

SpringerBriefs in Computer Science

Series editors

Stan Zdonik, Brown University, Providence, USA
Shashi Shekhar, University of Minnesota, Minneapolis, USA
Jonathan Katz, University of Maryland, College Park, USA
Xindong Wu, University of Vermont, Burlington, USA
Lakhmi C. Jain, University of South Australia, Adelaide, Australia
David Padua, University of Illinois Urbana-Champaign, Urbana, USA
Xuemin (Sherman) Shen, University of Waterloo, Waterloo, Canada
Borko Furht, Florida Atlantic University, Boca Raton, USA
V.S. Subrahmanian, University of Maryland, College Park, USA
Martial Hebert, Carnegie Mellon University, Pittsburgh, USA
Katsushi Ikeuchi, University of Tokyo, Tokyo, Japan
Bruno Siciliano, Università di Napoli Federico II, Napoli, Italy
Sushil Jajodia, George Mason University, Fairfax, USA
Newton Lee, Newton Lee Laboratories, LLC, Tujunga, USA

Mu Zhang • Heng Yin

Android Application Security

A Semantics and Context-Aware Approach



Springer

Mu Zhang
Computer Security Department
NEC Laboratories America, Inc.
Princeton, NJ, USA

Heng Yin
University of California, Riverside
Riverside, CA, USA

ISSN 2191-5768 ISSN 2191-5776 (electronic)
SpringerBriefs in Computer Science
ISBN 978-3-319-47811-1 ISBN 978-3-319-47812-8 (eBook)
DOI 10.1007/978-3-319-47812-8

Library of Congress Control Number: 2016959410

© The Author(s) 2016

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.

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

The authors would like to dedicate this book to their beloved families and friends and to those who overcome their frustration and persevere with resubmitting papers to top-tier computer security conferences.

Preface

This book is an introduction to the cutting-edge technologies for discovery, diagnosis, and defense of emerging security problems in modern Android applications.

With great power comes great threat. Recently, due to the popularity of Android smartphones, Android apps have attracted varieties of cyber attacks: some involve advanced anti-detection techniques; some exploit “genetic” defects in Android programs; some cover up identity theft with camouflage; some trick end users to fall into a trap using intriguing but misleading language. To defeat malicious attempts, researchers strike back. Many traditional techniques have been studied and practiced: malware classification, taint analysis, access control, etc. Yet, intrusive techniques also advance, and, unfortunately, existing defenses fall short, fundamentally due to the lack of sufficient interpretation of Android application behaviors.

To address this limitation, we look at the problem from a different angle. Android apps, no matter good, bad, or vulnerable, are in fact software programs. Their functionality is concretized through semantically meaningful code and varies under different circumstances. This reveals two essential factors for understanding Android application, semantics and contexts, which, we believe, are also the key to tackle security problems in Android apps. As a result, we have developed a series of semantics and context-aware techniques to fight against Android security threats. We have applied our idea to four significant areas, namely, malware detection, vulnerability patching, privacy leakage mitigation, and misleading app descriptions. This will be elaborated through the whole book.

Intended Audience

This book is suitable for security professionals and researchers. It will also be useful for graduate students who are interested in mobile application security.

Acknowledgments

The authors would like to thank Lok, Aravind, Andrew, Qian, Xunchao, Yue, Rundong, Jinghan, Manju, Eknath, and Curtis for the stimulating discussions and generous support.

Princeton, NJ, USA
Riverside, CA, USA
September 2016

Mu Zhang
Heng Yin

Contents

- 1 Introduction 1**
 - 1.1 Security Threats in Android Applications 1
 - 1.1.1 Malware Attacks 1
 - 1.1.2 Software Vulnerabilities 2
 - 1.1.3 Information Leakage..... 2
 - 1.1.4 Insecure Descriptions 2
 - 1.2 A Semantics and Context Aware Approach to Android Application Security 3
 - References 4
- 2 Background 7**
 - 2.1 Android Application 7
 - 2.1.1 Android Framework API 8
 - 2.1.2 Android Permission..... 8
 - 2.1.3 Android Component 8
 - 2.1.4 Android App Description..... 9
 - 2.2 Android Malware Detection 9
 - 2.2.1 Signature Detection and Malware Analysis 10
 - 2.2.2 Android Malware Classification 10
 - 2.3 Android Application Vulnerabilities 11
 - 2.3.1 Component Hijacking Vulnerabilities 11
 - 2.3.2 Automatic Patch and Signature Generation 12
 - 2.3.3 Bytecode Rewriting 12
 - 2.3.4 Instrumentation Code Optimization 13
 - 2.4 Privacy Leakage in Android Apps 13
 - 2.4.1 Privacy Leakage Detection 13
 - 2.4.2 Privacy Leak Mitigation 14
 - 2.4.3 Information Flow Control 14
 - 2.5 Text Analytics for Android Security 14
 - 2.5.1 Automated Generation of Software Description 15
 - References 15

3	Semantics-Aware Android Malware Classification	19
3.1	Introduction	19
3.2	Overview	21
3.2.1	Problem Statement	21
3.2.2	Architecture Overview	22
3.3	Weighted Contextual API Dependency Graph	23
3.3.1	Key Behavioral Aspects	23
3.3.2	Formal Definition	24
3.3.3	A Real Example	24
3.3.4	Graph Generation	26
3.4	Android Malware Classification	30
3.4.1	Graph Matching Score	30
3.4.2	Weight Assignment	31
3.4.3	Implementation and Graph Database Query	32
3.4.4	Malware Classification	33
3.5	Evaluation	34
3.5.1	Dataset and Experiment Setup	34
3.5.2	Summary of Graph Generation	34
3.5.3	Classification Results	36
3.5.4	Runtime Performance	40
3.5.5	Effectiveness of Weight Generation and Weighted Graph Matching	40
	References	42
4	Automatic Generation of Vulnerability-Specific Patches for Preventing Component Hijacking Attacks	45
4.1	Introduction	45
4.2	Problem Statement and Approach Overview	47
4.2.1	Running Example	47
4.2.2	Problem Statement	49
4.2.3	Approach Overview	50
4.3	Taint Slice Computation	51
4.3.1	Running Example	51
4.4	Patch Statement Placement	52
4.5	Patch Optimization	53
4.5.1	Optimized Patch for Running Example	54
4.6	Experimental Evaluation	56
4.6.1	Experiment Setup	56
4.6.2	Summarized Results	57
4.6.3	Detailed Analysis	58
	References	60
5	Efficient and Context-Aware Privacy Leakage Confinement	63
5.1	Introduction	63
5.2	Approach Overview	65
5.2.1	Key Techniques	65

5.3	Context-Aware Policy	66
5.3.1	Taint Propagation Trace	67
5.3.2	Source and Sink Call-Sites	67
5.3.3	Parameterized Source and Sink Pairs	68
5.3.4	Implementation	69
5.4	Experimental Evaluation.....	69
5.4.1	Summarized Analysis Results.....	70
5.4.2	Detailed Analysis	71
5.4.3	Runtime Performance	74
	References	75
6	Automatic Generation of Security-Centric Descriptions for Android Apps	77
6.1	Introduction	77
6.2	Overview	78
6.2.1	Problem Statement.....	78
6.2.2	Architecture Overview	80
6.3	Security Behavior Graph	82
6.3.1	Formal Definition	82
6.3.2	<i>SBG</i> of Motivating Example	82
6.3.3	Graph Generation	83
6.4	Behavior Mining and Graph Compression.....	86
6.5	Description Generation	87
6.5.1	Automatically Generated Descriptions	87
6.5.2	Behavior Description Model	88
6.5.3	Behavior Graph Translation	90
6.5.4	Motivating Example	91
6.6	Evaluation	92
6.6.1	Correctness and Security-Awareness	92
6.6.2	Readability and Effectiveness	95
	References	97
7	Limitation and Future Work	99
7.1	Android Malware Classification	99
7.2	Automated Vulnerability Patching	100
7.3	Context-Aware Privacy Protection	101
7.4	Automated Generation of Security-Centric Descriptions	102
	References	103
8	Conclusion	105