

Strategien gegen Spam

Ein Überblick

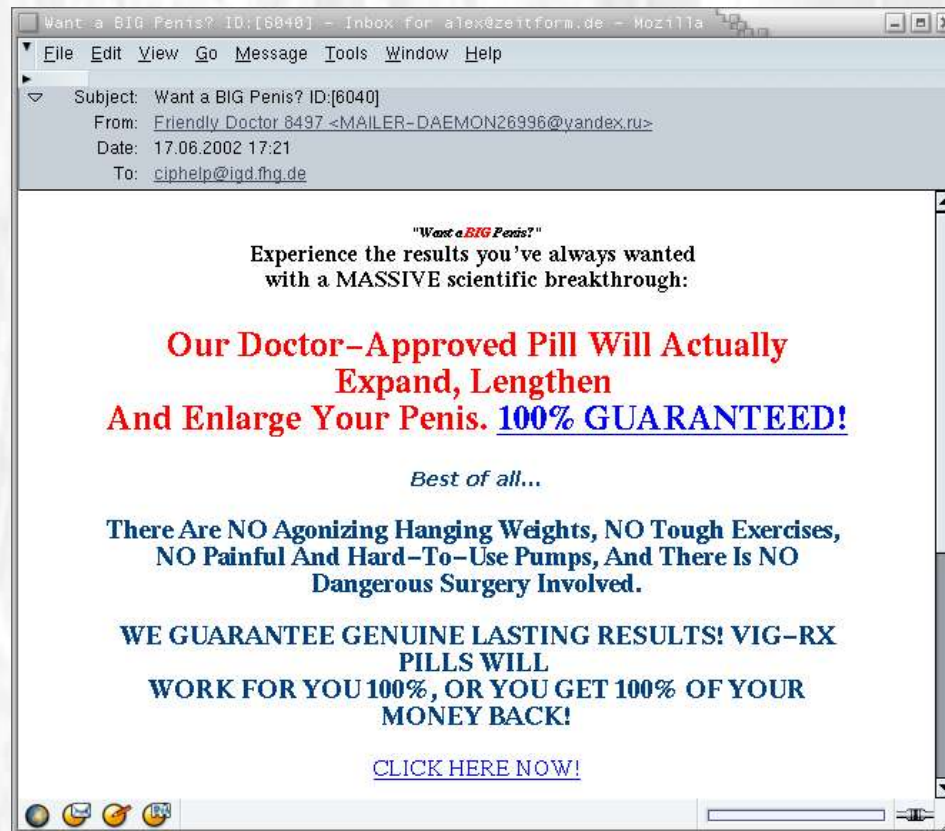


Was ist SPAM™?



Vikings: SPAM SPAM SPAM SPAM. Lovely SPAM! Wonderful SPAM!
SPAM SPA-A-A-A-A-AM SPAM SPA-A-A-A-A-AM SPAM. Lovely SPAM!
Lovely SPAM! Lovely SPAM! Lovely SPAM! Lovely SPAM!
SPAM SPAM SPAM SPAM!

Was ist Spam?



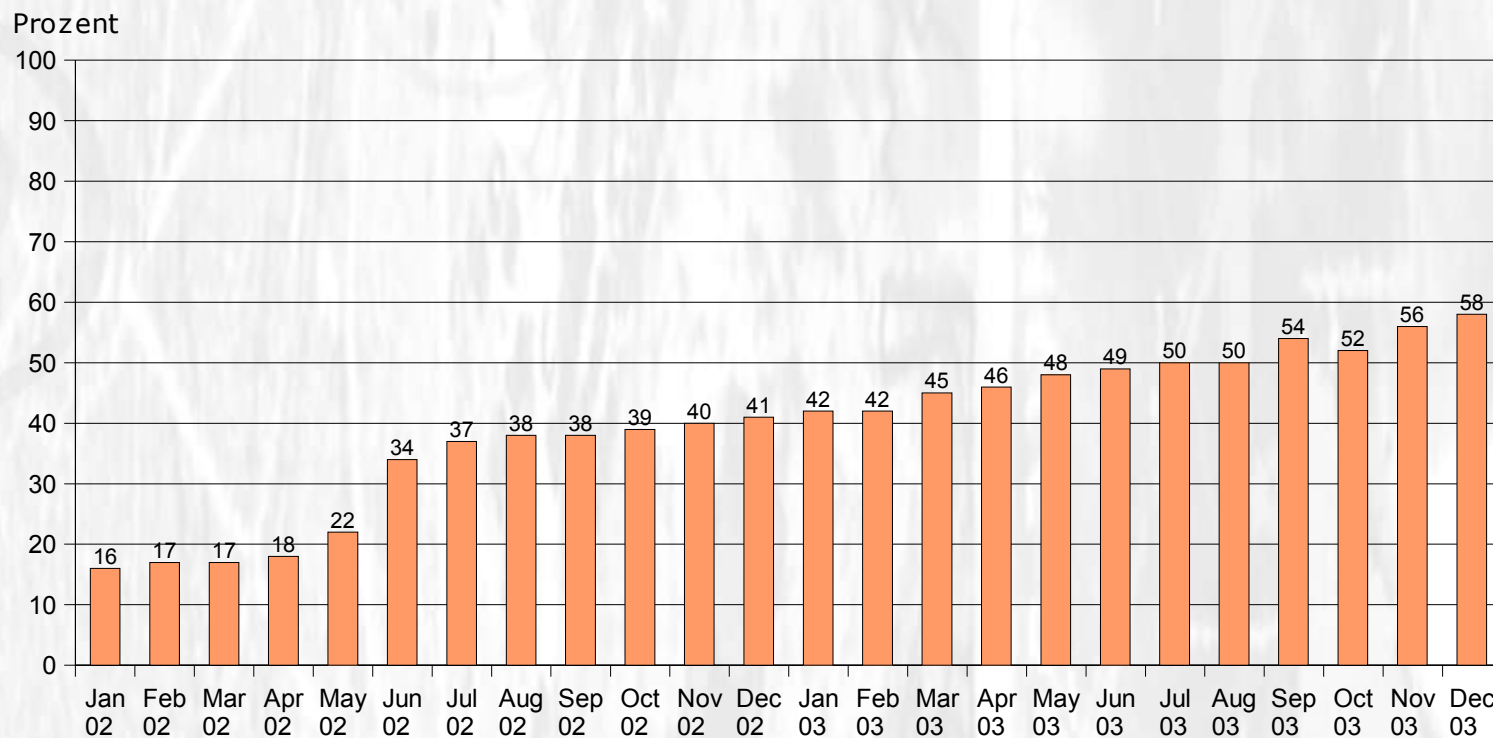
Was ist Spam?

UBE – Unsolicited Bulk Email

UCE – Unsolicited Commercial Email

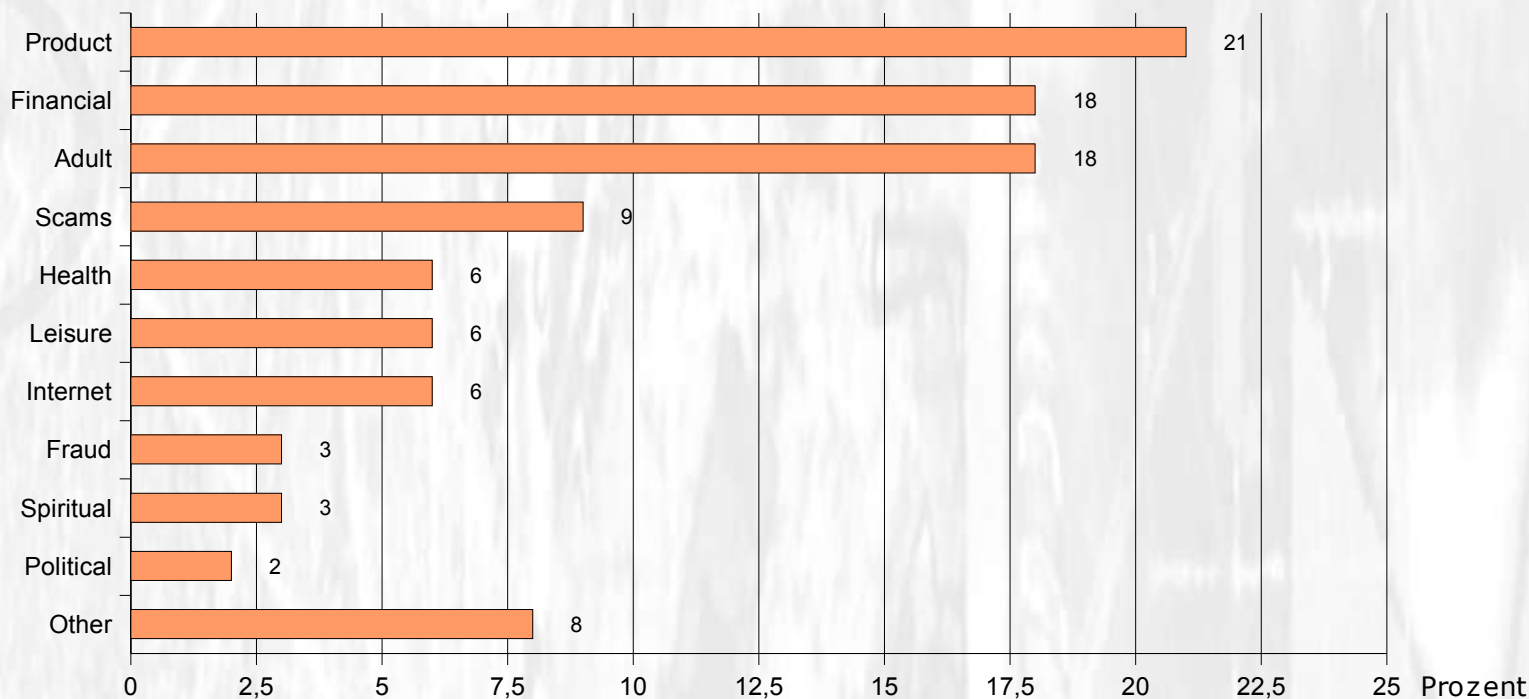
- **Unsolicited:** (dt. unverlangt, unerbeten, unaufgefordert): Die Zusendung der E-Mail wurde nicht explizit vom Empfänger angefordert oder erwartet.
- **Bulk** (dt. Masse, Menge): identische Nachrichten wurden an eine nicht-triviale Anzahl von Empfängern versendet. Ab welcher Zahl von Empfängern eine Nachricht als UBE gelten kann, ist nicht definiert.
- **Commercial** (dt. gewerblich): Die E-Mail wirbt i.d.R. für kommerzielle Produkte und Dienstleistungen.

Zunahme von Spam



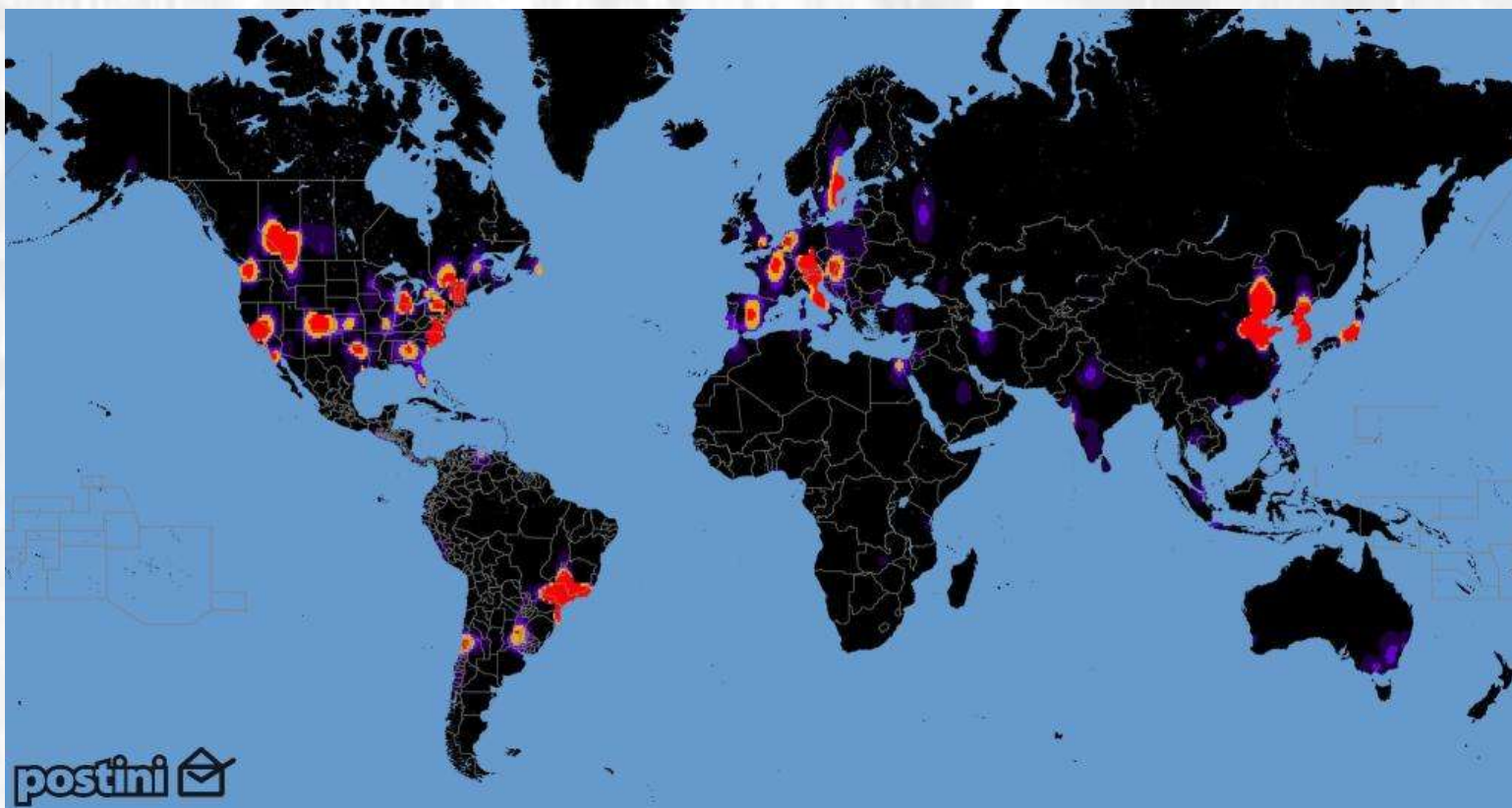
Quelle: <http://www.brightmail.com/spamstats.html>

Inhalte von Spam



Quelle: <http://www.brightmail.com/spamstats.html>

Herkunft von Spam



Quelle: <http://www.postini.com/>

Herkunft von Spam (ISPs)

1. uu.net (UUNET/MCI, USA)
2. chinanet-gd (Chinanet Guangdong, China)
3. above.net (AboveNet, USA)
4. kornet.net (Kornet/KT, Korea)
5. chinanet-cq (Chinanet Chongqing, China)
6. exodus.net (Exodus/Cable & Wireless, UK)
7. level3.net (Level 3, USA)
8. cw.net (Cable & Wireless, UK)
9. chinanet-fj (Chinanet Fujian, China)
10. hinet.net (HiNet, Taiwan)

TOP10 Worst Spam ISPs (January 2004)
Quelle: <http://www.spamhaus.org/>

AboveNet Customer Anti-Spam Policy

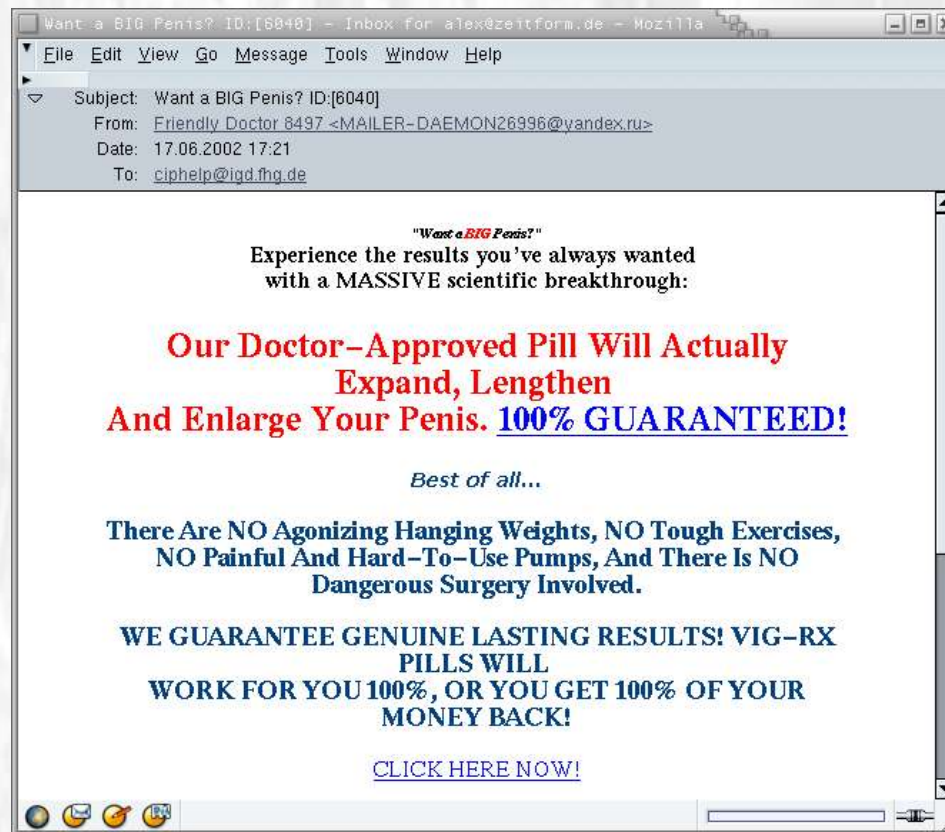
[...]

AboveNet, Inc. (AboveNet), has zero tolerance for Unsolicited Broadcast Email and Unsolicited Commercial Email ("UBE/UCE", commonly known as "Spam") whether originating from customers, from customers' customers, or from customers that provide services which are used to support UBE/UCE.

[...]

Quelle: <http://www.above.net/antispam.html>

Die Spam E-Mail



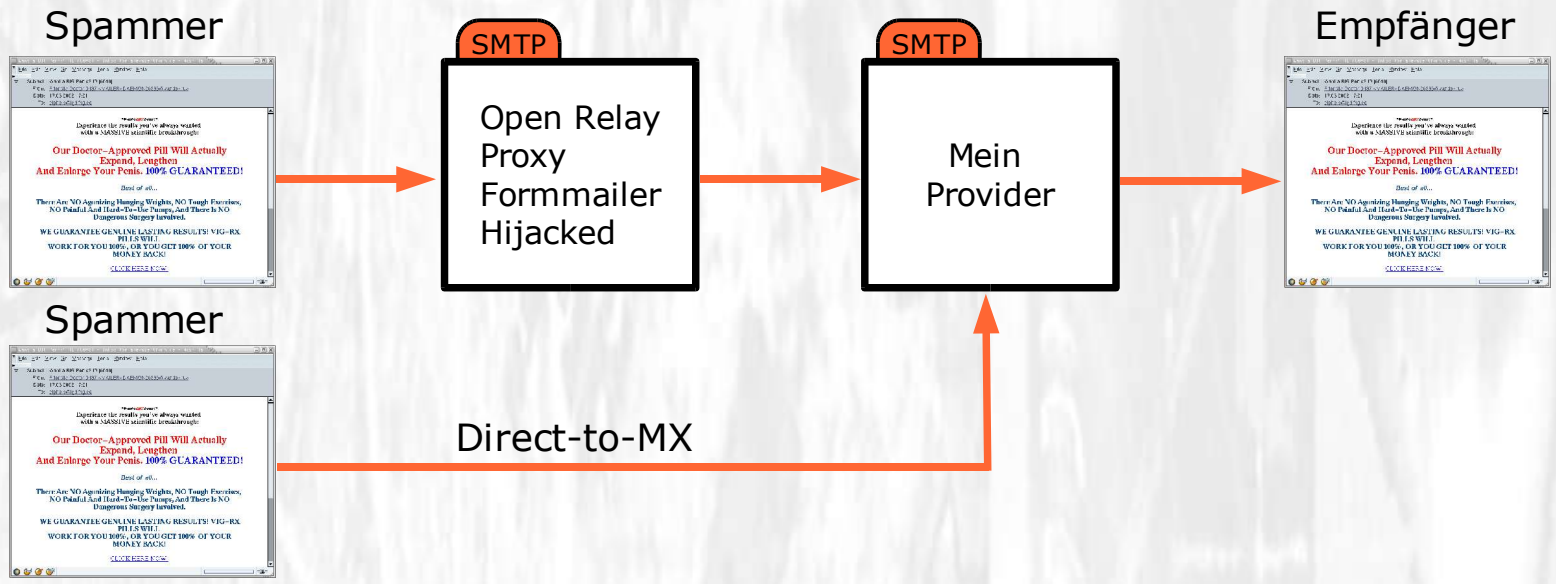
10

Die E-Mail im Rohformat

```
Return-Path: <MAILER-DAEMON11745@yandex.ru>
Received: from [...]
Received: from yandex.ru ([217.13.202.46]) [...]
From: "Friendly Doctor 8497" <MAILER-DAEMON26996@yandex.ru>
To: <ciphelp@igd.fhg.de>
Subject: Want a BIG Penis? ID:[6040]
Date: Mon, 17 Jun 2002 19:21:13 +0400
Mime-Version: 1.0
Content-Type: text/html; charset="ISO-8859-2"

<html>
<head>
<title>Want a BIG Penis?</title>
</head>
<body lang=RU link=blue vlink=purple style='tab-interval:35.4pt'>
[... ]
</body>
</html>
```

Der Versand von Spam (SMTP)



Der SMTP-Dialog

```
S: 220 mail.empfaenger.de ESMTP
C: HELO mail.yandex.ru
S: 250 mail.empfaenger.de
C: MAIL FROM: <MAILER-DAEMON26996@yandex.ru>
S: 250 ok
C: RCPT TO: <ciphelp@igd.fhg.de>
S: 250 ok
C: DATA
S: 354 go ahead
C: From: "Friendly Doctor 8497" <MAILER-DAEMON26996@yandex.ru>
C: To: <ciphelp@igd.fhg.de>
C: Subject: Want a BIG Penis? ID:[6040]
  [...]
C: .
S: 250 ok 997887680 qp 2592
C: QUIT
S: 221 mail.empfaenger.de
```

Verfügbare Informationen

- IP-Adresse des SMTP-Clients (verlässlich)
- HELO/EHLO-String: Hostname (gefälscht)
- Envelope MAIL FROM (gefälscht)
- Envelope RCPT TO (muss korrekt sein)
- Received: (gefälscht, nur teilweise verlässlich)
- From:, To:, Return-Path:, ... (gefälscht)
- Weitere Mail Header (gefälscht)
- Mail Body (Spam)

Blockieren von IP-Adressen

- IP-Adresse des SMTP-Clients ist verlässlich
- SMTP-Dialog wird mit Fehler quittiert:

```
451 Requested action aborted: error in processing
```

```
553 Requested action not taken: mailbox name not allowed
```

- Statische Blacklist (z.B. durch Spam-Traps)
- DNSBL/RBL (DNS-based Blocklist, Realtime Blocklist)
- Web-Interface: <http://openrbl.org>
- Frühestmögliche Abwehr von Spam
- Nachteil: Kollateralschaden (z.B. Dial-Up Adressen)

DNS-based Blocklist

- mail-abuse.org (MAPS, kommerziell, verschiedene Listen)
- spamhaus.org (SBL/XBL/ROKSO, bekannte Spammer)
- spews.org (spam-freundliche ISPs, sehr aggressiv)
- dsbl.org (unsichere Mailsysteme)
- relays.osirusoft.com (~~verschiedene Listen~~)
- spamcop.net (von Empfängern gemeldete Adressen)
- dnsbl.sorbs.net (Open Relays)
- ordb.org (Open Relays)
- monkeys.com (~~verschiedene Listen~~)
- dev.null.dk (Open Relays)
- blackholes.us (spam-freundliche ISPs und Staaten)

DNS-based Blocklist

Beispiel: 69.6.27.48

```
> host -t any 48.27.6.69.sbl.spamhaus.org
48.27.6.69.sbl.spamhaus.org has address 127.0.0.2
48.27.6.69.sbl.spamhaus.org descriptive text
"http://www.spamhaus.org/SBL/sbl.lasso?query=SBL6636"
```

Beispiel: 146.140.212.1

```
> host -t any 1.212.140.146.sbl.spamhaus.org
Host not found.
```

DNS-based Blocklist

```
; spfilter magic sources: SBL,SBL (0.57_031210)
; bind zone rbl.zeitform.de.
$TTL      43200
$ORIGIN   rbl.zeitform.de.
@         SOA      rbl.zeitform.de. root.rbl.zeitform.de. (
                2003121000 10800 3600 604800 21600)
@         NS       ns.zeitform.de.
@         MX       100 mail.zeitform.de.
; test-entries
about     TXT      "zone built by spfilter/0.59 (SBL, bind, 20031210)"
2.0.0.127 TXT      "Test bind rbl.zeitform.de [146.140.212.117]"
2.0.0.127 A        127.0.0.2      ; {every dnsbl should have that}
; listed addresses
0.142.135.12 TXT     "SBL http://www.spamhaus.org/SBL/sbl.lasso?query=SBL5845"
0.142.135.12 A       127.0.0.2
1.142.135.12 TXT     "SBL http://www.spamhaus.org/SBL/sbl.lasso?query=SBL5845"
1.142.135.12 A       127.0.0.2
10.142.135.12 TXT    "SBL http://www.spamhaus.org/SBL/sbl.lasso?query=SBL5845"
10.142.135.12 A      127.0.0.2
11.142.135.12 TXT    "SBL http://www.spamhaus.org/SBL/sbl.lasso?query=SBL5845"
11.142.135.12 A      127.0.0.2
[...]
```

Envelope-Check

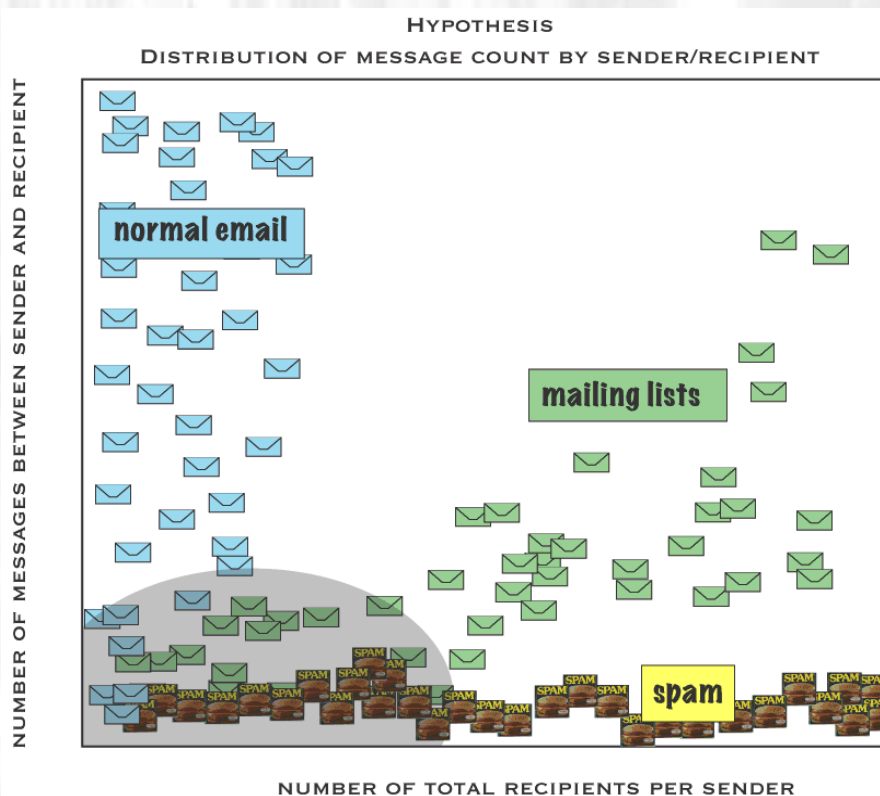
HELO host.domain.com

- Blacklist/Whitelist mit HELO-Strings
- Prüfung auf FQDN
- Prüfung der Domain über DNS und rDNS

MAIL FROM:<sender@domain.com>

- Blacklist/Whitelist mit Absendern/Absender-Domains
- Prüfung des Domain-Teils über DNS und rDNS
- Prüfung auf DSN (dsn.rfc-ignorant.org, MX takes bounces)

Grey-Listing



Quelle: <http://dumbo.pobox.com/spam-sensor/>

Grey-Listing

Annahmen:

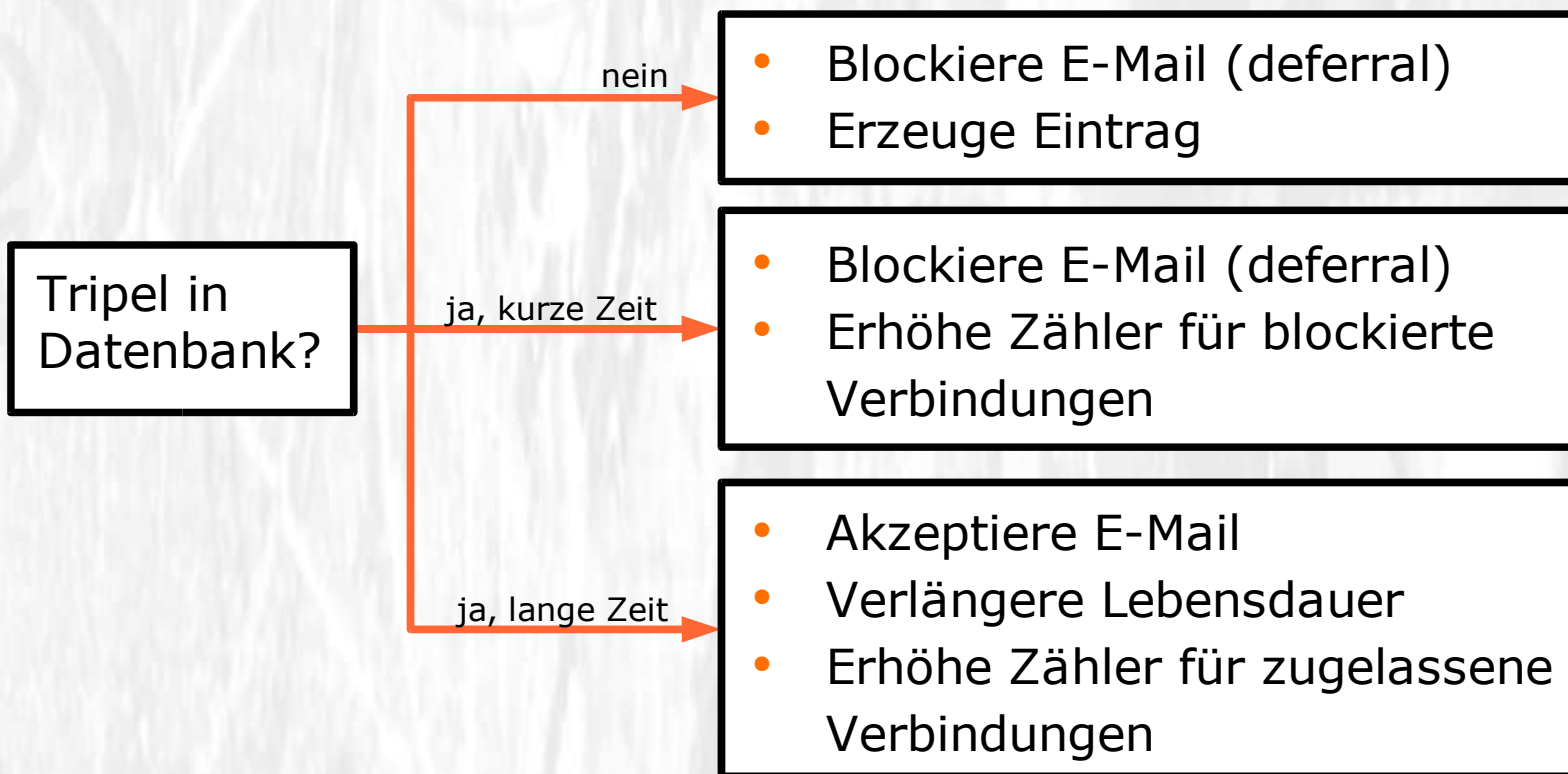
- Spammer geben den Versand von Spam nach einem Deferral auf.
- Spammer senden nicht mehrmals Nachrichten an den selben Empfänger (von der gleichen Absenderadresse)
- Spammer benutzen kein VERP (Variable Envelope Return Paths, <http://cr.yip.to/proto/verp.txt>)
- Spammer fälschen keine existierenden Absender-Adressen

Grey-Listing

Daten:

- **IP-Adresse des Sender-MTA**
 - **Envelope Sender (mail from:)**
 - **Envelope Empfänger (rcpt to:)**
 - Zeit der Erfassung
 - Verfallsdatum der Blockierung
 - Verfallsdatum des Eintrages
 - Anzahl geblockter Verbindungen
 - Anzahl zugelassener Verbindungen
- } Tripel

Grey-Listing



E-Mail-Analyse-Verfahren

- Typische Formulierungen im Betreff der E-Mail
- Typische Formulierungen im Body der E-Mail
- Typische Eigenschaften der E-Mail
- Mit Spam in Verbindung stehende Body URIs
- Checksummen-Verfahren (Razor, DCC, Pyzor)
- DNSBL/RBL-Prüfung der Received-Header
- Statistische Verfahren (Bayes)
- Absender in Whitelist/Blacklist

Typische Formulierungen für Spam

- „Viagra“, „Big Boobs“, „larger penis“, „ Impotence cure“
- „XXX Photos“, „Instant Access“, „Amateur Porn“
- „click here“, „for free“, „call now“, „unsubscribe“
- „Senate Bill 1618“, „This is not spam“, „opt-in“
- „Million Dollars“, „check or money order“, „pure profit“
- „University Diplomas“, „Toner Cartridge“, „Free DVD“
- „As seen on national TV!“, „No Medical Exams“
- „International driving license“, „NIGERIAN BANK“
- „gratis“, „Gewinnspiel“, „Hausfrauensex“, „Pu-Erh-Tee“
- „ist kein Spam“, „JETZT“, „seriös“, „anonym“, „abmelden“
- „ohne dialer“, „Sonderaktion“, „Vorteile sichern“

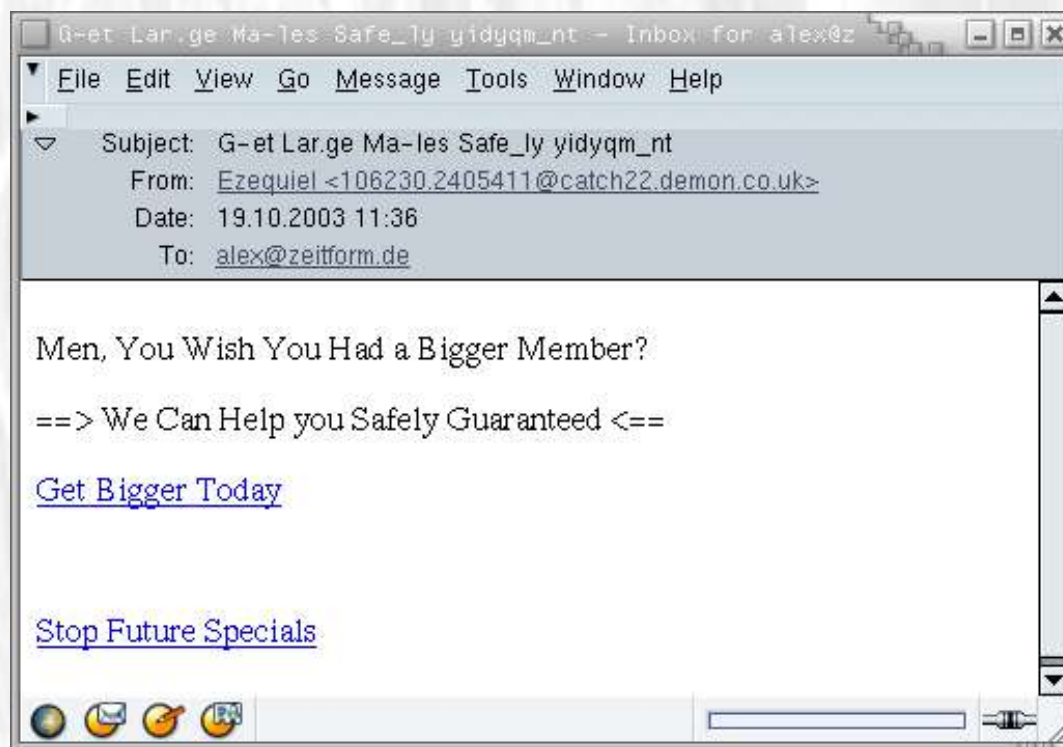
Typische Eigenschaften für Spam

- G.a.p.p.y T.e.x.t, GROSSBUCHSTABEN, Leerraum
- Fehlendes oder falsches Datum
- Gefälschter User-Agent oder X-Mailer
- Hinweise auf Spam-Tools
- Text ist base64-kodiert
- HTML-Mail mit font-Tags (size, color)
- HTML-Mail mit Bild(ern) und wenig/keinem Text
- HTML-Mail mit großgeschriebenen Tags
- HTML-Mail mit IP-Adressen oder Escaping in URLs
- ASCII-Formular [_____]
- Text enthält wirre Zeichenfolgen fgartz asdftr sfsdf dgd
- Received-Header sind gefälscht

Typische Eigenschaften legitimer E-Mail

- PGP-Signatur, Habeas Warrant Mark
- Text enthält Zitate „>“
- Gültiger User-Agent oder X-Mailer
- Eigenschaften von Mailinglisten-Software (majordomo)
- Vorhandene Signatur („-- “)
- Nachricht von Mailer-Daemon
- Hotmail- oder MSN-Footer
- In-Reply-To Header

Text-Analyse



28

Links: <http://www.getonitnow665.biz/bgite/>, <http://shapeisgood55.biz/re.php>

Text-Analyse

From: Ezequiel <106230.2405411@catch22.demon.co.uk>
Subject: G-et Lar.ge Ma-les Safe_ly yidyqm_nt
To: <alex@zeitform.de>

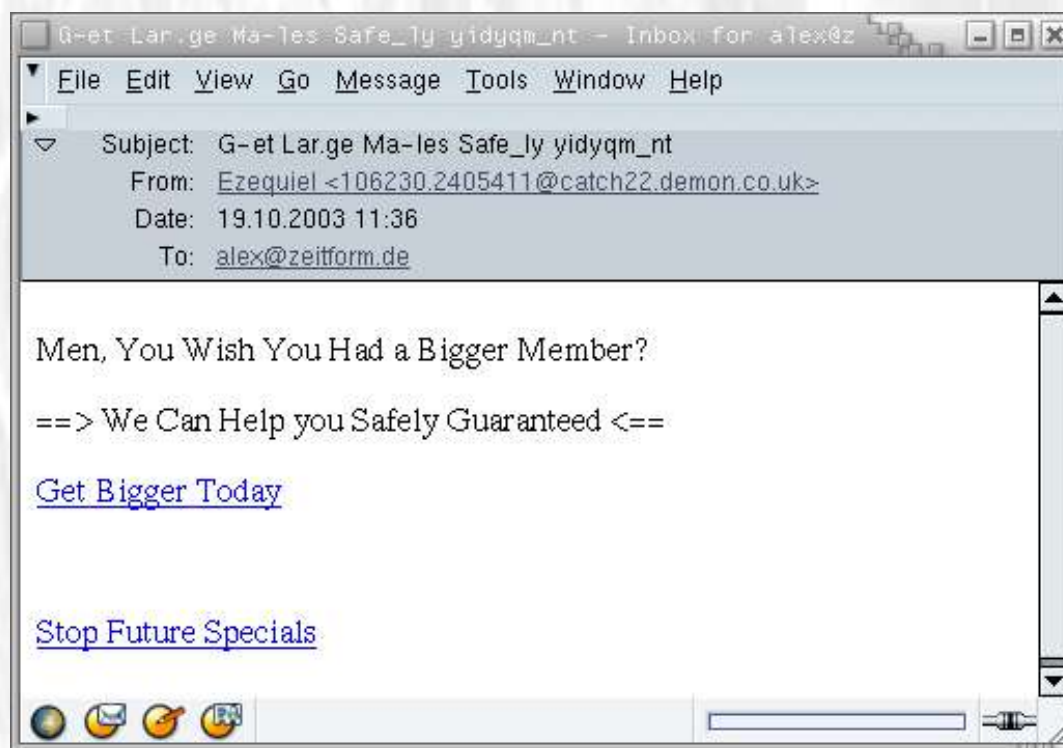
```
<html><body><font color=white>other writings whatsoever: because  
the subject where of they</font><br>Me<EauH>n, Yo<WuV>u Wi<my>sh  
Yo<Jlo>u Ha<Yz>d a<mB> Big<xiea>ger Mem<Mu>ber?<br><br>=<eZTN>=>  
W<ZjK>e Ca<jbtj>n He<Qx>lp yo<ftzP>u Saf<k>ely Guara<ner>nteed  
<<MKm>==<br><font color=white>Land and Conseil were slyly watching  
some of the ship's crew,</font><Br><a  
href="http://&#119;ww.g&#101;&#116;onit&#110;&#111;&#119;6&#54;&#5  
3;&#46;biz/b&#103;&#105;t/">G<w>et Big<XLA>ger To<TGvp>day</a>  
<br><br><br><br><a  
href="http://s&#104;apei&#115;&#103;o&#111;&#100;5&#53;. &#98;&#105  
&#101;&#112;&#104;&#112;">St<vTS>op Fut<MEzL>ure  
Spec<xR>ials</a><br><font color=white>To say that he risked his  
life twenty times before reaching</font>  
</html></body>
```

Text-Analyse

From: Ezequiel <106230.2405411@catch22.demon.co.uk>
Subject: G-et Lar.ge Ma-les Safe_ly yidyqm_nt
To: <alex@zeitform.de>

```
<html><body><font color=white>other writings whatsoever: because
the subject where of they</font><br>Me<EauH>n, Yo<WuV>u Wi<my>sh
Yo<Jlo>u Ha<Yz>d a<mB> Big<xiea>ger Mem<Mu>ber?<br><br>=<eZTN>=>
W<ZjK>e Ca<jbtj>n He<Qx>lp yo<ftzP>u Saf<k>ely Guara<ner>nteed
<<MKm>==<br><font color=white>Land and Conseil were slyly watching
some of the ship's crew,</font><Br><a
href="http://&#119;ww.g&#101;&#116;onit&#110;&#111;&#119;6&#54;&#5
3;&#46;biz/b&#103;&#105;t/">G<w>et Big<XLA>ger To<TGvp>day</a>
<br><br><br><br><a
href="http://s&#104;apei&#115;&#103;o&#111;&#100;5&#53;. &#98;&#105
;z/r&#101;. &#112;&#104;&#112;">St<vTS>op Fut<MEzL>ure
Spec<xR>ials</a><br><font color=white>To say that he risked his
life twenty times before reaching</font>
</html></body>
```

Body URIs



31

Links: <http://www.getonitnow665.biz/bgite/>, <http://shapeisgood55.biz/re.php>

Body URIs

- Links, Bilder, Skripte im Body von Spam
- Body URIs sollten verlässlich sein
- Kollateralschäden möglich

Beispiele:

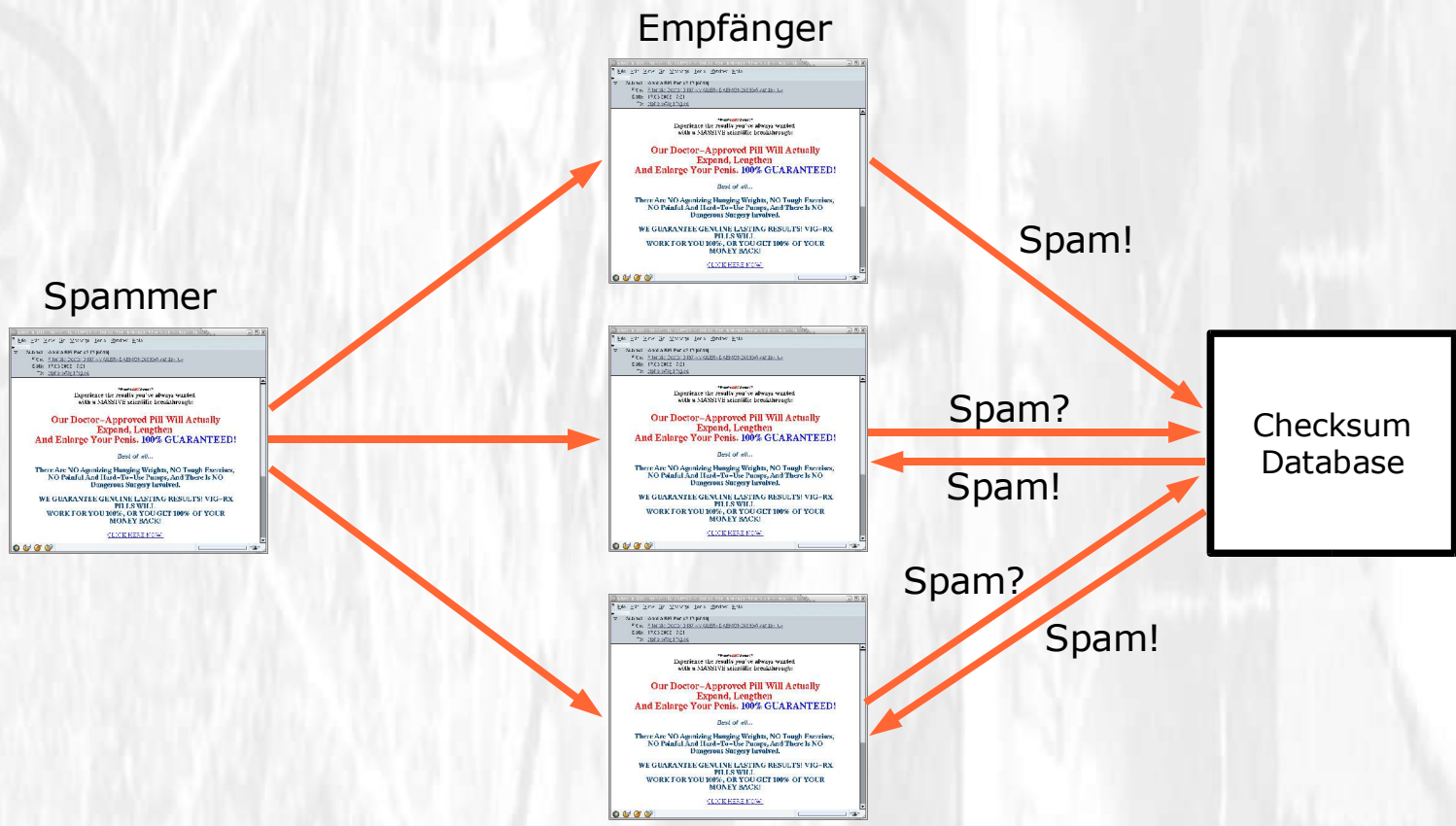
- bigevil.cf (SpamAssassin, Chris Santerre)
<http://www.merchantoverseas.com/wwwroot/gorilla/bigevil.cf>
- Filters That Fight Back (Paul Graham)
<http://www.paulgraham.com/ffb.html>
- RHSBL (analog DNSBL, vgl. RFC-Ignorant)
http://www.rfc-ignorant.org/how_to_domain.php

Checksummen-Verfahren

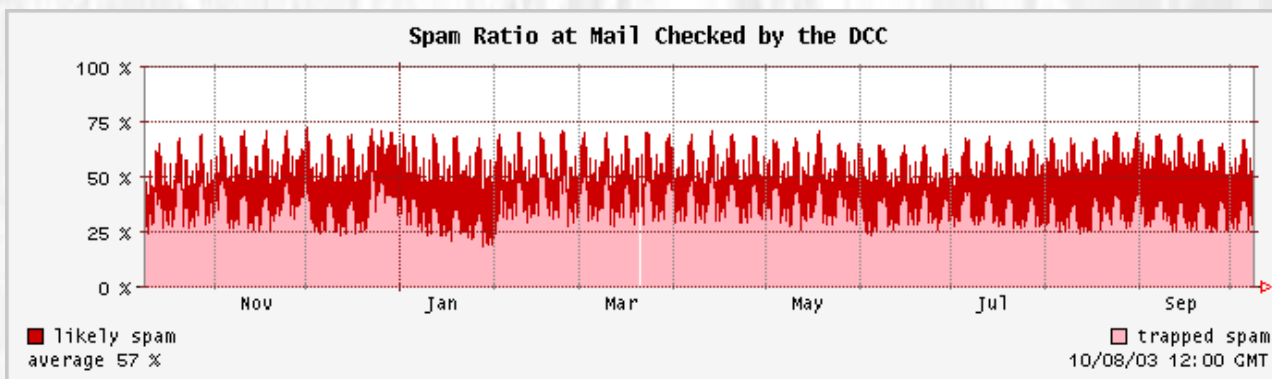
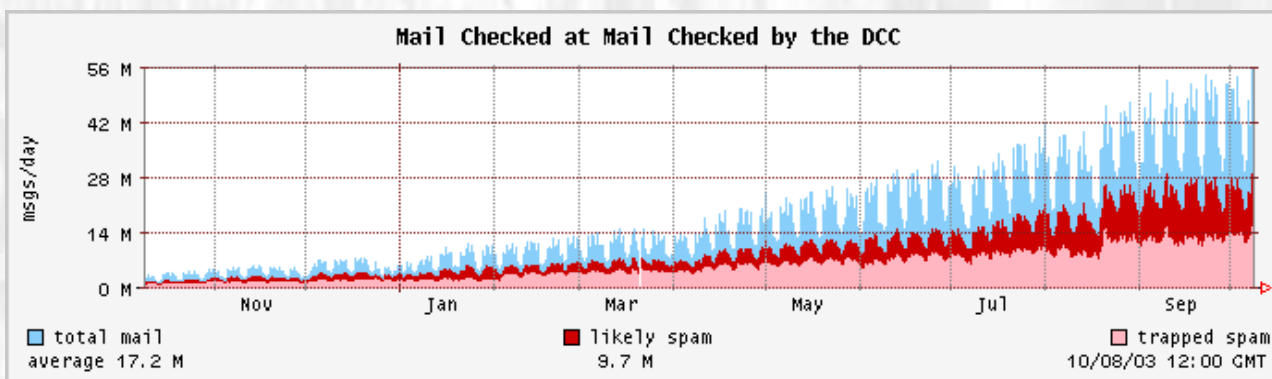
- DCC: <http://www.rhyolite.com/anti-spam/dcc/>
- Vipul's Razor: <http://razor.sourceforge.net/>
- Pyzor: <http://pyzor.sourceforge.net/>

- Öffentliche Datenbanken mit Fuzzy-Checksummen gemeldeter oder gefangener Spam-Mails
- Fuzzy-Checksummen-Algorithmen liefern bei geringen textuellen Änderungen gleichen Wert
- zusätzlicher Netzwerk-Verkehr für Abfrage

Checksummen-Verfahren



Checksummen-Verfahren



Quelle: <http://www.rhyolite.com/anti-spam/dcc/graphs/>

Razor2-Spam-Report

```
S: sn=N&srl=76&ep4=7542-10&a=1&a=cg
C: cn=razor-agents&cv=2.36
C: a=ai&cn=razor-agents&cv=2.36&user=alex%40zeitform.de
S: achal=gjiSrhu4_k_O3sCjulPKaIDNqaiX
C: a=auth&aresp=3WcztoZWFGU9GXF2PXh16SRMuSkA
S: res=1
C: -a=r&e=1&s=-kG19hd7RAXdxQB3NyQ8ZQ2_t9kA
C: a=r&e=4&ep4=7542-10&s=WkYZN7tIo2L_hrh5P-n5VFViVygA
C: a=r&e=4&ep4=7542-10&s=Imf__1UeBM3RhgTO_YRIW7v580UA
C: .
S: -res=1
S: err=230
S: err=230
S: .
C: -a=r&message=*
C: From joseph_ajakaiye@yahoo.com Wed Oct 8 11:40:46 2003
C: [...]
C: .
S: res=1
C: a=q
```

Razor2-Spam-Abfrage

Spam

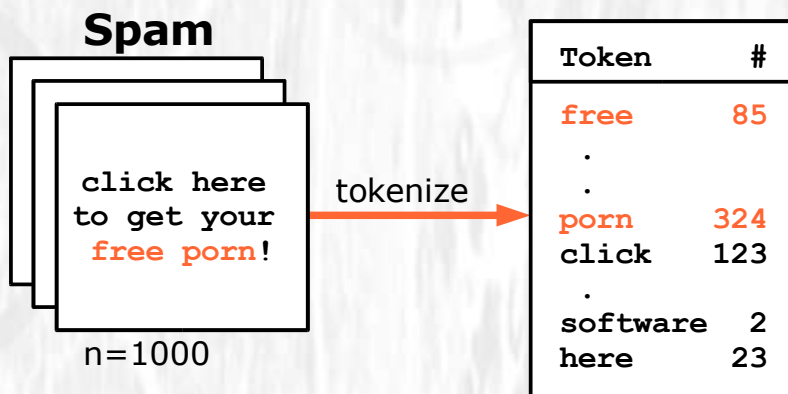
```
S: sn=C&sr1=72&ep4=7542-10&a=1  
C: a=c&e=4&ep4=7542-10&s=on7BdUVy4eW2TDWvVNpLfw1RUcoA  
S: p=1&cf=100  
C: a=q
```

Kein Spam

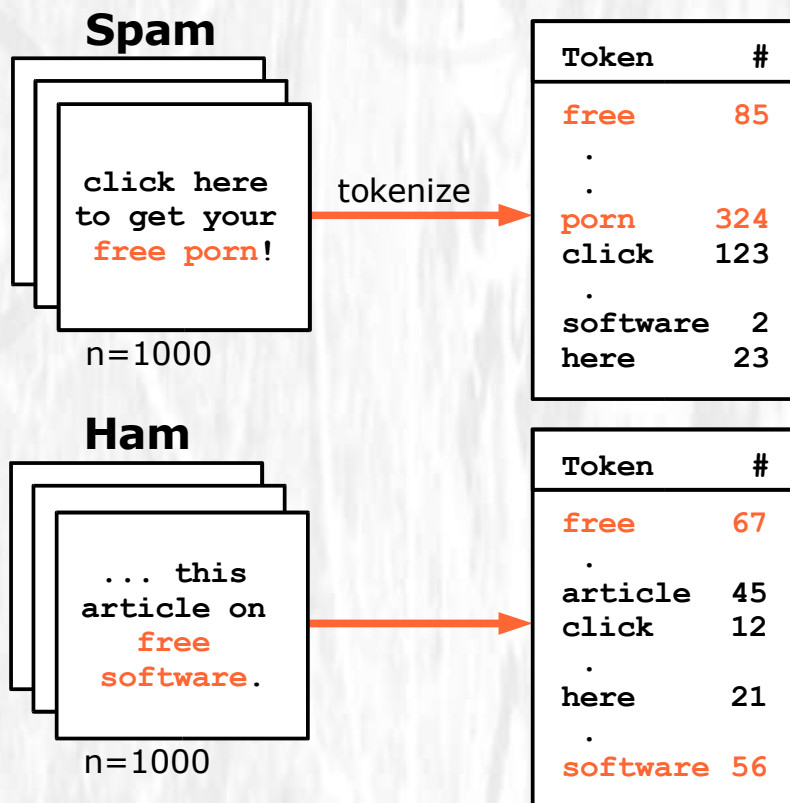
```
S: sn=C&sr1=72&ep4=7542-10&a=1  
C: a=c&e=4&ep4=7542-10&s=Id8u8ofcEDyRUht2D9DP6rMdlxYA  
S: p=0  
C: a=q
```

Info: <http://www.stearns.org/razor-caching-proxy/razor2-protocol>

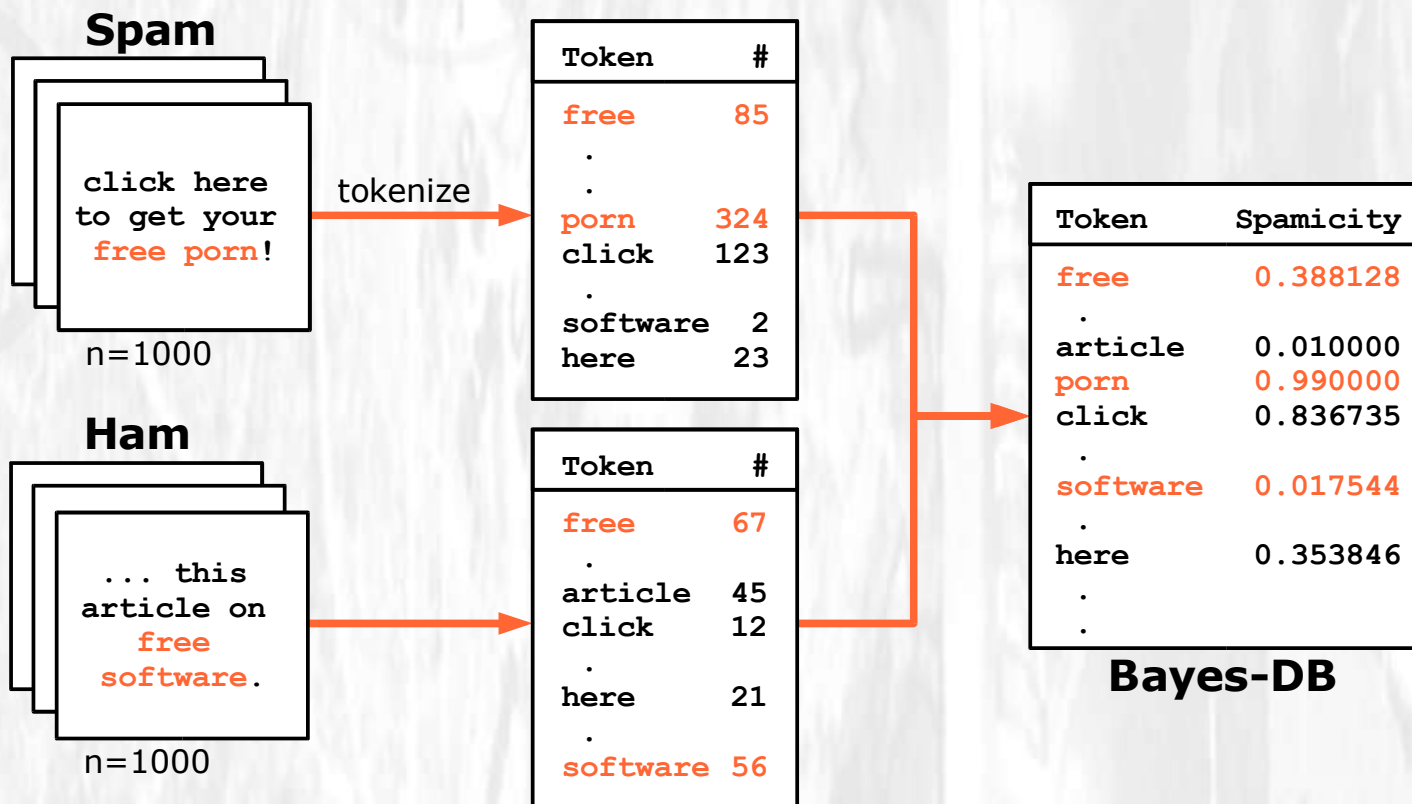
Bayes-Verfahren - Lernphase



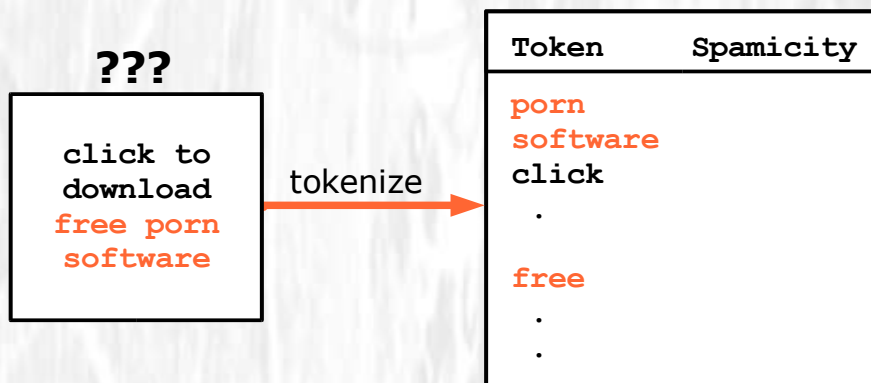
Bayes-Verfahren - Lernphase



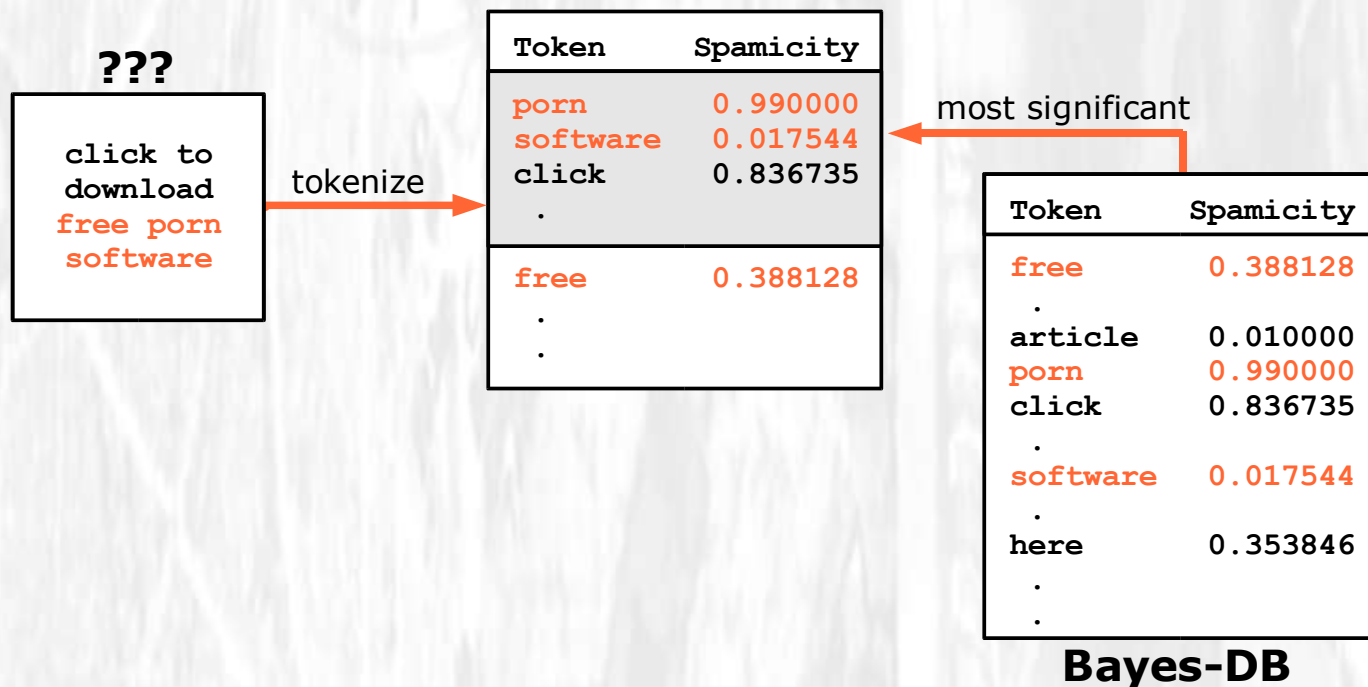
Bayes-Verfahren - Lernphase



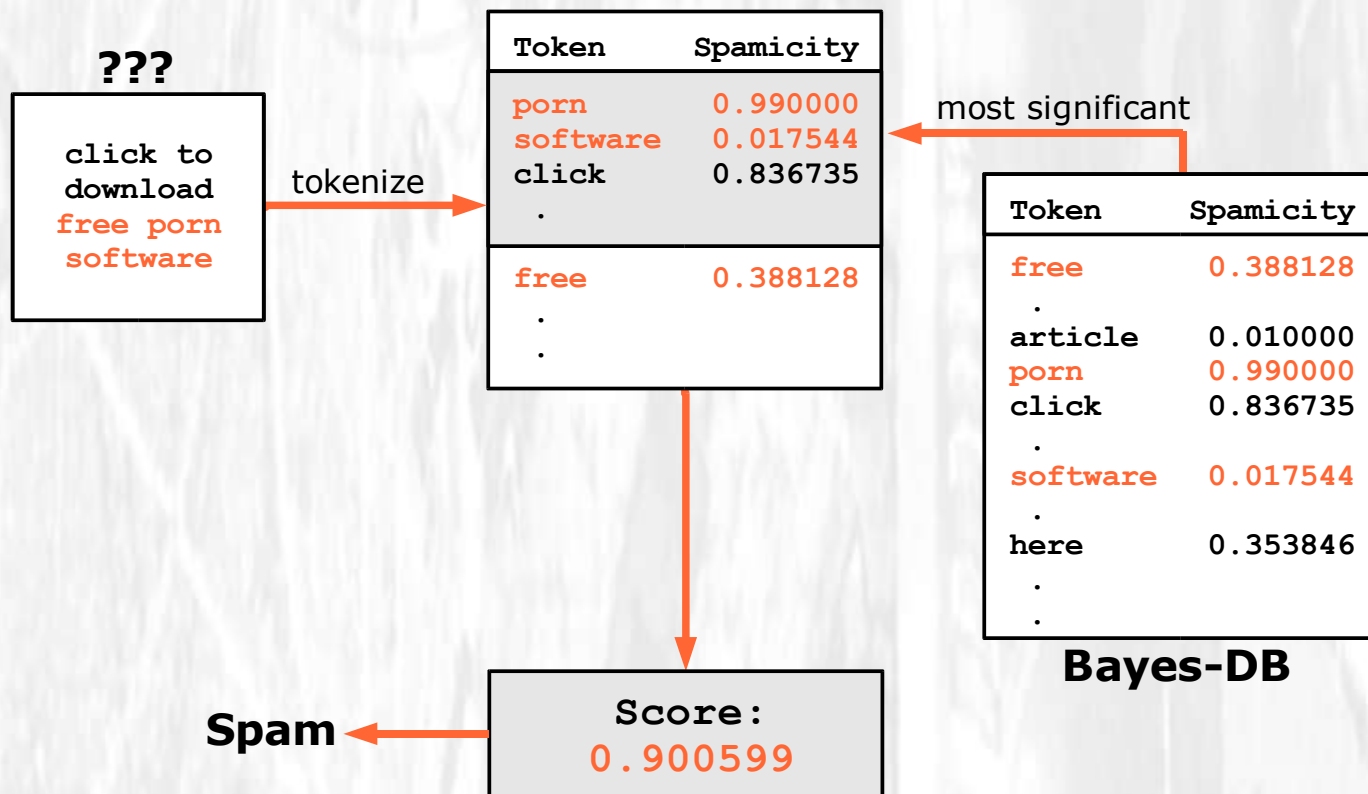
Bayes-Verfahren - Auswertungsphase



Bayes-Verfahren - Auswertungsphase



Bayes-Verfahren - Auswertungsphase



SpamAssassin

- <http://www.spamassassin.org/>
- Header Analyse
- Text Analyse
- Auto-Whitelist
- Bayes-Filter
- DNSBL, RHSBL
- Razor, Pyzor, DCC
- erweiterbar durch eigene Regeln (<http://www.exit0.us/>)
- flexible MTA Unterstützung
- Freie Software (Artistic License, APL ab Version 2.70)

SpamAssassin



SpamAssassin

X-Spam-Report:

- * 2.6 LOSE_POUNDS Subject talks about losing pounds
- * 0.6 J_CHICKENPOX_34 BODY: {3}Letter - punctuation - {4}Letter
- * 2.9 BANG_EXERCISE BODY: Talks about exercise with an exclamation!
- * 2.4 ALL_NATURAL BODY: Spam is 100% natural?!
- * 0.1 HTML_MESSAGE BODY: HTML included in message
- * 0.3 HTML_FONT_BIG BODY: HTML has a big font
- * 1.1 RAZOR2_CF_RANGE_51_100 BODY: Razor2 gives confidence between 51 and 100 [cf: 100]
- * 5.4 BAYES_99 BODY: Bayesian spam probability is 99 to 100%
- * 0.3 MIME_HTML_ONLY BODY: Message only has text/html MIME parts
- * 0.1 HTML_50_60 BODY: Message is 50% to 60% HTML
- * 0.1 HTML_FONTCOLOR_RED BODY: HTML font color is red
- * 3.0 BigEvilList_193 URI: Generated BigEvilList_193
- * 1.0 RAZOR2_CHECK Listed in Razor2 (<http://razor.sf.net/>)
- * 2.9 DCC_CHECK Listed in DCC (<http://rhyolite.com/anti-spam/dcc/>)
- * 2.0 DATE_IN_FUTURE_06_12 Date: is 6 to 12 hours after Received: date
- * 2.6 RCVD_IN_DYNABLOCK RBL: Sent directly from dynamic IP address [218.237.156.173 listed in dnsbl.sorbs.net]
- * 0.7 PLING_PLING Subject has lots of exclamation marks

SpamAssassin

```
X-Spam-Contact: Please contact postmaster@zeitform.de [...]  
X-Spam-Flag: YES  
X-Spam-Checker-Version: SpamAssassin 2.63-zeitform_1.11 (2004-01-11)  
on mail.zeitform.de  
X-Spam-Level: *****  
X-Spam-Status: Yes, hits=28.2 required=8.0 tests=[...] autolearn=spam  
version=2.63-zeitform_1.11  
X-Spam-Report: [...]
```

Empfänger-Verifikation

- Unsubscribe-Mail

For removal hit reply and put "Remove" in subject line

- „Click here“-Link

```
<a href="http://spammer.com/?id=your-email">Click here!</a>
```

- Unsichtbare Links

```

```

```

```

```
<iframe src="http://spammer.com/?id=your-email"></iframe>
```

```
<script language="JavaScript">....</script>
```

Empfänger-Verifikation

- Beispiel:

```

```

kriirccfix-ufdivx

tarrallorg-domreg (ROT17)



Adress-Sammler

- UseNet, Mailinglisten, Webseiten, Whois
- Online- und Offline-Formulare, Promotion
- Ident Daemon, Finger Daemon, IRC, Chat
- Web Browser (anon FTP, JavaScript, HTTP_FROM)
- AOL Profile (AOL Nutzer = Primärziele)
- „Yellow Pages“ (hotmail → bigfoot)
- Raten/Wörterbuchattacke
- Rechnerzugriff, Viren, Netzüberwachung, Social Engineering

Info: <http://www.private.org.il/harvest.html>

Spam-Viren

Beispiele: Mimapil, Sobig, Fizzer

- SMTP-Engine
- sammelt Adressen aus Outlook, Adressbuch, Dateien
- Joe Job/dDoS-Angriffe gegen Anti-Spam Community
- Fernadministration
- Bots, Würmer, HTTP-Server, Proxy
- Self-Update, Umgehung Anti-Virus-Schutz

Info: <http://www.spamhaus.org/cyberattacks/index.html>

Schutz vor Adress-Sammlern

- Verzicht auf Veröffentlichung der eigenen Adresse
- Adresse in Grafik einbetten (ganz/teilweise)

```
alex @ zeitform.de
```

- Adresse ausschreiben

```
alex at zeitform dot de, alex@zeitform.de-NOSPAM
```

- HTML-Entities, URL-Encoding

```
alex&#064;zeitform.de, alex%40zeitform.de
```

- JavaScript

```
<script language="JavaScript">  
  var Mailme = "alex@" + "zeitform.de";  
  document.write('<a href="mailto:' + Mailme + '>');  
  document.write(Mailme + '</a>');  
</script>
```

Adress-Sammler entdecken

- Verwendung einmaliger Adressen bzw. Erweiterungen
`alex+spamtrap@zeitform.de`
- Verwendung von SMTP-Kommentaren
`alex(spamtrap)@zeitform(comment).de`
- Verwendung dynamischer Adressen
`alex-78c1ed6da0322b3a@zeitform.de` (Spamtrap: IP, Datum)
- Access Log des Webservers
 - User-Agent
 - Keine angeforderten Bilder
 - Zugriffs-Häufigkeit und -Reihenfolge
 - CGI-Traps (z.B. wpoison)

Habeas Sender Warranted Email

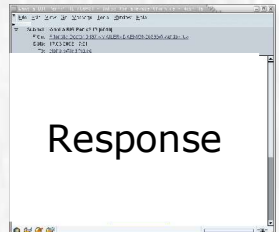
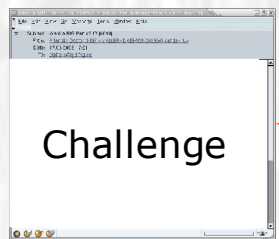
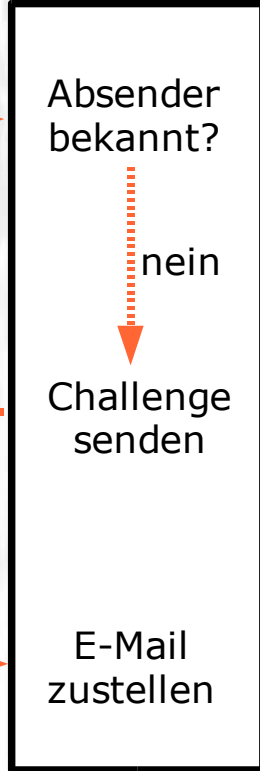
X-Habeas-SWE-1: winter into spring
X-Habeas-SWE-2: brightly anticipated
X-Habeas-SWE-3: like Habeas SWE (tm)
X-Habeas-SWE-4: Copyright 2002 Habeas (tm)
X-Habeas-SWE-5: Sender Warranted Email (SWE) (tm).
X-Habeas-SWE-6: email in exchange for a license for this Habeas
X-Habeas-SWE-7: warrant mark warrants that this is a Habeas Compliant
X-Habeas-SWE-8: Message (HCM) and not spam. Please report use of this
X-Habeas-SWE-9: mark in spam to <<http://www.habeas.com/report/>>.

Challenge-Response-Verfahren

Sender

C/R-System

Empfänger



Challenge-Response-Verfahren

- Reduziert Spam auf nahezu 0%
- Erhöhter Auswand für Absender
- Erhöhter Netzwerkverkehr
- Belästigt Unschuldige (gefälschter Absender, „Joe-Job“)
- Beispiel: TMDA (<http://tmda.net>)

Copyright (c) 2003-2004 zeitform Internet Dienste.
Permission is granted to copy, distribute and/or modify this document
under the terms of the GNU Free Documentation License, Version 1.2
or any later version published by the Free Software Foundation;
with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
A copy of the license can be found at: <http://www.gnu.org/licenses/fdl.txt>

History

v1.0: Erstveröffentlichung anlässlich des CAST-Workshops „Spam-Abwehr“
(<http://www.cast-forum.de/events/cast/2004/Spam-Abwehr/>)