Encrypted DBMS のための安全かつ高速な多属性索引の諸検討

理学専攻・情報科学コース　金子静花

1　はじめに

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の管理運用を請け負う Database as a Service (DBaaS) が注目を集めている．このような環境でユーザがデータ管理者から機密情報を守るための手段として，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暗号化データベースシステム (EDBMS) [1] が多く研究されてきた．EDBMS においては，暗号化されたデータに付与される検索用索引を，どのように安全かつ検索性能を落とさず生成するかが問題となる．

先行研究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 EDBMS[2] を提案してきた．本稿では，攻撃モデルとその攻撃に対する安全性の基準 $\gamma - FS^A$ を定め，$\gamma - FS^A$ を満たす，敵の持つ統計情報からの攻撃に堅固な検索用索引を提案する．

2　基本手法

EDBMS は，サーバプロバイダを通すデータをすべて暗号化することでサーバプロバイダからもデータの機密情報を守りつつ問合せを行うことのできるデータベース管理システムである．暗号化したデータとその索引をサーバに預けることで，データを暗号化したまま問合せを施すことができる．一般的な EDBMS のモデルを図 1 に示す．

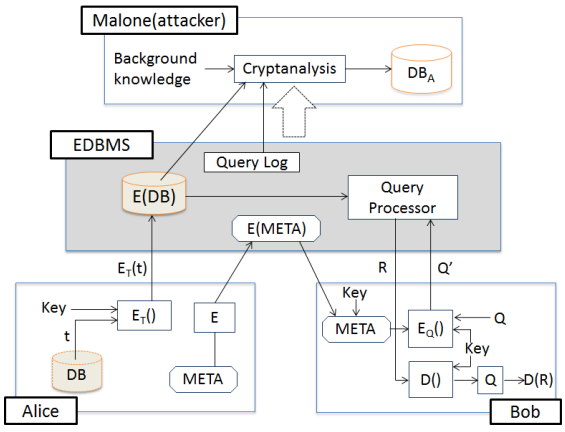

図 1: EDBMS のモデル

2.1　単属性索引（Single Attribute Index）

EDBMS では，元の値を暗号化したものに紐づく検索用の安全な検索用索引を用いて問合せを行っている．既存手法はタプルを列 (属性) 毎に暗号化した単属性索引を検索用索引として使用している．元の値とその単属性索引との間には 1 対 1 の対応関係がある．つまり，同一な元の値からは同一の単属性索引が生成されるため，このように生成された検索用索引は元リレーションの統計情報に関する大きな手がかりを含んでいる．

3　先行研究

3.1　多属性索引（Multi Attribute Index）

我々は先行研究 [2] にて，ブルームフィルタを用いた DBaaS におけるスキーマ情報と複合的な検索条件を隠ぺいしたプライバシ保護検索手法を提案した．この手法では，属性毎に検索用索引を生成するのではなく，タプル内の複数の属性を 1 つの多属性索引として生成することで安全性を担保しようと試みた．しかしながら素の多属性索引に対してもまた，多属性索引の値から元の値推測の可能性を否定できない．

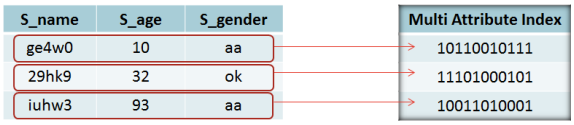

図 2: 多属性索引の生成

そこで先行研究では，ブルームフィルタを更にもう一段階適用することで多属性索引の安全性を担保した ShuffledBF を提案している．

ShuffledBF の生成方法を図 3 に示す．元テーブルの属性と語の集合“ ID:330 ”,“ 名前:Alice ”, “ 病気: 骨折 ”に対し，各々複数のハッシュ関数を適用したのち，タプル全体を暗号化したものである etuple をキーにしたハッシュ関数を更に適用（シャッフル）した結果を多属性索引として生成する．このように，タプル毎に生成される etuple をキーとして 2 段階目のハッシュ関数を適用しているので，安全性は完全に保たれている．

ShuffledBF の問合せ方法を図 3 に示す．索引生成時と同様に，検索したい元データの属性と語の集合“ 名前＝ Alice ”に対し，1 段階目の複数のハッシュ関数を適用した結果 QMAI ’を検索条件としてサーバへ送信する．(図 3 中 (1)) サーバ側では，受け取った QMAI ’にタプル全体を暗号化したものである etuple をキーにしたハッシュ関数を各タプル毎に適用（シャッフル）して QMAI ”とし，多属性索引と一致するかどうかを調べる．(図 3 中 (2))

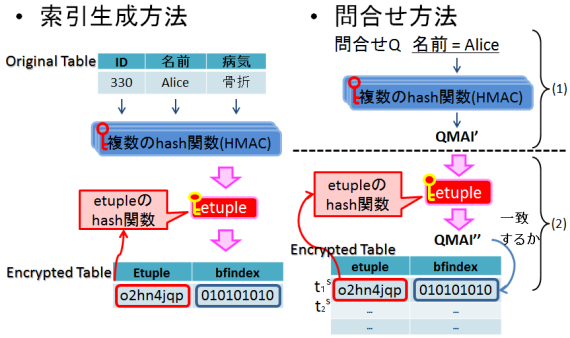

図 3: ShuffledBF の生成・問合せ方法

しかしながら，ShuffledBF は問合せの際タプル毎にハッシュ関数を適用しているため性能が低い．そこで，

安全かつ高速な多属性索引の諸検討を行った．

4 Encrypted DBMS のための安全かつ高速な多属性索引の諸検討

4.1 Semi-ShuffledBF

我々は，多属性索引と ShuffledBF(SBF) とを組み合わせることで ShuffledBF の性能を向上させた Semi-ShuffledBF(SSBF)[3] を提案した．

Semi-ShuffledBF の生成・問合せ方法を図 4 に示す．Semi-ShuffledBF の問合せでは，まず通常の多属性索引 (MAI) で検索を行った後一致したタプルに対してのみ etuple での 2 段階目のハッシュ関数適用（シャッフル）を行い，一致するかどうかを調べる．

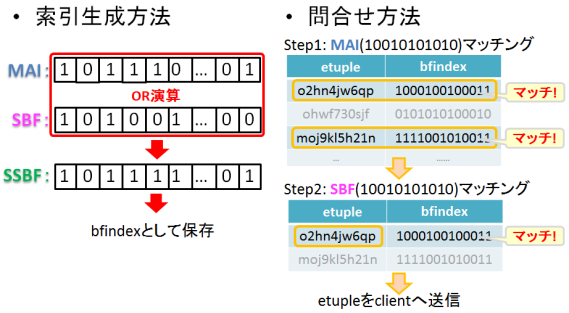

図 4: Semi-ShuffledBF の生成・問合せ方法

このようにして安全性を保ちつつ ShuffledBF の処理性能を改善することに成功したが，暗号化なしの問合せと比べ，依然最大 20 倍の性能劣化がある．

4.2 *PerturbedBF*

我々は多属性索引の原点に立ち返り，ノイズを加えることで性能を損なわずに安全性を保障できる EDBMS を考えた．安全性を保障するため，まず敵の攻撃モデル「攻撃者は，元テーブルの属性の統計情報を保持しており検索用索引の特徴と攻撃者の持つ統計情報とを照合して攻撃を行う」を定義した．次に，攻撃モデルに対する安全性の基準 ($\gamma - FS^A$) を定めた．その基準とは「検索用索引が少なくとも γ の特徴集合を持つ場合，その検索用索引は $\gamma - FS^A$ を満たす」である．以下 4 つのノイズ混入戦略を組み合わせて攻撃への攪乱を行い，$\gamma - FS^A$ を満たす検索索引を提案した．

$\gamma = 3$ とし、Emp(id, gender, country, blood, name) のデータ 500 レコードを用いた検証結果を図 5 に示す．

		攪乱なし	攪乱あり		
適応手法		手法1	手法2	手法3	手法4
適用戦略		Ⅰ	Ⅰ、Ⅱ	Ⅰ、Ⅱ、Ⅲ	Ⅰ、Ⅱ、Ⅳ
FS^A	gender	1	3	3	12
	blood	1	61	>100000	16
	country	>100000	>100000	>100000	>100000

図 5: 検証結果

- 戦略 Ⅰ ：頻度の類似した属性集合で多属性索引を生成

頻度が類似する属性値を多数多属性索引に含むことで，攻撃者の推測する特徴候補を増やすことができ，$\gamma - FS^A$ を増加させることができる．

- 戦略 Ⅱ ：偽属性の混入

各属性の頻度情報と頻度情報が全く等しい γ-1 の偽属性を生成し，多属性索引に含める．この場合，偽属性が $\gamma - 1$ 個含まれているので，$\gamma - FS^A$ を必ず満たすことができる．

- 戦略 Ⅲ ：誤検出率の調整

多属性索引の誤検出を増加させることで，実際にはそのタプルに存在しない値の特徴が発生するため，$\gamma - FS^A$ を増加させることができる．しかしながら誤検出率の増加は問合せコストの増加をも招くため，注意が必要である．

- 戦略 Ⅳ ：属性値の分割

頻度の高い属性値は特定しやすく，さらに類似した属性集合を見つけることが難しい．そこで，そのような属性を 2 つ以上の属性集合に分割し，属性集合の頻度を下げることとする．ただし，この場合複数の属性集合に対して問合せを行わなければならなくなるため，検索コストへの影響も考慮しなければならない．

PerturbedBF はシャッフル処理を行わないため，暗号化なしの問合せと比べて 2 倍程度の性能に抑えることができる．

5 まとめと今後の課題

本稿では，Encrypted DBMS のための安全かつ高速な多属性索引の諸検討として Semi-ShuffledBF と PerturbedBF について考察した．

$\gamma - FS^A$ を満たす性能の良い EDBMS の提案を行ったが，$\gamma - FS^A$ には攻撃者の所有する情報に属性間の相関関係が含まれていない．そのため，属性間の関係に相関ルールマイニングを用いることで元の値を推測されてしまう恐れがある．今後は実際に相関ルールマイニングを用いて安全性の検討を行い，新たな安全性の基準とそれを満たす手法を考える．

相関ルールの抽出には，たった 9 タプル 3 属性に対して 190 種類の相関ルールが存在するなど，莫大な量の組み合せを計算しなければならず，組み合せ爆発を起こすと考えられる．この組み合せを求めることは NP であるため，NP 完全になることが予想される．

さらに，選択演算以外にも対応させ，実際に EDBMS として機能するシステムを提案することが今後の課題である．

参考文献

[1] H. Hacigumus, B. Iyer, C. Li, and S. Mehrotra. : “Executing SQL over Encrypted Data in the Database-Service-Provider Model,” *Proceeding of the ACM SIGMOD International Conference on Management of Data*,(2002)

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

[3] Kaneko S., Watanabe C. and Amagasa T.: “Semi-ShuffledBF: Performance Improvement of a Privacy-Preserving Query Method for a DaaS Model Using a Bloom filterh,” *PDPTA, WorldComp*,(2011)