Guido Schryen

Anti-Spam Measures

Analysis and Design

With 50 Figures and 23 Tables

Springer

Contents

List of Figures XIII

List of Tables XV

Abbreviations XVII

1 Introduction 1
1.1 The problem 1
1.2 The history 2
1.3 Goals, methodology, and architecture 3

2 Spam and its economic significance 7
2.1 Definition 7
2.2 Spam statistics 9
2.3 Spam categories 12
2.3.1 Commercial advertising 13
2.3.2 Non-commercial advertising 16
2.3.3 Fraud and phishing 17
2.3.4 Hoaxes and chain e-mails 19
2.3.5 Joe jobs 19
2.3.6 Malware 21
2.3.7 Bounce messages 21
2.4 Economic harm 22
2.5 Economic benefit 26

3 The e-mail delivery process and its susceptibility to spam . 29
3.1 The e-mail delivery process 29
3.2 SMTP's susceptibility to spam 36

4 Anti-spam measures 43
4.1 Legislative measures 43
4.1.1 Parameters 44
4.1.2 Anti-spam laws 48
4.1.3 The effectiveness 52
4.2 Organizational measures 54
4.2.1 Abuse systems 54
4.2.2 International cooperation 55
4.3 Behavioral measures 56
4.3.1 The protection of e-mail addresses 56
4.3.2 The handling of received spam e-mails 58
4.4 Technological measures 59
4.4.1 IP blocking 61
4.4.2 Filtering 65
4.4.3 TCP blocking 71
4.4.4 Authentication 72
4.4.5 Verification 78
4.4.6 Payment-based approaches 80
4.4.7 Limitation of outgoing e-mails 86
4.4.8 Address obscuring techniques 87
4.4.9 Reputation-based approaches 90
4.4.10 Summary 91

5 A model-driven analysis of the effectiveness of technological anti-spam measures 95
5.1 A model of the Internet e-mail infrastructure 96
5.1.1 The definition 96
5.1.2 The appropriateness 101
5.2 Deriving and categorizing the spam delivery routes 105
5.2.1 Deriving the spam delivery routes 105
5.2.2 Categorizing the spam delivery routes 109
5.2.3 Some example delivery routes and their formal representations 111
5.3 The effectiveness of route-specific anti-spam measures 112
5.3.1 IP blocking 113
5.3.2 TCP blocking 113
5.3.3 SMTP extensions 115
5.3.4 Cryptographic authentication 115
5.3.5 Path authentication 115
5.3.6 Limitation of outgoing e-mails 116
5.3.7 Reputation-based 116
5.3.8 Conclusion 116

6 An infrastructure framework for addressing spam 119
6.1 Overview of the framework 120
6.2 Organizational solution 123
6.2.1 Integrating CMAAs into the Internet 124
6.2.2 Certificating an organization as a CMAA 125
6.2.3 Mapping organizations onto CMAAs 126
6.2.4 Registering for the usage of CMAA services 127
6.3 Technological solution 128
6.3.1 Databases 128
6.3.2 Database administration processes 131
6.3.3 E-mail delivery process 135
6.3.4 Abuse complaint process 138
6.4 Theoretical effectiveness 139
6.5 Deployment and impact on e-mail communication 140
6.6 Drawbacks and limitations 143

7 The empirical analysis of the abuse of e-mail addresses placed on the Internet 145
7.1 The relevance of inspecting e-mail address harvesting 145
7.2 Prior studies and findings 147
7.3 A methodology and honeypot conceptualization 149
7.3.1 A framework for seeding e-mail addresses 149
7.3.2 Data(base) models for storing e-mails 151
7.4 The prototypic implementation of an empirical study 165
7.4.1 The goals and the conceptualization of the seeding 166
7.4.2 The adaptation of the database model 167
7.4.3 The IT infrastructure of the honeypot 168
7.4.4 Empirical results and conclusions 169

8 Summary and outlook 175

A Process for parsing, classifying, and storing e-mails 185

B Locations seeded with addresses that attracted most spam 189

References 193

Index 205