

Spam und Viren - gemeinsam sind wir stark (im kommen)

Dr. Erwin Hoffmann

feh@fehcom.de

<http://www.fehcom.de>

[<http://www.fehcom.de/qmail/qmailbook.html>]

SAGE

Darmstadt, 2004-7-1

Abstrakt

- Ziel des Vortrags ist es, die aktuelle Spam- und Virensituation qualifiziert einzuschätzen,
- System-Administratoren die Möglichkeiten, Grenzen und Gefahren der Abwehr von Spam- und Viren-E-Mails aufzuzeigen,
- ein Konzept vorzustellen, dass auch unter den Bedingungen einer Virus-Attacke die E-Mail-Kommunikation der eigenen Mail-Gateways gewährleistet wird.
- Der Vortrag löst nicht das Problem: *Wie krieg' ich meine eigene Mailbox frei von Spam ?*

Viren, Würmer, Spam ...

- ... ist eine *Belästigung* für den End-Anwender,
- ... ist eine *Aufgabe* für den Administrator eines mittelständischen Unternehmens,
- ... ist eine *Herausforderung* für den Betreiber eines Mail-Gateways (ISP),
- ... stellt mittlerweile eine *Beeinträchtigung* der Internet-Kommunikation via E-Mail dar,
- ... hat das *Vertrauen* des Benutzers in das Kommunikationsmedium "E-Mail" *stark erschüttert*.

Viren, Würmer, Spam ...

- dürfen nicht mehr getrennt betrachtet werden, da mittlerweile über die mit Trojaner (Agobot, Phatbot et al.) infizierten Rechner (PCs) eine Infrastruktur von SMTP-Gateways zur Verfügung steht, die bei Bedarf genutzt werden können,
- führen in Kombination zu einer Distributed Denial of Service (DDoS) Attacke gegen die E-Mail-Systeme,
- müssen mit einer einheitlichen Strategie gemeinsam bekämpft werden.
 - Die Mitte letzten Jahres gross angekündigte Anti-Spam-Allianz der ECO hat bislang keine Früchte getragen; die Interessenlage der ISPs ist nicht einheitlich.
 - Das deutsche "Anti-Spam Gesetz" bezieht sich nur auf den unlauteren Wettbewerb zwischen Firmen.

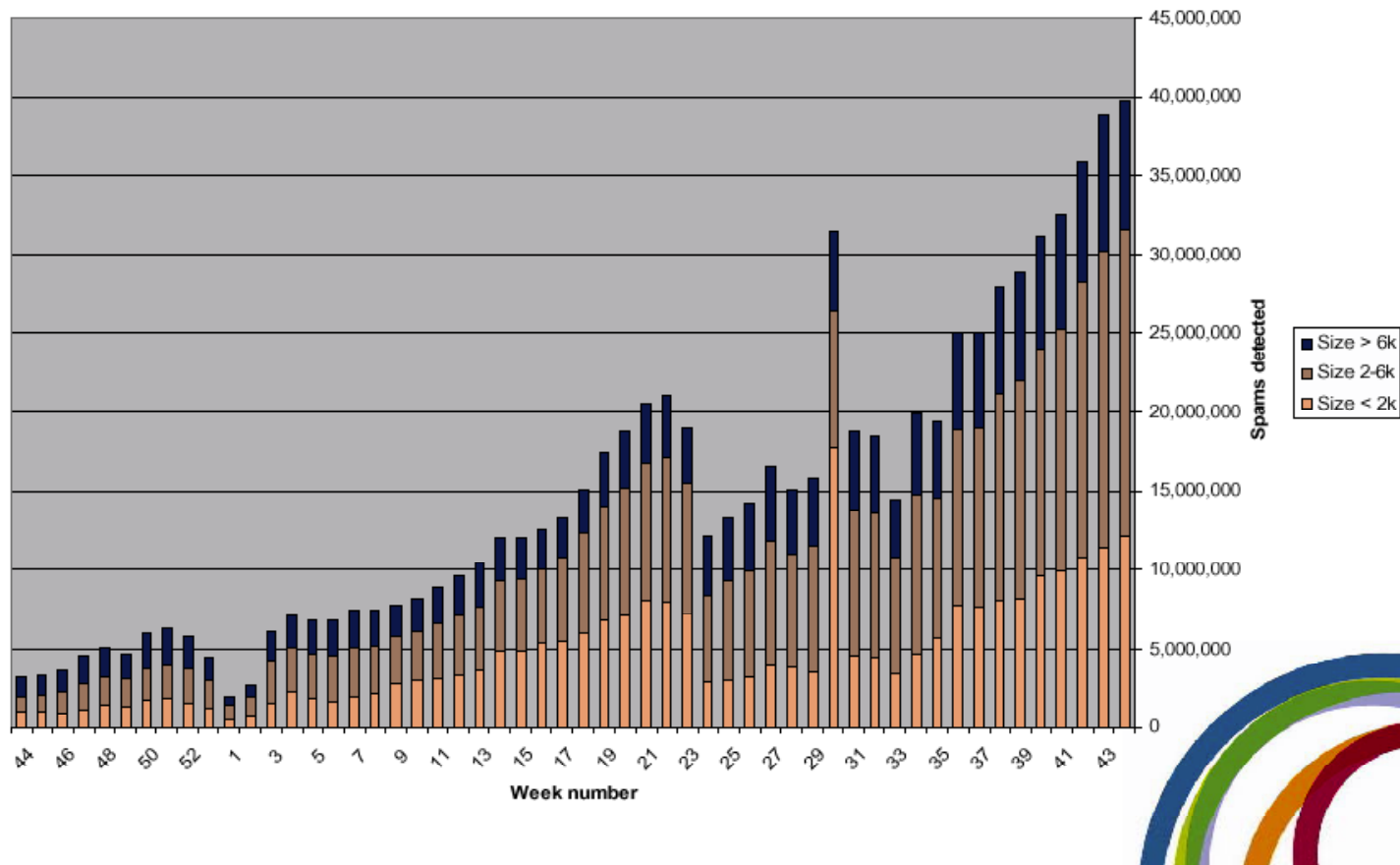
Rechtliche Lage und Pflichten des E-Mail Providers

Beim Einsatz von Anti-Viren/Spam-Lösungen müssen folgende Gesetze/Vorschriften berücksichtigt werden:

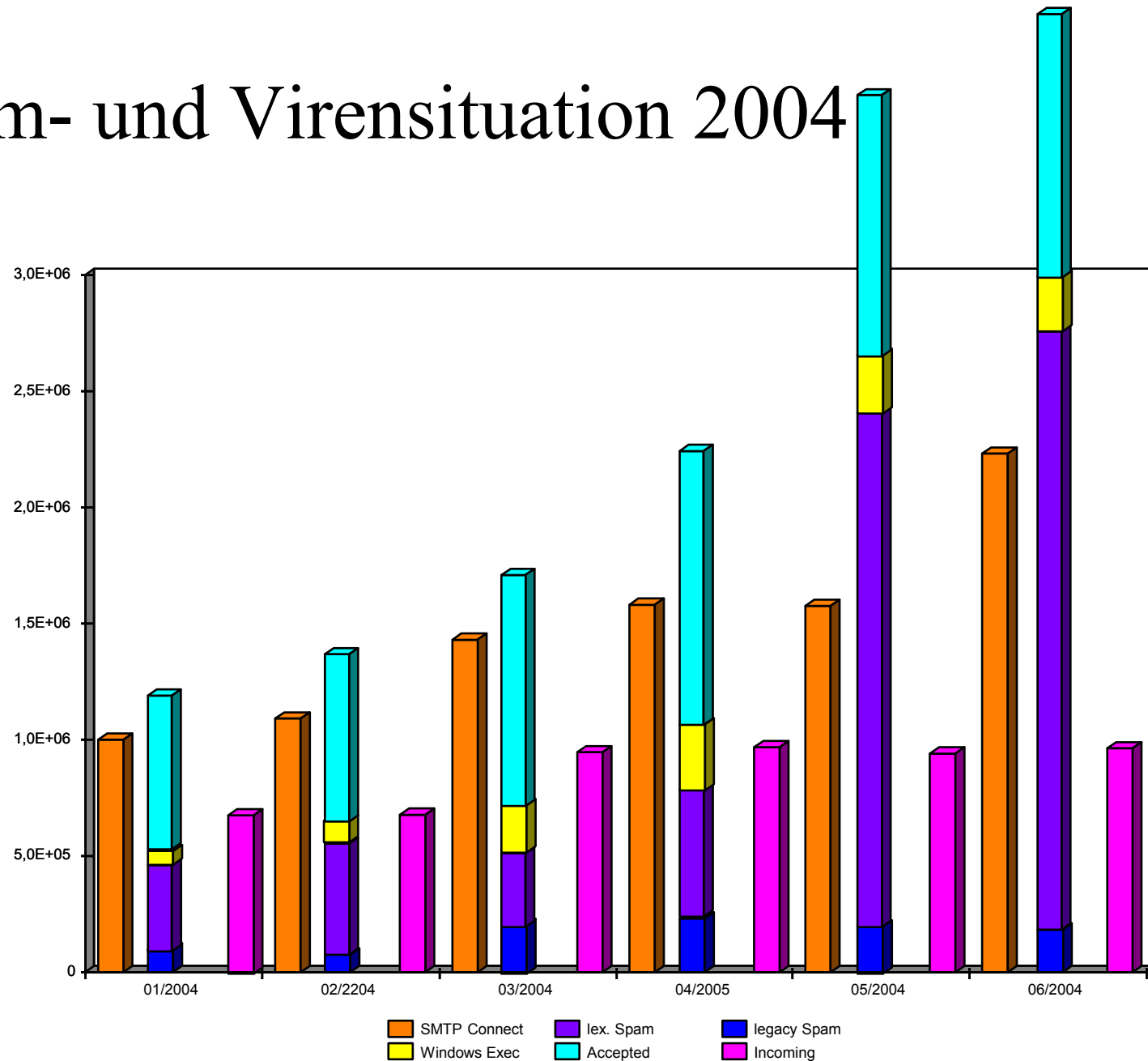
- Kanal B2B: Bundesdatenschutzgesetz (BDSG) - IT-Riskomanagment; Schaffung einer sicheren Netzinfrastruktur (FireWalls, AV-Software)
- Kanal B2C: Telekommunikationsgesetz (TK) - Empfänger hat ein Recht auf seine E-Mails (auch Spam)
- Kanal B2M: Betriebliche Vereinbarungen - Nutzung der Mail-Infrastruktur/Netiquette
- Kanal C2C/B: StGB (§ 303a/b) "Computersabotage"; BGB Zivilrechtlicher Schaden (Sasser)

Spamzuwachs ... 2003

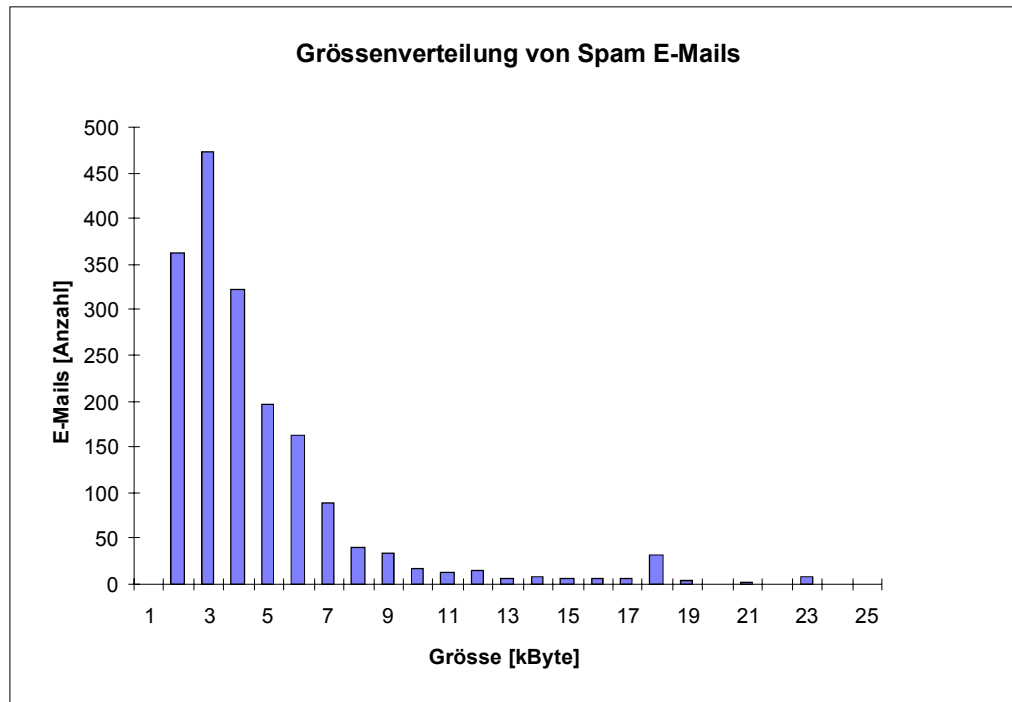
Wöchentlicher Spam Zuwachs



Spam- und Virensituation 2004



Größenverteilung von Spam E-Mails und Anteil am E-Mail-Volumen

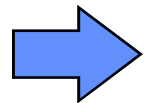


- Die mittlere Grösse einer Spam-E-Mail ist 5 kByte.
- Der Anteil an Spams in der SMTP-Kommunikation kann auf $> 80\%$ abgeschätzt werden, und zwar sowohl hinsichtlich der Anzahl der Mails als auch des Volumens.
- Hinzu kommt noch, dass zwischen 5% und 20% der E-Mails infiziert sind.

Wo und Wann soll die Abwehr erfolgen ?

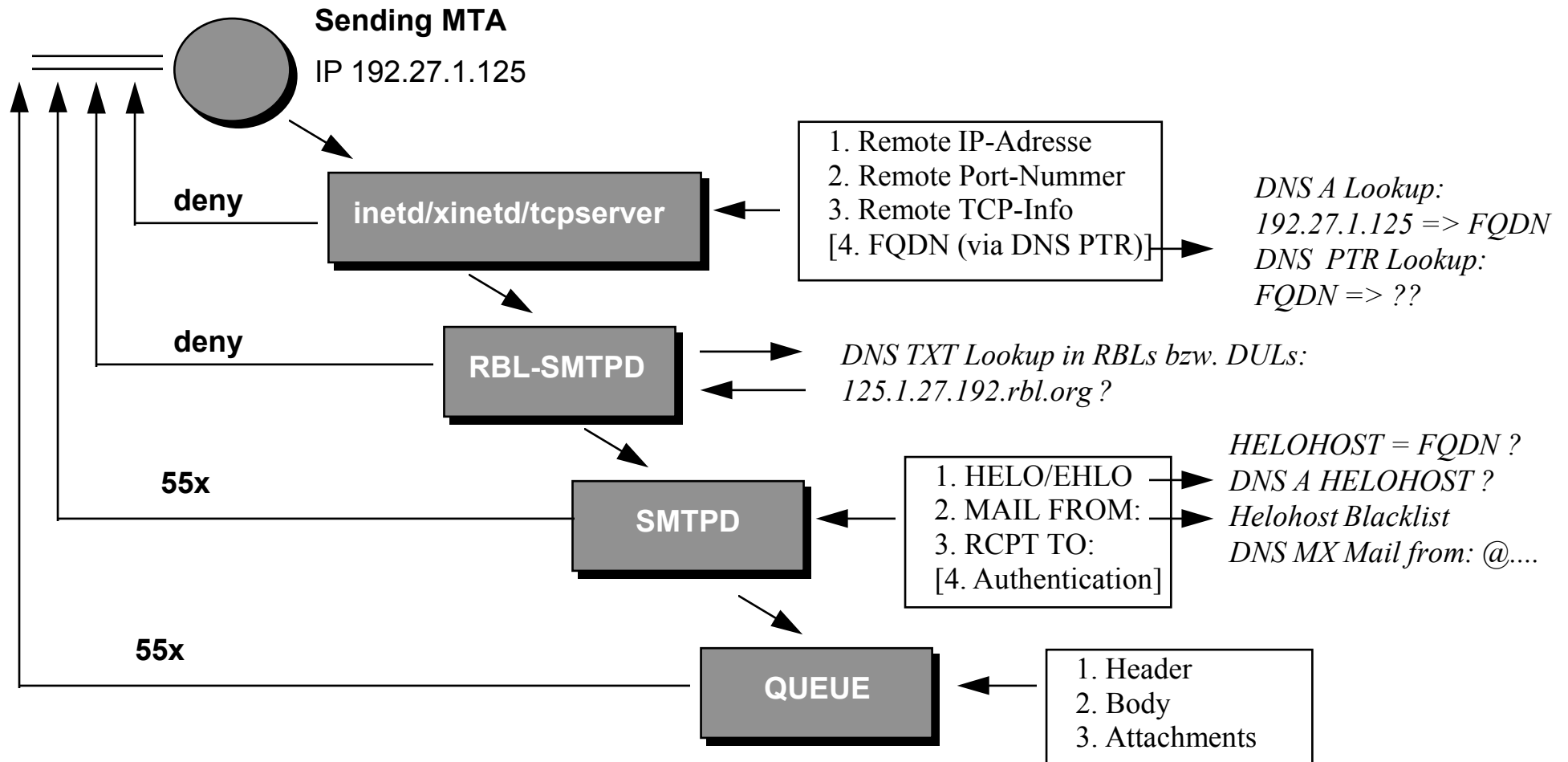
Die Abwehr kann zu drei Zeitpunkten im E-Mail-Zyklus erfolgen:

- Beim Empfang der E-Mail per SMTP - SMTP 55x Reply
- Bei der Zustellung ins lokale Postfach (a la procmail) - NDR (Bounce)
- Bei Herunterladen der E-Mail via POP3/IMAP4



Es sollte darauf verzichtet werden, bei der Spam- und Viren-Mails mit einer NDR (None Delivery Report) zu reagieren, da i.d.R. der Absender gefälscht ist und dies nur den falschen trifft, bzw. zu einer weiteren Bounce führt.

On-the-fly Blockieren unerwünschter E-Mails



Klassischer Spam

"Klassischer Spam" benutzt zwei Verbreitungsarten:

- Über offene E-Mail-Gateways bzw. Proxies
 - Hierzu zählen vor allem Windows-PC am DSL mit einem Mailproxy aber auch z.B. der Apache HTTPD, falls er als Proxy konfiguriert ist.
 - Schritt 1: Portscan auf Port 25; Schritt 2: Versenden einer Relay-Mail über diese Adresse; Schritt 3: Auswertung der Ergebnisse
- Über "Wegwerf-Domänen", die nur zum Versand der Spam-E-Mails genutzt werden.
 - Mittels einer Realtime-Blacklist (RBL) können einige dieser Sender ausgeschlossen werden, das Abweisen von E-Mails, die via Dial-Up-Verbindung eintreffen, ist sehr wirkungsvoll, unterdrückt aber auch einen Teil legitimer Mails.

Methoden der Spammer

Der Spammer hat sämtliche Elemente des SMTP-Envelope und fast alle des Header und natürlich auch des Body unter Kontrolle:

- Der SMTP-Envelope wird qualifiziert genutzt.
- Der E-Mail Header ist korrekt aufgebaut.
- Der E-Mail Body beinhaltet zufällige Wortfolgen, die die Bayesian-Filter verwirren sollen und darüber hinaus jede Spam-Mail mit einer unigen Checksumme versieht.
- Die eigentliche Spam-Information kann auch in einer embedded Graphik-Datei hinterlegt werden.

 Es gibt per se keine Möglichkeit, Spam von gewollter Mail zu unterscheiden!

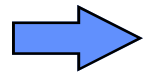
Methoden der Spam-Abwehr

Zeitpunkt	Schicht/(Schritt)	Informationen	Massnahmen
Aufbau der TCP/IP-Verbindung	IP (Layer 3)	Sender-IP, FQDN des Senders, Blocking List	Blockieren [per MTA]
SMTP-Dialog	SMTP (Layer 7)	Return-Path (MAIL From:), Forwarding-Path (RCPT To:) MX der Return-Path Domain, HELO/EHLO Angabe	Blockieren (SMTP Fehlercode 5xx), Deferral (Verzögern, SMTP Fehlercode 4xx) [per MTA]
Einfügen in die Queue	(Verarbeitung)	E-Mail-Header, E-Mail-Body (Inhalt/Text), MIME-Struktur, Footprint (remote)	Bounce, Verwerfen, Taggen [per MTA]
Ausliefern in Empfänger-Mailbox	(lokale Zustellung)	E-Mail-Header, E-Mail-Body (Inhalt/Text), MIME-Struktur, Footprint (remote); evtl. Spam-Tags	Bounce, Verwerfen, Ablage "Junkmail" Ordner, Taggen [per User]
Abholung durch Remote Mail User Agent	(entfernte Zustellung)	E-Mail-Header, E-Mail-Body (Inhalt/Text), MIME-Struktur, Footprint (remote); evtl. Spam-Tags	Abholen, Nicht-Abholen, Löschen nach <i>N</i> Tagen, lokaler "Junkmail" Ordner [per User]

Syntaktische Checks

Syntaktische Checks werden gegenüber dem SMTP-Envelope bzw. den Header-Zeilen vorgenommen:

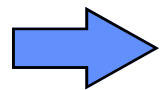
- HELO/EHLO: Qualifizierter Name; nicht die eigene IP-Adresse/Loopback
- MAIL From: Qualifizierter Name (@); keine % Zeichen; keine E-Mail Adresse aus eigenem Vorrat
- Message-ID: Vorhanden und mit korrekter Syntax
- Date: vorhanden und mit vernünftigem Wert



Die Wirksamkeit (Effektivität) dieser Methode liegt zwischen 5% und 10%.

Neben den Realtime-Blacklisten kann

- das Matching von IP und FQDN, für die IP-Verbindung,
- die Angabe des HELO/EHLO Names [DNS A] sowie
- die Existenz der Absender-Domain (Mail From:) [DNS MX] vorgenommen werden.



Während letzteres ein gängiges Verfahren ist (~ 10%), sind die ersten beiden Tests bei Dial-Up E-Mail-Nutzern problematisch und führen zu fehlerhafter Ablehnung.

Bayesian-Filter (SpamAssassin)

Bayesian-Filter (nach dem englischen Mathematiker Bayes) funktionieren nach dem Prinzip einer Bewertungsgrösse:

- $B = \frac{\sum W(x_i)}{\sum W(y_i)}$
 - x_i = "richtige Merkmale"; $W(x)$ = Wahrscheinlichkeit für x
 - y_i = "falsche Merkmale"; $W(y)$ = Wahrscheinlichkeit für y
- Die definierten Merkmale (Entscheidungsbäume!), z.B. Header-Zeilen, Formatierung, Worte, in der E-Mail werden mit einem Satz bekannter Spam-Merkmale verglichen und bewertet.
- Hieraus wird ein Spam-Score ermittelt; die Entscheidungsgrundlage muss immer den aktuellen Gegebenheiten angepasst werden (> 10k E-Mails).
- Prinzipiell kann eine E-Mail über einem bestimmten Spam-Score abgelehnt werden, typischerweise erhält sie aber eine neue Header-Zeile und man überlässt das Verwerfen dem Benutzer/MUA.

Realtime-Blocklisten / DUL

Realtime Blacklisten enthalten Informationen über IP-Adressen, von denen aus "gespamt" wird.

- Die Qualitäten dieser Listen ist unterschiedlich; zudem ist die Existenz dieser Listen "dynamisch". Ein guter Dienst ist z.B. MAPS (<http://www.mail-abuse.com/>). Unterschieden wird hier zwischen:
 - Bekannte IPs von Spammern (RBL)
 - Dial-Up (Dynamic) User Listen (DUL)
 - Bekante Open Relay (RSS)
 - Offene Proxy Server (OPS)
 - Non-Confirming Mail Lists (NML).
- Die Datenbank der Listen kann "gespooft" werden, d.h. es werden bewusst falsche Einträge (*false positive*) lanziert (z.B. gmx.de, hotmail.de).
- Einige andere aktive Listen:
 - sbl.spamhaus.org #Confirmed spammers
 - relays.ordb.org #Open Relay list
 - opm.blitzed.org #Open Proxy list

Kombinierte Spam-Filter

Zwei unterschiedliche Strategien:

1. Ablehnung der Spam-E-Mail beim Auftreten syntaktischer Merkmale.
2. Verwerfen der Spam-E-Mails bei entsprechendem Spam-Score.
 - Syntaktische Merkmale können auch in die Bayes-Bewertung mit eingeschlossen werden; unter Nutzung von IMAP bzw. WebMail, kann die verdächtige E-Mail dann sofort in ein "Junkbox"-Verzeichnis überführt werden.
 - Durch Verschieben von E-Mails zwischen "Inbox" und "Junkbox" kann das Bayes-Filter für den individuellen Benutzer "getunt" werden (GMX).

Address-Harvesting

Zum Versenden einer E-Mail braucht der Spammer eine E-Mail-Adresse. Neben der offensichtlichen Domain-Kennung muss auch der lokale Teil erzeugt werden bzw. bekannt sein:

- Brute-Force lexikalische Attacken.
- Address-Harvesting über E-Mail-Foren, Newsgroups ...
- Adressen-Handel (0.01 Cent/Adresse + Software).
- Bei Trojanern kann das lokale (Outlook) Adressbuch ausspioniert und E-Mail-Adressen aus dem Cache gelesen werden.
- Mail-Spider: Web-Seiten werden nach Adressen durchsucht.

E-Mail Netiquette

Einige Grundsätze:

- Auf eine Spam- und/oder Virenmail sollte nicht geantwortet werden, ausser mit einem SMTP-Reply Code.
- Beim Versenden von Mails an mehrere Benutzer ist die BCC-Funktion des E-Mail-Programms zu nutzen.
- E-Mail-Adressen (ausser funktionalen wie Webmaster) sollten nicht auf Web-Seiten veröffentlicht werden (überprüfen!).
- Wegwerf-Adressen haben sich nicht bewährt, da sie nur den Umfang des Spams vergrössern.
- Keine lokalen Namen mit wenigen Buchstaben verwenden (feh@ ...).

Spamtrapping

Bei lexikalischen Spam-Attacken (aa@domain , ab@domain ...) ist es möglich, funktionale Spamtrap Accounts aufzusetzen.

- Wird eine E-Mail an diese Adresse registriert, kann automatisch ein Regelwerk in Kraft treten, die Spam-E-Mail aufgrund
 - ihrer Herkunft (IP-Adresse),
 - den Absenderkennung (Mail From:) oder
 - ihrer Checksumme zu klassifizieren
- und die Annahme bzw. Zustellung weiterer E-Mails zu unterbinden.

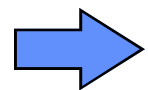
Razor, DCC und Checksummen

Im grossen Stil wird Spamtrapping bei den verteilten Diensten

- Razor und
- DCC Clearinghouse

genutzt.

- Hier werden die signifikanten Teile der E-Mail mit einer Checksumme belegt, aufgrund derer die E-Mail als Spam klassifiziert wird.
- Neben einem passiven Mode kann bei Razor auch die Datenbank aktiv ergänzt werden.



Nachteilig ist, dass hierdurch im gewissen Umfang die "Privacy" der empfangenen E-Mails verloren geht. Zwar kann aus der Checksummen nicht der eigentliche Inhalt der E-Mail zurückgerechnet werden, aber ...

RMX, SPF, DomainKeys

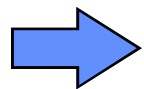
Seit einiger Zeit werden Erweiterungen des SMTP diskutiert, zu eine Absender-Validierung bzw. Authorisierung zu überprüfen: RMX, SPF (Sender Policy Framework), DomainKeys, MTAMark, MARID (MTA Authorization Records in DNS).

- Im IETF gibt es hinsichtlich der "richtigen" Methode (offene) Auseinandersetzungen, die vor allem politischer Natur sind (Microsoft XML basierte Caller-ID, AOL's SPF ...).
- Prinzipiell werden hierbei (invers zu RBLs) DNS TXT Einträge mit den pro Domain Sende(-erlaubten) MTAs hinterlegt.

... und was ist jetzt die Lösung ?

Die Vorschläge müssen mit folgenden Problem umgehen:

- Wie wird mit Bounces umgegangen (Mail From: <>) ?
- Wie wird die E-Mail-Weiterleitung unterstützt ?
- Wer ist Autoritiv für die Erstellung der DNS-Einträge ?
- Wie wird mit E-Mails umgegangen, deren Herkunft sich nicht validieren lässt ?



Spam ist kein technisches oder organisatorisches Problem, sondern eines der Authorisierung.

- M.E. lässt sich die Authorisierung nur durch eine Einführung einer qualifizierten DNSSec lösen.
- Zugleich muss das Problem der Authentität von E-Mails angegangen werden (hierzu habe ich auch einen Draft geschrieben).

Neue Virensituation

Mit den nun öffentlichen Virenbaukästen Phatbot und Agobot lassen sich hoch-effiziente Viren mit eigener SMTP-Engine herstellen. Die Viren übertragen sich per

- E-Mail,
- lokale Shares (KaZaA),
- HTTP.

Häufig werden die Viren in einem sog. UPX-Format verpackt. Dies wird von den üblichen Datei-Typen-Scannern als ZIP interpretiert (=> Tarnkappen-Viren) und die Anhänge in den E-Mails werden durchgelassen.

Trojaner und Backdoors

Besonders problematisch ist die Infektion von Windows-Rechnern, die per DSL quasi-permanent mit dem Internet verbunden sind.

- Die neuen Viren fungieren als Trojaner bzw. Backdoors, die die so infizierten Rechner selbst als Quelle weitere Virenattacken werden lassen.
- Dies entspricht einem gesteuerten DDoS-Angriff. Bekannt geworden sind solche auf die Mail-Infrastruktur von Microsoft bzw. SCO.
- Da sowohl mit gefälschten Absenderangaben, zufällig gewählten Empfängernamen als auch mit wechselnden IP-Adressen gearbeitet wird, entspricht die Abwehr dem Vorgehen gegen ein "moving Target".

Transport von Viren in E-Mails

Beim Transport von Viren in E-Mails kann von vier Möglichkeiten Gebrauch gemacht werden:

- SMTP E-Mail ist i.d.R. 8-bit Clean; das Virus kann direkt in den Body eingefügt werden - ungebräuchlich ausser für Skript-Viren (VBS, WSH).
- Das Virus kann per binhex-Kodierung transportiert werden; ungebräuchlich.
- Das Virus kann per Microsoft TNEF Encapsulation verbreitet werden; ungebräuchlich.
- Das Virus wird per Standard-MIME-Verfahren als BASE64 enkodiertes Attachment übertragen (aber nicht unbedingt "Content-Type: Application"); 99,99% aller Viren verhalten sich so.

Verpackung von Viren in E-Mails

Die meisten Viren werden "offen" übertragen:

- BASE64 enkodierter Anhang von Content-Type "Application". Jeder vernünftige E-Mail Client sollte die direkte Ausführung solcher Anhänge in der E-Mail prinzipiell unterbinden.
- BASE64 enkodierter Anhang mit falscher "Content-Type" Angabe (z.B. "Audio/x-wave").
- Transport-Stealth Viren nutzen das Windows UPX-Format, um sich als ZIP-Datei zu tarnen. Das UPX-Format verhält sich wie ein self-extracting ZIP, nur das als ZIP und nicht als Executable kodiert wird.
 - Diese Technik nutzte das MyDoom/NetSky-Virus, was sich dadurch schnell verbreiten konnte.

Klassische Anti-Virus-Lösungen

Die klassischen Anti-Virus-Lösungen bestehen darin, die E-Mail in ihre Bestandteile zu zerlegen und diese anschliessend durch einen oder mehrere AV-Scanner zu überprüfen. Beispiel hierfür ist AMaViS. Problematisch hierbei ist,

- dass die E-Mail Entpacker-Tools (metamail, reformime, ripmime) wesentlich pingeliger zu Werk gehen als die E-Mail Clients,
- die Trefferrate des/der AV-Scanner vom Zustand ihrer Pflege abhängen (Engine, Pattern),
- Viren in Archiven nicht unbedingt erkannt werden,
- der Ressourcen-Bedarf sehr hoch ist.

Client/Server-Virus-Scanner

Angefangen mit Sophos Sweep (Sophie/Trophie), haben mittlerweile viele AV-Hersteller auf ein Client/Server-Verfahren umgestellt:

- Hierbei arbeitet der Server permanent (~ 5MB Engine + ~ 5MB Pattern) [clamd] und stellt einen lokalen Socket bereit
- gegenüber dem Client (< 1 MB) bei Bedarf die Datei zur Virenprüfung übergibt [clamscan].

➡ Dies führt zu einer wesentlich effizienteren Nutzung des AV-Scanners (Sophos Sweep, F-Secure, Clam AV <=> McAfee, Inoculate IT). Die Effektivität wird aber weiterhin von der Engine und dem aktuellen Pattern bestimmt.

Blockieren von Anhängen

Bei einem Angriff mit neuartigen Viren/Würmern (bzw. auch Mutationen bestehender) versagen die üblichen AV-Scanner. Über die Reaktionszeiten der AV-Hersteller gibt es mittlerweile umfangreiche Untersuchungen; das prinzipielle Problem wird aber hierdurch nicht gelöst.

Daher werden häufig E-Mails bereits dann blockiert, wenn sie einen "verdächtigen" Anhang aufweisen:

- Der Dateiname enthält eine Endung wie ".exe".
- Die Datei wird über eine Magic-Check als ausführbare Datei erkannt.

MIME-Type-Erkennung

Die Erkennung des MIME-Types einer Datei kann entweder

- nach der BASE64 Dekodierung per "file" Kommando oder
- im BASE64 Datenstrom durch die ersten charakteristischen Byte erfolgen, was besonders Ressourcen-effizient ist:

his is a multi-part message in MIME format.

-----=_NextPart_000_0008_000058BC.00003703

Content-Type: text/plain;

charset="Windows-1252"

Content-Transfer-Encoding: 7bit

Please read the attached file.

-----=_NextPart_000_0008_000058BC.00003703

Content-Type: application/octet-stream;

name="document.pif"

Content-Transfer-Encoding: base64

Content-Disposition: attachment;

filename="document.pif"

TVqQAAMAAA EAAAA//8AALgAAAAAAAAAQAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAuAAAAKvnXsbvhjCV74Ywle+GMJVsmj6V44YwlQeZOpX2hjCV74YxlbiGMJVsjm2V

Loader-Type Identifizierung

Die Abweisung von Windows-Executables per MIME-Type Erkennung ist zwischen 75% und 85% effektiv.

- Will man nicht auch ZIP-Dateien blockieren (was technisch kein Problem ist), kann zusätzlich eine vorhandene "Loader" Anweisung im BASE64 kodierten Datenstrom herangezogen werden.
- Charakteristikum ist, dass in der Datei eine (BASE64 encodierte) Windows-spezifische Loader-Anweisung "KERNEL32.DLL" steckt.
- Hierdurch lassen sich 99,5% aller Viren direkt im SMTP-Datenstrom ohne zusätzliche Kosten erkennen und eliminieren; Windows-Executables können weiterhin ohne Schwierigkeiten im ZIP-Format (oder vergleichbar) übertragen werden.
- Den Code hierzu werden ich demnächst veröffentlichen (100 Zeilen) und auch für Sendmail, Postfix und Co. verfügbar machen (WARLORD).

Ressourcen-Bedarf der Spam- und Viren-Scanner

Der zusätzliche Einsatz eines AV-Scanners sowie eines Anti-Spam-Programms ist weiterhin unerlässlich; aber ausgesprochen Ressourcen-intensiv.

Unter den Bedingungen lexikalischer Spam-Attacken wird dies aber zum Problem:

- Die meisten MTAs akzeptieren E-Mails auf der Grundlage definierter Domain-Namen; dies macht die Systeme angreifbar für DDoS-Attacken.
- Erst durch das Filtern auf komplette Empfangs-Adressen kann die Flut einlaufender E-Mails unterbunden werden: Eine E-Mail die nicht angenommen wurde, braucht weder auf Viren noch auf Spam untersucht werden.

- Mittels geeigneter Mittel ist möglich, bereits während der SMTP Data-Phase die E-Mail auf Signaturen zu untersuchen und E-Mails mit gefährlichen Anhängen zu blockieren. Hierdurch werden 99,5% der infizierten E-Mails vermieden.
- Spam lässt sich im wesentlichen nicht durch technische Massnahmen sondern nur durch administrative reduzieren; dies wird aber Einschnitte in der Nutzung des Mediums "E-Mail" mit sich bringen. Welches administrative Verfahren sich hierbei durchsetzen wird, ist noch unklar.
- Der Inbound E-Mail-Verkehr vom Internet lässt sich nicht kontrollieren und ist seit den infizierten Rechnern mit Trojanern stark gestiegen. Sofern das E-Mail System die Zustellung von E-Mails Domain- und nicht Empfangs-Adressen-basiert zulässt, ist es angreifbar für umfangreiche DDoS-Attacken.