

# String Analysis for Software Verification and Security

Tevfik Bultan • Fang Yu • Muath Alkhalaf  
Abdulbaki Aydin

# String Analysis for Software Verification and Security

Tevfik Bultan  
Department of Computer Science  
University of California, Santa Barbara  
Santa Barbara, CA, USA

Muath Alkhalaf  
Computer Science Department  
King Saud University  
Riyadh, Saudi Arabia

Fang Yu  
Department of Management  
Information Systems  
National Chenchi University  
Taipei, Taiwan

Abdulkaki Aydin  
Microsoft (United States)  
Redmond, WA, USA

ISBN 978-3-319-68668-4      ISBN 978-3-319-68670-7 (eBook)  
<https://doi.org/10.1007/978-3-319-68670-7>

Library of Congress Control Number: 2017956344

© Springer International Publishing AG 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature  
The registered company is Springer International Publishing AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

This monograph is mainly based on the research that has been conducted in the Verification Laboratory at the University of California, Santa Barbara, in the last decade. String analysis has been an interesting and fruitful area to work on, leading to many research results some of which are discussed here. We observe that the research in analysis of string manipulating code is expanding due to the importance of the correctness of the string manipulation code for dependability and security of modern software systems. We hope that this monograph can inspire and motivate more research in this area and accelerate the transition of string analysis research to practice.

We would like to thank all current and past members of the Verification Laboratory for their help and support. We would also like to acknowledge the support provided by the National Science Foundation (NSF) and the Defense Advanced Research Projects Agency (DARPA) for the string analysis research at the Verification Laboratory.<sup>1</sup>

Santa Barbara, CA, USA  
Taipei, Taiwan  
Riyadh, Saudi Arabia  
Redmond, WA, USA

Tevfik Bultan  
Fang Yu  
Muath Alkhalaf  
Abdulkaki Aydin

---

<sup>1</sup>NSF under grants CCF 0916112, CNS-1116967, CCF-1423623, and CCF-1548848, and DARPA under agreement number FA8750-15-2-0087. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or the U.S. Government.

# Contents

<b>1</b>	<b>Introduction</b>	1
1.1	Common Vulnerabilities Due to String Manipulation Errors	4
1.2	Examples of String Manipulating Code and Errors	6
1.3	Overview	12
<b>2</b>	<b>String Manipulating Programs and Difficulty of Their Analysis</b>	15
2.1	A Simple String Manipulation Language	15
2.2	Automated and Precise Verification of String Programs Is Not Possible	16
2.3	A Richer String Manipulation Language	18
2.4	Summary	22
<b>3</b>	<b>State Space Exploration</b>	23
3.1	Semantics of String Manipulation Languages	23
3.2	Explicit State Space Exploration	26
3.2.1	Forward Reachability	27
3.2.2	Backward Reachability	28
3.3	Symbolic Exploration	29
3.3.1	Symbolic Reachability	30
3.3.2	Fixpoints	31
3.4	Summary	35
<b>4</b>	<b>Automata Based String Analysis</b>	37
4.1	A Lattice for Sets of Strings	37
4.2	Symbolic Reachability Analysis with Automata	38
4.3	Symbolic Automata Representation	41
4.3.1	MTBDD Representation	43
4.3.2	Modeling Non-Determinism Using Symbolic DFA	44
4.3.3	Symbolic vs. Explicit DFA	45
4.4	Post-Condition Computation	45
4.4.1	Concatenation and Replace Operations	46

4.4.2	Post-Condition of Concatenation .....	47
4.4.3	Post-Condition of Replacement .....	48
4.5	Pre-Condition Computation .....	51
4.5.1	Pre-Condition of Concatenation .....	51
4.5.2	Pre-Condition of Replacement .....	53
4.6	Summary .....	55
<b>5</b>	<b>Relational String Analysis .....</b>	<b>57</b>
5.1	Relations Among String Variables .....	57
5.2	Multi-Track DFAs .....	59
5.3	Relational String Analysis .....	59
5.3.1	Word Equations .....	60
5.4	Construction of Multi-Track DFAs for Basic Word Equations ....	61
5.5	Symbolic Reachability Analysis .....	65
5.5.1	Forward Fixpoint Computation .....	66
5.5.2	Summarization .....	68
5.6	Summary .....	68
<b>6</b>	<b>Abstraction and Approximation .....</b>	<b>69</b>
6.1	Regular Abstraction .....	69
6.2	Alphabet Abstraction .....	70
6.3	Relation Abstraction .....	74
6.4	Composing Abstractions .....	76
6.5	Automata Widening Operation .....	77
6.6	Summary .....	80
<b>7</b>	<b>Constraint-Based String Analysis .....</b>	<b>83</b>
7.1	Symbolic Execution with String Constraints .....	83
7.2	Automata-Based String Constraint Solving .....	87
7.2.1	Mapping Constraints to Automata .....	87
7.3	Relational Constraint Solving with Multi-Track DFA .....	93
7.3.1	Relational String Constraint Solving .....	94
7.3.2	Relational Integer Constraint Solving .....	95
7.3.3	Mixed String and Integer Constraint Solving .....	95
7.4	Model Counting .....	96
7.4.1	Automata-Based Model Counting .....	99
7.4.2	Counting All Solutions within a Given Bound .....	101
7.5	Summary .....	102
<b>8</b>	<b>Vulnerability Detection and Sanitization Synthesis .....</b>	<b>103</b>
8.1	Vulnerability Detection and Repair .....	104
8.2	Patching Algorithm .....	111
8.3	Vulnerability Analysis .....	112
8.4	Vulnerability Signatures .....	114
8.5	Relational Signatures .....	116
8.6	Sanitization Generation .....	119
8.7	Summary .....	122

<b>9</b>	<b>Differential String Analysis and Repair</b> .....	123
9.1	Formal Modeling of Validation and Sanitization Functions.....	126
9.1.1	Post- and Pre-Image of a Sanitizer.....	130
9.2	Discovering Client- and Server-Side Input Validation and Sanitization Inconsistencies.....	133
9.2.1	Extracting and Mapping Input Validation and Sanitization Functions at the Client- and the Server-Side.....	133
9.2.2	Inconsistency Identification.....	134
9.3	Semantic Differential Repair for Input Validation and Sanitization.....	136
9.3.1	Differential Repair Problem.....	140
9.3.2	Differential Repair Algorithm.....	141
9.4	Summary.....	147
<b>10</b>	<b>Tools</b> .....	149
10.1	LIBSTRANGER.....	149
10.2	STRANGER.....	150
10.3	SEMREP.....	151
10.4	ABC.....	153
<b>11</b>	<b>A Brief Survey of Related Work</b> .....	155
11.1	String Analysis.....	155
11.1.1	Grammar Based String Analysis.....	155
11.1.2	Automata-Based String Analysis.....	157
11.1.3	Hybrid String Analysis.....	158
11.2	String Constraint Solvers.....	158
11.2.1	Automata-Based Solvers.....	159
11.2.2	Bounded Solvers.....	160
11.2.3	Combination of Theories.....	161
11.2.4	Model Counting String Constraint Solvers.....	161
11.3	Bug and Vulnerability Detection in Web Applications.....	161
11.3.1	Client-Side Analysis.....	162
11.3.2	Server-Side Analysis.....	162
11.4	Differential Analysis and Repair.....	163
11.4.1	Differential Analysis.....	163
11.4.2	Differential Repair.....	164
<b>12</b>	<b>Conclusions</b> .....	165
	<b>References</b> .....	167