一対多マッピングに基づく リレーショナルデータベースの プライバシー保護検索

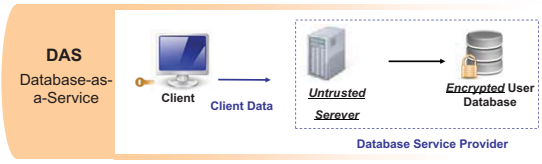
天笠俊之
筑波大学システム情報系

第81回知的コミュニティ基盤研究センター研究談話会
2011年11月17日 (木) 14時～15時

データベースにおけるセキュリティ機構

- 既存の主な手法
  - ビュー
  - アクセス制御
    - ユーザ認証
    - RBAC
    - ACL
    - …
  - 暗号化
- クラウドにおけるデータベース
  - Database as a Service (DAS)

DASモデルにおける プライバシー保護検索

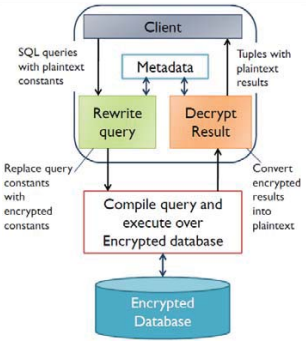


[Hacigumus et al., ICDE 2002]

- データベース
  - 暗号化した上でクラウド上のDSPに格納.
- 問合せ処理
  - 復号化にはデータベース全体のダウンロードが必要.
  - データベースのプライバシーを保護しつつ, DSPの計算機資源をなるべく利用した検索を実現したい.

想定される攻撃モデルとフレームワーク

- ストレージ上のデータは攻撃者に読まれる
  - 暗号化
- 主記憶上のデータは読み取れない
- 攻撃者は, データベースの平文と統計量を知っている可能性あり.

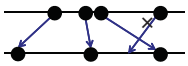


関連研究 データベースのプライバシー保護検索

- **数値データ**
  - Bucketization/Partitioning [Hacigumus, SIGMOD'02]
  - Order preserving encryption scheme [Agrawal, SIGMOD'04]
  - Structure preserving database encryption scheme [Yuval, SDM'04]
  - Anti-tamper database [Chung, ICDE WS'06]
- テキストデータ
  - [Boneh, EUROCRYPT'04], [Dong, IFIP WS 11.3 '08]
- XMLデータ
  - Efficient Secure Query Evaluation over Encrypted XML Databases [Wang, VLDB'06]

順序保存暗号化法

- Order preserving encryption for numeric data
  - Agrawal et al. SIGMOD 2004
- 考え方



$X_1 < X_2 < X_3 < .... < X_n$
➔$E^k(X_1) < E^k(X_2) < E^k(X_3) < ... < E^k(X_n)$
- 利点
  - 一致検索, 範囲検索などが, 暗号文上で処理可能.
  - B木などの索引をそのまま利用可能.
- 欠点
  - 統計的暗号攻撃に弱い.

一対一マッピングに基づく暗号化

- 一対一マッピングに基づく暗号化
  $X_1 = X_2 = … = X_n$
  ➔ $E^k(X_1) = E^k(X_2) = … = E^k(X_n)$
- 一対一マッピングに基づく暗号化は攻撃に対して脆弱
  - 暗号文の統計的性質を利用した攻撃.
  - カテゴリ属性の場合, 特に問題.
    - 性別
    - 国籍
    - 販売店コード

データベース暗号化の例

employee

E_ID	Salary	gender
111	100	1
222	200	1
333	350	1
444	200	1
555	200	2

E_ID	P_ID
111	31
222	33
333	33
444	34
222	34

emp_prj

project

P_ID	P_Name
31	projectAS
33	projectCZ
34	projectBX
35	projectMM

Plaintext database

emp^E

E_ID^E	$Salary^E$	$gender^E$
35634	4576456	67653
87685	586576	67653
32422	6756765	67653
64777	586576	67653
77547	586576	564564

E_ID^E	P_ID^E
35634	21133
87685	65874
32422	65874
64777	78554
87685	78554

emp_prj^E

$proj^E$

P_ID	P_Name
21133	$projectAS^E$
65874	$projectCZ^E$
78554	$projectBX^E$
96241	$projectMM^E$

Encrypted database

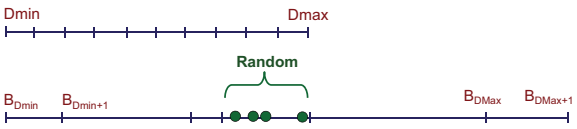
研究の目的

- 統計的暗号攻撃に対して頑健な，プライバシー保護データベース暗号化・問合せ法
- アプローチ
  - 順序保存暗号化法に基づく．
  - 一つの平文に対し，複数の暗号文を対応付ける．
    - ある平文に対し，区間内の値をランダムに割り当てる．
    - 統計的暗号攻撃に対応．
  - なるべく多くの関係代数演算をサポート．
    - 従来研究は，複数テーブルのジョインには非対応．
  - 実用可能なレベルでの実行速度．

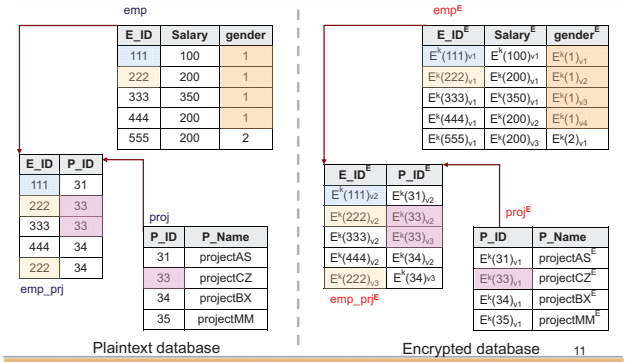
アイデア：一対多マッピング

Multivalued Order Preserving Encryption Scheme (MV-OPES)

- 順序保存
  - X1 < X2 < X3 < … < Xn
  - Ek(X1) < Ek(X2) < Ek(X3) < …. < Ek(Xn)
- 一対多マッピング
  - V1 = V2
  - 高い確率で Ek(V1) ≠ Ek(V2)



MV-OPESによるデータベース暗号化



MV-OPESによる暗号化

- バケツ境界の生成
  - サイズの異なるバケツを，ドメインのサイズだけ生成．
    - 初期値の選択
    - 増加／減少関数
- 乱数生成
  - 平文に対応するバケツを選択．
  - バケツ内の値をランダムに選択．

バケツ境界の生成

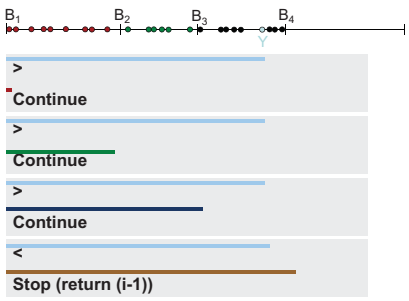
- ドメイン $[D_{min}, D_{max}]$
- 初期値選択
  - $B_{initial} = Enc^K(initial)$, $D_{min} <= initial <= D_{max}$
- 増加／減少関数
  - IS: バケツのベースサイズ，DP: 隣接バケツサイズ比
  - $B_i = B_{i+1} - [Enc^K(IS) + Enc^K(IS) * DP * R_i]$
    $D_{min} <= i < initial$
  - $B_i = B_{i-1} + [Enc^K(IS) + Enc^K(IS) * DP * R_i]$
    $initial < i <= (D_{max+1})$

暗号化

- 平文に対応するバケツ内の値をランダムに選択．
  - 一様分布
  - 正規分布
  - …
- 検索性能に与える影響
  - 隣接バケツとの境界付近に暗号文が分布すると，検索時に擬陽性の結果として報告される可能性あり．
  - 検索結果の精製が必要．

復号化

•暗号文の含まれるバケツを特定．

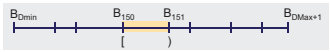


関係演算の実現

- 関係演算を，DASデータベースサーバの計算資源を利用しながら処理．
  - なるべく多くの計算をサーバ側で処理．
- いくつかの演算については，後処理が必要．
  - 復号化の後に，擬陽性の結果を取り除く必要あり．
- サポートする演算
  - **選択，結合，**射影，ソート，グルーピング，集約，集合差，集合和

関係演算の実現:選択演算

- **選択条件**
  - **Attribute θ Value**
- Attribute = Value
  - 書き換え前: A = v
  - 書き換え後: A^E between B_v and $(B_{v+1} - 1)$
- 例
  - A=150 → A^E between B_{150} and $(B_{151}-1)$

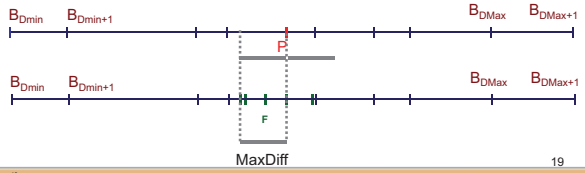


関係演算の実現:選択演算

- **関係演算の書き換え**
  - $\sigma_{A\ \theta\ v}(R)$
  - $\sigma_C(R) = D(\sigma^E(R^E))$
- **例**
  - $\sigma_{e\_id=150}(emp) = D(\sigma^E_{e\_id^E \text{ between B(150) and B(151)-1}}(emp^E))$

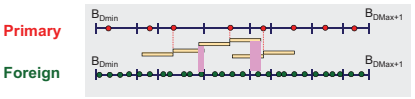
関係演算の実現:結合演算

- **結合条件**
  - **Attribute θ Attribute**
  - 異なる値にマップされているので，等号は使えない．
  - バケツ境界も利用不可．
- MaxDiffの導入
  - 暗号文とバケツ境界との最大値を保存して利用．



関係演算の実現:結合演算

- **結合条件**
  - **Attribute1 = Attribute2**
  - P = F,
  - F^E between $(P^E\text{-MaxDiff})$ and $(P^E\text{+Maxdiff})$
  - 擬陽性の結果を含む．



関係演算の実現:結合演算

- **関係演算の書き換え**

$R \bowtie_c T = \sigma_c(D(R^E \bowtie^E_{C^E} T^E))$

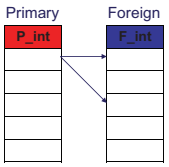
- 等結合
  - C: P = F,
  - C^E: F^E between $(P^E\text{-MaxDiff})$ and $(P^E\text{+Maxdiff})$
- 例

$emp \bowtie_{emp.e\_id=emp\_proj.e\_id} emp\_proj =$

$\sigma_{emp.e\_id=emp\_proj.e\_id}(D(emp^E \bowtie^E_{emp\_proj.e\_id^E \text{ between } (emp.e\_id^E\text{-MaxDiff}) \text{ and } (emp.e\_id^E\text{+Maxdiff})} emp\_proj^E))$

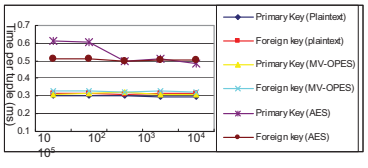
評価実験

- **提案手法の有効性を検証**
  - **暗号化**
  - **結合演算**
- データセット
  - ドメインのサイズ
    - 10, 100, 1000, 10000, 100000
  - テーブルサイズ
    - 100000 tuples.
  - DP
    - 5%, 15%, 25%, 35%, 45%, 55%



実験結果1

- **暗号化**

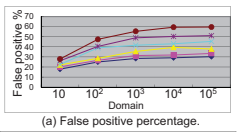


Time per tuple (in ms) required to insert tuples.

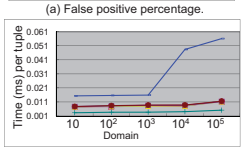
•提案手法は，平文とほぼ同等の性能．
•AESより高速．

実験結果2

- **等結合の結果**



(a) False positive percentage.

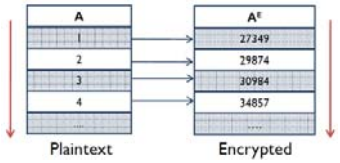


(b) Query execution time in server side.

(c) Decryption and filtering cost in client side.

これで十分？

- 攻撃者が平文・統計量を知っている場合．



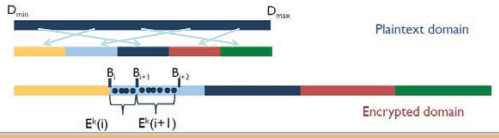
- 最大値，最小値と平文の組合せから，部分的に値が読まれる．

問題

- (MV-)OPESでは，値の大域的な順序を保存．
  - 平文＋統計量を使った攻撃に弱い．
- 要求
  - 大域的な順序を保存しない．
  - (MV-)OPESのような処理を行いたい．
- アプローチ
  - 部分的に順序を保存．
  - Multi-Valued Partial Order Preserving Scheme (MV-POPES)

MV-POPES

- ドメインをパーティション化，順序入れ替え
- 一対多マッピング
  - V1 = V2
  - 高い確率で Ek(V1) ≠ Ek(V2)
- パーティション内では順序を保存
  - X1 < X2 < X3 < … < Xn
  - Ek(X1) < Ek(X2) < Ek(X3) < …. < Ek(Xn)



メタデータ

- パーティション間の順序関係を記録．
- 問合せ書き換えに利用．

PID	EPID	PREV	F	L	NEXT
1	3	80	1	20	81
2	1	0	21	40	61
3	5	100	41	60	101
4	2	40	61	80	1
5	4	20	81	100	41

問合せの書き換え

- MV-OPESとは異なり，パーティションの順序を考慮する必要あり．
  - Attribute < Value (A < v)

select A from R where A<45 → select A^E from R^E where C^E

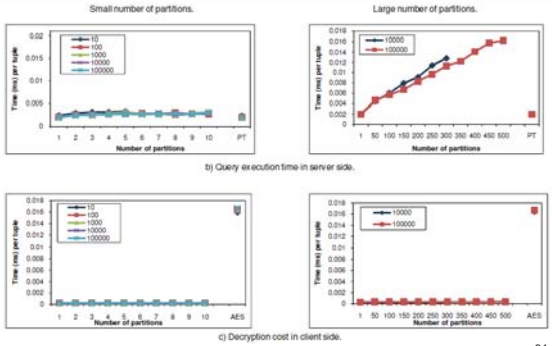
PID	EPID	PREV	F	L	NEXT
1	3	80	1	20	81
2	1	0	21	40	61
3	5	100	41	60	101
4	2	40	61	80	1
5	4	20	81	100	41

  - A < 45
  - C^E : $(A^E < B_{45} \wedge A^E \geq B_{41}) \vee (A^E \geq B_{21} \wedge A^E < B_{61}) \vee (A^E \geq B_1 \wedge A^E < B_{81})$

セキュリティ分析

	Full OPE	MV-POPES
Mapping Probability (μ)	$\eta = 1$ $\mu = 1$	k distinct PT, n distinct CT $\eta = \binom{n-1}{k-1} \sum_{i=1}^{k} \binom{k-1}{i-1} i!$ (Possible Frequencies; Possible # of partitions, size and order) $\mu = 1/\eta$ μ is significantly small even for small k, n
Similarity Degree of Statistics (Δ):	High	•Different order on PT and CT. •Max, Min are different on PT and CT. •Different frequencies on PT and CT. Δ is significantly small

実験（範囲問合せ）



まとめと今後の課題

- まとめ
  - 一対多マッピングに基づく順序保存暗号化方式MV-OPES
    - 多数の関係演算をサポート
    - 実験による評価
  - 部分的な順序を保存したMV-POPES
    - 多数の関係演算をサポート
    - 実験による評価
    - パーティションのランダム化方法を工夫することで，性能の劣化を防ぐ
- 今後の課題
  - MV-POPESの詳細な評価
  - 他のデータタイプのサポート

発表論文

- Hasan Kadhem, Toshiyuki Amagasa, and Hiroyuki Kitagawa, “A Secure and Efficient Order Preserving Encryption Scheme for Relational Databases,” Int’l Conf. on Knowledge Management and Information Sharing (KMIS 2010), Valencia, Spain, October 25-28, 2010. [Best Student Paper Award]
- Hasan Kadhem, Toshiyuki Amagasa, Hiroyuki Kitagawa, “A Secure and Efficient Order Preserving Encryption Scheme for Relational Databases“, iDBワークショップ2010, 東京, 2010.
- Hasan Kadhem, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "MV-OPES: Multivalued-Order Preserving Encryption Scheme: A Novel Scheme for Encrypting Integer Value to Many Different Values," IEICE Trans. Info. & Syst., Vol. E93-D, No. 9, pp.2520-2533, Sept. 2010.
- Hasan Kadhem, Toshiyuki Amagasa, Hiroyuki Kitagawa, "An Encryption Scheme to Prevent Statistical Attacks in the DAS Model", 第2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 2010), B5-5, 2010年2月28日～3月2日.