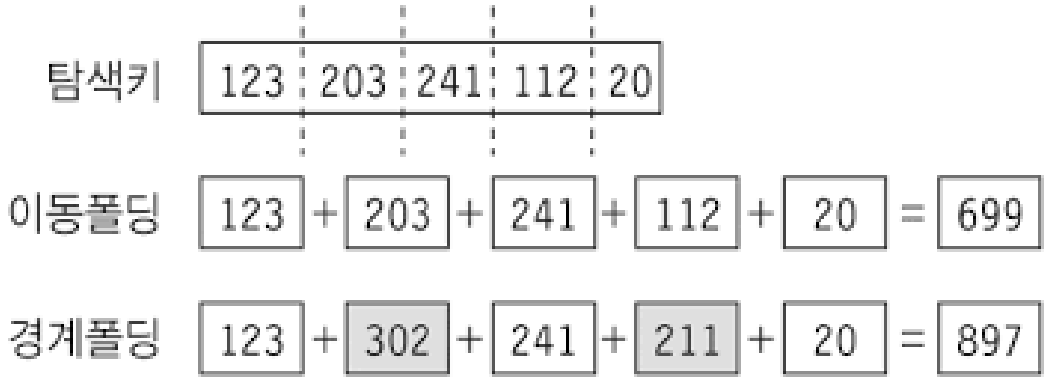
이때, 테이블 크기 M의 값을 소수(prime number)로 선택한다. M이 짝수일때 k가 짝수면 짝수 홀수면 홀수로 분포되므로 2의 배수인 메모리 주소가 한쪽으로 편향될 수 있다.

이 방법을 사용하면 어떠한 키 값이더라도 테이블 크기 범위 안으로 해시주소를 출력할 수 있다.

폴딩 함수

키 값을 여러 부분으로 나눠 모두 더한 값을 해시 주소로 사용하는 방식
(키 값이 해시 테이블의 크기보다 더 큰 정수일 때 사용)

-이동폴딩: 입력 키를 여러 부분으로 나눈 값들을 더하여 해시주소를 얻음
-경계폴딩: 입력 키의 이웃한 부분을 거꾸로 더하여 해시 주소를 얻음



그 외의 해시 주소 추출 방식

중간 제곱 합수

키 값을 제곱한 후 중간의 몇 비트의 취해서 해시 주소를 얻음.

비트 추출 방법

테이블 크기가 $M = 2^k$ 일 때, 탐색 키를 이진수로 간주하여 임의의 위치에서 k개의 비트를 해시 주소로 사용

숫자 분석 방법

키의 각각의 위치에 있는 숫자 중 편중되지 않은 수들을 해시 테이블 크기에 적합한 만큼 조합하여 해시주소로 사용 (수의 특징을 미리 알고 있을 때 유용)

충돌

충돌(collision) : 서로 다른 키가 같은 해시 주소를 갖는 현상.

해결책

-충돌이 일어난 항목을 해시 테이블의 다른 위치에 저장 : 선형조사법 (linear probing)

-해시테이블 하나의 위치에 여러 개의 항목을 저장할 수 있도록 해시테이블 구조 변경 : 체이닝 (chaining)

→보안 관점으로 볼 때의 해시함수는 충돌이 거의 일어나지 않도록 설계되었다. ex) SHA256

(충돌의 발생 확률이 낮을수록 좋은 함수로 평가됨.)

해시(Hash)와 암호화(Encryption)

해시(Hash)

단방향 암호화 기법 (평문을 암호화 함, 복호화X)

암호화(Encryption)

양방향 암호화 기법 (평문을 암호화 및 복호화)

복호화가 필요한 데이터(ex. 통신)에 사용

정보보안과 해시함수

정보보안 영역에서는 해시함수를 기밀성과 무결성을 위해 사용함.

1. 해시값을 통해 이전 입력값을 알아내기 어려움 (복호화 어려움)
2. 긴 입력값을 고정 길이의 값으로 출력 (문서 축약)
3. 하나의 값만이 바껴도 해시값 전체가 바뀜 (위변조 검증)

정보보안과 해시함수

암호화 해시함수가 가져야 하는 성질

역상 저항성: 주어진 해시 값으로, 그 해시 값을 생성하는 입력값을 찾는 것이 계산상 어렵다.

제 1 역상 공격 (해시값이 주어져 있을 때, 그 해시값을 출력하는 입력값 찾기)에 대해 안전 해야함.

제 2 역상 저항성 : 입력 값에 대해, 그 입력의 해시 값을 바꾸지 않으면서 입력을 변경하는 것이 계산상 어렵다.

제 2 역상 공격(입력값이 주어져 있을 때, 그 입력과 같은 해시값을 출력하는 다른 입력값 찾기)에 대해 안전 해야함.

충돌 저항성 : 해시 충돌에 대해 안전해야 함.

Q & A