様式 C-19

科学研究費補助金研究成果報告書

平成２３年５月１９日現在

機関番号：１４４０１
研究種目：基盤研究（C）
研究期間：2008～2010
課題番号：20500092
研究課題名（和文）　問合せ解像度に基づくデータベースの静的安全性確保に関する研究

研究課題名（英文）　A study on static security of databases based on query resolution

研究代表者
石原　靖哲（ISHIHARA　YASUNORI）
大阪大学・大学院情報科学研究科・准教授
研究者番号：００２６３４３４

研究成果の概要（和文）：本研究では，まず，データベースへの推論攻撃に対して，「問合せ解像度の高低関係」という概念に基づく静的な安全性定義を提案した．推論攻撃とは，ユーザが，実行を許可された問合せのみを用いて，許可されていない問合せの実行結果を推論することをいう．本研究ではさらに，連言問合せを対象とし，問合せ解像度に関連するいくつかの性質を示した．また，ある前提条件のもとで，問合せ間の解像度高低関係が成立するための判定可能な必要十分条件を与えた．

研究成果の概要（英文）： In this study, we first proposed static security definitions against inference attacks on databases, based on a concept of “query resolution”. Inference attacks mean that an attacker tries to infer the execution result of a query unauthorized to the attacker from the execution results of queries authorized to the attacker. Then, we focused on conjunctive queries and showed some properties of query resolution. Moreover, under some assumptions, we proposed decidable necessary and sufficient conditions for existence of resolution relation between two conjunctive queries.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2008 年度	1,200,000	360,000	1,560,000
2009 年度	1,000,000	300,000	1,300,000
2010 年度	800,000	240,000	1,040,000
年度			
年度			
総　計	3,000,000	900,000	3,900,000

研究分野：総合領域
科研費の分科・細目：情報学・メディア情報学・データベース
キーワード：データベースセキュリティ，安全性検証

１．研究開始当初の背景

データベースのセキュリティを達成する上で重要な課題の一つに，推論攻撃に対する安全性の確保がある．推論攻撃とは，ユーザが，実行を許可された問合せのみを用いて，許可されていない問合せの実行結果を推論して得たり，実行結果の候補を絞り込んだりすることをいう．もし，攻撃対象の機密データ（を引き出す問合せ）があるユーザによって推論攻撃され得ることが事前にわかったならば，許可された問合せの一部やそのユーザに対するアクセス権そのものを修正することにより，その攻撃を未然に防ぐことができる．

本研究の開始当初，推論攻撃に対する安全性検証に関する研究のほとんどは，「動的な

安全性」を扱っていた．すなわち，データベースインスタンス D と，許可された問合せ $q_1,\cdots, q_n$，および Dから機密情報を引き出す問合せ q_{sec} が与えられたときに，攻撃者が $q_1,\cdots, q_n$の定義および問合せ結果 $q_1(D),\cdots, q_n(D)$を用いて $q_{sec}(D)$の値を特定できるか（あるいは絞りこめるか）を判定するための研究である．しかし，動的安全性の検証法には大きな弱点がある．それは，別途逐次的な検証法が与えられていない限り，データベースインスタンス Dが更新される度に，安全性検証を最初からやり直す必要があるという点である．

一方，静的安全性検証では，データベースのスキーマ Sと,許可された問合せ $q_1,\cdots, q_n$,およびデータベースから機密情報を引き出す問合せ q_{sec} が入力として与えられる．そして，典型的には，「スキーマ S にしたがうあるインスタンス Dが存在して,攻撃者が $q_1,\cdots, q_n$ の定義および問合せ結果 $q_1(D),\cdots, q_n(D)$ を用いて $q_{sec}(D)$の値を特定できるか（あるいは絞りこめるか)」を判定する．静的安全性を用いる利点としては，インスタンスを入力としないため比較的短時間で安全性検証が行えることが期待できる点，データベーススキーマに従う範囲であればインスタンスをどのように更新しても検証結果が有効である点などが挙げられる．しかし，このように重要な研究課題であるにも関わらず，静的安全性はその問題構造の複雑さのためか，十分に研究されてきたとはいえなかった．

２．研究の目的

本研究では，関係データベースや XML データベースを対象とし，問合せの解像度という概念を導入して，静的安全性を判定する技術および静的安全性を確保する手法の開発を目指す．ここで，問合せ qが q'よりも解像度が高いとは，（ある範囲の）任意のデータベースインスタンス D, D'に対して，$q'(D)\neq q'(D')$ならば $q(D)\neq q(D')$であることを意味する．すなわち，直観的には，問合せ q'が Dと D'の違いを検出できるなら，問合せ qもそれらの違いを検出できることをいう．一般に，解像度の高い問合せを許可することは，機密情報が特定されるリスクを高くする．

本研究では以下の 3 点を達成することを目的とする．

(1) 問合せの解像度に関する基本的な性質を調査する．具体的には，関係データベースや XML データベースなどさまざまな問題設定のもとで，与えられた 2 つの問合せ q, q'について，q の解像度が q'より高いための十分条件を求めたり，q の解像度が q'より高いかどうかを判定する問題の計算複雑さを求める．また，問合せの包含関係判定問題など，他の問合せ静的解析問題との関連についても調査する．
(2) 関数従属性や多値従属性などさまざまな制約をもつ関係データベースや XML データベースを対象とし，解像度の概念およびその基本的な性質を用いて，データベースが静的に安全であること（の十分条件）を判定するアルゴリズムを提案する．
(3) 静的に安全ではない（かもしれない）データベースに対し，許可された問合せの解像度を適切に下げることにより，静的安全性を確保する手法を提案する．

３．研究の方法

本研究では，まず目的(2)に関し，問合せ解像度という概念を用いた静的安全性定義が実用上どのような意義をもつのかを検討した（学会発表の⑤と⑥）．次いで，目的(1)に関し，関係データベースにおける連言問合せというクラスを対象として，解像度の高低関係が成立するための判定可能な必要十分条件について検討した（学会発表の①）．また，これらと並行して，XML データベースへの推論攻撃に対する（静的ではない）安全性検証(雑誌論文の①, 学会発表の②)や, XPath 問合せ言語の充足可能性判定（解像度の高低関係の特殊な場合とみなせる．学会発表の③と④）についても検討を行った．なお，目的(1)の達成の困難さが当初の予想を大幅に上回っていたため，目的(3)に関しては十分に検討を行うことができなかった．

４．研究成果

まず，目的(2)に関する成果を述べる．

研究背景の節で述べた典型的な静的安全性定義は，推論攻撃が可能になってしまうインスタンスが１つでも存在してはいけないということを要求している．現実にはありそうにないどんな特殊なインスタンスについてもこの要求を満たそうとすると，機密情報とは無関係な問合せしか許可できないという事態になりがちとなってしまう．機密情報とは無関係な問合せしか許可できないという事態は可用性を大きく犠牲にしており，一般の安全性要求としては厳しすぎると考えられる．

本研究では，問合せ解像度の概念を用いることで，典型的な静的安全性よりも安全性要求を下げた安全性定義を提案した．これらの安全性定義を用いることで，アプリケーションなどに最低限必要と考えられるインスタンス独立な安全性の検証を，場合によっては高速に行うことができるようになると期待できる．また，より一般的な安全性を表現するため，データベースにパラメータや固定データの概念を取り入れた安全性定義の拡張についても検討を行った．

以下，提案した安全性定義のうちのいくつ

かを説明する．機密情報を取り出す問合せを q_{sec} とし，ユーザに許可された問合せを q とする．

(1) 常時特定可能性に着目した安全性定義：
q が q_{sec} よりも解像度が高くないとき，データベースは安全であると定義する．たとえば，病院における患者の疾患状態データベースにおいて，患者の患っている病気を機密情報と考える．この安全性定義に従うと，「患者がどんな病気にかかっている場合においても，それが必ず特定されてしまう」ようなデータベースは安全でない．

(2) 常時識別可能性に着目した安全性定義：
q_{sec} よりも解像度が低いどんな q' に対しても q が q' より解像度が高くないとき，データベースは安全であると定義する．この安全性定義に従うと，「病名が胃がんだと推論できてしまう場合がある」ようなデータベースは安全であるが，「がんに関係する病気にかかっているかそうでないかをいつでも推論できてしまう」ようなデータベースは安全でない．

(3) インスタンスのパラメータ化：
問合せの引数によりインスタンスをパラメータ化した安全性定義を与えた．これにより，たとえば「(見知らぬ誰かの病名を偶然推論できるのには目をつぶるが) 狙った患者の病名を確実に推論できることは許さない」ということを表現できる．

(4) 固定データの導入：
推論攻撃によりデータベースの更新の有無を検知できるかという問題を考える．更新が起こりえない部分（固定データ）がインスタンス中に存在することがわかっていると，より強力な推論が可能となることを示した．そして，固定データがある場合の安全性を定義した．

次いで，目的(1)に関する成果を述べる．

連言問合せは，変数や定数から成るテーブル T と，T 中の変数から成るタプル u の組 (T, u) で定義される．u をサマリと呼ぶ．

2 つのテーブル X, Y に対し，$Y/X = \{g \mid g$ は $g(X) \subseteq Y$ を満たす写像$\}$ と定義する．連言問合せ $q=(T, u)$ をインスタンス D において評価した結果 $q(D)$ は，$(D/T)(u) (=\{g(u) \mid g \in (D/T)\})$ と定義される．

本研究では，なるべく広い連言問い合わせクラスで，問合せ間の解像度の高低関係が成立するための必要十分条件を求めることを目指した．直観的には，問合せ q に比べて問合せ q' のテーブルサイズが小さく，かつ問合せ q' のサマリの変数集合が問合せ q のサマリの変数集合を包含していれば，問合せ q' の方が問合せ q より問合せ解像度が高くなり，またその逆も成り立つと予想される．なぜならば，テーブルのサイズが小さければより多くのインスタンスに対して答えを返すことができ，またサマリの変数集合に包含関係があればインスタンスの違いをより敏感に認識できるからである．しかし，この予想は一般には正しくないことが分かった．具体例を図 1 と図 2 で示す．

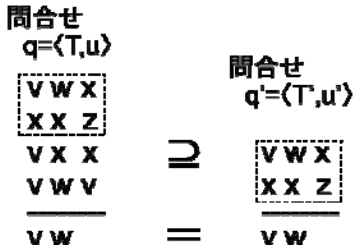

図 1. サマリが同一で，テーブル間に包含関係がある例

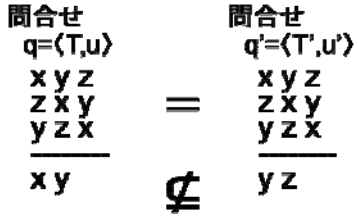

図 2. サマリに包含関係がなく，テーブルが同一である例

	q v w x x x z v x x v w v v w	q' v w x x x z v w
D={(1,2,3), (3,3,4), (1,3,3), (1,2,1)}	q(D)= {(1,2)}	q'(D)= {(1,2),(1,3)}
D' ={(1,2,3), (3,3,4), (1,3,3)}	q(D')= φ	q'(D')= {(1,2),(1,3)}

図 3. 図 1 中の問合せ q, q' に解像度高低関係がないことを示すインスタンス例

図 1 はサマリ u, u' を同一にし，テーブル T, T' に包含関係がある連言問合せの例である．しかし，この場合は問合せ間の解像度高低関係が成立しない．なぜなら図 3 のようなインスタンス D, D' を考えると，$q'(D)=q'(D')$ であるが $q(D) \neq q(D')$ となっており，すなわち q' の解像度は q よりも高くないからである．

一方，図 2 はテーブル T, T' を同一にし，サマリ u, u' 間に包含関係がない連言問合せ

の例である．しかしこのとき，u, u'間に包含関係がないにもかかわらず2つの問合せ解像度は一致する．なぜなら，すべての変数 x, y, zは対称であるため，q,q'は等価な問合せになっているからである．

このように，テーブル間およびサマリ間における単純な包含性だけでは問合せ間の解像度高低関係と一致しない．そこで本研究では，これらの考察をふまえ，以下のような3種類の前提条件をおいた上で，問合せ間の解像度高低関係が成立するための判定可能な必要十分条件をそれぞれ与えた．

(1) テーブルとサマリの変数集合が一致するという前提条件をおいた場合
(2) テーブルが一致するという前提条件をおいた場合
(3) 問合せ中に対称な変数が現れないという前提条件をおいた場合

今後の課題としては，これまでに得られた知見を活かして，前提条件をおかない場合の必要十分条件を求めることがあげられる．

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 Kenji Hashimoto, Kimihide Sakano, Fumikazu Takasuka, Yasunori Ishihara and Toru Fujiwara, “Verification of the Security against Inference Attacks on XML Databases,” IEICE Transactions on Information and Systems, Vol.E92-D, No.5, pp.1022-1032, 2009, 査読有.

〔学会発表〕（計19件）

① 松村 卓朗, 橋本 健二, 石原 靖哲, 藤原 融, “連言問合せ間の解像度高低関係成立のための必要十分条件に関する考察,” 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, 2011年2月28日, ラフォーレ修善寺（静岡県）.
② 川居 裕人, 橋本 健二, 石原 靖哲, 藤原 融, “XMLデータベースへの関数従属性を用いた推論攻撃に対する無限安全性検証法の提案,” 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, 2011年2月28日, ラフォーレ修善寺(静岡県）.
③ Yasunori Ishihara, Shogo Shimizu and Toru Fujiwara, “Extending the Tractability Results on XPath Satisfiability with Sibling Axes,” 7th International XML Database Symposium, 2010年9月17日, Grand Copthorne Waterfront Hotel, Singapore.
④ Yasunori Ishihara, Takuji Morimoto, Shougo Shimizu, Kenji Hashimoto and Toru Fujiwara, “A Tractable Subclass of DTDs for XPath Satisfiability with Sibling Axes,” 12th International Symposium on Database Programming Languages, 2009年8月24日, Cité Internationale, Lyon, France.
⑤ 廣田 祐一, 橋本 健二, 石原 靖哲, 藤原 融, “データベースへの推論攻撃に対する問合せ解像度の高低関係を用いたインスタンス独立の安全性定義の提案,” データ工学と情報マネジメントに関するフォーラム, 2009年3月9日, ヤマハリゾートつま恋（静岡県）.
⑥ 廣田 祐一, 橋本 健二, 石原 靖哲, 藤原 融, “データベースへの推論攻撃に対する問合せ解像度に基づいた安全性定義の提案,” コンピュータセキュリティシンポジウム2008, 2008年10月9日, 沖縄コンベンションセンター（沖縄県）.

６．研究組織
(1)研究代表者
石原　靖哲（ISHIHARA　YASUNORI）
大阪大学・大学院情報科学研究科・准教授
研究者番号：00263434