클라우드 스토리지 환경에서의 사용자 중심 암호데이터 중복제거 기술에 관한 연구

김원빈, 이임영*

순천향대학교 소프트웨어융합학과

wbkim29@sch.ac.kr, *imylee@sch.ac.kr

A Study on the User Centric Secure Data Deduplication in Cloud Storage Environment

Won-bin Kim, Im-Yeong Lee*

Department of Software Convergence, Soonchunhyang University

요 약

본 연구는 클라우드 스토리지 환경에서 사용자 중심의 암호화된 데이터의 중복제거를 수행하는 연구를 수행한다. 기존의 암호데이터 중복제거 연구에서는 서버를 중심으로 중복제거가 수행되었으며, 이를 통해 실질적으로 서버만이 이득을 취하는 형태를 보여주었다. 이러한 환경에서 사용자는 클라우드 스토리지에 종속된 서비스만을 제공받을 수 있다. 따라서 이러한 문제를 해결하기 위해 본 연구에서는 사용자가 직접 클라우드 스토리지에 저장된 데이터의 중복제거를 직접 제어하고 데이터를 직접 관리할 수 있는 사용자 중심의 암호데이터 중복제거 기술을 제안한다.

Ⅰ. 서 론

일반적으로 클라우드 스토리지는 Honest-but-curious라는 특성을 가지고 있다. 이는 사용자의 요청에 대해서는 정직한 응답을 제공하지만 호기심이 많아 사용자의 데이터를 보려고 하는 특성을 의미한다. 결국 클라우드 스토리지에 위탁한 데이터는 언제든지 유출될 수 있는 가능성을 내재한다는 것과 같다. 따라서 사용자는 클라우드 스토리지에 업로드하는 데이터를 암호화하여 데이터의 유출 시에도 내용이 노출되지 않도록 안전하게 보관하여야 한다.

데이터를 안전하면서도 효율적으로 보관하기 위해 암호데이터 중복제거가 제안되었다. 데이터 중복제거 기술은 스토리지상에 중복된 데이터의 저장을 제어하여 스토리지 공간 효율성을 높이는 기술이다. 데이터 중복제거 기술은 데이터의 업로드 과정에서 통신 환경에 따라 통신 능력과 연산능력 중 각 능력을 중점적으로 활용할 수 있는 형태가 제시되었다. 하지만 이러한 모든 형태는 결국 사용자가 저장한 데이터를 서버(클라우드 스토리지)가 중복제거하여 서버의 저장공간 효율성을 높이는 형태로 제공되고 있다. 결국 서버 중심(Server centric)의 데이터 중복제거가 이루어지고 있는 것으로 볼 수 있으며, 이는 데이터 저장 및 보관의 제어를 사용자가 직접 수행할 수 없다는 것을 의미한다. 본 연구에서는 이러한 형태를 개선하여 사용자가 직접 중복제거를 제어하고, 중복제거의 이점을 사용자가 이용할 수 있는 사용자 중심(User centric) 중복제거 방법을 제안한다.

Ⅱ. 관련연구

Ⅱ-1. 데이터 중복제거

데이터 중복제거는 데이터를 비교하는 방식을 이용하는 기술이다. 반면 데이터 암호화는 데이터의 내용을 알 수 없도록 바꾸는 기술이다. 따라서 일반적인 방법으로는 데이터 중복제거와 암호화를 동시에 적용할 수 없다. 이를 해결하기 위해 Convergent Encrypotion(CE)가 제안되었으며 [1], CE를 활용하여 다양한 목적에 맞게 활용할 수 있는 Message-Locked Encryption(MLE)이 제안되었다 [2]. 이후, 소유권 증명, 사전공격 등 다양한 위협에 대응하기 위한 여러 기술이 제안되었다 [3].

서버 측 중복제거는 사용자가 업로드한 데이터를 서버 내에서 중복제거하는 방법이다. 따라서 서버의 높은 처리 성능과 자원을 기반으로 빠르고 효율적인 중복제거가 가능하다.

클라이언트 측 중복제거는 사용자의 디바이스에서 중복제거를 수행한 후 중복되지 않은 데이터만을 업로드하는 방법이다. 따라서 사용자가 전송해야 하는 데이터량이 감소하는 장점을 갖는다. 하지만 중복제거되어 전송하지 않는 데이터를 사용자가 실제로 소유하는지를 확인하기 위한 소유권 증명 과정이 추가로 수행되어야 하는 등 사용자의 디바이스에서 추가적인 연산을 수행해야 한다.

Ⅱ-2. 사용자 중심 암호데이터 중복제거

본 연구에서는 사용자 중심의 암호데이터 중복제거를 제안한다. 기존의 서버 중심의 암호데이터 중복제거 형태에서는 사용자가 실질적으로 얻을 수 있는 중복제거의 효과가 없다. 또한 사용자가 중복제거 여부를 제어할 수 없다. 따라서 이러한 문제를 해결하기 위해 본 연구에서는 사용자 중심의 중복제거 방식을 제안한다. 이 방식은 데이터의 중복 여부 판단과 데이터의 관리 및 소유권 판단 등을 서버가 아닌 사용자가 수행할 수 있도록 하는 방식이다. 따라서 서버는 사용자의 데이터를 저장하는 원격 저장소의 역할만을 수행하고, 데이터의 관리 및 제어는 사용자가 직접 수행할 수 있도록 한다. 이에 대한 상세한 설명과 형태는 아래에서 설명한다.

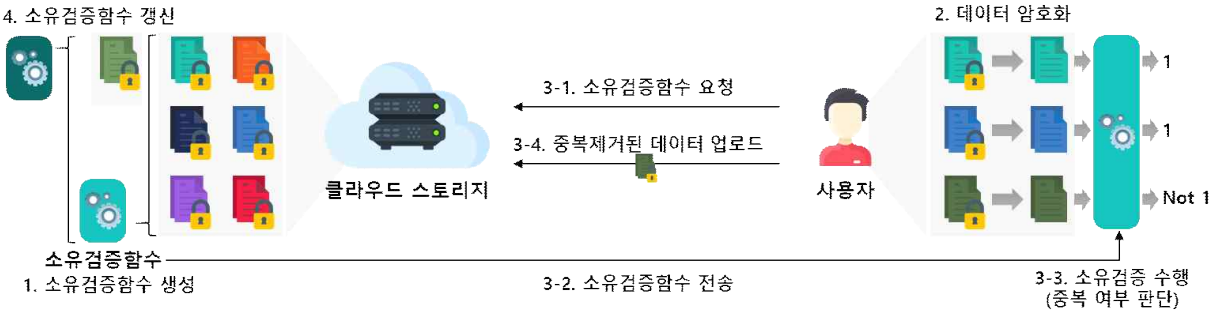

그림 1. 제안방식의 시나리오

Ⅲ. 제안방식

Ⅲ-1. 보안 요구사항

본 절에서는 본 연구의 보안 요구사항을 제시한다.

• 기밀성(Confidentiality) : 클라우드 스토리지에 보관된 데이터는 각 데이터에 대응되는 암호화키를 보유하지 않으면 데이터 내용을 알 수 없어야 한다.

• 소유검증(ownership verification) : 사용자는 클라우드 스토리지가 보관중인 데이터의 목록을 검증할 수 있어야 한다. 이 과정에서 사용자는 검증을 수행할 데이터를 클라우드 스토리지에게 알리지 않고 수행할 수 있어야 한다.

Ⅲ-2. 알고리즘

Ⅲ-2-1. 소유검증함수 생성 단계

Step1. 클라우드 스토리지는 보유한 모든 암호화된 데이터 조각 c_i를 해시화한 후 소유검증함수$f(x)$를 생성한다.

Ⅲ-2-2. 데이터 암호화 단계

Step1. 사용자는 업로드할 데이터를 특정한 크기의 조각 b_i으로 나눈다.

Step2. 사용자는 각 데이터 조각을 해시화하여 암호화 키 k를 생성한다.

Step3. 사용자는 생성한 암호화 키 k와 각 데이터 조각 b_i를 대칭키 암호화 방식으로 암호화하여 암호데이터 c_i를 획득한다.

Ⅲ-2-3. 중복제거 및 업로드 단계

Step1. 사용자는 클라우드 스토리지에게 소유검증함수를 요청하고, 클라우드 스토리지는 최신화된 소유검증함수 $f(x)$를 사용자에게 전달한다.

Step2. 사용자는 소유검증함수 $f(x)$에 업로드할 암호데이터의 해시값을 입력하여 해당 데이터의 중복 여부를 판단한다.

Step3. 사용자는 Step 2의 결과에 따라 중복되지 않은 데이터c_j만을 클라우드 스토리지에 전송한다.

Ⅲ-2-3. 소유검증함수 갱신 단계

Step1. 클라우드 스토리지는 사용자가 새로 업로드한 암호데이터 c_j를 이용하여 기존의 소유검증함수를 갱신한다.

Ⅳ. 제안방식 분석

• 기밀성(Confidentiality) : 본 제안방식에서 사용자는 CE를 이용하여 암호화를 수행한다. 따라서 동일한 데이터를 소유하지 않은 다른 사용자는 해당 데이터를 복호화 할 수 없다.

• 소유검증(ownership verification) : 클라우드 스토리지는 라그랑주 보간법을 응용한 방법을 통해 소유 검증 함수를 생성한다. 이를 통해 사용자는 소유검증함수에 업로드할 데이터를 입력함으로써 클라우드 스토리지에 중복된 데이터가 존재하는지를 확인할 수 있다.

Ⅴ. 결론

본 연구에서는 사용자 중심의 암호데이터 중복제거 기술을 제안하였다. 기존의 연구에서는 서버를 중심으로 한 데이터 중복제거 방식을 제공하였다. 이는 사용자가 중복제거를 직접 제어할 수 없는 방식이기 때문에 서버가 제공하는 서비스에 종속되는 결과를 낳았다. 본 연구에서는 이러한 문제를 해결하기 위해 사용자가 직접 데이터 중복제거를 제어하고 데이터를 관리할 수 있도록 하기 위해 소유검증함수라는 개념을 도입하였다. 이는 클라우드 스토리지가 보관 중인 데이터의 목록 중 사용자가 업로드하려는 데이터가 존재하는지를 확인할 수 있는 함수이다. 또한 이 과정에서 사용자는 자신이 업로드할 데이터를 클라우드에게 제공하지 않기 때문에 통신 효율성 및 보안성에서 장점을 갖는다. 이를 통해 본 연구는 사용자 중심으로 더욱 높은 효율성과 보안성을 갖는 암호데이터 중복제거 환경을 제공할 수 있다.

ACKNOWLEDGMENT

이 논문은 2019년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. NRF-2019R1A2C1085718)

참 고 문 헌

[1] Douceur, John R., et al. "Reclaiming space from duplicate files in a serverless distributed file system." Proceedings 22nd international conference on distributed computing systems. IEEE, pp. 617-624, 2002.

[2] Bellare, Mihir, Sriram Keelveedhi, and Thomas Ristenpart. "Message-locked encryption and secure deduplication." Annual international conference on the theory and applications of cryptographic techniques. Springer, Berlin, Heidelberg, pp. 296-312, 2013.

[3] Keelveedhi, Sriram, Mihir Bellare, and Thomas Ristenpart. "Dupless: Server-aided encryption for deduplicated storage." 22nd {USENIX} Security Symposium ({USENIX} Security 13). pp. 179-194, 2013.