様式 C-19

科学研究費補助金研究成果報告書

平成２３年６月１８日現在

機関番号：12611
研究種目：若手研究（Ｂ）
研究期間： 2009 ～ 2010
課題番号：21700099
研究課題名（和文）
クラウドコンピューティング環境におけるセキュアなアウトソーシングデータベース
研究課題名（英文） Secure Outsourcing Database in cloud computing environment
研究代表者 渡辺 知恵美 （WATANABE CHIEMI）
お茶の水女子大学・大学院人間文化創成科学研究科・講師
研究者番号：20362832

研究成果の概要（和文）：
クラウドコンピューティング環境にアウトソーシングデータベースシステムにおいて，データベース内部に格納されているプライバシ情報を一切管理者側に読み取られることなく検索する手法の提案を行い、その実装と検証を行った．特にデータのスキーマ情報を読み取られない方法としてブルームフィルタを用いた安全な索引 ShuffledBF を提案し，その安全性を確保するとともに，検索高速化のための拡張としてノイズ付与の試み，および部分的な ShuffleBF の適用（Semi-ShuffledBF）の提案および実装を行った．

研究成果の概要（英文）：
We proposed and implemented a new secure query processing methodology which don't steal sensitive data from the database administrator, in the case the data is stored in the database system on cloud computing environment. Especially, our proposed ShuffledBF, which can hide the schema information of the table, and it preserves the adversaries' inferring from the distribution of each attribute value. In addition, we investigated the following two improvements for speed up the query response time; (1) the perturbation of bloomfilter index withou shuffling (2) proposing Semi-ShuffledBF which is marged a shuffled bloomfilter and a non-shuffled bloomfilter .

交付決定額

（金額単位：円）

	直接経費	間接経費	合 計
2009 年度	2,200,000	660,000	2,860,000
2010 年度	1,100,000	330,000	1,430,000
総 計	3,300,000	990,000	4,290,000

研究分野：データベースシステム，プライバシ保護
科研費の分科・細目：情報学・データベース（ＤＢＭＳ）
キーワード：クラウドコンピューティング，アウトソーシングデータベース，プライバシ保護検索，ブルームフィルタ

１．研究開始当初の背景

近年 Database as a Service(DAS)がにわかに注目を集めている．DAS とはクラウドコンピューティング環境においてデータベース管理を請け負うサービスである．利用者はデータベースを設置・管理するための労力をかけることなく高性能なデータベース機能を利用することができるが，管理者がデータ所

有者と異なる第三者であるため，データ所有者が管理者に対して機密情報を守る必要性が生じてきた．そのための手法として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プライバシ保護検索手法についてこれまで多くの研究がなされてきたが，対象データのデータ型及び演算によって個別の手法が提案されており，実際に暗号化データベースに対する問合せを実現するためには，属性毎に別々の暗号化を施したり索引を作る必要があった．

２．研究の目的

そこで我々はデータベースのスキーマ情報および問合せ文に含まれる演算の種類を攻撃者に対して隠ぺいしたままで検索が可能なデータベース暗号化手法を提案し、それをベースにしたセキュアなアウトソーシングデータベースシステムの開発を行った。提案手法の概要を示すためにアウトソーシングデータベースに格納されるデータと発行される問合せの例を示す．であり，図１にて本提案手法におけるデータ格納および問合せ文の例を示す．
データはクライアントでタプル毎に暗号化され検索用の安全な索引を付与してサーバに格納する．問合せ文はサーバで実行される問合せ文に変換され，問合せ文の実行の結果該当するタプルの暗号化データがクライアントに返される．クライアントではそれらのタプルを復号化した後もう一度変換前の問合せ文を実行して正しい結果を得る．
図１に本提案手法におけるデータベース変換の例を示す．

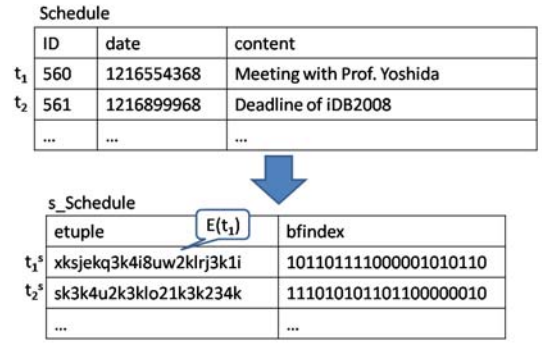


図１：元のテーブル（上）と暗号化後のテーブル

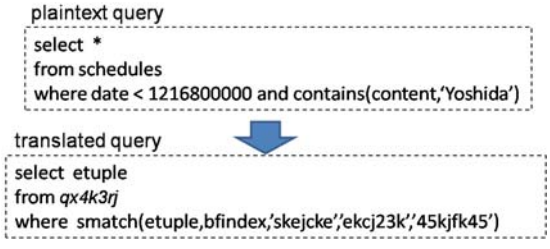


図２：ユーザが発行する問合せ文(上)とサーバに発行される問合せ

ID と日付(date)と内容(content)で構成されたテーブル Schedules が変換されて，サーバにはタプル t 全体を暗号化した t(etaple)と検索用の索引 t(bfindex)のみがアップロードされる．攻撃者はテーブル Schedules がどのような属性で構成されているかを推測することができない．図 2 は本提案手法による問合せ変換例である．date 属性に対する範囲検索条件や content 属性に対するテキスト検索が，索引 bfindex に対するマッチング関数 smatch に置き換えられている．そのため攻撃者は何の属性値を用いてどのような検索条件を指定したかを読み取ることができない．

３．研究の方法

(1) 基本手法-Non-ShuffledBF

本提案手法において，タプルからブルームフィルタ索引を生成する流れを図 3 に示す．まずタプル t から語の集合 Wt={w0,…,wn}を生成する．各語 wi 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を合わせて語とする．数値属性から語を生成する方法については 2.2.2 項にて述べる．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 3 では r 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．

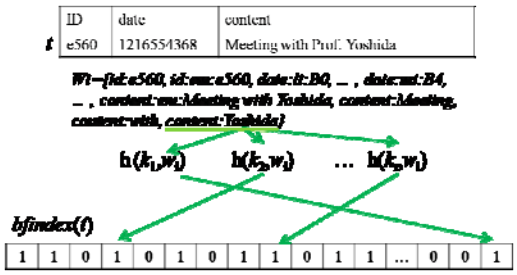
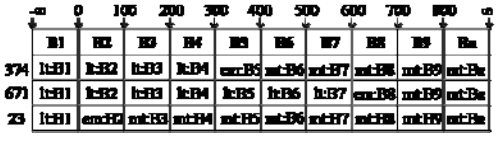


図３：タプル t から索引を生成する流れ

また我々は数値から語を生成する手法を提案した．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とする．図 4 は 374, 671, 23 とそれぞれに対する語の集合を表したものである．すべてのバケット 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 より大きな場合「<属性名>:mt:<バケット名>」それ以外の場合は「<属性名>:em:<バケット名>」という語を追



加する．

図 4：数値から語を求める

これを用いて値の代表比較をする場合，例えば 620 より大きな値を調べたい場合は，620 が含まれるバケット(B7)の左隣(B6)に注目し，<属性名>lt:B6 という語を持つタプルを探せばよい．逆に 420 より小さな値を探す場合には，右隣のバケット(B7)に注目し，<属性名>:mt:B7 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．

(2) ShuffledBF:
基本手法(Non-ShuffledBF)で索引を生成した場合，同じ語が含まれている複数のタプルはすべて同じ位置にビットが立っているためそこからデータの傾向などを推測される可能性がある．そこで ShuffleBF では基本手法のように一度ハッシュ関数を求めた後，タプルを暗号化した値（etuple）を鍵として再び HMAC を適用する．これにより，複数のタプルが同じ値を持っていた場合も 2 回目のハッシュ関数適用により異なる場所にビットが立ち，攻撃者はビットパターンから元データの特徴を推測することができない．この手法に関する問合せの手順を図 5 に示す．
まずクライアントで検索条件文から検索用の語を生成し，1 回目のハッシュ関数を適用しサーバに送る．サーバ側ではタプル毎に etuple を鍵にしてハッシュ関数を適用し，その値でブルームフィルタへのマッチングを行う．

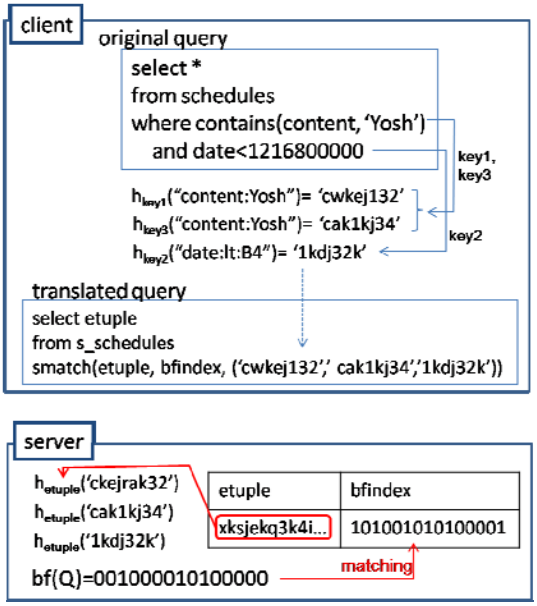


図 5：ShuffledBF による問合せの手順

(3) より高速な検索手法の検討

ShuffledBF ではサーバ側でタプル毎にクライアントから渡されたハッシュ値の数だけハッシュ関数を適用しなければならず検索時間がかかるという問題がある．Google App Engine でタプル数 1000，検索用のハッシュ値数 5，検索に対する正解率 60%の場合 0.4 秒かかることがわかった．その場合，タプル数が 20000 以上になるとタイムアウト(8 秒)せずに回答できる保証がなくなってしまう．そこで我々は以下の 2 種類の検索高速化の検討を行った．

(a) Non-ShuffledBF における攻撃可能性の検討とそこからのノイズ付与による安全性確保の検討

(b) Semi-ShuffledBF: ShuffleBF に一部 Non-ShuffledBF をマージした，安全性は多少劣るが高速な問合せ索引の提案

結論としては(a)の検討の結果，Non-ShuffledBF において数値属性の抽出が高確率で可能であることが判明し，ノイズを大量に付与しなければならないということから現実的でないため，Non-ShuffledBF からのノイズ付与は適切でないということが明らかになった．

(4) Semi-ShuffledBF
Semi-ShuffledBF は，ハッシュ関数を適用するタプル数を絞り込むため，ShuffledBF に加えて，Non-ShuffledBF による索引を付与する．問合せ時には，まず Non-ShuffledBF 索引による絞り込みの後 ShuffledBF 索引による絞り込みを行うことで，SFB 索引によるハッシュ関数の適用回数を減らす．
Non-ShuffledBF 索引では以下のハッシュ関数を適用する．

… (1)

ここで m はブルームフィルタのビット長，l は 2 以上の整数をとるパラメタとし，関数 gj は元テーブル R((A1,…,An)) における属性 Aj 毎に設定する．ShuffledBF ではハッシュ値に対してブルームフィルタのビット長である m の剰余を求めていたのに対し，Non-ShuffledBF では⌈m/l⌉の剰余を求めている．l の値を大きくすることで Non-ShuffledBF による擬陽性をあえて高くし，ブルームフィルタによる元データの推測が難しくする．関数 gj は，属性 Aj ごとにビットの立つ位置を定めるための関数で，単一のブルームフィルタ内でのビットを立てる位置の重複を防ぐ．また，Non-ShuffledBF と ShuffledBF を別々に用意するのではなく，これらを合成し一つのブルームフィルタとすることによって Non-ShuffledBF における元データの推測をより難しくすること

ができる． Semi-ShuffledBF の問合せの方法は以下の 2 段階で行われる．

（a） Non-ShuffledBF で検索

（b） ShuffledBF で検索

まず Non-ShuffledBF を各タプルに適用し，問合せ条件に該当すると判断したもののみ ShuffledBF を適用する．

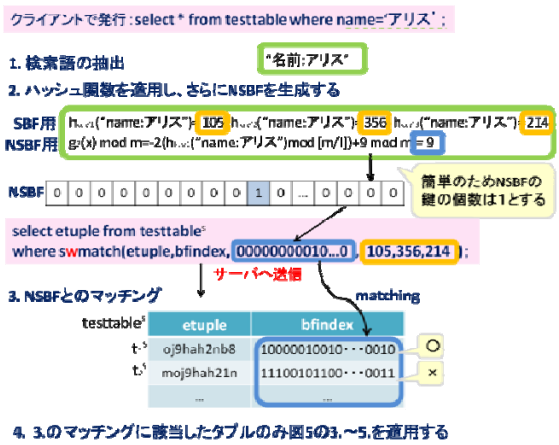


図 6：Semi-ShuffledBF による問合せ例

まずクライアントで検索条件文から検索用の語を生成し，Non-ShuffledBF 用ハッシュ関数，ShuffledBF 用ハッシュ関数を各々適用しサーバに送る．サーバ側ではまず Non-ShuffledBF 用ハッシュ関数から得られた値（NSBF）でマッチングを行い，そこから得られたものにだけ ShuffledBF 用ハッシュ関数から得られたブルームフィルタ索引タプル毎に etuple を鍵にしてハッシュ関数を適用し，その値（SBF）でブルームフィルタへのマッチングを行う．

図 7 に Non-ShuffledBF(NSBF)，ShuffledBF(SBF)，Semi-ShuffledBF(SSBF)における問合せ処理時間の比較を示す．

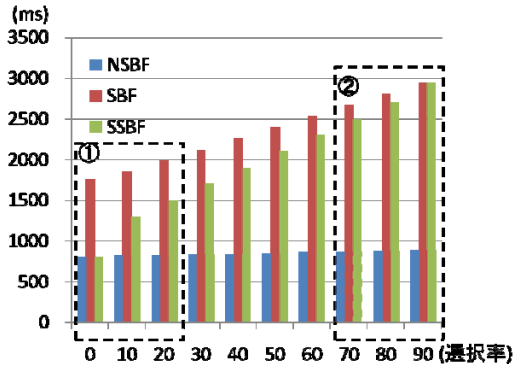


図 7：問合せ処理時間の比較

この結果選択率が 20%程度までは十分に問合せ処理時間が短縮できていることが分かった．実際の問合せでは複数の問合せ条件の適用により選択率は十分に低くなると考えられるため，選択率が低いケースでの結果が表れていることで成果が表れていると言える．ただしその改善率は劇的なものではなく，さらなる改善が必要と考えられる．

４．研究成果

クラウドコンピューティング環境において，データの内容及びスキーマ情報を隠ぺいまま問合せが可能な索引として ShuffledBF 提案を行い，その安全性の検証を行った後，問合せ高速化の改良として Semi-ShuffledBF を提案した．その成果は国内の研究会および国際会議にて発表し評価を得ている．国内での発表は 2 件（うち 1 件は査読付き研究会）国際会議での発表は 3 件，論文誌掲載 1 篇と多くの結果を残している．

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 1 件）

①渡辺知恵美，新井裕子，天笠俊之，ブルームフィルタを用いたプライバシ保護検索における攻撃モデルとデータ撹乱法の一検討，日本データベース学会論文誌，査読あり，Vol.8, No.1，2009，pp.113-118.

〔学会発表〕（計 5 件）

①Chiemi Watanabe, Yuko Arai: Privacy Preserving Queries for a DAS model using Two-Phase Encrypted Bloomfilter, The 14th International Conferences on Database Systems for Advanced Applications, 2009 年 4 月 22 日，Queensland University, Brisbane, Australia.

② 新井裕子，渡辺知恵美，プライバシ保護検索におけるクエリログを利用した攻撃とその防御に関する一手法，Web とデータベースに関するフォーラム（WebDB Forum 2009），査読あり，2009 年 11 月 20 日，慶應義塾大学，神奈川，日本

③Yuko Arai, Chiemi Watanabe, Query Log Perturbation Method for Privacy Preserving Query, The 4th International Conference on Ubiquitous Information Management and Communication, 2010 年 1 月 15 日, Suwon, Korea

④金子静花，渡辺知恵美，Semi-ShuffledBF：ブルームフィルタを用いた安全かつより高速なプライバシ保護検索手法の提案，第 3 回データ工学と情報マネジメントに関するフォーラム(DEIM 2011)，2011 年 2 月 28 日，ラフォーレ修善寺，静岡県伊豆市修善寺町，

⑤ Shizuka Kaneko, Chiemi Watanabe, Semi-ShuffledBF: Performance Improvement of a Privacy-Preserving Query Method for a DaaS Model Using a Bloom filter, The 2011 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'11), 2011年7月18日, モンテカルロホテル, ラスベガス, アメリカ

６．研究組織

(1)研究代表者

渡辺 知恵美 (WATANABE CHIEMI)

お茶の水女子大学・大学院人間文化創成科学研究科・講師

研究者番号：20362832

(2)研究分担者

無し

(3)連携研究者

無し