# 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な安全性定義の提案

廣田　祐一† 　橋本　健二† 　石原　靖哲† 　藤原　融†

† 大阪大学大学院情報科学研究科　〒 565-0871 吹田市山田丘 1-5
E-mail: †{y-hirota,k-hasimt,ishihara,fujiwara}@ist.osaka-u.ac.jp

**あらまし**　データベース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することをいう．これまでに推論攻撃に対する安全性の検証法がいくつか提案されており，検証に用いられている安全性の定義は，大別すると「インスタンス依存」と「インスタンス独立」の 2 種類に分けられる．インスタンス独立な安全性は更新が頻繁に行われるデータベースに有用であるが，全てのインスタンスにおいてインスタンス依存の安全性が成立するものと単純に定義すると，機密情報とは無関係の問合せしか許可できないという事態になりがちである．これは可用性を大きく下げ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．本稿では筆者らが提案している問合せ解像度の高低関係を用いて，安全性要求を下げたインスタンス独立な安全性定義を提案する．また，提案する安全性が決定可能となる条件についても検討する．

**キーワード**　セキュリティ，推論攻撃，問合せ解像度，安全性，インスタンス独立

# Instance-independent security definitions using query resolution against inference attacks on databases

Yuichi HIROTA†, Kenji HASHIMOTO†, Yasunori ISHIHARA†, and Toru FUJIWARA†

† Graduate School of Information Science and Technology，Osaka University　1-5 Yamadaoka，Suita，Osaka，565-0871，Japan
E-mail: †{y-hirota,k-hasimt,ishihara,fujiwara}@ist.osaka-u.ac.jp

**Abstract**　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. So far, some security verification methods against inference attacks have been proposed. Security definitions verified by the methods are classified into “instance-dependent” ones and “instance-independent” ones. Instance-independent security is useful in a situation that database instances are frequently updated. However, if an instance-independent security is defined as that every possible database instance satisfies some instance-dependent security, it is often the case that only the queries completely independent of the secret information can be allowed. In this case, the availability of the database is low, and so the instance-independent security is too strong as security requirements. In this paper, we propose some suitably-relaxed instance-independent security definitions based on a concept of “query resolution”. Moreover, we explore the condition where the security is decidable.

**Key words**　security, inference attacks, query resolution, instance-independent

## 1. ま え が き

一般に企業などの組織がかかえるデータベースの中には機密情報が含まれており，アクセス制御により権限のないのユーザは直接アクセスできないよう処理がなされている．しかし，アクセス制御により直接的なアクセスを防ぐだけでは，推論攻撃による機密情報の漏えいを防ぐことができない．推論攻撃とは，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ようとする試みのことをいう．

これまでに推論攻撃に対する安全性の検証法がいくつか提案されており，検証に用いられている安全性の定義は，大別すると「インスタンス依存」と「インスタンス独立」の 2 種類に分

けられる．

インスタンス依存の安全性とは，許可された問合せ $q_1, \ldots, q_n$ と機密情報を取り出す問合せ $q_{\mathrm{sec}}$ の定義およびデータベースインスタンス $D$ に対する問合せ結果 $q_1(D), \ldots, q_n(D)$ とデータベース制約情報 (スキーマ，関数従属性など) から，機密情報の値を特定または絞り込むことができないことをいう．安全性検証の際に，検証するインスタンスに対する問合せ結果を用いるため，データベースが更新される度に検証のやりなおしが必要となる．

一方，インスタンス独立な安全性とは，広義にはデータベースインスタンスが安全性定義のパラメータ (入力) になっていない安全性を指す．しかし，通常は，なんらかのインスタンス依存の安全性定義に基づいて，以下のように定義されている: データベース制約に従うどんなインスタンス $D$ においても，許可された問合せ $q_1, \ldots, q_n$ と機密情報を取り出す問合せ $q_{\mathrm{sec}}$ の定義およびデータベースインスタンス $D$ に対する問合せ結果 $q_1(D), \ldots, q_n(D)$ とデータベース制約情報から機密情報の値を特定または絞り込むことができない．この形の定義をもつインスタンス独立な安全性の利点として，インスタンスを入力としないのでインスタンスのサイズによらず比較的少ない時間で安全性の検証が可能な点がある．また，インスタンス独立な安全性は，データベース制約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となるため重要な研究課題である．しかし，問題構造の複雑さもあり，これまで十分に研究されてきたとはいえない．また，狭義のインスタンス独立な安全性は，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．そのような例を次に示す．

［例 1］ ある学校における学生の個人情報を含む関係データベースについて考える．データベースは属性として「名前」「出身地」「成績」のみをもつとする．今，学生の名前と成績の組を機密情報として考え，⟨ 名前，出身地 ⟩，⟨ 出身地，成績 ⟩ の組一覧を返すものを許可された問合せとしてそれぞれを $q_1$，$q_2$ とする．ここで，表 1，2，3 に示すようなインスタンス $D_1$ とその問合せ結果 $q_1(D_1)$，$q_2(D_1)$ について考える．今，インスタンス $D_1$ を見ることができず，タプルの並び順は関係ないとして考えると，これらの問合せ結果 $q_1(D_1)$，$q_2(D_1)$ のみからは出身地が大阪である学生も京都である学生も成績を特定することができない．しかし，表 4，5，6 に示すようなインスタンス $D_2$ とその問合せ結果 $q_1(D_2)$，$q_2(D_2)$ について考えると，出身地が京都である学生は D しか存在しないため，D の成績は良であることが特定されてしまう．よって特定可能性の意味で安全でない．このように，任意の可能なインスタンスについての安全性を保証する必要があるため，機密情報と許可された問合せに関係がある限りは，安全でないインスタンスが 1 つでも存在する可能性は高い．

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．本稿では文献 [1] で提案されている問合せ解像度の概念に基づく解像度の高低関係を用いて，機密情報の値を特定しようとする攻撃に対する安全性要求を下げた広義のインスタンス独立な安全性定義の提案を行う．問合せ解像度とは，データベースインスタンスの全体集合の，$q$ の結果に基づく同値類分割である．$q$ による同値類分割より $q'$ による同値類分割の方が細かいとき，$q'$ の解像度は $q$ の解像度より高いという．直観的には，問合せ結果ごとにインスタンスのグループ分けを行い，$q$ がインスタンス $D, D'$ を区別できるならば $q'$ も $D, D'$ を区別できることを表す．提案する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考えられる．

| 名前 | 出身地 | 成績 |
|---|---|---|
| A | 大阪 | 優 |
| B | 大阪 | 良 |
| C | 京都 | 優 |
| D | 京都 | 良 |

表 1　インスタンス $D_1$

| 名前 | 出身地 |
|---|---|
| A | 大阪 |
| B | 大阪 |
| C | 京都 |
| D | 京都 |

表 2　$q_1(D_1)$

| 出身地 | 成績 |
|---|---|
| 大阪 | 優 |
| 大阪 | 良 |
| 京都 | 優 |
| 京都 | 良 |

表 3　$q_2(D_1)$

| 名前 | 出身地 | 成績 |
|---|---|---|
| A | 大阪 | 優 |
| B | 大阪 | 良 |
| C | 大阪 | 優 |
| D | 京都 | 良 |

表 4　インスタンス $D_2$

| 名前 | 出身地 |
|---|---|
| A | 大阪 |
| B | 大阪 |
| C | 大阪 |
| D | 京都 |

表 5　$q_1(D_2)$

| 出身地 | 成績 |
|---|---|
| 大阪 | 優 |
| 大阪 | 良 |
| 大阪 | 優 |
| 京都 | 良 |

表 6　$q_2(D_2)$

本稿の構成は以下の通りである．2 節では関連研究に関して述べる．3 節では文献 [1] の問合せ解像度に関する諸定義を述べる．4 節では問合せ解像度の高低関係を用いた安全性定義を述べる．5 節では問合せ解像度の高低関係の決定可能性について述べる．6 節では結論と今後の課題を述べる．

## 2. 関 連 研 究

これまでに研究されてきた，推論攻撃に対する安全性の多くは，インスタンス依存の安全性である．古くは文献 [2] で紹介されている，集約操作を用いた攻撃とそれに対する安全性がある．また最近提案された $k$ 匿名性 [3] や $\ell$ 多様性 [4] は，特定の個人に対応付け可能なデータベース中のタプルを攻撃者がどれだけ絞り込めるかという観点に基づく安全性定義である．また，$t$ 接近性 [5] は，$k$ 匿名性や $\ell$ 多様性の欠点を補う安全性定

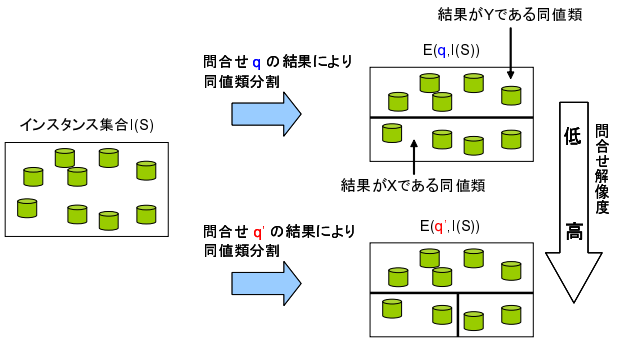


図 1　$q \preceq_{I(S)} q'$ となる例

義として提案されたものであり，特定の個人に対応付け可能なタプル集合における機密情報の値の分布が，データベース全体における値の分布に近いことを要求している．文献 [6] は XML データベースにおける推論攻撃に対する安全性定義および検証法を提案している．また文献 [7] では，本稿と同じく同値類分割の考え方を用いて，攻撃者が機密情報の値をどれだけ絞り込み可能かを表す指標値を提案している．

一方，インスタンス独立な安全性を扱った研究として，文献 [8] は，関数従属性を持つ関係データベースにおけるインスタンス依存の安全性とインスタンス独立な安全性の両方について検討しており，選択と射影のみからなる問合せクラスについての安全性検証法を提案している．また，文献 [9] はオブジェクト指向データベースにおける安全性を扱っている．文献 [10] では，値の分布に関する知識の変化に基づく，安全性定義を提案している．そして，本稿の先行研究である文献 [1] では，問合せ解像度の高低関係を用いて機密情報の特定可能性と変更検知可能性に着目した安全性定義を提案している．

インスタンス独立な安全性定義はいくつか提案されてはいるが，筆者らの知る限りではどれも安全性要求が厳しいものであると考える．

## 3. 問合せ解像度の諸定義と既知の性質

本節では文献 [1] で述べた問合せ解像度の諸定義と，既知の性質について述べる．

データベース制約情報 (スキーマ，関数従属性など) $S$ に従うインスタンスの集合を $I(S)$ と書く．問合せ $q$ に対し，$q(D) = q(D')$ $(D, D' \in I(S))$ のとき $D \equiv_{q,I(S)} D'$ と書く．$\equiv_{q,I(S)}$ により定まる $I(S)$ 上の同値類の集合を $I(S)$ における $q$ の解像度といい，$E(q, I(S))$ と書く．なお，問合せ集合 $Q = \{q_1, \ldots, q_n\}$ に対しても，$Q$ は $D$ に対して $\langle q_1(D), \ldots, q_n(D)\rangle$ という組を返す 1 つの問合せとみなし，$\equiv_{Q,I(S)}$ や $E(Q, I(S))$ を同様に定義する．

[定義 1]（問合せ解像度の高低関係）　任意の $C' \in E(q', I(S))$ に対してある $C \in E(q, I(S))$ が存在して $C' \subseteqq C$ (すなわち，任意の $D, D' \in I(S)$ について $q'(D) = q'(D')$ ならば $q(D) = q(D')$) ならば，$I(S)$ において $q'$ は $q$ より解像度が高いといい，$q \preceq_{I(S)} q'$ と書く (図 1)．インスタンス集合が自明な時，または特に指定されない時は単に $q \preceq q'$ と書く．

直観的には，解像度が高い問合せの方がより敏感にインスタンスの違いを検出できることを示している．定義 1 から以下の性質が導かれる．

- 問合せ $q$，$q'$ の合成を $q \circ q'$(ただし $q \circ q'(D) = q(q'(D))$) と書くと，$q' \preceq q \circ q'$.
- $q \preceq_{I(S)} q'$ かつ $I(S') \subseteqq I(S)$ ならば，$q \preceq_{I(S')} q'$.

次に，問合せ解像度の高低関係と，それと似た定義を持ち決定可能な問合せクラスが多く知られている問合せ包含関係との関係について述べる．

[定理 1]　以下の 3 条件を満たす問合せ $q$，$q'$ とスキーマ $S$ において，問合せの包含関係と問合せ解像度の高低関係は必要十分条件の関係にある．すなわち，

$$\forall D \in I(S), q(D) \subseteqq q'(D) \Leftrightarrow q \preceq_{I(S)} q'.$$

（1）　任意の $D \in I(S)$ について $q(D) \subseteqq D$, $q'(D) \subseteqq D$.

（2）　任意の $D, D' \in I(S)$ について $q(D \cup D') = q(D) \cup q(D'), q'(D \cup D') = q'(D) \cup q'(D')$.

（3）　$I(S)$ は部分集合について閉じている．すなわち，$D \in I(S)$ かつ $D' \subseteqq D$ ならば $D' \in I(S)$.

なお，定理 1 よりもより広い問合せクラスである，連言問合せ (conjunctive query [11]) のクラスにおける問合せ解像度の高低関係と問合せ包含関係の関係についても現在考察を行っている．

## 4. 問合せ解像度の高低関係を用いた安全性定義

本節では，従来のインスタンス独立な安全性定義より安全性要求を下げた安全性定義を提案する．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う．

### 4.1 常時特定可能性に着目した安全性定義

ある病院における患者の疾患状態データベースについて，患者の患っている病気を機密情報と考える．従来のインスタンス独立な安全性では患者の病気が特定されるようなインスタンスが 1 つでもある場合は安全でないとされていた．つまり，患者がいかなる病気にかかっている場合においても，それが特定される可能性が少しでもあると安全でないということになる．しかし，例えばアンビエント情報社会のように，極めて多くのユーザが，それぞれ異なるプライバシポリシをもちつつ，一つの巨大なシステムを利用しているような状況を考えた場合，「何らかのプライバシが推論されるユーザが一人でもいれば安全ではない」という定義を採用するのは，可用性を大きく損なうと考えられる．ここでは，1 節で述べた動機付けと同様に，「あらゆるユーザについてそのプライバシが推論されるならば安全ではない」という安全性要求を下げた定義を考える．例えば，患者がいかなる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時を安全でないと考える．そのような安全性は，機密情報を取り出す問合せを $q_{\mathrm{sec}}$ とし，ユーザに許可された問合せを $q$ として，問合せ解像度の高低関係を使った式では「$q_{\mathrm{sec}} \preceq_{I(S)} q$ ならば安全でない」と定義できる．しかし，全てのインスタンス集合 $I(S)$ において機密情報問合せと許可問合せに問合せ解像度の高低関係が成り立つ場合は多くはないと考

えられる．

### 4.2 常時識別可能性に着目した安全性定義

前節の安全性定義では安全性要求を下げすぎているため，おそらくほとんどの状況を安全と判定することになってしまう．本節では適度に安全性要求をさげたものを考える．

例えば可用性の確保のため「病名が胃癌だと推論できてしまう場合がある」のは目をつぶり「癌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」のは防ぎたい場合がありうる．このように安全性要求を下げた安全性は解像度の高低関係を用いることで表現することができる．

［定義 2］（常時識別可能性に着目した安全性定義） ある $q'_{\mathrm{sec}} \preceq_{I(S)} q_{\mathrm{sec}}$ が存在して，$q'_{\mathrm{sec}} \preceq_{I(S)} q$ であるとき，$q_{\mathrm{sec}}$ は $q$ を用いた推論攻撃に対して安全でない．(ただし問合せ結果が 1 種類しかない $q'_{\mathrm{sec}}$ は除外する)

この定義の直観的な意味は，$q$ や $q_{\mathrm{sec}}$ などの問合せによって分割される同値類のグループによってインスタンス集合に仕切りが入るものだと考えると，$q$ と $q_{\mathrm{sec}}$ の間に部分的に一致する仕切りが存在するとき安全でないということになる．安全性要求が下がったことを示すため，次のような例を考える．

［例 2］ ある病院における患者の疾患状態データベースについて考える．患者は 1 つだけ病気を患っているとする．問合せ $q_{\mathrm{sec}}, q_1, q_2$ を以下に示すものとする．

$q_{\mathrm{sec}}$： 患者 A の患っている病気を返す

$q_1$： 患者 A が入院している病棟を返す

$q_2$： B 棟に入院している全ての患者の病気を返す

このとき，患者 A が B 棟に入院していてかつ B 棟に入院している患者の病気が 1 種類であるときのみ，患者 A の病気が特定されてしまうにも関わらず，従来の (狭義の) インスタンス独立な安全性定義では，少しでも安全でない場合があるとき全体として安全でないと定義されているため，このような状況を安全でないと判定する．定義 2 ではこの状況を安全であると判定する．

また，定義 2 において安全でないと判定される例を以下に示す．

［例 3］ ある病院における患者の疾患状態データベースについて考える．患者と病気の対ごとに担当医師が 1 人定まり，その対応は変更されないとする．また，医師は 1 人につき 1 つの病気を担当するとし，患者は 1 つだけ病気を患っているとする．問合せ $q_{\mathrm{sec}}, q_1, q_2$ を以下に示すものとする．

$q_{\mathrm{sec}}$： 患者 A の患っている病気を返す

$q_1$： 患者 A を担当する医師を返す

$q_2$： 癌担当の医師全員を返す

このとき，患者 A の患っている病気が癌である場合は，患者 A の患っている病気が癌であることがいつでも特定されてしまう．反対に，患者 A の患っている病気が癌でない場合は，患者 A の患っている病気が癌でないことがいつでも特定されてしまう．定義 2 ではこの状況を安全でないと判定する．

### 4.3 安全性定義の拡張

本節では前節で提案した安全性定義をより一般的なものに拡張するため，データベースインスタンスのパラメータ化と，データベース内の固定データの存在について検討を行う．

#### 4.3.1 パラメータの導入

例 2 や例 3 では，A という患者に固定した例を考えているが，患者名は問合せの引数として与えられると想定するのが自然である．これを実現するために，データベースインスタンスを以下のとおりパラメータ化する．

$U$ を可能なパラメータ値の無限集合とする．各データベースインスタンス $D$ は，各パラメータ値 $u \in U$ とそれに対応する部分インスタンス $D_u$ から成っているとする．部分インスタンス $D_u$ への問合せ $q$ の結果 $q(D_u)$ を $q^u(D)$ で表す．このとき，安全性は形式的には以下のように定義できる．

［定義 3］（パラメータを導入した安全性定義） ある $q'_{\mathrm{sec}} \preceq q_{\mathrm{sec}}$ が存在して，任意の $u \in U$ について，ある $\{u_1, \ldots, u_n\} \subseteqq U$ が存在して $q'^{u}_{\mathrm{sec}} \preceq \{q_1^{u_1}, \ldots, q_n^{u_n}\}$ である時，安全でない．

この定義の直観的な意味は，「(見知らぬ誰かのプライバシを偶然推論できるのには目をつぶるが) 狙ったユーザのプライバシを確実に推論できることは許さない」ということに相当している．

#### 4.3.2 固定データの導入

より一般的なデータベースを考えるため，前節のようにパラメータ化したデータインスタンスにおいて，文献 [1] でも扱った固定データの概念を表現することを考える．固定データとは全ての部分インスタンスにおいて変化のない部分を表すとする．すなわち，インスタンス $D$ が集合で表すことができるデータベースだとすると，すべての $u \in U, D_u$ の交わりであると定義できる．すなわち，固定データを $D_f$ とすると $D_f = \bigcap_{\forall u \in U} D_u$ と表される．ここで，固定データを考慮した上で定義 3 を以下のように再定義する．

［定義 4］（パラメータを導入した安全性定義 2） 任意の $C_f \in E(q^f, I(S))$ において，ある $q'_{\mathrm{sec}} \preceq_{C_f} q_{\mathrm{sec}}$ が存在して，任意の $u \in U$ について，ある $\{u_1, \ldots, u_n\} \subseteqq U$ が存在して $q'^{u}_{\mathrm{sec}} \preceq_{C_f} \{q_1^{u_1}, \ldots, q_n^{u_n}\}$ である時，安全でない．

これは攻撃者がデータベース内に変化のない部分 (医師が担当する患者の病気等) の情報が与えられている場合を想定しており，攻撃者にとってより有利な状況となっている．

## 5. 解像度の高低関係の決定可能性

本節では定義 1 で定めた解像度の高低関係の決定可能性について検討を行い，問合せを選択演算と射影演算に限定した上で，提案した安全性定義が決定可能となる条件を示す．そして提案した安全性定義が決定可能であるか検討する．

まず，インスタンス集合や問合せクラスに特に制限をかけない場合について検討を行う．このとき問合せ解像度の高低関係の決定可能性について以下の定理が導かれる

［定理 2］ $I(S)$ を多項式時間認識可能な集合とし，$q$，$q'$ を多項式時間計算可能な関数とする．$I(S)$ において $q \preceq_{I(S)} q'$ は決定不能である．

（証明）Post の対応問題（PCP）から帰着する．PCP のインスタンスを $\langle u_1, \ldots, u_n, v_1, \ldots, v_n \rangle$ とする．$\alpha, \beta$ を新たな記号とし，$I(S) = \alpha \cdot \{1, \ldots, n\}^+ \cup \beta \cdot \{1, \ldots, n\}^+$ とする．$q$，$q'$

を以下のように定義する．

$$q(\alpha\cdot i_1\cdots i_k)=\alpha\cdot i_1\cdots i_k,$$
$$q(\beta\cdot i_1\cdots i_k)=\begin{cases}\beta\cdot i_1\cdots i_k & \text{if } u_{i_1}\cdots u_{i_k}=v_{i_1}\cdots v_{i_k}\\ \alpha\cdot i_1\cdots i_k & \text{otherwise.}\end{cases}$$

$$q(\alpha\cdot i_1\cdots i_k)=\alpha\cdot i_1\cdots i_k,$$
$$q(\beta\cdot i_1\cdots i_k)=\begin{cases}\alpha\cdot i_1\cdots i_k & \text{if } u_{i_1}\cdots u_{i_k}=v_{i_1}\cdots v_{i_k}\\ \beta\cdot i_1\cdots i_k & \text{otherwise.}\end{cases}$$

$I(S)$ は多項式時間認識可能であり，$q$ と $q_{\rm sec}$ はどちらも多項式時間計算可能である．

PCP インスタンスが解をもつ時，PCP の解となる $i_1\cdots i_k$ について，$q(\alpha\cdot i_1\cdots i_k)\neq q(\beta\cdot i_1\cdots i_k)$，$q'(\alpha\cdot i_1\cdots i_k)=q'(\beta\cdot i_1\cdots i_k)$，PCP の解とならない $i_1\cdots i_k$ について，$q(\alpha\cdot i_1\cdots i_k)=q(\beta\cdot i_1\cdots i_k)$，$q'(\alpha\cdot i_1\cdots i_k)\neq q'(\beta\cdot i_1\cdots i_k)$，すなわち $q\not\preceq_{I(S)}q'$ となる．

逆に，PCP インスタンスが解をもたない時，任意の $i_1\cdots i_k$ について $q(\alpha\cdot i_1\cdots i_k)=q(\beta\cdot i_1\cdots i_k)$，$q'(\alpha\cdot i_1\cdots i_k)\neq q'(\beta\cdot i_1\cdots i_k)$，すなわち $q\preceq_{I(S)}q'$ となる． □

定理 2 によりインスタンス集合や問合せクラスに特に制限をかけない場合は，問合せ解像度の高低関係は決定不能であることが示された．そこで次は，インスタンス集合や問合せクラスに制限をかけて，問合せ解像度の高低関係が判定できるような条件について検討を行う．具体的には，次のようにモデル化したデータベース，選択演算問合せ，射影演算問合せにおいて検討を行う．

- データベースは，$\Sigma$ を有限アルファベットとして，

$$I(S)=\{D|D\text{ は}\Sigma^*\text{の任意の有限部分集合 }\}$$

と表す．

- 選択演算問合せは次の条件を満たす $q$ とし，$D$ の部分要素を受理するようなオートマトンとする．
  - 任意の $D\in I(S)$ について $q(D)\subseteqq D$.
  - 任意の $D,D'\in I(S)$ について $q(D\cup D')=q(D)\cup q(D')$.
- 射影演算問合せは $D$ の要素の一部を特別な文字 \$ に書き換える順序機械とする．

### 5.1 選択演算問合せについて

モデル化を行った選択演算問合せは定理 1 の条件 (1)，(2) と一致している．定理 1 の条件 (1)，(2) を満たす問合せの結果は $q(D)=\bigcup_{e_D\in D}q(\{e_D\})$ と書ける．これは直観的には，$q$ が $D$ の中の要素一つ一つについてみていき，$q$ の定義に従ってその要素を残すか今，定理 1 の条件 (3) よりも本節のデータベースのモデル化の条件の方が制限が強いので，このモデル化の下では選択演算問合せにおいて問合せ解像度の高低関係と問合せの包含関係は一致している．よって，言語の包含判定問題に帰着することにより，次の定理が導かれる．

［定理 3］ 定理 1 の 3 条件を満たす問合せ $q$，$q'$ とスキーマ $S$ において，$q$，$q'$ が有限オートマトンで表現できる問合せクラスである時，$q\preceq_{I(S)}q'$ は決定可能である．

(証明) 問合せの包含関係の判定を正規言語の包含判定問題に帰着する．仮定より $q$，$q'$ をそれぞれ有限オートマトン $A$，$A'$ で表現できるとし，$A$，$A'$ が受理する言語をそれぞれ $L(A)$，$L(A')$ と書くとする．$D$ は言語として考えることができ，$q(D)$ は $D$ の要素のうち $q$ で受理される要素が残るので，$q(D)=D\cap L(A)$ となる．同様に $q'(D)=D\cap L(A')$ となる．正規言語の包含判定問題は判定可能であることが知られており，今 $D\cap L(A)\subseteqq D\cap L(A')$ を判定することで $q(D)\subseteqq q'(D)$ は判定可能である．よって $q\preceq_{I(S)}q'$ も判定可能である． □

### 5.2 射影演算問合せについて

続いて射影演算の問合せについて検討する．問合せ $q$ を $D$ の要素の一部を特別な文字 \$ に置き換えるような順序機械とモデル化すれば射影演算とみなすことができる．

［定理 4］ 本節でモデル化した射影演算問合せ $q$，$q'$ とインスタンス集合 $I(S)$ において，$q$，$q'$ が有限状態順序機械で表現できる問合せクラスである時，$q\preceq_{I(S)}q'$ は決定可能である．

(証明) 問合せ $q$ が $D$ の中の系列 $x$ について，$x$ の $i$ 番目の文字をそのまま残すか，\$ に変換するかどうかが，$x$ の 1 番目から $i$ 番目までの文字に依存している場合「$q$ が $x$ の $i$ 番目の文字を \$ に変換する $\Leftrightarrow$ $x$ の 1 番目から $i$ 番目までの文字列が $L$ に属さない」として，$q$，$q'$ に対応する言語をそれぞれ $L$，$L'$ とする．ある語 $wa(w\in\Sigma^*,a\in\Sigma)$ が存在して $wa\in L-L'$ とする．すると，$q'$ は $wa$ と $wb$ を区別できる．したがって，$L'\subseteqq L$ ならば「$L$ のほうが解像度が真に低い」ということはありえない．ある語 $w'a(w'\in\Sigma^*)$ が存在して $w'a\in L\cap L'$ であり，かつ，$q(\{wa\})=q(\{w'a\})$ が成り立つ場合，高低関係はなくなる．それ以外に高低関係が成り立たない場合はない．よって，$q'\preceq q\Leftrightarrow L'\subseteqq L$ かつ任意の対 $w\in L-L',w'\in L\cap L'$ について $q(\{w\})\neq q(\{w'\})$．後ろの式は，$q(L-L')\cap q(L\cap L')=\emptyset$ と書ける．今，$L$ が正規言語の場合，$q$ は GSM 写像となる．よって，$L'$ も正規言語の場合，$q(L-L')$ も $q(L\cap L')$ も正規言語となる．したがって，$q$ が有限状態順序機械であれば積の空判定は可能となる． □

### 5.3 部分的な解像度の高低関係の一致について

本節では，定義 2 にあるような，部分的なしきりの一致を判定することができるのか，つまり $q''\preceq q'$ が存在して，$q''\preceq q$ となることが決定可能かを検討する．定理 3 により，選択演算問合せではオートマトンが認識する言語の包含判定問題に帰着することにより，問合せ解像度の高低関係が決定可能であることを示した．よって $q'_{\rm sec}\preceq q_{\rm sec}$ かつ $q'_{\rm sec}\preceq q$ となるような $q'_{\rm sec}$ があることと $q_{\rm sec}$ の言語と $q$ の言語とが互いに素ではないこととが同値となる．これは記憶領域が有限であるオートマトンでは言語の積の空判定は決定可能である．なお，射影演算問合せについては現在検討中である．

## 6. あ と が き

本稿では問合せ解像度という概念に基づく解像度の高低関係を用いることで，従来のインスタンス独立な安全性よりも安全性要求の低い安全性定義の提案を行った．また，解像度の高低関係が決定可能となる問合せクラスについて検討した．

今後の課題としては，より具体的なモデル化を行い，関係データベースや XML データベースのモデル上で，安全性検証法を確立することがあげられる．問合せ解像度の高低関係の判定を高速に行える問合せなどのクラスを明らかにできれば，最低限必要となるインスタンス独立な安全性検証法を高速に行うことが可能となる．また，アプリケーションに応じて必要となる安全性要求に対応できるように，安全性要求をパラメータ化して表現することも検討する．

## 謝　　辞

本研究の一部は科学研究費補助金 (課題番号 20500092) によるものである．

## 文　　献

[1] 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, 第 2008 巻, pp. 467–472 (2008).

[2] D. E. R. Denning: “Cryptography and Data Security”, Addison-Wesley (1982).

[3] L. Sweeney: “$k$-anonymity: A model for protecting privacy”, International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, **10**, 5, pp. 557–570 (2002).

[4] A. Machanavajjhala, J. Gehrke, D. Kifer and M. Venkitasubramaniam: “$\ell$-diversity: Privacy beyond $k$-anonymity”, Proceedings of the 22nd International Conference on Data Engineering, p. 24 (2006).

[5] N. Li, T. Li and S. Venkatasubramanian: “$t$-closeness: Privacy beyond $k$-anonymity and $\ell$-diversity”, Proceedings of the 23rd International Conference on Data Engineering, pp. 106–115 (2007).

[6] K. Hashimoto, F. Takasuka, K. Sakano, Y. Ishihara and T. Fujiwara: “Verification of the security against inference attacks on XML databases”, Proceedings of the Asia Pacific Web Conference (APWeb 2008), Vol. 4976 of LNCS, pp. 359–370 (2008).

[7] K. Zhang: “IRI: A quantitative approach to inference analysis in relational databases”, Database Security XI, pp. 279–290 (1998).

[8] A. Brodsky, C. Farkas and S. Jajodia: “Secure databases: Constraints, inference channels, and monitoring disclosures”, IEEE Transactions on Knowledge and Data Engineering, **12**, 6, pp. 900–919 (2000).

[9] Y. Ishihara, T. Morita, H. Seki and M. Ito: “An equational logic based approach to the security problem against inference attacks on object-oriented databases”, Journal of Computer and System Sciences, **73**, pp. 788–817 (2007).

[10] G. Miklau and D. Suciu: “A formal analysis of information disclosure in data exchange”, Journal of Computer and System Sciences, **73**, 3, pp. 507–534 (2007).

[11] S. Abiteboul, R. Hull and V. Vianu: “Foundations of Databases”, Addison-Wesley (1995).