6W-02

クラウド環境におけるプライバシ保護を考慮した検索処理のための軽量な階層索引

篠塚 千愛† 渡辺 知恵美‡ 北川 博之‡

†筑波大学情報学群情報科学類 ‡筑波大学システム情報系

1 はじめに

近年，Database as a Service (DaaS) を利用した検索サービスが増加している．DaaS とは，インターネットを介してデータベース機能を提供するクラウドコンピューティングサービスである．データ所有者は DaaS を利用することで安価にデータベースを使用できる上，コストのかかるデータベース管理を委託できる．

しかしながら，DaaS を利用した検索サービスでは，第三者であるクラウドのサーバ管理者に対するデータとクエリ双方のプライバシ保護が求められる．既存手法として，Hu らは紛失通信によるデータとクエリ双方をサーバに秘匿可能な検索手法を提案した [1] が，我々の調査により，この手法では安全性やクライアントの負荷，処理コストの点で問題があることが分かった．

そこで本研究では，クライアントの負荷や安全性の観点から，既存手法における索引の構造や保管場所，システム構成を見直し，クライアントの負荷を軽減した安全な秘匿検索手法を提案する．また，処理コストのより小さな探索アルゴリズムを検討し，検索処理の高速化を図る．

2 基本事項

2.1 準同型暗号

準同型暗号は，2 つの平文に対する演算を 2 つの暗号文に対する演算に変換できる準同型性を有する暗号である．本稿では，準同型暗号に Paillier 暗号を使用する．Paiilier 暗号は，2 つの平文 m_1 および m_2 の暗号文から $m_1 + m_2$ の暗号文を計算できる加法準同型性を有する．

A Lightweight Hierarchical Indexes for Query Processing Preserving Privacy in the Cloud Environment

Chisato Shinozuka†(sn@kde.cs.tsukuba.ac.jp),
Chiemi Watanabe‡(chiemi@cs.tsukuba.ac.jp) and
Hiroyuki Kitagawa‡(kitagawa@cs.tsukuba.ac.jp)
†College of Information Sciences University of Tsukuba
‡Faculty of Engineering, Information and Systems, University of Tsukuba

2.2 GT-SCOT

GT-SCOT (Strong Conditional Oblivious Transfer for “Greater Than” predicate)[2] は，条件付き紛失通信プロトコルの一種で，受信者が持つ値 x と送信者が持つ y の値をお互いが知ることなく，その大小関係を受信者のみが知ることができる．GT-SCOT では，受信者が暗号化された値 $E(x)$，大小関係が $x < y$ であったことを示す値 s_0, および $x > y$ であったことを示す値 s_1 を送信者に送信すると，送信者によって x と y の大小比較として $E(s_0)$ または $E(s_1)$ が得られる計算アルゴリズムが適用され，結果が返される．受信者は自身のもつ復号化関数 $D(\cdot)$ を適用することで，s_0 もしくは s_1 を取得できる．送信者側で大小比較を行なっているものの，比較結果は暗号化された状態で導出されるため，送信者は大小関係を知ることができない．

3 先行研究

文献 [1] では，ノードの各エントリを Pailier 暗号および可換暗号で暗号化した B^+ 木索引をクライアントで保持し，探索の際に暗号化されたエントリ $E(p)$ と暗号化されていないクエリ q の大小比較を GT-SCOT を用いて求めることで，データとクエリ双方を秘匿する検索手法を提案している．GT-SCOT による大小比較の様子を図 1 に示す．初めにランダム値 r を送り，準同型性を利用して $E(p)$ および q を書き換えることで，サーバにクエリを秘匿する．

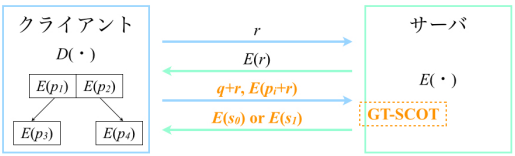

図 1: GT-SCOT による大小比較の様子

しかしながら，この手法では索引を保持するクライアントの負荷が大きい上，探索毎に通信を行なうため処理時間が長い．また，クライアントが GT-SCOT で

比較結果の取得時に適用する復号化関数で索引のエントリも復号化できることは安全性に関わる問題である．

4 提案手法

本研究では，クライアントの負荷の軽減および安全性の向上を図るため，索引をサーバで保持した秘匿検索手法を提案する．システムの構成および探索の流れを図 2 に示す．サーバで索引を保持するにあたり探索経路によるクエリ推定を防ぐため，サーバを 2 台用意して索引保持と大小比較をそれぞれ別のサーバで担うようにしている．以下，索引とデータを保管するサーバをデータサーバ，大小比較を行なうサーバを計算サーバと呼ぶ．提案手法では，索引を構造情報とエントリ情報とに分離し，クライアントで構造情報を，データサーバでエントリ情報を保持する．また索引のデータ構造についても見直した．DBMS ではページ IO が検索時間に影響するが，この手法ではクエリとエントリの比較に通信が必要なため，比較回数が検索時間に影響する．そこで提案手法では，比較回数を減らすために二分探索木を採用する．

クライアントは構造情報を用いて比較したいエントリ番号とランダム値 $E(r)$ をデータサーバに送信する．データサーバはクライアントが指定したエントリの書き換え後の情報 $E(p+r)$ を，クライアントは書き換え後のクエリ $q+r$ をそれぞれ計算サーバに送って GT-SCOT を行ない，計算サーバは結果をクライアントに返す．クライアントは得られた比較結果を復号化し，新たに比較したいエントリ番号をデータサーバに送信するということを繰り返して探索を行なっていく．

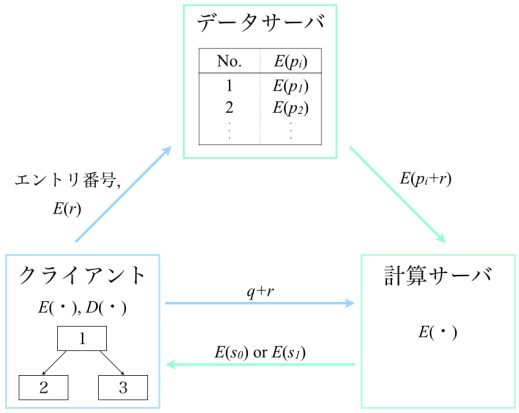

図 2: 提案手法のシステムの構成および探索の流れ

5 評価実験

前節で述べた提案手法の性能を評価するため，様々なデータ数 N における平均クエリ応答時間を測定し，既存手法との比較を行なった．各データ数 N における既存手法・提案手法での平均クエリ応答時間の比較結果を図 3 に示す．提案手法では N が大きくなっても処理時間が急激に長くならないことが分かる．

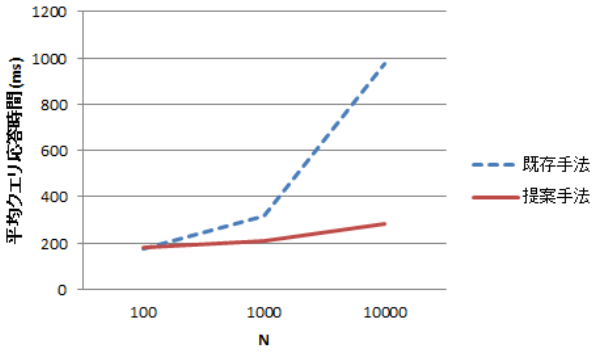

図 3: 各データ数 N における平均クエリ応答時間

6 まとめと今後の課題

本稿では，クライアントの負荷が小さく，効率的で安全なデータとクエリ双方の秘匿検索手法を提案した．また，評価実験より，クライアントの負荷軽減および高速化に対する有効性が示された．

今後の課題としては，データサーバによる構造情報の推定を防ぐため，二分探索にランダム性を組み込むことを検討する．

謝辞

本研究の一部は，文部科学省 “実社会ビックデータ利活用のためのデータ統合・解析技術の研究開発” による．

参考文献

[1] H. Hu *et al.*, “Private search on key-value stores with hierarchical indexes,” Data Engineering (ICDE), 2014 IEEE 30th International Conference on, pp. 628–639 (2014).

[2] I.F. Blake *et al.*, “Strong Conditional Oblivious Transfer and Computing on Intervals,” Advances in Cryptology – ASIACRYPT 2004, pp. 515–529 (2004).