

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

The right to copy the material presented in this thesis is fully owned by the Author of this thesis, while saying this individuals are permitted to download a copy for their individual research and private study excluding any reproduction without a written consent of the Author.

Copyright ©

By

Mohammad Rashid Ourahman

2015

*To my
Beloved father and grandfather who left me during the writing of this thesis, may their
souls rest in peace; also to my beloved mother, wife
and lovely children, from whom I have been separated for the
past year of work on this thesis.*



AutoURL: Automatic URL Tracking
to identify rogue advertising

This thesis is submitted in fulfillment of the requirements for the degree of

MASTER OF INFORMATION SCIENCES in
SOFTWARE ENGINEERING

School of Engineering and Advanced Technology (S.E.A.T)
at Massey University, [Albany],
New Zealand.

by

Mohammad Rashid Ourahman, BEng (Massey University)

January 2015

Acknowledgements

I express my deepest appreciation to Prof. Paul Watters for giving me the opportunity to complete this thesis through helping, guiding and motivating me during the entire processes of researching and completion. Furthermore I also acknowledge with much appreciation the love and support of family and friends who were not given their rights to their affection and left alone for almost one year. Last but not least, many thanks go to my friends and staff at Massey University who shared lovely moments and didn't hesitate to show me the right directions whenever I asked them any question during the entire process of writing this thesis.

Abstract

Maintaining Cyber Security has been one of the biggest challenges of a modern era which has seen the extensive emergence of internet advertisers, and in which some promote their malicious contents through rogue websites.

Internet rogue advertisers penetrate through cybercrime in various forms of advertisement banners which are displayed within any parts of a website. Tracking these rogue advertisers is important to the Cyber Security cause, where in an ideal scenario individuals are exposed to correct information as is their basic right, along with their reaction toward the sensitivity of any content.

In the past manual tracking has been the commonest method of checking but in some cases manual tracking could fail, other than time parameters the accuracy is also questionable, the solution to this the concept of Automatic URL Tracking.

This thesis represents an analytical method of Automatic URL Tracking, according to this approach, where various pages are checked for advertising banners, these are clicked until the final URL or its destination is reached.

To achieve various concrete results a significant work has been done to develop an Automatic URL Tracking Software which is run when connected through internet while holding the reported URLs databases where each of these are tracked to its final destination.

The Automatic URL Tracking Software was run for the total of 2500 URL samples, upon manually tracking these URLs the two processes showed 87.7 % agreement which can be reliable result considering the presence of various blocking techniques adopted by hosting sites and site developers but there are chances for further development where the application is enhanced specifically to overcome these obstacles.

Automatic URL Tracking overcomes the difficulties and challenges of manual tracking, allowing larger data volumes to be tested, identified and verified, but having said that it also comes with the challenges of rapidly changing internet technologies, in which more comprehensive strategies need to be built to overcome this challenge.

Table of Contents

Introduction	1
1.2 Cyber Security:.....	4
1.3 Internet Advertising:	12
1.3.1 Types of Internet Advertising:.....	15
1.3.2 Mainstream Advertising:	23
1.3.3 High-Risk Advertising:.....	24
1.3.4 Advertising Networks:.....	24
1.3.5 Internet Advertisers:	25
1.3.6 Rogue websites:.....	25
1.4 Google Transparency Report:.....	27
1.4.1 Digital Millennium Copyright Act (DMCA):	28
1.4.2 Chilling Effects Database	28
1.5 Automatic URL Tracking:	6
1.6 Research Outline:	7
1.6 Thesis Scope:	9
1.7 Contribution	10
1.8 Summary:	10
Literature Review	11
2.1 Background:	11
2.2 Increasing Online Spending:.....	29
2.2 Online Advertisement Revenue:.....	30
2.3 Dangers of Rogue Sites:	31
2.4 Tracking Rogue Sites:	32
2.5 Summary:.....	33
Research Methodology	34
2.1 Background:	34
2.3 Algorithm:.....	36
2.4Summary:	37
Automatic URL Tracking	38
4.1 Background:.....	38

4.2 Auto URL Track:	39
4.3 Auto URL Track Software:	39
4.3.1 Technical Specifications:	41
4.3.2 Data Type and Source:.....	42
4.3.3 Auto URL Tracking Technical Features:.....	43
4.3.4 Auto URL Tracking Functionalities:.....	45
4.3.4.1 Main Window:.....	45
4.3.4.2 Insert Function:.....	46
4.3.4.3 Display Function:	47
4.3.4.4 Exit Function:	48
4.3.4.5 Search Function:	49
4.3.4.6 Reset Function:	49
4.3.4.7 Export Function:	49
4.4 Auto URL Track Function:	52
4.4 Summary:	65
Experiments and Test Results.....	66
Conclusion and Discussion	70
Future Work.....	73
References	74
Appendix A.....	83
Appendix B.....	90
Appendix C.....	93
Appendix D.....	94
Appendix E.....	96
Appendix F.....	97
Appendix G.....	98

List of Figures

Figure 1: Shows various forms of Cyber Crimes ⁵	5
Figure 2: An Example of Google Ads.....	18
Figure 3: An Example of AdWord Ads	19
Figure 4: An Example of Facebook Ads	19
Figure 5: An Example of Facebook Exchange Ads	20
Figure 6: An Example of Twitter Ads	20
Figure 7: An Example of Retargeting ads	21
Figure 8: Examples of Banner Ads	22
Figure 9: An Example of Mobile Ads.....	23
Figure 10: Example of rogue site.....	26
Figure 11: Screenshot of Auto URL Tracking software	40

List of Tables

Table 1: List of common form of Internet Advertisements ⁶	17
Table 2: Data Structure for requests.csv file	43
Table 3: Test results produced by automation Vs Manual Tracking	67