Journal of Korea Multimedia Society Vol. 17, No. 10, October 2014(pp. 1182-1188)
http://dx.doi.org/10.9717/kmms.2014.17.10.1182

암호화 데이터를 위한 힐버트 커브 기반 다차원 색인 키 생성 및 질의처리 알고리즘

김태훈[†], 장미영[††], 장재우[†††]

Hilbert-curve based Multi-dimensional Indexing Key Generation Scheme and Query Processing Algorithm for Encrypted Databases

Taehoon Kim[†], Miyoung Jang[††], Jae-Woo Chang[†††]

ABSTRACT

Recently, the research on database outsourcing has been actively done with the popularity of cloud computing. However, because users' data may contain sensitive personal information, such as health, financial and location information, the data encryption methods have attracted much interest. Existing data encryption schemes process a query without decrypting the encrypted databases in order to support user privacy protection. On the other hand, to efficiently handle the large amount of data in cloud computing, it is necessary to study the distributed index structure. However, existing index structure and query processing algorithms have a limitation that they only consider single-column query processing. In this paper, we propose a grid-based multi column indexing scheme and an encrypted query processing algorithm. In order to support multi-column query processing, the multi-dimensional index keys are generated by using a space decomposition method, i.e. grid index. To support encrypted query processing over encrypted data, we adopt the Hilbert curve when generating a index key. Finally, we prove that the proposed scheme is more efficient than existing scheme for processing the exact and range query.

Key words: Cloud computing, Encrypted query processing, Distributed index structure, Multi-column query processing

1. 서 론

최근 정부 기관, 은행, 병원 등 다양한 기관/기업에서 대용량의 사용자 데이터를 분석을 통해 효율적인 개인별 맞춤 서비스 제공하기 위한 새로운 IT 기술에 대한 요구가 증대되고 있다. 이러한 대용량 데이터 처리를 위해서는 대규모의 컴퓨팅 자원이 필요하기 때문에, 개인 및 중소규모의 기업에서 직접 시스템을 구축하기에는 한계점이 존재한다. 이를 해결하기 위해, 클라우드 컴퓨팅을 활용한 데이터베이스 아웃소싱(Outsourcing)에 대한 연구가 활발히 진행되고 있다. 데이터베이스 아웃소싱이란 데이터 소유자와 서비스 제공자를 분리하여, 데이터 소유자는 데이터베이스를 구축하고, 서비스 제공자는 데이터 소유자로

※ Corresponding Author : Jae-Woo Chang, Address: (561-756) Dept of Information Technology & Engineering, Jeonbuk National University, Baek-Jae Dae Ro, Duckjin-Gu, Chonju city, Chonbuk, 561-756, South Korea, TEL : +82-63-270-2414, E-mail : jwchang@chonbuk.ac.kr
Receipt date : July 15, 2014, Revision date : Sep. 3, 2014
Approval date : Sep. 11, 2014
[†] Dept of Computer Engineering, Jeonbuk National University (E-mail : taehun3718@jbnu.ac.kr)
[††] Dept of Computer Engineering, Jeonbuk National University (E-mail : brilliant@jbnu.ac.kr)
[†††] Dept of Information Technology & Engineering, Jeonbuk National University
※ This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(grant number 2013010099)

부터 전송받은 데이터베이스를 관리하는 시스템 구조를 말한다. 인증된 사용자는 서비스 제공자를 통해 데이터베이스에 접근하여 정보를 검색한다. 이를 통해, 개인 또는 회사는 IT 인프라에 대한 초기투자 비용 및 유지보수 비용을 절감할 수 있으며, 저렴한 비용으로 대용량 데이터 관리 및 원할한 서비스 제공이 가능하다. 그러나 원본 데이터를 그대로 아웃소싱 할 경우, 서비스 제공자에 의해 데이터 소유자의 의료/질병 정보, 금융 정보 등의 센시티브 데이터(sensitive data)가 노출될 위험이 존재한다. 아웃소싱 수행 시 대용량 센시티브 데이터 유출 문제를 해결하고, 원본 데이터를 보호하기 위해 데이터 암호화 기법이 다양하게 연구되었다[1-9]. 그러나 기존 암호화 기법[1-3]의 경우, 질의처리에 요구되는 데이터 순서 및 값 등의 정보가 유지되지 않기 때문에 암호화 데이터에 대한 질의처리가 제한되며, 다양한 질의를 처리하기에 어려운 문제점이 존재한다. 이러한 환경에서 질의 처리를 수행하기 위해서는 암호화된 데이터를 복호화하여, 복호화 된 원본 데이터 상에서 질의 처리를 수행해야 한다. 초기 데이터 암호화 연구는 데이터 복호화로 인해 서비스 제공자에게 원본 데이터가 노출되는 치명적인 문제점이 존재하며, 전체 데이터베이스를 복호화하기 때문에 질의 수행비용이 증가하는 문제점이 존재한다. 따라서 대용량 센시티브 데이터의 데이터 보호도 향상 및 질의 처리 성능향상을 위해서는 암호화 질의처리 지원이 필수적이다.

대표적인 암호화 연구로, R. A. Popa et al.[9]는 암호화된 데이터 상에서 질의 처리를 할 수 있는 CryptDB를 제안하였다. 제안하는 방법은 암호화 데이터 상에서 SQL 질의를 수행하며 반환된 질의 결과는 암호화 되어 있기 때문에 질의 결과를 유추하는 것이 불가능하다. 그러나 CryptDB는 아래와 같은 두 가지 문제점을 보인다. 첫째, 단일 서버 환경에서 질의를 처리하기 때문에 사용자가 요청하는 질의 처리를 효율적으로 처리할 수 없다. 이는 제한되어 있는 서버 자원을 이용해서 질의를 처리하기 때문에 대용량 데이터를 처리하는데 부적합하다. 둘째, 다중 컬럼에 대한 질의를 지원하지 않기 때문에, 데이터 분석 및 마이닝을 지원하지 못하는 문제점이 존재한다. 따라서 대용량 데이터를 효율적으로 처리할 수 있는 데이터 분산에서 다중 컬럼을 지원하는 연구가 필요하다.

이러한 문제를 해결하기 위해, 본 논문에서는 클라우드 환경에서 그리드 기반 암호화 질의 색인키 생성 기법 및 질의처리 알고리즘을 제안한다. 제안하는 분산 암호화 색인 구조는 암호화 데이터에 대한 효율적인 색인을 지원하기 위해 클러스터링 기반 데이터 색인키를 생성한다. 이때, 컬럼 사이의 연관성을 측정하고, 낮은 연관성을 지닌 컬럼들을 하나의 그룹으로 묶어 클러스터링을 수행함으로써 데이터 보호를 수행한다. 아울러, 색인키 정보 보호를 위해 힐버트 커브(Hilbert curve)를 적용한다. 각 클러스터에 대한 비트맵 ID을 생성함으로써 암호화 데이터를 위한 색인 키를 생성한다. 마지막으로, 생성된 색인키(비트맵 ID)를 이용하여 데이터를 Prefix-Tree에 삽입한다. Prefix-Tree의 말단모드는 분산 환경에서 하나의 데이터 노드를 의미한다. 이와같이, 제안하는 기법은 분산 환경에서 대용량 센시티브 데이터에 대한 강화된 보안을 제공하는 동시에 암호화 질의처리를 빠르게 수행하는 것이 가능하다.

본 논문의 구성은 다음과 같다. 2장에서는 기존의 CryptDB가 질의 처리하는 기법에 대해서 소개한다. 다음으로 3장에서는 제안하는 클라우드 환경에서 그리드 기반 색인키 생성 기법 및 암호화 알고리즘의 구조에 대해서 설명하고, 4장에서는 성능평가를 통해 제안하는 기법이 기존 기법에 비해 질의 처리 대비 성능이 우수함을 보인다. 마지막으로 5장에서는 결론 및 향후 연구에 대해 기술한다.

2. 관련연구

암호화된 데이터베이스 상에서 질의 처리를 수행하기 위한 연구는 크게 i) 센시티브 데이터 보호를 위한 암호화 기법, ii) 암호화 질의 처리 알고리즘으로 구분된다. 본 장에서는 대표적인 데이터 암호화 기법 및 암호화 질의처리 알고리즘에 대해 기술하고, 문제점을 분석한다.

2.1 데이터 암호화 기법

A. Desai.의 연구[4]는 데이터 입력 시 이전 영역과 다른 암호화 영역(ciphertext plain)을 선택하여 매핑(mapping) 함으로써 암호화를 수행하는 random symmetric 암호화 기법을 제안하였다. 해당 기법은 임의의 <암호문, 평문> 쌍을 이용한 암호화 키

유출 공격을 방지한다. 둘째, O. Goldreich의 연구[5]는 데이터 블록단위의 비트변환을 이용하여 데이터를 암호화하는 deterministic symmetric 암호화 기법을 제안하였다. 아울러, 김보선 외[6]의 연구는 교무업무시스템 상에서 학생의 개인 정보 보호를 위해 ARIA 블록 암호화 알고리즘을 적용한 데이터베이스 시스템 및 개선된 SQL질의 처리 기법을 제안하였다. Deterministic symmetric 암호화 기법을 이용한 기법들은 입력 값 x에 대해 하나의 암호화 값 Encx를 생성하기 때문에, 복호화 없이 정확매칭 질의를 지원할 수 있는 장점이 존재한다.

셋째, T. Ge et al.의 연구[7]는 지수/로그 등의 함수를 이용하여 평문 공간과 암호문 공간에 정의된 연산을 보존하는 homomorphic 암호화 기법을 제안하였다. 해당 기법은 데이터 복호화 없이 산술 연산을 수행하는 장점이 존재한다. 넷째, A. Boldyreva et al.의 연구[8]는 암호화 함수를 사용하여 원본 데이터의 순서를 보존하는 order-preserving 암호화(OPES) 기법을 제안하였다. 이는 원본 데이터의 순서가 보존되기 때문에, 복호화 없이 정확매칭/범위 질의처리를 수행할 수 있는 장점이 존재한다. 그러나 기존 데이터 암호화 기법은 다음과 같은 문제점을 지닌다. 첫째, random symmetric 암호화 기법은 복호화 없이 질의 수행이 불가능하여 서비스 제공자에 의해 원본 데이터가 노출되는 문제점이 존재한다. 둘째, deterministic symmetric 기법, homomorphic 암호화 기법은 암호화된 데이터의 순서가 유지되지 않기 때문에 범위 질의 및 Top-k 질의 처리가 불가능하며, 공격자가 데이터의 빈도수를 파악하여 원본 데이터를 유추하는 데이터 카운트 공격에 취약하다. 따라서 본 논문에서는 데이터 암호화 기법으로 OPES를 적용하여 암호화를 수행하고, 이를 기반으로 암호화 질의 처리를 위한 분산 인덱스 및 암호화 질의처리 알고리즘을 제안한다.

2.2 암호화 질의 처리 기법

암호화된 데이터 상에서의 대표적인 SQL-like 질의 처리 기법은 R. Popa et. al.의 연구[9]에서 제안한 CryptDB가 존재한다. CryptDB는 데이터의 타입에 따라 독립적으로 암호화를 수행함으로써 정확 매칭, 범위 질의 및 문자 검색 등 사용자의 요구사항을 만족하는 다양한 질의를 지원한다. CryptDB는 데이터의 타입 및 적용가능한 질의 타입을 고려하여, 컬럼별로 독립적인 암호화를 수행한다. 예를 들어 정확매칭을 지원하기 위해서는 해당 컬럼에 AES[2]와 같은 Deterministic 암호화 기법을 적용하여 데이터를 암호화한다. CryptDB는 다양한 암호화 기법을 사용하여 정확 매칭, 범위 질의 및 문자 검색 등 사용자의 요구사항을 만족하는 다양한 질의를 지원하는 장점이 존재한다. 그러나 CryptDB는 질의 유형에 따라 다른 암호화 기법을 적용하여 중복 저장하기 때문에, 저장 오버헤드가 존재한다. 아울러, 서로 다른 암호화 기법으로 암호화된 컬럼에 대해서는 질의 처리를 수행하는 것이 불가능하며, 다중 컬럼 질의를 지원하지 못하는 문제점이 존재한다.

3. 클라우드 환경에서 그리드 기반 색인키 생성 기법 및 암호화 질의처리 알고리즘

3.1 시스템 구조

암호화된 단일 컬럼 상에 질의 처리를 수행하는 CryptDB는 질의 처리 시 단일 서버 환경에서 처리하였기 때문에, 질의 수행이 지연되는 단점이 있다. 또한 다중 컬럼 질의를 지원하지 않기 때문에, 해당 질의를 처리하기 위해서는 질의를 포함하는 컬럼들을 서버에서 복호화를 해야 한다. 현재까지 암호화 된 데이터 상에서 다중 컬럼 질의 지원을 위한 색인키 생성 및 질의처리 알고리즘에 대한 연구는 전무한 실정이다. 이를 해결하기 위해, 다중 컬럼 질의 지원을 위한 색인키 설계가 필요하며, 설계된 색인키의 정보 보호를 위해 암호화 색인키를 생성해야 한다. 또한 생성된 색인키를 질의 처리 할 때 성능을 보장받을 수 있어야 한다.

본 논문에서는 클라우드 환경에서 그리드 기반 색인키 생성 기법 및 암호화 질의처리 알고리즘을 제안한다. 제안하는 기법의 시스템 구조는 Fig. 1과 같다. 정확 매칭 질의와 범위 매칭 질의 지원을 위한 순서 보존 암호화 모듈, 암호화 된 데이터의 다중 컬럼 색인키 질의 지원을 위한 힐버트 기반 그리드 모듈, 분산된 환경에서 질의 처리 성능 보장을 위한 Prefix-Tree로 구성된다. 클라우드 서비스 환경에서 대용량 센시티브 데이터를 아웃소싱하기 위한 과정은 다음과 같다. 첫째, 데이터 소유자는 데이터 보호도 향상 및 효율적인 분산 저장관리를 지원하기 위해, 순서

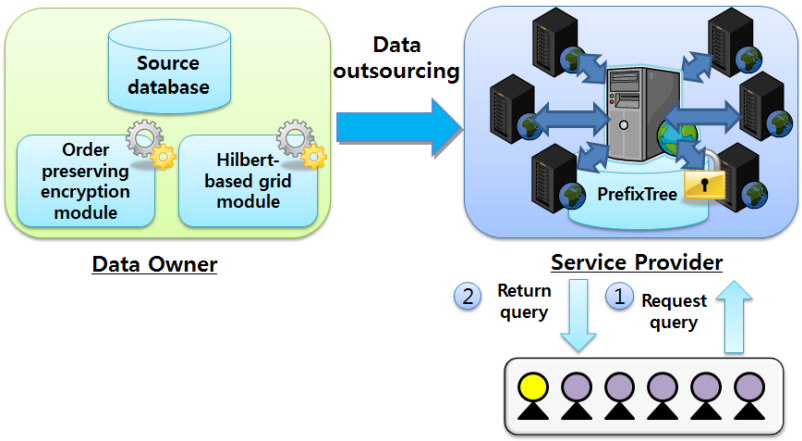


Fig 1. System architecture of the proposed system.

보존 암호화 모듈과 힐버트 기반 그리드 모듈을 이용해 데이터 암호화 및 상위 색인키 생성을 수행한다. 둘째, 서비스 제공자는 전송받은 색인키를 기반으로 데이터 분산저장을 위한 Prefix트리 기반 분산 색인 구조를 구축한다. 구축된 색인을 이용하여 데이터 소유자로부터 전송받은 암호화 데이터를 분산 데이터 서버에 저장한다. 마지막으로 인증된 사용자는 질의를 암호화하여 서비스 제공자에게 질의처리를 요청한다. 서비스 제공자는 분산 색인키를 탐색함으로써 암호화 질의 처리를 수행하고, 검색된 후보 결과 집합을 인증된 사용자에게 전송한다. 이후 사용자는 전송받은 후보 결과 집합을 복호화하고, 최종 결과를 선정한다.

3.2 클라우드 환경에서 그리드 기반 색인키 생성 기법 및 암호화 질의처리 알고리즘

본 절에서는 클라우드 컴퓨팅 환경에서 다중 컬럼 기반 정확 매칭, 범위 매칭 질의를 위한 그리드 기반 색인키 생성 기법 및 암호화 질의처리 알고리즘을 제안한다. 알고리즘은 크게 컬럼 유사도 측정 통한 그리드 조합 선정, 힐버트 커브 기반 그리드 매칭 및 색인 키 생성 및 검색으로 구성된다.

Step 1. 컬럼 유사도 측정을 통한 그리드 조합 선정

데이터 소유자는 먼저 원본 데이터의 각 컬럼을 POPIS[8]로 암호화를 수행한다. 아울러, 암호화 데이터에 대한 효율적인 클러스터링을 위해, 컬럼간 상관분석을 수행하여 산개형에 가까운 데이터 분포도를 나타내도록 컬럼 그룹을 선정한다. 컬럼간 상관분석이란, 서로 다른 컬럼 내의 데이터들이 얼마나 밀접하게 직접적으로 관련되어있는가 하는 정도를 분석하는 통계적 분석방법이다. 이를 고려하면, 서로 상관 계수가 낮은 컬럼들을 선정하여 클러스터링을 수행하는 것이 가능하다. 본 논문에서는 컬럼간 상관계수 측정을 위해, 대표적인 상관계수 측정기법인 피어슨 상관 계수(Pearson correlation coefficient)[10]를 이용한다. 대용량 센시티브 데이터에서 모집단을 통해 상관 계수를 측정하는 것은 매우 높은 계산 비용을 요구하기 때문에, 표본 집단을 통해 모집단의 상관 계수를 추정(estimation) 하는 것이 필수적이다. 이를 위한 피어슨 표본 상관 계수(sample correlation coefficient)는 Expression 1을 통해 계산한다.

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X}) - (Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2} = \frac{S_{xy}}{S_x S_x} \quad (1)$$

예를 들어, 질의에 사용되는 센시티브 컬럼이 {0, 1, 3, 5}이라 가정했을 때, 해당 컬럼들의 모든 조합에 따른 피어슨 상관 계수를 측정한 결과는 Table 1과 같다. 그 중, 피어슨 상관 계수 값이 가장 작은 컬럼 조합 (0, 3)을 선정하여 데이터 클러스터링을 수행한다.

Step 2. 힐버트 커브 기반 그리드 매칭 및 색인키 생성 및 검색

암호화된 데이터의 보호도 향상을 위해, 주어진

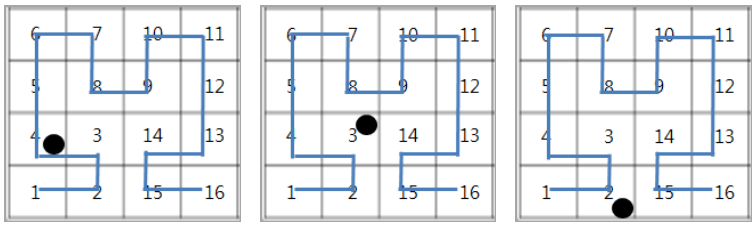


Fig. 2. Grid matching based on the Hilbert curve.

Table 1. Examples of measurements of the Pearson correlation coefficients corresponding to the combination of the column

Combination of column		Pearson's correlation coefficient
0	3	0.663843
1	3	0.666327
3	5	0.677279

데이터 색인키는 원본 데이터에 대한 상세 정보를 포함하지 않아야 한다. 즉, 서비스 제공자를 비롯한 어떠한 공격자도 변환된 데이터를 이용하여 원본 데이터를 유추할 수 없어야 한다. 이를 위해, 제안하는 기법에서는 Step 1에서 선정한 컬럼 조합에 대해 k개의 그리드 인덱스를 생성하고, 힐버트 커브(Hilbert curve)를 이용하여 그리드 id를 변환하고, 변환된 id를 조합하여 색인 키를 생성한다. 해당 예제에서는 컬럼 조합 중 상위 3개의 조합을 이용하여 3개의 그리드 인덱스를 생성한다. 아울러, 각 그리드 인덱스에서 생성된 id를 힐버드 커브에 매칭하고, 이를 순차적으로 조합하여 색인키를 생성한다. 예를 들어, Fig. 2는 하나의 레코드를 컬럼 조합에 따라 각 그리드 인덱스에 삽입한 결과를 나타낸다. 삽입한 데이터의 그리드 셀 id를 힐버트 커브에 매칭한 경우 셀 id는 4, 3, 2이며, 이를 비트맵으로 연속 저장하여 색인키를 생성하면 0100 0011 0010의 색인키가 생성된다. 생성된 키는 분산된 환경의 PrefixTree에 저장된다.

Algorithm 1. Grid-based Index Generation and Query Processing Algorithm

Input : original database D
Output : encrypted database D′, bit
/*data owner*/
1: popis = EncPOPIS(D);
2: column = CalculatePearson(D);
3: grid = CreateGrid(k, column);
4: For each data d of D
5: GridMaching(grid, k);
6: grid = MatchHilbertCurve(grid);
7: bit = GetHilbertCurveID(grid, d);
/*send to service provider*/
8: D′ = popis + bit

Fig. 3. The proposed index key generation algorithm.

제안하는 색인키 생성 알고리즘은 Fig. 3과 같다. 첫째, 원본 데이터베이스 각 컬럼을 POPIS로 암호화하여 암호화된 데이터베이스 D′를 생성한다(line 1). 둘째, 원본 데이터 D의 모든 컬럼에 대해 컬럼 유사도 측정을 통한 그리드 조합을 선정한다.(line 2). 셋째, 선정된 컬럼 조합에 대해 k개의 그리드를 생성한다.(line 3) 넷째, 각 레코드의 그리드 셀 id를 힐버트 커브 id로 변환시킨다. 다섯째, 변환된 힐버트 셀 id를 비트로 변환하고, 이를 연속적으로 저장하여 최종적 색인키를 생성 및 저장 한다.(line 4-7) 다섯째, 각 컬럼에 대한 색인키와 OPES 암호화를 서비스 제공자에게 전송하고 알고리즘을 종료한다.

4. 성능 평가

본 장에서는 기존 암호화 질의처리 기법인 CryptDB와 제안하는 기법의 질의 처리 성능을 분석한다. 성능평가의 실험 환경은 Table 2와 같다. 아울러, 실험 데이터는 UC Irvine 대학에서 제공하는 US Census Data Database[11]를 이용하였다. Census-Original 데이터는 이름, 결혼여부, 자녀수, 성별, 나이, 학력, 직업 및 전문 분야, 직업에 의한 소득, 재산에 의한 소득 및 지출 등을 포함한다. 이 중 4개의 컬럼을 이용하여 질의를 수행하였으며, 총 100회의 질의를 수행한 결과의 평균을 측정하였다. 정확 매칭 질의의 경우, 데이터 크기를 0.5GB, 1GB, 1.5GB, 2GB로 변화하면서 성능을 측정하였으며, 범위 매칭 질의는 데이터 크기를 2G로 고정하고, 질의 범위는 전체 데이

Table 2. Experimental environment parameters

CPU	Intel®CoreTM i3-2100 CPU 3.10Ghz
Memory	2GB
O/S	Windows 7 64bit
Compiler	Visual Studio 2010 C++

터의 크기에 대한 백분율로 변화하여 수행하였다. 해당 데이터가 대용량 데이터이므로 질의 영역을 0.0001%~ 0.001%의 질의 영역을 포함하도록 설정하여 수행하였다.

Fig. 4는 데이터 크기 변화에 따른 정확 매칭 질의를 비교한 것이다. 기본 기법은 단일 서버 환경에서 질의 처리를 수행하기 때문에 질의 처리에 필요한 시간이 증가하며, 다중 컬럼 질의를 위해서는 서로 다른 암호화를 적용한 컬럼을 복호화 하여 데이터를 검색하기 때문에 데이터복호화 비용이 추가로 필요하다. 한편, 제안하는 기법은 분산 암호화 색인 키 생성을 통해 복호화 없이 빠르게 탐색이 가능하여 질의 처리 성능이 향상되었음을 알 수 있다. Fig. 4에서, 데이터 크기가 2G인 경우, CryptDB는 약 0.4초의 정확 매칭 질의 처리 시간을 소요한다. 한편, 제안하는 기법의 경우 평균 질의 처리 시간이 약 0.12초로 약 4배 향상된 질의 처리 성능을 지원한다.

Fig. 5는 질의 영역 크기 변화에 따른 범위 매칭 질의 성능을 비교한 것이다. 정확 매칭 질의와 마찬가지로 제안하는 기법이 기존 CrypDB에 대해 개선된 질의 처리 성능을 제공함을 알 수 있다. 이는 제안하는 기법이 분산 암호화 색인 키를 이용하여 복호화 없이 Prefix-Tree 탐색만을 통해 빠르게 접근하여 데이터를 반환하기 때문이다. Fig. 5에서 0.001%의 질의 영역에 대한 데이터 탐색의 경우, CryptDB는

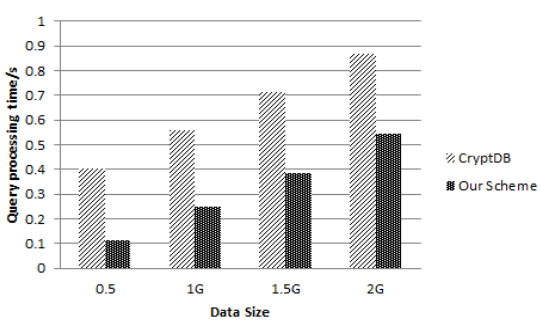


Fig. 4. Exact matching performance evaluation associated with changes in the data set.

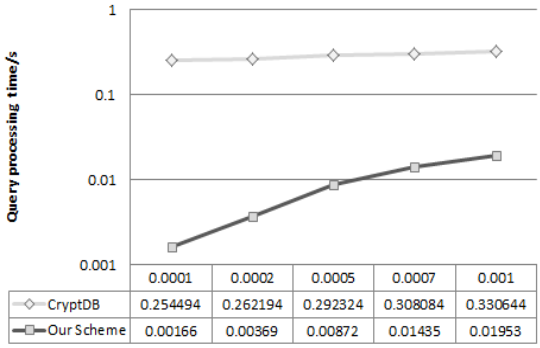


Fig. 5. Range matching performance evaluation associated with change in the number of quality.

약 0.33초의 질의 처리 시간을 소요하며, 제안하는 기법의 질의 처리시간은 약 0.02초로 기존 기법에 비해 약 15배 성능이 향상되었음을 알 수 있다.

5. 결 론

본 논문에서는 클라우드 환경에서 데이터 보호 및 분석 질의처리를 위한 그리드 기반 색인 키 생성 기법 및 암호화 질의처리 알고리즘을 제안 하였다. 제안하는 기법은 다중 컬럼 질의를 위해 그리드 기반 다중 컬럼 질의 색인 키를 생성하며, 색인 키 정보 보호를 위해 힐버트 커브를 적용한다. 아울러 분산된 환경에서 질의 처리 성능 보장을 위해 Prefix-Tree를 사용한다. 이를 통해 다중 컬럼 질의를 지원하며, 분산된 환경 내에서 빠르게 질의를 수행하는 것이 가능하다. 기존 암호화 질의처리 연구인 CryptDB와 성능 비교를 통해 제안하는 기법이 정확 매칭 질의 및 범위 질의 처리 측면에서 기존 기법에 비해 우수한 질의 처리 성능을 지원함을 검증하였다.

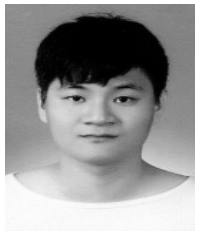
향후 연구 방향은 제안하는 기법을 top-k 및 집계 질의를 지원하는 알고리즘으로 확장하여 연구하는 것이다.

REFERENCE

[1] NIST, *Digital Signature Standard(DSS),* Federal Information Processing Standards Publication 186-3, 2009.

[2] NIST, *Advanced Encryption Standard(AES),* Federal Information Processing Standards Publication 197, 2001.

[3] RSA Laboratories, RSAREF, A Cryptographic Toolkit Version 2.0(1994). http://www.csm.ornl.gov/~dunigan/rsaref.txt (accessed Sep., 19, 2014)
[4] A. Desai, *New Paradigms for Constructing Symmetric Encryption Schemes Secure against Chosen-ciphertext Attack*, Springer, Berlin Heidelberg, 2000.
[5] O. Goldreich, *Foundations of Cryptography: Volume I Basic Tools*, Cambridge University Press New York, 2001.
[6] B. Kim, E. Hong, "Implementation and Performance Evaluation of Database Encryption for Academic Affairs System", Journal of Korea Multimedia Society Vol. 11, No1. pp. 1-12, 2008.
[7] T. Ge and Z. Stan, "Answering Aggregation Queries in a Secure System Model," *Proceeding of the 33rd International Conference on Very Large Data Bases*, pp. 519-530, 2007.
[8] A. Boldyreva, N. Chenette, and A. O'Neill, *Order-Preserving Encryption Revisited: Improved Security Analysis and Alternative Solutions*, Springer, Berlin Heidelberg, 2011.
[9] R.A. Popa, C.M.S. Redfield, N. Zeldovich, and H. Balakrish-nan, "CryptDB: Protecting Confidentiality with Encrypted Query Processing," *Proceeding of the Twenty-Third ACM Symposium on Operating Systems Principles*, pp. 85-100, 2011.
[10] J. Wang, *Pearson Correlation Coefficient*, Springer, New York, 2013.
[11] US Census Data(1990), https://archive.ics.uci.edu/ml/datasets/US+Census+Data+(1990) (accessed Sep., 19, 2014)



김 태 훈

2013년 전북대학교 컴퓨터공학과 학사
2013년~현재 전북대학교 컴퓨터공학과 석사과정

장 미 영

2009년 전북대학교 컴퓨터공학과 학사
2011년 전북대학교 컴퓨터공학과 석사
2011년~현재 전북대학교 컴퓨터공학과 박사과정

장 재 우

1984년 서울대학교 전자계산기공학과 학사
1986년 한국과학기술원 전산학과 석사
1991년 한국과학기술원 전산학과 공학 박사
1996년~1997년 Univ. of Minnesota, Visiting Scholar.
2003년~2004년 Penn State Univ., Visiting Scholar.