

SpringerBriefs in Computer Science

Series editors

Stan Zdonik, Brown University, Providence, Rhode Island, USA

Shashi Shekhar, University of Minnesota, Minneapolis, Minnesota, USA

Jonathan Katz, University of Maryland, College Park, Maryland, USA

Xindong Wu, University of Vermont, Burlington, Vermont, USA

Lakhmi C. Jain, University of South Australia, Adelaide, South Australia, Australia

David Padua, University of Illinois Urbana-Champaign, Urbana, Illinois, USA

Xuemin (Sherman) Shen, University of Waterloo, Waterloo, Ontario, Canada

Borko Furht, Florida Atlantic University, Boca Raton, Florida, USA

V.S. Subrahmanian, University of Maryland, College Park, Maryland, USA

Martial Hebert, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Katsushi Ikeuchi, University of Tokyo, Tokyo, Japan

Bruno Siciliano, Università di Napoli Federico II, Napoli, Campania, Italy

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

Newton Lee, Newton Lee Laboratories, LLC, Tujunga, California, USA

More information about this series at <http://www.springer.com/series/10028>

Xingquan Zhu • Haicheng Tao • Zhiang Wu
Jie Cao • Kristopher Kalish • Jeremy Kayne

Fraud Prevention in Online Digital Advertising

Xingquan Zhu
Dept. of Computer & Electrical
Engineering and Computer Science
Florida Atlantic University
Boca Raton, FL, USA

Haicheng Tao
Dept. of Computer & Electrical
Engineering and Computer Science
Florida Atlantic University
Boca Raton, FL, USA

Zhiang Wu
College of Information Engineering
Nanjing University of Finance
and Economics
Nanjing, China

Jie Cao
College of Information Engineering
Nanjing University of Finance
and Economics
Nanjing, China

Kristopher Kalish
Bidtellect, Inc.
Delray Beach, FL, USA

Jeremy Kayne
Bidtellect, Inc.
Delray Beach, FL, USA

ISSN 2191-5768 ISSN 2191-5776 (electronic)
SpringerBriefs in Computer Science
ISBN 978-3-319-56792-1 ISBN 978-3-319-56793-8 (eBook)
DOI 10.1007/978-3-319-56793-8

Library of Congress Control Number: 2017942765

© The Author(s) 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

Foreword

Advertising has a long history which can be traced back to ancient civilization. Four thousand years ago in early Egypt, Greece, China, and India, people used wall or rock paintings for commercial advertising. In England during the eighteenth century, advertisements started to appear in weekly newspapers which marked the start of modern advertising. Advertising continued to advance with technology and offered increased audience targeting and reach via newspapers, radio, TV, and many other modern devices. The birth of the Internet in the twentieth century was a ground-breaking platform allowing advertisers to immediately reach out to individuals, collecting their response and feedback in real time. Such evolution marks a new discipline, computational advertising, which uses computing technology to offer more efficient and effective advertising.

In digital advertising, advertisers, publishers, and audiences are connected over the Internet. The lack of physical interaction and the requirement of real-time response make our advertising ecosystem vulnerable to fraud. These fraudulent attacks come in many forms including non-human traffic, Bot clicks, unintended clicks, or stacked and unviewable Ad banners. The need for countermeasures has never been so urgent and challenging given that both advertisers and publishers are facing huge volumes of traffic on a daily basis and fraudsters are stealing a significant portion of revenue and continuously deteriorating the whole ecosystem (online ad fraud driven by bots costs brands over \$7 billion globally in 2016 alone).

This book offers the first comprehensive study of fraud in digital advertising. The fraud map, or taxonomy as noted in the book, provides a clear view of most known fraud in the industry. The study of fraud prevention and commercial systems allows academic researchers and industrial developers to fully understand the problems from both scientific and practical perspectives. As we are embracing the big data era, fraud detection and prevention represents a big challenge that the whole industry is facing. I am glad that this book timely captures such challenges and is moving the industry ahead to provide better and safer advertising.

Delray Beach, FL, USA
January 17, 2017

Lon Otremba

Preface

Computational advertising refers to computing platforms or systems, which use computational approaches to optimally match audiences and advertisement. In advertising ecosystems, revenue is the driving force fueling the whole system. Publishers, advertisers, and third-party service providers are all trying to maximize the return by following business rules and procedures. Yet, the highly specialized business models and the nature of the cyberspace information switch without physical interactions make it difficult to directly evaluate whether the advertisements are indeed served to genuine human viewers, or clicks are actually generated by human users. As a result, it creates an environment for fraudsters to use deceptive approaches to attack advertising systems for illicit returns. This reality has raised serious integrity concerns, because a fraud flooded advertising market hardly has any value to stakeholders, and fraud must be well controlled, if not completely removed, for the healthy growth of the digital advertising market.

In this monograph, we systematically review forms of online digital advertising (Ad) fraud and the techniques to prevent and defeat them. We categorize Ad fraud into three major categories, including (1) placement fraud, (2) traffic fraud, and (3) action fraud. We summarize major features of each type of fraud and also outline measures and resources to detect each type of fraud.

The book provides a comprehensive guide to help researchers comprehend the state of the art in Ad fraud detection. It also serves as a technical reference for industry to design new techniques and solutions to win the battle against fraud.

Boca Raton, FL, USA
Boca Raton, FL, USA
Delray Beach, FL, USA
Delray Beach, FL, USA
January 15

Xingquan Zhu
Haicheng Tao
Kristopher Kalish
Jeremy Kayne

Acknowledgements

On January 26, 2016, Florida Atlantic University's College of Engineering and Computer Science announced that it received a \$300,000 gift from Bidtellect, a global leader in native advertising technologies and solutions. The gift will allow the Department of Computer and Electrical Engineering and Computer Science (CEECS) to engage in research, develop and establish curriculum, facilitate laboratory improvements, fund research assistantships, and provide an annual award to the top Bidtellect Laboratory researcher.

We are grateful to the Bidtellect for their continuous support in establishing FAU Bidtellect Laboratory and funding provided for the research projects. "This generous gift from Bidtellect enforces the mission of the College of Engineering and Computer Science to offer innovative educational experiences to its students alongside our faculty researchers," said Mohammad Ilyas, Ph.D., dean of the College of Engineering and Computer Science.

H. Tao, Z. Wu, and J. Cao are partially supported by National Natural Science Foundation of China (91646204, 71571093, 71372188) and National Center for International Joint Research on E-Business Information Processing (2013B01035).

Contents

- 1 Introduction 1**
 - 1.1 Computational Advertising 1
 - 1.2 Display Advertising 2
 - 1.3 Fraud in Digital Advertising..... 3
 - 1.4 Book Objective 5
 - References 5
- 2 Ad Ecosystems and Key Components 7**
 - 2.1 Ad Banner and Viewability 8
 - 2.2 Ad Business Model 12
 - 2.3 Revenue Models..... 13
 - 2.4 Ad Advertisement Concept Flow..... 14
 - References 17
- 3 Ad Fraud Taxonomy and Prevention Mechanisms..... 19**
 - 3.1 Ad Fraud Taxonomy 19
 - 3.2 Ad Fraud Prevention Mechanism 21
 - References 22
- 4 Ad Fraud Categorization and Detection Methods 25**
 - 4.1 Placement Fraud 25
 - 4.1.1 Stuffing or Stacking..... 26
 - 4.1.2 Fake Sites 26
 - 4.1.3 Domain Spoofing 27
 - 4.1.4 Ad Injection and Malware..... 28
 - 4.2 Traffic Fraud..... 29
 - 4.2.1 Impression Fraud 30
 - 4.2.2 Click Fraud..... 30

4.3	Action Fraud	33
4.3.1	Conversion Fraud	33
4.3.2	Re-targeting Fraud	34
4.3.3	Affiliate Fraud	35
	References	36
5	Ad Fraud Measure and Benchmark	39
5.1	Measures with Ground Truth	39
5.2	Measures Without Ground Truth	41
5.3	Real World Datasets	42
	References	43
6	Ad Fraud Detection Tools and Systems.....	45
6.1	Commercial Ad Fraud Detection Systems	45
6.2	Ad Fraud Detection Systems in Academia	47
	References	48
7	Conclusion	51
	Glossary	53

Acronyms

Ad	The abbreviation for advertising or advertisement.
CPM	Cost Per Mille (CPM) is the unit price based on one thousand impression.
CPC	Cost Per Click (CPC) is the unit price based on one click event.
CPA	Cost Per Action (CPA) is the unit price based on one conversion event.
CTR	Click Through Rate (CTR) is the ratio of clicks on an Ad (or a page, site etc.) divided by the number of impressions of the Ad (or the page, site etc).
DSP	Demand Side Platform (DSP) is for advertisers to manage/plan advertising campaigns.
DMP	Data Management Platform (DMP) is a system for creating audiences based on first-party and third-party data so that advertisers may target those audiences.
eCPC	Effective Cost Per Click (eCPC) is calculated using the costs of buying the impressions divided by the number of clicks generated from the impressions. CPC is normally used for defining price of buying clicks, whereas eCPC is a metric defining the equivalent CPC prices of buying impressions.
eCPM	Effective Cost Per Mille (eCPM) is calculated using the total costs spent on an Ad campaign divided by number of impressions resulted from such costs (normalized in 1000 impressions). CPM is normally used for defining price of buying impressions, whereas eCPM is a metric defining the equivalent CPM prices of buying methods (e.g. fixed price, CPM, CPC, or CPA).
KPI	Key Performance Indicator (KPI) may refer to any metric, such as CPA or CTR, used to evaluate the performance.
SSP	Supplier Side Platform (SSP) is an analytics tool for publishers to manage advertisement placements.