A

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

Fig. A.1: UML activity diagram for parsing, classifying, and storing e-mails (1)
A Process for parsing, classifying, and storing e-mails

Fig. A.2: UML activity diagram for parsing, classifying, and storing e-mails (2)
**B**

Locations seeded with addresses that attracted most spam

<table>
<thead>
<tr>
<th>#spams</th>
<th>web location</th>
</tr>
</thead>
<tbody>
<tr>
<td>829</td>
<td><a href="http://jeepbrokers.com/jeepbrokers_guestbook.htm">http://jeepbrokers.com/jeepbrokers_guestbook.htm</a></td>
</tr>
<tr>
<td>536</td>
<td><a href="http://www.theaterhaus.com/easync/easync_page.php?id=1,4,1&amp;page=forum_geastebuch.htm">http://www.theaterhaus.com/easync/easync_page.php?id=1,4,1&amp;page=forum_geastebuch.htm</a></td>
</tr>
<tr>
<td>534</td>
<td><a href="http://www.theaterhaus.com/easync/easync_page.php?id=1,4,1&amp;page=forum_geastebuch.htm">http://www.theaterhaus.com/easync/easync_page.php?id=1,4,1&amp;page=forum_geastebuch.htm</a></td>
</tr>
<tr>
<td>518</td>
<td><a href="http://www.la-palma24.net/de_visitas/guestbook.php3">http://www.la-palma24.net/de_visitas/guestbook.php3</a></td>
</tr>
<tr>
<td>510</td>
<td><a href="http://www.la-palma24.net/de_visitas/guestbook.php3">http://www.la-palma24.net/de_visitas/guestbook.php3</a></td>
</tr>
<tr>
<td>499</td>
<td><a href="http://www.la-palma24.net/de_visitas/guestbook.php3">http://www.la-palma24.net/de_visitas/guestbook.php3</a></td>
</tr>
<tr>
<td>453</td>
<td><a href="http://www.beaufortrlty.com/guestbook.html">http://www.beaufortrlty.com/guestbook.html</a></td>
</tr>
<tr>
<td>412</td>
<td><a href="http://www.cyber-kitchen.com/cgibin/gbook/guestbook.cgi">http://www.cyber-kitchen.com/cgibin/gbook/guestbook.cgi</a></td>
</tr>
<tr>
<td>396</td>
<td><a href="http://www.germantownnews.com/guestbook">http://www.germantownnews.com/guestbook</a></td>
</tr>
<tr>
<td>393</td>
<td><a href="http://www.cyber-kitchen.com/cgibin/gbook/guestbook.cgi">http://www.cyber-kitchen.com/cgibin/gbook/guestbook.cgi</a></td>
</tr>
<tr>
<td>383</td>
<td><a href="http://www.kelso.gov/">http://www.kelso.gov/</a></td>
</tr>
<tr>
<td>378</td>
<td><a href="http://www.cyber-kitchen.com/cgibin/gbook/guestbook.cgi">http://www.cyber-kitchen.com/cgibin/gbook/guestbook.cgi</a></td>
</tr>
<tr>
<td>337</td>
<td><a href="http://www.bowlsengland.com/efgbk00.htm">http://www.bowlsengland.com/efgbk00.htm</a></td>
</tr>
<tr>
<td>337</td>
<td><a href="http://jeepbrokers.com/jeepbrokers_guestbook.htm">http://jeepbrokers.com/jeepbrokers_guestbook.htm</a></td>
</tr>
<tr>
<td>332</td>
<td><a href="http://jeepbrokers.com/jeepbrokers_guestbook.htm">http://jeepbrokers.com/jeepbrokers_guestbook.htm</a></td>
</tr>
<tr>
<td>304</td>
<td><a href="http://jeepbrokers.com/jeepbrokers_guestbook.htm">http://jeepbrokers.com/jeepbrokers_guestbook.htm</a></td>
</tr>
<tr>
<td>275</td>
<td><a href="http://www.metager.de">http://www.metager.de</a></td>
</tr>
<tr>
<td>253</td>
<td><a href="http://www.ourchurch.com/view/?pageID=111918">http://www.ourchurch.com/view/?pageID=111918</a></td>
</tr>
<tr>
<td>222</td>
<td><a href="http://books.dreambook.com/dawsadopt/main.html">http://books.dreambook.com/dawsadopt/main.html</a></td>
</tr>
</tbody>
</table>

*multiple occurrence of web locations refers to the placement of different e-mail addresses on these locations

**Fig. B.1:** Web locations seeded with addresses that attracted most spam
## B Locations seeded with addresses that attracted most spam

<table>
<thead>
<tr>
<th>#spams</th>
<th>newsgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td>69</td>
<td>de.rec.sport.paintball</td>
</tr>
<tr>
<td>67</td>
<td>de.rec.tv.buffy</td>
</tr>
<tr>
<td>60</td>
<td>alt.drugs</td>
</tr>
<tr>
<td>52</td>
<td>de.sci.misc</td>
</tr>
<tr>
<td>50</td>
<td>alt.america</td>
</tr>
<tr>
<td>41</td>
<td>alt.airports</td>
</tr>
<tr>
<td>35</td>
<td>alt.fan.brad-pitt</td>
</tr>
<tr>
<td>30</td>
<td>de.rec.sport.misc</td>
</tr>
<tr>
<td>29</td>
<td>de.talk.jokes</td>
</tr>
<tr>
<td>29</td>
<td>de.org.ccc</td>
</tr>
<tr>
<td>28</td>
<td>de.soc.weltanschauung.misc</td>
</tr>
<tr>
<td>27</td>
<td>de.rec.tv.technik</td>
</tr>
<tr>
<td>26</td>
<td>alt.fan.shania-twain</td>
</tr>
<tr>
<td>26</td>
<td>alt.games.microsoft.age-of-empires</td>
</tr>
<tr>
<td>26</td>
<td>microsoft.public.microsoft.transaction.server.integration</td>
</tr>
<tr>
<td>25</td>
<td>de.etc.selbsthilfe.angst</td>
</tr>
<tr>
<td>23</td>
<td>de.talk.jokes.d</td>
</tr>
<tr>
<td>23</td>
<td>alt.windows-me</td>
</tr>
<tr>
<td>22</td>
<td>alt.off-topic</td>
</tr>
</tbody>
</table>

*Fig. B.2: Usegroups seeded with addresses that attracted most spam*
<table>
<thead>
<tr>
<th>#spams</th>
<th>Newsletter*</th>
</tr>
</thead>
<tbody>
<tr>
<td>66</td>
<td>Churchill College, University of Cambridge; <a href="http://www.opendays.com/newsletter/">http://www.opendays.com/newsletter/</a></td>
</tr>
<tr>
<td>45</td>
<td>Jayde B2B Search Engine; <a href="http://www.jayde.de">http://www.jayde.de</a></td>
</tr>
<tr>
<td>44</td>
<td>Jayde B2B Search Engine; <a href="http://www.jayde.de">http://www.jayde.de</a></td>
</tr>
<tr>
<td>40</td>
<td>Jayde B2B Search Engine; <a href="http://www.jayde.de">http://www.jayde.de</a></td>
</tr>
<tr>
<td>22</td>
<td>Jayde B2B Search Engine; <a href="http://www.jayde.de">http://www.jayde.de</a></td>
</tr>
<tr>
<td>21</td>
<td>Churchill College, University of Cambridge; <a href="http://www.opendays.com/newsletter/">http://www.opendays.com/newsletter/</a></td>
</tr>
<tr>
<td>6</td>
<td>French Erotic Site; <a href="http://www.sexy.legratuit.com">http://www.sexy.legratuit.com</a></td>
</tr>
<tr>
<td>6</td>
<td>Churchill College, University of Cambridge; <a href="http://www.opendays.com/newsletter/">http://www.opendays.com/newsletter/</a></td>
</tr>
<tr>
<td>6</td>
<td>French Erotic Site; <a href="http://www.sexy.legratuit.com">http://www.sexy.legratuit.com</a></td>
</tr>
<tr>
<td>6</td>
<td>Churchill College, University of Cambridge; <a href="http://www.opendays.com/newsletter/">http://www.opendays.com/newsletter/</a></td>
</tr>
<tr>
<td>5</td>
<td>French Erotic Site; <a href="http://www.sexy.legratuit.com">http://www.sexy.legratuit.com</a></td>
</tr>
<tr>
<td>3</td>
<td>Edvisors.com: International Student Newsletter; <a href="http://www.edvisors.com/cgi/page.cgi?p=newsletter">http://www.edvisors.com/cgi/page.cgi?p=newsletter</a></td>
</tr>
<tr>
<td>3</td>
<td>Edvisors.com: International Student Newsletter; <a href="http://www.edvisors.com/cgi/page.cgi?p=newsletter">http://www.edvisors.com/cgi/page.cgi?p=newsletter</a></td>
</tr>
<tr>
<td>2</td>
<td>Edvisors.com: International Student Newsletter; <a href="http://www.edvisors.com/cgi/page.cgi?p=newsletter">http://www.edvisors.com/cgi/page.cgi?p=newsletter</a></td>
</tr>
</tbody>
</table>

* multiple occurrence of newsletters refers to multiple subscription to these newsletters (each subscription uses a unique e-mail address)

Fig. B.3: Newsletters seeded with addresses that attracted most spam
References


References


[81] InfoSec: 2004, Email Spamming (include scam), Technical report.


[85] ITU: 2003, Memorandum of Understanding Between the Korea Information Security Agency and the Australian Communications Authority and the National Office for the Information Economy of Australia Concerning Cooperation in the Regulation of Spam.


[145] Sandvine: 2004, Zombie PCs spew out 80% of spam.


Abuse Database (ADB) 121, 128, 130
Abuse system 54, 121, 138
Address book 57
Address Obscuring/Obfuscating
    Technique (AOT) 58, 87, 145
Adware 21
Anti-Phishing Working Group 59
Anti-spam
    laws 48, 50, 51
    legislation 49
    opt-in 54
    opt-out 49, 54, 59
    policy 53
Anti-spam measures 1
    behavioral 4, 56
    classification 60
    CPU-based 80
    effectiveness 95
    legislative 3, 43
    confirmed opt-in 46
    double out-in 45
    opt-in 45
    opt-out 45, 46
    plain opt-in 46
    memory-based 82
    monetary 83
    organisational 54
    organizational 3
    route-invariant 96
    route-specific 95, 112
    technological 4, 59
Anti-Spam Technical Alliance (ASTA) 2, 56, 72, 86, 115
Attention bond mechanism 84
Australian Coalition Against Unsolici
dited Bulk Email (CAUBE.AU) 147
Australian Communications Authority 55
Austrian Law of Telecommunications 47
Authentication
    cryptographic 72, 115
    DNS-based 76
    path 76, 115
Automata theory 105
Bayes theorem 69
Blacklist 40, 55, 61, 62, 113, 125
    Arbitrary black hole list 62
    DNS-based 63
    Domain Name System Blacklist
        (DNSBL) 62
    Domain Name System Real-time
        Blacklist (DNSRBL) 62
    Exploits Block List 62
    MAPS Dial-up User List (MAPS-
        DUL) 63
    Open Relay Database (ORDB) 62
    Right Hand Side Blocklist 64
    Spam Prevention Early Warning
        System (SPEWS) 63
    Spamhaus Block List (SBL) 62
    Uniform Resource Identifier Realtime
        Blacklist 64
Blocking
    IP 61, 113
TCP 71, 113
Blog 57
Bot 53, 110
Botnet 110, 117
Bounce Address Tag Validation (BATV) 65
Bounce e-mail 22
Brute force attack 145, 172
Bulk e-mail 1
Business Model 120, 183
Camram 80
CANSPAM Act of 2003 47–49
Center for Democracy & Technology 148
Central Organization (CO) 124, 126, 127
Certified Server Validation (CSV) 76
Chain e-mail 19
Challenge-response 79, 83
Charity hoax 19
Chat room 57
Commercial advertising 13
Commercially oriented hoax 19
Completely Automated Public Turing Test to Tell Computers and Humans Apart (CAPTCHA) 79, 86, 116, 128, 133
Computer Fraud and Abuse Act 46
Cooperation 55
bilateral government-to-government 55
government-to-private sector 55
multilateral 55
privat sector groups 55
Counter Database (CDB) 121, 128, 131, 144
Counter Managing & Abuse Authority (CMAA) 120, 122, 123, 134, 135, 140, 144
certificate 125
role 120, 121, 123
services 124, 127, 130, 143
Crawler 57
Decentralized Object Location and Routing System (DOLR) 69
Denial of Service (DoS) attacks 40, 42
Designated Mailers Protocol (DMP) 76
Deterministic Finite Automaton (DFA) 105
Dictionary attack 145, 172
Discussion forum 57
Distributed Checksum Clearinghouse (DCC) 69
Distributed Denial of Service (DDoS) attack 144
Domain Name System (DNS) 34, 42, 62, 91, 127, 133, 143
reverse lookup 38
spoofing 64
DomainKeys 73
DomainKeys Identified Mail (DKIM) 73
E-mail
body 36, 66, 151
composition 11
envelope 36
header 36, 38, 66, 151
infrastructure 96
E-mail address
extended e-mail address 87
taxonomy 146
throw-away 57
E-mail Service Provider (ESP) 10
E-stamp 83, 85
Economic benefit 26
Economic harm 22
Electronic Communications Privacy Act (ECPA) 46
Email Service Provider Coalition 91
Entity Relationship Diagram (ERD) 161
Extended e-mail address 87
EZ-Gimpsy 79
False-negative 10, 62, 66, 67
False-Positive 10, 62, 64–68
Federal Trade Commission (FTC) 147, 148
Fiat-Shamir based scheme 81
Filter 59, 65
Artificial Neural Network 66, 71
Bayesian 66, 69
naïve 70
Boosting Tree 66, 71
collaborative 68, 71
Markov Random Field Model 66, 71
non-collaborative 68
rule-based 66, 68
signature-based 66, 68
statistical 66, 69
Support Vector Machine 66, 71
Finger 57
Fraud 17
Frequency analysis 63
Fully Qualified Domain Name (FQDN) 31, 37, 40
Gateway 30
German Civil Code 47
German Criminal Code 46
German Law against Unfair Competition 48
Greylisting 61, 65
Guestbook 57
Ham e-mails 2
Harvester 56–58, 150
Hashcash 80
Hoax 19
Honeypot 148, 149, 165, 169
HTTP over SSL 100
HTTP over TLS 100
Hypertext Transfer Protocol (HTTP) 30, 100
Identified Internet Mail (IIM) 73
Identity theft 26
Infrastructure framework 119, 124, 128
Internet Architecture Board (IAB) 124
Internet Corporation for Assigned Names and Numbers (ICANN) 91, 124, 127, 142
Internet Engineering Steering Group (IESG) 124
Internet Engineering Task Force (IETF) 124
Internet Message Access Protocol (IMAP) 30, 72, 100
Internet Protocol (IP) 30
adress restrictions 30
dynamic adress 38, 40
secure 30
spoofing 37, 62
Internet Relay Chat (IRC) 57
Internet Service Provider (ISP) 9, 41
Internet Society (ISOC) 124, 142
JavaScript 57, 58
Joe job 19
Joke hoax 19
Keystroke logger 86
Korea Information Security Agency 55
Lightweight Currency Protocol (LCP) 85
Lightweight MTA Authentication Protocol (LMAP) 72, 76, 116, 127, 144
Local Delivery Agent (LDA) 101
Local Mail Transfer Protocol (LMTP) 101
London Action Plan 55
LUMOS 123
Lumos 91
Mail Delivery Agent (MDA) 30
Mail Transfer Agent (MTA) 29, 41
last trustable 38
Mail User Agent (MUA) 29
Mailing list 56
Malware 21
Market paradigm 126
Marking Mail Transfer Agents (MTAMARK) 76
Memorandum of Understanding (MoU) 55
Message Authentication Signature Standards (MASS) 73
Message checksum 69
Message Resource Allocation Policy (MRAP) 90
Message Scheduling Policy (MSP) 90
Message submission 30
Message Submission Agent (MSA) 100
META Signatures 73
Microsoft Postmarks 73
Monty Python’s Flying Circus 2
MS-Sender-ID 76
MTA Authorization Records in DNS (MARID) 76
Multi-variate Bernoulli model 71
Multinominal model 71
Multipurpose Internet Mail Extensions (MIME) 151
MX record 35
National Office for the Information Economy of Australia 55
Network Abuse Clearinghouse 59
Network Address Translation (NAT) 38
Newsletter 56
Nigerian scam 17

OECD Spam Task Force 56
Open Pretty Good Privacy (PGP) 73, 151
Organization Database (ODB) 121, 128, 130
Peer-to-Peer (P2P) 68, 69
Penny Black project 80
Phishing 17, 86
Post Office Protocol (POP) 30, 72, 100
authentication 30
Pricing function 80
Privacy Enhancement for Internet Electronic Mail (PEM) 151
Project Honey Pot 148
Proof-of-work 80
Proxy 102
SOCKS 4 112
SOCKS 5 112
Public key
  cryptography 72
  Public Key Infrastructure (PKI) 60
  Public Key Infrastructure (PKI) 75, 85, 115
Racketeer Influenced and Corrupt Organizations Act (RICO) 46
Rate limit 86
Receiving Organization (RO) 97, 109, 126, 137

Register Of Known Spam Operations (ROKSO) 52
Regulation paradigm 126
Relay 30, 40, 102
Reputation 90, 116
Reverse MX (RMX) 76
Secure MIME (S/MIME) 73, 151
Sender Address Verification Extension (SAVE) 79
Sender Policy Framework (SPF) 76
Sending Organization (SO) 97, 109, 127, 128, 134
Service Level Agreement Policy (SLAP) 90
Simple Authentication and Security Layer (SASL) 72, 137
Simple Mail Transfer Protocol (SMTP) 29, 31, 36
authentication 72
cryptographic 72
path 72
commands 34
gateway 102
service extension 29, 72, 100, 115
Deliver By SMTP Service Extension 29, 100
SMTP after IMAP 72
SMTP after POP 72, 100
SMTP Service Extension for Authentication 29, 100
SMTP Service Extension for Returning Enhanced Error 29, 100
SMTP Service Extension for Secure SMTP over Transport Layer Security 29, 100
SMTP-AUTH 72
service extension 72
tunneling 102
Simple Object Access Protocol (SOAP) 131
Single-purpose address (SPA) 87, 88
Social engineering attack 57
SPA block (SPAB) 89
SPA block encoded and encrypted (SPABEE) 89
Spam 1, 3
categories 12
definition 7
measurement 9, 11
non-commercial advertising 16
origin 11
portion 10
types 12
SpamAssassin 67
Spamhaus 123
Sponsored Top Level Domain (sTLD) 91
SpotSpam 54, 56
Spyware 21

Tagged Message Delivery Agent (TMDA) 87, 90
The Honeynet Project & Research Alliance 82, 110, 117
Transmission Control Protocol (TCP) 30
Trojan horse 19, 21, 47, 86, 110, 115, 152

Unified Modeling Language (UML) 31, 151
activity diagram 131, 132, 136
Unsolicited Bulk E-mail (UBE) 9
Unsolicited Commercial E-mail (UCE) 9, 13
UseNet 56
Verification 78
Vipul’s Razor 68
Virtual channel 87
Virus 19, 21, 47, 57, 86, 152

Web page 57
Whitelist 61, 63, 64
Domain Name System Whitelist (DNSWL) 64
Whois service 57
Worm 19, 21, 47, 57, 86, 152
Zombie PC 21, 91, 102, 110, 115