ERNW
Enno Rey Netzwerke GmbH

ERNW Newsletter 3 / Juli 2003

Liebe Partner, liebe Kollegen,

willkommen zur neuen Ausgabe des ERNW-Newsletters, die sich diesmal mit der allgegenwärtigen und stetig ansteigenden Spam-Flut beschäftigt.

Der Begriff "Spam" geht nach herrschender Meinung auf ein amerikanisches Dosenfleisch umstrittener Geschmacksgüte namens "SPAM" ["Spiced HAM", der Hersteller legt zur Abgrenzung seines Produkts Wert auf die Schreibweise in Grossbuchstaben] und einen Monty Python-Sketch der 70er Jahre zurück.

Dort versucht ein Ehepaar in einem Restaurant erfolglos, ein Gericht ohne SPAM zu bestellen, während im Hintergrund eine Horde Wikinger die Dialoge zwischen Kellnerin und Ehepaar immer stärker durch "spam spam spam"-Gesänge übertönt.

Die eigentliche Kommunikation wird damit unmöglich, vergleichbar mit der zunehmenden Beeinträchtigung des Kommunikationsmediums Mail durch unerwünschte Massenmail (Spam).

Spam beansprucht dabei nicht nur die Arbeitszeit und Nerven der Empfänger, sondern verbraucht auch Ressourcen wie Bandbreite, Traffic-Volumen und Speicherplatz oder Rechenleistung von Mailservern. Gleichzeitig werden die Adress-Datenbanken der Versender immer umfangreicher und ihre Mechanismen immer raffinierter.

Methoden der Spam-Bekämpfung

Spam-Mails zeichnen sich üblicherweise durch verschiedene Merkmale aus, die einen Hinweis darauf liefern, dass es sich eben um Spam handeln könnte. Zu diesen Merkmalen zählen die Quelle der Mails [die möglicherweise keinen Reverse DNS-Eintrag aufweist oder in einer sog. Blackhole List offener Mailrelays auftaucht], fehlende RFC-Konformität bei der Mail-Einlieferung [etwa ein fehlendes HELO/EHLO oder syntaktisch falsche bzw. nicht voll-qualifizierte Hostnamen], auffällige Mail-Header [bei denen z.B. das "Reply-To"-Feld leer ist oder das Datum in die Zukunft verlegt ist, um beim empfangenden Client entsprechend einsortiert zu werden] oder natürlich Mail-Inhalte mit Schlüsselwörtern ["Viagra", "Nigeria"] oder eindeutigen HTML-Bestandteilen ["Click here"].

Jedes dieser Merkmale _kann_ ein Indiz für ein Spam sein; viele können jedoch auch in durchaus reellen Mails auftreten [Mails an einen Urologen enthalten vielleicht gewünschte Viagra-Anfragen und auch die Mailer von GMX sind unlängst auf einer Blackhole List gelandet].

Die meisten Spam-Filter versuchen, Spam im wesentlichen anhand eines dieser Merkmale zu erkennen [etwa durch eine Prüfung von Blackhole Lists oder durch einen Vergleich mit vorkonfigurierten Textmustern ("Viagra"). Dies führt jedoch zu einer steigenden Zahl an "false positives" [als Spam klassifizierte Mails, die gar kein Spam sind (GMX/spamcop)] und "false negatives" [Spam-Mails, die nicht als solche erkannt werden ("V*i*a*g*r*a")].

Die (zumindest für den einzelnen Nutzer) viel-versprechendste Vorgehensweise ist das sog. "Bayesian Filtering", benannt nach dem englischen Pfarrer Thomas Bayes (1702-1761) und seinem "Essay towards solving a Problem in the Doctrine of Chances".

Es geht darin um die Wahrscheinlichkeit von Ereignissen vor dem Hintergrund bestimmter Bedingungen (und deren Veränderung). Für unseren Kontext somit in etwa um die Frage: wenn ein Mail das Wort "Viagra" enthält (das ja auch in diesem Mail mehrfach vorkommt), wie wahrscheinlich ist dann, dass es sich um Spam handelt? Und wie ändert sich diese Wahrscheinlichkeit, wenn man einen grösseren Mail-Korpus betrachtet [bspw. im Falle des Urologen]? Bayes-Filter zur Spam-Abwehr setzen üblicherweise voraus, dass man dem Filter möglichst viele nach den Kategorien "Spam" und "Ham" [= erwünschte Mails] vorsortierte Mails zum Lernen übergibt, die dann nach statistischen Kriterien ausgewertet werden [etwa der Häufigkeit von Wörtern oder Buchstabenkombinationen].

Diese Methode gilt als die akkurateste (und ist daher z.B. im Mozilla integriert); mit grösserer Bandbreite des verarbeiteten Mailkorpus (also Mails von vielen Usern mit unterschiedlichem Mail-Verhalten) verliert sie jedoch an Treffer-Genauigkeit (weshalb sie auch unter Providern umstritten ist).

Alle genannten Methoden weisen also Schwächen auf, die zu "false positives" oder "false negatives" führen können. Bei vielen Filtern werden zudem Merkmale sequentiell geprüft ["Gibt es einen Reverse DNS Eintrag? Falls ja, steht die Absender-IP in einer Blackhole List?" usw.], so dass schon ein einziges Merkmal zur Ablehnung eines Mail führen kann.

Zugleich werden Spams immer mehr den gängigen Filter-Methoden angepasst: zum Versand werden offene Web-Proxies (und nicht offene

Mailrelays) verwendet, die Subject-Zeilen werden mit polymorphen unsinnigen Zeichen aufgefüllt oder die famosen Millionen-Summen der Nachkommen verunglückter Diktatoren sollen inzwischen aus Liberia (und nicht mehr aus Nigeria) transferiert werden.

Zur effektiven Anti-Spam Filterung sind daher Werkzeuge erforderlich, die verschiedene Erkennungsmethoden kombinieren können und die idealerweise anstelle einer (Spam|Ham)-Bewertung eher mit Wahrscheinlichkeiten operieren ("das Mail weist diese oder jene Merkmale auf, daher ist die Spam-Wahrscheinlichkeit soundso gross").

Das bekannteste und (nicht nur) unserer Meinung nach beste ist der SpamAssassin (www.spamassassin.org). Dieses Open Source (GPL) Tool führt pro eingelieferter Mail eine Reihe von intelligenten Prüfungen durch und vergibt für jedes erkannte Merkmal einen Punktwert unterschiedlicher Höhe.

Wenn die Gesamt-Punktzahl eines Mails einen (konfigurierbaren) Schwellen-Wert überschreitet, wird das Mail im Header (und optional im Subject) als Spam markiert und zunächst normal weitergeleitet. Anhand dieses Headers kann dann eine weitere Prozessierung erfolgen, etwa ein Einsortieren in einen bestimmten Mail-Ordner oder die Weiterleitung an eine bestimmte Adresse.

Die Installation des SpamAssassin und Einbindung in vorhandene Mail-Abläufe ist relativ leicht möglich und schon die Default-Konfiguration führt zu einer enorm hohen Erkennungs-Rate empfangener Spams bei ausgesprochen wenig "false positives". Wir selbst setzen den SpamAssassin seit einiger Zeit mit sehr guten Ergebnissen ein und viele kommerzielle Lösungen wie der Spam-Filter der WebShield Appliance von NAI oder die Anti-Spam Mechanismen vieler Freemail-Anbieter basieren darauf.

Die Kombination vieler unterschiedlicher Tests und einer darauf basierenden "Spam-Wahrscheinlichkeit" ist sicher die beste, aktuell verfuegbare Methode der Spam-Bekämpfung.

Wir werden in den kommenden Monaten in unseren Heidelberger Räumen einige Workshops zum Thema Anti-Spam durchführen und dabei auch eine Lösung auf Basis SpamAssassin demonstrieren. Wir würden uns freuen, Sie bei dieser Gelegenheit begrüssen zu können. Inhalte, Termine und Preise erhalten Sie unter http://www.ernw.de/training/antispam.pdf oder einfach per Mail an training@ernw.de.

Zum Abschluss noch ein kurzer Nachtrag zum letzten Newsletter: auch Check Point hat ein Advisory zu dem von uns beschriebenen Problem herausgegeben (zu finden unter [2]) und den Schweregrad mit 'High' bewertet.

[1] http://www.detritus.org/spam/skit.html
[2] http://www.checkpoint.com/securitycenter/advisories/2003/cpsa-2003-04.html

Die bisherigen Ausgaben des Newsletters finden Sie unter:
http://www.ernw.de/download/ERNWnewsletter1.pdf
http://www.ernw.de/download/ERNWnewsletter2.pdf

Es grüssen Sie herzlich

Enno Rey & das Team von ERNW

ERNW Enno Rey Netzwerke GmbH - Zaehringerstr. 46 - 69115 Heidelberg
Tel. +49 6221 480390 - Fax 6221 419008 - Mobil +49 173 6745902
www.ernw.de