データベースアウトソーシングにおける プライバシー保護に考慮した範囲検索法

新井裕子　(指導教員：渡辺 知恵美)

1　はじめに

近年，データベースの管理運用を外部の技術者に委託するサービスの普及に伴い，データプライバシーに対する関心が高まっている．このサービスにおいて管理者は第 3 者であるため，不十分な管理を行ったり，データを意図的に悪用することも考えられる．そこで，データを暗号化してからサービスプロバイダに格納することで，機密を保持する研究が行われてきた．その研究の一つに，2 段階暗号を用いた完全一致検索がある [1]．この手法を用いると，管理者に問合せ条件やその結果を知られることなく，サービスプロバイダ側で問合せを実行することができる．本稿では，文献 [1] を拡張した範囲検索法を提案する．

2　2 段階暗号による暗号化データへの検索

図 1 に Yang 氏らによって提案された 2 段階暗号 [1] を用いた問合せの流れを示す．元データの暗号化はセル単位に行われ，その際 2 つの異なる鍵を用いる．暗号化処理は全てクライアント側で行うため，管理者でもデータの内容を知ることはできない．また $data1$ を生成する際，元データに乱数を加え暗号化することで，分布解析による大まかな内容把握を防いでいる．

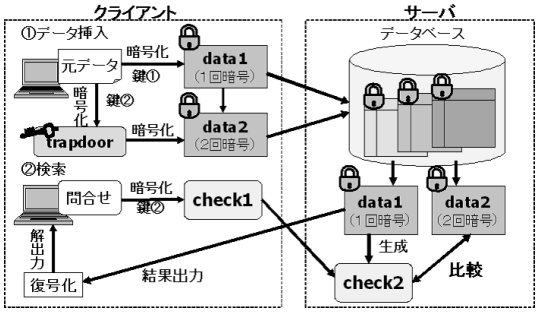

図 1: 2 段階暗号を用いた問合せの流れ

テーブル T の i 行 j 列目のセルを T_{ij}，暗号化関数 $E()$，乱数 r_i，2 種類の鍵 $k1$,$k2$，問合せの $where$ 文に含まれる値を x とし，各要素の式を以下に示す．

$$
\begin{aligned}
&data1 = E_{k1}(T_{ij} + r_i)\\
&trapdoor = E_{k2}(T_{ij})\\
&data2 = E_{trapdoor}(E_{k1}(T_{ij} + r_i))\\
&check1 = \overline{E_{k2}(x)}\\
&check2 = E_{\underline{check1}}(E_{k1}(T_{ij} + r_i))
\end{aligned}
$$

サーバで比較される data2 と $check2$ の違いは，鍵にある (上式下線部)．$trapdoor$ と $check1$ はそれぞれ T_{ij}，x を鍵②で暗号化したものであるから，$T_{ij} = x$(T_{ij} が解) であれば，$trapdoor = check1$ となるため，$data2 = check2$ となり，問合せの解と判定される．一方 $T_{ij} \neq x$ のとき，$trapdoor \neq check1$ となるため，データ $2 \neq check2$ となり，問合せの解でないとわかる．

3　暗号化データベースに対する範囲検索

本研究では，前節で紹介した 2 段階暗号をベースとし，ブルームフィルタを用いることで範囲検索へと拡張した．また文献 [1] 同様，問合せの一部をサーバで行うことでクライアントの負担を減らしている．

3.1　ブルームフィルタ

ブルームフィルタとは，ある要素がある集合に含まれるかどうかをテストする際に用いられるビット列のことである．あらかじめ適当な数のハッシュ関数と全て 0 に設定された空のブルームフィルタを用意しておく．集合の各要素に対するハッシュ値を計算し，空のブルームフィルタに 1 を立てていく．次に各要素のブルームフィルタの論理和をとり，集合のブルームフィルタとする．テストは集合のブルームフィルタと検索語句のハッシュ値を用いて行われる．検索語句のハッシュ値の全ての位置に 1 が立っていれば，その要素は集合に含まれていると判断される．

集合 S=｛東京都, 文京区, 大塚｝，3 種類のハッシュ関数 $hash1$, $hash2$, $hash3$，8 ビットの空のブルームフィルタを例に説明する．図 2 に集合 S の各要素である『東京都』『文京区』『大塚』から得られたブルームフィルタを示す．集合の各要素のブルームフィルタの論理和をとり，集合 S のブルームフィルタを生成する．次に，このブルームフィルタと検索語句『豊島区』のハッシュ値を比較しテストを行う．検索語句『豊島区』から次のようなハッシュ値が得られたとする．

$$
\begin{aligned}
&hash1(\text{豊島区}) = 2\\
&hash2(\text{豊島区}) = 8\\
&hash3(\text{豊島区}) = 1
\end{aligned}
$$

集合 S のブルームフィルタの 2,8,1 番目のビット列を見ると，2 ビット目が 0 となっているため豊島区は集合 S に含まれないとわかる．

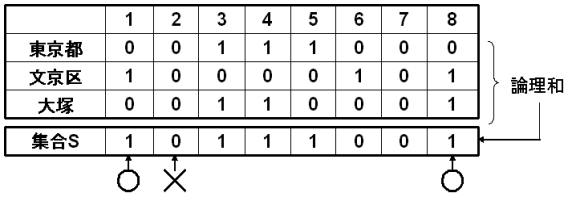

図 2: ブルームフィルタの様子

3.2　提案する範囲検索法

図 3 に提案する範囲検索法の流れを示す．暗号化には 2 節で紹介した手法を用いる．また各属性のドメインを複数の領域に分割し，属性値をドメインの最小値からその値までの領域と見なす．範囲検索は，領域同士の包含関係を調べることで実現している．元データを鍵①で暗号化したものを $data1$，バケット番号の各要素を鍵②で暗号化したものを $trapdoor$，各 $trapdoor$

を鍵とし $data1$ を暗号化したものを $data2$ とする．値によっては $trapdoor$・$data2$ は複数となる．また $data1$ を生成する際，元データに乱数を加えた後，暗号化することで，データ分布解析からの内容把握を防ぐことができる．次に $data2$ からブルームフィルタを生成し，そのブルームフィルタと $data1$ をサーバのデータベースに格納しておく (図 3①)．検索は，サーバに格納されたブルームフィルタに対して $check2$ の包含テストを行うことで実現する．$check1$ は各バケット番号を鍵②で暗号化したものであり，$check2$ は $check1$ を鍵としサーバに格納された $data1$ を暗号化したものである．問合せの解であるとされたタップルからユーザが必要とする属性の $data1$ のみクライアントに返す．$data1$ はクライアントで復号化され，解精製の後ユーザに送られる (図 3②)．

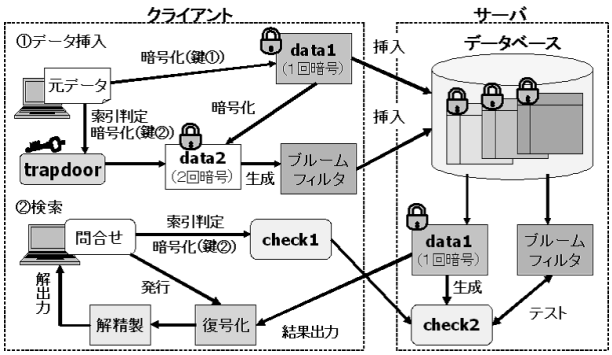

図 3: 提案する範囲検索法の流れ

3.3 挿入例

図 4 に属性『在庫』の挿入例を示す．元データに適当な乱数 r_i を加え, 鍵①で暗号化しデータ 1 を生成しておく．次にあらかじめ用意しておいた索引情報からバケット番号を割り出す (図 4①)．20, 65 のバケット番号はそれぞれ $\{a\}$, $\{a, b, c\}$ である．各バケット番号を鍵②で暗号化し (図 4②)，$trapdoor$ を生成する (図 4③)．元データが大きい値である程，生成される $trapdoor$ の数は増えることになる．$trapdoor$ を鍵とし $data1$ を暗号化し $data2$ を生成する (図 4④)．$data2$ からブルームフィルタを生成し，サーバにそのブルームフィルタと $data1$ を格納する..

3.4 検索例

ユーザから『$select$ id $from$ 商品 $where$ 在庫≧ 30』が発行されたとする (図 5)．$where$ 文の 30 のバケット番号 {a,b} の b から $check1$ を生成し, 問合せを書き換え, サーバに発行する．書き換え後の問合せに含まれる match() 関数は範囲検索を行う関数である．クライアントから渡された $check1$ を鍵として各行の data1 を暗号化し, $check2$ を生成する．1 つ目の $check2$ からハッシュ値 6,8,10 が得られたとする．1 行目のブルームフィルタの 6 番目には 1 が立っておらず，解ではないと判断される．一方，2 つ目の check2 からハッシュ値 3,5,6 が得られたとすると，2 行目のブルームフィルタには，いずれも 1 が立っているので $check2$ は含まれ，この元データは 30 より大きいと判断される．結果はクライアントに返され，復号化・解精製の後，ユーザに渡される．

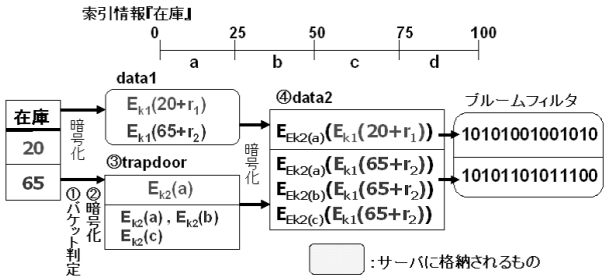

図 4: 挿入例

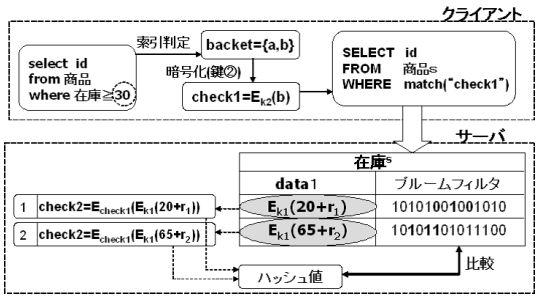

図 5: 検索例

4 実装

範囲検索システムはクライアントで実行される暗号モジュールと reSQL モジュール，サーバで実行される検索モジュールの 3 つから構成される (図 6)．クライアント部分の実装は ruby 言語で行い，RDBMS には PostgreSQL のバージョン 8.2 を用いている．各モジュールの詳細は文献 [2] を参照されたい．

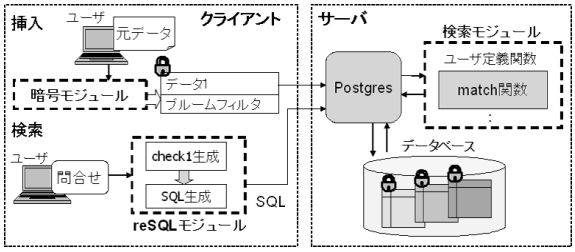

図 6: 範囲検索システムの構成

5 まとめと今後の課題

本稿では，データベースアウトソーシングにおけるプライバシー保護に考慮した範囲検索法を提案した．データベースの閲覧など強い権利を持つ管理者に対する機密保持に有効であると考えられる．しかし処理数が多く実行時間が膨大になる可能性がある．実装は今後の課題であり，高速化についても考える必要がある．

参考文献

[1] Zhiqiang Yang et al: “Privacy-Preserving Queries on Encrypted Data”, *Proceedings of the 11th European Symposium On Research In Computer Security*, LNCS4189, pp.479-495, 2006.

[2] 新井裕子，渡辺知恵美: “データベースアウトソーシングにおけるプライバシー保護に考慮した範囲検索法”，第 19 回データ工学ワークショップ (DEWS2008)(投稿中)