

Advances in Information Security

Volume 73

Series editor

Sushil Jajodia, George Mason University, Fairfax, VA, USA

More information about this series at <http://www.springer.com/series/5576>

Jiaojiao Jiang • Sheng Wen • Bo Liu • Shui Yu
Yang Xiang • Wanlei Zhou

Malicious Attack Propagation and Source Identification

Jiaojiao Jiang
Swinburne University of Technology
Hawthorne, Melbourne
VIC, Australia

Sheng Wen
Swinburne University of Technology
Hawthorne, Melbourne
VIC, Australia

Bo Liu
La Trobe University
Bundoora, VIC, Australia

Shui Yu
University of Technology Sydney
Ultimo, NSW, Australia

Yang Xiang
Digital Research & Innovation Capability
Swinburne University of Technology
Hawthorn, Melbourne
VIC, Australia

Wanlei Zhou
University of Technology Sydney
Ultimo, NSW, Australia

ISSN 1568-2633

Advances in Information Security

ISBN 978-3-030-02178-8

ISBN 978-3-030-02179-5 (eBook)

<https://doi.org/10.1007/978-3-030-02179-5>

Library of Congress Control Number: 2018959747

© Springer Nature Switzerland AG 2019

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.

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

Preface

In the modern world, the ubiquity of networks has made us vulnerable to various malicious attacks. For instance, computer viruses propagate throughout the Internet and infect millions of computers. Misinformation spreads incredibly fast in online social networks, such as Facebook and Twitter. Experts say that “fake news” on social media platforms influenced US election voters. Researchers and manufacturers evolve new methods to produce detection systems to detect suspicious attacks. However, how can we detect the propagation source of the attacks so as to protect network assets from fast acting attacks? Moreover, how can we build up effective and efficient prevention systems to stop malicious attacks before they do damage and have a chance to infect a system.

So far, extensive work has been done to develop new approaches, to effectively identify the propagation source of malicious attacks and to efficiently restrain the malicious attacks. The goal of this book is to summarize and analyze the state-of-the-art research and investigations in the field of identifying propagation sources and preventing malicious propagation, so as to provide an approachable strategy for researchers and engineers to implement this new framework in real-world applications. The striking features of the book can be illustrated from three basic aspects:

- A detailed coverage on analyzing and preventing the propagation of malicious attacks in complex networks. On the one hand, a practical problem in malicious attack propagation is the spreading influence of initial spreaders. This book presents and analyzes different methods for influential spreader detection. On the other hand, various strategies have been proposed for preventing malicious attack propagation. This book numerically analyzes these strategies, concludes the equivalences of these strategies, and presents a hybrid strategy by combining different strategies.
- A rich collection of contemporary research results in identifying the propagation source of malicious attacks. According to the categories of observations on malicious attacks, current research can be divided into three types. For each

type, we particularly present one representative method and the theory behind each method. A comprehensive theoretical analysis of current methods is further presented. Apart from the theoretical analysis, the book numerically analyzes their pros and cons based on real-world datasets.

- A comprehensive study of critical research issues in identifying the propagation source of malicious attacks. For each issue, the book presents a brief introduction to the problem and its challenges, and a detailed state-of-the-art method to solve the problem.

This book intends to enable readers, especially postgraduate and senior undergraduate students, to study up-to-date concepts, methods, algorithms, and analytic skills for building modern detection and prevention systems through analyzing the propagation of malicious attacks. It enables students not only to master the concepts and theories in relation to malicious attack propagation and source identification but also to readily use the material introduced into implementation practices.

The book is divided into three parts: malicious attack propagation, propagation source identification, and critical research issues in source identification. In the first part, after an introduction of the preliminaries of malicious attack propagation, the book presents detailed descriptions on areas of detecting influential spreaders and restraining the propagation of malicious attacks. In the second part, after a summary on the techniques involved in propagation source identification under different categories of observations about malicious attack propagation, the book then presents a comprehensive study of these techniques and uses real-world datasets to numerically analyze their pros and cons. In the third part, the book explores three critical research issues in the research area of propagation source identification. The most difficult one is the complex spatiotemporal diffusion process of malicious attacks in time-varying networks, which is the bottleneck of current approaches. The second issue lies in the expensively computational complexity of identifying multiple propagation sources. The third important issue is the huge scale of the underlying networks, which makes it difficult to develop efficient strategies to quickly and accurately identify propagation sources. These weaknesses prevent propagation source identification from being applied in a broader range of real-world applications. This book systematically analyzes the state of the art in addressing these issues and aims at making propagation source identification more effective and applicable.

Hawthorne, Melbourne, VIC, Australia
 Hawthorne, Melbourne, VIC, Australia
 Bundoora, VIC, Australia
 Ultimo, NSW, Australia
 Hawthorne, Melbourne, VIC, Australia
 Ultimo, NSW, Australia
 September 2018

Jiaojiao Jiang
 Sheng Wen
 Bo Liu
 Shui Yu
 Yang Xiang
 Wanlei Zhou

Acknowledgments

We are grateful to many research students and colleagues at Swinburne University of Technology in Melbourne and the University of Technology Sydney in Sydney, who have made a lot of comments to our presentations as their comments inspire us to write this book. We would like to acknowledge some support from research grants we have received, in particular the Australian Research Council Grant no. LP120200266, DP140103649, and DP180102828. Some interesting research results presented in the book are taken from our research papers that indeed (partially) were supported through these grants. We also would like to express our appreciations to the editors at Springer, especially Susan Lagerstrom-Fife and Caroline Flanagan, for the excellent professional support. Finally we are grateful to the family of each of us for their consistent and persistent supports. Without their support, the book may just become some unpublished discussions.

Hawthorne, Melbourne, VIC, Australia
Hawthorne, Melbourne, VIC, Australia
Bundoora, VIC, Australia
Ultimo, NSW, Australia
Hawthorne, Melbourne, VIC, Australia
Ultimo, NSW, Australia
September 2018

Jiaojiao Jiang
Sheng Wen
Bo Liu
Shui Yu
Yang Xiang
Wanlei Zhou

Contents

- 1 Introduction** 1
 - 1.1 Malicious Attacks 1
 - 1.2 Examples of Malicious Attacks..... 2
 - 1.3 Propagation Mechanism of Malicious Attacks 4
 - 1.4 Source Identification of Malicious Attack Propagation 6
 - 1.5 Outline and Book Overview 7
- Part I Malicious Attack Propagation**
- 2 Preliminary of Modeling Malicious Attack Propagation**..... 11
 - 2.1 Graph Theory 11
 - 2.2 Network Topologies 14
 - 2.3 Community Structure 16
 - 2.4 Information Diffusion Models 18
- 3 User Influence in the Propagation of Malicious Attacks** 21
 - 3.1 Introduction 21
 - 3.2 Problem Statement 23
 - 3.3 Epidemic Betweenness..... 25
 - 3.3.1 Information Propagation Model 26
 - 3.3.2 Epidemic Influence 27
 - 3.3.3 Computation of Epidemic Betweenness 28
 - 3.3.4 Simple Examples 29
 - 3.3.5 Computational Complexity 30
 - 3.4 Evaluations 32
 - 3.4.1 Accuracy in Measuring Influence..... 32
 - 3.4.2 Comparison with Other Measures of Influence 34
 - 3.5 Correlation Analysis 35
 - 3.5.1 Correlation with Traditional Betweenness 35
 - 3.5.2 Correlation with Classic Centrality Measures..... 37
 - 3.6 Related Work 38
 - 3.7 Summary..... 39

| | | |
|----------|--|----|
| 4 | Restrain Malicious Attack Propagation | 41 |
| 4.1 | Introduction | 41 |
| 4.2 | Methods of Restraining Rumors | 42 |
| 4.2.1 | Controlling Influential Users | 43 |
| 4.2.2 | Controlling Community Bridges | 44 |
| 4.2.3 | Clarification Through Spreading Truths | 45 |
| 4.3 | Propagation Modeling Primer | 46 |
| 4.3.1 | Modeling Nodes, Topology and Social Factors | 46 |
| 4.3.2 | Modeling Propagation Dynamics | 47 |
| 4.3.3 | Modeling People Making Choices | 49 |
| 4.3.4 | The Accuracy of the Modelling | 49 |
| 4.4 | Block Rumors at Important Users | 50 |
| 4.4.1 | Empirical Studies | 50 |
| 4.4.2 | Theoretical Studies | 53 |
| 4.5 | Clarify Rumors Using Truths | 55 |
| 4.5.1 | Impact of the Truth Injection Time | 55 |
| 4.5.2 | Impact of the Truth Propagation Probability | 57 |
| 4.6 | A Hybrid Measure of Restraining Rumors | 58 |
| 4.6.1 | Measures Working Together | 58 |
| 4.6.2 | Equivalence of Measures | 60 |
| 4.7 | Summary | 61 |
| 4.7.1 | The Robustness of the Contagious Ability | 61 |
| 4.7.2 | The Fairness to the Community Bridges | 62 |

Part II Source Identification of Malicious Attack Propagation

| | | |
|----------|---|----|
| 5 | Preliminary of Identifying Propagation Sources | 65 |
| 5.1 | Observations on Malicious Attack Propagation | 65 |
| 5.2 | Maximum-Likelihood Estimation | 66 |
| 5.3 | Efficiency Measures | 67 |
| 6 | Source Identification Under Complete Observations: | |
| | A Maximum Likelihood (ML) Source Estimator | 69 |
| 6.1 | Introduction | 69 |
| 6.2 | The SI Model for Information Propagation | 70 |
| 6.3 | Rumor Source Estimator: Maximum Likelihood (ML) | 70 |
| 6.4 | Rumor Source Estimator: ML for Regular Trees | 71 |
| 6.5 | Rumor Source Estimator: ML for General Trees | 72 |
| 6.6 | Rumor Source Estimator: ML for General Graphs | 73 |
| 6.7 | Rumor Centrality | 74 |
| 6.7.1 | Rumor Centrality: Succinct Representation | 75 |
| 6.7.2 | Rumor Centrality Versus Distance Centrality | 76 |

| | | |
|---|---|-----|
| 7 | Source Identification Under Snapshots: A Sample Path Based Source Estimator | 79 |
| 7.1 | Introduction | 79 |
| 7.2 | The SIR Model for Information Propagation | 80 |
| 7.3 | Maximum Likelihood Detection | 81 |
| 7.4 | Sample Path Based Detection | 82 |
| 7.5 | The Sample Path Based Estimator | 83 |
| 7.6 | Reverse Infection Algorithm | 87 |
| 8 | Source Identification Under Sensor Observations: A Gaussian Source Estimator | 89 |
| 8.1 | Introduction | 89 |
| 8.2 | Network Model | 90 |
| 8.3 | Source Estimator | 91 |
| 8.4 | Source Estimator on a Tree Graph | 91 |
| 8.5 | Source Estimator on a General Graph | 93 |
| 9 | Comparative Study and Numerical Analysis | 95 |
| 9.1 | Comparative Study | 95 |
| 9.1.1 | Methods Based on Complete Observations | 95 |
| 9.1.2 | Methods Based on Snapshots | 100 |
| 9.1.3 | Methods Based on Sensor Observations | 102 |
| 9.2 | Numerical Analysis | 105 |
| 9.2.1 | Comparison on Synthetic Networks | 106 |
| 9.2.2 | Comparison on Real-World Networks | 112 |
| 9.3 | Summary | 113 |
| Part III Critical Research Issues in Source Identification | | |
| 10 | Identifying Propagation Source in Time-Varying Networks | 117 |
| 10.1 | Introduction | 117 |
| 10.2 | Time-Varying Social Networks | 118 |
| 10.2.1 | Time-Varying Topology | 118 |
| 10.2.2 | Security States of Individuals | 119 |
| 10.2.3 | Observations on Time-Varying Social Networks | 120 |
| 10.3 | Narrowing Down the Suspects | 122 |
| 10.3.1 | Reverse Dissemination Method | 122 |
| 10.3.2 | Performance Evaluation | 125 |
| 10.4 | Determining the Real Source | 126 |
| 10.4.1 | A Maximum-Likelihood (ML) Based Method | 127 |
| 10.4.2 | Propagation Model | 129 |
| 10.5 | Evaluation | 130 |
| 10.5.1 | Accuracy of Rumor Source Identification | 130 |
| 10.5.2 | Effectiveness Justification | 132 |
| 10.6 | Summary | 136 |

| | | |
|-----------|---|------------|
| 11 | Identifying Multiple Propagation Sources | 139 |
| 11.1 | Introduction | 139 |
| 11.2 | Preliminaries | 141 |
| 11.2.1 | The Epidemic Model | 141 |
| 11.2.2 | The Effective Distance | 142 |
| 11.3 | Problem Formulation | 143 |
| 11.4 | The K-Center Method | 144 |
| 11.4.1 | Network Partitioning with Multiple Sources | 145 |
| 11.4.2 | Identifying Diffusion Sources and Regions | 145 |
| 11.4.3 | Predicting Spreading Time | 148 |
| 11.4.4 | Unknown Number of Diffusion Sources | 149 |
| 11.5 | Evaluation | 149 |
| 11.5.1 | Accuracy of Identifying Rumor Sources | 151 |
| 11.5.2 | Estimation of Source Number and Spreading Time | 153 |
| 11.5.3 | Effectiveness Justification | 154 |
| 11.6 | Summary | 157 |
| 12 | Identifying Propagation Source in Large-Scale Networks | 159 |
| 12.1 | Introduction | 159 |
| 12.2 | Community Structure | 161 |
| 12.3 | Community-Based Method | 162 |
| 12.3.1 | Assigning Sensors | 162 |
| 12.3.2 | Community Structure Based Approach | 163 |
| 12.3.3 | Computational Complexity | 165 |
| 12.4 | Evaluation | 167 |
| 12.4.1 | Identifying Diffusion Sources in Large Networks | 168 |
| 12.4.2 | Influence of the Average Community Size | 169 |
| 12.4.3 | Effectiveness Justification | 171 |
| 12.4.4 | Comparison with Current Methods | 173 |
| 12.5 | Summary | 178 |
| 13 | Future Directions and Conclusion | 179 |
| 13.1 | Continuous Time-Varying Networks | 179 |
| 13.2 | Multiple Attacks on One Network | 180 |
| 13.3 | Interconnected Networks | 180 |
| 13.4 | Conclusion | 180 |
| | References | 183 |