SGX를 활용한 클라우드 환경에서의 프라이버시 보존 데이터 검색 효율성에 대한 고찰

구동영*, 허준범**
*한성대학교 기계전자공학부
**고려대학교 정보대학 컴퓨터학과
dykoo@hansung.ac.kr, jbhur@korea.ac.kr

A Study on Efficiency of Privacy-preserving Search in Cloud Storage using SGX

Dongyoung Koo*, Junbeom Hur**
*School of Mechanical and Electronics Engineering, Hansung University
**Dept. of Computer Science and Engineering, Korea University

요 약

네트워크에 존재하는 저장 공간을 필요에 따라 유연하게 대여하여 사용할 수 있는 클라우드 스토리지 서비스는 데이터의 일관성 유지, 저렴한 유지관리 비용 등 여러 장점에 힘입어 널리 활용되고 있다. 하지만 클라우드 시스템은 데이터 소유자에 의한 관리가 이루어지지 않으므로 민감한 데이터의 노출에 의한 피해 또한 다수 발생하고 있는데, 이를 해결하기 위하여 암호화 등을 통한 프라이버시 보존을 위한 연구가 꾸준히 진행되고 있다. 본 연구에서는 프라이버시가 보존된 상태에서 클라우드에 저장된 데이터를 검색함에 있어, 대수적 난제에 근거를 둔 접근 제어 기능을 내포한 소프트웨어 기반의 검색 가능한 암호화 (searchable encryption) 기법과 최근 많은 관심을 받고 있는 하드웨어 기반 클라우드 데이터 검색의 효율성 및 기능에 대한 비교 분석을 수행한다. 이를 통하여 하드웨어 기반 기법의 활용을 통한 성능 향상 가능성을 확인하고 잠재적 보안 위협을 검토한다.

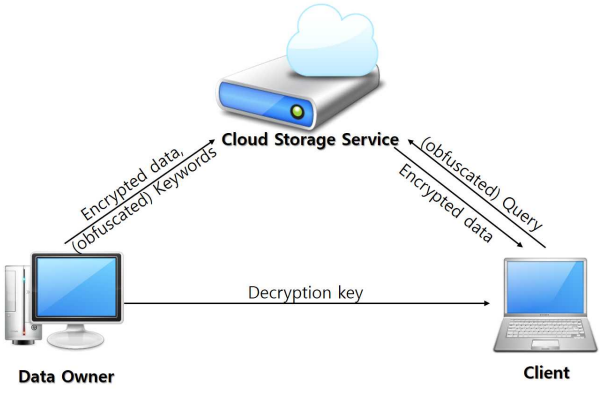
1. 서론

클라우드 서비스의 등장과 함께 개인용 컴퓨터를 통해서만 데이터가 저장되고 처리되던 환경은 네트워크 기능을 가진 단말만 있으면 언제, 어디에서나 인터넷 접속을 통하여 업무 및 오락을 비롯한 모든 컴퓨팅 기능을 수행할 수 있는 환경으로 변하였다. 특히, 클라우드 스토리지 서비스는 다양한 기기에 분산되어 있던 데이터를 저렴한 유지·관리 비용으로 일관성 있게 관리가능하게 하였다.

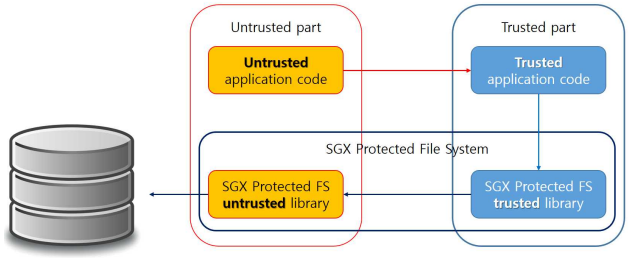
하지만 클라우드 스토리지 서비스의 이용은 데이터 소유자가 관리하던 데이터의 관리 권한이 제3의 클라우드 서비스 제공자에게 위임되는 것을 의미하며, 클라우드 스토리지에 저장된 민감한 데이터의 의도치 않은 노출 등 프라이버시 침해에 대응하기 위한 암호화 기법 등 다양한 연구가 수행되고 있다.

특히, 클라우드와 같은 원격저장소에서 관리되는 데이터의 기밀성 보존을 위하여, 저장되는 데이터 자체의 암호화와 더불어 수많은 암호문 사이에서 적법한 사용자가 암호문을 검색하고 활용하는 과정에서 발생 가능한 검색 키워드 프라이버시 보존의 중요성에 대한 연구가 활발히 이루어지고 있다.

본 연구에서는 클라우드 환경에 저장된 암호화된 데이터를 프라이버시가 보존된 상태에서 효율적으로 검색하기 위하여 제시된 대표적인 검색 가능한 암호화 기법 (searchable encryption)과 최근 활발한 연구 및 개발이 이루어지고 있는 신뢰 실행 환경 (trusted execution environment, TEE)에서의 검색 성능을 실험을 통하여 비교 분석하고, 수행 가능한 기능의 특징 및 제한사항에 대하여 살펴본다.



(그림 1) 프라이버시 보존 데이터 검색 환경



(그림 2) SGX 보안 모듈에서의 안전한 데이터 관리

2. 프라이버시 보존 암호문 검색 기법

암호화는 원문의 내용을 알 수 없도록 난독화를 수행하므로 프라이버시 보존을 위하여 암호화되어 저장된 다수의 클라우드 데이터로부터 특정 키워드에 대응되는 데이터의 검색은 사실상 불가능하다. 암호문과 관련 키워드를 함께 저장하는 방법을 고려할 수 있으나, 평문 형태로 존재하는 키워드로부터 암호문에 대한 정보 유추가 가능하므로 그림 1과 같이 검색 키워드 자체에 대한 프라이버시 보존 또한 고려할 필요가 있다.

평문에 대한 접근 권한이 없는 사용자로 하여금 대응되는 암호문에 대한 검색 기능을 제공하기 위하여 Boneh et al.은 암호문에 키워드를 이용하여 검색 가능한 부가정보를 함께 저장하는 기법 (public key encryption with keyword search, PEKS)을 제시하였다 [1]. 공개키 암호화 기법을 활용한 PEKS는 트랩도어 함수를 이용하여 특정 암호문이 검색하고자 하는 키워드를 포함하고 있는지 여부만을 확인할 수 있도록 함으로써, 검색 키워드에 대한 프라이버시를 보존하는 대표적인 기법이다.

최근에는 보안 영역을 물리적으로 분리하여 하드웨어 상에서 민감 정보를 관리할 수 있는 신뢰 실행 환경 (trusted execution environment, TEE)에 대한 연구 및 개발이 활발하게 이루어지고 있다. ARM TrustZone 및 Intel SGX로 대표되는 이러한 신뢰 실행 환경을 활용하는 경우, 그림 2와 같이 클라우드에 존재하는 데이터의 저장 및 관리, 검색 작업을 보안 영역 안에서 수행함으로써 일반 영역의 공격에 의한 정보 유출을 차단할 수 있다. 본 연구에서 사용한 Intel SGX [2]에서는 인클레이브 (enclave)라고 불리는 보안 영역에 클라우드 스토리지에 저장된 데이터 파일명과 연관 키워드 데이터베이스를 그림 3과 같이 간단한 테이블로 구성하여 저장하고 인클레이브 내에서만 검색을 가능하도록 함으로써 키워드 검색의 프라이버시를 보장하도록 한다.

Filename	Keywords
File000001.data	Korean
File000002.data	Institute, Information
...	...

(그림 3) SGX에서의 파일명-키워드 인덱스 테이블

3. 암호문 검색 성능 및 기능 비교 분석

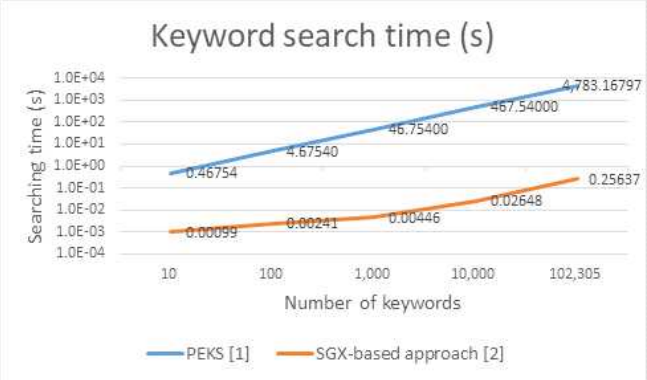
본 절에서는 클라우드에 존재하는 암호문 검색을 위하여 Boneh et al.의 PEKS [1]와 Intel SGX [2]를 활용한 신뢰 실행 환경에서의 성능을 실험을 통하여 확인한다. 키워드를 이용한 암호문 검색은 클라우드 스토리지에 저장된 모든 데이터에 대하여 전수조사가 이루어져야 하므로 키워드 검색에 소요되는 시간을 반복 측정한 후 평균을 구하여 비교하도록 한다.

두 실험군의 성능 비교를 위하여 SGX를 지원하는 Intel NUC Kits (NUC7CJYH, Intel Celeron J4005 CPU, 4GB RAM)에 Ubuntu 18.04 Desktop (64bit) 운영체제를 설치하여 실험을 수행하였다. 키워드 집합은 '/usr/share/dict' 디렉터리에 존재하는 words 파일을 활용하였고, Ubuntu 18.04에서는 총 102,305개의 단어를 포함하고 있다. 실험에서는 하나의 파일이 하나의 키워드로만 검색할 수 있다는 가정에서 저장된 파일의 개수를 증가시키면서 수행하였다.

PEKS는 Github에 공개된 키워드 비교 검색 라이브러리[1]를 활용하였고, SGX 기반 신뢰 실행 환경에서의 실험은 인클레이브 내에서 생성되고 저장된 키워드 데이터베이스 파일로부터 키워드를 검색하도록 구현하였다.

클라우드에 저장된 파일 개수에 따른 검색 소요 시간은 그림 4와 같다. 하나의 키워드 검색을 위한 연산 소요 시간은 PEKS [1]에서 평균 0.46754초이고, SGX 구현 [2]에서는 평균 0.00100초가 소요되었다. 이는 PEKS가 큰 수에 대한 대수적 연산을 수행하는 공개키 기반 페어링 연산 (pairing operation)을 수행하는 반면, SGX에서는 분리된 인클레이브 영역에서의 암호화된 인덱스 테이블 파일을 읽어들인 후 복호화하는 시간과 SGX 라이브러리 로딩 시간을 제외하면 평문에서의 연산과 동일한 과정이 메모리에서 수행되므로 상대적으로 매우 빠른 연산이 가능하기 때문이다. 또한, 파일 (및 검색 키워드)의 개수 증가에 비례하여 키워드 검색에 소요되는 시간이 증가하므로, 약 10만개의 파일이 저장된 시스템에서

1) https://github.com/atulmahind/PEKS



(그림 4) 기법별 키워드 검색 소요 시간

실험에 사용한 미니PC로는 하나의 파일을 검색하는 데 평균 39.8분이 소요되어 검색 효율성이 현저하게 낮아짐을 알 수 있다.

검색 키워드 수에 따른 성능 비교와 더불어, 각 기법에서 지원 가능한 검색의 추가 기능에 대한 비교는 표 1과 같다. PEKS [1]에서는 단순 키워드의 일치 여부에 기반한 검색만을 지원하기 때문에, 다수의 키워드와 연관된 (암호화된) 데이터에 대해서는 독립적인 키워드 비교 연산이 이루어져야 하고 키워드 A, B, C에 대하여 '(A or B) and C'와 같은 여러 키워드 조합에 대한 검색이 불가능하다. 또한, 부분문자열 일치 및 유사도 검색, 범위 검색 등이 지원되지 않는다. 검색 표현의 유연성 증가를 위한 다수의 연구가 진행되어 왔으나, 공개키 기반의 큰 수에 대한 대수적 연산을 수행하기 때문에 상대적으로 높은 연산 복잡도를 가진다. 이에 반하여, SGX를 이용한 신뢰 실행 환경에서의 키워드 검색은 키워드를 포함하고 있는 암호화된 인덱스 테이블을 메모리에서 복호화하여 활용하므로, 평문에 대한 키워드 연산에 대응되는 유연한 연산이 대부분 가능할 것으로 예측할 수 있다.

4. 결론

클라우드 환경과 같이 원격저장소에서 관리되는 데이터의 기밀성 보장을 위하여 암호문이 저장되는 환경에서, 검색 키워드에 대한 프라이버시를 보장하면서도 검색 가능한 대표적인 소프트웨어 기반 및 하드웨어 기반 보안 기법을 PEKS [1]와 SGX [2] 구현을 통하여 살펴보았다.

실험을 통하여 살펴본 바와 같이 하드웨어 기반 신뢰 실행 환경을 활용함으로써 검색 기능의 유연한 확장과 성능 효율을 꾀할 수 있을 것으로 보이지만, 신뢰 실행 환경에는 Intel과 같은 하드웨어 제조사에 대한 전적인 신뢰를 요구하고 있으므로, 제3의 기관인 제조사에 의한 위협을 견제할 수 있는 방법에 대한 고찰 또한 필요하다. 신뢰 실행 환경에 대한 공격 기법 또한 새롭게 발견되고 보완되고 있는 상황을 고려할 때, 보안 기능과 더불어 효율성 제고 방안으로 사용 가능함을 확인하였으며, 부채널을 비롯한 발생 가능한 위협에 대한 향후 연구 수행 또한 이루어질 필요가 있다.

기법	비교 기능
PEKS [1]	동일성 (equality)
SGX-based [2]	동일성, 크기, 키워드 결합 포함관계 (일부 일치) 등

(표 1) PEKS 및 SGX 기반 키워드 검색 기능 비교

사사 (Acknowledgement)

본 연구는 고려대 암호기술 특화연구센터 (UD170109ED)를 통한 방위사업청과 국방과학연구소의 연구비 지원으로 수행되었습니다.

참고문헌

[1] Dan Boneh, Giovanni Di Crescenzo, Rafail Ostrovsky and Giuseppe Persiano, "Public Key Encryption with Keyword Search," Advances in Cryptology - EUROCRYPT, pp. 506-522, 2004.

[2] Victor Costan and Srinivas Devadas, "UIntel SGX Explained," IACR Cryptology, ePrint Archive, pp.86:1-118, 2016.