2017년 12월

17ZH1700-01-2110P

10

암호화된 데이터베이스에서의 데이터 저장 및 검색을 위한 암호 원천 기술 개발

Development of Storage and Search Technologies over Encrypted Database

한국전자통신연구원
Electronics and Telecommunications Research Institute

인 사 말 씀

최근 ICT 기술의 발전과 스마트 단말기의 확산으로 촉발된 스마트 혁명은 사회적, 경제적, 문화적 생활 전반에 걸쳐 매일 매일 다양한 경험을 우리에게 새롭게 선사하고 있습니다. 스마트 환경에서 사용자는 소형화되고 경량화된 스마트 기기들을 통해 보다 다양한 데이터 서비스를 장소와 시간에 구애받지 않고 자유롭게 제공 받을 수 있을 것으로 전망됩니다.

그러나 이에 따른 역기능 또한 함께 커져가고 있어 사용자 데이터가 집중되는 데이터베이스에서의 빈번한 개인정보 유출은 국내외적으로 큰 우려를 낳고 있으며, 이를 근본적으로 방지하기 위한 데이터베이스 암호화에 대한 관심도 커지고 있습니다. 우리 정부 또한 정책적으로 중요 데이터에 대한 암호화를 의무화하고 있으며, 데이터 암호화 대상 또한 점차 확대하여 궁극적으로 모든 데이터가 암호화될 것으로 예상되고 있습니다. 하지만, 기존의 데이터 암호화 기술 적용에 따른 성능적/기능적 제약으로 인해 데이터베이스 암호화 적용은 현재 미진한 실정이며 향후 전망 또한 부정적인 상황으로, 이를 해결하기 위한 새로운 기술 개발이 요구되고 있습니다.

이에 본 과제에서는 데이터베이스 암호화에 따른 데이터 기밀성을 유지하면서, 암호화된 데이터에 대한 활용을 극대화 할 수 있는 암호 원천 기술에 대한 연구를 수행합니다. 특히, 암호화된 데이터베이스 활용을 위한 기본이 되는 데이터 저장, 열람 및 검색 기술을 주요 연구 목표로 설정하였습니다. 본 과제의 연구 결과들은 다양한 정보 통신 서비스의 기반 프리미티브 기술로 활용될 것으로 기대되고 있으며, 또한 다양한 신규 정보 통신 서비스의 신뢰성을 향상시켜 관련 사업의 활성화 및 국가 경쟁력 강화에 기여할 것으로 확신합니다.

끝으로 연구개발 과제에 참여한 연구원 및 공동연구 기관 관계자 여러분들의 노고를 치하하는 바입니다. 앞으로 여러분 개개인의 열정으로 개발된 연구 결과물이 우리나라 정보통신 및 정보보호 기술 발전에 큰 기여가 있기를 기대합니다.

2017 년 12 월

한국전자통신연구원 원장 이 상 훈

제 출 문

본 연구보고서는 주요사업인 “암호화된 데이터베이스에서의 데이터 저장 및 검색을 위한 암호 원천 기술 개발”의 결과로서, 본 과제에 참여한 아래의 연구팀이 작성한 것입니다.

2017 년 12 월

주관연구기관 : 한국전자통신연구원

연구책임자 : 책임연구원 장구영(정보보호연구본부)

연구참여자 : 책임연구원 김건우(정보보호연구본부)

책임연구원 이상수(정보보호연구본부)

선임연구원 윤택영(정보보호연구본부)

선임연구원 조남수(정보보호연구본부)

공동연구기관 : 공주대학교 산학협력단

연구책임자 : 책임연구원 홍도원

연구참여자 : 책임연구원 서창호

연구원 강명모

연구원 김현일

연구원 남승수

연구원 류권상

연구원 박철희
연구원 윤보람
연구원 홍은주
연구원 정수용
연구원 박소희

요 약 문

Ⅰ. 제 목

암호화된 데이터베이스에서의 데이터 저장 및 검색을 위한 암호 원천 기술 개발

Ⅱ. 연구목적 및 중요성

가. 연구개발의 목적

데이터 유출의 원천적인 방지를 위해 데이터베이스가 암호화된 상태로 데이터의 저장, 열람 및 검색이 가능한 암호 원천 기술 개발

나. 연구개발의 중요성

- ㅇ 기존의 방어 체계를 우회하는 지능적인 공격의 발달로 공격의 목표가 되는 데이터를 원천적으로 보호하기 위한 암호 기술에 대한 중요성이 증가하고 있음
- ㅇ 하지만, 데이터베이스 암호화에 따른 기능적/성능적 제약이 암호화 사용에 걸림돌로 작용하고 있음
- ㅇ 데이터 유출의 원천적인 방지를 위한 데이터베이스 암호화를 통한 기밀성 보장을 기반으로 암호화된 데이터 활용을 위한 암호 원천 기술 개발이 중요함

Ⅲ. 연구내용 및 범위

본 과제는 2015년부터 2017년까지 3년간 진행되며 최종 목표는 다음과 같다.

○ 데이터 유출의 원천적인 방지를 위해 데이터베이스가 암호화된 상태로 데이터 저장, 열람 및 검색이 가능한 암호 원천 기술 개발
- 암호데이터 중복 처리 기술 개발
- 암호데이터 소유권 검증 기술 개발
- 암호데이터 검색 기술 개발

연도별 내용 및 범위는 다음과 같다.

가. 1차년도(2015년) : 암호데이터 저장 및 검색을 위한 핵심 프리미티브 설계

○ 암호데이터 중복 처리를 위한 핵심 설계 논리 개발
- Message-locked encryption 최신 기술 및 요구 사항 분석
- 중복 처리에 따른 데이터 손실 공격 분석 및 암호데이터 중복 처리 기술 안전성 모델 연구
- 메시지 기반 암호화 핵심 설계 논리 개발

○ 암호데이터 검색을 위한 핵심 알고리즘 설계
- 암호데이터 검색 최신 기술 및 요구 사항 분석
- 수용 가능 안전성 모델 연구
- 암호데이터 키워드 검색 알고리즘 설계

나. 2차년도(2016년) : 암호데이터 저장, 열람 및 검색 기술 설계 및 안전성 검증

○ 암호데이터 중복 처리 기술 설계
- 파일 단위 암호데이터 중복 처리 기술 설계
- 암호데이터 중복 처리 안전성 검증
- 암호데이터 중복 처리 성능 최적화

○ 암호데이터 소유권 검증 모델 연구
- 적용 환경 분석을 통한 요구 사항 정의
- 데이터 기반의 소유권 부여, 유지 및 검증을 위한 관리 모델 설계

○ 동적 환경을 위한 암호데이터 검색 기술 개발
- 동적 암호데이터 검색 기술 요구 사항 분석 및 안전성 모델 정립
- 데이터 추가, 삭제 기능을 제공하는 암호데이터 검색 기술 개발 및 안전성 검증

다. 3차년도(2017년) : 암호데이터 저장, 열람 및 검색 기술 개발

○ 암호데이터 중복 처리 기술 개발
- 블록 단위 암호데이터 중복 처리 기술 개발
- 다중 사용자 기반 암호데이터 중복 처리 기술 개발
- 암호데이터 중복 처리 기술 성능 분석 및 개선 연구

○ 암호데이터 소유권 검증 기술 개발
- 데이터 기반의 암호데이터 소유권 검증 기술 설계

○ 암호데이터 검색 실용화 기술 개발
- 부가 기능 제공 암호데이터 검색 기술 개발
- 수용 가능 안전성 기반의 암호데이터 검색 기술 최적화

Ⅳ. 연구결과

본 과제의 수행을 통하여 확보된 주요 연구개발 결과는 아래와 같다.

○ 암호데이터 중복 처리 기술 개발

- 메시지 기반의 Client-side 암호데이터 중복 처리 기술 개발
  - 메시지 기반 암호화 프리미티브 설계
  - Client-side 암호데이터 중복 처리 프로토콜 개발
  - 암호데이터 중복 처리 SW 개발
- 파일 신원 확인 공격에 안전한 암호데이터 중복 처리 기술 개발

○ 암호데이터 소유권 검증 기술 개발

- 데이터 기반의 소유권 확인 기술 개발
- 데이터 기반의 저장 데이터 검증 기술 개발

○ 암호데이터 검색 기술 개발

- 링크드 체인 구조 기반 확장 검색 가능 암호데이터 검색 프리미티브 알고리즘 및 이를 바탕으로 한 범위 검색 기술 설계
- Bloom Filter Tree 기반 동적 암호데이터 검색 기술 설계

○ 논문

- SCI(E) 논문 10건 게재 및 4건 게재 승인

○ 특허

- 국내 특허 : 출원 7건, 제출 3건(출원 중)
- 국제 특허 : 출원 3건, 제출 3건(출원 중)

Ⅴ. 연구개발결과의 활용 계획

- ○ 금융, 의료, 교육 등 공공 부문의 데이터를 위탁 활용하는 서비스 환경을 위한 데이터 암호화 및 저장/검색 기술로 활용
- ○ 암호데이터에 대한 저장 및 검색 기능을 평문과 유사한 수준으로 제공함으로써 현재 평문 데이터 대상으로만 제공되는 서비스 영역을 암호데이터로 확장, 나아가서 데이터 서비스 전체의 신뢰성을 향상시키기 위한 원천 기술로 활용
- ○ 암호데이터 중복 처리 기술은 암호데이터 저장 단계에서 스토리지 및 네트워크 비용 절감을 위한 원천 기술로 활용
- ○ 클라우드/빅데이터 환경의 동적으로 변화하는 데이터를 활용하는 환경의 데이터 보호를 위한 핵심 기술로 활용
- ○ 안전한 초연결 서비스 보안 인프라 구축 및 신규 지능형 사이버 보안 시장 창출에 활용

Ⅵ. 기대성과

- ○ 기존 보안 기술이 지니는 한계를 극복한 새로운 데이터 중심 보안으로의 변화를 선도하기 위한 핵심 원천 기술로 활용
- ○ 클라우드/빅데이터 서비스 확산에 요구되는 프라이버시 보호 관련 기술의 고도화 견인
- ○ 미래 IoT 환경에서 예상되는 대규모 데이터에 대한 저장 시스템의 안전성과 저장 성능 향상을 위한 원천 기술 선점 가능
- ○ 암호화된 데이터베이스 활용 기술의 고도화를 통해 정보의 기밀성과 활용성을 동시에 제공하여 그 동안 구축하기 힘들었던 공공데이터 활용 서

비스 산업의 조속한 활성화 및 이에 따른 고용 촉진을 기대할 수 있음

- ○ 클라우드/빅데이터 환경에서 데이터 활용을 기반으로 한 서비스 확대 및 새로운 보안 시장 창출을 견인할 수 있음
- ○ 국내외에서 발생하고 있는 개인정보 유출에 따른 사회적 피해 규모는 수치로 표현할 수 없을 정도로 심각하며, 신용 사회의 근간을 위협한다는 점에서 데이터 유출 방지를 위한 원천 기술 개발은 파급 효과가 매우 높음
- ○ 해킹이나 내부 공모자에 의한 데이터 유출의 근본적인 해결책을 제공함으로써, 관련 분야의 사건 발생에 따르는 사회적 피해 복구 비용 절감

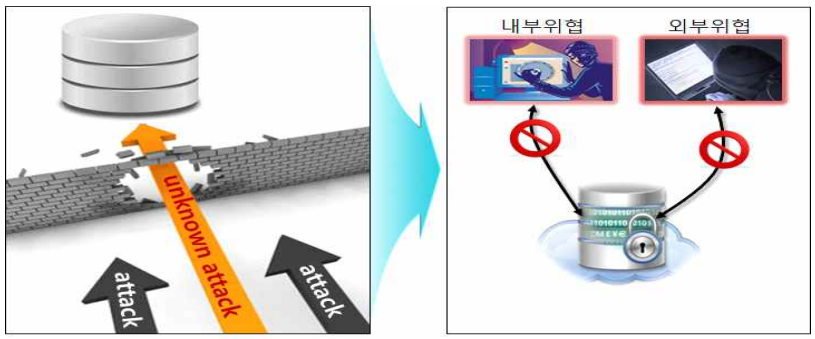
목 차

제 1 장 서 론

제 1 절 연구개발의 중요성

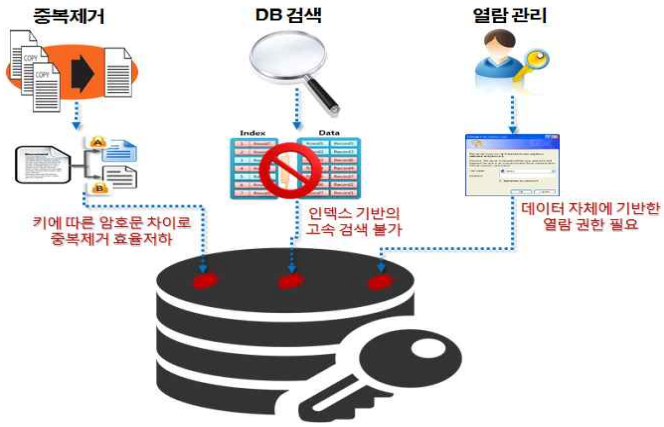
○ 기존의 방어 체계를 우회하는 지능적인 공격의 발달로 공격의 목표가 되는 데이터를 원천적으로 보호하기 위한 암호 기술에 대한 중요성이 증가하고 있음

- 개인정보의 저장·유통이 대량화, 광역화, 네트워크화 되면서 저장·유통되는 개인정보가 더욱 많은 위협에 쉽게 노출되고 있음
- 알려진 공격 유형에 단기적으로 대응하는 현존 방어 체계로는 새로운 공격 기술에 근원적으로 취약함
- 데이터 자체를 근본적으로 보호할 수 있는 데이터 중심적 보안으로 패러다임이 변화하고 있으며, 이를 해결할 수 있는 암호 원천 기술 개발이 중요해지고 있음
- 클라우드 컴퓨팅 보안으로 가장 잘 알려져 있는 조직인 CSA(Cloud Security Alliance)가 암호데이터 검색, 암호화 기반 접근 제어, 암호데이터 연산, 스토리지에 저장된 데이터에 대한 무결성 검증 등 빅데이터 환경에 필요한 암호 기술에 대한 10대 챌린지를 발표하는 등, 암호데이터 활용을 위한 원천 기술이 많은 주목을 받고 있음



○ 하지만, 데이터베이스 암호화에 따른 기능적/성능적 제약이 암호화 사용에 걸림돌로 작용하고 있음

- 현재의 단순 데이터베이스 암호화 기술은 암호데이터 활용을 위해 데이터 전체에 대한 복호화가 요구되어, 성능 저하 및 서버 관리자에 의한 데이터 프라이버시 침해 방지 어려움 등의 문제가 존재
- SW 방식의 DB 암호화는 평균 7배의 성능 저하가 발생하여, 3천여 개에 달하는 우리나라 금융회사 가운데 고객 정보를 암호화해 관리하는 곳은 47개사에 불과할 정도로 암호화 적용이 저조한 상황임(인텔코리아, 2014)



< DB 암호화에 따른 주요 문제점 >

○ 데이터 유출의 원천적인 방지를 위한 데이터베이스 암호화를 통한 기밀성 보장을 기반으로 암호화된 데이터 활용을 위한 암호 원천 기술 개발이 중요함

- 암호데이터 활용을 위한 암호 원천 기술은 데이터 유출 사고 방지를 위한 데이터베이스 보호 기술 뿐 만 아니라, 클라우드/빅데이터 서비스 등과 같은 다양한 신규 서비스에서 사용자 데이터 프라이버시 보호를 위한 핵심 기술로 확대 적용이 가능해 중요성이 더욱 커지고 있음
- 따라서, 데이터 유출의 원천적인 방지를 위해 암호화된 데이터베이스에서 데이터 저장, 열람 및 검색과 같은 평문 데이터베이스의 기능적/성능적 요구 사항을 만족할 수 있는 암호 원천 기술 개발이 중요함

제 2 절 연구내용 및 범위

1. 연구개발의 목표

암호화된 데이터베이스에서의 데이터 저장 및 검색을 위한 암호 원천 기술 개발 사업의 최종 연구목표 및 세부목표는 다음과 같다.

가. 과제의 최종목표

○ 데이터 유출의 원천적인 방지를 위해 데이터베이스가 암호화된 상태로 데이터의 저장, 열람 및 검색이 가능한 암호 원천 기술 개발

- 암호데이터 중복 처리 기술 개발
- 암호데이터 소유권 검증 기술 개발
- 암호데이터 검색 기술 개발

나. 과제의 세부목표

○ 암호데이터 중복 처리 기술 개발

- 메시지 기반 암호화 핵심 프리미티브 설계
- 파일 단위 암호데이터 중복 처리 기술 개발
- 블록 단위/다중 사용자 기반 암호데이터 중복 처리 기술 개발

○ 암호데이터 소유권 검증 기술 개발

- 데이터 기반의 암호데이터 소유권 검증 기술 개발

○ 암호데이터 검색 기술 개발

- 암호데이터 키워드 검색 기술 설계
- 데이터 추가·삭제 기능을 제공하는 동적 암호데이터 검색 기술 설계
- 부가 기능을 제공하는 암호데이터 검색 기술 개발

2. 연구개발의 내용

암호화된 데이터베이스에서의 데이터 저장 및 검색을 위한 암호 원천 기술 개발 사업의 주요 연구내용은 다음과 같다.

○ 암호데이터 중복 처리 기술 개발

- 메시지 기반의 Client-side 암호데이터 중복 처리 기술 개발
  - 메시지 기반 암호화 기술 핵심 설계 논리 개발
    * 안전성과 효율성을 동시에 만족하는 Client-side 중복 처리 논리 설계
    * 네트워크 전송량 및 서버 계산량 감소
    * 데이터 위조 공격 및 데이터 위조 공격 방지 기술 개발
  - 파일/블록 단위 및 다중 사용자 처리 암호데이터 중복 처리 기술 개발

- 암호데이터 중복 처리 SW 개발
  * Server-side 및 Client-side 중복 처리 알고리즘 10종 구현
- 데이터 손실 공격을 방지하는 암호데이터 중복 처리 성능(1MB 평문 파일 대비 암호 파일 업로드 시 부가 시간)
  * 61ms/1MB(데이터가 중복되지 않는 경우)
  * 50ms/1MB(데이터가 중복된 경우)

• 파일 신원 확인 공격에 안전한 암호데이터 중복 처리 기술 개발
- 파일 신원 확인 공격에 안전한 Time-Lock 제어 가능 중복 처리 프로토콜 개발
- 파일 신원 확인 공격에 안전한 암호데이터 중복 처리를 위한 사용자 관리 기술 개발

○ 암호데이터 소유권 검증 기술 개발
• 안전한 암호데이터 중복 처리를 위한 데이터 소유권 상호 확인 기술 개발
• 권한에 기반한 암호데이터 소유권 검증 및 중복 처리 기술 개발
• 소유권 검증 및 공개 감사 동시에 지원하는 데이터 권한 관리 프로토콜 개발

○ 암호데이터 검색 기술 개발
• 암호데이터 검색 기술 설계를 위한 기반 기술 연구
- 기존 공개키/대칭키 암호데이터 검색 핵심 기술 분석 및 프리미티브 알고리즘 설계
- 동적 데이터 관리 환경을 위한 서버 간 암호데이터 비교 기술 개발
• Bloom Filter Tree 기반의 검색 인덱스를 활용한 동적 암호데이터 검색 기술 설계
- Tree 구조의 각 노드를 bloom filter로 구성한 bloom filter tree 검색 인덱스 설계

- Bloom Filter의 특성을 활용한 동적 암호데이터 검색 인덱스 구성 및 bloom filter 연산을 통한 확장 검색 방법 연구

• 다중 링크드 체인 구조 설계를 통한 효율적인 범위 검색 기법 설계
  - 범위 검색의 특성을 고려한 다중 링크드 체인 구조 설계를 통한 O(m)의 범위 검색 성능 제공

• 인덱스 재구성을 통한 동적 암호데이터에 대한 효율적인 검색 인덱스 관리 기법 제시

제 3 절 국내외 기술개발 동향

1. 세계 기술현황

○ 급증하는 데이터의 저장 비용 절감을 위한 데이터 중복 처리 기술에 대한 연구가 진행되어 왔으나, 암호화에 의한 기밀성 유지 상태에서도 저장의 효율성을 제공하는 암호데이터 중복 처리 기술 개발은 최근 본격적으로 시작되고 있음

- 급증하는 데이터에 대응하기 위해 데이터베이스 스토리지 비용 절감이 중요해짐에 따라, 평문 데이터에 대한 중복 처리 기술은 Drop Box, Google Drive, Mozy 등과 같은 클라우드 기반 스토리지 서비스에 적용되고 있음
- 암호데이터 중복 처리 기술은 2002년 최초로 기술 개념이 도입된 이후 단편적인 연구 결과에 머무르다가, 2013년도에 최초로 암호학적 안전성 개념이 제안되고 이를 기반으로 증명 가능한 안전성이 제공되면서 본격적인 연구가 시작되고 있음
- 현재 데이터에 의해 암호화 키가 결정되는 message-locked encryption을 중심으로 연구가 수행 중에 있으나, 기본적인 안전성 요구 사항만 만족하는 기술로 향후 다양한 공격 위협에 대응할 수 있는 원천 기술의 연구가 필요한 상황임
- 낮은 엔트로피를 가지는 데이터의 경우, 현재 알려진 암호데이터 중복 처리 기술로는 안전성을 보장하기 어려워 안전성 강화를 위한 연구가 시도되고 있음
- 향후 암호데이터 중복 처리 기술에 대한 연구는 암호학적 안전성에 개념 고도화, 효율성 개선, 다양한 목적 및 응용 환경 지원이 가능한 기능 부가형 원천 기술 개발 방향으로 진행될 것으로 전망됨

○ 암호데이터 소유권 검증 기술은 데이터 자체를 기반으로 데이터의 소유권 부여, 유지 및 검증을 제공하는 기술로 전 세계적으로 관련 연구가 전무한 상황임

- 기존 사용자 정보 기반 인증을 통한 권한 관리나 접근 제어와 같은 시스템적인 소유권 검증 기술은 권한 이상의 데이터 열람이 가능하여 데이터 프라이버시 침해 우려를 낳고 있음
- 평문데이터에 대한 소유권 검증 기술을 위한 안전성 모델 및 설계 개념을 암호데이터 소유권 검증 기술에 단순 적용하기에는 구조적 차이로 인한 어려움이 존재함
- 현재 암호데이터 중복 처리 과정에서 데이터의 실제 보유 여부 검증을 위한 제한적인 목적으로 일부 이론적인 연구가 수행
- 암호화에 의한 기밀성을 유지하면서 데이터 기반의 소유권 부여, 유지 및 검증이 가능한 새로운 암호데이터 소유권 관리 모델의 정립 및 이를 처리할 수 있는 원천 기술 개발이 필요함

○ 암호데이터 검색 기술은 높은 안전성 제공을 목적으로 한 이론적인 방향으로 연구가 진행되어 왔으나, 데이터베이스 보안에 대한 필요성 증가로 실제 서비스 적용 가능한 암호데이터 검색 기술에 대한 요구가 증가하여 성능 현실화를 위한 실용화 가능한 구조 설계 및 기존 기술의 성능 개선 방향의 연구가 수행되고 있음

- 검색 가능 암호 기술은 암호화된 데이터베이스에서 효율적인 검색을 목적으로 2000년에 처음으로 제안되었으며, 초기에는 높은 안전성을 추구하는 이론적인 연구가 주로 수행됨
- 2006년 최초로 검색 시간이 데이터베이스에 저장되어 있는 전체 데이터 수에 무관한 대칭키 기반 암호데이터 검색 기술인 SSE가 제안된 이후, 관련 기술 개선을 위한 연구가 진행되고 있음
- 대칭키 기반 암호데이터 검색 기술은 빠른 검색 성능을 제공하는 키워드

검색 위주 연구 결과가 제안되어 왔으나, 데이터 추가/삭제와 같은 동적 데이터 처리를 위한 기본 기능 및 범위 검색 등과 같은 부가 기능 제공의 어려움이 존재

- 2004년 데이터 저장 및 검색 주체가 상이한 환경에서도 암호데이터 검색이 가능한 공개키 기반의 기술이 제안된 이후, 다양한 부가 기능 제공 및 안전성 강화를 위한 연구가 진행되었으나 실제 응용 환경에 적용되기에는 제한된 성능을 제공함
- 공개키 기반 암호데이터 검색 기술은 높은 안전성 및 부가 기능 제공이 용이하지만, 데이터 검색 시간이 데이터베이스에 저장된 데이터 총량에 비례하는 비효율성으로 인해 현실적으로 실제 응용 환경에 적용하기에는 성능적 한계를 지님
- 향후 암호데이터 검색 기술은 이론적인 안전성 위주의 연구에서 벗어나 현실 적용 가능한 기술 개발을 위해, 효율성을 강조한 대칭키 기반의 암호데이터 검색 기술을 바탕으로 데이터 추가/삭제와 같은 동적 데이터 처리 및 다양한 부가 기능을 제공하는 방향으로 연구가 진행될 것으로 전망됨

○ 단순한 암호화 기반의 기밀성 제공에서 벗어난 다양한 암호데이터 처리 서비스의 실현을 위해 이론적으로 다루어졌던 암호데이터 중복 처리 기술, 암호데이터 검색 기술들에 대한 구현 연구가 시도되고 있음

- 신생 업체인 Bitcasa사는 클라우드 스토리지에서 초기 수준의 암호데이터 중복 처리 기술인 convergent encryption을 활용해 파일 단위의 암호데이터 중복 처리 기술을 일부 구현하였으나, 안전성 관점의 문제가 존재함
- Elastic Security사는 클라우드 스토리지 서비스에서의 블록 단위 중복 제거 및 데이터의 기밀성을 보장하기 위한 기본 보안 아키텍처를 제안함
- CipherCloud사는 클라우드 환경에서 데이터 암호화 및 인덱싱을 통한 검색 기술을 개발하였으나, 관련 기술의 주요 작업이 게이트웨이에 집중되

어 있어 가용성 및 보안성이 취약함

- 미쓰비시사는 클라우드 환경에서 키워드 검색을 수행하는 검색 가능한 암호화 플랫폼을 개발하였으나, 공개키 방식의 검색 기술 활용으로 인해 데이터 처리 성능에 한계가 존재함
- Fujitsu사는 동형 암호를 이용한 암호데이터 검색 기술을 발표하였으나, 암호화된 상태로 검색을 수행한 후 전체를 복호화하는 구조에서 발생하는 성능 문제로 실제 응용 환경 적용이 어려운 상황임

2. 국내 기술현황

○ 개인정보 유출 사례의 증가로 인해 데이터 프라이버시 보호를 위한 안전한 데이터 관리에 대한 관심 증가

- 개인정보보호법에 따라 주민등록번호, 여권 번호 등과 같은 개인 식별 번호를 포함한 다양한 사용자 주요 정보에 대한 암호화가 의무화되고 있으며, 개인 프라이버시에 대한 인식 제고로 암호화 적용 범위가 점차 늘어날 것으로 예상되고 있음
- 최근 단순한 접근 제어나 암/복호화 처리를 통한 민감 데이터 보호 기술에서 벗어나, 기능 부가형 데이터 서비스를 제공하기 위해 원본 데이터의 길이와 형태를 보존할 수 있는 형태 보존 암호화와 같은 확장형 암호화 기법 개발에 대한 관심이 늘어나고 있음
- 특히, 암호화 대상 데이터의 증가에 따라 암호데이터 저장, 열람 및 검색과 같이 암호데이터를 적극적으로 활용하기 위한 암호 기술에 대한 관심이 급증하고 있음

○ 개인정보보호법 시행됨에 따라, 금융, 의료, 국방, 교육, 공공과 같은 다양한 분야에서 데이터베이스 저장 정보를 기반으로 다양한 데이터 서비스를 제공하면서도 데이터 프라이버시를 보호하기 위한 응용 환경에 적합한 기술에

대한 요구가 급증할 것으로 예상됨

- 암호화된 데이터에 대한 저장, 열람 및 검색과 같은 주요 암호데이터 처리 기술을 지원하는 데이터베이스 보안 시스템 분야가 국내 보안 전문 기업의 주요 사업 영역으로 새롭게 부각될 것으로 전망
- 기밀성을 유지하면서도 고효율의 복합 검색 기능을 제공하는 암호화 기반 데이터베이스 처리 솔루션, 비용 효과적인 데이터베이스 구축 및 운영 기술 개발이 활발히 수행될 것으로 예상

○ 암호데이터 중복 처리 기술은 2013년도에 최초로 암호학적 안전성에 대한 개념이 제안된 이후 세계적으로 본격적인 연구가 시작되고 있는 분야지만, 현재 국내 연구 결과는 미미한 상황임

- 최근 클라우드 서비스의 도입 및 클라우드 스토리지에 저장된 데이터의 급증으로 인해 국내에서도 암호데이터 중복 처리 기술 연구가 관심을 받고 있음
- 국내에서는 평문 및 암호데이터에 대한 중복 처리 기술에 대한 동향 분석 수준의 결과가 일부 발표되고 있음

○ 암호데이터 소유권 검증 기술은 데이터 자체를 기반으로 사용자 소유권을 검증하는 기술로 국내외적으로 연구 결과가 전무한 상황임

- 기존의 사용자 정보 기반의 시스템적인 보안 기술에서 발생하는 권한 이상의 데이터 열람을 방지하기 위해, 데이터 자체를 기반으로 한 새로운 개념의 암호데이터 소유권 모델 정립 및 원천 기술에 대한 수요가 발생할 것으로 예상됨

○ 국내의 암호데이터 검색 기술에 대한 연구는 대학 및 연구소를 중심으로 단편적인 연구가 수행되고 있음

- 고려대, 포항공대 등 대학을 중심으로 암호데이터 검색 방법에 대한 이론

적인 연구가 일부 수행되고 있으나, 단편적인 결과에 머무르고 있음

- 대부분 기존 기법들의 단순 변형 위주의 단편적인 연구에 그쳐 활용 및 확장에 한계가 있으며, 효율성에 대한 근본적인 취약점이 존재함
- ETRI는 순서 보존 암호, 암호데이터 conjunctive 키워드 검색 기술에 대한 연구 수행 및 관련 IPR을 확보하고 있음

3. 국내외 표준화 현황

○ 2000년대 초 선진 각국은 정부 주도하에 암호 알고리즘에 대한 표준화 및 개발을 추진하였음

- 미국은 NIST를 중심으로 차세대 미국 표준 암호 알고리즘인 AES를 선정하여, 2001년 11월 미 연방 표준 FIPS 197로 제정
- 유럽 및 일본은 미국의 AES 프로젝트와 유사한 작업으로 NESSIE 프로젝트와 CRYPTREC 프로젝트를 통해 다양한 암호 알고리즘을 표준으로 제정
- NIST를 중심으로 새로운 해쉬 함수 선정을 위한 SHA-3 프로젝트를 진행하여, 2014년 5월 미 연방 표준 FIPS 202로 제정

○ 암호 알고리즘의 실용성을 고려하여 부가 기능 제공형 암호 기술에 대한 표준화 활동이 진행 중에 있음

- NIST는 운영 모드 표준화를 통해 평문데이터의 형태를 보존하는 포맷 보존 암호화 기술에 대한 표준을 진행 중에 있음
- ISO/IEC SC27 WG2에서는 2014년 홍콩 회의에서 Homomorphic encryption을 18033-6 Encryption algorithms의 Part6 주제로 정하고 표준화를 추진 중

○ 국내의 경우, 암호데이터 활용을 위한 표준화 활동은 미미한 상황임

- TTA PG 501을 통해 암호 알고리즘, 서명, 인증 등 다양한 암호 기술에 대한 표준화를 진행하고 있음

- TTA PG 502에서는 개인정보보호 및 ID 관리 관점에서 데이터베이스 관련 보안 요구 사항에 대한 표준화만 진행된 상황임
- 최근 개인정보보호 강화를 위한 데이터베이스 암호 기술에 대한 범정부적 관심에 따라 향후 많은 활동이 이루어질 것으로 예상됨

○ 빅데이터/클라우드 시장 활성화의 최대 걸림돌이 프라이버시 노출에 대한 위협으로 지적되고 있어, 암호화된 상태의 데이터 활용을 위한 기술 선점을 위해 MS, Google, IBM 등 글로벌 업체들의 적극적인 표준화 추진이 예상됨

○ ISO/IEC, IEEE, ITU 등 국제기구에서도 암호데이터 저장 및 검색 기술에 대한 표준화 추진이 예상되고 있으며, 이에 대응할 수 있는 선제적인 원천 기술의 개발이 요구되고 있음

제 4 절 연구 수행방법 및 보고서 체계

1. 연구 추진체계 및 수행방법

○ 한국전자통신연구원 주도로 암호데이터 저장, 검색을 위한 암호 기술 연구에 대한 방향 설정 및 원천 기술 개발

- 한국전자통신연구원은 암호데이터 중복 처리 및 암호데이터 검색 기술에 대한 원천 기술 설계
- 공동연구기관(공주대학교)은 암호데이터 중복 처리 기반 기술 연구를 수행함으로써, 중복 처리 핵심 알고리즘 개발을 위해 한국전자통신연구원과 협력
- 암호데이터 중복 처리 및 검색에 대한 핵심 원천 기술 개발 및 우수 IPR의 전략적인 확보

○ 응용 환경 분석과 다양한 암호 알고리즘 및 프로토콜 설계 경험을 활용하여 암호데이터 중복 처리 및 암호데이터 검색에 대한 핵심 기술 설계

- 암호데이터 중복 처리 기술은 안전성 개념이 명확히 정립되지 않아 다양한 공격 위협에 노출될 수 있어, 이를 해결하기 위한 안전성 강화된 신규 프리미티브 제공. 또한 기존의 암호데이터 중복 처리 대비 성능이 향상된 암호데이터 중복 처리 기술 제공
- 암호학적 안전성 위주의 이론적인 연구가 주를 이루고 있어 현실 시스템에 적용하기에 무리가 있는 기존의 암호데이터 검색 기술의 한계를 극복하여, 수용 가능 안전성을 기반으로 현실 적용 가능성을 암호데이터 검색 기술을 제공

○ 미국, 유럽, 일본 등 선진국의 연구 프로젝트, 각종 국제 학회 및 저널 논

문, 국내외 특허 등을 면밀히 검토하여 차별화된 연구 개발 수행

○ 위탁연구기관 및 전문가 초청 등을 적극 활용하여 학계의 우수한 기술 확보

○ SCI(E) 저널 및 국제 유수 학회 논문 기고를 통해 연구 결과물의 국제적 검증 수행

2. 보고서 체계

본 보고서는 2015.1.1 ~ 2017.12.31 기간에 수행된 “암호화된 데이터베이스에서의 데이터 저장 및 검색을 위한 암호 원천 기술 개발” 사업의 연구내용 및 결과에 대하여 요약 작성하였다. 1장은 서론으로 연구 개발의 중요성, 연구내용 및 범위, 국내외 기술개발 동향을 소개한다. 2장에서는 과제의 최종 목표와 각 연차별 세부 목표를 제시하고, 이에 대한 평가 지표 및 달성 방법을 기술한다. 3장은 암호데이터 저장 및 검색 기술에 대한 연구 결과를 상세히 기술하고, 4장에서는 연구개발 결과의 활용 계획을 제시한다. 마지막으로 5장에서는 본 보고서의 결론을 맺는다.

제 2 장 기술 개발 내용 및 방법

제 1 절 최종 목표 및 평가 방법

1. 최종 목표

구 분	내 용
최종목표	○ 데이터 유출의 원천적인 방지를 위해 데이터베이스가 암호화된 상태로 데이터의 저장, 열람 및 검색이 가능한 암호 원천 기술 개발 - 암호데이터 중복 처리 기술 개발 - 암호데이터 소유권 검증 기술 개발 - 암호데이터 검색 기술 개발 
세부목표	○ 암호데이터 중복 처리 기술 개발 - 메시지 기반 암호화 핵심 프리미티브 설계 - 파일 단위 암호데이터 중복 처리 기술 개발 - 블록 단위/다중 사용자 기반 암호데이터 중복 처리 기술 개발

	- 데이터 손실 공격 방지를 위한 암호데이터 중복 처리 기술 안전성 모델 정립 및 검증 - 암호데이터 중복 처리 성능 : 460ms 이하/1MB (1MB 파일 단위 암호데이터 중복 처리를 위한 평문데이터 중복 처리 대비 부가 시간) ○ 암호데이터 소유권 검증 기술 개발 - 데이터 기반의 소유권 부여, 유지 및 검증을 위한 관리 모델 설계 - 데이터 기반의 암호데이터 소유권 검증 기술 개발 ○ 암호데이터 검색 기술 개발 - 암호데이터 키워드 검색 기술 설계 - 데이터 추가·삭제 기능을 제공하는 동적 암호데이터 검색 기술 설계 - 부가 기능을 제공하는 암호데이터 검색 기술 개발 - 수용 가능 안전성 모델 정립 및 이를 이용한 암호데이터 검색 기술 최적화 - 검색 성능 : O(m), m : 검색 키워드를 포함하는 데이터 수

2. 연차별 연구개발 목표

구 분	목 표	내 용
1차년도 (2015)	암호데이터 저장 및 검색을 위한 핵심 프리미티브	○ 암호데이터 중복 처리를 위한 핵심 설계 논리 개발 - Message-locked encryption 최신 기술 및 요구사항 분석 - 중복 처리에 따른 데이터 손실 공격 분석 및 암호데이터 중복 처리 기술 안전성 모델 연구

	설계	- 메시지 기반 암호화 핵심 설계 논리 개발 ○ 암호데이터 검색을 위한 핵심 알고리즘 설계 - 암호데이터 검색 최신 기술 및 요구 사항 분석 - 수용 가능 안전성 모델 연구 - 암호데이터 키워드 검색 알고리즘 설계
2차년도 (2016)	암호데이터 저장, 열람 및 검색 기술 설계 및 안전성 검증	○ 암호데이터 중복 처리 기술 설계 - 파일 단위 암호데이터 중복 처리 기술 설계 - 암호데이터 중복 처리 안전성 검증 - 암호데이터 중복 처리 성능 최적화 ○ 암호데이터 소유권 검증 모델 연구 - 적용 환경 분석을 통한 요구 사항 정의 - 데이터 기반의 소유권 부여, 유지 및 검증을 위한 관리 모델 설계 ○ 동적 환경을 위한 암호데이터 검색 기술 개발 - 동적 암호데이터 검색 기술 요구 사항 분석 및 안전성 모델 정립 - 데이터 추가·삭제 기능을 제공하는 암호데이터 검색 기술 개발 및 안전성 검증
3차년도 (2017)	암호데이터 저장, 열람 및 검색 기술 개발	○ 암호데이터 중복 처리 기술 개발 - 블록 단위 암호데이터 중복 처리 기술 개발 - 다중 사용자 기반 암호데이터 중복 처리 기술 개발 - 암호데이터 중복 처리 기술 성능 분석 및 개선 연구 ○ 암호데이터 소유권 검증 기술 개발 - 데이터 기반의 암호데이터 소유권 검증 기술 설계 - 암호데이터 소유권 검증 기술 안전성 검증

		○ 암호데이터 검색 실용화 기술 개발 - 부가 기능 제공 암호데이터 검색 기술 개발 - 실용화를 위한 수용 가능 안전성 적용 방안 연구 - 수용 가능 안전성 기반의 암호데이터 검색 기술 최적화

3. 성과 목표 및 평가

가. 성과목표의 개요

○ 개요

- 데이터 유출의 원천적인 방지를 위해 데이터베이스 암호화를 통한 기밀성을 기반으로 데이터의 저장, 열람 및 검색이 가능한 암호 원천 기술 개발

○ 설정 근거

- 개인 정보 유출의 증가 및 피해 확산으로 사용자 데이터를 원천적으로 보호할 수 있는 암호 기술에 대한 요구가 증가하고 있으나, 데이터베이스 암호화에 따른 기능적/성능적 제약이 암호화 사용에 걸림돌이 되고 있음
- 이러한 상황을 극복하기 위해 암호화된 데이터베이스에서 데이터 저장, 열람 및 검색과 같은 평문 데이터베이스의 필수 요구 사항을 만족할 수 있는 암호 원천 기술 개발 및 핵심 IPR 확보를 위해 성과목표를 설정하였음

나. 성과지표

○ 기술 개발 성과 지표(‘17년도)

성과지표 (주요성능 Spec)	세계최고 수준	기술개발 목표치	목표치 산출근거
암호데이터 중복 처리 기술	데이터 손실 공격 가능 (460ms/ 1MB)[1]	데이터 손실 공격 방지 (데이터가 중복되지 않는 경우 : 460ms 이하/ 1MB, 데이터가 중복된 경우 :100ms 이하/1MB)	암호데이터 중복 처리 과정에서 발생하는 데이터 손실공격을 방지하면서, 기존 기술 대비 성능이 향상된 기술 개발을 위한 목표 설정
암호데이터 검색 기술	검색 성능 O(n)[2]	검색 성능 O(m)[3]	전체 데이터 수가 아닌 검색 키워드를 포함하는 데이터 수에 비례하는 검색 성능 제공을 통한 암호데이터 검색 성능 최적화를 목표로 설정

1) 성능 비교치 : 1MB 파일 단위 암호데이터 중복 처리를 위한 평문데이터 중복 처리 대비 부가 시간

2) n : 전체 데이터 수

3) m : 검색 키워드를 포함하는 데이터 수

○ 연구 산출물 성과 지표

<table>
<tr><td colspan="3">공통지표(필수제시)</td><td colspan="2">공통지표(필수제시)
* 2017년 이전</td></tr>
<tr><td colspan="2">지표명</td><td>총사업연도</td><td>지표명</td><td>총사업연도</td></tr>
<tr><td>과학적 성과</td><td>표준화된 IF 상위 20% SCI 논문(건)</td><td>1건 (제출 이상)</td><td rowspan="3">SCI 논문(건)</td><td rowspan="3">9건 (게재 승인 이상)</td></tr>
<tr><td rowspan="4">기술적 성과</td><td>특허활용률
(기술이전건수/
특허등록보유건수)</td><td>-</td></tr>
<tr><td>국제표준특허(건)</td><td>-</td></tr>
<tr><td>국제표준승인표준기고서 (건)</td><td>-</td><td>국내 특허 출원 (건)</td><td>9건 (출원 및 제출)</td></tr>
<tr><td>3극 특허(건)</td><td>-</td><td>국제 특허 출원(건)</td><td>6건 (출원 및 제출)</td></tr>
<tr><td rowspan="3">경제적 성과</td><td rowspan="3">연구비 대비 기술료 수입(%)</td><td rowspan="3">-</td><td colspan="2">특성 지표(자율제시)</td></tr>
<tr><td>지표명</td><td>총사업연도</td></tr>
<tr><td>논문 피인용도(IF)</td><td>게재(게재 승인) 논문 평균 0.8 이상</td></tr>
</table>

제 2 절 접근방법

1. 핵심요소 및 접근방법

○ 데이터 유출의 원천적인 방지를 위한 암호화된 데이터베이스 환경에서 데이터를 자유롭게 활용하기 위한 핵심 요소인 저장, 열람 및 검색의 3가지 기술 분야로 나누어 접근함

○ 3가지 핵심 분야에 대한 핵심 기술 및 접근 방법은 다음과 같음

핵심 요소		접근 방법
암호데이터 저장/열람 기술	다양한 데이터 활용 환경에 대한 암호데이터 중복 처리 기술 개발	- Message-locked encryption 최신 기술 및 요구 사항 분석을 통한 암호데이터 중복 처리 기술 안전성 모델 연구 - 메시지 기반 암호화 핵심 설계 논리 개발 - 개발된 핵심 설계 논리를 바탕으로 파일 단위 암호데이터 중복 처리 기술 설계 - 다양한 데이터 활용 환경의 목적 달성을 위한 블록 단위/다중 사용자 처리 등의 기술 고도화를 통한 최적화 및 세분화 - 암호데이터 중복 처리 기술 성능 분석 및 개선 방안 연구
	데이터 기반의 소유권 검증 모델 연구 및 기술 개발	- 기존 데이터 소유권 관리 기술 및 응용 환경의 요구사항 분석을 통한 새로운 데이터 기반 소유권 검증 모델 설계 - 이를 바탕으로 데이터 기반의 암호데이터 소유권 검증 기술 설계 및 안전성 증명 - 프라이버시 강화를 위한 데이터 소유권 검증 기술 고도화

암호데이터 검색 기술	동적 환경에서의 암호데이터 검색 기술 개발 및 부가기능 제공	- 기존 암호데이터 검색 기술 및 현실 데이터베이스 구조 분석을 통해 현실 적용 가능한 기술 개발의 토대 마련 - 수용 가능 안전성 모델 연구를 통한 암호데이터 검색 기술에의 적용 방안 모색 - 암호데이터 키워드 검색 알고리즘 설계 및 이를 바탕으로 동적 환경의 요구 사항을 반영한 동적 암호데이터 검색 기술 개발 - 부가 기능 제공을 위한 암호데이터 검색 기술 개발 및 수용 가능 안전성 적용을 통한 기술 최적화

2. 혁신성과 독창성

○ 데이터 위탁 서비스의 활성화와 더불어 데이터 중복 처리 기술의 필요성이 급증하고 있지만, 데이터 암호화에 따른 기술적 어려움으로 인해 암호데이터에 대한 중복 처리 기술은 최근에야 알고리즘 설계 및 안전성 개념이 정립되고 있는 분야로 향후 기술적 발전 방향을 결정지을 수 있는 원천 기술 선점이 중요한 기술 분야임. 또한 현재 데이터 단순 저장 환경에만 적용되고 있는 중복 처리 기술을 클라우드 서비스를 포함한 다양한 응용 환경에서의 데이터 활용 목적에 따라 세분화된 암호데이터 중복 처리 기술로 확대

○ 암호데이터 소유권 검증 기술은 데이터 자체의 열람 권한을 할당하여 현존 데이터 소유권 관리 기술에서 해결하지 못하는 내부자에 의한 데이터 유출 방지 등의 다양한 문제 해결을 위한 새로운 기술 분야의 제시가 목표임. 또한, 현재 관련 연구가 진행되고 있지 않아 원천 IPR 및 기술 선점을 통해 기술 선도 가능

○ 암호데이터 검색 기술은 데이터 활용을 위해 필수적으로 요구되는 기술로 비교적 많은 연구가 이루어졌으나, 이론적인 위협까지 모두 반영한 과도한 암호학적 안전성 위주의 기술 개발이 주를 이루고 있음. 이에 과도한 암호학적 안전성에 대한 재분석과 함께 현실적인 안전성과 효율성 제공을 목표로 하는 수용 가능 안전성 모델을 적용하여 현실 데이터베이스에 적용 가능한 실용적인 검색 가능 암호화 기술 개발 기대

○ 이러한 암호데이터 저장, 열람 및 검색 기술을 통해 데이터의 기밀성 보장을 위한 암호화를 기반으로 암호화된 데이터를 평문데이터처럼 자유롭게 활용할 수 있는 새로운 데이터 보안 패러다임인 CipherData 트렌드를 선도할 수 있는 혁신적인 개발 목표임

제 3 장 연구 수행 내용 및 결과

제 1 절 성과 목표 달성도

성과지표 (주요성능 Spec)	기술개발 목표치	성과	달성도 (%)
암호데이터 중복 처리 기술	데이터 손실 공격 방지 (데이터가 중복되지 않는 경우 : 460ms 이하/1MB, 데이터가 중복된 경우 : 100m 이하/1MB)	○ 메시지 기반 암호화 프리미티브 설계 • 네트워크 전송량 및 서버 계산량 감소 • 데이터 위조 공격 및 데이터 삭제 공격 방지 ○ Client-side 암호 데이터 중복 처리 기술 개발 • 파일/블록 단위 및 다중 사용자 처리 가능 암호 데이터 중복 처리 프로토콜 설계 • 요구되는 보안 수준 선택에 따른 역기능 방지 및 안전성 제공 ○ 암호데이터 중복 처리 SW 개발 • Server-side 및 Client-side 중복 처리 알고리즘 10종 구현 • 클라이언트, 키 서버, 서버, Swift 스토리지 서버로 구성된 중복 처리 테스트 프로그램 개발 ○ 데이터 손실 공격을 방지하는 암호데이터 중복 처리 성능(1MB 평문 파일 대비 암호 파일 업로드 시 부가 시간) • 61ms/1MB (데이터가 중복되지 않은 경우) • 50ms/1MB (데이터가 중복된 경우) ○ 파일 업로드 시간 보다 중복 처리를 위한 부가 시간이 훨씬 더 작으므로, 파일 크기가 클수록 효과적임 • 1MB 대비 33MB 암호파일 업로드 시간 : 30배 (중복처리 안할 때)/3배(중복처리 할 때) 증가	100%
암호데이터 검색 기술	검색 성능 O(m)	○ Bloom Filter Tree 기반 동적 암호데이터 검색 기술 설계 • 가공이 용이한 Bloom Filter를 검색 인덱스로	100%

		활용하여 Bloom Filter Tree 구조의 검색 인덱스 설계 방식 제공 • 전체 데이터 수에 대한 sublinear 검색 성능 O(m log n) 제공 * n : 전체 데이터 수, m : 검색된 데이터 수 • Bloom Filter의 중첩 성질과 Counting Bloom Filter 기술 적용을 통해 효율적인 데이터 추가/삭제 기법 제시 ○ 다중 링크트 체인 구조 기반의 범위 검색 기술 설계 • 링크드 체인 기반 기술 적용을 통한 sublinear 검색 성능 제공 • 범위 검색의 특성을 바탕으로 링크드 체인 사이의 연관성을 정의하고, 이를 활용한 다중 링크드 체인 구조 설계 • 암호데이터에 대한 효율적인 범위 검색 기능 제공 O(m) ○ 동적 암호데이터 범위 검색을 위한 인덱스 재구성 기법 설계 • 임의의 암호데이터 범위 검색 기법에 적용하여 효율적인 동적데이터 추가/삭제가 가능한 인덱스 재구성 기법 설계 • 키워드에 대한 속성 기반 암호 기술 적용을 통해 기반 범위 검색 기술과 동일한 안전성을 보장하는 서버의 인덱스 재구성 과정 설계 • 다중 링크드 체인 기반의 범위 검색 기법과 결합하여 동적 암호데이터에 대한 O(m)의 범위 검색 성능 제공, O(1) 삭제 기능 제공	
SCI(E) 논문(건)	9건 (게재승인 이상)	○ 14건(SCI(E) 논문 10건 게재, SCI(E) 논문 4건 게재 승인)	155%
표준화된 IF 상위 20% SCI 논문(건)	1건(제출 이상)	○ 2건 게재(Information Sciences, IEEE Transactions on Computers)	200%
논문 피인용도(IF)	게재(게재승인) 논문 평균 0.8 이상	○ 1.370 • 14건 전체 IF 합/14 = 19.182/14 = 1.370	100%

국내 특허 출원 (건)	9건 (출원 및 제출)	○ 10건(출원 7건, 제출(출원 중) 3건)	111%
국제 특허 출원 (건)	6건(출원 및 제출)	○ 6건(출원 3건, 제출(출원 중) 3건)	100%

제 2 절 암호데이터 중복 처리 기술 개발

1. 메시지 기반의 Client-side 암호데이터 중복 처리 기술 개발

○ 메시지 기반 암호화 방식을 사용한 새로운 프리미티브 설계

- 안전성과 효율성을 위해 두 가지 타입의 태그 사용
  - 짧은 키로부터 유도된 태그와 긴 메시지로부터 유도된 태그가 모두 일치하는 경우에만 중복 제거 발생

	Client(U)		Server(S)
[Case I]	(1-1) $K=H(M)$, $t=H(K)$ (1-2) Compute $C=E(M,K)$ (1-3) Send t to server	t →	(2-1) Search t (2-2) Confirm that there is no tag that matches t (2-3) Request C
	(3-1) Send C (3-2) Store K, t	C →	(4-1) Compute $T'=H(C)$ (4-2) Store C, t, T'
[Case II]	(1-1) $K=H(M)$, $t=H(K)$ (1-2) Compute $C=E(M,K)$ (1-3) Send t to server	t →	(2-1) Search t (2-2) Confirm that there is a tag that matches t (2-3) Request T
	(3-1) Compute $T=H(C)$ (3-2) Send T to server	T →	(4-1) Search T (4-2) Confirm that there is a tag that matches T (4-3) Update U
	(5-1) Store K, t		
[Case III]	(1-1) Set $K=H(M)$, $t=H(K)$ (1-2) Compute $C=E(M,K)$ (1-3) Send t to server	t →	(2-1) Search t (2-2) Confirm that there is a tag that matches t, (2-3) Request T
	(3-1) Compute $T=H(C)$ (3-2) Send T to server	T →	(4-1) Search T (4-2) Confirm that there is no tag that matches T (4-3) Request C
	(5-1) Send C (5-2) Store K, t	C →	(6-1) Compute $T'=H(C)$ (6-2) Store C, t, T'

<메시지 기반 암호데이터 중복처리 프로토콜>

- Poison 공격에 안전 & 네트워크 자원 절약하는 최초의 프리미티브
  - 기존의 MLE 알고리즘(CE, HCE1, HCE2, RCE)은 트래픽 비효율 또는 Poison 공격에 취약
  - 안전성/효율성/성능 비교

	Items to be compared			CE	HCE1	HCE2	RCE	Our primitive
Security	Against duplicate-faking attack (TC secure)			Yes	No	Yes	Yes	Yes
	Against erasure attack (STC secure)			Yes	No	No	No	Yes
Network	Comm. traffic required to upload	server-side		$N+n\cdot\|t\|$	$N+n\cdot\|t\|$	$N+n\cdot\|t\|$	$N+n\cdot(\|t\|+\|K\|)$	$N-M+$ $n\cdot\|t\|+m\cdot\|T\|$
		client-side			$N-M$ $+n\cdot\|t\|$	$N-M$ $+n\cdot\|t\|$	$N-M+n\cdot\|t\|$ $+(n-m)\cdot\|K\|$	
Perf.	Cryptographic operations	upload	at server	$n\times H$	-	-	-	$(n-m)\times H$
			at client	$n\times E$	$n\times G$, $(n-m)\times E$	$n\times G$, $(n-m)\times E$	$n\times G$, $(n-m)\times E$	$n\times G$, $m\times H$, $n\times E$
		download	at client	$n\times D$	$n\times D$	$n\times D$, $n\times H$	$n\times D$, $n\times H$	$n\times D$

n : 저장할 파일 개수, m : 중복된 파일 개수, N : n개의 파일 전체 크기

M : m개의 중복된 파일 전체 크기, t : 첫 번째 태그, T : 두 번째 태그

$|t|$: 첫 번째 태그의 크기, $|T|$: 두 번째 태그의 크기, $|K|$: 메시지 기반 암호키의 크기

H, G : 해쉬 함수, E, D : 암복호용 블록 암호

<안전성, 네트워크 트래픽, 암호 연산 횟수 비교>

○ 다중 사용자를 위한 파일/블록 단위 Client-side 암호데이터 중복 처리 프로토콜 개발

- Client-side 중복 처리 기능 제공
  - 클라이언트는 중복된 파일을 서버로 전송하지 않음
  - 서버는 중복되지 않는 파일에 대해서 태그 생성 및 검증을 통한 안전성 강화
- 파일/블록 단위의 암호데이터 중복 처리 지원
  - 데이터의 형태는 파일이거나 블록인 경우에도 중복 처리가 가능한 암호학적 프리미티브를 이용하여 프로토콜 설계
- 다중 사용자에 의한 데이터 중복 처리 가능
  - 단일 사용자 뿐 아니라 여러 사용자가 동일한 암호 파일을 공유하는 경우도 데이터 중복 제거 기능 포함
- 데이터 손실 공격(Poison attack)에 대한 안전성 제공

- 데이터 위조 공격에 대한 방지 기능 제공

  * 클라이언트는 공격자에 의해 다른 파일로 대체된 위조된 파일을 다운로드 하지 않음

- 데이터 삭제 공격에 대한 방지 기능 제공

  * 공격자에 의해 사용자의 원본 파일이 서버에서 삭제되지 않음

- 기존의 Client-side MLE 알고리즘은 데이터 위조 공격과 데이터 삭제 공격에 취약하지만, 본 연구 결과는 두 가지 공격 모두에 안전

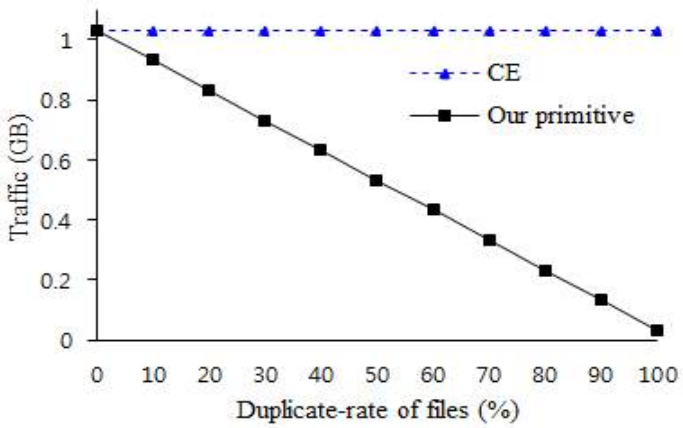
- 해쉬 함수의 Collision-Resistance 성질에 의해 안전성 보장

• 통신 비용(전송되는 트래픽) 감소

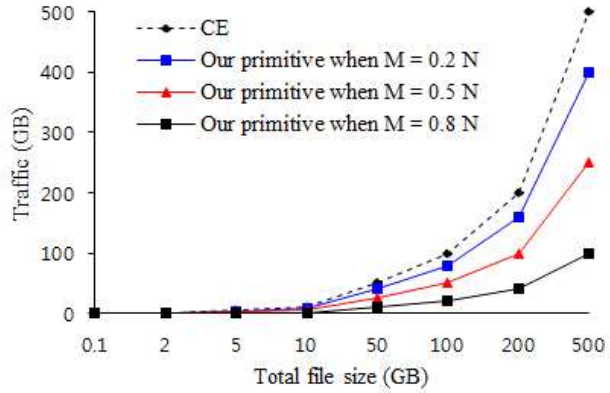
- 기존의 가장 대표적인 CE 기법과 비교

- 중복되는 전체 데이터(N) 중에서 중복되는 데이터(M) 만큼 비례하여 전송되는 트래픽 감소

- 업로드 단계에서 서버는 중복되지 않는 데이터만 태그 계산



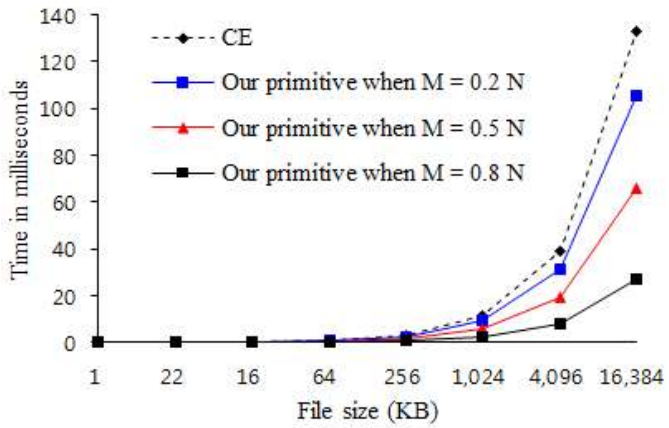
<업로드 시 데이터가 1GB 일 때, 전송되는 트래픽 양>



<저장할 데이터가 증가할 때, 중복 제거율에 따라 전송되는 트래픽 양>

- 암호 연산(해쉬 및 블록 암호) 성능 분석

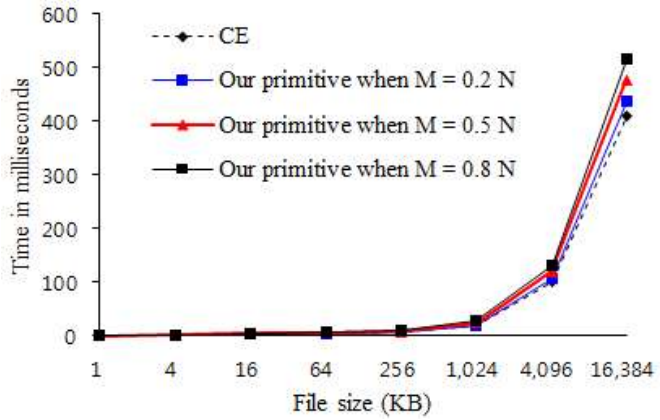
  - 업로드 단계에서 서버의 암호 연산(해쉬(H)) 감소



<업로드할 때 서버에서의 해쉬 연산 시간>

  - 다운로드 단계에서 클라이언트는 해쉬(H) 연산 불필요

  - 업로드 단계에서 클라이언트의 암호 연산 (해쉬(H), 블록암호(E)) 비교



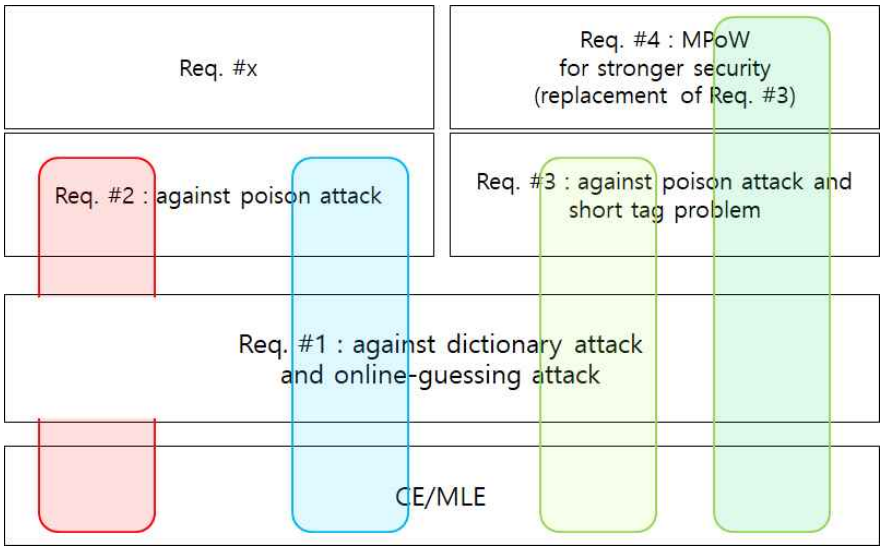
<업로드할 때 클라이언트에서의 해쉬 + 블록암호 연산 시간>

○ Lego 타입의 암호데이터 중복 처리 기술 개발 (요구 사항에 따라 중복 처리 알고리즘 및 컴포넌트 선택)

- 키 서버 사용, PoW(Proofs of Ownership) 등의 중복 처리 기법과의 자유로운 결합을 통해 효율성 저하 없이 다양한 중복 처리 기법을 수용하는 프리미티브 설계 및 구현
  - 클라이언트와 키 서버가 협력하여 사전 공격(Dictionary attack)에 안전한 키 생성 기능 구현
  - Challenge-Response 형식의 클라이언트 파일 소유권 검증으로 불법 사용자의 파일 권한 획득 금지
  - 상호 소유권 검증을 통한 안전한 중복 처리 프로토콜 설계 및 구현
- 사용자의 보안 수준 선택에 따라 Contents guessing attack, Poison attack 등에 대한 역기능을 방지하는 안전성 제공
  - 다양한 안전성과 효율성을 제공하고 사용자는 응용 환경에 따라 중복 처리 모델 선택 가능
  - 보안 요구 사항에 따라 레고 블록을 조립하는 구조
    * Req. #1) Dictionary attack 방지
    * Req. #2) Poison attack 방지

* Req. #3) Poison attack과 짧은 태그 사용 문제 방지

* Req. #4) Req. #3 이상의 강한 안전성 요구



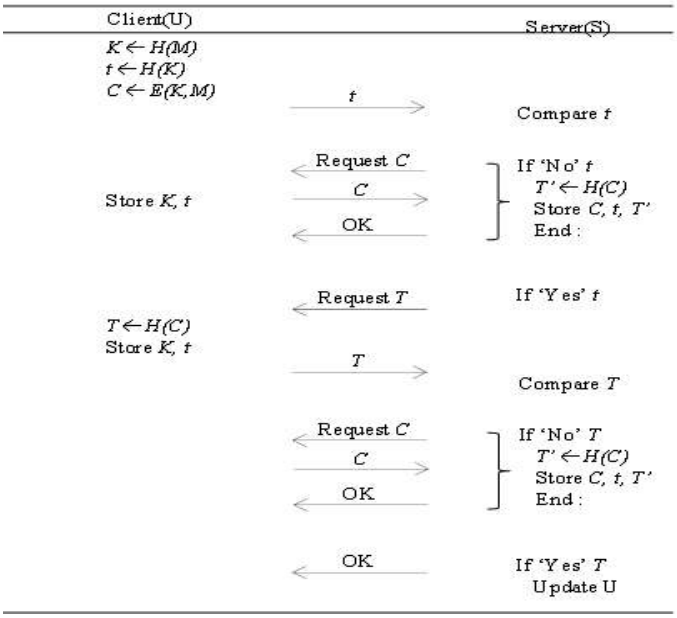
<요구 보안 레벨에 따른 안전성 모델 개념도>

- 보안 요구 사항에 대응하는 구체적인 해결 방법

  * Sol. #1) 동일한 데이터는 동일한 키를 사용하는 결정적 암호 문제를 해결하기 위해 RSA-OPRF 프로토콜 적용한 Dupless의 키 서버 사용

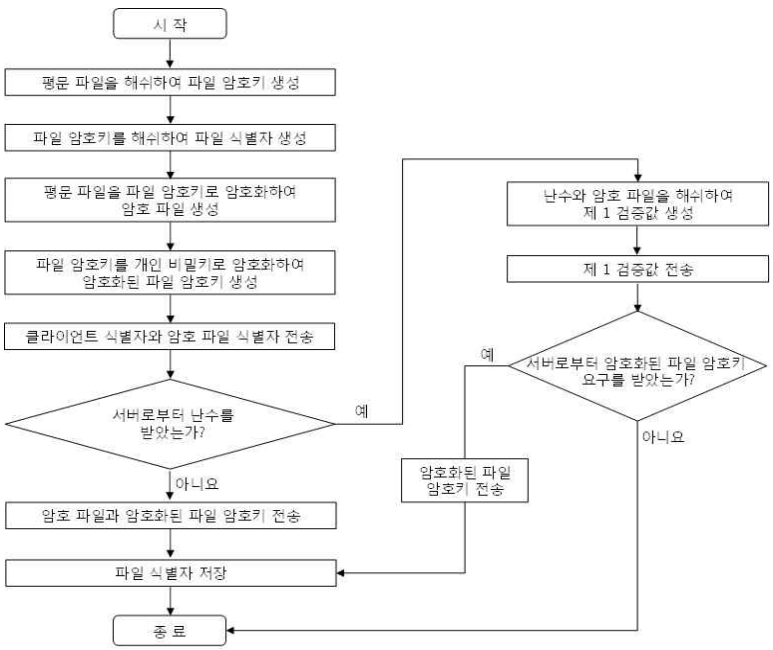
    Cloudedup 방식의 부가적인 암호화, IBM의 인덱스 서버, Privilege key 및 속성 기반 암호 적용 가능성 연구

  * Sol. #2) 두 가지 타입의 태그(짧은 키로부터 유도된 태그, 긴 메시지로부터 유도된 태그)를 사용하고, 두 가지 모두 일치하는 경우에만 중복 제거 실행

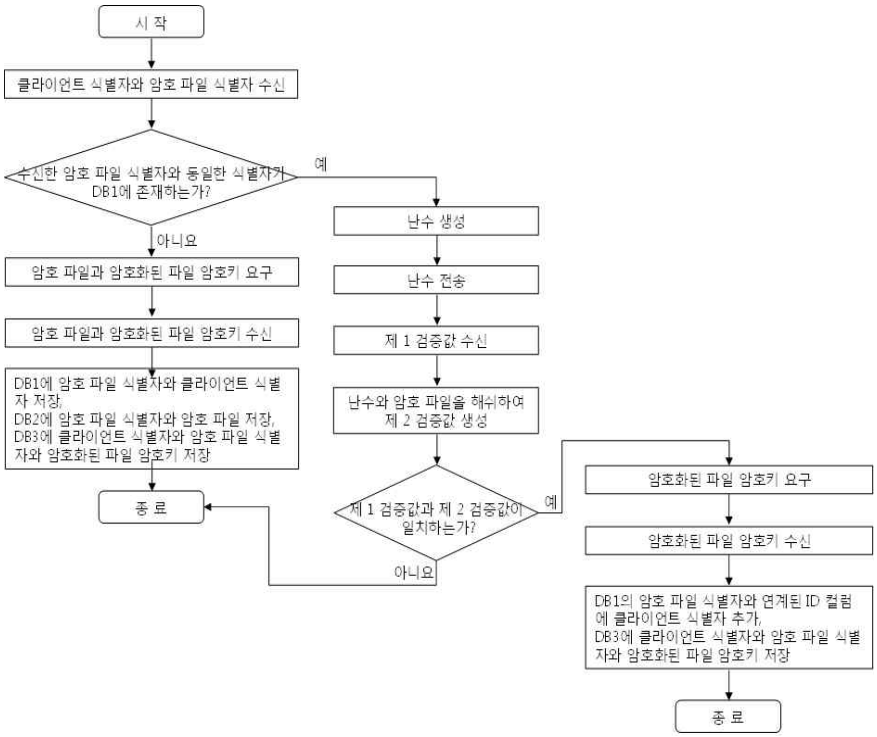


<두 가지 태그 검증을 이용한 암호 파일 중복 처리 알고리즘>

* Sol. #3) Challenge-Response 형태의 클라이언트 파일 소유 유무 검증

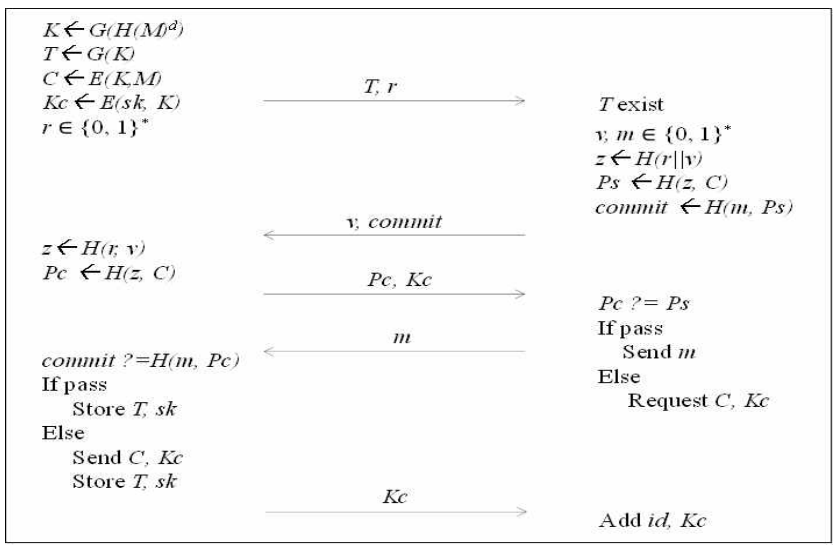


<클라이언트에서의 파일 소유 검증 비교에 의한 암호 파일 중복 처리 알고리즘>



<서버에서의 파일 소유 검증 비교에 의한 암호 파일 중복 처리 알고리즘>

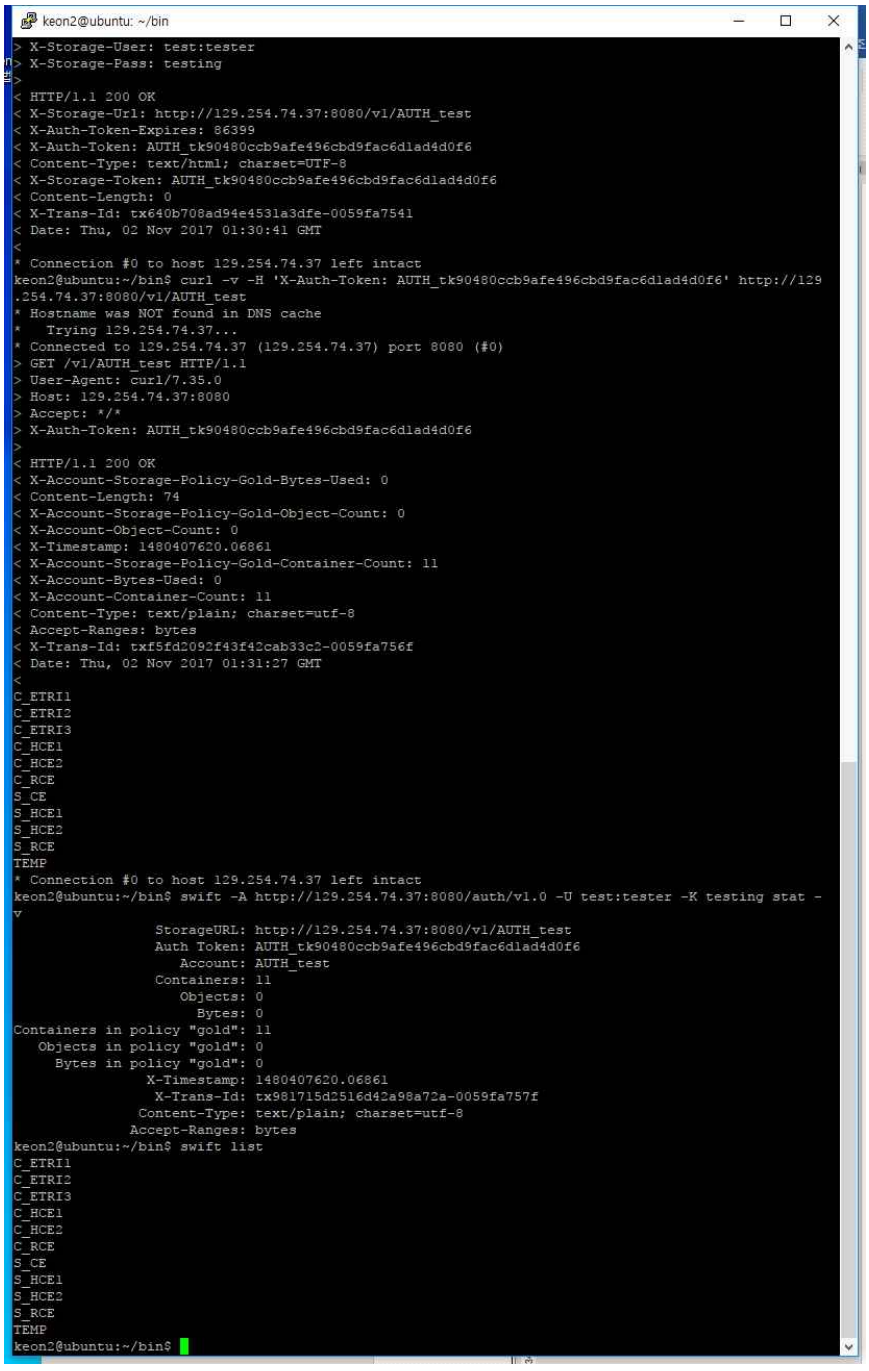
* Sol. #4) Commitment 형태의 양방향 상호 인증 방식



<상호 인증을 이용한 암호 파일 중복 처리 알고리즘>

○ 암호데이터 중복처리 SW 개발

- 암호데이터 중복처리 알고리즘 10종 구현
  - Server-side CE, HCE1, HCE2, RCE를 포함한 4종
  - Client-side HCE1, HCE2, RCE를 포함한 3종
  - 신규 논리를 포함한 중복 처리 알고리즘 3종
- 암호 데이터 중복 처리 알고리즘 성능 평가
  - 평문 파일과 대비하여 암호 파일 중복처리에 필요한 부가 시간 측정
    . 평문 파일 업로드/다운로드 시간 측정
    . 암호 파일 업로드/다운로드/중복 처리 시간 측정
    . 암호 파일 중복 처리에 필요한 암호학적 연산 시간 측정
  - 타입별/알고리즘별 업로드/중복처리時 데이터 전송량 및 스토리지 사용량 시각화 기능 제공
- 프로그램 시험 환경
  - Client : 암호데이터 중복 처리 기능을 테스트하는 클라이언트 SW
  - Key Server : Client와 함께 파일의 키 생성에 참여하는 키 서버 SW
  - Server : Client 접속을 제어하고 중복 여부 판단 등의 업로드된 파일 관리하는 서버 SW
  - Swift Storage Server : Openstack Swift Swfit 서비스를 사용하여 업로드된 파일을 저장하는 오브젝트 스토리지 서비스



<Swift 스토리지 서비스 화면>

- 데이터 손실 공격을 방지하는 중복 처리 성능(신규 개발 프리미티브 적용)

  - 1MB 평문 파일 대비 암호 파일 업로드 부가시간(중복이 없을 때)

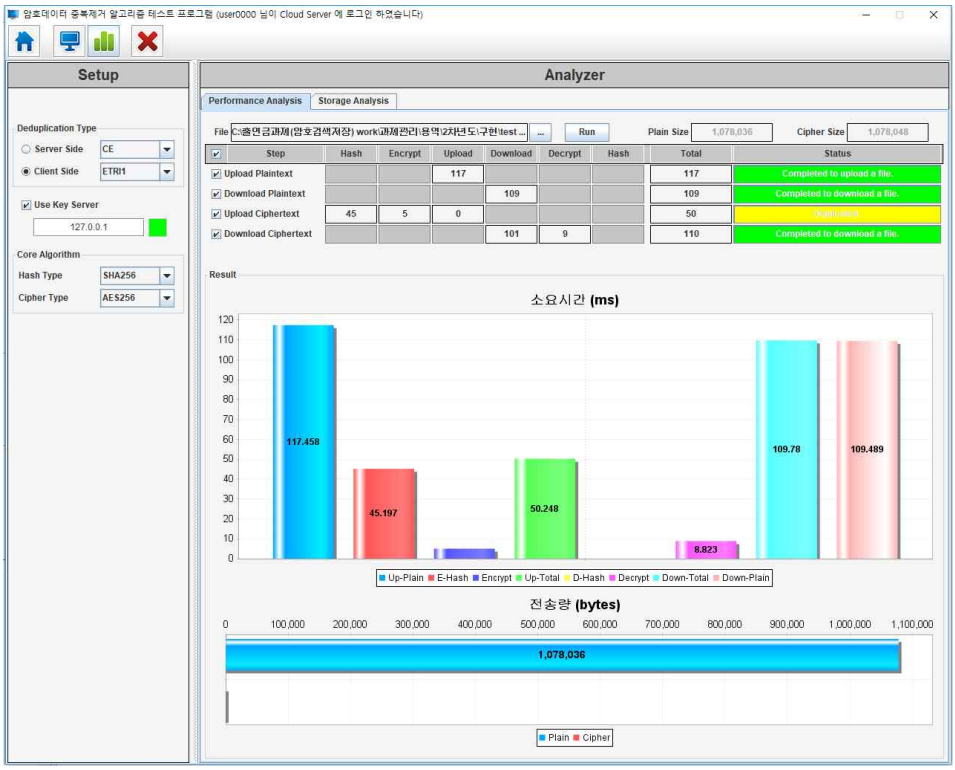
<중복처리 성능 측정을 위한 클라이언트 화면>

. 평문 파일 업로드 119ms

. 암호 파일 업로드 177ms(해쉬 55ms, 암호 6ms, 순수 업로드 117ms)

-> 평문 파일 업로드 대비 암호 파일 업로드를 위한 부가 시간 : 61ms

- 1MB 평문 파일 대비 암호 파일 업로드 부가 시간(중복이 있을 때)



<중복처리 성능 측정을 위한 클라이언트 화면>

. 평문 파일 업로드 117ms

. 암호 파일 중복처리 부가 시간 50ms(해쉬 45ms, 암호 5ms, 업로드 0)

-> 1MB 암호 파일을 중복 처리 안할 때와 할 때의 업로드 시간 : 177ms(중복 처리 안하는 경우), 50ms(중복 처리 하는 경우)

- 크기가 큰 파일의 업로드/중복 처리 시간

. 33MB 암호 파일을 중복 처리 안할 때 업로드 시간 : 3,195ms

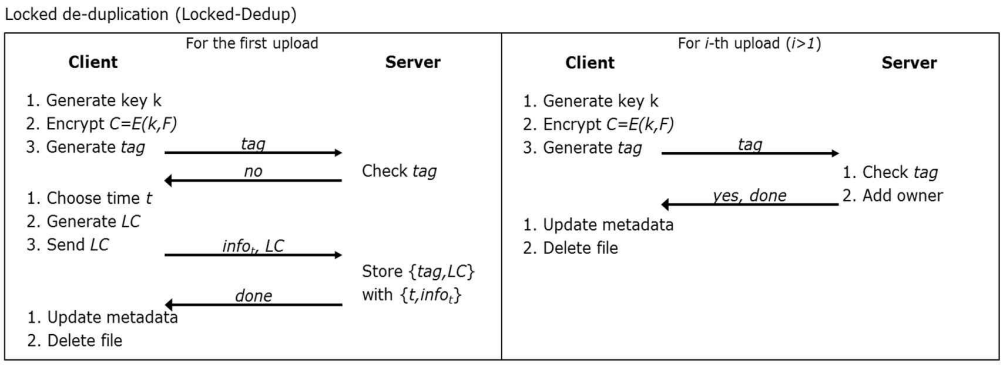
. 33MB 암호 파일을 중복 처리 할 때의 업로드 시간 : 155ms

-> 파일 크기가 클수록 중복 처리를 위한 부가 시간 증가보다는 업로드 시간이 매우 증가하므로, 데이터 중복 처리의 필요성이 부감됨

2. 파일 신원 공격에 안전한 암호데이터 중복 처리 기술 개발

가. 파일 신원 공격에 안전한 Time-Locked 제어 가능 중복 처리 프로토콜 설계

○ 본 연구 결과로 개발된 신원 확인 공격 대응 기술은 사용자가 본인의 프라이버시 강도를 조절할 수 있는 구조로 되어 있어 사용자가 자신의 프라이버시 성향에 따라 안전성 강도를 조정할 수 있음



<제어가능 중복 처리 프로토콜>

- 사용자가 본인이 저장하는 데이터에 대해 중복 처리 제한 시간을 선택할 수 있도록 구성된 기술로써 일종의 시간봉인(time-lock) 기능을 활용하여 설계됨
- 카운터 기반의 대응 기술에서는 서버가 관리하는 카운터를 통해 특정 파일이 중복 처리에 활용되는 시점을 확인하는 형태이기 때문에 중복 처리 제한에 대한 강제성이 없었으나, 본 연구 개발 기술의 경우 정해진 시간 간격으로 특정 정보를 생성하는 시간 서버(time server)에서 매 시간 공

개하는 정보를 기반으로 중복 처리 활용이 가능하도록 기술적으로 강제하는 기술임

- 시간 서버가 공개하는 정보는 각 시간에 대응되는 일종의 타원곡선 기반의 서명 값에 해당하는 정보로 시간 서버의 비밀키를 알지 못하면 생성할 수 없는 값임. 안전성 관점으로, 시간에 대응되는 비밀 정보를 시간 서버의 도움 없이 생성하는 것은 서명 위조와 동일한 강도의 안전성 제공함

○ 임의의 Client-Side 중복 처리 기법에 쉽게 적용할 수 있는 기술로, 파일 신원 확인 공격에 취약한 각 파일별 초기 업로드 시점을 제외하고 기반이 되는 중복 처리 기법의 성능을 유지하는 구조로 설계되어 성능 면에서는 일반적인 중복 처리 기법과 거의 유사한 성능 제공

- 파일별로 초기 업로드하는 사용자의 경우, 시간을 기준으로 중복 처리 제한을 설정하기 위한 부가 비용이 발생함
- 중복 처리 제한이 설정된 파일의 경우 중복 처리 가능한 데이터에 추후 설정 변경을 지원하기 위한 데이터가 추가되나, 이는 고정된 짧은 상수 길이의 데이터로 전체 저장량에 큰 영향을 미치지 않음

	Cost for the first upload				
	Computational Cost		Size of Message		# of Round
	Server	Client	Server	Client	
No-Dedup	-	C_E+C_{Tg}	-	$\ell_T+\ell_F$	2
SS-Dedup	C_{Tv}	$C_K+C_E+C_{Tg}$	-	$\ell_T+\ell_F$	2
CS-Dedup	C_S+C_{Tv}	$C_K+C_E+C_{Tg}$	-	$\ell_T+\ell_F$	4
CS-Dedup$^+$	C_S+C_{Tv}	$C_{KS}+C_E+C_{Tg}$	-	$\ell_k+\ell_T+\ell_F$	6
Locked-Dedup	C_S	$C_K+C_E+C_{Tg}+C_L$	-	$\ell_T+\ell_F+\ell_t$	4
Locked0Dedup$^+$	C_S	$C_{KS}+C_E+C_{Tg}+C_L$	-	$\ell_k+\ell_T+\ell_F+\ell_t$	6

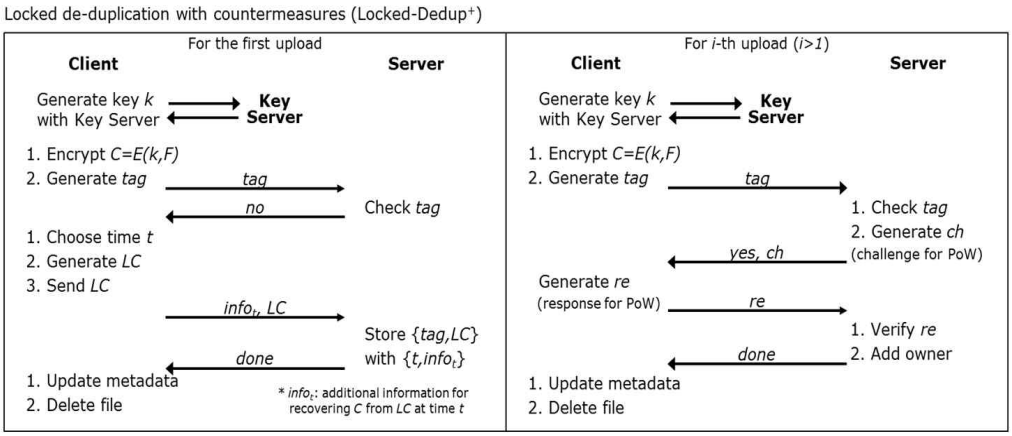
<Time-Locked 중복 처리에서의 최초 사용자 처리 비용>

• 중복 처리 가능한 형태로 저장된 데이터에 대한 업로드 요청 시, 일반적인 데이터 업로드 과정과 동일하게 수행되어 연산량 및 전송 데이터의 길이 등의 비용이 일반적인 중복 처리 과정과 동일하게 소요됨

	Cost for the i-th upload ($i > 1$)				
	Computational Cost		Size of Message		# of Round
	Server	Client	Server	Client	
No-Dedup	-	C_E+C_{Tg}	-	ℓ_T+ℓ_F	2
SS-Dedup	C_{Tv}	C_K+C_E+C_{Tg}	-	ℓ_T+ℓ_F	2
CS-Dedup	C_S	C_K+C_E+C_{Tg}	-	ℓ_T	2
CS-Dedup$^+$	C_S+$C_{PoW\text{-}S}$	C_{KS}+C_E+C_{Tg}+$C_{PoW\text{-}C}$	ℓ_c	ℓ_k+ℓ_T+ℓ_r	6
Locked-Dedup	C_S	C_K+C_E+C_{Tg}	-	ℓ_T	2
Locked0Dedup$^+$	C_S+$C_{PoW\text{-}S}$	C_{KS}+C_E+C_{Tg}+$C_{PoW\text{-}C}$	ℓ_c	ℓ_k+ℓ_T+ℓ_r	6

<Time-Locked 중복 처리에서의 사용자 처리비용>

○ 파일 신원 확인 공격 외의 위협에 대응하기 위해 기존에 알려진 대응 기술들을 동시에 적용할 수 있어 기존 보안성 강화 기술들과 함께 사용하여 다양한 공격에 대응하는 안전한 중복 처리 서비스 제공 가능



<알려진 공격에 안전한 제어 가능 중복 처리 프로토콜>

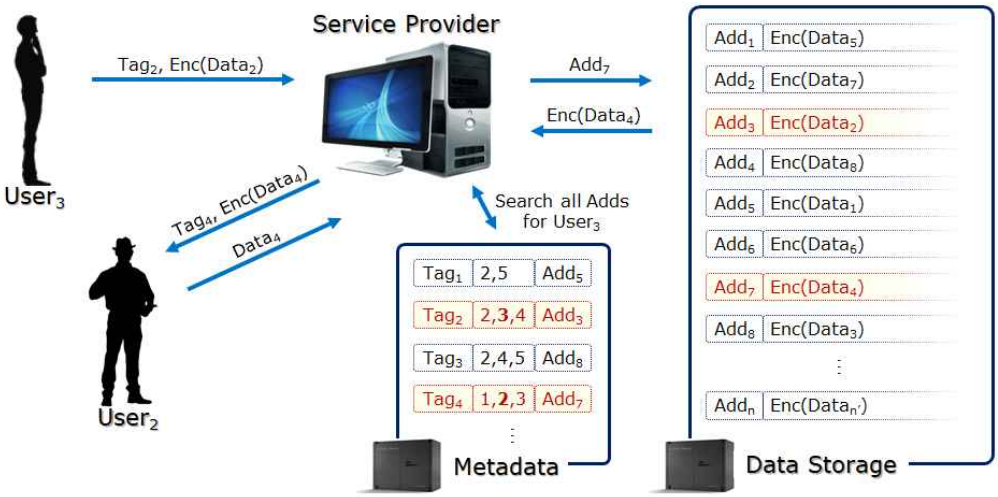
- 제안 기술은 기존에 동일 파일이 저장되어 있지 않은 경우에 한하여 동작하는 것이므로 중복 처리 기능에 영향을 미치지 않고, 동일 파일이 저장되어 있는 경우에는 일반적인 중복 처리 기법이 동일하게 동작하므로 기존의 중복 처리 기술의 안전성 향상 기술을 적용함에 있어 제약이 따르지 않음
- 전수 조사 형태의 공격에 대응하기 위한 기술로, 파일에 대응되는 키 생성을 지원하는 보조 서버를 도입하여 데이터에 대한 키를 무작위로 추정하여 시도할 수 있는 다양한 형태의 공격을 약화시킬 수 있음
- 실제 데이터 소유하지 않은 사용자가 특정 데이터에 대한 소유권 획득하는 문제를 대응하기 위해 소유권 증명 기술을 도입하여 적법한 사용자에게만 소유권 부여하는 기능 제공 가능

나. 파일 신원 확인 공격에 안전한 암호데이터 중복 처리를 위한 사용자 계정 관리 기술 설계

○ 스토리지 서비스에 대한 파일 신원 공격은, 특정 파일과 해당 파일 소유자에 대한 관계 정보 확인이 가능한 경우 위협적인 공격이 됨

- 공격자는 특정 파일이 스토리지 서버에 존재하는지 주기적으로 확인함으로써 공격 대상 사용자가 해당 파일을 소유하고 있는지 확인함으로써 파일 신원 공격을 수행할 수 있으나, 공격 실행 난이도가 높음
- 다양한 공격 방법으로 스토리지 서버에 저장된 파일-사용자 관계 정보가 공격에 활용되는 경우 상기와 같은 공격자의 노력 없이도 프라이버시 정보가 크게 훼손됨
- 실제로 서비스 제공자의 관리 소홀 등으로 인한 사용자 관련 정보 유출이 많이 발생하고 있어 기술적인 대응책이 요구됨

- 특히, 실제 정보 소유자가 공격자로 활동하는 경우 피해자의 보유 데이터 리스트 및 실제 저장 데이터를 모두 확인 가능한 프라이버시 침해 공격을 수행할 수 있음
  - 아래 공격 시나리오에서, $User_2$는 스토리지 서버와 함께 $User_3$이 보유한 데이터를 확인하는 공격을 수행하고 있음



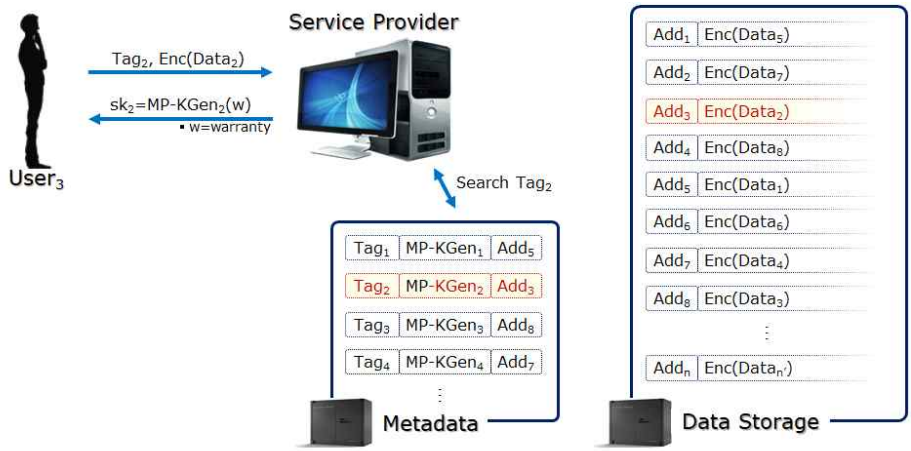
<사용자-데이터 연관 정보를 이용한 스토리지 서비스 공격 시나리오>

○ 본 기술은 그룹 서명 기법을 기반으로 실제 사용자 관련 정보 노출은 최소화하면서 스토리지 서비스 사용자를 관리할 수 있는 방법을 제시함

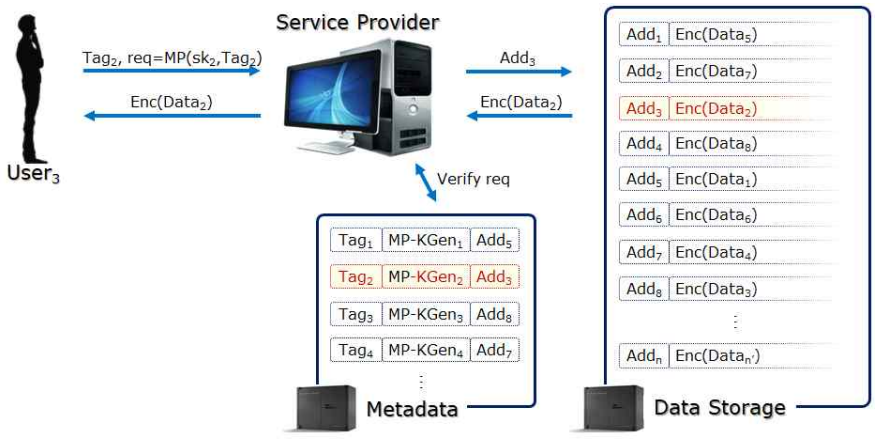
- 기본적인 사용자 관리 기술인 리스트 기반의 사용자-데이터 정보 관리 방법에서는 특정 사용자가 접근 가능한 데이터 정보 및 동일 데이터 보유한 다른 사용자 관련 정보들이 쉽게 노출되어 상기 도시된 공격 시나리오 시행이 용이함
- 이와 같은 문제점을 해결하기 위해 그룹 서명 기법을 기반으로 해결 방법을 제시하는데, 기본적으로 서명자를 특정 짓지 않으면서도 적법한 서명

생성이 가능한 그룹의 구성원임을 확인할 수 있는 그룹 서명 기법의 기능을 기본으로 스토리지 서비스에 적합한 데이터 관리 방법을 제공함

- 스토리지 서비스에서 특정 파일에 대한 소유 권한이 있는 사용자에게 해당 파일에 할당된 그룹 서명 파라미터에 대한 서명 생성 권한을 제공함으로써 해당 파일에 대한 권한 검증 제공



<제안하는 사용자-데이터 정보 관리 기술 개념도, 데이터 업로드 단계>



<제안하는 사용자-데이터 정보 관리 기술 개념도, 데이터 다운로드 단계>

○ 데이터 업로드와 다운로드 과정에서 사용자 연관 정보는 일체 저장 또는 사용되지 않고, 대상 파일에 대응되는 그룹 서명을 통해 사용자 권한 관리를 제공함

- 최초 업로드 되는 파일의 경우, 해당 파일에 대응되는 그룹 서명 파라미터를 생성하고 서명키를 생성하여 파일 업로드 수행한 사용자에게 제공
- 이미 존재하는 파일에 대한 업로드를 수행하는 경우 실제 소유여부 검증 후에 저장되어 있는 그룹 서명 파라미터로 서명키를 생성해서 추후 소유권 확인을 위해 사용하도록 제공함
- 파일 다운로드 수행을 위해 기존에 제공받은 그룹 서명키로 인증 정보를 생성하여 저장된 파일에 대한 소유권을 증명함

제 3 절 암호데이터 소유권 검증 기술 개발

1. 암호데이터 소유권 검증 기술 분석

○ 기존의 암호데이터 중복 처리 서비스 기술에서는 동일 파일을 가지고 있는 사용자의 경우 동일한 키를 생성할 수 있는 특성을 기반으로 설계됨

- 동일한 데이터도 다른 키를 사용하면 상이한 암호문으로 생성되기 때문에 이론적으로 중복되는 데이터가 발생할 수 없기 때문에 데이터에서 키가 생성되는 방식으로 설계되어 있음

○ 중복 처리 서비스 제공을 위해 기본적으로 요구되는 특성이 해당 서비스의 취약점의 근본적인 원인이 됨

- 동일한 파일에 동일한 키가 할당되는 특성으로 인해, 낮은 엔트로피를 가지는 파일의 경우 파일을 추측하고 대응되는 키를 생성할 수 있음
- 무분별한 데이터 추측을 통한 키 생성으로 인한 공격에 대응하기 위한 기술로 키 생성 서버를 사용하기도 함
- 키 서버는 파일에 대응되는 키를 생성하는 방식을 함수 형태로 공개되어 누구나 계산 가능한 형태로 제공하지 않고, 키 서버에 의해서만 계산 가능한 값을 생성

○ 데이터에 동일한 키가 대응되는 구조는 특정 스토리지 서버가 다수의 사용 그룹에 스토리지 서비스를 제공하는 서비스 그룹 간 보안 특성 차이에 의해 의도치 않은 취약점이 발생할 수 있음

- 상기 기술된 취약점이 동일 서비스 그룹 내의 문제에서 타 서비스 그룹을 포함한 대상에서의 문제로 위협이 파급됨
- 서비스 그룹별로 다른 수준의 보안 서비스를 제공할 수 있어서, 동일 스

토리지를 사용하는 서비스 그룹 중에서 낮은 수준의 보안을 제공하는 서비스 그룹에서 제공하는 수준으로 해당 스토리지 사용하는 전체 서비스 그룹의 보안 강도가 낮아질 수 있음

- 능동적인 공격자의 경우, 악의적으로 공격 대상 서비스 제공 서버와 동일한 스토리지 서버에 접근해 데이터 중복 처리 기능에서 발생하는 특성 기반의 공격을 시도할 수 있음
- 기대하는 수준의 보안 수준과 실제 제공될 수 있는 보안 수준이 다른 것은 스토리지 서비스의 안전성 측면에서 매우 큰 문제점이 될 수 있음

2. 안전한 암호데이터 중복 처리를 위한 데이터 소유권 상호 확인 기술 설계

○ 안전한 client-side 암호데이터 중복 처리를 위한 원천 기술로써 동시에 서버와 클라이언트의 데이터 소유권 증명하는 최초의 기법 설계

- 서버는 클라이언트에게 특정 데이터를 실제로 전송받지 않고 소유권을 부여하기 위해 클라이언트가 실제로 보유하고 있는지 검증해야함
- 클라이언트는 실제로 데이터를 서버에게 전송하지 않고 본인의 파일을 삭제하므로, 데이터 삭제 전에 서버에 저장된 데이터가 업로드 하려는 데이터와 동일한지 검증해야함

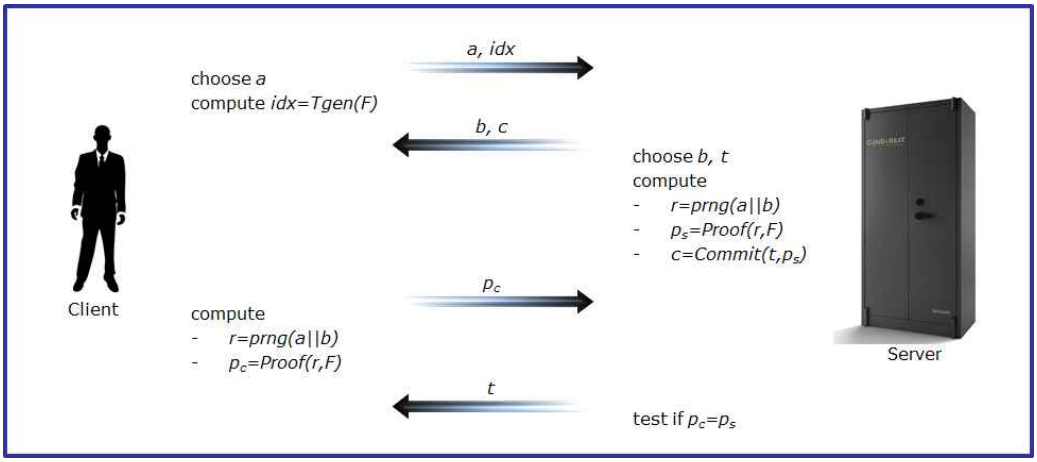
○ 기존의 client-side 암호데이터 중복 처리를 위한 소유권 증명 기술보다 확장된 소유권 증명 기능을 제공하여 데이터 서비스의 신뢰성 향상시킴

- 기존에 설계된 안전한 client-side 암호데이터 중복 처리를 위한 소유권 증명 기술의 경우 클라이언트의 소유 여부를 증명하는 것으로 한정됨
- 데이터 업로드 이후 서버의 데이터 보유 여부 확인 할 수 있는 기술은 존

재했으나 가장 중요한 업로드 시점에 서버와 클라이언트 두 주체의 데이터 소유 여부를 동시에 증명할 수 있는 기술은 본 개발이 최초임

○ 암호데이터 중복 처리를 위한 데이터 소유권 상호 검증 기술의 일반적인 설계 방법 제시

- 알려져 있는 태그 생성 함수, 의사 난수 생성 함수, (단방향) 소유권 증명 기법, 위임 기법을 사용하여 소유권 상호 검증이 가능한 지식 증명 기술의 일반적인 설계 방법을 제시함



<데이터 소유권 상호 검증 기술 일반적인 설계 기법>

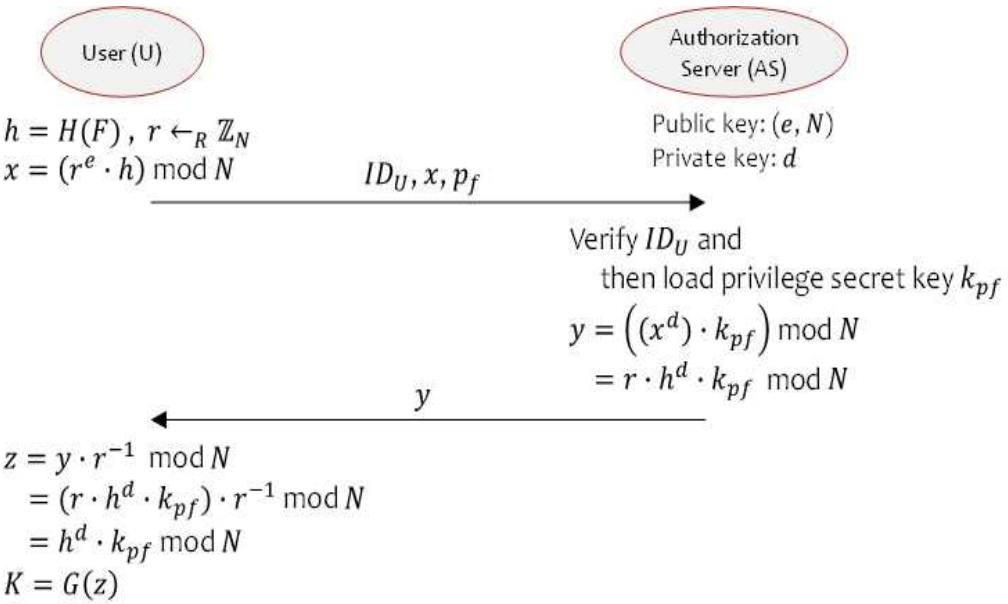
○ 일반적인 설계 기술을 해쉬 함수만 사용하여 구현한 효율적 기법의 경우 매우 적은 연산으로 서버와 클라이언트의 데이터 소유권을 동시에 검증할 수 있음

- 소유권 증명 제공하지 않는 단순한 client-side 암호데이터 중복 처리의 경우에도 태그를 생성하기 위해 최소 해쉬 1번의 연산을 수행해야 함
- 클라이언트의 데이터 소유권만 검증하는 경우에도 증명 정보 생성하기 위해 최소 해쉬 1번 이상의 연산이 수행됨

- 본 연구를 통해 개발된 기술의 경우 서버와 클라이언트가 해쉬 1번의 연산만 계산하여 자신의 소유권 증명 생성 및 상대방의 소유권 검증을 동시에 수행할 수 있음
  - 고정된 짧은 길이를 가지는 입력에 대한 해쉬 계산은 파일 길이에 준하여 증가하는 해쉬 계산에 비해 매우 작아서 파일 길이에 비례하여 증가하는 해쉬 계산만 비용으로 언급하였음

3. 권한에 기반한 암호데이터 소유권 검증 및 중복 처리 프로토콜 설계

○ 제안 기술은 동일 스토리지 서버 내에서도 서비스 그룹에 따라 별도의 권한을 부여함으로써, 다른 서비스 그룹에 존재하는 동일 파일에 의한 안전성 훼손을 방지하고 독립적인 보안 서비스 유지가 가능함
- 파일에 대응되는 키를 생성하는 방식으로는 DupLess에서 도입된 키 서버 기반의 방식을 사용함으로써 동일 파일에 동일 키 값이 생성되도록 구성
- 서비스 그룹별로 상이한 키를 생성함으로써 그룹 간에 동일한 키가 사용되는 것을 방지하여, 타 그룹에서 저장한 동일 데이터에 의한 보안 위협을 방지할 수 있음 (아래 그림은 권한별 키 생성 방법)



<암호데이터 중복 처리를 위한 권한별 키 생성 프로토콜>

○ 각 파일이 저장되는 최대 개수는 서비스 그룹의 수와 동일하여 실제 중복 처리로 인한 저장 공간 관리 비용 개선 효과는 감소하나, 타 그룹과 동일 파일 저장함으로 인해 발생하는 안전성 문제가 해결됨

- 서버가 각 서비스 그룹별로 별도의 데이터 관리를 하는 정책적인 해결 방법도 존재하나, 해당 정책의 반영을 통해 비용 증가가 야기되어 기술적으로 강제할 수 있는 방법이 바람직함
- 제안 기술에서는 키 생성 서버와 스토리지 서버가 상이하게 구성되어 있어, 스토리지 서버가 다른 권한으로 암호화된 파일은 동일 파일에 대해 타 권한으로 암호화된 파일과 다른 파일로 인식하게 되어 권한별 스토리지 분리가 기술적으로 제공됨

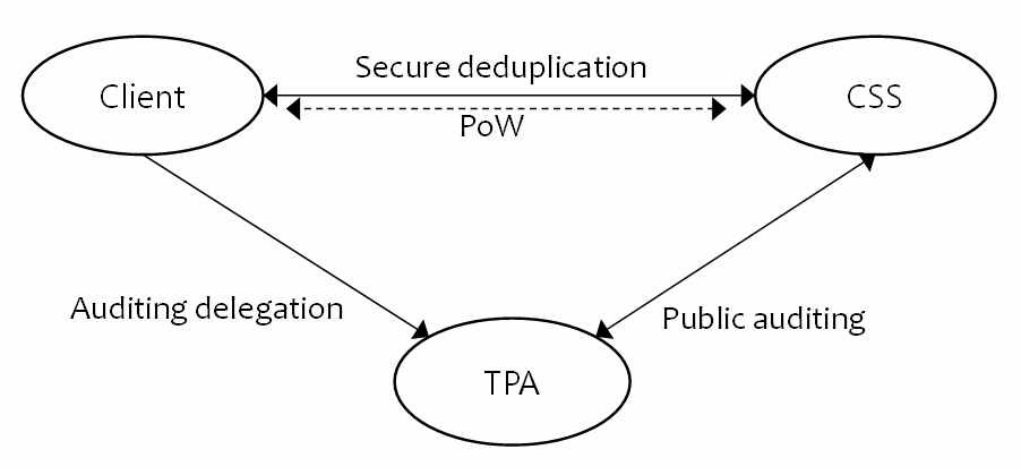
4. 소유권 검증 및 공개 감사 동시에 지원하는 데이터 권한 관리 프로토콜 설계

○ 클라우드 스토리지 서비스에서, 소유권 검증 기능과 공개 감사 기능은 서버와 클라이언트가 안전하고 효율적인 서비스 제공 및 사용을 위해 가장 기본적으로 요구하는 기능임

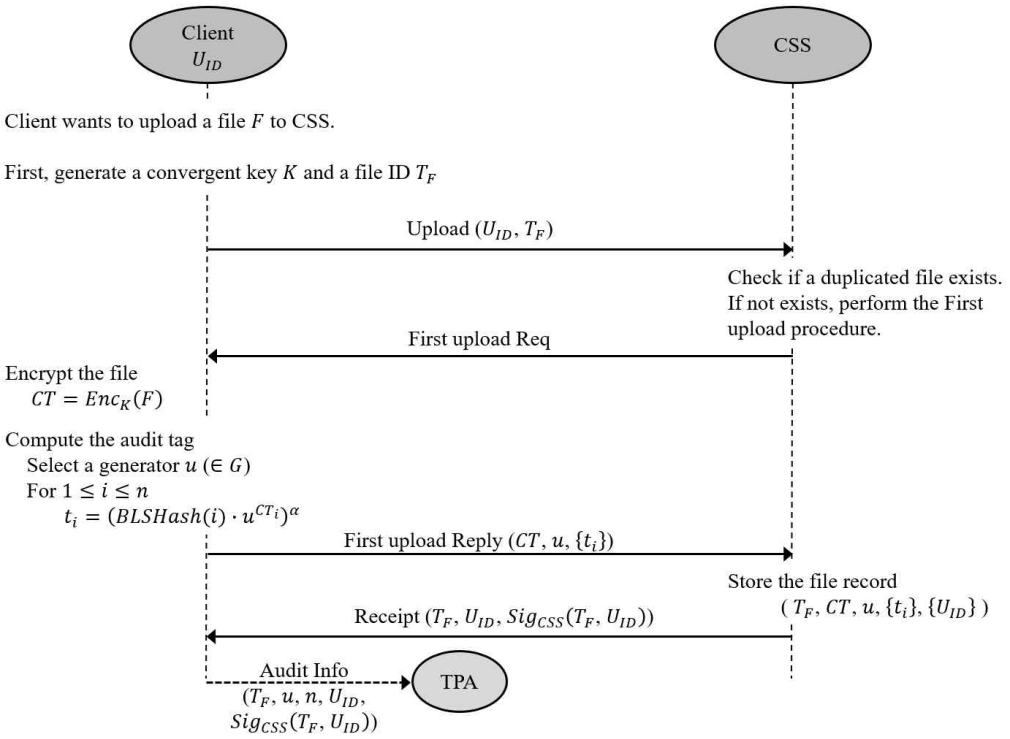
- 사용자는 신뢰되지 않는 클라우드 서비스 제공자(CSP)에게 자신의 중요 데이터를 노출시키지 않기를 원하고, 동시에 아웃소싱된 데이터가 원격의 스토리지에 온전히 잘 보관되어 있다는 것을 확인할 수 있기를 기대함
- CSP 는 자신의 스토리지 공간을 효과적으로 사용하기 위해 중복된 데이터를 저장하지 않으려고 함

○ 제안 기술은 소유권 검증 기능과 공개 감사 기능을 동시에 제공하는 기법을 BLS 서명 기법을 기반으로 제안함

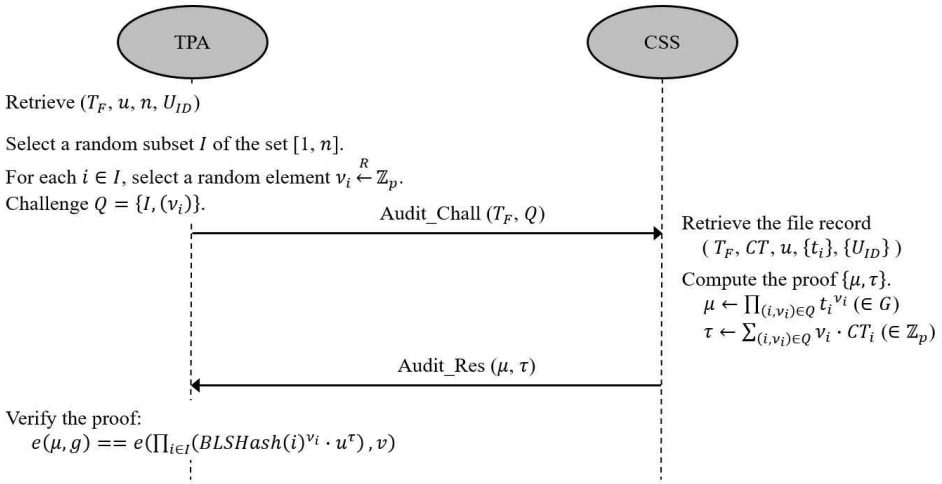
- BLS 서명 기법을 기반으로 설계된 homomorphic authenticator 사용하여 상이한 두 기능에서 계산되는 인증 정보 생성
- 스토리지 서버는 클라이언트와 BLS 서명 기반의 소유권 증명을 수행하여 특정 데이터에 대한 소유권 검증
- 클라이언트는 신뢰 기관인 TPA에 공개 감사 기능 위탁할 수 있는 구조로 설계하여 실제 공개 감사 수행은 TPA와 스토리지 서버 사이에서 수행됨 (아래 개념도 참고)



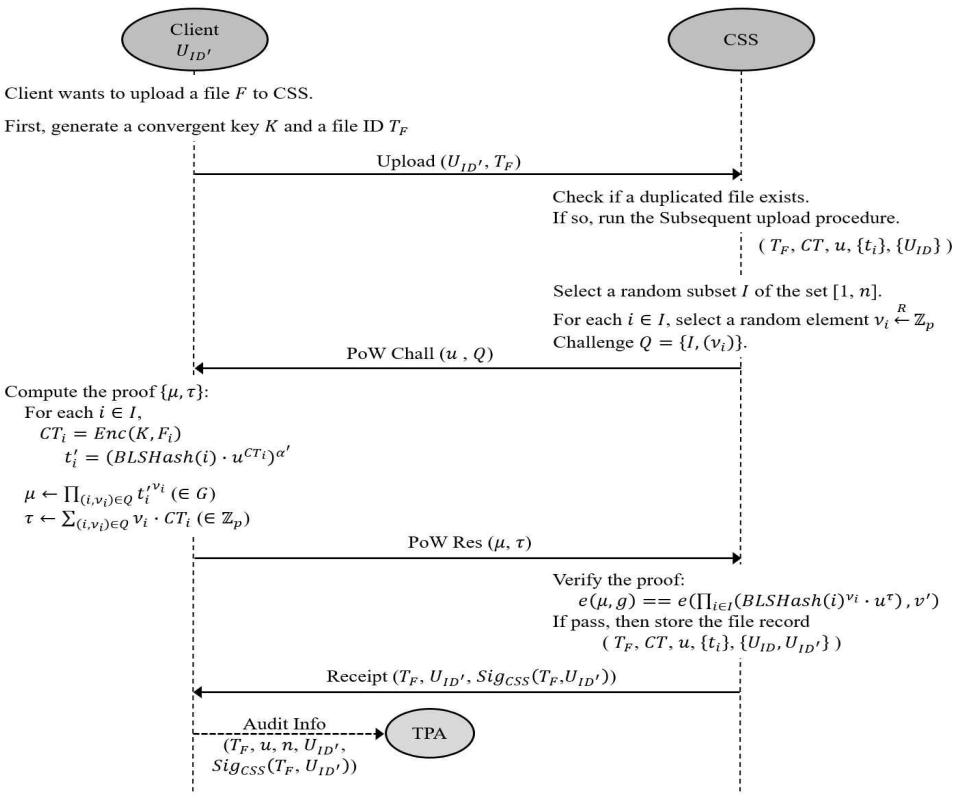
<소유권 검증 및 공개 감사 동시에 지원하는 권한 관리 프로토콜 시스템 모델>



<스토리지 저장되지 않은 파일 최초 업로드 절차도>



<TPA를 통한 공개 감사 절차도>



<동일 클라우드 스토리지에 저장되어 있는 파일 업로드 절차도>

○ 기존 기술과는 달리 특정 정보의 보유 여부를 증명하는 두 기능을 BLS 기반 기술 하나로 제공하여 높은 효율성을 제공하고, 스토리지 서버나 TPA에 사용자 보유 데이터의 평문 정보가 저장되지 않아 높은 프라이버시가 제공됨

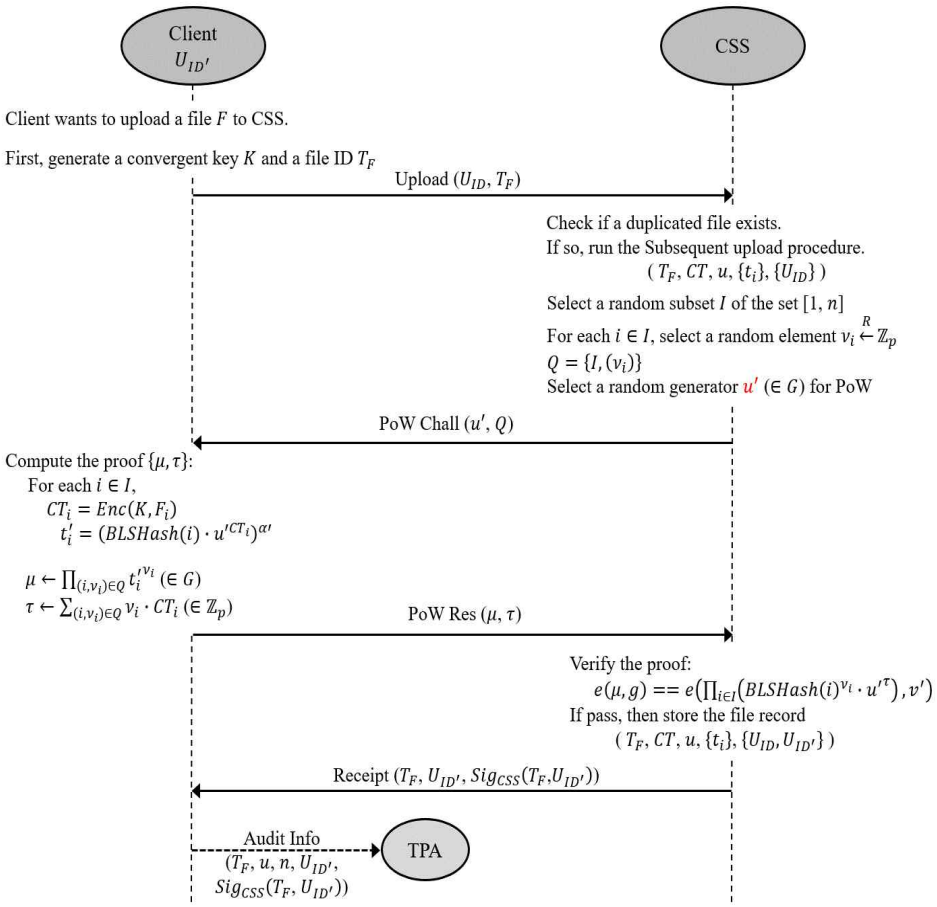
- [YY]는 스토리지 서버가 사용자의 데이터를 평문 형태로 보유하고 있어 안전한 중복 처리 서비스 제공에 제약이 따르고, 각 기능별로 별도의 인증값 생성 알고리즘을 사용하여 두 기능을 하나로 통합하여 제공하는 성능상의 장점이 없음
- [LL]는 공개 감사 기능을 TPA에 위탁하기 위해 사용자가 파일 전체를 TPA에 전달해야 하므로 공개 감사 기능에 프라이버시 관점의 취약점이 존재하고, 효율성 면에서도 클라이언트가 전체 데이터를 TPA에 업로드 해야하는 제약 사항이 존재함

	Secure deduplication	Private public auditing	Efficiency
[YY]	X	O	기능별 별도의 인증 기술이 사용되어 연산 효율성 개선 효과가 없음
[LL]	O	X	사용자가 TPA에 전체 데이터를 업로드하는 통신 비용 발생
제안 기법	O	O	상기의 기술적 제약사항 해결

<제안 기술과 기존 기술의 기능 및 성능 비교표>

○ 가장 기본적인 안전성 요구 사항보다 높은 강력한 공격 모델에서 안전성을 제공하기 위한 대응 기술 제공

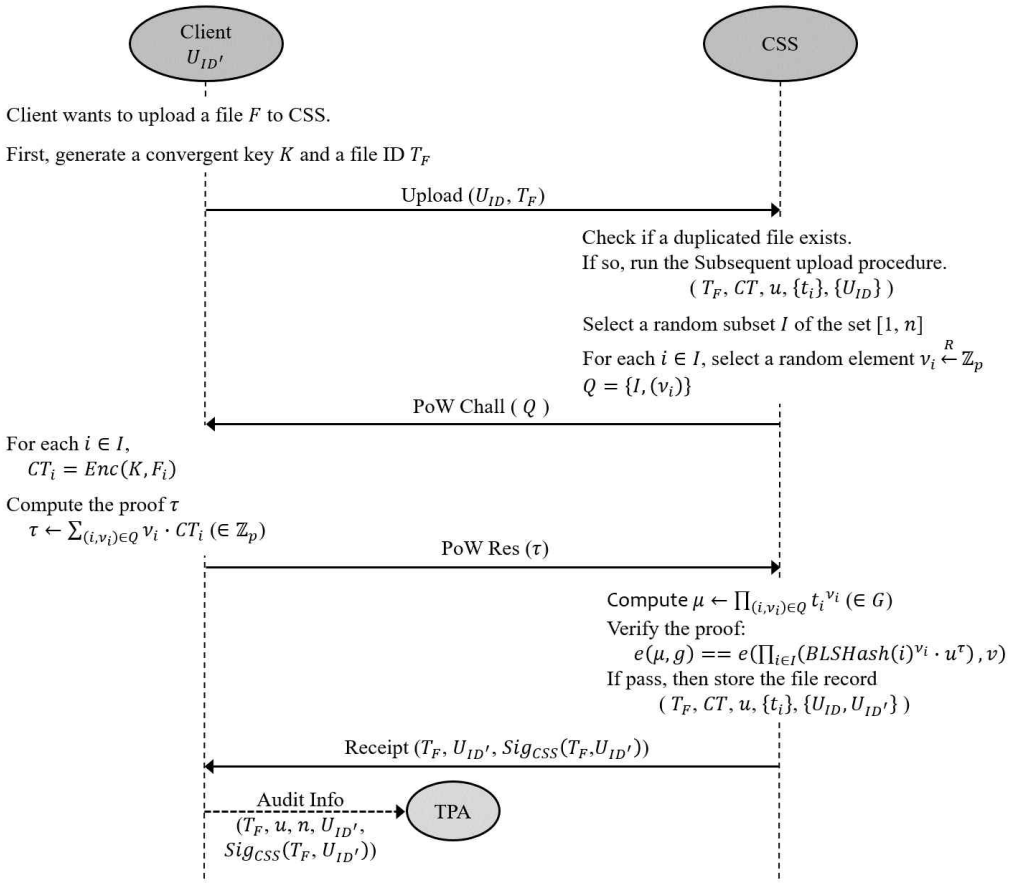
- 실제 데이터 소유한 사용자가 악의적으로 스토리지 서비스를 데이터 분산 네트워크(Data distribution network)로 악용하는 공격자가 존재하는 공격 모델을 가정할 수 있음
- 상기와 같은 강력한 공격자를 감안하고 서비스 제공해야 하는 경우에 충분한 안전성 제공할 수 있는 기술 제안



<안전성 향상을 위한 개선 기법의 소유권 증명 프로토콜 절차도>

○ 저(低)사양 디바이스 기반의 스토리지 서비스에서 소유권 검증 기능과 공개 감사 기능을 효율적으로 제공하기 위한 경량화 기법 제공

- TPA에 위임할 수 있는 사용자 보유 정보를 최대로 확보하고, 이를 기반으로 상당수의 연산을 TPA에게 위임함으로써 주요 프로토콜 수행 비용을 TPA에 전가함
- TPA가 수행해야 하는 연산량이 증가하는 단점이 발생하므로 저사양 디바이스 기반의 서비스 환경 등과 같은 특정 환경을 위한 기술임



<클라이언트 효율성 향상을 위한 개선 기법의 소유권 증명 프로토콜 절차도>

제 4 절 암호데이터 검색 기술 개발

1. 암호데이터 검색 프리미티브 알고리즘 설계

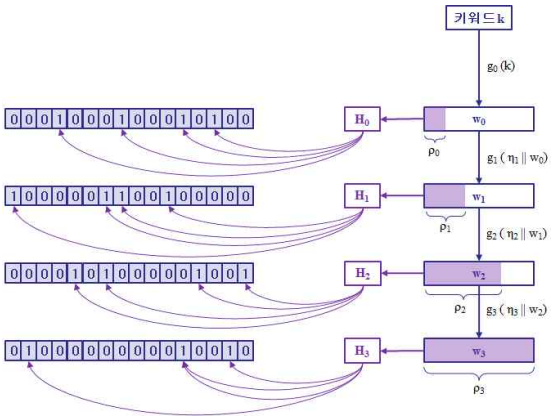
○ 링크드 체인 구조의 검색 인덱스 기법에 대한 확장 검색 기능 연구를 통한 신규 프리미티브 알고리즘 설계

- 동일한 검색 키워드를 포함한 암호데이터를 체인 형태로 구성하는 링크드 체인 검색 인덱스는 공개키 기반 검색 인덱스에 비해 효율적인 키워드 검색이 가능
- 링크드 체인 구조의 검색 인덱스는 검색 시간이 전체 데이터 수에 무관한, sublinear 검색 시간 제공 가능
- 기존의 링크드 체인 구조의 검색 인덱스는 하나의 검색 키워드에 하나의 링크드 체인을 독립적으로 구성하는 방식으로 검색이 제한적인 단점을 지님
- 검색 키워드 사이의 포함 관계를 활용하여, 다수의 링크드 체인 사이의 연결 관계를 정의하고, 링크드 체인 사이의 외부 링크를 통한 연결성 제공을 통해, 다양한 확장 검색 기능 제공을 위한 프리미티브 확보

○ Bloom Filter에 기반한 암호데이터 검색 인덱스 생성 방안 연구

- 집합 포함 관계에 대한 효율적인 검증이 가능한 bloom filter를 기반 기술로 활용한 암호데이터 키워드 검색 기법 연구
- 기존 bloom filter를 이용한 암호데이터 검색 기법 설계 시도가 있었으나, 각각의 데이터에 bloom filter를 단순 적용하는 방식으로 전체 암호데이터에 linear한 검색 시간을 제공하는 비효율적인 기법임
- 각각의 내부 노드가 bloom filter로 구성된 bloom filter tree를 검색 인덱스로 활용하는 암호데이터 키워드 검색 알고리즘 설계

- Tree 구조와 bloom filter의 장점을 결합하여 전체 데이터에 대한 sublinear 검색 시간 제공 가능

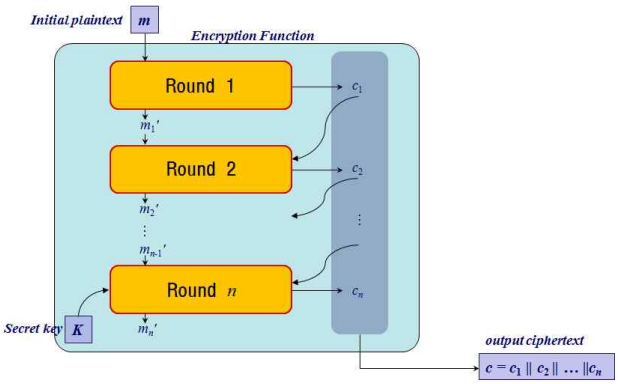


<Bloom filter 기반 암호데이터 검색 인덱스 생성 알고리즘>

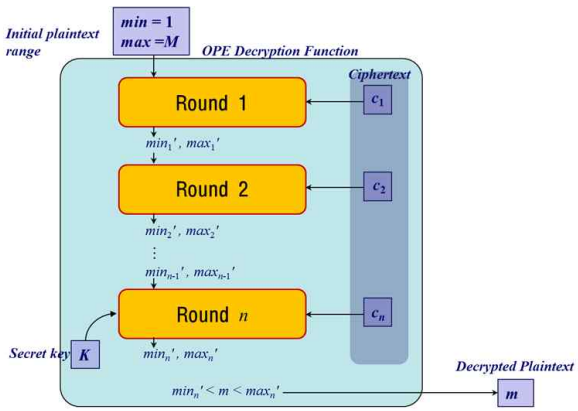
2. 실용적인 암호데이터 검색을 위한 안전성-효율성 조정 가능 암호데이터 검색 기술 설계

○ 수용가능 안전성 적용을 통한 암호데이터 검색 기술에 대한 안전성-효율성 조정 방안 연구

- 제한적인 안전성을 바탕으로 평문의 순서 정보를 보존하는 순서 보존 암호화 기술 설계
- 복호화를 포함한 추가적인 연산 없이 암호데이터에 대한 대소 비교 및 범위 검색 기능 제공
- 수용가능 안전성 적용을 통해 요구 안전성 수준에 따른 암호문/평문 비율 조정 가능
- 순서 보존 암호화 기술에 대한 증명 가능 안전성 개념 적용을 통한 안전성 증명



<유사난수생성기를 활용한 데이터 암호화 알고리즘 구조>



<유사난수생성기를 활용한 데이터 복호화 알고리즘 구조>

○ 구현 효율성 강화를 통한 효율적인 설계 방식 제공

- 임의의 유사 난수 생성기(해쉬 함수, 블록암호 등을 유사 난수 생성기로 사용 가능)가 구현된 환경에서 추가적인 알고리즘 구현 없이 데이터 암/복호화 가능
- 암호문 길이 자유 선택 가능 : 암호화 함수에서의 추가적인 변수 선택을 통해 원하는 수준의 암호문/평문 비율 선택 가능

3. 동적 암호데이터 검색을 위한 안전성 모델 정립

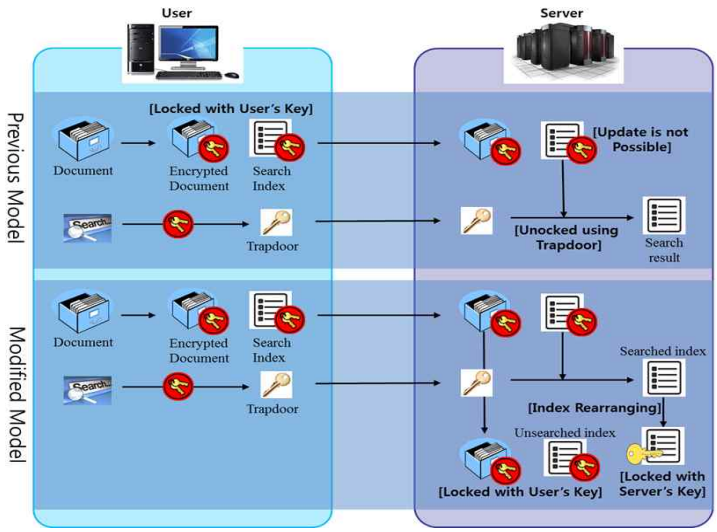
○ 기존 암호데이터 검색 기술의 안전성 모델 분석 및 동적 암호데이터 검색 환경에서의 문제점 도출

- 암호데이터 검색 기술은 데이터에 대한 암호화와 동시에 검색을 위한 ‘검색인덱스’를 추가로 생성
- 검색인덱스는 저장된 데이터의 핵심 키워드에 대한 정보를 포함하고 있기 때문에 데이터 기밀성 보장을 위해 별도의 암호화 과정이 요구됨
- 기존 암호데이터 검색 기술은 저장된 데이터에 대한 높은 기밀성 제공을 목적으로 설계되어 검색인덱스의 생성/확인/수정의 모든 과정이 사용자의 비밀키에 기반하여 처리
- 공개키 기반의 암호데이터 검색 기술은 각각의 데이터에 대한 검색 인덱스를 개별로 생성하기 때문에 검색인덱스 구조면에서 높은 확장성을 지니고 있으나, 검색 과정에서 모든 검색인덱스를 확인해야 하는 비효율성을 포함
- 데이터 사이의 연관성에 기반한 검색인덱스 설계 방식은 높은 검색 효율성을 제공하지만, 데이터가 동적으로 추가/삭제되는 경우 해당 데이터와 연관된 모든 검색인덱스에 대한 수정이 요구되기 때문에 검색인덱스의 수정 과정에 사용자가 참여하는 기존의 모델은 비효율적임

○ 동적 암호데이터 검색 기술 적용 환경 및 안전성 요구조건 분석을 통해 새로운 안전성 모델 제시

- 기존 암호데이터 검색 기술에서는 데이터 서버를 잠재적인 공격자로 설정하며, 높은 안전성 제공을 위해 데이터 서버에 제공되는 비밀 정보를 최소화하도록 설계됨
- 하지만, 데이터 검색이 서버에 의해서 이루어지기 때문에 한 번 검색된 데이터에 대한 일부 정보는 서버에 의해 수집 가능

- 사용자가 제공한 검색인덱스에서 한 번 검색이 수행된 부분의 인덱스를 분리하여 관리하는 안전성 모델 제시



<동적 암호데이터 검색 기술을 위한 안전성 모델>

- 서버는 검색인덱스에서 검색에 활용된 부분 인덱스를 자체적으로 재구성하고 서버의 비밀키로 기밀성을 제공하여 동적인 데이터 활용 환경에서 검색인덱스 수정 과정을 서버 단독으로 수행
- 기존 암호데이터 검색 기술의 안전성을 훼손하지 않으면서 서버에 의한 검색인덱스 관리 기능 제공을 통해 높은 활용성 제공이 가능

4. 동적 데이터 관리 환경을 위한 서버 간 암호데이터 비교 기술 개발

○ 동일성 검사 기능을 제공하는 공개키 암호화 기술(PKEET - Public Key Encryption with Equality Test)에 대한 안전성 모델 정립

- Equality test 기능을 제공하는 공개키 암호화 기술(PKEET - Public Key Encryption with Equality Test)은 서로 다른 공개키로 암호화된 데이터 사이의 데이터 활용을 가능하게 하는 기술로 2010년 이후 연구가 활발히 이루어지고 있음
- 데이터를 하나의 서버에 집중하지 않고 복수의 데이터베이스에 분산 저장 및 활용하는 동적 데이터 관리 환경을 위한 프리미티브 보호 기술
- 서로 다른 기관이 보유하는 민감 정보를 데이터 프라이버시 침해 없이 활용 가능
- 기존의 PKEET 기술은 랜덤 오라클 모델을 사용한 안전성 증명에 기반하고 있어, 현실적 활용을 고려한 표준 모델에 기반한 PKEET 기술 설계 연구 수행
- Post-quantum 암호를 고려한 Lattice 기반 문제 등을 활용한 PKEET 기술 설계 연구 수행
- 기존 PKEET 기술에 대한 안전성 분석을 통해 취약점을 제시하고 이에 대한 계산 효율성과 저장 효율성을 유지하면서 안전성을 개선한 기법 설계

○ 임의의 CDH 가정에 기반한 PKEET 기술 설계를 위한 일반적인 구성 방법 제시

- 임의의 CDH 가정에 기반한 공개키 암호데이터 검색 기술에 대해, 서버 간 암호데이터 비교 기능을 제공하는 일반적인 구성 방법 설계
- Trapdoor 제공 여부를 기반으로 Type-I, Type-II 공격자 모델 제시
- 랜덤 오라클 모델에서 Type-1 공격자에 대한 OW-ID-CCA2 안전성과 Type-2

공격자에 대한 IND-ID-CCA2 안전성을 최초로 달성하는 PKEET 설계

		[19]	[18]	[12]†	Ours
Comp of	Enc	3Exp	5Exp	6Exp	6Exp + 2SE
	Dec	3Exp	2Exp	5Exp	3Exp + 2SE
	Test	2Pairing	4Exp	2Pairing + 2Exp	2Exp + 2SE
Size of	PK	$\|\mathbb{G}\|$	$2\|\mathbb{G}\|$	$3\|\mathbb{G}\|$	$3\|\mathbb{G}\|$
	CT	$3\|\mathbb{G}\| + \|\mathbb{Z}_p\|$	$4\|\mathbb{G}\| + \|\mathbb{Z}_p\| + 2\lambda$	$5\|\mathbb{G}\| + \|\mathbb{Z}_p\|$	$2\|\mathbb{G}\| + 10\lambda$
	TD	–	$\|\mathbb{Z}_p\|$	$\|\mathbb{Z}_p\|$	$\|\mathbb{Z}_p\|$
Security	Type-I	OW-CCA2	OW-CCA2	OW-CCA2	OW-CCA2
	Type-II	–	IND-CCA2	IND-CCA2	IND-CCA2
Assumptions		CDH	CDH	CDH	CDH

<PKEET 기술 성능 및 안전성 비교>

- 동일성 검사 기능을 제공하는 ID 기반 암호화 기술(IBEET : ID-Based Encryption with Equality Test)로 확장 설계 방식 제시

		IBEKS ([1]+[14])	IBEET ([11])	Ours (with BF-IBE [4])
Comp of	Enc	7Exp	6Exp	6Exp
	Dec	-	2Pairing+2Exp	3Pairing+2Exp
	Test	4Pairing	4Pairing	2Pairing+2Exp
Size of	PK	15$\mathbb{G}$	2$\mathbb{G}$	4$\mathbb{G}$
	CT	4$\mathbb{G}$+$\mathbb{G}_T$	4$\mathbb{G}$+$\mathbb{Z}_p$	2$\mathbb{G}$+5$\|\mathcal{H}\|$
	TD	12$\mathbb{G}$	$\mathbb{G}$	$\mathbb{G}$
Fun	KS	Yes	Yes	Yes
	ET	No	Yes	Yes
Security		IND-ID-CPA	OW-ID-CCA2	IND-ID-CCA2
ROM		No	Yes	Yes
Assumption		ℓ-DBDHE & aug ℓ-DL	BDH	BDH

<IBEET 기술 성능 및 안전성 비교>

5. Bloom Filter 기반의 동적 암호데이터 검색 기술 설계

○ 동적 암호데이터 검색을 위한 bloom filter tree 구조 최적화 연구

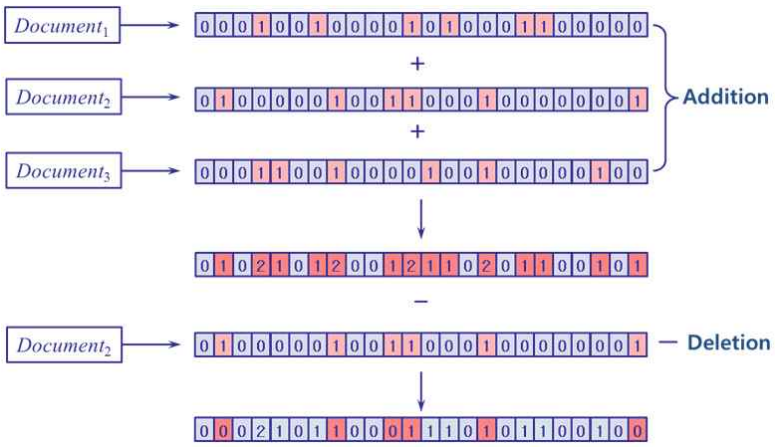
- 수용가능 안전성 개념 적용을 통해 요구 안전성 수준 및 저장 데이터 규모에 따른 bloom filter 검색 인덱스에 대한 parameter 최적화 연구
- Tree형태의 bloom filter 결합에서 발생하는 filter saturating 문제를 해결하기 위해 tree의 상위 노드에 대한 modified 키워드 집합 기반의 filter 생성을 통한 해결 방안 제시
- Bloom filter 기반 암호데이터 검색 기술에 대한 체계적인 안전성 분석 수행
- Bloom filter tree를 구성하는 각 filter 사이의 연산을 통한 확장 검색 기능 연구 및 counting filter 등의 개념 도입을 통한 동적 환경 최적화 연구 수행

○ Bloom Filter Tree 기반의 검색 인덱스를 활용한 동적 암호데이터 검색 기술 설계

- Bloom filter는 집합에서 원소의 포함 관계를 효율적으로 확인 가능한 데이터 구조로 tree의 각 노드가 bloom filter로 구성되는 bloom filter tree 구조의 검색 인덱스 구성을 통해 O(m log n)의 암호데이터 검색이 가능

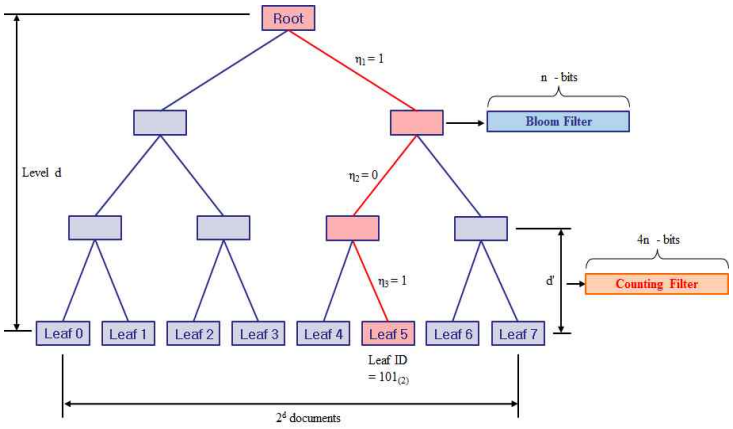
  - n : 전체 데이터 수, m : 검색된 데이터 수
- Bloom filter 기반의 검색 인덱스는 각각의 키워드에 대한 해쉬 연산을 통해 얻어진 bloom filter의 한 원소의 값을 1로 변환하는 방식으로 구성되기 때문에, 서로 다른 키워드를 포함하는 두 bloom filter에 대한 Bitwise-OR 연산을 통해 두 집합의 union 연산 가능
- Bitwise-OR 연산을 통한 Bloom Filter 중첩을 활용하여 자유로운 데이터 추가 기능을 제공

- 각각의 원소가 하나의 bit 정보로 표현되는 기본적인 bloom filter를 각각의 원소가 일정 범위의 정수로 표현되는 counting bloom filter로 확장하고, 각 원소 단위의 덧셈/뺄셈을 활용하여 데이터에 대한 추가/삭제 기능을 제공하는 검색 인덱스 갱신 방법 제시



<Counting Bloom filter를 활용한 데이터 추가/삭제>

- 검색 인덱스에 대한 저장 효율성을 위해 일반 bloom filter와 counting bloom filter를 융합한 bloom filter tree 구성



<Bloom filter tree 구조>

- Bloom filter tree를 활용한 검색 인덱스 구성으로 O(m log n)의 검색 성능을 제공하면서 자유로운 데이터 추가/삭제가 가능한 암호데이터 검색 기술 설계

	[Goh04]	[CGK06]	[KPR12]	Ours
Index Size	O(n)	O(n+k)	O(n+k)	O(n)
Search Time	O(n)	O(m)	O(m)	O(m log n)
Basic Structure	Bloom Filter	Linked Chain	Linked Chain	Bloom Filter Tree
Dynamic Update	Static	Static	O(1)	O(1)
			- O(n) size deletion table 필요 - linked chain의 update를 과정에서 일부 security 훼손	- Counting bloom filter 사용을 통한 deletion

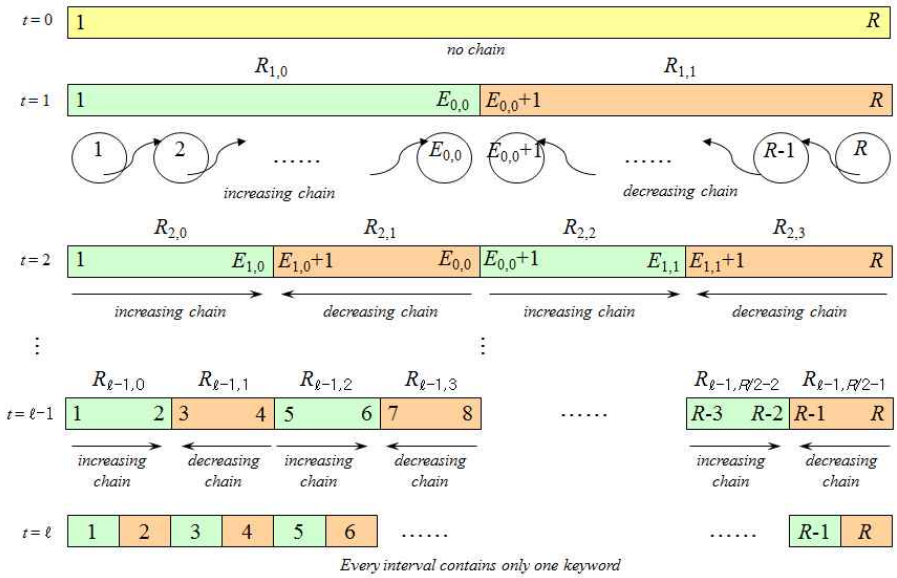
<Bloom filter tree를 활용한 동적 암호데이터 검색 기술 비교>

6. 다중 링크드 체인 구조 기반의 범위 검색 기술 설계

○ 링크트 체인 구조에 기반한 확장 검색 기술 연구

- 링크드 체인을 통해 sublinear 검색 성능을 제공하는 효율적인 검색 인덱스 생성이 가능하지만, 데이터 사이의 연관성에 기반한 설계 방식으로 확장 검색 제공이 극히 제한적임
- 기존 링크드 체인 구성 방법은 하나의 검색 키워드를 포함하는 데이터 집합을 하나의 링크드 체인 형식으로 구성하며, 각각의 링크드 체인 사이의 연관성을 활용하지 않는 방식임
- 범위 검색의 경우, 검색 키워드는 순서를 정의할 수 있는 수치화된 키워드이며 따라서 각 키워드를 포함하는 링크드 체인 사이에도 순서 관계가 성립
- 복수의 링크드 체인에 대한 검색을 하나의 링크드 체인에 대한 검색처럼 활용이 가능한 외부 링크(external link) 기법 제시

- 검색 범위 R을 최대 log R개의 부분 검색 범위로 구분하여 임의의 검색을 처리하기 위한 다중 링크드 체인 구조 설계



<범위 검색을 위한 다중 링크드 체인 구조>

○ 다중 링크드 체인 구조를 이용한 암호데이터에 대한 효율적인 범위 검색 기법 연구

- 기존의 암호데이터 범위 검색은 공개키 방식의 기법이 주류를 이루고 있으며, 전체 저장된 데이터에 대한 linear 검색 시간으로 현실 데이터 활용 환경에 부적합
- 대칭키 암호 프리미티브를 이용한 다중 링크드 체인 구조를 설계하고, 이에 기반한 암호데이터 범위 검색 기법 제시
- 전체 검색 시간이 전체 암호데이터 수에 무관한 sublinear 검색 시간을 제공하는 범위 검색 기법 제공

	Enc. Type	Range Query	Efficiency		
			Index Size	Trapdoor Size	Search Time
Curtmola, et al.	Symmetric	N	O(N)	1 Link (address, s.key)	m symmetric decryptions
Boneh, Waters	Public	Y	O(NR)	3 points over EC	(2R+1)N Pairings
Shi, et al.	Public	Y	O(N log R)	5 log R points over EC	5N log R Pairings
Ours	Symmetric	Y	O((N+R) log R)	2 Links (address, s.key)	m+R symmetric decryptions

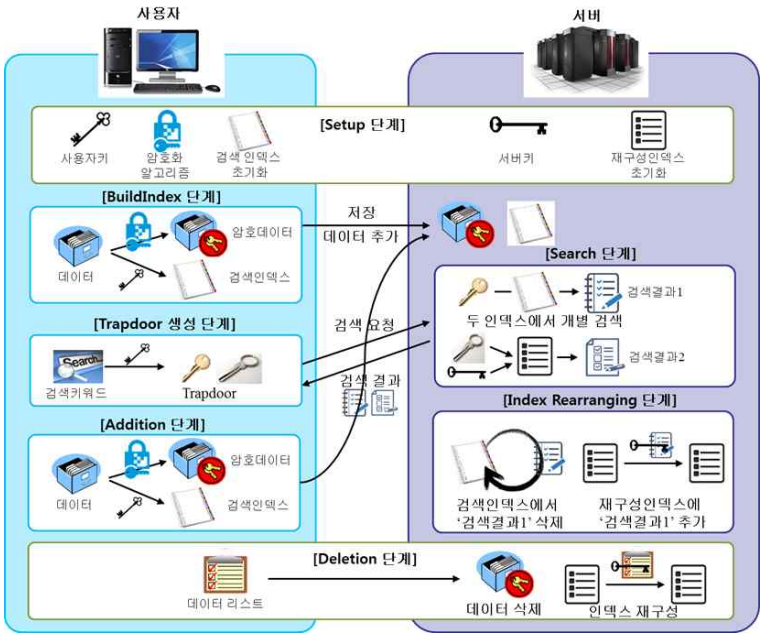
N : number of all documents
m : number of searched documents
R : total range
EC : elliptic curves

<암호데이터 범위 검색 기법 비교>

7. 암호데이터 범위 검색 기법에서의 인덱스 재구성을 통한 동적 데이터 처리 기술 설계

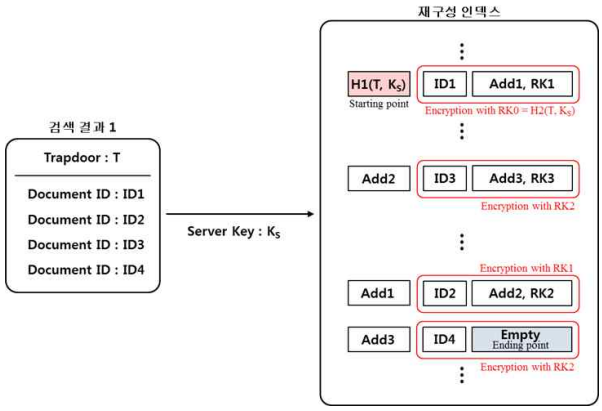
○ 개선된 안전성 모델을 바탕으로 검색 인덱스 재구성을 통한 효율적인 데이터 추가/삭제가 가능한 암호데이터 검색 기술 설계안 제시

- 기존 암호데이터 검색 기술에서 데이터에 대한 기밀성 보장을 위해 검색 인덱스는 사용자의 비밀키에 의해서 암호화되며, 서버 등의 다른 주체에 의한 수정이 용이하지 않음
- 암호데이터에 대한 검색 과정에서 검색에 활용된 검색인덱스에 대한 일부 정보는 서버에 공개되기 때문에 공개된 부분 인덱스에 대한 안전성 모델 수정을 통해 효율적으로 검색인덱스를 관리할 수 있는 기법 설계
- 검색에 활용된 부분 인덱스를 기존의 검색인덱스에서 분리하여 서버가 재구성하는 인덱스 재구성 기법 제시
- 사용자가 제공한 검색인덱스와 서버가 구성한 재구성인덱스를 동시에 활용하는 검색 기법 설계를 통해 동일한 키워드 검색에 대해 사용자가 제공한 검색인덱스 만을 사용한 검색 보다 빠른 검색 가능



<검색 인덱스 재구성을 통한 암호데이터 검색 기술 개념도>

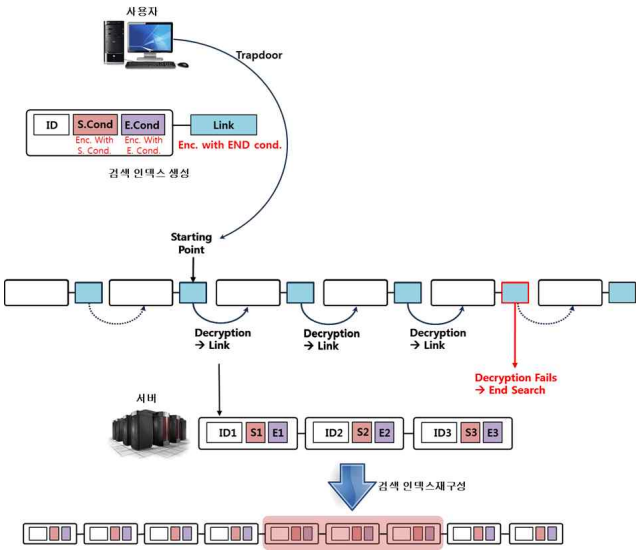
- 한 번 검색된 데이터에 대한 인덱스 재구성 과정에서 기존의 검색인덱스에서 삭제되기 때문에 데이터 추가의 경우 사용자는 신규 데이터와 동일한 검색인덱스 생성
- 재구성된 인덱스는 서버의 비밀키를 통해 기밀성이 보장되어 재구성인덱스에 포함된 데이터 대한 삭제 과정은 서버에 의한 독자적인 인덱스 수정 과정으로 수행
- 검색 인덱스 재구성 방식 변화를 통해 다양한 추가 기능 제공을 위한 기술적 토대 제공



<Linked List를 이용한 검색인덱스 재구성 기법>

○ 동적 암호데이터 범위 검색을 위한 인덱스 재구성 기법 설계

- 임의의 암호데이터 범위 검색 기법에 적용하여 효율적인 동적데이터 추가/삭제 처리가 가능한 인덱스 재구성 기법 설계
- 기반 기법에서 검색된 데이터를 효율적인 범위 검색이 가능한 링크드 리스트 형태의 확장 인덱스로 재구성



<Linked List를 이용한 검색인덱스 재구성 기법>

- 인덱스 생성 과정에서 각 데이터에 대해서 기반 기법의 인덱스에 추가적으로 범위 검색 질의에 대한 포함관계를 확인하기 위한 'starting condition'과 'end condition'의 두 조건에 대한 확장 인덱스 생성
- 속성 기반 암호 기술 적용을 통해 서버의 확장 인덱스 재구성 과정에서 검색 키워드에 대한 정보 유출 수준을 최소화하고, 기반 범위 검색 기술과 동일한 안전성 제공
- 한 번 검색이 수행된 데이터에 대한 기반 인덱스는 삭제, 서버는 확장 인덱스를 하나의 링크드 체인으로 재구성
- 검색은 기반 범위 검색 기법을 통한 검색과 확장 인덱스 검색으로 구성
- 확장 인덱스를 통한 검색은 최대 O(m)이며, 기반 기법에 따라서는 $O(2\log N_2)$로 검색 가능

		Curtmola, et al. [CGK06]	Boneh, Waters [BW07]	Shi, et al. [SBC07]	Ours Linked Chain	Ours Ext. Index
Index Size	Basic Index	O(N)	O(NR)	O(N log R)	O((N+R) log R)	Same to Basic Scheme (with N_1 docu.)
	Ext Index	-	-	-	-	$O(N_2)$
Search Time	Basic (Range Query)	O(m) (not range query)	(2R+1)N	5N log R	m+R	Same to Basic Scheme (with N_1 docu.)
	Ext Search (Range Query)	-	-	-	-	O(m)
Dynamic Update		None	O(1)	O(1)	None	Init. Step of Basic Scheme
Deletion		None	O(1)	O(1)	None	O(1)

N : number of all documents (= N_1 + N_2)
N_1 : number of documents which are not searched ever
N_2 : number of documents which are searched at least once
m : number of queried documents
R : total range

<인덱스 재구성을 통한 동적 암호데이터 처리 효율성>

제 5 절 연구 성과

1. 연구 결과의 질적 우수성

○ 암호데이터 중복 처리 기술 및 소유권 검증 개발

- 안전성과 효율성을 동시에 만족하는 암호데이터 중복 처리 알고리즘 개발
  - 메시지 기반의 암호화 프리미티브 설계
    * 안전성과 효율성을 동시에 만족하는 client-side 메시지 기반 암호화 프리미티브 설계
    * 기존 CE 알고리즘 대비 네트워크 전송량 및 서버 계산량 감소
    * 데이터 위조 공격 및 데이터 삭제 공격 방지 기능 제공
  - Lego 타입의 암호데이터 중복 처리 기술 설계
    * 키 서버, PoW 등 다양한 중복 처리 기법과 결합 가능
    * 사용자의 보안 수준 선택에 따라 Contents guessing attack, Poison attack 등에 대한 역기능 방지
    * 실제 응용 환경(Openstack Swift 스토리지)에서 알고리즘 구현 및 성능 평가
  - 데이터 손실 공격을 방지하는 우수한 중복 처리 성능
    * 1MB 평문 파일 대비 암호 파일 중복 처리를 위한 부가 시간 : 61ms/1MB (경쟁기술 460ms/1MB)
    * 파일 업로드 시간 보다 중복처리를 위한 부가 시간이 훨씬 더 작으므로, 파일 크기가 클수록 효과적임
    * 1MB 대비 33MB 암호파일 업로드 시간 : 30배(중복 처리 안할 때) / 3배(중복 처리 할 때) 증가

- 파일 신원 확인 공격에 대한 근본적인 해결 방안에 대한 연구를 수행하여 기존 기술이 제공할 수 있던 안전성 강도보다 높은 안전성을 제공할 수 있는 대응 기술 설계
  - 서비스 제공자가 안전 강도를 충분히 높이지 않아도 확인하기 어려운 기존의 파일 신원 확인 공격 대응 기술과 달리 사용자가 직접 안전성 강도를 선택할 수 있는 기술을 최초로 개발함
  - 프로토콜 분석을 통한 파일 신원 확인 공격보다 강력한 내부자 공격이나 공모 공격에 대한 안전성 제공할 수 있는 사용자 관리 기술 개발
- 스토리지 서비스에서 발생하는 다양한 보안 문제의 해결을 위해 접근 제어 방식을 탈피한 데이터 기반의 소유권 관리 기술 제공
  - 권한 부여를 위한 사용자의 소유권 검증, 업로드 과정에서 저장 데이터 존재 확인을 위한 스토리지 서버의 소유 여부 검증, 주기적인 저장 데이터 무결성 검증을 위한 스토리지 보유 데이터 검증, 다양한 권한 분류에 따른 소유권 관리 등 다양한 데이터 소유권 관련 요소 기술 개발
  - 안전한 데이터 소유권 관리를 위해 개발된 총 3종의 기술은 기존 제공하지 않는 최초의 기술이거나 기존 기술 대비 향상된 안전성 제공

○ 암호데이터 검색 기술 개발
- 동적 암호데이터 활용 환경을 새로운 안전성 모델 정립을 통한 암호데이터 검색 프리미티브 기술 확보
  - Bloom filter tree, 다중 링크드 체인 등의 독창적인 검색 인덱스 구조 설계 및 다양한 부가 검색 기능 제공을 위한 핵심 프리미티브 기술 확보
  - 암호데이터 검색 기술에 대한 안전성 모델 수정을 통한 서버의 자체적인 검색 인덱스 재구성 기법 제시
  - 동적 데이터 관리 환경을 위한 서로 다른 공개키로 암호화된 데이터에 대한 동일성 검사가 가능한 공개키 암호화 기술 안전성 모델 정립 및 기술 설계

- 검색 시간이 전체 암호데이터 수에 무관한 sublinear 검색 성능 제공 암호데이터 범위 검색 기술 설계
  - 기존 범위 검색 기술은 공개키 위주의 방식으로 검색 성능이 전체 데이터의 수에 비례
  - Sublinear 검색 성능을 제공하는 암호데이터 검색 기술은 키워드 사이의 연관성에 기반하는 형태적 특성에 의한 제약으로 효율적인 부가 기능 제공이 어려움
  - 범위 검색의 특성을 바탕으로 링크드 체인 사이의 연관성을 정의하고, 이를 활용한 다중 링크드 체인 구조 설계하여 sublinear 검색 성능을 제공하는 범위 검색 기술 제시
- 동적 암호데이터 범위 검색을 위한 인덱스 재구성 기법 설계
  - 임의의 암호데이터 범위 검색 기법에 적용하여 효율적인 동적데이터 추가/삭제가 가능한 인덱스 재구성 기법 설계
  - 키워드에 대한 속성 기반 암호 기술 적용을 통해 기반 범위 검색 기술과 동일한 안전성을 보장하는 서버의 인덱스 재구성 과정 설계
  - 다중 링크드 체인 기반의 범위 검색 기법과 결합하여 동적 암호데이터에 대한 O(m)의 범위 검색 성능 제공, O(1) 삭제 기능 제공

○ Information Sciences, IEEE Transactions on Computer(IF 상위 20% SCI) 게재 등 SCI(E) 논문 10건 게재 및 4건 게재 승인

2. 정량적 연구 성과

○ SCI(E) 논문 실적

번호	구분	논문명	논문발표학회명 또는 게재지	년도, 권호	SCI(E) 여부	IF
1	게재	Efficient construction of order-preserving encryption using pseudo random function	IEICE Transactions on Communications	2015, E98-B(7)	SCI(E)	0.827
2	게재	Low complexity multiplier based on Dickson basis using efficient Toeplitz matrix-vector product	IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences	2015, E98-A(11)	SCI(E)	0.274
3	게재	Semi-generic transformation of revocable hierarchical identity-based encryption and its DBDH instantiation	IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences	2016, E99-A(1)	SCI(E)	0.274
4	게재	Explicit formulae for Mastrovito matrix and its corresponding Toeplitz matrix for all irreducible pentanomials using shifted polynomial basis	Integration - the VLSI journal	2016, 53	SCI(E)	1.000
5	게재	Symmetric searchable encryption with efficient range query using multi-layered linked chains	Journal of Supercomputing	2016, 72(11)	SCI(E)	1.326
6	게재	Efficient multiplication based on Dickson bases over any finite fields	IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences	2016, E99-A(11)	SCI(E)	0.274

7	게재	CCA2 attack and modification of Huang et al's public key encryption with authorized equality test	Computer Journal	2016, 59(11)	SCI(E)	0.711
8	게재	Semi-generic construction of public key encryption and identity-based encryption with equality test	Information Sciences	2016, 372	SCI(E)	4.832
9	게재	Client-Side deduplication to enhance security and reduce communication costs	ETRI Journal	2017, 39(1)	SCI(E)	1.116
10	게재	New block recombination for subquadratic space complexity polynomial multiplication based on overlap-free approach	IEEE Transactions on Computers	2017, 66(8)	SCI(E)	2.916
11	게재 승인	Design of additive homomorphic encryption with multiple message spaces for secure and practical storage services over encrypted data	Journal of Supercomputing	2018 예정	SCI(E)	1.326
12	게재 승인	Bi-directional and concurrent proof of ownership for stronger storage services with de-duplication	Science China-Information Sciences	2018 예정	SCI(E)	1.628
13	게재 승인	Authorized client-side deduplication using access policy-based convergent encryption	Journal of Internet Technology	2018 예정	SCI(E)	1.930
14	게재 승인	Efficient multiplier based on hybrid approach for Toeplitz matrix-vector product	Information Processing Letters	2018 예정	SCI(E)	0.748

○ 특허 실적

번호	구분	특허명	출원번호	출원국	출원일
1	국내		10-2016-0000615	한국	2016.1.4
2	국내		10-2016-0000623	한국	2016.1.4
3	국내		10-2016-0025951	한국	2016.3.3
4	국내		10-2016-0051127	한국	2016.4.26
5	국내		10-2016-0143636	한국	2016.10.31
6	국내		10-2016-0149839	한국	2016.11.10
7	국내		10-2017-0026599	한국	2017.2.28
8	국내		PR20170827 (출원 중)	한국	2017.10.30. (제출일)
9	국내		PR20170966 (출원 중)	한국	2017.11.17. (제출일)
10	국내		PR20170977 (출원 중)	한국	2017.11.20. (제출일)
11	국외		15/051574	미국	2016.2.23
12	국외		15/179375	미국	2016.6.10

13	국외		15/586180	미국	2017.5.3
14	국외		PR20170827 (출원 중)	미국	2017.10.30. (제출일)
15	국외		PR20170966 (출원 중)	미국	2017.11.17. (제출일)
16	국외		PR20170977 (출원 중)	미국	2017.11.20. (제출일)

제 4 장 연구 개발 결과의 활용 계획

제 1 절 연구결과의 활용 가능성

○ 성과 활용 방안

- 안전하고 스마트한 초연결 사회로 진화하는 환경에서 성능 문제로 적용이 지연되고 있는 데이터 암호화 도입을 위한 원천 기술로 활용하여, 초연결 서비스의 신뢰성 향상 및 사회적 현안 해결
  - 암호데이터에 대한 저장 및 검색 기능을 평문과 유사한 수준으로 제공함으로써 현재 평문 데이터 대상으로만 제공되는 서비스 영역을 암호데이터로 확장, 나아가서 데이터 서비스 전체의 신뢰성을 향상시키기 위한 원천 기술로 활용
  - 암호데이터 중복 처리 기술은 암호데이터 저장 단계에서 스토리지 및 네트워크 비용 절감을 위한 원천 기술로 활용
  - 금융, 의료, 교육 등 공공 부문의 데이터를 위탁 활용하는 서비스 환경을 위한 데이터 암호화 및 저장/검색 기술로 활용
  - 클라우드/빅데이터 환경의 동적으로 변화하는 데이터를 활용하는 환경의 데이터 보호를 위한 핵심 기술로 활용
  - 안전한 초연결 서비스 보안 인프라 구축 및 신규 지능형 사이버 보안 시장 창출에 활용
- 전 세계적으로 추진되고 있는 데이터 위탁 서비스 환경을 대비하여 향후 암호화된 데이터를 중심으로 한 신 데이터 보안 패러다임인 CipherData 트렌드 선도
  - 정책적으로 금융, 의료, 교육을 비롯한 공공 부문 데이터에 대한 위탁 환경을 통한 서비스가 추진되고 있으며, 민간 부문의 데이터 또한 이러

한 변화를 따를 것으로 전망

- 데이터 프라이버시에 대한 관심이 커지면서 민감 데이터에 대한 암호화 적용은 사회적인 요구 사항으로 발전하고 있으나, 기술적인 성능 문제로 인해 현실 서비스에 적용이 지연될 것으로 예상
- 이러한 상황을 해결하기 위해 관련 기관 및 기업의 효율적인 암호화된 데이터 활용 기술에 대한 요구가 커지고 있어, 본 과제의 암호데이터 저장/검색 기술이 다양한 방면의 암호데이터 활용 환경에 적용될 수 있을 것으로 기대함
- 또한, 이러한 핵심 기술을 바탕으로 암호데이터를 평문 데이터처럼 활용하는 신 데이터 보안 패러다임인 CipherData 트렌드를 창출 및 선도할 것으로 기대

○ 연구목표 달성 시 활용분야 파급성

- CipherData 패러다임의 데이터베이스 적용 시스템인 CipherDB 시스템은 광범위하고 난해한 원천 기술 및 실제 데이터베이스 시스템 구축을 동시에 요하는 분야로 메가 프로젝트로 기획 가능
  - 본 연구에서 개발된 암호데이터 저장, 열람 및 검색 기술을 기반으로 암호데이터 연산, 처리, 무결성 검증 등의 추가 요소 기술 개발, 이종 기능 통합 알고리즘 개발, 다자간 환경으로 적용 환경 확대를 통해 CipherDB 원천 기술로 확대 개발
  - 실 데이터베이스 환경 적용을 위한 통합 시스템 구축을 통해 CipherDB 시스템 기획 추진 가능
- 상용화를 위한 정부 수탁과제로 기획 가능
  - 현재 국내외 데이터베이스 보안 제품은 접근 제어나 단순 암복호화 기능 위주로 적용
  - 본 연구를 통해 개발된 암호데이터 저장, 열람 및 검색 기술을 상용 데

이터베이스나 클라우드 시스템에 적용함으로써 사업화 성공 가능성을 높일 수 있을 것으로 사료됨

- 데이터베이스 보안 업체와의 협력을 통한 기술 고도화 및 상용화 과제로 기획 가능

○ 성과확산계획

구분	내용
기술	○ 암호데이터 중복 처리 기술 ○ 암호데이터 검색 기술
성과확산 전략	○ 관련 유관 기관과의 긴밀한 협력을 통해 실 서비스 환경의 요구 사항 반영한 핵심 원천 기술 및 IPR을 확보 ○ 확보된 원천 기술을 기반으로 기술 고도화 및 상용화 과제 기획을 통한 사업화 추진
활용	○ 금융, 공공기관 및 민간 데이터베이스 보안 기술로 활용 ○ 클라우드 서비스에서 스토리지 절감 및 데이터 보호 분야에 활용

제 2 절 기대 효과

○ 기술적 기대효과

- 기존 보안 기술이 지니는 한계를 극복한 새로운 데이터 중심 보안으로의 변화를 선도하기 위한 핵심 원천 기술로 활용
  - 정부, 금융 기관 등 대량의 개인정보를 관리하는 기관의 데이터베이스에 대한 근본적인 데이터 유출 방지 시스템 구축을 위한 핵심 기술로 활용

- 데이터베이스 단순 암복호화 기능에서 벗어난 암호데이터 저장, 열람 및 검색 등과 같은 암호데이터 활용 기술에 대한 원천 IPR 확보를 통한 핵심 기술 선점 및 선도 가능

• 클라우드/빅데이터 서비스 확산에 요구되는 프라이버시 보호 관련 기술의 고도화 견인

- 암호데이터 저장, 열람 및 검색 기술은 클라우드/빅데이터 서비스 등과 같은 다양한 신규 서비스로의 확대 적용이 가능해 기술적 파급 효과가 매우 높음
- 개인 데이터의 안전성을 보장하면서, 이를 활용하여 새로운 부가가치를 창출할 수 있는 데이터 공유 및 거래 프레임워크 설정을 위한 핵심 기술로 활용 가능

• 미래 IoT 환경에서 예상되는 대규모 데이터에 대한 저장 시스템의 안전성과 저장 성능 향상을 위한 원천 기술 선점 가능

-

○ 경제적 기대효과

• 국내 DB 산업은 2011년 이미 10조 이상의 규모를 형성하고 있어, 데이터 보호가 의무화될 시 DB 보호를 위한 기술 필요성이 증대되어 DB 보안 솔루션 산업도 폭발적으로 성장할 것으로 예상됨 (출처 : 한국데이터베이스진흥원, 2011년도 국내 데이터베이스 산업 시장 분석 결과보고서, 2011.12)

•

하는 가운데 빅데이터 관련 산업 활성화의 최대 장애 요소로 보안 문제가

제기되고 있음

- 2013년 ‘10대 국가정보화 트렌드’에 따르면, 대량 데이터 활용에 따른 개인정보 유출 가능성을 우려하여 클라우드/빅데이터 활성화의 핵심 선결 과제로 보안 문제를 지적하고 있음 (출처 : 한국정보화진흥원, 2013년 10대 국가정보화 트렌드, 2013.2)
- 빅데이터/클라우드 환경에서도 신뢰할 수 있는 데이터 보호 기술의 개발은 결과적으로 해당 산업 시장의 확대 및 조기 정착을 견인할 수 있는 새로운 동력원이 될 것으로 판단됨

• 암호화된 데이터베이스 활용 기술의 고도화를 통해 정보의 기밀성과 활용성을 동시에 제공하여 그 동안 구축하기 힘들었던 공공데이터 활용 서비스 산업의 조속한 활성화 및 이에 따른 고용 촉진을 기대할 수 있음

- 국내에서는 재난 전조 감지, 구제역 예방, 맞춤형 복지 서비스 제공, 물가 관리, DNA, 개인맞춤형 의료 시스템 구축의 분야까지 활용 범위를 넓혀간다는 구상을 수립하고 있음 (출처 : 국가정보화전략위원회, 빅데이터를 활용한 스마트 정부 구현(안), 2011.10)
- 영국은 공공데이터 개방에 따라 약 150억 파운드의 경제적 효과와 2017년까지 58,000개의 신규 일자리가 창출될 것으로 예측 (출처 : 한국정보화진흥원, 창조경제 기반조성을 위한 공공데이터 개방과 활용 사례, 통권 제72호, 2014.3)

○ 사회문제해결 기대효과

• 국내외에서 발생하고 있는 개인정보 유출에 따른 사회적 피해 규모는 수치로 표현할 수 없을 정도로 심각하며, 신용 사회의 근간을 위협한다는 점에서 데이터 유출 방지를 위한 원천 기술 개발은 파급 효과가 매우 높음

• 해킹이나 내부 공모자에 의한 데이터 유출의 근본적인 해결책을 제공함으로써, 관련 사건 발생에 따르는 사회적 피해 복구 비용 절감

-

- 산업연구원의 분석에 따르면, 국내에서 개인정보 유출이나 해킹에 의한 피해액은 2015년 13조 4,000억원에서 2030년 26조, 7,000억원으로 피해 규모가 대폭 증가할 것으로 예상 (출처 : 산업연구원, e-KEIT 산업경제 정보, 제 586호, 2014.4)

• 지식정보보안 산업의 트렌드가 개인 및 사회 안전으로 빠르게 진화하면서 주요 사회 인프라 및 공공 서비스의 프라이버시 정보 노출 위협에 대한 국민적 불안감 해소

제 5 장 결론

최근 ICT 기술의 발전과 스마트 단말기의 확산으로 촉발된 스마트 혁명은 사회적, 경제적, 문화적 생활 전반에 걸쳐 매일 매일 다양한 새로운 경험을 우리에게 선사하고 있으며, 스마트 환경에서 사용자는 소형화되고 경량화된 스마트 기기들을 통해 보다 다양한 환경의 새로운 데이터 활용 서비스를 받을 수 있을 것으로 전망된다. 그러나, 이에 따른 역기능 또한 함께 커져가고 있어 사용자 데이터가 집중되는 데이터베이스에서의 빈번한 개인정보 유출은 대내외적으로 큰 우려를 낳고 있다.

이에 본 과제는 데이터베이스 암호화를 통한 데이터 기밀성을 제공하면서, 암호화된 데이터에 대한 활용을 극대화 할 수 있는 암호 원천 기술에 대한 연구를 최종 목표로 하고 있다. 특히, 암호화된 데이터베이스 활용을 위해 기본이 되는 데이터 저장 및 검색 기능을 제공할 수 있는 핵심 기술인 암호데이터 중복 처리 기술과 암호데이터 검색을 기술에 대한 연구를 수행하였다.

주요 연구 추진 실적으로 암호데이터 중복 처리 기술 분야에서는 메시지 기반 암호화 신규 프리미티브를 설계하고, 이를 기반으로 Client-side 암호데이터 중복 처리 프로토콜을 개발하였다. 본 기술은 파일 단위 또는 블록 단위 암호데이터 중복 처리 기능을 제공하여, 다용 사용자에 대한 중복 처리도 가능하다. 또한, 데이터 손실 공격에 안전성을 제공하면서 기존 기술 대비 우수한 중복 처리 성능을 제공한다. 암호데이터 검색 기술에 대한 연구에서는 Bloom Filter Tree, 다중 링크드 체인 등의 독창적인 검색 인덱스 구조를 제시하고, 이를 활용한 암호데이터 검색 프리미티브 알고리즘을 설계하였다. 또한 검색 인덱스 재구성 개념을 제시하고 안전성을 훼손하지 않으면서 동적으로 변화하는 암호데이터에 대한 검색 인덱스를 효과적으로 관리할 수 있는 기술을 설계하여, 최종적으로 동적으로 변화하는 암호데이터에 대한 자유로운 추가/삭제가 가능한 sublinear 검색 성능의 암호데이터 검색 기술을 설계하였다.

본 과제의 연구 결과들은 다양한 정보 통신 서비스의 기반 프리미티브 기술로 활용될 것으로 기대되며, 또한 다양한 신규 정보 통신 서비스의 신뢰성을 향상시켜 관련 산업 발전 및 활성화에 기여할 것으로 전망된다.

참고문헌

[ABC05] M. Abdalla, M. Bellare, D. Catalano, E. Kiltz, T. Kohno, T. Lange, J. Malone-Lee, G. Neven, P. Paillier, and H. Shi, “Searchable encryption revisited: consistency properties, relation to anonymous IBE, and extensions”, Proceedings of Crypto 2005, pp. 205-222, 2005.

[AKK09] G. Ateniese, S. Kamara, and J. Katz, “Proofs of storage from homomorphic identification protocols”, Proceedings of Asiacrypt 2009, LNCS 5912, pp. 319-333, 2009.

[AKS04] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, “Order-preserving encryption for numeric data”, Proceedings of SIGMOD 2004, pp. 563-574, 2004.

[ATE08] G. Ateniese, R. Di Pietro, L. Mancini, and G. Tsudik, “Scalable and efficient provable data possession”, Proceedings of SecureComm 2008, pp. 1-10, 2008.

[BCL09] A. Boldyreva, N. Chenette, Y. Lee, and A. O'Neill, “Order-preserving symmetric encryption”, Proceedings of Eurocrypt 2009, LNCS 5479, pp. 224-241, 2009.

[BCO04] D. Boneh, G. Crescenzo, R. Ostrovsky, and G. Persiano, “Public key encryption with keyword search”, Proceedings of Eurocrypt 2004, pp. 506-522, 2004.

[BCO11] A. Boldyreva, N. Chenette, and A. O'Neill, “Order-preserving encryption revisited: Improved security analysis and alternative solutions”, Proceedings of Crypto 2011, LNCS 6841, pp. 578-595, 2011.

[Beb02] G. Bebek, “Anti-tamper database research: inference control techniques”, Technical Report EECS 433 Final Report, Case Western Research University, 2002.

[BJO09] K. D. Bowers, A. Juels, and A. Oprea, “Proofs of retrievability: Theory and implementation”, Proceedings of CCSW 2009, pp. 43-54, 2009.

[BKR13a] M. Bellare, S. Keelveedhi, and T. Ristenpart, “Message-locked encryption and secure deduplication”, Proceedings of EUROCRYPT 2013, LNCS 7881, pp. 296-312, 2013.

[BKR13b] M. Bellare, S. Keelveedhi, and T. Ristenpart, “DupLESS: Server-aided encryption for deduplicated storage”, Proceedings of the 22nd USENIX conference on Security, pp. 179-194, 2013.

[BLA14] J. Blasco, R. Di Pietro, A. Orfila, and A. Sorniotti, “A tunable proof of ownership scheme for deduplication using bloom filters”, Proceedings of CNS 2014, pp. 481-489, 2014.

[BW07] D. Boneh, and B. Waters, “Conjunctive, subset, and range queries on encrypted data”, Proceedings of IEEE Trans. on Cloud Computing, vol. 4392, pp. 535-554, 2007.

[CCK13] J.H. Cheon, J.-S. Coron, J. Kim, M.S. Lee, T. Lepoint, M. Tibouchi, and A. Yun, “Batch fully homomorphic encryption over the integers”, Proceedings of Eurocrypt 2013, LNCS 7881, pp. 315-335, 2013.

[CGK06] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, “Searchable symmetric encryption: Improved definitions and efficient constructions”, Proceedings of ACM CCS 2006, 2006.

[CKL12] D. Choi, S.H. Kim, and Y. Lee, “Address permutation for privacy-preserving searchable symmetric encryption”, ETRI Journal, Vol. 34, no. 1, pp. 66-75, 2012.

[CM05] Y.C. Chang, and M. Mitzenmacher, “Privacy preserving keyword searches on remote encrypted data”, Proceedings of Applied Cryptography and Network Security Conference, 2005.

[DOU02] J. R. Douceur, A. Adya, W.J. Bolosky, D. Simon, and M. Theimer,

"Reclaiming space from duplicate files in a serverless distributed file system", Proceedings of International Conference of Distributed Computing Systems 2002, pp. 617-624, 2002.

[DVJ03] E. Damiani, S.D.C. di Vimercati, S. Jajodia, S. Paraboschi, and P. Samarati, "Balancing confidentiality and efficiency in untrusted relational DBMSs", Proceedings of 10th ACM Conference on Computer and Communication Security 2003, pp. 93-102, 2003.

[DVW09] Y. Dodis, S. Vadhan, and D. Wichs, "Proofs of retrievability via hardness amplification", Proceedings of TCC 2009, LNCS 5444, pp. 109-127, 2009.

[Gen09] C. Gentry, "Fully homomorphic encryption using ideal lattices", Proceedings of 41st Annual ACM Symposium on Theory of Computing 2009, pp. 169-178, 2009.

[Goh03] E.J. Goh, "Secure indexes", IACR ePrint Archive, Technical report 2003/216.

[GS03] S.C. Gultekin and D. Singer, "Anti-tamper database: Querying encrypted databases", Proceedings of 17th Annual IFIP WG 11.3 Working Conferences on Database and Applications Security, 2003.

[HAL11] S. Halevi, D. Harnik, B. Pinkas, A. Shulman-Peleg, "Proofs of ownership in remote storage systems", Proceedings of 18th ACM. Conference on Computer and Communications Security, pp. 491-500, 2011.

[HIL02] H. Hacigümüş, B.R. Iyer, C. Li, and S. Mehrotra, "Executing SQL over encrypted data in the database-service-provider model," Proceedings of ACM SIGMOD International Conference on Management of Data, pp. 216-227, 2002.

[HPS10] D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Side channels in cloud services : Deduplication in cloud storage", IEEE Security and Privacy Magazine, vol. 8, pp. 40-47, 2010.

[HUS14] M. I. Husain, S. Y. Ko, S. Uurtamo, A. Rudra, and R. Sridhar, “Bidirectional data verification for cloud storage”, Journal of Network and Computer Applications, Vol. 45, pp. 96-107, 2014.

[JK07] A. Juels and B. Kaliski, “PORs: Proofs of retrievability for large files”, Proceedings of CCS 2007, pp. 584-597, 2007.

[LI14a] J. Li, X. Chen, M. Li, J. Li, P. PC Lee, and W. Lou, “Secure deduplication with efficient and reliable convergent key management”, IEEE Transactions on Parallel and Distributed Systems, Vol. 25, no. 6, pp. 1615-1625, 2014.

[LI14b] J. Li, X. Chen, F. Xhafa, and L. Barolli, “Secure deduplication storage systems with keyword search”, Proceedings of AINA 2014, pp. 971-977, 2014.

[LI15] J. Li, Y. K. Li, X. Chen, P. PC Lee, and W. Lou, “A hybrid cloud approach for secure authorized deduplication”, IEEE Transactions on Parallel and Distributed Systems, Vol. 26, no. 5, pp. 1206-1216, 2015.

[Lu12] Y. Lu, “Privacy-preserving logarithmic-time search on encrypted data in cloud”, Proceedings of NDSS 2012, 2012.

[LL] J. Li, J. Li, D. Xie, and Z. Cai. Secure auditing and deduplicating data in cloud. IEEE Transactions on Computers, 65(8):2386−2396, Aug. 2016.

[MC11] L. Marques and C. Costa, “Secure deduplication on mobile devices”, Proceedings of the 2011 Workshop on Open Source and Design of Communication, pp. 19-26, 2011.

[PS12] R. Di Pietro and A. Sorniotti, “Boosting efficiency and security in proof of ownership for deduplication”, Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security, pp. 81-82, 2012.

[RAS13] S. Rass, “Dynamic proofs of retrievability from chameleon-hashes”, Proceedings of SECRYPT 2013, pp. 1-9, 2013.

[SBC07] E. Shi, J. Bethencourt, T.-H.H. Chan, D. Song, and A. Perrig, “Multi-dimensional range query over encrypted data”, Proceedings of IEEE Symposium on Security and Privacy 2007, pp. 350-364, 2007.

[SK14] Y. Shin, and K. Kim, “Efficient and secure file deduplication in cloud storage”, IEICE Transactions on Inf. & Syst., Vol. E97-D, NO. 2, pp. 184-197, 2014.

[STO08] M. Storer, K. Greenan, D. Long, and E. Miller, “Secure data deduplication” Proceedings of the 4th ACM International Workshop on Storage Security and Survivability, pp. 1-10, ACM, 2008.

[SWP00] D. Song, D. Wagner, and A. Perrig, “Practical techniques for searching on encrypted data”, Proceedings of IEEE symposium on security and privacy 2000, pp. 44-55, 2000.

[XCZ13] J. Xu, E. C. Chang, and J. Zhou, “Weak leakage-resilient client-side deduplication of encrypted data in cloud storage”, Proceedings of ASIA-CCS 2013, pp. 195-206, 2013.

[XZ14] J. Xu and J. Zhou, “Leakage resilient proofs of ownership in cloud storage, revisited”, Proceedings of ACNS 2014, pp. 97-115, 2014.

[YCC15] C.-M. Yu, C.-Y. Chen, and H.-C. Chao, “Proof of ownership in deduplicated cloud storage with mobile device efficiency”, IEEE Network, Vol. 29, no. 2, pp. 51-55, 2015.

[YY] J. Yuan and S. Yu. Secure and constant cost public cloud storage auditing with deduplication. In Communications and Network Security (CNS), 2013 IEEE Conference on, National Harbor, MD, USA, pages 145–153. IEEE, Oct. 2013

약어표

AES	Advanced Encryption Standard
BF	Bloom Filter
CE	Convergent Encryption
CSA	Cloud Security Alliance
DB	Data Base
FHE	Fully Homomorphic Encryption
HCE1	Hash and CE without tag check
HCE2	Hash and CE with tag check
HE	Homomorphic Encryption
IoT	Internet of Things
MLE	Message Locked Encryption
OPE	Order Preserving Encryption
PoW	Proof of Ownership
RCE	Randomized Convergent Encryption
SMC	Secure Multiparty Computation
SSE	Symmetric Searchable Encryption

주 의

1. 이 연구보고서는 한국전자통신연구원의 주요사업으로 수행한 최종연구 결과입니다.

2. 이 보고서의 내용을 발표할 때에는 반드시 한국전자통신연구원에서 수행한 주요사업 결과임을 밝혀야 합니다.