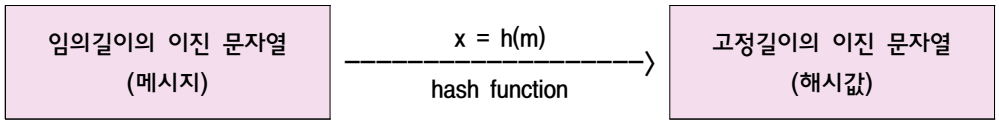
제4장 암호학적 해시함수

1. 해시함수 개요

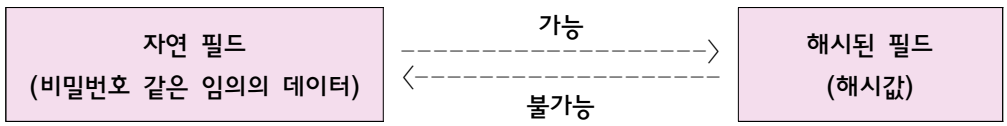
영어 단어 hash는 '자료에서 뺄 수 없는 토막을 긁어 모은다'는 뜻을 가진다.



- 해시함수의 계산 결과를 "해시값, 메시지 지문, 메시지 **다이제스트(digest)**"이라 한다.
- 해시함수는 임의 길이의 이진 문자열을 **고정길이**의 이진 문자열로 매핑하는 함수이다.
- 해시함수는 임의 길이의 메시지를 **대신**할 수 있는 고정길이의 값을 계산하는 함수이다.
- 해시함수는 원래 메시지에서 1bit만 변경되어도 해시값은 매우 크게 달라진다.**(눈사태 효과)**

◈ **해시함수 특징**

① 변환은 한 쪽 방향뿐이다.(일방향 함수)



- 해시함수가 암호화 기술로 사용될 때, 암호화된 정보가 복호화 될 수 **없는 이유**이다.
- 해시함수의 결과 값은 복원할 필요가 없다. 복원이 가능해서도 안 된다.

② 서로 다른 자연 필드가 동일한 값으로 해시될 수 있다.(충돌, collision)

- h(30) = 30 % 7 = 2
- h(37) = 37 % 7 = 2

③ 충돌이 많을수록 다른 데이터를 구별하기 어려워지고, 검색 등의 비용은 증가한다.

④ 2개의 해시값이 서로 다르면, 그 해시값에 대한 원래 데이터는 무언가 서로 다르다.
• 해시함수 성질이 '결정적'이기 때문이다. '결정적'은 예측한 결과가 도출되는 것이다.
• 역은 성립하지 않는다.

⑤ 동일한 해시값을 가지면, 원래의 데이터가 서로 같을 수 있다는 것을 암시할 뿐이다.
• 동일한 해시값을 가진다고, 원래의 데이터가 서로 같다는 것을 보장하지는 않는다.
• 동일한 해시값을 가지는 동의어(synonym)는 여러 개 존재할 수 있다.

⑥ 키를 사용하지 않는 해시함수는 같은 입력에 대해서 항상 같은 출력이 나온다.
• 해시함수의 **주목적은 메시지 무결성**을 제공하기 위한 것이다.
• 메시지에 대한 해시값(증거값)을 추출하여 메시지의 오류나 변조를 탐지할 수 있다.

⑦ 키를 사용하는 해시함수는 특별한 경우이다.(MAC 해시함수 – 뒤에서 설명)

기출문제 분석

1. 암호학적 해시함수가 가져야 할 특성으로 옳지 않은 것은? [2016년 지방 9급]

① 서로 다른 두 입력 메시지에 대해 같은 해시값이 나올 가능성은 있으나, 계산적으로 같은 해시값을 갖는 서로 다른 두 입력 메시지를 찾는 것은 불가능해야 한다.
② 해시값을 이용하여 원래의 입력 메시지를 찾는 것은 계산상으로 불가능해야 한다.
③ 입력 메시지의 길이에 따라 출력되는 해시값의 길이는 비례해야 한다.
④ 입력 메시지와 그 해시값이 주어졌을 때, 이와 동일한 해시값을 갖는 다른 메시지를 찾는 것은 계산상으로 불가능해야 한다.

♤ **암호학적 해시함수**

• 입력 메시지의 길이에 따라 출력되는 해시값의 길이는 비례해야 한다.**(×)**
→ 임의 길이의 이진 문자열을 **고정길이**의 이진 문자열로 변환해야 한다.

정답 : ③

2. 메시지 인증에 사용되는 해시함수의 요건으로 옳지 않은 것은? [2018년 국가 9급]

① 임의 크기의 메시지에 적용될 수 있어야 한다.
② 해시를 생성하는 계산이 비교적 쉬워야 한다.
③ 다양한 길이의 출력을 생성할 수 있어야 한다.
④ 하드웨어 및 소프트웨어에 모두 실용적이어야 한다.

♤ **해시함수 요건**

- 다양한 길이의 출력을 생성할 수 있어야 한다.(×)
  → 임의 길이의 이진 문자열을 **고정길이**의 이진 문자열로 변환해야 한다.

정답 : ③

3. 해시함수의 충돌에 대한 설명으로 옳은 것은? [2019년 지방 9급]

① 해시함수의 입력 메시지가 길어짐에 따라 생성되는 해시값이 길어지는 것을 의미한다.
② 서로 다른 해시함수가 서로 다른 입력 값에 대해 동일한 출력 값을 내는 것을 의미한다.
③ 동일한 해시함수가 서로 다른 두 개의 입력 값에 대해 동일한 출력 값을 내는 것을 의미한다.
④ 동일한 해시함수가 동일한 입력 값에 대해 다른 출력 값을 내는 것을 의미한다.

♤ **해시함수의 충돌**

- 충돌은 동일한 해시함수가 서로 다른 2개의 입력에 대해 동일한 출력을 내는 것이다.

정답 : ③

4. 다음 중 문서의 무결성 비교를 위하여 사용되는 해시값(함수)의 성질에 대한 설명으로 가장 옳지 않은 것은? [2022년 군무원 9급]

① 입력되는 가변의 데이터에 대해서 고정길이의 해시값이 발생한다.
② 입력되는 데이터가 다르면 해시값도 다르다.
③ 해시값 복호화를 위해서는 대칭키 알고리즘만 가능하다.
④ 해시값으로부터 원래의 데이터 복구가 불가능하다.

♤ **해시값(함수)**

- 해시값 복호화를 위해서는 대칭키 알고리즘만 가능하다.(×)
  → 해시값은 복호화가 **불가능**하다.**(해시함수의 일방향성)**

정답 : ③