

# **Netzwerksicherheit A**

## **Spam und Anti-Spam**

Eine technologische Betrachtung

**Christian J Dietrich**  
**dietrich@internet-sicherheit.de**

**Institut für Internet-Sicherheit**  
**<https://www.internet-sicherheit.de>**  
**Fachhochschule Gelsenkirchen**



**if(is)**  
internet-sicherheit.



- **Einleitung**
- **Definition und Voraussetzungen**
- **Quellen von Spam**
- **Anti-Spam-Techniken**
- **Auswirkungen in der Praxis**
- **Verhinderung / Vermeidung**
- **Ausblick**

- **Einleitung**
- Definition und Voraussetzungen
- Quellen von Spam
- Anti-Spam-Techniken
- Auswirkungen in der Praxis
- Verhinderung / Vermeidung
- Ausblick

# Begriffsklärung

- SPAM ≠ Spam!
  - SPAM – Markenname für Dosenfleisch der Fa. Hormel Foods
  - In den USA auch Software-Marke, also SPAM™ / SPAM®
- Ursprung ist ein Sketch aus Monty Python's Flying Circus
- UCE = Unsolicited Commercial Email
  - Unerwünschte kommerzielle Werbe-Email
- UBE = Unsolicited Bulk Email
  - Unerwünschte Massen-Email

# Definition: Spam

- Spam ist **unerwünschte, für den Empfänger wert-, nutz- und sinnlose E-Mail**
- „Unerwünscht“ ist individuell...
  - 92% bezeichnen unerwünschte Werbung als Spam
  - Werbung von politischen Gruppen oder Bürgervertretung: 74%
  - ... von Nonprofit- oder Wohltätigkeitsorganisationen nur noch 65%
- aber: Spam-Nachrichten haben gemeinsam:
  - Spam wird in Massen versendet
  - Es gibt einen geschäftlichen Hintergrund
  - Denial of Service

# Definition: False Negative/False Positive

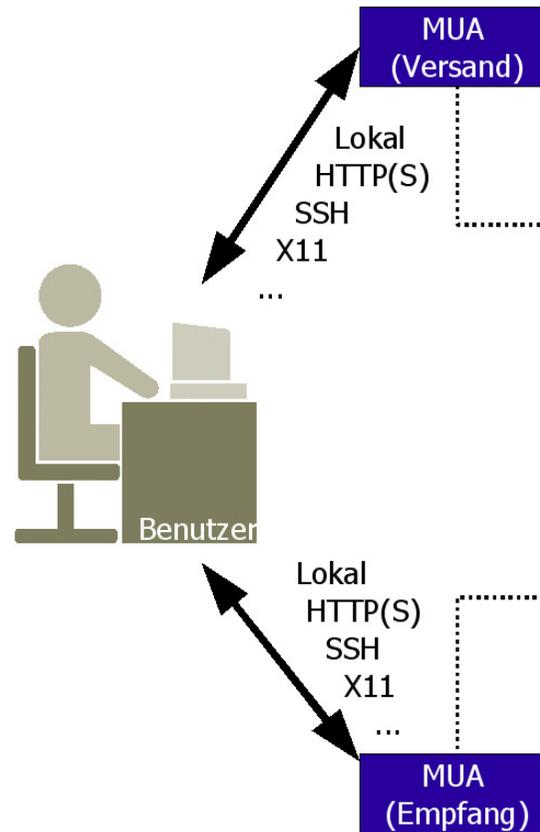
- Im Kontext Anti-Spam bezeichnet
- **False Negative**  
Eine Spam-Nachricht wird nicht als Spam erkannt (der Test fiel **fälschlicherweise negativ** aus)
- Problem: False Negative wird nicht gefiltert (landet in der Inbox)
- Typischerweise bei modernen Anti-Spam-Mechanismen < 0,1%
  
- **False Positive**  
Eine Ham-Nachricht wird fälschlicherweise als Spam erkannt (der Test fiel **fälschlicherweise positiv** aus)
- Problem: Legitime E-Mail wird gefiltert (evtl. destruktiv) – inakzeptabel
- Typischerweise 0,001% – 0,0001%  
(z.B. MessageLabs garantiert per SLA maximal 0,0004%)

# E-Mail als Infrastruktur

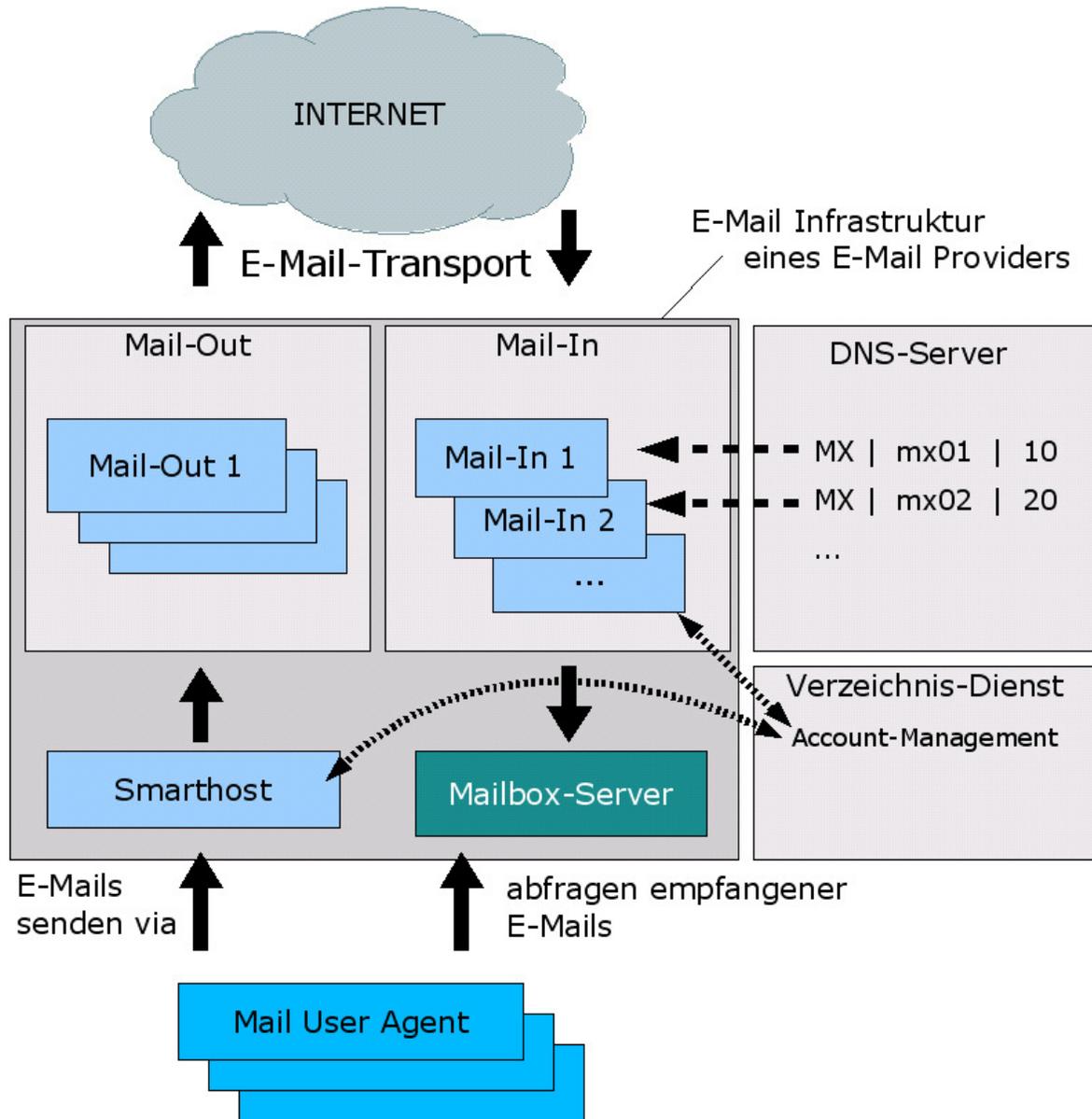
- E-Mail-Infrastruktur im Internet
- Infrastruktur aus Sicht eines E-Mail Service Providers

# Infrastruktur E-Mail im Internet

E-Mail-Zustellung im Internet bedeutet Übertragung per SMTP



# Infrastruktur eines E-Mail Service Providers



- Mail User Agent interagiert mit Smarthost und Mailbox-Server (hohe Verfügbarkeit aus Providersicht)
- Provider X versendet über den Mail-Out ausgehende E-Mails mit Empfängern bei Provider Y an Mail-In von Provider Y
- Account- und Identity-Management über Verzeichnisdienst
- MX RRs zeigen mit verschiedenen Prioritäten auf Mail-Ins

# Inhalt

- Einleitung
- Definition und Voraussetzungen
- **Quellen von Spam**
- Anti-Spam-Techniken
- Auswirkungen in der Praxis
- Verhinderung / Vermeidung
- Ausblick

# Quellen von Spam

- Spam-Server
- Open Relays
- Open Proxies
- Unsichere CGI-Skripte (formmail)
- Zombie PCs und Botnets
- Mailserver der Provider

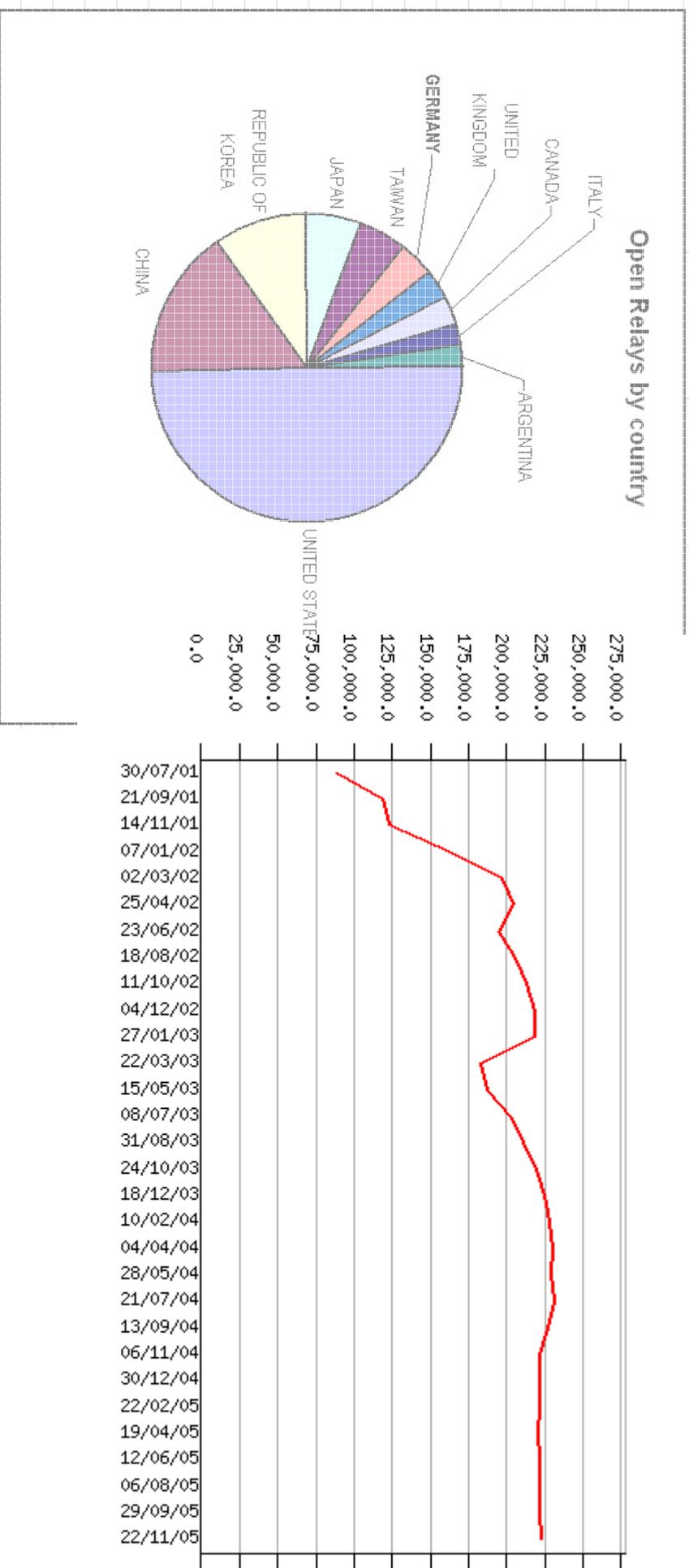
- Mailserver, die von Spammern zum Zweck des Spammings betrieben werden
  - Verursachen heutzutage immer noch einen gewissen Grad an Spam
  - Üblicherweise auf einfache Art und Weise anhand der festen IP-Adresse per Blacklist zu blockieren
- ➔ **rückläufig, da teurer als Missbrauch fremder Systeme**
- aktuell: ca. 5 Mio. IP-Adressen in gängigen Blacklists**

# Missbrauchte / schlecht konfigurierte Mail-Server

- Mailserver, die schlecht konfiguriert sind und z.B. basierend auf Absender-Domains im SMTP-Dialog als Open Relay agieren
  - Gekaperte Netzbereiche
  - Üblicherweise auf einfache Art und Weise anhand der festen IP-Adresse per Blacklist zu blockieren
- ➔ rückläufig (immerhin ca. 2,5 Mio. hijacked IP-Adressen!)

- Open Relay = Mailserver, der Emails weiterleitet, bei denen weder der Absender noch der Empfänger ein local user ist
- Problem: Spammer können Open Relays missbrauchen, ohne erkannt zu werden
- Heutzutage entweder
  - durch fehlerhafte Konfiguration oder
  - absichtlich – im Sinne des Internet („freie Infrastruktur“)
- Üblicherweise anhand der IP-Adresse z.B. per Blacklist zu blockieren

# Open Relays



- **Klassisch**
    - Fehlerhafte konfigurierte Proxies (z.B. SOCKS), die von Spammern zur Verschleierung des E-Mail-Pfades genutzt werden
    - Anonymizing Proxies
      - Z.B. JAP, TU Dresden, <http://anon.inf.tu-dresden.de/>
  - **Modern**
    - Überbegriff für Bots auf Zombie PCs
  - **Kein Received-Header in der Spam-Mail**
  - **Heute: ganze Proxy-Ketten zur Verschleierung**
- ➔ **klassisch (ca. 6,4 Mio. Hosts – Jan 2006) rückläufig  
modern auf dem Vormarsch**

# Unsichere CGI-Skripte / formmail

- (Frei verfügbare) Skripts zum Versand von E-Mails aus HTML-Formularen
- Problem: Verantwortung liegt beim Web-Server-Betreiber

→ rückläufig, da Versandprobe aufwändig und Anpassungsarbeit notwendig

**ca. 2,44 Mio. IP-Adressen**

(Stand Jan. 2006)

- Zombie PC = fernsteuerbarer, gekapeter Computer (Aldi-PCs)  
"Nur wegen der ‚Aldi-PC-Besitzer‘ kann sich Sober so stark verbreiten",  
sagt Prof. Dr. Norbert Pohlmann vom Institut für Internetsicherheit der  
Fachhochschule Gelsenkirchen."  
[Sueddeutsche]
  - Botnet = Zusammenschluss von Zombie PCs
  - Kontrollkanäle sind z.B. IRC oder HTTPS
  - Problem:
    - Always on (DSL-Flatrate) = hohe Verfügbarkeit
    - Verantwortung liegt beim PC-Besitzer
    - Spam-Versand wird vom Besitzer/Nutzer nicht bemerkt
  - Weltweit mehr als 1.000.000 Zombies (Honeynet Research)
  - Pro Botnet bis zu 400.000 (aktueller Trend: mehrere kleine – unauffälliger)
- ➔ **auf dem Vormarsch!**

# Zombie PCs / Botnets

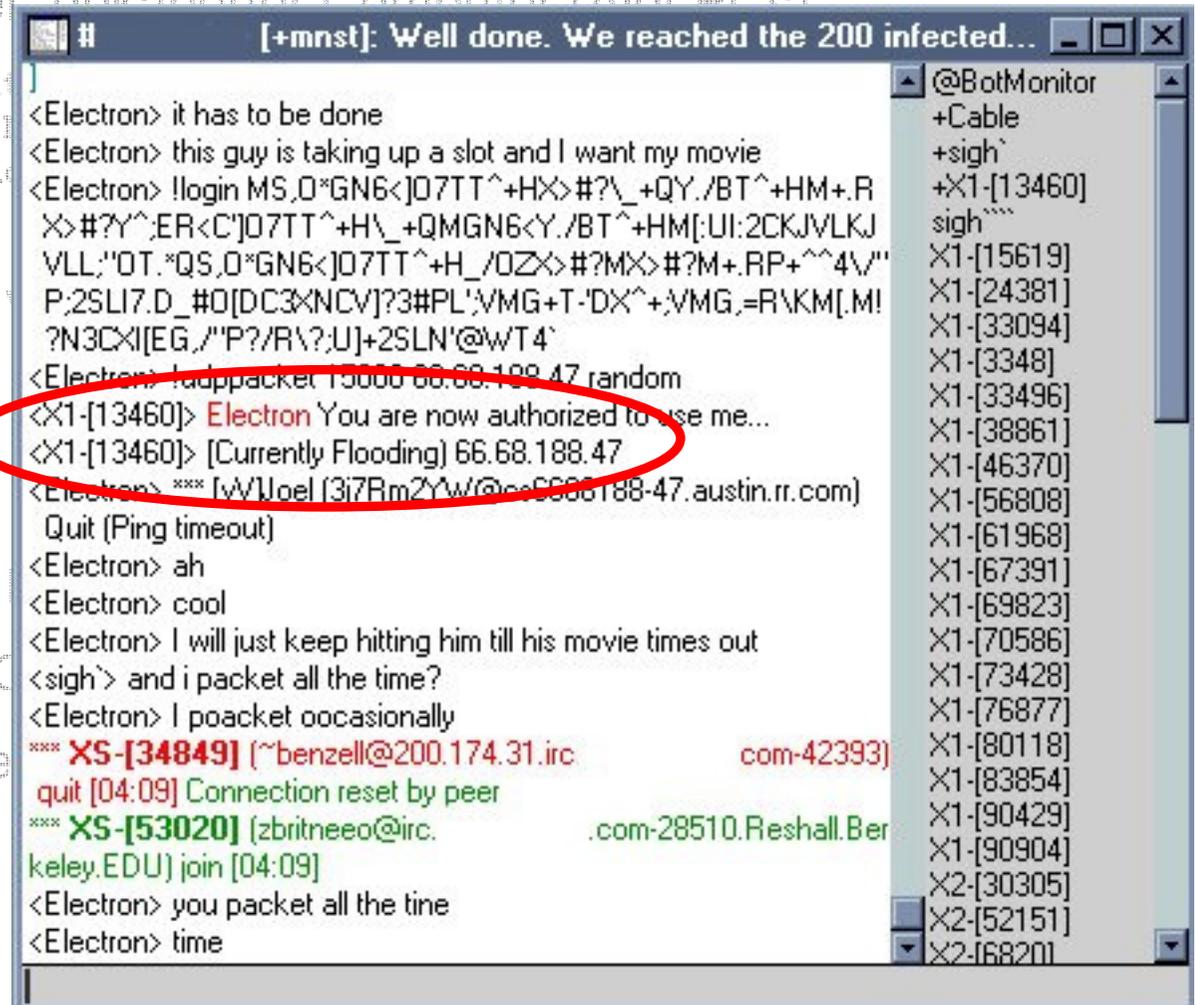
- Zombie PC = fernsteuerbarer, gekapneter Computer (Aldi-PCs)

"Nur wegen der ,Aldi-PC-Besitz' verbreiten", sagt Prof. Dr. Internet-sicherheit der Fachhochschule

- Botnet = Zusammenschluss von
- Kontrollkanäle sind z.B. IRC
- Problem:

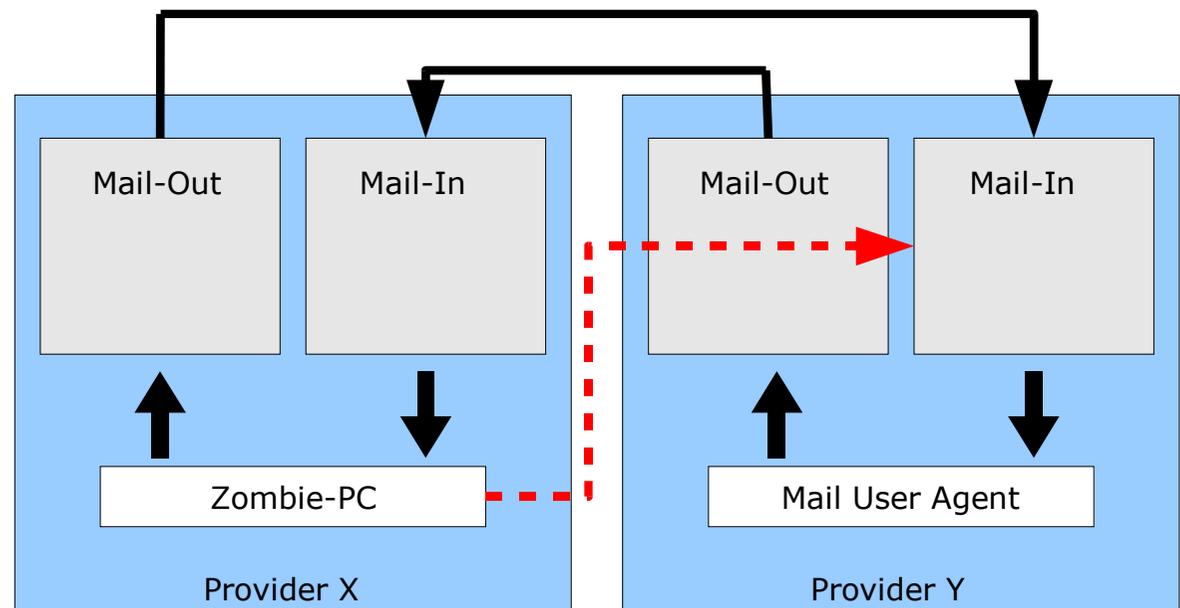
- Always on (DSL-Flatrate) =
- Verantwortung liegt beim PC
- Spam-Versand wird vom Be

→ auf dem Vormarsch!



# Das Zombie-Problem

- Zombie (PC) = gekapert, fernsteuerbarer PC (der Benutzer merkt nicht, dass sein PC missbraucht wird)
- Zombies werden zu sog. Botnetzen zusammengefasst
- Botnetze sind Quelle eines Großteils an Spam
- Zombies fluktuieren stark und sind aufgrund wechselnder dynamischer Dialup-IP-Adressen schwer „greifbar“
- Wenige ISPs bzw. NSPs geben ihre Dialup-IP-Adressbereiche freiwillig in öffentliche Blacklists



- „Jeder Provider hat früher oder später einen Spammer als Kunden“
  - **„pink contracts“**  
Vertragsverhältnis, der zu deutlich höheren Kosten den Verstoß u.a. gegen AGBs (z.B. durch Spam-Versand) toleriert 
  - Beispiele: PSINet, AT&T
  - Konflikt zwischen Network Service Provider und E-Mail Service Provider
- ➔ **Heutzutage größtenteils als Spam-Quelle kein Problem mehr**

# Inhalt

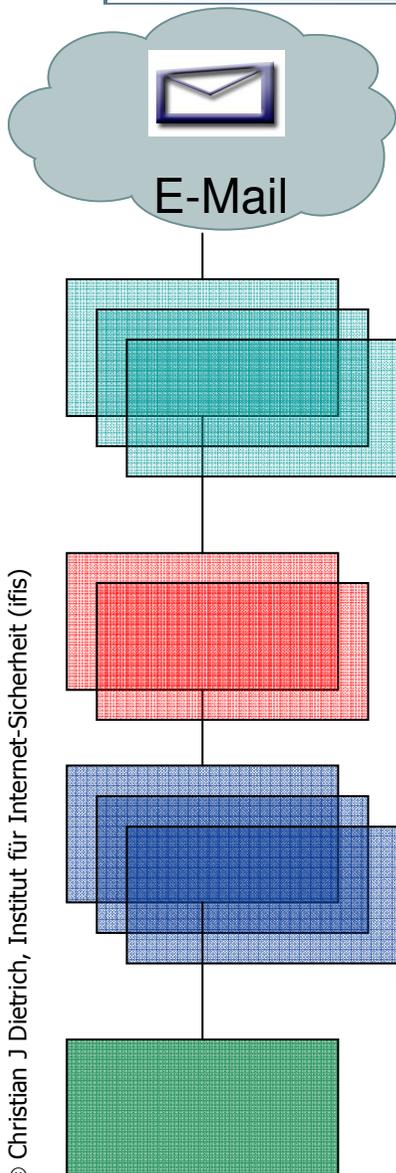
- Einleitung
- Definition und Voraussetzungen
- Quellen von Spam
- **Anti-Spam-Techniken**
- Auswirkungen in der Praxis
- Verhinderung / Vermeidung
- Ausblick

# Anti-Spam-Techniken

- Das Ebenen-Modell
- TCP/IP-Ebene
- SMTP-Ebene
- MARID-Verfahren
- Inhaltsbasierte Verfahren
- Praxis

# Anti-Spam-Techniken

## Das Ebenen-Modell



### Ext. E-Mail-Gateway / E-Mail-Proxy als Teil einer Firewall

- Checks auf IP-Ebene (IP-Adresse)
  - Blacklists (RBLs, Dynamische-/Dial-Up-IP, open relay, ...)
  - Reverse MX
  - Frequenzmessung
- Checks auf SMTP-Ebene
  - Überprüfen der HELO-Angabe
  - Überprüfen der Absender-E-Mail-Adresse (Black-/White-/Greylist)
  - Existenz der Empfänger-E-Mail-Adresse (DB, Verzeichnisdienst)

1

### Spam-Filter

- Checks auf Nachrichten-Ebene
  - Wortliste
  - Hash/Signatur

2

### Viren-Filter

- Check Nachricht und Anhänge auf Virenbefall

3

### Interner E-Mail-Server

Ressourcenverbrauch

- Verfahren:  
IP-Adresse der konnektierenden Partei wird in einer schwarzen Liste (Blacklist) nachgeschlagen (kann genauso für Whitelist benutzt werden)
- Methode ist lange bekannt
- In der Praxis:
  - In der Regel DNS-basiert (sog. DNSBLs)
  - Anfrage nach 194.94.127.5: 5.127.94.194.dns.bl.betreiber
- Interessante Blacklists:
  - Spamhaus Block List Manuelle Prüfung (known spammers)
  - XCL/CBL Kombinationslisten
  - SORBS Open Relays, Open Proxies, Dyn IPs
  - ORDB Open Relays
  - RFC-Ignorant.org Listet Nicht-RFC-konforme Hosts
  - Siehe auch <http://rbls.org>

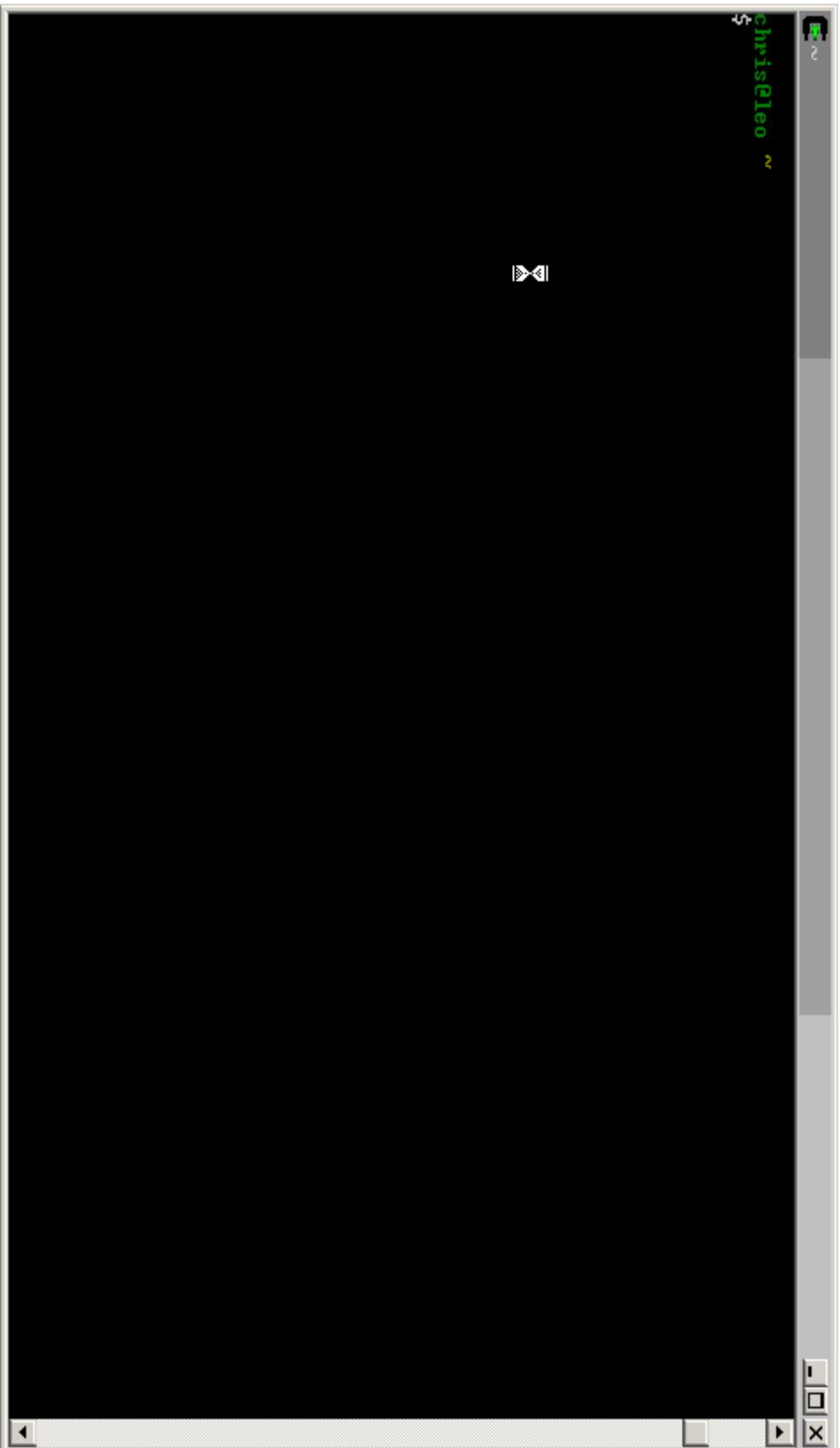
# TCP/IP-Ebene

- Wirksamste Technik: Frequenzanalyse
- Anzahl an Verbindungsaufbauten (SYN-Pakete) / Zeit
- Bei Überschreiten eines Limits wird die Verbindung zwar auf TCP-Ebene zugelassen, aber im SMTP-Dialog ein Fehler der Kategorie Fatal / Non-Temporary zurückgeliefert

- Beispiel:

```
Dec 6 19:23:12 pluto postfix/smtp[24897]: 52948354079:
host mx-ha01.web.de[217.72.192.149] refused to talk to
me: 421 mx19.web.de: Too many concurrent SMTP
connections; please try again later
...
Dec 6 19:23:18 pluto postfix/smtp[24897]: 52948354079:
to=[xxxx@web.de], relay=mx-ha02.web.de[217.72.192.188],
delay=7, status=deferred (host mx-
ha02.web.de[217.72.192.188] refused to talk to me: 421
mx05.web.de: Too much load; please try again later)
```

# DNSBL – Beispiel



# URI basierte Blacklists

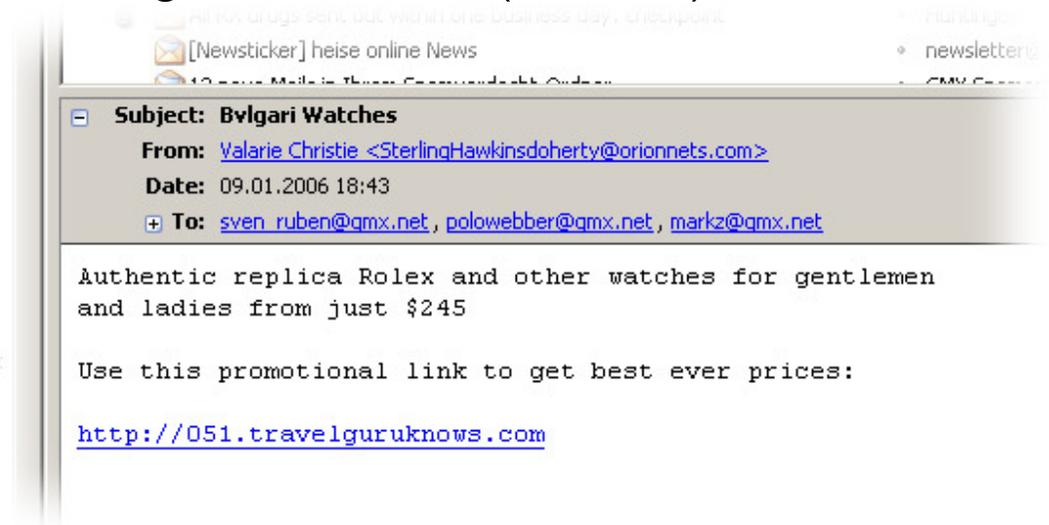
- Keine klassische Blacklist auf der 1. Ebene
- Basiert auf dem Inhalt einer Nachricht, genauer URIs (SURBL)

- Vorteile:

- Spammer müssen URIs in den Nachrichten verwenden
- Einfaches Verfahren: pattern matching

- Nachteile:

- URIs können dynamisch sein bzw.
- auf Wegwerf-Domains verweisen
- (Text in Bildern wird (ohne OCR) nicht erkannt)
- DoS-Szenario denkbar



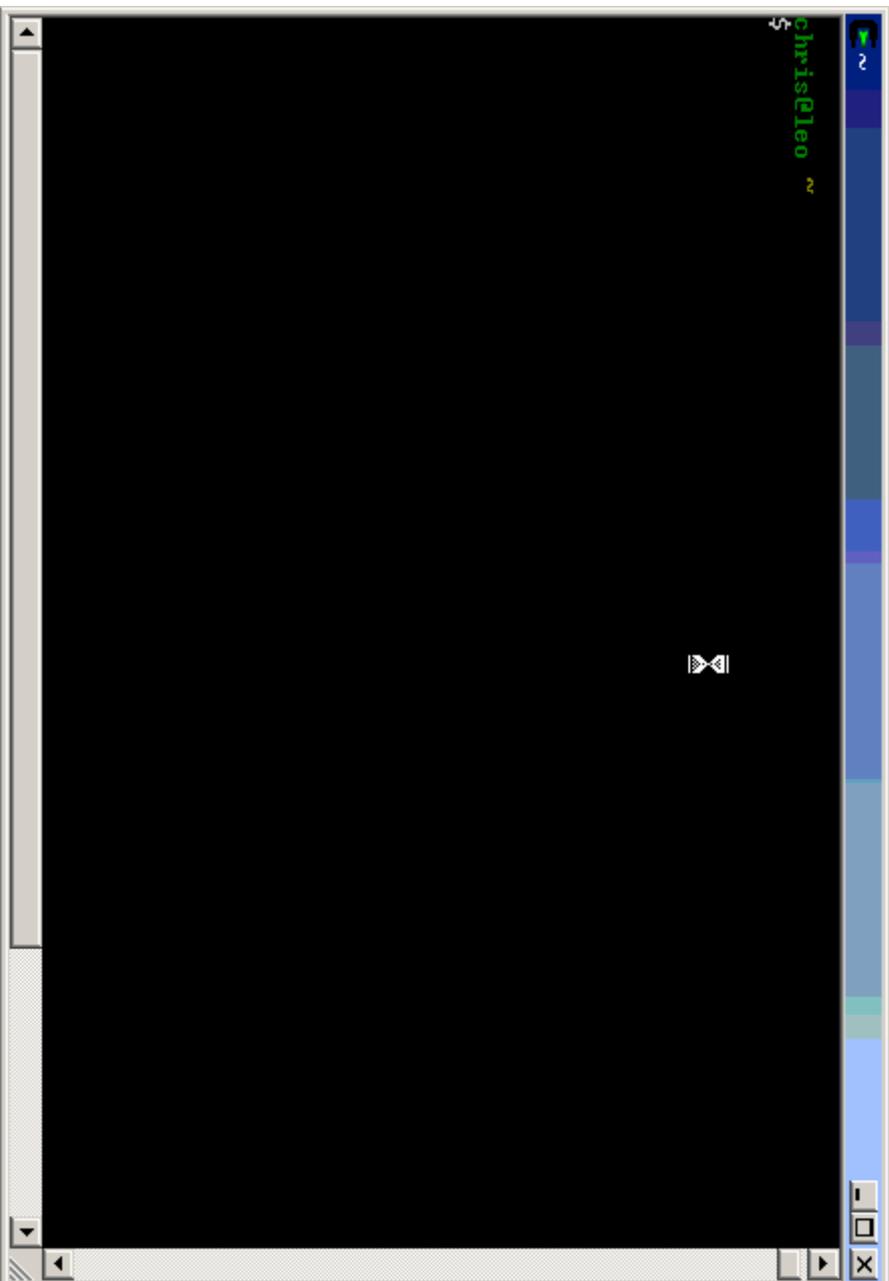
# Blacklists – Bewertung

- Vorteile
  - Hohe Geschwindigkeit, geringe Performance notwendig
  - Einfache Implementierung (gängige MTAs können mit DNSBLs umgehen)
  - weit verbreitet
  
- Nachteile
  - Entscheidung auf Host-Ebene, nicht Nachrichten-Ebene (kann für ISPs problematisch sein)
  - Quellinformationen sind aufwändig und nicht immer zweifelsfrei zu eruieren (z.B. dynamische IP-Adressbereiche)

# Checks auf SMTP-Ebene

- SMTP hat viele Details
- SMTP lässt sich streng auslegen
- Spam-Software hält sich nicht immer an RFCs
- Und vieles mehr...

# SMTP-Dialog (Beispiel)



# SMTP-Dialog (Beispiel)

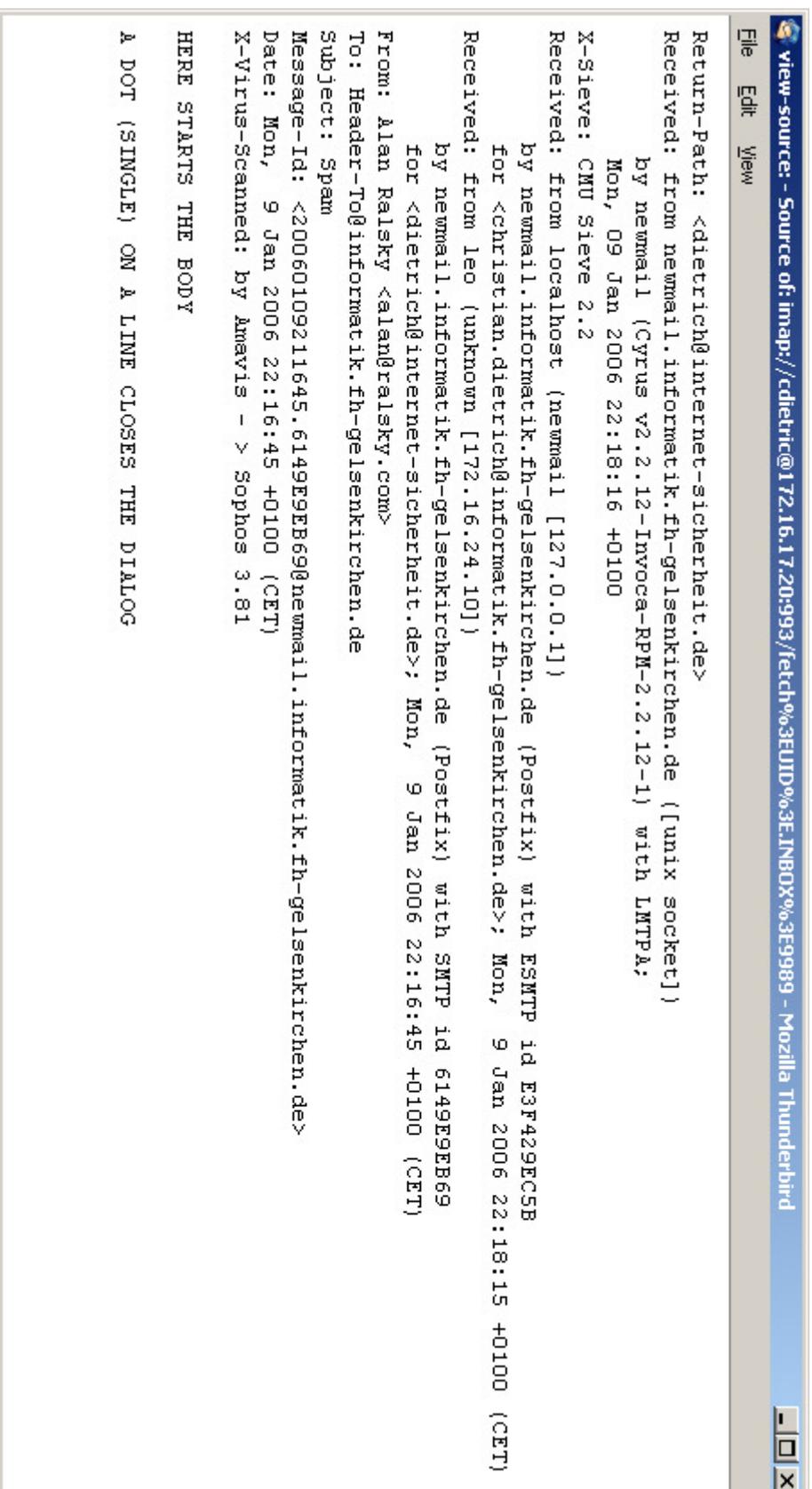
```
chris@leo ~
$ nc 172.16.17.20 25
220 newmail.informatik.fh-gelsenkirchen.de ESMTP Postfix
HELO leo
250 newmail.informatik.fh-gelsenkirchen.de
MAIL FROM:<dietch@internet-sicherheit.de>
250 Ok
RCPT TO:<dietch@internet-sicherheit.de>
250 Ok
DATA
354 End data with <CR><LF>.<CR><LF>
From: Alan Ralsky <alan@ralsky.com>
To: Header-To
Subject: Spam

HERE STARTS THE BODY
A DOT <SINGLE> ON A LINE CLOSSES THE DIALOG
.
250 Ok: queued as 6149E9EB69
QUIT
221 Bye

chris@leo ~
$
```



# SMTP-Dialog (Beispiel)

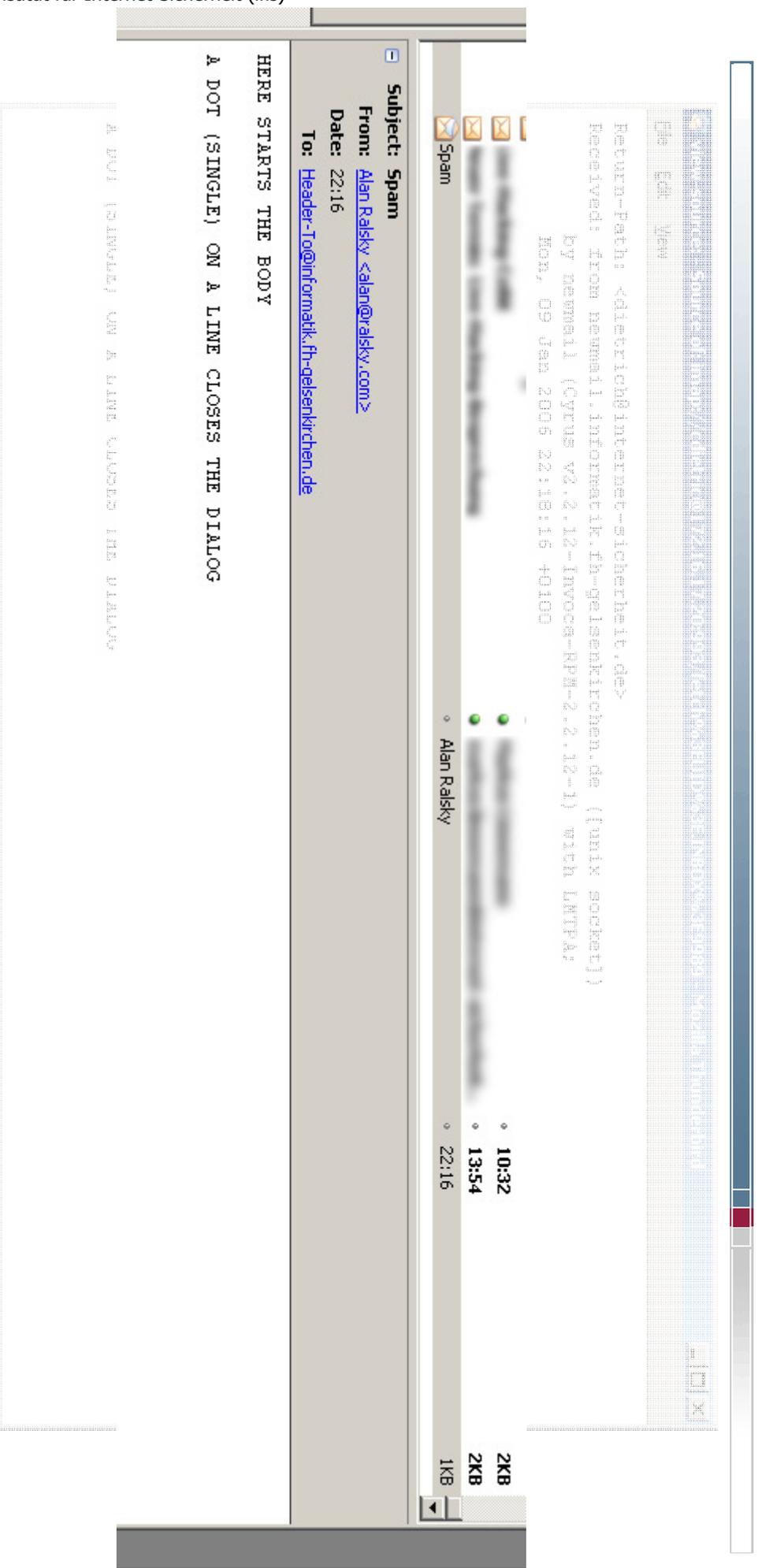


```
Return-Path: <dietrich@internet-sicherheit.de>
Received: from newmail.informatik.fh-gelsenkirchen.de ([unix socket])
    by newmail (Cyrus v2.2.12-Invoca-RPM-2.2.12-1) with LMTPA;
    Mon, 09 Jan 2006 22:18:16 +0100
X-Sieve: CMU Sieve 2.2
Received: from localhost (newmail [127.0.0.1])
    by newmail.informatik.fh-gelsenkirchen.de (Postfix) with ESMTP id E3F429EC5B
    for <christian.dietrich@informatik.fh-gelsenkirchen.de>; Mon,  9 Jan 2006 22:18:15 +0100 (CET)
Received: from leo (unknown [172.16.24.10])
    by newmail.informatik.fh-gelsenkirchen.de (Postfix) with SMTP id 6149E9EB69
    for <dietrich@internet-sicherheit.de>; Mon,  9 Jan 2006 22:16:45 +0100 (CET)
From: Alan Ralsky <alan@ralsky.com>
To: Header-To@informatik.fh-gelsenkirchen.de
Subject: Spam
Message-Id: <20060109211645.6149E9EB69@newmail.informatik.fh-gelsenkirchen.de>
Date: Mon,  9 Jan 2006 22:16:45 +0100 (CET)
X-Virus-Scanned: by Amavis - > Sophos 3.81

HERE STARTS THE BODY

A DOT (SINGLE) ON A LINE CLOSES THE DIALOG
```

# SMTP-Dialog (Beispiel)



Return-Path: <christian@internet-sicherheit.de>  
Received: from remote1.informatik.fh-gelsenkirchen.de (mail.socke1)  
by remote1 ( Cyrus v2.2.12-Invoker-RPM-2.2.12-1) with SMTP  
Mon, 09 Jan 2006 22:18:16 +0100

Subject: Spam  
From: Alan Ralsky <alan@ralsky.com>  
Date: 22:16  
To: [Header-To@informatik.fh-gelsenkirchen.de](mailto:Header-To@informatik.fh-gelsenkirchen.de)

HERE STARTS THE BODY

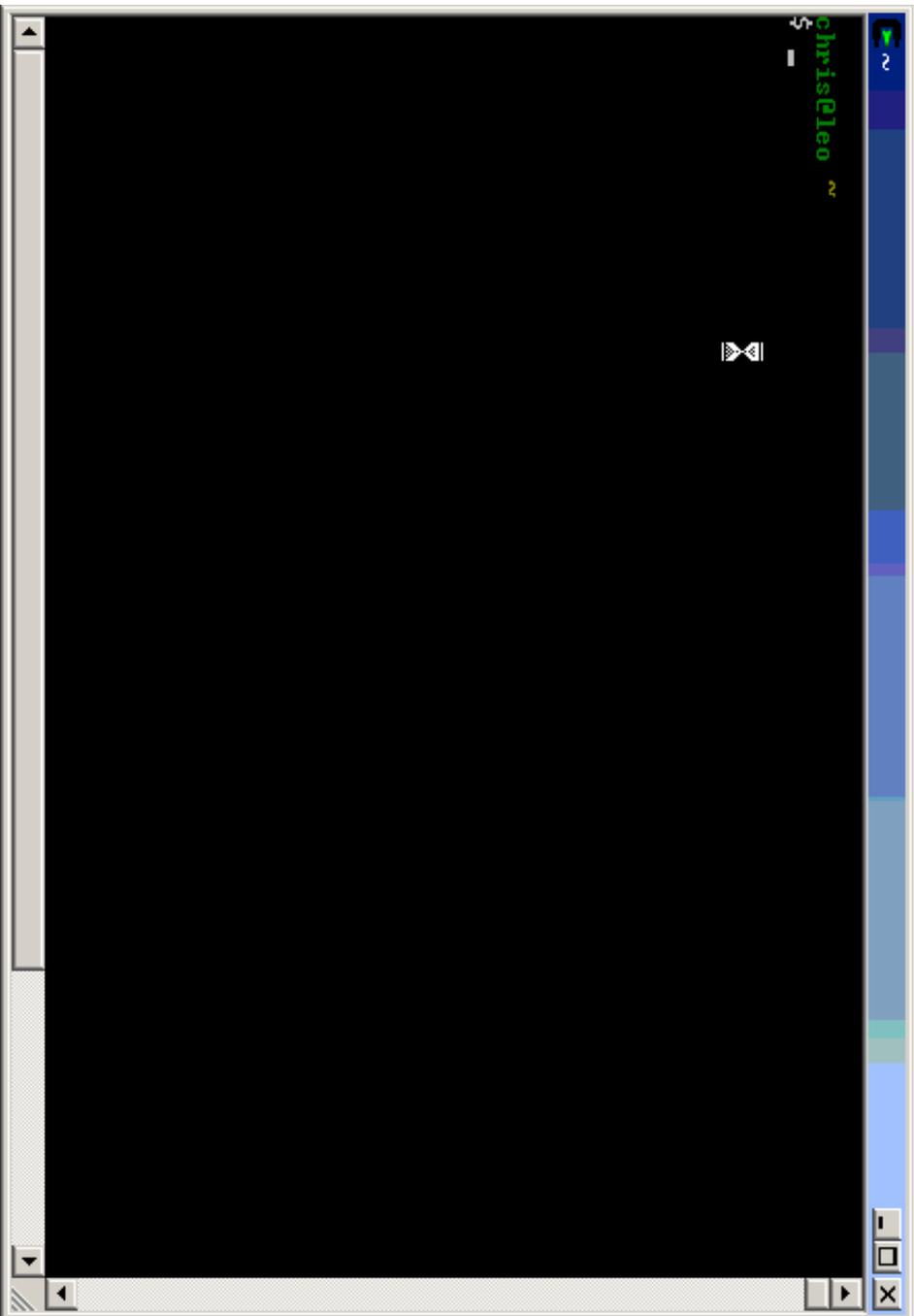
A DOT (SINGLE) ON A LINE CLOSES THE DIALOG

A DOT (SINGLE) ON A LINE CLOSES THE DIALOG

# Checks auf SMTP-Ebene: HELO

- Der erste Befehl einer SMTP-Sitzung ist HELO (Begrüßung)
- Das Argument ist laut RFC der Hostname des Clients
- Gibt der Client einen falschen oder syntaktisch inkorrekten Hostname an, verstößt er gegen RFC2821.
- > 60% der Client-Software überträgt falsches HELO-Argument!
- **Wenige Provider filtern dieses Kriterium**  
Problem: NAT-maskierte Rechner (Hostname muss von außen aufgelöst werden)

# Checks auf SMTP-Ebene: HELO



# SMTP-Feinheiten / Parlez-vous RFC?

- ESMTP Pipelining  
Erweiterung um mehrere Kommandos abzusetzen
- RFC-konforme Implementierung muss nach HELO/EHLO und DATA warten (Spam-Software – wenn sie überhaupt Pipelining verwendet – tut dies in der Regel nicht)
- Transport Layer Security (TLS)
- Leeres Mail-From (laut RFC nur bei bounces erlaubt)

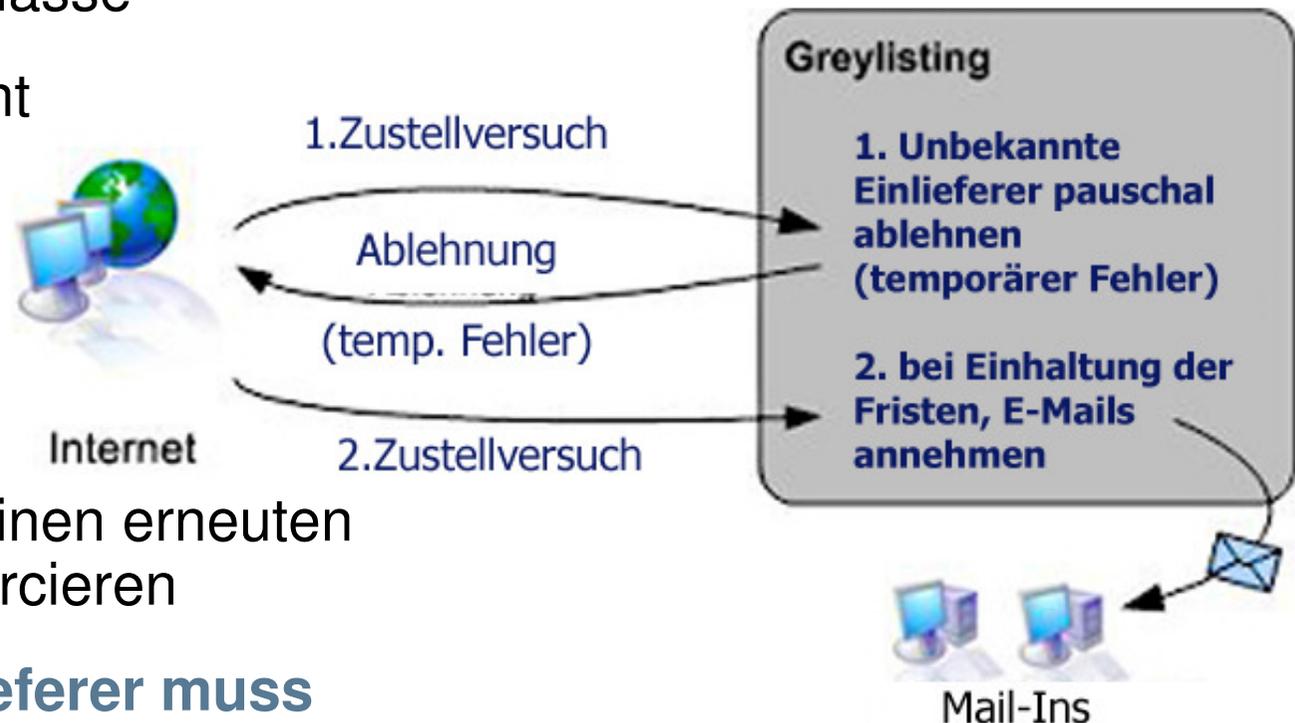
```
Nicht autorisierte Antwort:  
web.de MX preference = 110, mail exchanger = mx-ha02.web.de  
web.de MX preference = 100, mail exchanger = mx-ha01.web.de
```

- Besonderes Filtern auf Secondary MX
- Kurze Timeouts
- Greylisting

```
chris@leo ~  
$ nc mx-ha01.web.de 25  
220 WEB.DE  
550 Connection timed out.  
  
chris@leo ~  
$
```

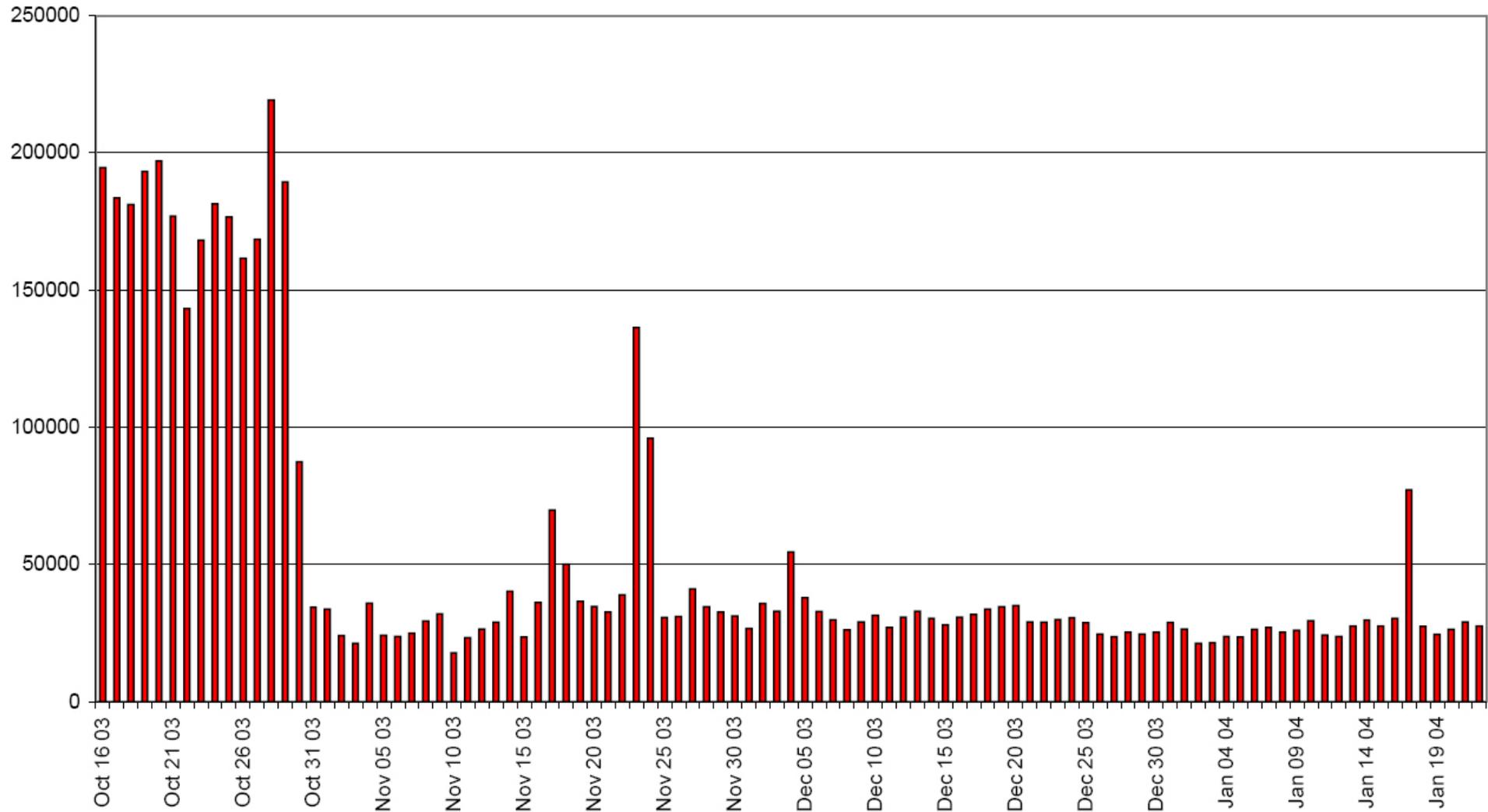
# Greylisting

- SMTP definiert eine non-fatale (temporäre) Fehlerklasse
- Ursprünglich gedacht für kurzfristige Probleme auf Seite des Empfängers
- Bei Greylisting wird diese Fehlerklasse „missbraucht“, um einen erneuten Zustellversuch zu forcieren
- Kernidee: **Der Einlieferer muss die Fristen laut RFC einhalten, nur dann werden seine E-Mails angenommen**

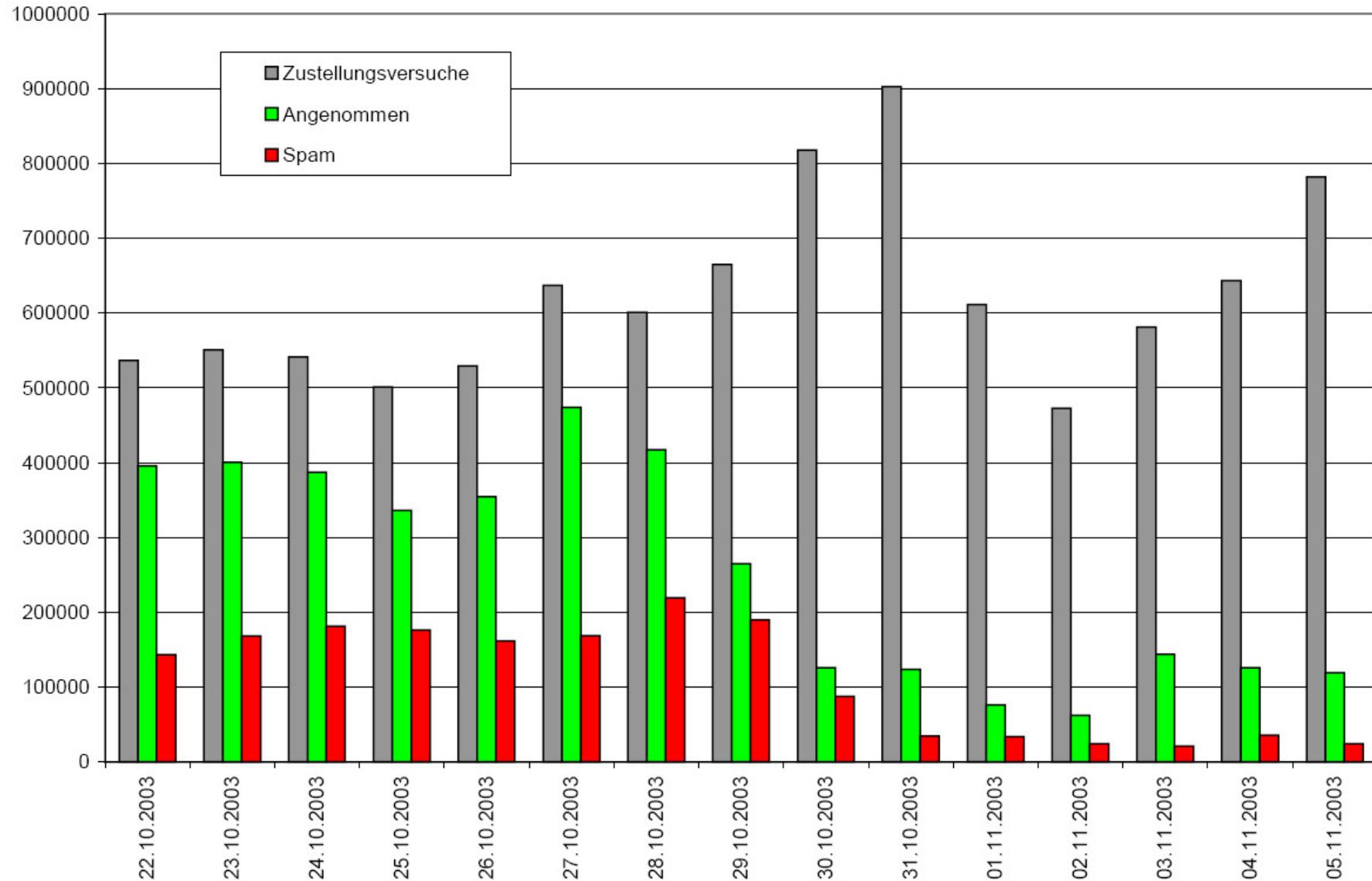


# Greylisting – ein paar Zahlen

Tägliches Spamaufkommen auf den Eingangsrelays



# Greylisting – ein paar Zahlen



# MARID-Verfahren

- MARID = MTA Authorization Records in DNS
- Arbeitsgruppe der IETF
- Im Frühjahr 2004 gegründet
- Im Herbst 2004 gescheitert (aufgelöst)
- Ziel war ein Standardverfahren zur Überprüfung von Mailabsendern im DNS zu verabschieden

# Reverse MX

- Mail-Ins sind per MX RRs im DNS hinterlegt.

```
Nicht autorisierte Antwort:  
web.de MX preference = 110, mail exchanger = mx-ha02.web.de  
web.de MX preference = 100, mail exchanger = mx-ha01.web.de
```

- Warum nicht auch Mail-Outs dort angeben?
- Reverse MX (Hadmut Danisch):
  - Auflösung der Domain der Envelope-From-Domain
  - dort wird in RMX-Records hinterlegt, welche IP-Adressen Mails senden dürfen
  - Probleme: RMX als eigene Resource Records schwierig (DNS Libraries und MTA-Software müssen angepasst werden)

# MTAMARK

- Sendende IP-Adresse wird rückwärts aufgelöst
- Subdomäne: `_send._smtp._srv` mit Querytype TXT
- 1=MTA, 0=kein MTA
- Im Beispiel: `mail.space.net (195.30.0.8)`, Markus Stumpf

```
chris@leo ~  
$ nslookup -q=txt _send._smtp._srv.8.0.30.195.in-addr.arpa  
Server: icarus.home  
Address: 192.168.28.2  
  
Nicht autorisierte Antwort:  
_send._smtp._srv.8.0.30.195.in-addr.arpa      text =  
"1"
```

```
$ nslookup -q=txt _send._smtp._srv.5.127.94.194.in-addr.arpa  
Server: icarus.home  
Address: 192.168.28.2  
  
*** _send._smtp._srv.5.127.94.194.in-addr.arpa wurde von icarus.home nicht gefunden: Non-existent do  
main  
chris@leo ~  
$
```

# Sender Policy Framework und Sender ID

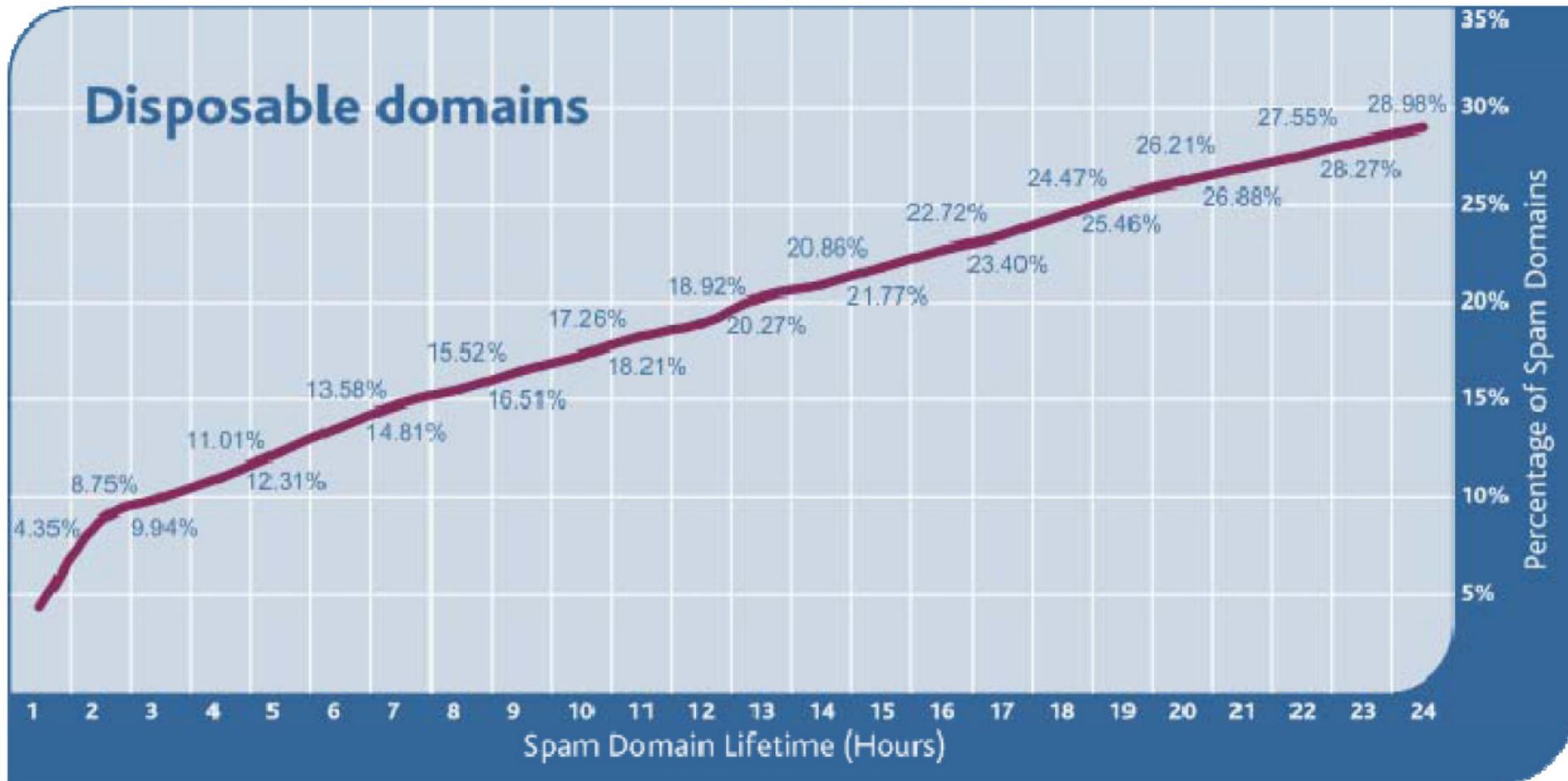
- DNS-basierte Überprüfung, ob die einliefernde IP-Adresse die Erlaubnis hat, E-Mails mit gegebenen Absender-Domains zu versenden
- Problem: Heutzutage fast normal von mehreren Relays Mails zu versenden (u.a. Forward, Mailing-Listen)
- Unterscheidung hinsichtlich...
  - der Absender-Adresse, die zur Authentifizierung herangezogen wird
  - der DNS-Records, die verwendet werden (bzw. der Grammatik)
- Beispiel (Domain gmx.net):
  - SPF Version 1
  - Server mit IP-Addr. Aus 213.165.64.0/23 dürfen Mails im Namen von gmx.net versenden
  - Sonst keiner (-all)

```
chris@leo  
$ nslookup -q=txt gmx.net 217.237.151.225  
Server: www-proxy.D01.srv.t-online.de  
Address: 217.237.151.225  
  
Nicht autorisierte Antwort:  
gmx.net text =  
  
"v=spf1 ip4:213.165.64.0/23 -all"  
  
chris@leo ~  
$ nslookup mail.gmx.net  
Server: icarus.home  
Address: 192.168.28.2  
  
Nicht autorisierte Antwort:  
Name: mail.gmx.net  
Addresses: 213.165.64.21, 213.165.64.20
```

# Sender Policy Framework und Sender ID

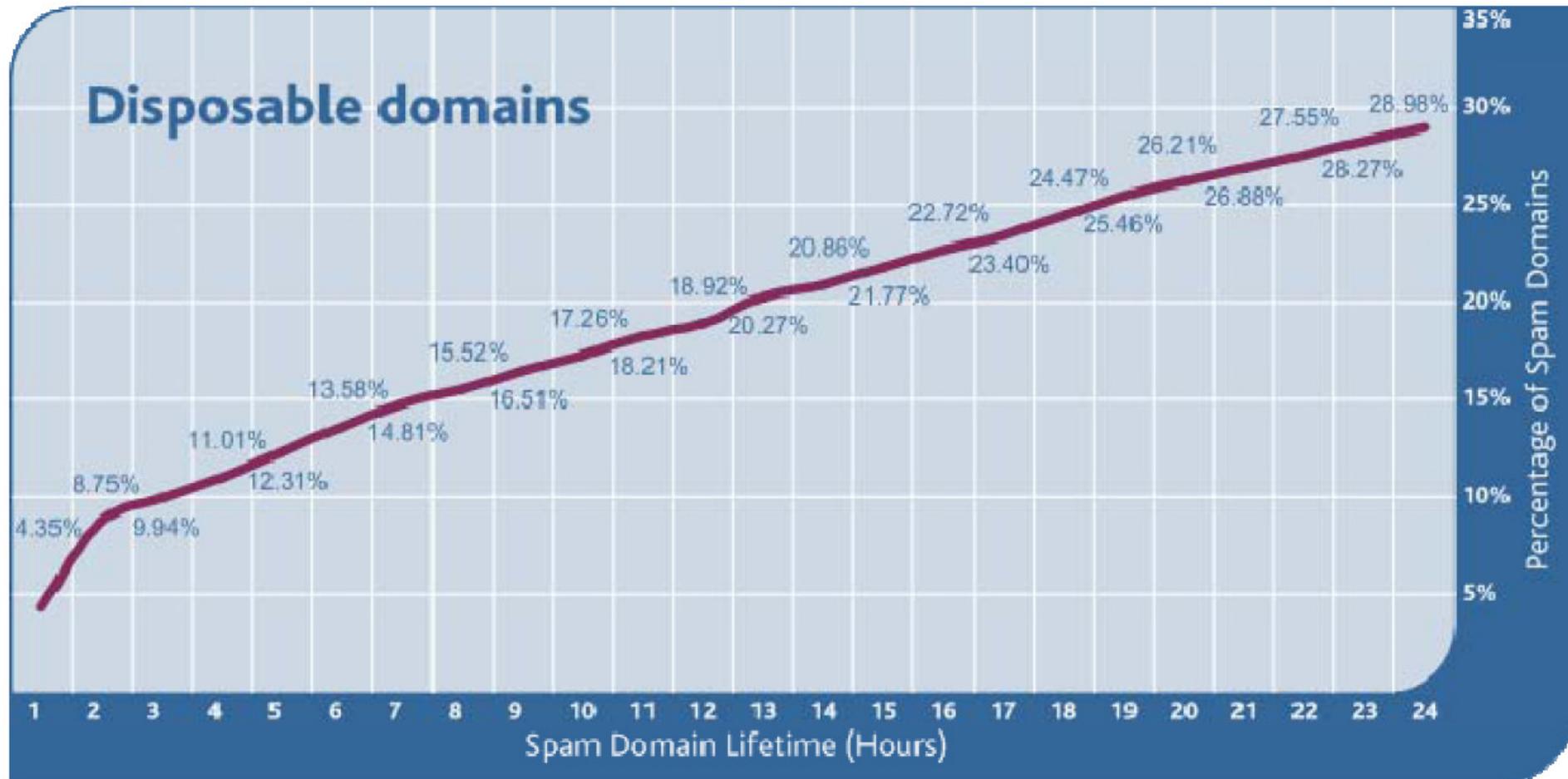
- Nachteile:
  - Forwarding funktioniert nicht mehr!
  - Lösungsansatz Sender Rewriting Scheme (SRS)
  - Verschiedene Absender-Adressen problematisch
  - Erfordert DNS-Maintenance
- Weitere Infos: <http://www.openspf.org/>
- Viele weitere
  - teilweise mit Krypto-Hintergrund: Yahoo! DomainKeys
  - siehe <http://www.ietf.org/html.charters/OLD/marid-charter.html>

# MARID – Probleme



- Spam wird häufig über Wegwerf-Domains versandt.
- Im DNS können beliebige Einträge vorgenommen werden – auch MARID

# MARID – Probleme



- ~29% aller Spam-Domains werden < 24h benutzt
- ~10% aller Spam-Domains werden < 3h benutzt!
- Maximale Dauer der Verwendung liegt bei nur 12 Tagen!

# Inhaltsbasierte Verfahren

- Header-Checks
- Prüfsummen-Verfahren (DCC, Pyzor, Razor)
- Bayesscher Filter

- Merkmale von Spam-Nachrichten werden zur Filterung herangezogen
- Beispiele:
  - HTML-Inhalt
  - Gefälschter UserAgent String (MS Outlook Ver 6.10.51214.1241)
  - Nachricht besteht (fast) nur aus Bildern
  - Buchstaben-Spielchen (GROSS, u\_n\_d\_e\_r\_s\_c\_o\_r\_e, gappy)
  - zufällige Strings (sidufwbw skfjhawer), um Signaturverfahren zu täuschen

# Header-Checks

- Received-Zeile
- Datum und Zeit (auch in Kombination mit Received)
- UserAgent-Strings
- ‚Subject‘ enthält cialis, viagra, phentermine, soma, valium, ...
- ‚From‘ enthält Ziffern und Buchstaben gemischt
- Sehr gute Übersicht:  
[http://spamassassin.apache.org/tests\\_3\\_1\\_x.html](http://spamassassin.apache.org/tests_3_1_x.html)

- Verfahren:
  - Über die Nachricht wird ein Hash gebildet
  - Hash wird von vielen Anwendern an einen zentralen Server übermittelt
  - Anhand der Häufigkeit mit der ein bestimmter Hash auftritt kann auf Spam geschlossen werden
- Probleme: Mailing-Listen, Newsletter
- Daher Fuzzy-Hashing (Auslassen bestimmter Informationen, z.B. Absender-Adresse)
- Beispiele:
  - Distributed Checksum Clearinghouse (<http://www.rhyolite.com/anti-spam/dcc/>)
  - Vipul's Razor (<http://razor.sourceforge.net/>)
  - Pyzor (<http://pyzor.sourceforge.net/>)

# Bayesscher Filter

- Benannt nach Thomas Bayes (ca. 1702 – 1761), engl. Mathematiker
- Rechnet mit bedingten Wahrscheinlichkeiten, Beispiel: Von charakteristischen Wörtern in einer E-Mail (Ereignis) wird auf die Eigenschaft Spam (Ursache) geschlossen.
- In Bezug auf Spam zuerst auf einem AAAI-Workshop vorgeschlagen und durch ‚A Plan for Spam‘ von Paul Graham bekannt gemacht
- Arbeitet mit Häufigkeiten von klassifizierten Wörtern
- häufig lernend implementiert

Subject*FREE	0.9999
free!!	0.9999
To*free	0.9998
Subject*free	0.9782
free!	0.9199
Free	0.9198
Url*free	0.9091
FREE	0.8747
From*free	0.7636
free	0.6546

Patrick Pantel and Dekang Lin. ``SpamCop - A Spam Classification & Organization Program.'' Proceedings of AAAI-98 Workshop on Learning for Text Categorization.

Mehran Sahami, Susan Dumais, David Heckerman and Eric Horvitz. ``A Bayesian Approach to Filtering Junk E-Mail.'' Proceedings of AAAI-98 Workshop on Learning for Text Categorization.

# Lösungen des Zombie-Problems

- **Lösung: Blockieren von E-Mails, die von Hosts aus fremden Dialup-IP-Adressbereichen eingeliefert werden**

Problem: Veröffentlichung der Dialup-IP-Adressen = Aufwand für den Provider und evtl. mangelnde Flexibilität

- Einige ISPs blockieren Outbound-Traffic mit Zielport 25 bzw. leiten auf eigenen Proxy mit SMTP-Authentifizierungszwang um (Comcast)
- E-Mail Service Provider können Zombies durch Analysen des Kommunikationsverhaltens erkennen. Erkennungsmerkmale sind
  - Massenversand (kein Alleinstellungsmerkmal)
  - hohe sog. Bounce-Rate (Verhältnis von zustellbaren zu nicht-zustellbaren E-Mails)

- Mails über 250K werden nicht gescannt
- SpamAssassin ist weit verbreitet
  - Funktionalität durch Plugins oder eigene Textregeln
- SMTP after POP / geringe Verbreitung von SMTP AUTH

SMTP Erweiterung / Art des E-Mail-Servers	Verbreitungsgrad
SIZE / generell	81,00%
<b>SMTP AUTH / Smarthost</b>	<b>72,00%</b>
PIPELINING / generell	71,00%
8BITMIME / generell	64,50%
ENHANCED STATUS CODES / generell	27,30%
<b>STARTTLS / Smarthost</b>	<b>15,10%</b>
STARTTLS / Mail-In	4,00%

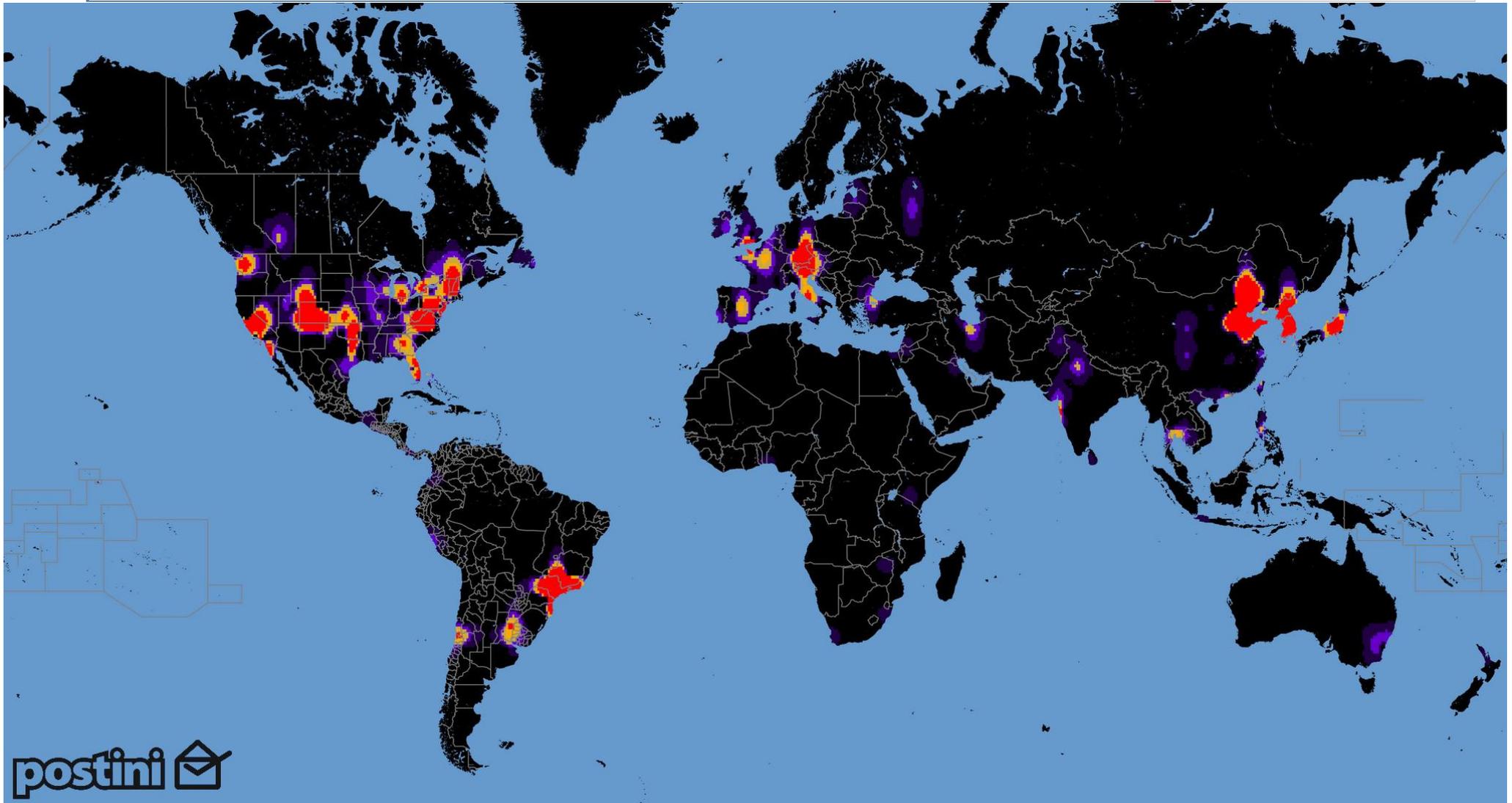
- Teergruben

```
C: RCPT TO: <user1@domain.com>
S: 250 ok
C: RCPT TO: <user2@domain.com>
S: [3 Sekunden Wartezeit] 250 ok
C: RCPT TO: <user3@domain.com>
S: [6 Sekunden Wartezeit] 250 ok
C: RCPT TO: <user4@domain.com>
S: [9 Sekunden Wartezeit] 250 ok
C: RCPT TO: <user5@domain.com>
S: [12 Sekunden Wartezeit] 250 ok
... usw.
```

# SpamAssassin Report

```
X-Spam-Report: 22.50 hits, 9 required;
* 0.3 -- From: does not include a real name
* 2.6 -- Bulk email software fingerprint (screwup 1) found in headers
* 1.9 -- Subject: contains advertising tag
* 2.4 -- BODY: List removal information
* 2.2 -- BODY: Click-to-remove with mailto: found beforehand
* 1.9 -- BODY: Claims you can be removed from the list
* 0.3 -- BODY: Asks you to click below
* 1.4 -- BODY: JavaScript code
* 0.4 -- BODY: FONT Size +2 and up or 3 and up
* 2.6 -- BODY: Spam phrases score is 34 to 55 (high) [score: 36]
* 0.9 -- BODY: Message is 70-90% HTML tags
* 0.8 -- BODY: HTML font color is blue
* 0.6 -- BODY: Includes a URL link to send an email
* 4.1 -- BODY: Frontpage used to create the message
* 1.6 -- BODY: Tells you to click on a URL
* 1.7 -- URI: Includes a link to send a mail with a subject
* 0.8 -- URI: Includes a URL link to send an email with the subject 'remove'
* 0.7 -- URI: Includes a 'remove' email address
* 0.7 -- HTML-only mail, with no text version
```

# Herkunft von Spam



- Lokalisierung von IP-Adressen nicht verlässlich!

# Herkunft von Spam (ISPs)

The 10 Worst Spam Service ISPs		As at 16 January 2006
Rank	Network	Number of Current Known Spam Issues
1	mci.com	<a href="#">243</a>
2	sbc.com	<a href="#">90</a>
3	comcast.net	<a href="#">82</a>
4	hinet.net	<a href="#">42</a>
5	seed.net.tw	<a href="#">40</a>
6	yahoo.com	<a href="#">40</a>
7	ocn.ne.jp	<a href="#">37</a>
8	twtelecom.net	<a href="#">36</a>
9	fdcservers.net	<a href="#">36</a>
10	telekom.de	<a href="#">35</a>

- Paul Graham – A Plan for Spam, August 2002, <http://www.paulgraham.com/spam.html>
- Christian J Dietrich, Norbert Pohlmann – E-Mail-Verlässlichkeit: Verbreitung und Evaluation, März 2005, Konferenzband DACH Security 2005
- Christian J Dietrich, Norbert Pohlmann – Spam: Situation und Hintergründe, April 2004, Konferenzband BSI Kongress 2005
- Christian J Dietrich, Norbert Pohlmann – IP Blacklisting zur effektiven Spam-Abwehr, September 2005, Datenschutz und Datensicherheit (DuD) 29, Ausgabe 09/2005, S. 548 ff.
- Christian J Dietrich, Norbert Pohlmann – Spam auf dem Rückmarsch?, Oktober 2005, IT-Sicherheit, Ausgabe 04/2005
- Bundesamt für Sicherheit in der Informationstechnik (BSI), Antispam-Strategien, 2005, Studie, <http://www.bsi.de/literat/studien/antispam/>
- RFC 2821 (SMTP) et al.
- <http://www.heise.de/newsticker/meldung/51379>
- <http://www.space.net/~maex/Drafts/dns-mtamark/draft-stumpf-dns-mtamark-03.html>
- <http://www.danisch.de/work/security/antispam.html>
- <https://www.internet-sicherheit.de>
- Praxis: Peer Heinlein – Das Postfix-Buch, 2004, OpenSource Press
- Praxis: <http://spamassassin.apache.org/doc.html>
- Interessante Webseiten:  
[www.antispam.de](http://www.antispam.de), [www.spamhaus.org](http://www.spamhaus.org), [www.message-labs.com](http://www.message-labs.com), [www.openrbl.org](http://www.openrbl.org), [www.ordb.org](http://www.ordb.org), [www.surbl.org](http://www.surbl.org),  
[spamassassin.apache.org](http://spamassassin.apache.org), [www.greylisting.org](http://www.greylisting.org), <http://www.rrze.uni-erlangen.de/dienste/e-mail/spam-analyse/>,  
<http://www.rz.rwth-aachen.de/infodienste/email/>, [www.senderbase.org](http://www.senderbase.org), [www.trustedsource.org](http://www.trustedsource.org)
- <http://honeynet.org/papers/bots/>

## **Netzwerksicherheit A Spam und Anti-Spam**

**Vielen Dank für Ihre Aufmerksamkeit**

**Fragen ?**

**Christian J Dietrich  
dietrich@internet-sicherheit.de**

**Institut für Internet-Sicherheit  
<https://www.internet-sicherheit.de>  
Fachhochschule Gelsenkirchen**

