랭킹기반의 검색가능한 암호화 알고리즘 분석

민정기*, 허준범**
고려대학교 컴퓨터학과
eternalray@korea.ac.kr*
jbhur@korea.ac.kr**

Analysis of ranked searchable encryption schemes

Junggi Min, Junbeom Hur
Dept. of Computer Science and Engineering, Korea University

요 약

Ranked searchable encryption 은 모바일 환경 등의 자원 제약적인 환경에 적용 가능한 랭킹기반의 검색가능한 암호화 알고리즘이다. 본 연구에서는 기존 암호화 기법에 대한 안전성과 효율성에 대한 분석을 하고, 이를 바탕으로 각 기법의 장단점을 분석한 후 향후 연구에 대한 고찰을 한다.

1. 서론

많은 기업들이 클라우드 서비스의 이점에도 불구하고 클라우드 서비스를 사용하려 하지 않는 이유는 보안 문제이다. 이를 해결하기 위해서 데이터 소유자는 자신의 정보를 클라우드 서버에 평문이 아니라 암호문의 형태로 저장하게 된다. 그 결과 데이터 사용자가 암호문으로 저장된 데이터에 손쉽게 접근할 수 있도록 하는 Searchable Encryption(SE)가 필요해졌다.

본문에서는 데이터 사용자의 모바일 환경을 가정하여, 데이터 사용자가 제한된 대역폭만을 사용할 수 있다고 할 때 원하는 결과에 해당하는 모든 데이터를 반환하는 것이 아니라, 가장 관련성이 높은 결과를 k 개만큼 반환하는 ranked searchable encryption scheme 들을 조사하고 기능, 안전성, 효율성의 관점에서 비교분석 하여 각 기법들의 특징과 현재 연구의 진행정도를 알아보고자 한다.

2. Ranked Searchable Encryption Schemes

기법들을 자세히 알아보기 전에, 각 기법들에서 공통으로 가정하고 있는 system model 은 세개의 entity 를 가정하고 있다. 이는 데이터 소유자, 데이터 사용자, 클라우드 서버이다.

데이터 소유자는 문서 집합 F = $\{f_1, f_2, \ldots, f_n\}$,과 검색에 사용될 인덱스 I를 F와 키워드의 집합 W = $\{w_1, w_2, \ldots, w_m\}$를 이용해 만들어 암호화해서 클라우드 서버에 올리게 된다. 그리고 trapdoor 생성에 필요한 키 SK = $\{S, M_1, M_2\}$ 와 기타 정보들을 권한이 있는 데이터 사용자에게 분배해 준다.

데이터 사용자는 trapdoor TD를 통해 k 개의 암호화된 문서를 클라우드 서버로부터 반환 받아 자신이 갖고 있는 키로 복호화할 수 있다.

클라우드 서버는 데이터 사용자로부터 받은 암호화된 문서 집합과 인덱스를 저장하고 있다가, 데이터 사용자로부터 TD 를 받으면 검색 알고리즘을 통해 점수가 높은 순서대로 k 개의 문서를 반환한다. 여기서 클라우드 서버는 honest-but-curious 한 entity 이다. 이는 클라우드 서버가 정확하게 프로토콜을 수행하지만, 추가적인 정보를 얻기 위해 프로토콜 수행 과정에서 다뤄지는 데이터들을 분석하려 하는 것을 의미한다.

2.1 Multi-Keyword Ranked Searchable Encryption

Cao et al.[1]은 최초로 Multi-Keyword Ranked Searchable Encryption scheme 을 제시한 논문이다. 매우 기초적인 기법으로서 공통으로 가정하고 있는 과정과 거의 흡사하게 동작한다. 기본적으로 데이터와 인덱스를 암호화하기 때문에 data privacy 를 제공하며, SK 를 생성할 때 임의의 값이 선택되어 M_1, M_2 행렬의 원소로 들어가기 때문에 trapdoor unlinkability 와 keyword privacy 가 보장된다.

2.2 Efficient Multi-Keyword Ranked Search

Li et al.[2]은 blind storage 를 사용해서 다른 보안요소 뿐만 아니라, access pattern 을 숨긴 것이 특징이다. 보통 일반적인 SE 기법의 경우, access pattern 을 숨기지 않는데, access pattern 을 숨기기 위해서는 많은 비용을 지불해야 하기 때문이다. 이 기법은 데이터 사용자가 암호화된 문서 집합과 인덱스를 블록 단위로 분할해 blind storage 에 저장하고, Attribute-based Encryption 을 통해 특정 블록에 대해 권한을 갖고 있는 데이터 사용자만 TD 를 생성하여 blind storage 에 접근할 수 있다. 이 때 클라우드 서버와 데이터 사용자간에 통신은, 먼저 데이터 사용자가 TD 를 보내고 (1) 클라우드 서버는 그에 맞는 블록 descriptor 를 k

개만큼 보내면(2) 데이터 사용자는 descriptor 를 받아서 복호화를 하고 이에 해당하는 블록들을 blind storage 로부터 반환 받아서 복호화 한다(3). 따라서 통신 비용이 굉장히 높다고 할 수 있다. 또한 blind storage 에서의 추가적인 연산에 대한 비용을 고려해야 하기 때문에 성능이 다른 기법보다 비교적 안 좋을 수 있다. 검색 성능은 O(mn)으로, 문서의 수와 키워드의 수에 linear 하다.

2.3 Authorized and Ranked Multi-Keyword Search

Li et al.[3]은 Certificate Authority 와 Third-Party Auditor 라는 추가적인 entity 들을 통해 데이터 사용자를 인증한다는 점이 특징이다. 처음 CA 를 통해 각 사용자에게 attribute 키 AK 가 발급되고, 사용자는 이를 이용해 토큰 TK를 만들어서 검색을 한다. 클라우드 서버는 받은 TK를 TPA 에 보내 사용자가 access policy 를 만족하는지 확인한 후에 데이터 사용자에게 또 다른 키 k 를 전달해 문서를 복호화 할 수 있게 한다. Attribute 키를 만드는 과정에서 사용자마다 다른 임의의 값이 들어가기 때문에 collusion resistance 를 갖게 된다. 검색 성능은 O(mn)으로, 문서의 수와 키워드의 수에 linear 하다.

2.4 Enhanced Dynamic Multi-Keyword Ranked Search

Xia et al.[4]는 Tree-based search 를 통해 검색 시간을 단축했다는 점과 데이터의 dynamic operation 을 지원한다는 점이 특징이다. Leaf node 에 문서의 각 키워드마다 TF 값을 저장하고, 이를 바탕으로 balanced binary tree 를 구성하여 인덱스 I를 만든다. 그리고 검색시에 Greedy Depth First Search 알고리즘을 통해 관련 없는 대부분의 node 를 탐색하지 않도록 하고 병렬 연산을 가능하게 해 효율성을 높였다. 또한 데이터의 변동이 있어 클라우드 서버에 저장된 암호화된 문서 집합과 I를 변경해야 할 경우에 사용할 수 있는 insert, delete, update operation 을 지원한다. 검색 성능은 $O(\theta m \log n)$인데, 여기서 θ는 쿼리에 포함된 키워드를 하나 이상 포함하고 있는 leaf node 의 개수로 $\theta \ll n$이다. 여기에 processor 의 개수를 ω라 하면, 시간복잡도는 $O(\frac{\theta m \log n}{\omega})$이다.

3. 기존 연구에 대한 고찰

앞서 분석한 내용을 표로 정리하면 표 1 과 같다. [1],[2],[3]의 경우 검색 성능이 문서의 수와 키워드의 수에 linear 하게 증가하기 때문에, 실제 일반적인 클라우드 환경에 적용하기에는 현실적으로 무리가 있다. 따라서 [4]와 같이 검색 속도를 최적화 하는 것이 매우 중요하다. 그러나 [4]의 경우, 일반적인 symmetric encryption 을 사용해서 키를 생성하기 때문에 사용자 개개인 마다 세부적인 접근 제어를 할 수 없다. 따라서 [2]나 [3]과 같이 attribute-based encryption 을 사용해서 궁극적으로는 검색속도를 최적화하는 한편, 사용자별로 접근 권한을 세부적으로 설정할 수 있도록 하는 기법이 필요하다.

	[1]	[2]	[3]	[4]
data privacy	**O**	**O**	**O**	**O**
trapdoor unlinkability	**O**	**O**	**O**	**O**
Keyword-privacy	**O**	**O**	**O**	**O**
access pattern	**X**	**O**	**X**	**X**
collusion resistance	**N/A**	**X**	**O**	**N/A**
search time	O(mn)	O(mn)	O(mn)	$O(\frac{\theta m \log n}{\omega})$

표 1. 안전성 및 효율성 분석 1

4. 결론

Ranked Searchable Encryption 은 저장하고자 하는 문서에 점수를 매겨 데이터 사용자가 검색을 할 때 상위 k 개 만큼의 문서만 반환하는 기법이다. 본문에서는 네 개의 Ranked Searchable Encryption scheme 을 기능, 안전성, 효율성의 측면에서 비교 분석하였다. 안전성 측면에서 기본적인 보안 요소를 모두 제공하였고, [2]은 access pattern, [3]은 collusion resistance 를 추가로 제공하였다. 효율성은 검색 성능을 비교하였는데, 최근에 와서 tree-based search 를 이용해 최적화가 되고 있다. 앞으로는 attribute-based encryption 을 통해 사용자의 세부적인 접근 제어를 할 수 있도록 하는 기법을 연구할 필요가 있다.

사사

이 논문은 2016 년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2016R1A2A2A05005402).

참고문헌

[1] Ning Cao, Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data. IEEE Transactions on Parallel and Distributed Systems, Vol.25, 2014.

[2] Hongwei Li, Enabling Efficient Multi-Keyword Ranked Search Over Encrypted Mobile Cloud Data Through Blind Storage. IEEE Transactions on Emerging Topics in Computing, 2014.

[3] Hongwei Li, Achieving Authorized and Ranked Multi-keyword Search over Encrypted Cloud Data. IEEE ICC, 2015.

[4] Zhihua Xia, A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data, IEEE Transactions on Parallel and Distributed Systems, Vol.27, 2016.