様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ３１ 年　 ５ 月　 １ 日現在

機関番号：１３９０１
研究種目：基盤研究(B)（一般）
研究期間：2015～2018
課題番号：１５Ｈ０２６８４

研究課題名（和文）ソフトウェアセキュリティ・プライバシーのための静的解析・動的検査法

研究課題名（英文）Static Analysis and Dynamic Monitoring Methods for Software Security and Privacy

研究代表者
関　浩之（Seki, Hiroyuki）

名古屋大学・情報学研究科・教授

研究者番号：８０１９６９４８

交付決定額（研究期間全体）：（直接経費）　　13,800,000円

研究成果の概要（和文）：非決定性木変換器における問合せ保存性について考察した．木文法およびトップ木に基づく圧縮法，圧縮データを解凍することなく直接問合せおよび更新する手法を提案・実装した．文脈自由文法にデータ値を扱う能力を限定的に加えたレジスタ文脈自由文法(RCFG)について，所属問題と空問題の計算量がEXPTIME完全であることなどを示した．また，ガード式において任意の２項関係を指定できるようにRCFGを一般化した．重み付きレジスタオートマトン(WRA)に関する基本問題の計算量を考察し，最小重み経路問題に帰着して求めるアルゴリズムを提案した．さらに，マルウェアの意味理解やSMTソルバに関する基礎的研究を行った．

研究成果の学術的意義や社会的意義
セキュアなソフトウェアの設計運用技術を発展させるべく，ソフトウェア検証・解析などの形式手法からのアプローチが必要となっている．同時に，大規模構造化データが計算機システム上でますます蓄積，交換されるようになり，それら大規模データの安全で効率的な処理のための計算モデルの導入，基本問題を解くアルゴリズムの設計や計算量の解析，実用上効率的に動作するツールの開発等が求められている．本研究成果はこれらの要請に答えるものである．

研究成果の概要（英文）：We investigated query preservation of nondeterministic tree transducers. We proposed methods of compressing large trees and graphs based on tree grammars or top trees and directly evaluating a query on them without decompressing the compressed data. Register context-free grammar (RCFG) is an extension of CFG by adding limited power of manipulating data values. We showed that both the membership and emptiness problems for RCFG are EXPTIME-complete. We analyzed the computational complexity of basic problems for weighted register automata (WRA) and proposed an algorithm that computes a minimum-weight run of a given WRA. We also conducted a fundamental study on SMT solvers and an empirical study on understanding the semantics of malware.

研究分野：情報学

キーワード：セキュリティ　木オートマトン　木変換器　文法圧縮　マルウェア解析　SMTソルバ　レジスタオートマトン　レジスタ文脈自由文法

様　式　Ｃ－１９，Ｆ－１９－１，Ｚ－１９，ＣＫ－１９（共通）

１．研究開始当初の背景

ネットワーク社会における機密保持・プライバシー保護の問題には，それらに対する社会通念の変化，ビジネスロジックなどのシステムサービス要件，暗号の頑健性，ヒューマンエラーに対する組織的保全など様々な社会的および工学的要件が関わっている．これらの問題を個々に検討するのではなく，それらの要件を抽象化し，セキュアなソフトウェアの設計運用技術を発展させるべく，ソフトウェア検証･解析などの形式手法からのアプローチが必要となっている．同時に，大規模構造化データが計算機システム上でますます蓄積，交換されるようになり，それら大規模データの安全で効率的な処理のための計算モデルの導入，基本問題を解くアルゴリズムの設計や計算量の解析，実用上効率的に動作するツールの開発等が求められている．

２．研究の目的

情報システムの利便性と，システムのセキュリティ・ユーザのプライバシー保護との間のよりよいトレードオフを達成するため，ソフトウェアの設計運用技術の改善が求められている．そのためには可用性と安全性のトレードオフ関係を定量化すること，および，ソフトウェアリリース前の静的検査と稼働後の動的監視を適切に役割分担させて実用に耐える技術を開発することが必要である．本研究では，研究代表者らが過去四半世紀にわたって蓄積してきた形式言語理論とそのソフトウェア検証への応用研究の知見に基づき，セキュリティ保全・プライバシー保護の観点からソフトウェアの自動解析技術を理論的に深化させ，それに基づく解析系・自動生成系の実装を通して，提案手法の有効性を実証することを目的とする．

３．研究の方法

(1) 構造化文書のための木計算モデル：木オートマトンと木変換器を基本モデルとして用い，情報問合せ保存性という概念に基づき文書の機密保持と情報保存性のトレードオフについての理論的基盤を構築する．
(2) 構造化文書の圧縮法：データ圧縮は，単に記憶領域の節約のためでなく安全なデータ管理の基盤技術である．木文法に基づく圧縮法を前提にし，XML 文書や大規模グラフデータの圧縮法と直接問合せ法を提案し，ベンチマークを用いた実証実験を行う．
(3) マルウェア解析：マルウェアの意味理解のためのディスアセンブルに基づく解析を行う．
(4) データ値を扱う計算モデル：大規模構造化文書処理の基本モデルとして，有限オートマトンや形式文法にデータ値（整数値など）を扱う能力を限定的に加えた計算モデルが再注目されている．これらのモデルに関する基本問題の計算量の解析などを行う．
(5) その他：共通基盤技術としての SMT ソルバに関する研究，SMT ソルバを利用した有界モデル検査の事例研究などを行う．

４．研究成果

(1) 構造化文書のための木計算モデル

① 非決定性木変換器における問合せ保存性

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(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$v(t)$からも得られることを意味する．これまでに知られている問合せ保存性はビューが決定性関数であることを前提としている．そこで本研究では，ビューが非決定性関数であることも許すように定義した問合せ保存性について考察した．具体的に，非決定性ビューにおける問合せ保存性として，全称保存性と存在保存性を新たに定義した．そして，ビューが先読み付き拡張線形トップダウン木変換器の合成で与えられ，問合せが決定性単項２階木変換器で与えられる場合に，問合せ全称保存性が判定可能であることを証明した．さらに，問合せが非決定性関数である場合も考慮した問合せ保存の定義とその判定可能性についても議論した．

②XPath式から木オートマトンへの変換

XML文書に対する問合せ言語としてXPathが知られている．本研究ではXPathの部分クラスを定義し，そのクラスのXPath式を先読み付き決定性選択木オートマトンへ変換する手法を提案した．提案法ではXPath式の述語部分をボトムアップ先読み動作に変換し，式全体を先読み結果を用いたトップダウン動作に変換する．また，XPath式とともにXMLのスキーマ情報を入力として与え，そのスキーマに従うXML文書に対しては使われない無駄な規則を削減する変換法も提案した．提案法に基づく変換ツールの実装と評価実験も行った．

③いくつかの選択木オートマトンの表現能力の比較

非決定性選択木オートマトン(NSTA)は，非決定性木オートマトン(NTA)と頂点選択指定の対であり，与えられた木に対する受理実行において頂点選択指定を満たす頂点集合を問合せ結果とする簡潔な計算モデルである．任意の NTA を受理言語の等しい決定性ボトムアップ木オートマトン(DBTA)に変換できることはよく知られているが，与えられた NSTA を選択頂点集合の等しい決定性選択ボトムアップ木オートマトンに変換できるとは限らない(決定性選択トップダウン木オートマトンについても同様)．本研究では，ボトムアップ先読み付き決定性トップダウン選択木オートマトン(DSTAB)が NSTA と頂点選択能力が等価であることを示した．さらに XML 文書の

問合せ言語として知られる XPath の部分クラスから DSTAB への変換法の提案と変換ツールの実装をおこなった．DSTAB は，DBTA を先読みとして用い，それが割当てた状態を参照しながら決定的トップダウンに動作する選択木オートマトンである．NSTA と DSTAB の等価性を示すために，与えられた NSTA からそれと等価な DSTAB への変換法とその正当性の証明，およびその逆方向の変換法とその正当性の証明を与えた．ここで，等価とは実行結果の選択頂点集合が等しいことを意味する．NSTA から変換する DSTAB の先読み部は NSTA の部分集合構成によって構成する．トップダウン部は，先読み部で根頂点に割り当てられた受理状態を開始状態とし，先読み部の受理状態への遷移ではその頂点に割り当てられない状態を消すようにして状態を割り当てる．また，DSTAB の先読み部の順序を逆にしたトップダウン先読み付き決定性選択ボトムアップ木オートマトンでは NSTA を模倣できないことがある例も示した．

(2) 構造化文書の文法圧縮法と圧縮データの直接更新法

XMLは構造化データを容易に記述できる文書形式であり高い汎用性をもつが，実用的なXML文書はサイズが膨大になりやすい．そこでXML文書は圧縮された状態で保存するのが望ましい．XML文書はタグによる階層構造をなしているため，木で表現することができる．様々な木圧縮法が研究されているが，ここでは，木文法およびトップ木に基づく圧縮法を前提とし，圧縮データを解凍することなく直接問合せおよび更新する手法についていくつかの研究成果を得た．さらにこれらの発展として，大規模グラフデータの木への分解による圧縮法の開発も行った．

①木文法に基づく圧縮XML文書に対するデータ値を考慮した直接更新手法

XML文書に対する様々なデータ圧縮法が提案されている．その中で木文法に基づく圧縮法は，圧縮した状態で走査可能であるという特徴をもつ．本研究では，直線的文脈自由木文法によって圧縮されたXML文書を，解凍を行わずに更新する手法を提案した．更新箇所の指定には決定性選択トップダウン木オートマトンを用い，更新操作としてはタグ名の変更，部分構造の削除，部分構造の挿入のいずれかを指定する．また，自然言語処理用コーパスTreebank，DBLPを含む実用規模のXML文書を用いて行った実験結果に基づき，提案手法の有効性を示した．

次に，上記の圧縮法および直接更新法を拡張し，構造情報だけでなく，データ値の参照，およびそれらに依存した構造やデータ値の更新が可能な手法を提案した．具体的に，データ値をもつデータ木を導入し，木文法に基づく圧縮法を用いてデータ木の木構造部を圧縮したとき，圧縮文書に対して直接問合せおよび更新を行う手法を提案した．圧縮文書における重複計算の回避やデータ値を格納するファイルの適切な分割により高速化を図った結果，既存のXML DBMSであるBaseXよりも時間，領域両面から効率よく処理が行えることを実証した．

②トップ木に基づくデータ圧縮法および問合せ処理法

構造化データの圧縮法としてトップ木を用いる方法が提案されている．この方法は木文法を用いる方法と比較して圧縮率では劣る一方，圧縮データの直接問合せが容易であるという利点がある．本研究では圧縮時に行う併合と呼ばれる操作におけるクラスタの選択順序に着目し，簡単な基準でクラスタ対を併合する手法と，出現数の多い隣接クラスタ対から優先して併合する手法とを実装した．さらに，圧縮データを解凍することなく直接問合せ処理する手法を実装し，実データに対して評価実験を行った．

次に，トップ木に基づく圧縮データであるトップDAGに対して，(1)頂点選択法，(2)更新法の2つを提案した．まず(1)について，トップDAGと選択木オートマトンが与えられたとき，トップDAGを解凍することなく深さ優先順で巡回し，選択位置を取得するアルゴリズムを提案した．選択木オートマトンはトップダウンなDSTTAとボトムアップなDSBTAの2つで，それぞれ先祖，子孫を条件とした頂点選択を行える．次に(2)について，トップDAGと更新位置と更新操作（ラベル変更，木挿入，木削除，頂点削除）が与えられたとき，トップDAGを解凍することなく深さ優先順で巡回し，指定された位置に更新を適用するアルゴリズムを提案した．提案手法に基づいて実装したシステムを用いていくつかのXML文書に対して評価実験を行った．更新に要した時間・メモリ使用量を計測し，素朴な更新法（圧縮データを一度解凍し，元のデータを更新し，再圧縮）と比較して提案した直接更新法はより100倍高速かつ1/5のメモリ使用量で実行できることが分かった．さらに，代表的なXMLデータ管理システムであるBaseX上での更新との比較実験を行ったところ，提案した直接更新法の方が30倍高速かつ1/40のメモリ使用量で実行できることが分かった．

③　木文法に基づくグラフデータの圧縮

最後に，木よりさらに一般的なデータ構造であるグラフに着目した．既に様々なグラフ圧縮の研究が行われているが，圧縮された有向グラフに対する直接問合せ評価法については報告されていない．本研究では有向グラフに対して合流頂点分解という手法を導入することで辺の向きも保存するような，直線的文脈自由木文法に基づく圧縮法と，圧縮グラフに対して解凍することなく問合せを行う手法を提案し，提案手法に基づくツールの試作し実験的評価を行った．

(3) マルウェア解析

マルウェアは対象となる機器により，PC マルウェアと IoT マルウェアに大別される．前者はプラットフォームは x86 などに限られるが（アンチウィルスソフトをかいくぐるため）制御構造隠蔽手法を多く用いる．後者は，アップデート等が稀なため制御構造隠蔽手法は殆ど用いられず単純であるが，プラットフォームは 20 種以上と多岐にわたる．本研究では前者を対象として，バイナリ記号実行器 BE-PUM の実装を進めた．（国際会議 FPS15）自己改変やジャンプ先の直接

書換えなどの手法のため，PC マルウェアにおいては，通常のプログラミング言語では構文解析により直接得られるコントロールフローグラフ（モデル）生成が非常に困難となる．そのため，動的記号実行を用いて漸増的モデル生成を行うことで PC マルウェアのディスアセンブリ・コントロールフローグラフ生成が可能となることを示した．その応用として，現在 85%以上の PC マルウェア作成に用いたれるといわれるパッカーの同定のために，個別の制御構造隠蔽手法の形式的定義を与え，その検出を通して CFFexplorer，PEiD や VirusTotal などの商用サイト・ツールにおけるパッカー同定の誤りを 13000 弱の PC マルウェア中，約 400 発見した．（国際会議 SSPREW17）

(4) データ値を扱う計算モデル

① レジスタ付き文脈自由文法

有限オートマトンにデータ値を扱う能力を加える試みが数多くなされているが，データ値の演算能力を加えると計算能力がチューリング機械と等価まで上がってしまうことが知られている．そこでデータ値を扱う能力を抑え，所属問題などの判定問題や受理言語クラスの閉包性に関するよい性質を持つようなクラスへの拡張が試みられている．その一つとして，レジスタオートマトン(register automaton，RA と略)が知られている．RA は所属問題，空問題が判定可能であり，それらの計算量も解析されている．RA は，XML などの大規模構造化データへの問い合わせ言語のモデルとして注目されており，木構造データの経路に関する問合せを表現可能であるが，一方で XPath のように部分木に関する問合せを記述する能力を有していない．そこで，RA の文脈自由文法，木オートマトンへの拡張がなされ，それぞれ，レジスタ付き文脈自由文法(RCFG)，レジスタ木オートマトン(RTA)と呼ばれている．RCFG や RTA はデータ値を含む構造化データに対する質問言語のモデルとして有用であり，所属問題，空問題が判定可能であることが知られているが，その計算量については知られていない．本研究では，一般の RCFG，ε規則なし RCFG，成長的 RCFG に対する所属問題が順に，EXPTIME 完全，PSPACE 完全，NP 完全であること，空問題はこれらどのクラスについても EXPTIME 完全であることを示した．また，RTA についてはこれら２つの問題の計算量が NP 完全，EXPTIME 完全であることを示した．

次に，レジスタ値を抽象化したレジスタ型という概念を導入し，これを利用して，RCFG に対するε規則除去法を示した．さらに，RCFG においてガード条件に記述できる比較演算を等号から一般の関係に拡張した一般化 RCFG(GRCFG)を提案し，GRCFG に対する所属問題，空問題が判定可能となるような十分条件を与えた．

②重み付きレジスタオートマトン

重みによって状態遷移やデータ値操作に要するコストを表現するため，RA に重みを導入した重み付き RA(WRA)が提案されている．本研究では，WRA に関する実行重み計算問題，重み実現可能性問題がそれぞれ，NP 完全，PSPACE 完全であることを示し，さらに，レジスタ型のみで演算重みが確定する WRA のある部分クラスを導入し，このクラスに属する WRA の重み最小実行を，有向グラフの最小重み経路問題に帰着して求めるアルゴリズムを提案した．

(5) その他

①有界モデル検査法などの事例研究

有界モデル検査法の事例研究として，背景理論付き命題論理式の重み付き最大充足可能性(weighted partial Max-SAT)ツールを用いて，自動車の電子制御ソフトウェアのハザード解析および時系列解析を行った．

プログラムセキュリティの定量的尺度である動的情報漏洩量を，抽象解釈法を用いて近似計算するアルゴリズムの提案やその計算量解析を行い，さらに，適応的制御入力最適化への拡張も試みた．

オープンソースソフトウェア開発における大規模ソフトウェアメトリクスの自動収集を行い，パッチの特徴量に対する統計的解析に基づき，ソフトウェア品質の予測，具体的に再投稿要求の有無の予測実験を行った．

②実数上の非線形制約（多項式制約）を対象とした SMT ソルバ開発・実装

SMT ソルバには，線形制約・アレイ・ビット演算・関数の構文解釈など，様々な背景理論があり，多くは標準的な実装法が定まっている．近年，ループ不変式生成やマルウェアの opaque predicate（複雑だが常に真偽値が変わらない述語）などの応用において，非線形制約が注目されている．非線形制約は，整数上ではヒルベルトの第１０問題として知られ，決定不能問題として知られている．実数上では CAD (Cylindrical Algebraic Decomposition) を用いることで決定可能となることが知られているが，二重指数時間となる．SMT ソルバにおいて実数上の非線形制約はいまだ標準的実装法が定まっておらず，CAD のほか，区間演算，ビットコーディング，線形制約による近似など様々な手法が乱立している．本研究では，主に区間演算とテスト手法を組み合わせた SMT ソルバ raSAT の開発・実装を進め，２０１４年より国際競技会の SMT-COMP (www.smtcomp.org) QFNRA 部門(Quantifier-free nonlinear arithmetic over reals)に参加を続け，２０１４年４位，２０１５年３位，２０１６～８年２位の成績をおさめた．（雑誌論文②）特に，2018 年はロレーヌ大学 LORIA と共同で，Pascal Fontaine 准教授が開発する SMT ソルバフレームワーク VeriT と raSAT を融合し，参加した．

５．主な発表論文等

〔雑誌論文〕（計３件）

① Kenji Hashimoto, Ryunosuke Takayama and Hiroyuki Seki, Direct Update of XML Documents with Data Values Compressed by Tree Grammars, IEICE Transactions on Information and Systems, 査読有, Vol.E101-D, No.6, pp.1467-1478, June 2018.
DOI: 10.1587/transinf.2017FOP0002

② Vu Xuan Tung, To Van Khanh and Mizuhito Ogawa, raSAT: An SMT Solver for Polynomial Constraints, Formal Methods in System Design, 査読有, Vol.51, No.3, pp.462-499, 2017.
DOI: 10.1007/s10703-017-0284-9

③ Kazuki Miyahara, Kenji Hashimoto and Hiroyuki Seki, Query Rewriting for Nondeterministic Tree Transducers, IEICE Transactions on Information and Systems, 査読有, Vol.E99-D, No.6, 1410-1419, June 2016.
DOI: 10.1587/transinf.2015FOP0007

〔学会発表〕（計２５件）

① Ryoma Senda, Yoshiaki Takata and Hiroyuki Seki, Generalized Register Context-Free Grammars, 13th International Conference on Language and Automata Theory and Applications (LATA 2019), 査読有, Sankt Petersburg, LNCS 11417, pp.259-271, March 2019.

② Takeshi Takeda, Kenji Hashimoto and Hiroyuki Seki, Graph Compression by Tree Grammars and Direct Evaluation of Regular Path Query, 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS 2019), 査読有, Singapore, CD104, pp.257-261, Feb 2019.

③ Ryoma Senda, Yoshiaki Takata and Hiroyuki Seki, Complexity Results on Register Context-Free Grammars and Register Tree Automata, 15th International Colloquium on Theoretical Aspects of Computing (ICTAC 2018), 査読有, Stellenbosch, LNCS 11187, pp.415-434, Oct 2018.

④ Nguyen Minh Hai, Mizuhito Ogawa and Quan Thanh Tho, Packer Identification Based on Metadata Signature, 7th Software Security, Protection, and Reverse Engineering Workshop (ACM SSPREW 2017), 査読有, Dec 2017.

⑤ Pascal Fontaine, Mizuhito Ogawa, Thomas Sturm and Xuan Tung Vu, Subtropical Satisfiability, 11th International Symposium on Frontiers of Combining Systems (FroCoS 2017), 査読有, LNCS 10483, pp.189-206, Aug 2017.

⑥ Shuichi Sato, Shogo Hattori, Hiroyuki Seki, Yutaka Inamori and Shoji Yuen, Automating Time Series Safety Analysis for Automotive Control Systems in STPA using Weighted Partial Max-SMT, 5th International Workshop on Formal Techniques for Safety-Critical Systems (FTSCS 2016), 査読有, pp.39-54, Tokyo, Nov 2016.

⑦ Kenji Hashimoto, Suguru Nishimura and Hiroyuki Seki, Direct Evaluation of Selecting Tree Automata on XML Documents Compressed with Top Trees, 4th International Workshop on Trends in Tree Automata and Tree Transducers (TTATT 2016), 査読有, pp. 29-36, Seoul, South Korea, July 2016.

⑧ Vu Xuan Tung, To Van Khanh and Mizuhito Ogawa, raSAT: An SMT Solver for Polynomial Constraints, 8th International Joint Conference on Automated Reasoning (IJCAR 2016), 査読有, LNCS 9706, pp.228-237, June-July 2016.

⑨ Nguyen Minh Hai, Mizuhito Ogawa and Quan Thanh Tho, Obfuscation Code Localization Based on CFG Generation of Malware, 8th International Symposium on Foundations and Practice of Security (FPS 2015), 査読有, LNCS 9482, pp.229-247, Oct 2015.

⑩ Shogo Hattori, Shoji Yuen, Hiroyuki Seki and Shuichi Sato, Automated Hazard Analysis

with pMAX-SMT for Automobile Systems, 15th International Workshop on Automated Verification on Critical Systems (AVOCS 2015), 査読有, Edinburgh, U.K., Sept 2015.

〔その他〕
ホームページ等
https://sites.google.com/a/sqlab.jp/sk-lab/publication

６．研究組織

(1)研究分担者
研究分担者氏名：小川　瑞史
ローマ字氏名：(OGAWA, Mizuhito)
所属研究機関名：北陸先端科学技術大学院大学
部局名：先端科学技術研究科
職名：教授
研究者番号（8桁）：４０３６２０２４

研究分担者氏名：結縁　祥治
ローマ字氏名：(YUEN, Shoji)
所属研究機関名：名古屋大学
部局名：大学院情報学研究科
職名：教授
研究者番号（8桁）：７０２３０６１２

研究分担者氏名：橋本　健二
ローマ字氏名：(HASHIMOTO, Kenji)
所属研究機関名：名古屋大学
部局名：大学院情報学研究科
職名：助教
研究者番号（8桁）：９０５４８４４７