Heuristische Verfahren zur Spamabwehr

Albert Weichselbraun

In den letzten Jahren hat die Spamflut ein kaum mehr erträgliches Ausmaß erreicht. Dies hat jedoch auch dazu geführt, dass beträchtliche Ressourcen in den Kampf gegen unerwünschte E-Mails investiert wurden, sodass inzwischen bereits sehr wirksame Werkzeuge zur Abwehr von Spammern zur Verfügung stehen.

1 Open Relay Datenbanken

Einer der wohl naheliegendsten Ansätze zur Bekämpfung von Spam ist die Blockade von jenen Servern, über welche ein Großteil der Werbeflut an den "Endkunden" weitergeleitet wird.

Meist handelt es sich hierbei um Systeme, die von ihren Administratoren nicht hinreichend abgesichert wurden und deshalb als "Offene Relays" Spammern aus aller Welt für den Versand ihrer unpopulären Mails zur Verfügung stehen.

Mittels entsprechender Testprogramme - durch gezielte Suche oder aufgrund von Beschwerden - ist es jedoch möglich, solche Systeme zu identifzieren und in öffentliche Open-Relay Datenbanken ***[5]*** einzutragen. Diese Datenbanken werden von den Administratoren gerne für eine Vorselektion von "guten" und "bösen" Mails verwendet.

Die Schwachstellen dieses Ansatzes zeigen sich jedoch spätestens dann, wenn erwünschte Mails nicht ankommen, weil der Administrator des Absenders seinen Mailserver nicht entsprechend gesichert hat.

Dies mag für eine Privatperson nicht besonders dramatisch sein - wenn jedoch zum Beispiel Mails von Ministerien an Schulen (wohlgemerkt ein Beispiel aus der Praxis!) blockiert werden, so interessiert es die Verantwortlichen meist wenig, dass auf ihrer Seite ein "schlecht konfiguriertes" System vorliegt.

Weiters passiert es leider immer wieder, dass die entsprechenden Datenbanken vom Netz gehen. Gründe hierfür sind rechtliche Probleme oder auch das Bestreben, das ehemalig frei zugängliche Service kommerziell zu vermarkten ***[4]***.

2 Prüfsummen

Open Relay-Datenbanken sind jedoch nicht der Weisheit letzter Schluss - sodass wir uns einer wesentlich vielversprechenderen Gruppe, nämlich den Prüfsummen-Verfahren, zuwenden können.

Diese berechnen eine eindeutige Prüfsumme (zum Beispiel SHA1) über den Inhalt von Werbemails und legen diese in einer Datenbank ab.

Die Folge ist, dass Mails mit dieser Prüfsumme von allen Rechnern, welche auf die entsprechenden Datenbanken zugreifen, abgewiesen werden.

Leider sind einfache Prüfsummenverfahren sehr leicht zu umgehen. Bereits eine winzige ¨Änderung des Textes genügt, um die Prüfsumme zu ändern und somit die Mails wieder - allen Blockaden zum Trotz - zustellen zu lassen.

Ein Umstand, den auch Spammer sehr gerne durch entsprechende "persönliche" Zusätze Rechnung tragen (Vergleich ***Abbildung 1***).

Glücklicherweise haben die Entwickler (wie zum Beispiel ***[2]*** und ***[3]***) schnell auf dieses Problem reagiert und Prüfsummenalgorithmen in ihre Software integriert, die nicht mehr auf eine identische Nachricht angewiesen sind (wie zum Beispiel Nilsimsa und Ephemeral).

***Abbildung 1**: Ein "persönlicher" Link zum Ausschalten von einfachen Prüfsummenverfahren.*

```
If you do not wish to receive future mailings from us, please go to the link below:
http://p1j2m3a4.pdhost.com/pdsvr/www/r?1000002317.949.3.PRTCUC++nYGkU4
```

***Abbildung 2**: Auszug aus dem Prüfbericht eines als Spam klassifizierten E-Mails.*

```
SPAM: 36.10 hits, 5 required;
SPAM: * 2.4 -- MiME-Version header (oddly capitalized)
SPAM: * 1.3 -- From: does not include a real name
SPAM: * 0.9 -- From: ends in numbers
SPAM: * 3.5 -- BODY: Claims compliance with Senate Bill 1618
SPAM: * 2.7 -- BODY: Offers a free consultation
SPAM: * 1.3 -- BODY: Incorporates a tracking ID number
SPAM: * 1.0 -- BODY: No such thing as a free lunch (3)
SPAM: * 1.3 -- BODY: Spam phrases score is 13 to 21 (high)
...
SPAM: * 0.2 -- BODY: JavaScript code
SPAM: * 0.2 -- BODY: Image tag with an ID code to identify you
SPAM: * 0.6 -- BODY: Tells you to click on a URL (in caps)
SPAM: * 1.4 -- RAW: Message text disguised using base-64 encoding
SPAM: * 0.7 -- URI: Includes a link to a likely spammer email address
SPAM: * 3.9 -- Listed in Razor2, see http://razor.sf.net/
SPAM: * -0.1 -- Email came from some known mailing list software
SPAM: * 3.2 -- RBL: Received via a relay in list.dsbl.org
SPAM: [RBL check: found 130.71.151.203.list.dsbl.org]
SPAM: * 0.5 -- Message has X-MSMail-Priority, but no X-MimeOLE
...
```

3 Heuristische Verfahren

Als Non-Plus-Ultra haben sich jedoch heuristische Verfahren erwiesen, welche die Wahrscheinlichkeit, dass es sich bei einer bestimmten Mail um Spam handelt, anhand einer Kombination von Einzelkriterien bestimmen. Im Falle von Spamassassin ***[1]*** sind das weit über 100 solcher Kriterien!

Hierbei kommen Techniken wie

- Bewertung des in der Mail verwendeten Vokabulars
- Analyse des Formates der Nachricht (Mailheader, gültiges HTML, verwendetes E-Mailprogramm, etc.)
- Abfrage von Open Relay Datenbanken
- Prüfsummendatenbanken

zur Anwendung.

Anschließend werden die Ergebnisse der Einzeltests entsprechend der Vorgaben des Benutzers gewichtet und eine zur Wahrscheinlichkeit, dass es sich bei der Nachricht um Spam handelt, proportionale Punktezahl dem Kopf der Nachricht hinzugefügt.

Der Benutzer kann nun Grenzen definieren, innerhalb deren er Mails akzeptiert beziehungsweise löschen/filtern lässt. Weiters ist es möglich, E-Mails, die als Spam klassifiziert wurden, um einen Prüfbericht zu erweitern ***(Abbildung 2)***.

Möchte man heuristische Mailfilter zum Schutz des eigenen Mailservers einsetzen, so empfiehlt es sich in den meisten Fällen, Klassifikationssystem und Mailsoftware auf unterschiedlichen Rechnern laufen zu lassen, da die Klassifikation der Nachrichten relativ ressourcenintensiv ist und das System somit bei größeren Mailvolumina nicht mehr einsetzbar wäre.

Das Klassifikationssystem kann dabei auch von externen Anbietern zugekauft werden ***(Abbildung 3)***.

Weiters sollten die Einstellungen des Spamfilters von den Benutzern - zum Beispiel über ein Webinterface ***(Abbildung 4)*** - individuell verwaltet werden können, sodass niemand mit einem "Zensurierungssystem" zwangsbeglückt wird.

Die Integration des heuristischen Filters in bestehende Mailsysteme kann bei neuerer Software problemlos über entsprechende Content-Filter erfolgen, sodass alle Benutzer in den Genuss der mit zwischen 95 und 99 % extrem hohen Erkennungsrate des Systems kommen können.

Literatur

[1] Spamassassin - ein heuristischer Mailfilter
http://www.spamassassin.org

[2] Vipul's Razor
http://razor.sourceforge.net

[3] Spamnet Outlook Add-In
http://www.spamnet.com

[4] Mail Abuse Prevention System
http://www.mail-abuse.org

[5] Open Relay Database (ORDB)
http://www.ordb.org

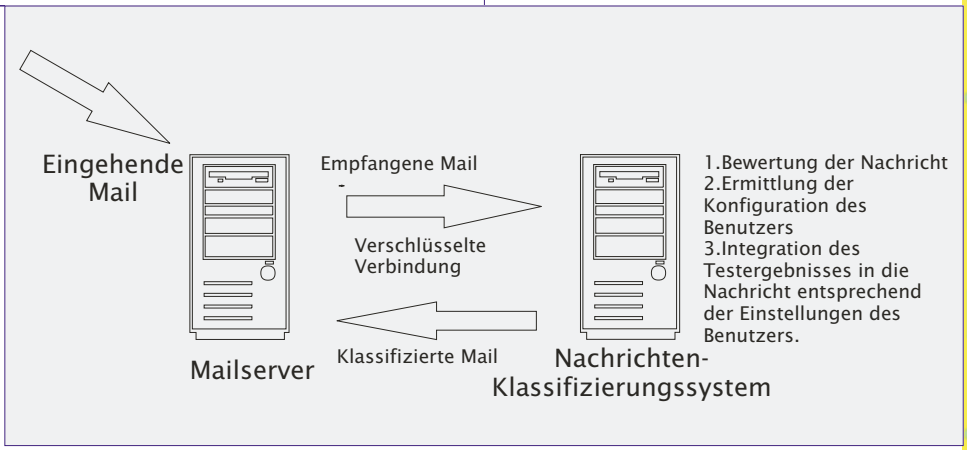

***Abbildung 3**: Integration von Mailklassifikationssystemen*

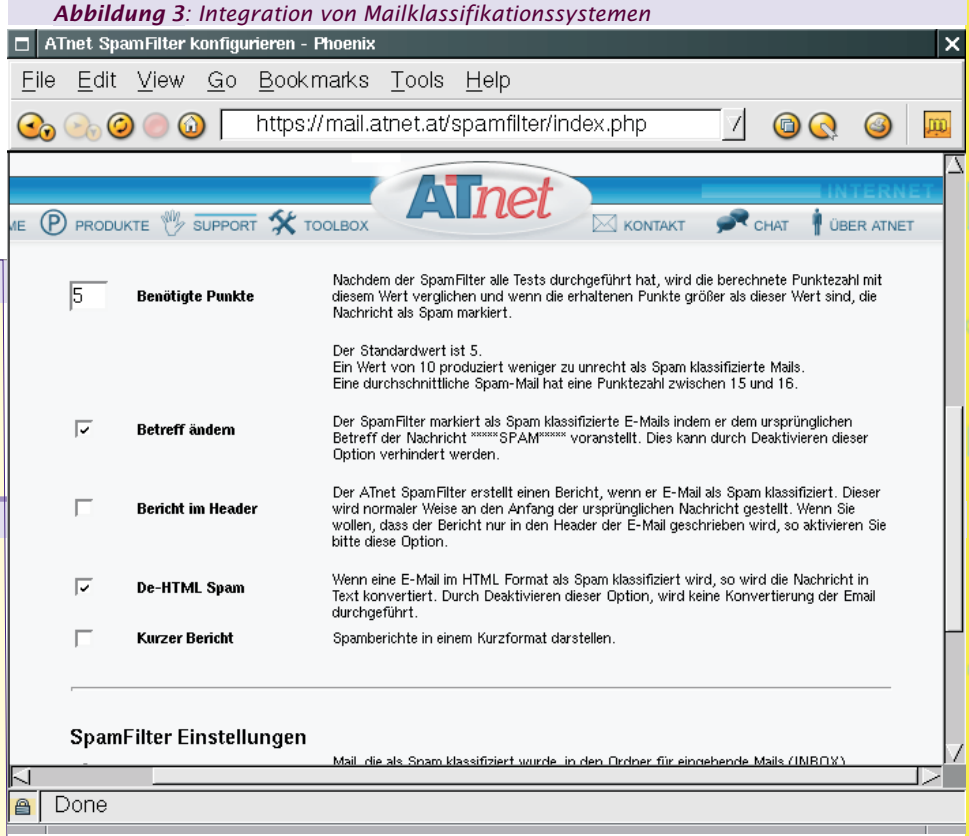

***Abbildung 4:** Spamfilter Konfiguration*

Security

European E-Mail Security End-user Study

Frost & Sullivan

Inhaltsfilterung von E-Mails: Anbieter müssen besser über Vorteile informieren

Potenzielle Kunden wissen noch nicht ausreichend Bescheid über die Möglichkeit der Inhaltsfilterung von E-Mails, weshalb entsprechende Softwaresysteme noch nicht in adäquatem Umfang nachgefragt werden. Eine neue Analyse der Unternehmensberatung Frost & Sullivan geht allerdings davon aus, dass die jüngsten heftigen Diskussionen um den Missbrauch unternehmenseigener E-Mail-Systeme das Bewusstsein für Sicherheitslösungen schärfen und damit einen erhöhten Bedarf generieren werden.

Diese Erkenntnisse beruhen auf ausführlichen Interviews mit mehr als 200 IT-Entscheidungsträgern in mittleren und großen Unternehmen in Großbritannien, Skandinavien, Deutschland und Frankreich, Europas wichtigsten Ländermärkten für Sicherheitssoftware. Gefragt wurde nach Markenbewusstsein sowie nach der qualitativen Einsstufung von Marktakteuren wie Clearswift, Symantec, Trend Micro, Computer Associates, SurfControl, Message Labs, Tumbleweed, Network Associates, Brightmail, Marshal Software, Invisimail und Aladdin.

Ein Fünftel der potenziellen Kunden kennt die Technologie nicht

Laut Analyse müssen sich die Anbieter von Filtersystemen in der nächsten Zukunft hinsichtlich Kundenbewusstsein und Nutzung von Software zur Inhaltsüberwachung zentralen Herausforderungen stellen. So gaben nahezu 20 Prozent der Befragten an, über das Angebot entsprechender Lösungen nicht informiert zu sein, und ein weiterer Prozentsatz wusste nur bedingt über die erhältlichen Produkte Bescheid. Von den Befragten, die die angebotenen Lösungen kennen, setzen allerdings knapp 53 Prozent sie nicht ein. Die große Mehrheit hatte der Installation solcher Sicherheitssysteme zudem innerhalb ihres Unternehmens niedrige Priorität eingeräumt.

Nutzwert muss besser kommuniziert werden

Grund zum Optimismus gibt es dennoch, meint Jose Lopez, Sicherheitsexperte bei Frost & Sullivan: "Es muss den Anbietern nur gelingen, die Zielgruppe vom hohen Nutzen der Inhaltsfilterung zu überzeugen. Dabei werden die zunehmende Anzahl von Negativmeldungen über die unerlaubte Nutzung bzw. Belastung von Firmen-E-Mail-Systemen und die sich daraus ergebenden Sicherheitsprobleme sicherlich als wichtige Argumente dienen."

Um Kaufinteresse zu wecken, sollten die angebotenen Lösungen aus der Perspektive der potenziellen Endnutzer vor allem drei Aspekte abdecken: erstens die Bedrohung durch Viren und unautorisierte Weitergabe unternehmensinterner Informationen, zweitens die unnötige Belastung der Systeme durch Spam oder Junk Mails (Massenmails wie Werbung, Kettenbriefe, Pyramidenspiele etc.) und drittens E-Mails mit beleidigendem oder pornographischem Inhalt.

Virenabwehr ist Hauptgrund für Einsatz

"An erster Stelle geht es den Firmen immer noch um die Viren," kommentiert Lopez, "doch gibt es in Bezug auf E-Mail-Sicherheit noch andere Probleme, die nicht einfach durch Installation einer Anti-Virus-Software gelöst werden können. Da gibt es beispielsweise vertrauliche Informationen, die über undichte Stellen an die Konkurrenz gelangen könnten. Das ist vor allem in der Finanzbranche oder in der Luftfahrt- und Verteidigungsindustrie relevant, aber im Prinzip ist jedes Unternehmen betroffen, das Forschung betreibt. Das sind die Punkte, an denen die Anbieter von E-Mail-Inhaltsfilterung ansetzen müssen."