ブルームフィルタを用いたプライバシ保護検索システムの実装

合田智美　(指導教員：渡辺知恵美)

1　はじめに

近年 Database as a Service（DaaS）が注目を集めている．しかし，DaaS ではデータ管理者はインターネット上の外部の第 3 者であるため，機密情報を守りたいというユーザの要求が生じる．その為，プライバシ保護検索というデータを暗号化した状態でデータベースに保存し，暗号化したまま問合せを施す手法がこれまで多くの研究で提案されてきた．

我々もまた先行研究 [1][2] にて，ブルームフィルタを用いたスキーマ情報を隠蔽するプライバシ保護手法を提案してきたが，これまでは手法の提案と検証がメインだった．本研究では，この手法を用いてクライアント側でスキーマ情報や検索条件等を隠蔽した上でデータベースサーバと通信を行うようなプライバシ保護検索システムの開発を行う．

2　先行研究

先行研究 [1][2] では，ブルームフィルタを用い文字列属性と数値属性を対象としたプライバシ保護検索手法を提案した．この提案手法の特徴は属性毎ではなくタプル毎の索引構成である．また，文字列属性に対する完全一致，部分一致や数値属性に対する範囲検索をブルームフィルタによるマッチングという形で統一的に行うことで，複数の検索条件を一つの検索条件に変換している．

2.1　ブルームフィルタ索引

先行研究 [1][2] にて提案した手法におけるテーブルと問合せの変換例は以下の図 1 である．

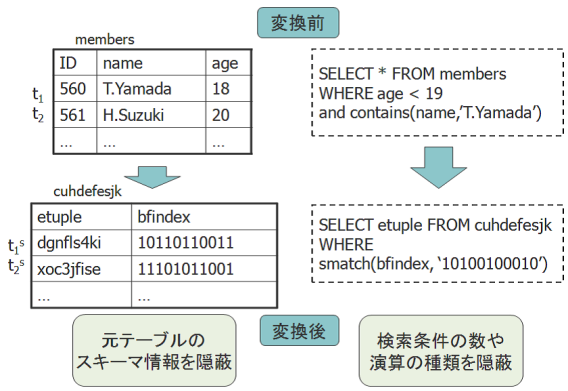

図 1: 変換前後のテーブルと問合せ例

このように変換することで，データベース管理者はテーブル構成や指定された検索条件を読み取ることができない．本稿では以降このテーブルを例に進めていく．

変換の基本手法であるタプルからブルームフィルタ索引を生成する流れを図 2 に示す．まず，タプル t から語の集合 $W_t = w_0, \cdots, w_n$ を生成する．各語 w_i は属性名と属性値からなる．属性値が文字列である場合は，部分一致用に単語毎や n-gram に従って分割し属性名と合わせて語とする．数値属性から語を生成する方法は 2.2 で述べる．これらの語に対して HMAC（鍵付きハッシュによるメッセージ認証関数）を複数（図 2 では 3 個としている）適用し，その値に基づいてブルームフィルタにビットを立てる．

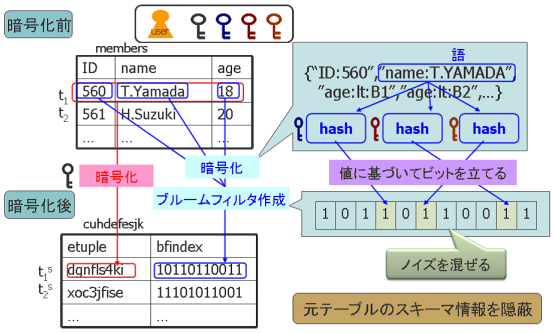

図 2: タプルの暗号化

問合せの変換に関しても同様で，WHERE 節に含まれる各問合せ条件からブルームフィルタを生成する．条件が論理演算子 AND で結合されている場合は，それらのブルームフィルタは論理和でまとめる．変換後の WHERE 節にある smatch 関数は，各タプルのビットマップである bfindex 属性値と問合せ用のビットマップをマッチングさせて真か偽かを返す関数として定義する．このようにして，テーブルのスキーマ情報や検索条件の数や演算の種類を隠蔽することができる．

2.2　数値からの語の生成

基本的には数値属性のドメインを複数のバケットに分割し，該当するバケットの名前を属性名と合わせて語にする．図 3 は 47,62,13,103,77 とそれぞれに対する語の集合を現したものである．全てのバケット B=B1,⋯,Ba に対して，バケットの上限が値 v より小さい場合「<属性名>:lt:<バケット名>」，バケットの下限が値 v より大きい場合「<属性名>:mt:<バケット名>」，それ以外の場合は「<属性名>:eq:<バケット名>」という語を追加する．

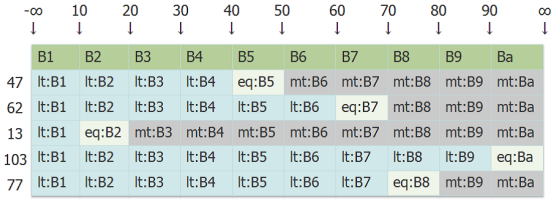

図 3: 数値属性のバケットによる表現

これを用いて値の代表を比較する場合，例えば 75 より大きな値を調べたい場合は，75 が含まれるバケット（B8）の左隣（B7）に注目し，「<属性名>:lt:B7」という語を持つタプルを探せば良い．逆に 35 より小さい値を探す場合には，右隣のバケット（B5）に注目し，「<属性名>:mt:B5」という語を持つタプルを探す．このようにして，数値の比較演算を文字列のマッチングと同様に扱う．

3　プライバシ保護検索システムの実装

本研究のシステムは，図 4 で示されるように主にクライアント側で動作する．

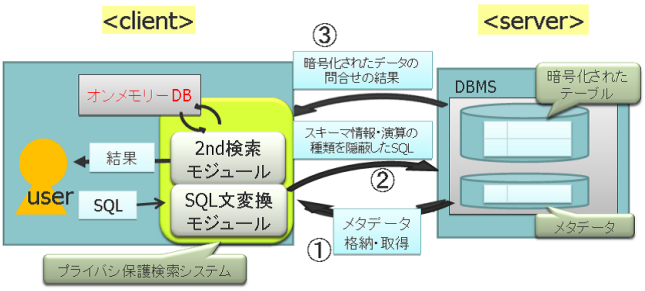

図 4: システム概要

ユーザが SQL 文を発行すると，本システムの SQL 文変換モジュールがそれを受け取り，安全に暗号化された SQL 文を用いてデータベースサーバとやり取りを行う．検索時のサーバから返る結果は，本システムの 2nd 検索モジュールが受け取り精製した解をユーザに返す，本研究では，データベースを扱う上でユーザが行う処理の内，テーブルの生成・データの挿入・データの検索の 3 つの処理に関して実装した．各処理を行う上で必要な実装項目に関して，3.1 で create 文の拡張, 3.2 でメタデータの格納, 3.3 で SQL 文変換と 2nd 検索モジュールについて述べる．

3.1　create 文の拡張

先行研究の暗号化手法を適用するために，SQL 文を拡張する必要がある．そこで，数値を語の集合に置き換えるために，各数値属性において値の取りうる範囲とバケット数を指定する ***backetize*** **[最小値, 最大値, 分割数]** というオプションを受け取れるように拡張した．(図 5(a))

3.2　メタデータの格納

データベースサーバに送信されるのは常に (etuple,bfindex) というペアの属性を持つ形なので，ユーザが入力したテーブル情報を安全に保存する必要がある．その為，テーブル情報を XML 形式の文字列に変換し，これを暗号化して変換テーブル名とともにデータベースに格納する．図 5(b) が (a) から得られる XML 形式のメタデータである．サーバにメタデータを格納することによって，別のユーザが同じデータベースを使うことが出来る．

```
create table members (
    id int ,
    name varchar(32),
    age int
    backetize[20,5,5]
) ;
```

(a) 拡張 create 文による記述例

```
<table>
  <tn>members</tn>
  <dl>
    <dmn>
      <name>id</name>
      <typ>int</typ>
    </dmn>
    <dmn>
      <name>name</name>
      <typ>varchar(32)</typ>
    </dmn>
    <dmn>
      <name>age</name>
      <typ>int</typ>
      <opt>backetize[5,20,5]</opt>
    </dmn>
  </dl>
</table>
```

(b) (a) のテーブルのメタデータ

図 5: 拡張 create とメタデータ

3.3　SQL 文変換と 2nd 検索モジュール

図 4 を用いて，ユーザの入力を受け取った時の SQL 文変換, 2nd 検索モジュールの処理手順について述べる．

＜テーブル生成＞

拡張 create 文を受け取ると，XML 形式のメタデータを生成し，データベースに格納する (①)．etuple と bfindex の属性を持つ安全な create 文を発行する (②)．

＜データの挿入＞

insert 文を受け取ると，まずメタデータを取得し，それを復号化する (①)．取得してきたテーブル情報に従い，語とブルームフィルタを生成する．またタプル全体を暗号化したものは insert 文の values 以下を暗号化したものとし，これらをデータベースサーバに格納する (②)．

＜データの検索＞

挿入と同様に，select 文を受け取るとまずメタデータを取得・復号化する (①)．取得してきたテーブル情報に従い，語とブルームフィルタを生成する．WHERE 節に bfindex と問合せ用ビットマップが引数の smatch 関数を含む select 文を発行する (②)．サーバ側では smatch 関数に従い該当する暗号化されたデータである etuple をクライアント側にある 2nd 検索モジュールに返し (③)，そこで得られた結果をユーザに返す．

この 2nd 検索モジュールを通して，誤検索したものを取り除いたり，数値属性の絞り込み検索を行ない解を精製する．

＜ 2nd 検索モジュール＞

受け取った etuple の復号化，insert 文の生成を行ない，メモリデータベースに一時的なテーブルを作成する．そのテーブルに対してユーザが発行したオリジナルの問合せを行ない，その結果をユーザに返す．

4　まとめと今後の課題

先行研究で提案された手法を用いた DaaS 環境におけるプライバシ保護検索システムを構築した．今後はプログラムを改良，機能を拡張し，実際に提供されている DaaS サービス上で動作させたいと考えている．

参考文献

[1] Watanabe C. and Arai Y.: “Privacy-Preserving Queries for a DAS model using Two-Phase Encrypted Bloomfilter,” *International Conference on Database Systems for Advanced Applications*, pp.491-495 (2009)

[2] 渡辺知恵美, 新井裕子: “DaaS におけるスキーマ情報と複合的検索条件を隠蔽したプライバシ保護検索法,” 情報処理学会研究報告, 2008-DBS-146, Vol.2008, pp.163-168 (2008)