3. 암호학적 해시함수 요구사항

암호학적 해시함수와 자료구조 분야에서 다루는 해시함수는 관점이 다소 다르다.

① 압축

• **임의 길이**의 이진 문자열을 **고정길이**의 이진 문자열로 변환해야 한다.

② 계산의 용이성

• 입력 **m**이 주어지면, 해시함수 h(**m**)을 계산하기 쉬워야 한다.

③ 일방향성(one-way)

• h(**m**) = x에 대한 h(**m'**) = x를 만족하는 임의의 메시지 **m'**를 찾는 것은 어려워야 한다.

예	• h(12) = 5처럼 해시함수와 해시값 **5**가 주어졌을 때, • h(19) = 5처럼 해시값 **5**를 출력하는 임의의 입력 **19**를 찾는 것은 어려워야 한다.

④ 약한 충돌 저항성(weak collision-resistance)

• **m**이 주어졌을 때, h(**m**) = h(**m'**)인 m'(≠m)를 찾는 것은 계산적으로 어려워야 한다.

• **미리 주어진 입력 m**에 대해서 **m'**을 찾는 것

• 입력 **m**에 대해 같은 해시값을 내는 **다른 입력 m'를 찾는 것**이 계산상 어려워야 한다.

예	• **m** = 12가 미리 주어졌을 때, • h(12) = 12 % 7 = 5 • h(19) = 19 % 7 = 5이므로, h(12) = h(19)이다. • 입력 **12**가 주어졌을 때, **또 다른 입력 19**를 찾아내는 것이 계산상 불가능해야 한다.

⑤ 강한 충돌 저항성(strong collision-resistance) - 약한 충돌 저항성 일반화(조건 완화)

• h(**m**) = h(**m'**)인 서로 다른 임의의 **m**과 **m'**를 찾는 것은 계산적으로 어려워야 한다.

• 순서쌍 **(m, m')**을 찾는 것

• 같은 해시값을 내는 임의의 서로 다른 두 입력 m≠m'를 찾는 것은 계산상 어려워야 한다.

• 강한 충돌 저항성은 약한 충돌 저항성보다 충돌을 찾는 것이 **쉽다**는 뜻이다.

예	h(12) = h(19) 같은 서로 다른 **임의의 순서쌍 (12, 19)**를 찾는 것이 불가능해야 한다.

// 암호학적 해시함수 정리

해시함수가 가져야 할 성질	구체적인 설명(2가지 씩)	성질과 연관된 것
preimage resistance 역상 저항성 프리이미지 저항성	① **임의의 출력**(해시값)에 대한 입력(메시지)을 구하는 것이 계산상 불가능해야 한다. ② 주어진 해시값에 대해, 그 해시값을 생성하는 입력값을 찾는 것이 계산상 어렵다.	일방향함수 (일방향성)
second preimage resistance 제2역상 저항성 제2프리이미지 저항성	① **미리 주어진 입력**(메시지)에 대하여, 같은 출력(해시값)을 내는 또 다른 입력을 찾아내는 것이 계산상 불가능해야 한다. ② 입력값에 대해, 그 입력의 해시값을 바꾸지 않으면서 입력값을 변경하는 것이 계산상 어려워야 한다.	약한 충돌 저항성
collision resistance 충돌 저항성	① **같은 출력(해시값)**을 내는 임의의 서로 다른 두 입력(메시지)을 찾는 것은 계산상으로 불가능해야 한다. ② 같은 해시값을 생성하는 두 개의 입력값을 찾는 것이 계산상 어려워야 한다.	강한 충돌 저항성 ‖ 생일 패러독스 (paradox)

• preimage를 **역상**으로 해석한다.

① 프리이미지 저항성(pre-image resistance) ; 일방향성

• 프리이미지에 대한 공격의 어려움 정도는 2^n에 비례한다.

• 공격자가 50%이상 성공하려면 2^n에 비례하는 수의 다이제스트를 생성해야 한다.

• 여기서, n은 다이제스트 크기(비트 수)이다.

② 제2프리이미지 저항성(second pre-image resistance) ; 약한 충돌 저항성

• 제2프리이미지에 대한 공격의 어려움 정도는 2^n에 비례한다.(프리이미지와 같다)

• 공격자가 50%이상 성공하려면 2^n에 비례하는 수의 다이제스트를 생성해야 한다.

③ 충돌 저항성(collision resistance) ; 강한 충돌 저항성

• 충돌에 대한 공격의 어려움 정도는 $2^{n/2}$에 비례한다.

• 공격자가 50%이상 성공하려면 $2^{n/2}$에 비례하는 수의 다이제스트를 생성해야 한다.

• 충돌 저항성은 제2프리이미지 저항성보다 공격이 **쉽다**는 뜻이다.

역상 공격

역상 공격은 암호화 해시함수에 대한 공격 방식을 말한다.
해시함수에 대한 공격은 역상 공격과 충돌 공격으로 구분하기도 한다.

제1역상 공격 **(first preimage attack)**	• 해시값이 주어져 있을 때, 그 해시값을 출력하는 입력값을 찾는 공격이다. • 예를 들면, 12 % 7 = 5에서 • 5가 주어졌을 때, 5를 출력하는 **임의의 입력값**을 찾는 공격 • 제1역상 공격에 안전한 것을 **역상 저항성**이라 한다. • 역상 저항성을 **일방향성**이라 한다.
제2역상 공격 **(second preimage attack)**	• 입력값이 주어져 있을 때, 그 **입력과 같은 해시값**을 출력하는 다른 입력값을 찾는 공격이다. • 예를 들면, 12 % 7 = 5에서 • 12가 주어졌을 때, 12처럼 5를 출력하는 다른 값을 찾는 공격 • 제2역상 공격에 안전한 것을 **제2역상 저항성**이라 한다. • 제2역상 저항성을 **약한 충돌 저항성**이라 한다.

• 제2역상 공격은 제1역상 공격에서 **원본 메시지**가 주어져 있는 상황이다.

// 약한 충돌 저항성에 보안 취약성이 있을 경우
• 메시지를 위조하여도 검사자는 위조 여부를 확인할 수 없다.(치명적인 일)
• 따라서, 공인인증시스템 뿐만 아니라 보안과 관련된 전 분야에 심각한 위협 초래

// 충돌 공격

• 충돌 공격은 해시 충돌이 일어나는 **임의의 두 입력값**을 찾는 공격이다.
• 충돌 공격이 역상 공격과 다른 점은 해시함수의 출력값이 고정되어 있지 않다.

• 충돌 공격에 안전한 것을 **강한 충돌 저항성**이라 한다.

• 역상 공격은 충돌 공격보다 더 어렵다.
• 충돌 공격이 역상 공격보다 더 쉽다.

해시함수 공격 알고리즘

◉ 제1역상 공격 – 일방향성

```
제1역상_attack(d)          //d는 메시지 다이제스트, d=h(M)은 공격자가 가지는 정보
{
  for(k=1 to n)            //임의의 메시지 M 구하기, n은 다이제스트 크기이다.
  {
    create(M[k]);          //임의의 문서 생성
    imsi = h(M[k]);        //생성된 문서에 대한 해시값
    if(d == imsi) return M[k]; //공격 성공, 공격자가 구하려고 임의의 문서 M
  }
  return 0;                //공격 실패
}
```

• 제1역상 공격은 preimage 공격이라고도 한다.

◉ 제2역상 공격 – 약한 충돌 저항성

```
제2역상_attack(d, M)       //d는 메시지 다이제스트(해시값), M은 메시지
{
  for(k=1 to n-1)          //또 다른 메시지 M'를 구하므로 반복이 n-1까지 이다.
  {
    create(M[k]);          //임의의 문서 생성
    imsi = h(M[k]);        //생성된 문서에 대한 해시값
    if(d == imsi) return M[k]; //공격 성공, 공격자가 구하려고 하는 문서 M'(≠M)
  }
  return 0;                //공격 실패
}
```

• 제2역상 공격은 제1역상 공격에서 **원본 메시지 M**이 추가로 주어져 있는 상황이다.

기출문제 분석

1. 해시함수(hash function)에 대한 설명으로 옳지 않은 것은? [2014년 국가 9급]

① 임의 길이의 문자열을 고정된 길이의 문자열로 출력하는 함수이다.
② 대표적인 해시함수는 MD5, SHA-1, HAS-160 등이 있다.
③ 해시함수는 메시지 인증과 메시지 부인방지 서비스에 이용된다.
④ 해시함수의 충돌 회피성은 동일한 출력을 산출하는 서로 다른 두 입력을 계산적으로 찾기 가능한 성질을 나타낸다.

♤ **해시함수의 충돌 회피성(collision resistance)**

- 충돌 회피성은 같은 출력(해시값)을 산출하는 임의의 서로 다른 두 입력(메시지)을 찾는 것은 계산상으로 **불가능**해야 한다는 것이다.
- 예 : $12 \equiv 19 \pmod 7$

정답 : ④

2. 해시 알고리즘의 특징에 대한 설명으로 가장 옳은 것은? [2018년 서울 7급]

① 해시값이 같으면서 입력 값이 서로 다른 충돌 쌍을 찾는 것은 계산상 불가능하다.
② 고정길이의 입력 메시지를 임의 길이의 출력 값으로 압축시킨 함수이다.
③ 주어진 해시값 y에 대해 hash(x)=y 식을 만족하는 x를 찾는 것이 계산적으로 가능하다.
④ 메시지의 거대화 방지 및 데이터의 은닉에 사용된다.

♤ **해시 알고리즘**

② 고정길이의 입력 메시지를 임의 길이의 출력 값으로 압축시킨 함수이다.**(×)**
→ 임의 길이의 입력 메시지를 고정길이의 출력 값으로 압축시킨 함수이다.
③ 주어진 해시값 y에 대해 hash(x)=y 식을 만족하는 x를 찾는 것이 계산적으로 가능하다.**(×)**
→ 계산적으로 불가능해야 한다.
④ 메시지의 거대화 방지 및 데이터의 은닉에 사용된다.**(×)**
→ 해시 알고리즘과는 무관한 내용이다.
→ 데이터의 은닉은 스테가노그래피(steganography) 기술이다.

정답 : ①

3. 다음에서 설명하는 해시함수(H)의 특성은? [2019년 국가 7급]

주어진 메시지 x에 대해,
H(y) = H(x)를 만족하면서 y≠x 인 y를 찾는 것이 계산상 매우 어려워야 한다.

---

① 의사난수성(pseudo-randomness)
② 역상 저항성(pre-image resistance)
③ 약한 충돌 저항성(weak collision resistance)
④ 강한 충돌 저항성(strong collision resistance)

♤ **해시함수(H)의 특성**

◈ **일방향성(one-way)**

• h(**m**) = x에 대한 h(**m'**) = x를 만족하는 임의의 메시지 **m'**를 찾는 것은 어려워야 한다.

예	• h(12) = 5처럼 해시함수와 해시값 **5**가 주어졌을 때, • h(19) = 5처럼 해시값 **5**를 출력하는 임의의 입력 **19**를 찾는 것은 어려워야 한다.

◈ **약한 충돌 저항성(weak collision-resistance)**

• **m**이 주어졌을 때, h(**m**) = h(**m'**)인 m'(≠m)를 찾는 것은 계산적으로 어려워야 한다.
• **미리 주어진 입력 m**에 대해서 **m'**을 찾는 것
• 입력 **m**에 대해 같은 해시값을 내는 **다른 입력 m'를 찾는 것**이 계산상 어려워야 한다.

예	• **m** = 12가 미리 주어졌을 때, • h(12) = 12 % 7 = 5 • h(19) = 19 % 7 = 5이므로, h(12) = h(19)이다. • 입력 **12**가 주어졌을 때, **또 다른 입력 19**를 찾아내는 것이 계산상 불가능해야 한다.

◈ **강한 충돌 저항성(strong collision-resistance) – 약한 충돌 저항성 일반화(조건완화)**

• h(**m**) = h(**m'**)인 서로 다른 임의의 **m**과 **m'**를 찾는 것은 계산적으로 어려워야 한다.
• 순서쌍 **(m, m')**을 찾는 것
• 같은 해시값을 내는 임의의 서로 다른 두 입력 m≠m'를 찾는 것은 계산상 어려워야 한다.
• 강한 충돌 저항성은 약한 충돌 저항성보다 충돌을 찾는 것이 **쉽다**는 뜻이다.

예	h(12) = h(19) 같은 서로 다른 **임의의 순서쌍 (12, 19)**를 찾는 것이 불가능해야 한다.

정답 : ③

4. 보안 해시함수가 가져야 하는 성질 중 하나인 강한충돌저항성(strong collision resistance)에 대한 설명으로 옳은 것은? [2015년 지방 9급]

① 주어진 해시값에 대해, 그 해시값을 생성하는 입력값을 찾는 것이 어렵다.
② 주어진 입력값과 그 입력값에 해당하는 해시값에 대해, 동일한 해시값을 생성하는 다른 입력값을 찾는 것이 어렵다.
③ 같은 해시값을 생성하는 임의의 서로 다른 두 개의 입력값을 찾는 것이 어렵다.
④ 해시함수의 출력은 의사난수이어야 한다.

♤ **강한충돌저항성(strong collision resistance)**

① 주어진 해시값에 대해, 그 해시값을 생성하는 입력값을 찾는 것이 어렵다.**(×)**
→ **일방향성**에 대한 설명
② 주어진 입력값과 그 입력값에 해당하는 해시값에 대해, 동일한 해시값을 생성하는 다른 입력값을 찾는 것이 어렵다.**(×)** → **약한충돌저항성**에 대한 설명
④ 해시함수의 출력은 **의사난수이어야 한다.(×)** → 해시함수의 출력은 **의사난수일 필요가 없다.**

정답 : ③

5. 〈보기〉에서 설명하는 해시함수(H)의 특성으로 옳은 것은? [2017년 국회 9급]

주어진 메시지 x에 대해, H(x)=H(y)인 x≠y를 만족하는
두 개의 메시지 x, y를 찾는 것이 어려울 때, 해시함수가 이 성질을 가지고 있다고 한다.

① Second Pre-image Resistance
② Collision Resistance
③ Integrity
④ Onewayness
⑤ Uniform Distribution

♤ **second preimage resistance(제2프리이미지 저항성, 제2역상 저항성, 약한 충돌 저항성)**

- **m**이 주어졌을 때, h(**m**) = h(**m'**)인 m'(≠m)를 찾는 것은 계산적으로 어려워야 한다.
- **미리 주어진 입력 m**에 대해서 **m'**을 찾는 것
- 입력 **m**에 대해 같은 해시값을 내는 **다른 입력 m'를 찾는 것**이 계산상 어려워야 한다.

예	• **m** = 12가 미리 주어졌을 때, • h(12) = 12 % 7 = 5 • h(19) = 19 % 7 = 5이므로, h(12) = h(19)이다. • 입력 **12**가 주어졌을 때, **또 다른 입력 19**를 찾아내는 것이 계산상 불가능해야 한다.

정답 : ①

6. 다음 중 암호학적 해시함수(h)에 대한 설명으로 가장 옳지 않은 것은? [2022년 군무원 7급]

① 주어진 출력 y에 대하여 h(x)=y를 만족하는 x를 구하기 어렵다.
② 임의 메시지에 대하여 동일한 해시값을 가지는 메시지가 없다.
③ 동일한 해시값을 가지는 서로 다른 x와 x'을 구하기 어렵다.
④ 주어진 입력 x에 대하여 h(x')=h(x), x'≠x를 만족하는 x가 아닌 x'을 구하기 어렵다.

♤ **해시함수**

· **임의 메시지**에 대하여 **동일한 해시값**을 가지는 메시지가 **없다.(×)**
→ **동일한 해시값**을 가지는 메시지는 많을 수 **있다.**
→ **예** : h(7) mod 5 = **2**, h(12) mod 5 = **2**, h(17) mod 5 = **2**, h(22) mod 5 = **2**, …

정답 : ②