様式Ｃ－１９

# 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成２４年　5月 15 日現在

機関番号：12102
研究種目：基盤研究（C）
研究期間：2009 ～ 2011
課題番号：21500028
研究課題名（和文）　文字列解析に基づくウェブソフトウェアの検証

研究課題名（英文）　Verification of Web Software Based on String Analysis

研究代表者
南出　靖彦（MINAMIDE YASUHIKO）
筑波大学・システム情報系・准教授
研究者番号：50252531

研究成果の概要（和文）：
ウェブソフトウェアの脆弱性検査に応用される文字列解析と呼ばれるプログラム解析を以下の点で改良した．検査結果として得られる反例を特殊な文脈自由文法として構成することで，反例の可読性を高めた．スクリプト言語で重要な役割を果たす正規表現マッチングの意味論をモナドを用いて定式化し，その意味論からトランスデューサを構成することで，正規表現マッチングの正確な解析を可能とした．データベースと連携するサーバサイドプログラムの検査おいては，データベースに関する制約を求めることで解析の精密を大きく向上した．

研究成果の概要（英文）：
We have improved a program analysis called string analysis that can be applied to the detection of Web software vulnerabilities in the following respects. The readability of counter examples generated by the analysis is improved by constructing them as context-free grammars in a specific form. We have formulated the semantics of regular expression matching in programming languages, and enabled their precise analysis through the precise translation to transducers. For the analysis of a server-side program utilizing a database, we have improved its analysis by analyzing the constraint on data imposed by the program storing the data.

交付決定額

（金額単位：円）

| | 直接経費 | 間接経費 | 合　計 |
|---|---|---|---|
| 2009 年度 | 1,200,000 | 360,000 | 1,560,000 |
| 2010 年度 | 1,100,000 | 330,000 | 1,430,000 |
| 2011 年度 | 1,000,000 | 300,000 | 1,300,000 |
| 総　計 | 3,300,000 | 990,000 | 4,290,000 |

研究分野：総合領域
科研費の分科・細目：情報学、ソフトウェア
キーワード：プログラム処理系，ソフトウェア検証，ウェブ

## １．研究開始当初の背景

ウェブソフトウェアは，ウェブブラウザで解釈・実行される HTML 及び JavaScript，データベース上で実行される SQL 問合わせを，サーバ側プログラムで動的に生成する技術によって非常に柔軟性なシステムとなっている．しかし，動的にスクリプトを生成することによって得られる柔軟性は，ウェブシステムの構成要素（サーバ，ブラウザ，データベース）間のインタラクションの不整合の原因となることがある．動的に生成される SQL 問合わせで生じる不整合は，SQL インジェクションと呼ばれる脆弱性の原因となり，動的に生成されるウェブページでの不整合は，ク

ロスサイトスクリプティング脆弱性の原因となっている．これらの脆弱性が，ウェブソフトウェアの信頼性を損なう重大な問題となっている．
このようなウェブソフトウェアの信頼性における問題を解決するために，本研究の代表者はプログラム解析の技術を用いて，ウェブソフトウェアを検証する研究を行ってきた．先行研究では，検証の基礎として，プログラムの文字列出力を文脈自由文法を用いて保守的に近似するプログラム解析を開発した（WWW2005）．このプログラム解析をサーバサイドプログラムに適用することで，生成されうるウェブページの近似を得ることができ，サーバサイドプログラムの脆弱性の検出や生成されるウェブページの妥当性検証（プログラムが常に文法的に正しいウェブページを生成するかの検証）が可能になる．

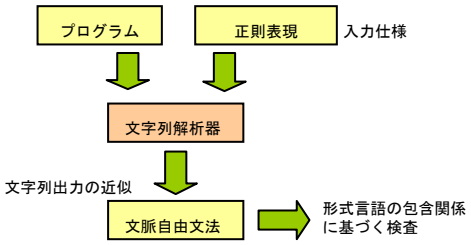


図 1: 文字列解析による
ウェブソフトウェアの検査

上の図１が文字列解析によるウェブソフトウェアの検査の概略を示したものである．プログラムと入力の仕様を表す正則表現から生成されうるウェブページを近似した文脈自由文法が得られる．その文法に対して，形式言語の包含関係の検査を適用することで，脆弱性等の検査が可能となっている．

２．研究の目的
文字列解析によりウェブソフトウェアを検証する基礎的な技術を開発してきたが，実用レベルの検証技術としては未成熟であり，解析の精度不足による問題（脆弱性）の誤検知，クロスサイトスクリプティングなど複雑な脆弱性の検出などが問題となっている．本研究は，これらの問題を解決し，文字列解析の技術に基づく実用的なウェブソフトウェア検証システムを構築することを目指す．また，ウェブソフトウェア検証の基礎となるスクリプト言語の意味論，HTML の仕様の形式化の研究も行う．以下の点について重点的に研究を行う．

（１）Wassermann らによる文字列解析と情報流解析を融合したプログラム解析は，クロスサイトスクリプティング脆弱性の検出にも応用できる．しかし，クロスサイトスクリプティングでは，SQL インジェクションよりもはるかに複雑な攻撃が考えられ，Wassermann らの方法のみでは十分な精度で脆弱性を検出できない．本研究では，Wassermann らの方法と妥当性検証の技術を組み合わせることで，クロスサイトスクリプティング脆弱性を検出する技術を開発する．

（２）サーバサイドプログラムとデータベースとのインタラクションが間接的な要因となるプログラムの誤りや脆弱性がある．データベースに保存されたデータが原因となって生じる保存型クロスサイトスクリプティングが，その典型的な例である．このような問題に対する検証を実現するために，データベースとのインタラクションを精密に考慮したサーバサイドプログラムの検証技術を開発する．

（３）文字列解析に基づくプログラムの検証の実用性は，解析の精度に大きく依存する．モデル検査や抽象解釈の研究で開発されてきた技術を応用することで，文字列解析の精度を向上することを目指す．

３．研究の方法
ウェブソフトウェア検証の基礎技術として，研究代表者は，本研究の基礎となる文脈自由言語に基づく文字列解析，文字列解析による妥当性検証の技術を開発し，実験レベルではウェブソフトウェアの妥当性の検証や脆弱性の検出に成功している．本研究では，代表者のこれまでの研究を発展し，形式言語理論及びプログラム解析の理論を駆使し，文字列解析によるウェブソフトウェアの実用的な検証を可能にする理論・技術を開発する．また，理論・技術の実証のため，先行研究で開発した PHP 文字列解析器をベースとしたウェブソフトウェア検証システムの開発を並行して進める．

現在の PHP 文字列解析器は，次の三点で実用性に問題があり，改善の必要がある．

- クロスサイトスクリプティング脆弱性を検出する手法が，十分に確立されておらず，単純な攻撃しか検出できない．
- 文字列解析の精度が不十分でありプログラムの問題（脆弱性）が誤検出されことが多い．
- 検出された問題が誤検出であるかを検証者が判断する場合に，それを支援する機能がない．

４．研究成果
文字列解析に基づくウェブソフトウェアの

検証に関して以下の研究を行った.
(1) 検査結果として得られる反例を特殊な文脈自由文法として構成することで，反例の可読性を高めた.検出した脆弱性を示す反例 HTML 文書を唯一の文字列を生成する文脈自由文法（Straight Line Program）として生成し，反例のどの部分がプログラムのどの部分で生成されたかインタラクティブに調べられるようにした．これにより文字列解析によって検出された脆弱性が真の脆弱性かを効率的に確かめられるようになった．この技術を実装した反例インスペクタを図２に示す.

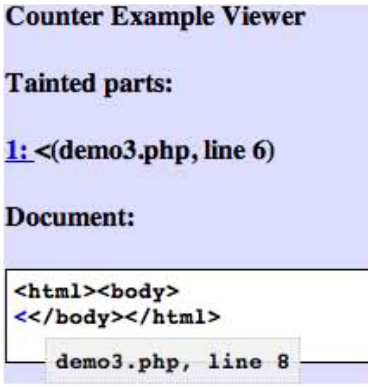


図 2：SLP に基づく反例生成

demp3.php, line8 の部分は，マウスポインタが指している部分がプログラムのどの部分から生成されたかを示している.

（２）Ruby の操作的意味論及び制御フロー解析の研究を行った．クラスやメソッドの定義，イテレータブロック，大域脱出など，主要な言語機能を含むサブセットに対し，操作的意味論を与えた．さらに，文字列解析などの高度なプログラム解析の基礎となる，動的なメソッド定義を精密に解析する制御フロー解析を開発した．制御フロー解析の結果に大きな影響を与えるメソッド定義について，フロー依存な解析を行うことで高い精度の解析を可能にした．この制御フロー解析の健全性を，操作的意味論に基づき証明した.

（３）文字列解析を応用したクロスサイトスクリプティング脆弱生検査を，いくつかの観点から改良した．データベースと連携するサーバサイドプログラムの検査おいては，データベースのスキーマ及びデータを格納するプログラムから，データベースに関する制約を求めることで，これまでに比べ格段に精密な解析が可能になった．また，クロスサイトスクリプティング脆弱生の検査と反例生成の仕組みを，トランスデューサを用いて再構成し柔軟な検査を可能とした．これらの改良を PHP 文字列解析器上に実装し，評価を行った.

（４）スクリプト言語で記述されたプログラムでは，正則表現を用いたマッチングや文字列の置き換えが重要な役割を果たす．既存の文字列解析では，このような文字列操作をマッチングの戦略を無視した近似を用いることで解析を行っていた．そのため，このような操作に対して高い精度の解析結果が得られていなかった．本研究では，戦略を考慮にいれた正則表現マッチングの意味論の定式化をモナドの概念を用いて行い，その定式化に基づき出力付きオートマトンを構成する方法を与えた．さらに，モナド射やモナドトランスフォーマを用いて，より洗練された意味論を構築した．モナドトランスフォーマを用いて意味論を段階的に詳細化することで，出力付きオートマトンの構成法に関して，見通しの良い正当性証明を与えることができた.

（５）HTML5 構文仕様を形式化する研究を行った．HTML5 構文仕様は，スタックを用いた複雑なアルゴリズムとして自然言語により記述されている．本研究では，遷移の条件としてスタックの内容を正則言語で検査できる条件付きプシュダウンオートマトンとして，このアルゴリズムの形式化を行った．この形式化は，文脈自由言語の HTML5 構文に対する妥当性検査アルゴリズムの設計の基礎となる.

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２件）
① Yuto Sakuma, Yasuhiko Minamide, Andrei Voronkov Translating Regular Expression Matching into Transducers, Journal of Applied Logic, 査読有, Vol.10, 2012, pp.32-51. DOI: 10.1016/j.jal.2011.11.003
② 松本宗太郎, 南出靖彦, Ruby プログラムの制御フロー解析とその健全性の証明, 情報処理学会論文誌:プログラミング, 査読有, Vol.3, 2010, pp.9-25. URL:http://id.nii.ac.jp/1001/00068444/

〔学会発表〕（計９件）
① 森 俊介, 南出 靖彦, HTML5 構文解析のプッシュダウンオートマトンを用いた検証, プログラミングおよびプログラミング言語ワークショップ （ポスタ

ー), 2012 年 3 月 8 日, 南紀白浜むさし(和歌山県)

② 武井 裕也, 南出 靖彦, 抽象 DPLL の Isabelle/HOL による形式化と検証, プログラミングおよびプログラミング言語ワークショップ (ポスター), 2012 年 3 月 8 日, 南紀白浜むさし(和歌山県)

③ 木村 将人, 南出 靖彦, 文字列解析によるクロスサイトスクリプティング脆弱性検査の改良, プログラミングおよびプログラミング言語ワークショップ(ポスター), 2012 年 3 月 8 日, 南紀白浜むさし(和歌山県)

④ Yasuhiko Minamide, Semantics and Implementations of Regular Expression Matching, The Eighth Asian Workshop on Foundation of Software, 2011 年 5 月 13 日, 上海交通大学(中国)

⑤ Yasuhiko Minamide, Formalizing Regular Expression Matching in Isabelle/HOL, TPP'10: 6th Theorem Proving and Provers Meeting, 2010 年 11 月 26 日, 名古屋大学(愛知県)

⑥ Yasuhiko Minamide, Yuto Sakuma, Andrei Voronkov, Translating Regular Expression Matching into Transducers, 12th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, Timisoara, Romania, September 23-26, 2010

⑥ Yasuhiko Minamide, Formalizing Regular Expression Matching in Isabelle/HOL, TPP'10: 6th Theorem Proving and Provers Meeting, 2010 年 11 月 26 日, 名古屋大学(愛知県)

⑦ Yasuhiko Minamide, The PHP String Analyzer, IFIP Working Group 2.8, April 13, 2010, Shirahama, Japan

⑧ 藤原拓也, 南出靖彦, 証明支援系 Isabelle/HOL によるごみ集めアルゴリズムの形式化と安全性検証, 日本ソフトウェア科学会第 26 回大会, 2009 年 9 月 16 日, 島根大学(島根県)

⑨ 松本宗太郎, 南出靖彦, Ruby のコア言語の操作的意味論, 日本ソフトウェア科学会第 26 回大会, 2009 年 9 月 16 日, 島根大学(島根県)

〔その他〕
ホームページ
http://www.score.cs.tsukuba.ac.jp/~minamide/phpsa

南出靖彦, 講義「Web プログラムの脆弱性と静的検査」, 2012 年 1 月 5 日, お茶の水女子大学理学部情報科学科

南出靖彦, 特別講義「Web プログラムの脆弱性とその自動検査」, 2010 年 10 月 15 日, 香川大学工学部信頼性情報システム工学科

６．研究組織
(1)研究代表者
南出　靖彦　(MINAMIDE YASUHIKO)
筑波大学・システム情報系・准教授
研究者番号：５０２５２５３１