様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２８ 年　 ６ 月 ２２ 日現在

機関番号：１３９０１
研究種目：基盤研究(B)（一般）
研究期間：2011～2014
課題番号：２３３００００８

研究課題名（和文）形式言語理論に基づく静的解析法とその安全性検査への応用

研究課題名（英文）Software Analysis based on Formaly Language Theory and Its Application to Security Verification

研究代表者
関　浩之（Seki, Hiroyuki）

名古屋大学・情報科学研究科・教授

研究者番号：８０１９６９４８

交付決定額（研究期間全体）：（直接経費）　　15,400,000円

研究成果の概要（和文）：木言語理論を用い，XML文書等の構造化データに対する情報保存性およびセキュリティに関して以下の成果を挙げた．変換vが問合せqを保存するとは、任意のデータtに対して，　tへの問合せqの結果を，変換の結果v(t)からも得ることができることをいう．木変換器や頂点問合せ等のモデルに基づき問合せ保存性問題の判定可能性や判定に要する計算量を明らかにした．許可問合せの結果や公開情報を利用して禁止問合せの結果(機密情報)を得ようとする行為を推論攻撃と呼ぶ．本研究ではk-安全性，l-多様性という推論攻撃に対する２つの安全性に着目・導入し，それらの判定可能性の理論的考察および判定手法の実験的評価を行った．

研究成果の概要（英文）：We obtained the following research results on information preservation and security of structured data, especially XML documents, based on tree language theory. A translation v is said to preserve a query q if there is a query q' that can obtain from v(t) the same result when q is applied to t. We obtained decidability and complexity results on the problem of deciding preservation based on tree transducers and n-ary node queries. An inference attack is a behavior that tries to obtain the result of an unauthorized query by combining the result of authorized queries and other public information. We focused on k-secrecy and l-diversity as security notions against inference attacks. We discussed the decidability of schema k-secrecy problem and also compared the effectiveness of our two proposed methods of deciding l-diversity.

研究分野：ソフトウェア基礎理論

キーワード：ソフトウェア検証　形式言語理論　モデル検査　XML　情報保存性　木変換器　セキュリティ

１．研究開始当初の背景

ソフトウェアの静的解析問題とは、ソフトウェア S がそれに求められる要求を満たして動作するかどうかを判定する問題である。要求は安全性と生存性に分類できるが、以下では簡単のため安全性を例として説明する。（生存性の場合は以下で到達可能集合の代わりに無限長実行系列集合を考える。）状態集合 X から S の実行によって到達できる状態集合を $S^*(X)$と表し、初期状態集合を I で表す。すると静的解析問題は次のように定式化することができる。「ソフトウェア S に対する要求を満たす状態集合を R で表すとき、$S^*(I)\subseteq R$ が成り立つかどうかを判定せよ。」例えば S が有限状態遷移系の場合、$S^*(I)$も R も有限集合となるので静的解析問題は明らかに自動判定可能である。これが現在広く利用されているモデル検査器の動作原理である。しかし、現実のソフトウェアはデータ構造、制御構造（繰返し・再帰）および実時間性・確率的振舞い等の観点から有限状態遷移系としてモデル化することが適切でないことも多い。S を無限状態遷移系によってモデル化する場合でも、$S^*(I)\subseteq R$　が判定可能であるような部分クラスがいくつか知られている（例えばプッシュダウンオートマトン(PDA)）。PDA の他にも、データ構造（木、グラフ）、時間、確率等の観点から有限状態遷移系を拡張した種々の計算モデルにおいて、静的解析問題の判定可能性や計算量が考察されている。しかし現在のマルチスレッド化、マルチコア CPU や多段キャッシュへの対応など、複雑化するプログラムの挙動に適した形式モデルや諸性質の考察は依然十分とはいえない。

研究代表者の関は既に、PDA と能力が等価であり再帰プログラム形を自然にモデル化できる文脈自由文法(context-free grammar, CFG)に基づく静的解析法を用い、言語組み込みアクセス制御とよばれる機構をもったプログラムのモデル検査、セキュリティ仕様からアクセス制御文の自動生成に関する研究等を行ってきた。また、研究分担者の小川は重み付きプッシュダウンシステム(PDS = 入力なし PDA) に基づくモデル検査法のスケーラビリティの飛躍的向上、PDA の部分クラスに対する静的解析法の理論的検討, 椙と橋本は木言語の枠組みを用い、暗号を用いたプロトコルの安全性検証、XML データベースにおける推論攻撃に対する安全性検証を行ってきた。これらの経験をふまえ、大規模化とマルチスレッド化が進展するソフトウェア信頼性向上のための現実的な計算モデルの提案と、それに基づく検証法の開発を行う必要性を認識するに至った。

２．研究の目的

インターネットの普及により、クラウドと総称される大規模かつハイブリッド化した計算機システムが構築されている。その基盤ソフトウェアは、マルチスレッド化による処理効率向上、CPU のマルチコア化・キャッシュの多段化への対応のため複雑化し、従来のテスト手法では十分なデバッグが行えず開発が困難になっている。静的解析法は信頼性向上の有用な手法であるが、プログラムの複雑化に対応する形式モデルの拡張・スケーラビリティの向上が必須である。本研究では、研究代表者らが過去約 25 年にわたって蓄積してきた形式言語理論の知見に基づき、現実のソフトウェアのモデル化に適する無限状態遷移系を設定し、静的解析の自動化を理論的に考察する。次に検証系を実装し、セキュリティ解析等の具体的事例を用いて提案手法の有効性を実証する。

３．研究の方法

23 年度は主として検証対象を表現するための計算モデルの設定とそれに関する数学的諸性質（演算閉包性、基本問題の判定可能性と計算量）を理論的に究明する。24 年度以降はより具体的な問題設定の下での理論的研究，事例研究ならびに解析系や検証系の実装とそれに基づく実証実験を行う。具体的に、構造化データに対する情報保存性に関する研究、構造化データに対する安全性に関する研究，木文法によって圧縮された構造化文書の直接問合せ・更新に関する研究等を実施する。

４．研究成果

(1) 構造化データに対する情報保存性に関する研究

問合せ保存性はデータ変換における情報保存性の一定式化である。変換（またはビュー）v が問合せ q を保存するとは、ある問合せ q' が存在して、任意のデータ t に対して q(t)=q'(v(t))を満たすことをいう。すなわち、ソースデータ t への問合せ q の結果を、ビューの結果 v(t)からも得ることができることを意味する。主にデータベース理論分野でデータベース統合に関連した問題として、問合せ保存性の判定可能性についての考察が盛んに行われている。本研究課題では、単値ボトムアップ木変換器に対する問合せ保存性問題の判定可能性について考察し、次に、頂点問合せに対する保存性問題の判定可能性について考察を行った。

① 単値ボトムアップ木変換器の問合せ保存性問題

v と q を木変換器とする。ある部分関数 q'があって、任意の t に対して q(t)=q'(v(t))が成り立つとき、v は q を保存する（または決定する）という。また、v が q を保存し q が q と同じ変換器のクラスに属するとき、v は q を包摂するという。本研究では、v が単値線形拡張ボトムアップ木変換（sl-xbot と略記）、q が単値ボトムアップ木変換器（s-bot と略記）で与えられるとき、保存性、包摂性の双方が coNEXPTIME で判定可能であることを示した。保存性の証明方針は次の通りである。

sl-xbot v, s-bot q が入力として与えられるとする。まず、v の逆変換を表す木変換器 v^{inv}を構成する。次に、q'' = q(v^{inv}) を満たす木変換器 q''を構成する。最後に、q''が単値であるかどうかを判定する。q''が単値であるときかつそのときに限り v は q を保存する。上の判定手続きにおいて、sl-xbot の逆変換およびそれらの合成を表すため、木変換器に grafting という機能を導入した。我々の知る限り、sl-xbot のクラスと s-bot のクラスという組み合わせは、保存性の判定に要する計算量の自明でない上界が明らかにされた最大のクラスである。

② 木変換器の頂点問合せに対する保存性問題

各頂点がデータ値をもつ木構造データ(データ木)を想定し、問合せをデータ木からデータ値の組の集合を返す関数とみなしたときの問合せ保存性について考察を行った。問合せ保存性に関する多くの既存研究はビューと問合せは同じ型の関数を前提としている。これに対し本研究では、ビューはデータ木からデータ木への決定性関数、問合せはデータ木からデータ値の組の集合を返す関数とし、問合せの強保存性と弱保存性を定義した。ビュー v が問合せ q を強保存するとは、ある問合せ q'が存在して、任意のデータ t に対して q(t)と q(v(t))が集合等価であることをいう。一方、v が q を弱保存するとは、ある q'が存在して、任意の t に対して q(t)が q'(v(t))に包含されることをいう。本研究では、ビューがデータ付き決定性線形トップダウン木変換器、問合せが木オートマトンの実行に基づく n 項問合せで与えられるとき、弱保存性問題は coNP-完全、強保存性問題は二重指数時間可解であることを証明した。また、ビューのクラスを先読み付き決定性線形トップダウン木変換器に拡張してもどちらの問合せ保存性も判定可能であることを示した。

(2) 構造化データに対する安全性に関する研究

アクセス制御はデータベース管理において不正なユーザからのアクセスを防ぐための最も重要な機能の一つである。アクセス制御を実現する典型的な方法は、許可された問合せ（許可問合せとよぶ）の集合と許可されない問合せ（不許可問合せとよぶ）の集合を切り分け、ユーザはデータベースインスタンスに対する許可問合せの結果しか得られないようにすることである。一見するとこのアクセス制御ポリシーで十分に思われるが、不正なユーザは許可問合せの結果や問合せのコード(意味)、その他に利用可能な外部情報を巧妙に利用することで、不許可問合せの結果(すなわち、機密情報)を得ることが可能である場合がある。このような攻撃を推論攻撃と呼ぶ。本研究課題では、XML データベースおよび関係データベースにおける推論攻撃に対する安全性の検証について、以下の 2 つの成果を得た。

① XML データベーススキーマに対する k-安全性問題

推論攻撃に対する安全性の尺度として k-安全性に着目し、XML データベーススキーマに対する k-安全性問題の判定可能性を考察した。直観的に、k-安全性は、許可問合せとそれらの結果などの利用可能な情報を用いて、インスタンスに対する不許可問合せの結果の候補の数が k-1 個以下に絞り込まれることがないことを意味する。本研究では、次のように定義されるスキーマ k-安全性の判定可能性について考察を行った：XML データベーススキーマと許可問合せ、不許可問合せが与えられたときに、そのスキーマに従うすべてのデータベースインスタンスが k-安全である。成果として、問合せが線形決定性トップダウン木変換器(dl-top)のシンプルなサブクラスで表現される場合でも、任意の有限の値 k>1 についてスキーマ k-安全性問題が判定不能であることを証明した。一方で、dl-top のクラスに対するスキーマ∞-安全問題が決定性指数時間完全であることを証明した。さらに、dl-top と同様に、正規先読み付きの dl-top に対してもスキーマ∞-安全性問題が判定可能であることを示した。

② 関係データベースにおける問合せに基づく l-多様性問題

関係データベースにおける l-多様性を、問合せのアクセス制御を考慮した場合に拡張して、推論攻撃に対するインスタンスレベルの安全性の概念を導入した。l-多様性は Machanavajjhala らによって導入された概念であり、準識別子の値によってインスタンスを同値類に分割したとき、どの同値類も秘匿属性の異なる値を l 個以上含むことをいう。しかしこの定義においては許可問合せが複数存在しそれらに対して推論攻撃が行われることが想定されていない。この問題点を解決するため、本研究では、問合せに基づく l-多様性と呼ばれるプライバシーの概念を提案した。データベースインスタンス t が許可問合せに関して l-多様性をもつとは、攻撃者がインスタンス t に対する許可問合せ結果とその問合せの意味を利用しても、機密情報の値の候補を l よりも少ない数までは絞り込むことができないことをいう。次に、この性質を判定する 2 つの手法を提案し評価実験によりその有効性について考察した。1 つ目の方法は、関係データベース管理システム、たとえば SQL を用いて秘匿属性値の種類を直接計数する手法であり、高速であるが問合せとして射影演算しか取り扱えない。2 つ目の方法は、入力を命題論理式に変換し、\#SAT ソルバを用いてその命題論理式のモデル計数を行うことによって判定を行うものである。この方法は実行時間が大きくなるものの、自己結合と否定以外の任意の関係演算を用いた問合せを取り扱えるため適用範囲が広いことが利点である。これらの 2 つの方法の有効性とスケーラビリティについて実験結果に

基づいて議論を行った。

(3) その他

① 木文法によって圧縮された構造化文書の直接問合せ・更新

木文法によって圧縮された XML 文書に対して、圧縮した状態で問合せや更新を行う手法を提案し、実装ツールを用いた実験結果に基づいて、提案手法の有効性を評価した。本研究は、Sebastian らが提案し実装している TreeRepair と呼ばれる圧縮法を前提としている。TreeRepair では Straight Line Context-Free Tree Grammar と呼ばれる木文法を用いて、XML 文書を圧縮する。本研究では、ボトムアップ木オートマトン(DBTA)に基づく頂点問合せ、または、トップダウン選択木オートマトン(DSTA)に基づく頂点問合せによって頂点位置を指定し、更新の場合は指定位置に対して置換、挿入、削除のいずれかを行う操作を前提とした。DBTA, DSTA のいずれにおいても、提案手法では、いったん解凍し問合せ・更新処理後、再圧縮を行う場合と比較して、約 15%～0.1%の実行時間、約 30%～0.1%のメモリ使用量で問合せ・更新を実行できることが分かった。

② ロールに基づく新しいアクセス制御法

ロールに基づくアクセス制御 (RBAC) では個人とロール間の関係は単一組織内で閉じており、多組織間で共有することができない。そこで本研究では、異なる組織のロール間関係だけを定義し、個人がどのロールをもつかを階層的 ID ベース暗号で認識できる仕組みを導入することにより、個人が所属する組織のロールを用いて他組織でアクセス制御を行う機構を提案した。

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３件）

(1) Chittaphone Phonharath, Kenji Hashimoto, and Hiroyuki Seki, Deciding Schema k-Secrecy for XML Databases, IEICE Transactions on Information and Systems, 査読有, Vol.E96-D, No.6, pp.1268-1277, June 2013.

(2) Kazuki Miyahara, Kenji Hashimoto and Hiroyuki Seki, Node Query Preservation for Deterministic Linear Top-Down Tree Transducers, IEICE Transactions on Information and Systems, 査読有, Vol.E98-D, No.3, pp.512-523, March 2015.

(3) Kenji Hashimoto, Ryuta Swada, Yasunori Ishihara, Hiroyuki Seki and Toru Fujiwara, Determinacy and Subsumption of Single-valued Bottom-up Tree Transducers, IEICE Transactions on Information and Systems, 査読有, Vol.E99-D, No.3, 575－587, March 2016.

〔学会発表〕（計１２件）

(1) Ramon Mejia, Yuichi Kaji and Hiroyuki Seki, Trans-Organizational Role-Based Access Control, ACM Computer and Communications Security (ACM CCS) 2011, Poster, Chicago, IL, Oct 17-21, 2011.

(2) Chittaphone Phonharath, Kenji Hashimoto and Hiroyuki Seki, Verification of the Security against Inference Attacks on XML Databases, 1st International Workshop on Trends in Tree Automata and Tree Transducers (TTATT 2012), pp.11-22, Nagoya, June 2, 2012.

(3) Kenji Hashimoto, Ryuta Sawada, Yasunori Ishihara, Hiroyuki Seki and Toru Fujiwara, Determinacy and Subsumption for Single-valued Bottom-up Tree Transducers, 7th International Conference on Language and Automata Theory and Applications (LATA 2013), Bilbao, Spain, April 2013, Lecture Notes in Computer Science 7810, pp.335-346.

(4) Kazuki Miyahara, Kenji Hashimoto, Hiroyuki Seki, Node Query Preservation for Deterministic Linear Top-Down Tree Transducers, 2nd International Workshop on Trends in Tree Automata and Tree Transducers (TTATT 2013), EPTCS 134, pp.27-37, Hanoi, Oct 19, 2013.

(5) Chittaphone Phonharath, Ryonosuke Takayama, Kenji Hashimoto and Hiroyuki Seki, Query-based l-diversity, 7th International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA 2015), pp.15-20, May 25, 2015. ISBN: 978-1-61208-408-4

(6) Chittaphone Phonharath, Kenji Hashimoto and Hiroyuki Seki, Static Analysis for k-secrecy against Inference Attacks, Korea-Japan Joint Workshop on Software Science and Engineering, June 2011.

(7) Hiroyuki Seki, Multiple Context-Free Grammars: Basic Properties and Complexity, the Second Workshop on Multiple Context-Free Grammars and Related Formalisms (MCFG+2), Nara,

Sept 2011.

(8) 宮原 一喜, 橋本 健二, 関 浩之, 決定性線形下降木変換器における頂点問合せ保存, 電子情報通信学会技術研究報告, 112(275), SS2012-38, 13-18, Nov 1, 2012.

(9) 尾上栄浩, 橋本健二, 関浩之, 木文法による圧縮 XML 文書に対する問合せと更新手法, 電子情報通信学会技術研究報告, 114(271), SS2014-28, 17-22, Oct 23, 2014.

(10) 後藤健志, 尾上栄浩, 橋本健二, 関浩之, 木文法に基づく圧縮 XML 文書に対する直接更新手法の評価, 電子情報通信学会技術研究報告, 114(416), SS2014-45, 73-78, Jan 27, 2015.

(11) Chittaphone Phonharath, Ryunosuke Takayama, Kenji Hashimoto and Hiroyuki Seki, Query-based l-diversity, IEICE Technical Report, 115(20), SS2015-14, 65-70, May 12, 2015.

(12) 石原 鷹, 橋本 健二, 関 浩之, 酒井 正彦, 拡張線形ボトムアップ木変換器の関数性の多項式時間判定,第 104 回情報処理学会・プログラミング研究会, 2015-1-(1), June 4, 2015.

〔その他〕

(1) 宮原 一喜, 橋本 健二, 関 浩之, 平成 24 年度電子情報通信学会ソフトウェアサイエンス研究会研究奨励賞, May 9, 2013.

(2) 後藤健志, 尾上栄浩, 橋本健二, 関浩之, 平成 26 年度電子情報通信学会ソフトウェアサイエンス研究会研究奨励賞,

(3) May 11, 2015.

６．研究組織

(1)研究代表者

関　浩之（HIROYUKI SEKI）
名古屋大学・大学院情報科学研究科・教授
研究者番号：８０１９６９４８

(2)研究分担者

小川　瑞史（MIZUHITO OGAWA）
北陸先端科学技術大学院大学・情報科学研究科・教授
研究者番号：４０３６２０２４

楫　勇一（YUICHI KAJI）
奈良先端科学技術大学院大学・情報科学研究科・准教授
研究者番号：７０２６３４３１

橋本　健二（KENJI HASHIMOTO）
名古屋大学・大学院情報科学研究科・助教
研究者番号：９０５４８４４７