



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

In Search of Search Privacy

by

Wesley Antonio Brandi

Submitted in partial fulfilment
of the requirements for the degree

Doctor of Philosophy

in the

Faculty of Computer Science

of the

University of Pretoria

July, 2011

In Search of Search Privacy

by

Wesley Antonio Brandi

Supervisor: Prof. Martin S Olivier
Department: Computer Science
Degree: Doctor of Philosophy

Summary

Search engines have become integral to the way in which we use the Web of today. Not only are they an important real time source of links to relevant information, but they also serve as a starting point to the Web. A veritable treasure trove of the latest news, satellite images, directions from anywhere to anywhere, local traffic updates and global trends ranging from the spread of influenza to which celebrity happens to be the most popular at a particular time.

The more popular search engines are collecting incredible amounts of information. In addition to indexing significant portions of the Web they record what hundreds of millions of users around the world are searching for. As more people use a particular search engine, it has the potential to record more information on what is deemed relevant (and in doing so provide better relevance in the future, thereby attracting more users). Unfortunately, the relevance derived from this cycle between the search user and the search engine comes at a cost: privacy. In this work, we take an in depth look at what privacy means within the context of search. We discuss why it is that the search engine must be considered a threat to search privacy. We then investigate potential solutions and eventually propose our own in a bid to enhance search privacy.

Acknowledgements

My thanks to everybody who stood by me through the years that it took to get this done. Prof. Olivier for believing in me way back in the day and providing subtle hints and tips in a way only he knows how. Olman and olady for their unwavering support. My sisters for putting up with me in general and my brother for the bizarre ideas and crazy conversations.

The whole experience: full time, part time, on campus, off campus, jobs, sundays, bursaries, friends, labs, exams, tutorials, workgroups, papers, reviews, conferences, questions, rejections, acceptance, and on and on. All of it an incredible journey that I cherish and remember fondly.

Happy days.

Contents

1	Aims and Scope	1
1.1	Introduction	2
1.2	Outline	3
2	Privacy Enhancing Technologies	6
2.1	Introduction	7
2.2	Privacy	7
2.2.1	Privacy in this research	9
2.3	Anonymity	10
2.4	Privacy Enhancing Technologies	11
2.4.1	Private Communication	12
2.4.2	Anonymity	14
2.4.3	Personal Control	16
2.4.4	Organisational Safeguards	19
2.5	Conclusion	20
3	Search Privacy	21
3.1	Introduction	22
3.2	Search	22
3.3	Search Profiles	25
3.3.1	Offline Profiles	26
3.3.2	A Problem	27
3.4	PETs and Search Privacy	30
3.4.1	Private Communication	30
3.4.2	Anonymity	31
3.4.3	Personal Control	31
3.4.4	Organisational Safeguards	32
3.4.5	Summary	32
3.5	Conclusion	32

4	Search Privacy Through Anonymity	35
4.1	Introduction	36
4.2	Motive	36
4.3	Background	38
4.4	The Case for an External Attack	39
4.4.1	Principle 1 - an unlikely sender	40
4.4.2	Principle 2 - recognising related requests	40
4.5	Assumptions	41
4.6	The Attack	42
4.7	A Simulation	44
4.8	Conclusion	47
5	Search Privacy Through Personal Control	49
5.1	Introduction	50
5.2	P3P	51
5.3	Trust and P3P	52
5.3.1	Reputation Systems	53
5.3.2	The Need for Trust in P3P	54
5.3.3	Prerequisites of a Reputation System	55
5.3.4	A Reputation System in P3P	56
5.3.5	Reputations Systems Bring Trust to P3P	61
5.4	Proxies and P3P	64
5.4.1	Dealing with Web proxies in P3P	64
5.4.2	P3P Web Proxy Problems	66
5.4.3	Transparent and Chained Proxies	70
5.4.4	Identification and Separation	73
5.5	Search Privacy and Personal Control	74
6	Search Privacy Through Private Communication	76
6.1	Introduction	77
6.2	TrackMeNot	77
6.3	Recognising TMN	78
6.4	Obfuscation and Search	79
6.5	Conclusion	81
7	A Search Network	82
7.1	Introduction	83
7.2	A Case for Sharing	83
7.2.1	Analysis of Search Data	84
7.3	A Search Network	86

7.3.1	A Scalable Network	87
7.3.2	A Fast Network	88
7.3.3	A Privacy Preserving Network	88
7.3.4	Submitting a Query	89
7.4	Formalisation	90
7.4.1	Submitting Queries Directly	91
7.4.2	Submitting Queries Through a Proxy	91
7.4.3	Using a Proxy and Direct Submission	92
7.4.4	Using Multiple Proxies and Direct Submission	92
7.4.5	Submission of Queries Through a DHT	93
7.5	Conclusion	94
8	Conclusion	96
8.1	Does this enhance search privacy?	100
8.2	Is the search engine still a threat?	101

List of Figures

2.1	IE8 InPrivate Blocking feature	18
3.1	A search engine processes a user’s query with the goal of delivering a set of links to relevant content on the Web that was crawled, indexed and processed earlier.	24
4.1	Correct assumptions made by the colluding servers	45
4.2	Percentage of crowd members that were exposed through an attack.	46
4.3	Percentage of the original crowd that has been exposed through an attack. A growth factor of 1.0 indicates that the crowd grew by 100% of its original size.	47
5.1	Reputation Ring of a new site.	59
5.2	A reputation that has improved over time and a similar reputation that has had a confirmed incident reported against it.	60
5.3	The effect of v ranging from 0.1 to 1.0 using a 12 month reputation with a single confirmed report against it.	62
5.4	The effect of v ranging from 0.1 to 1.0 using a 12 month reputation with a single confirmed report against it.	63
5.5	User accesses a Web site through a P3P compliant Proxy. If applicable, the proxy will load and adhere to a policy specific to the site being accessed.	69
7.1	The top 1,000 queries ranked in descending order.	84
7.2	Each user is plotted against the number of shared and unique queries that he/she submitted.	85
7.3	In this figure the number of users that used each shared query is plotted.	86

List of Tables

3.1	An evaluation of the PET categories	33
4.1	A map of senders and participants responsible for delivering the message.	42
4.2	A map of senders and participants upon a new jondo joining the crowd.	43
5.1	An optional parameter extension to the POLICY-REF element.	68
5.2	Optional addition to the P3P Header.	71
5.3	The PROXY element.	72
5.4	An optional parameter extension to the POLICY-REF element.	72
7.1	A comparison of the different approaches used when querying a search engine.	95