WHITE PAPER

Effektive Bekämpfung von Spam mit Hilfe der Bayes'schen Filtertechnologie

Wie ein mathematischer Ansatz Erkennungsraten von über 98% garantiert

In diesem White Paper erhalten Sie Details zur Funktionsweise der Bayes'schen Filterung und erfahren, warum die Bayes'sche Methode am besten zur Abwehr unerwünschter Werbemitteilungen geeignet ist.

Einführung

In diesem White Paper wird erläutert, wie dem Problem der Spam-Mitteilungen mit Bayes'schen Verfahren begegnet werden kann und warum diese adaptive Technik mit ihrer "statistischen Intelligenz" sehr hohe Spam-Erkennungsraten bietet.

Zudem wird hervorgehoben, warum der Bayes'sche Ansatz der Spam-Bekämpfung allen anderen Methoden, die lediglich auf statischen Technologien basieren, darunter die Blacklist-Kontrolle, Überprüfungen von Datenbanken zu bekannten Spammern und die Stichwort-Kontrolle, weitaus überlegen ist. Diese Technologien sind zwar nicht veraltet, jedoch längst nicht so zuverlässig wie der Einsatz eines Bayes'schen Filters.

Aktuelle Techniken der Spam-Identifizierung

Unerwünschte Werbemitteilungen haben sich zu einem ernsthaften Problem entwickelt. Die Anzahl der Spam-Mitteilungen wächst täglich um ein Vielfaches – Studien haben ergeben, dass über 50% aller E-Mails mittlerweile aus Spam bestehen. Laut Forschern der US-amerikanischen Radicati Group wird dieser Wert bis zum Jahr 2007 auf 70% anwachsen. Zudem wenden Spammer immer geschicktere Methoden zum Umgehen von Software-Lösungen an, die Spam mit Hilfe “statischer“ Methoden zu blockieren versuchen.

“Statisch” in diesem Zusammenhang bedeutet, dass Anti-Spam-Software sehr leicht zu überlisten ist, indem der Absender der Spam-Mitteilung Format und Inhalt seiner Nachricht einfach an aktuelle Abwehrmaßnahmen anpasst, um die starren Kontrollmechanismen vieler Anti-Spam-Lösungen zu umgehen und die Werbenachricht unerkannt in das Postfach des Anwenders zu transportieren.

Für eine erfolgreiche Bekämpfung von Spam ist daher eine adaptive Technologie erforderlich. Lösungen zur Spam-Abwehr müssen anpassungsfähig sein, um sofort auf neu eingesetzte Taktiken von Spammern reagieren und Anti-Spam-Regeln entsprechend aktualisieren zu können. Zudem muss sich eine solche Anti-Spam-Lösung aber auch an die speziellen Anforderungen des Unternehmens, in dem sie eingesetzt wird, anpassen lassen. Die Lösung hierfür: Bayes’sche Filter.

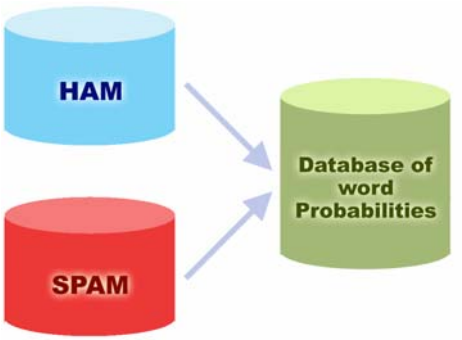
Funktionsweise des Bayes'schen Spam-Filters

Die Bayes'sche Filtertechnologie basiert auf dem mathematischen Prinzip, dass die meisten Ereignisse voneinander abhängig sind und dass die Wahrscheinlichkeit eines zukünftigen Ereignisses aus vorherigen Ereigniseintritten abgeleitet werden kann. (Weitere Hintergrundinformationen zur Berechnungsgrundlage, die bei der Bayes'schen Filterung verwendet wird, stehen zur Verfügung in dem Beitrag "Bayesian Parameter Estimation" unter http://www-ccrma.stanford.edu/~jos/bayes/Bayesian_Parameter_Estimation.html und "An Introduction to Bayesian Networks and their Contemporary Applications" unter http://www.niedermayer.ca/papers/bayesian/bayes.html).

Diese Technik eignet sich ideal zum Identifizieren von Spam. Einige Begriffe oder Inhalte werden oft nur in Spam-Mitteilungen verwendet und kommen in "normalen" E-Mails nicht vor. Werden dieselben Begriffe auch in einer neuen E-Mail gefunden, handelt es sich bei dieser Nachricht mit größter Wahrscheinlichkeit um Spam.

Erstellen einer angepassten Bayes'schen Begriffsdatenbank

Bevor Nachrichten mit Hilfe diese Methode gefiltert werden können, muss eine Datenbank mit Wörtern oder Token (Dollar-Zeichen, IP-Adressen, Domänen etc.) erstellt werden, die aus Spam-Beispielen und gültigen Mitteilungen (auch als "Ham" bezeichnet) stammen.



Erstellung einer Begriffsdatenbank für den Filter

Danach wird jedem Wort oder Token ein Wahrscheinlichkeitswert zugewiesen. Die Wahrscheinlichkeit basiert hierbei auf Berechnungen, in die einfließt, wie oft das in der Spam-Mail verwendete Wort/der Token im Vergleich mit erwünschten Nachrichten (Ham) vorkommt. Als Datenmaterial dienen ausgehende Benutzerpost und Analyseergebnisse bekannter Spam-Mitteilungen. Hierbei werden sämtliche Wörter und Token in beiden E-Mail-Kategorien analysiert, um bestimmten Wörtern einen Wahrscheinlichkeitswert zuzuweisen, der sie als Spam-typisch klassifiziert.

Der Wahrscheinlichkeitswert wird wie folgt berechnet: Beim Wort "Hypothek", das in 3.000 Spam-Mails 400 Mal vorkommt und in 300 erwünschten E-Mails 5 Mal verwendet wird, liegt die Spam-Wahrscheinlichkeit bei 0,8889. Berechnung: (400/3000) / (5/300 + 400/3000) = 0.8889.

Erstellung einer an ein Unternehmen angepassten Ham-Datenbank

Im Zusammenhang mit der Ham-Datenbank muss beachtet werden, dass die Analyse von Ham-Mitteilungen speziell für die E-Mail-Korrespondenz eines einzelnen Unternehmens vorgenommen wird und die Ergebnisse somit nicht allgemeingültig sind. Beispielsweise würde die Anwendung einer allgemeinen Anti-Spam-Regel auf die E-Mail-Korrespondenz eines Finanzunternehmens sehr viele Fehlalarme auslösen, da von diesem das sonst für Spam übliche Wort "Hypothek" sehr häufig und legitim verwendet wird. Nach einer Lernphase und Anpassung an die Unternehmenskorrespondenz zu Beginn seines Einsatzes überprüft der Bayes'sche Filter jedoch die gültige ausgehende Post des Unternehmens, erkennt dabei "Hypothek" als häufig eingesetzten Begriff in ordnungsgemäßen Mitteilungen und vermeidet somit eine zu hohe Anzahl von Fehlalarmen – die Erkennungsrate "echter" Spam-Mitteilungen ist somit viel höher.

Bitte beachten Sie, dass bei einigen Anti-Spam-Lösungen die Bayes'sche Filtertechnologie nicht sehr umfassend integriert wurde, wie beim Spam-Filter von Outlook oder dem Internet Message Filter von Exchange Server. Diese Lösungen bieten lediglich eine standardmäßige Ham-Datei, die nicht dynamisch erweitert wird und sich nicht an die Anforderungen Ihres Unternehmens anpassen lässt. Obwohl bei dieser Technologie keine Lernphase erforderlich ist, weist sie zwei grundlegende Mängel auf:

1. Die Datei mit den Ham-Daten ist öffentlich zugänglich und kann daher von professionellen Spammern gehackt und somit umgangen werden. Ist die Ham-Datei jedoch an Ihr Unternehmen angepasst und somit einzigartig, hat das Hacken der Ham-Datei keinen Sinn. Auch beispielsweise für den Spam-Filter von Outlook 2003 oder Exchange Server sind bereits Hacks aufgetaucht.

2. Der Inhalt der Ham-Datendatei ist sehr allgemein gehalten. Da die speziellen Eigenheiten der für Ihr Unternehmen üblichen Korrespondenz nicht berücksichtigt werden, ist diese Datei längst nicht so effektiv wie eine individuell angepasste Ham-Datei und hat eine höhere Anzahl von Fehlalarmen zur Folge.

Erstellung einer Spam-Datenbank

Zusätzlich zur Ham-Datenbank greift der Bayes'sche Filter zur Kontrolle von Mitteilungen auch auf eine Datei mit Spam-Daten zurück. In dieser Datei muss für eine optimale Erkennung eine große Auswahl bekannter Spam-Mitteilungen gespeichert sein. Daher ist es erforderlich, dass diese Datei von der Anti-Spam-Lösung kontinuierlich mit den neuesten Spam-Nachrichten aktualisiert wird. Dadurch wird sichergestellt, dass der Bayes'sche Filter auch sämtliche neue Spam-Tricks erkennt und eine hohe Erkennungsrate erzielt (Hinweis: Optimale Erkennungsraten werden erst nach der zweiwöchigen Lernphase erzielt.)

Vorgehensweise bei der Filterung

Sind die Datenbanken für Ham- und Spam-Nachrichten erstellt, kann der Spam-

Wahrscheinlichkeitswert für die einzelnen Wörter berechnet werden, und der Filter ist einsatzbereit.

Trifft eine neue E-Mail ein, wird sie in ihre einzelnen Wörter aufgeschlüsselt, wobei die wichtigsten herausgegriffen werden, d. h. solche, die für die Klassifizierung von E-Mail als Spam von größter Bedeutung sind. Unter Berücksichtigung diese Wörter errechnet der Bayes'sche Filter die Wahrscheinlichkeit, ob eine neue Mitteilung als Spam eingestuft werden muss oder nicht. Ist die Wahrscheinlichkeit größer als ein bestimmter Schwellenwert, z. B. 0,9, wird die Nachricht als Spam gekennzeichnet.

Der Bayes'sche Ansatz der Spam-Bekämpfung hat sich als äußerst effektiv erwiesen. In einer Meldung der britischen BBC wurde hervorgehoben, dass mit dieser Methode über 99,7% aller Spam-Mitteilungen erkannt werden und gleichzeitig eine sehr geringe Anzahl von Fehlalarmen auftritt.

Vorteile der Bayes'schen Filtertechnologie

1. Die Bayes'sche Methode berücksichtigt die gesamte Mitteilung. Hierbei werden nicht nur Spam-typische Stichwörter erkannt, sondern auch solche, die in gültiger E-Mail-Korrespondenz vorkommen. Beispiel: Nicht jede E-Mail, in der die Wörter "kostenlos" und "bares Geld" vorkommen, ist als Spam einzustufen. Der Vorteil der Bayes'schen Methode besteht darin, dass Wörter, die am auffälligsten sind (ermittelt durch ihre Abweichung vom gängigen Wortschatz), kontrolliert werden, um dann die Wahrscheinlichkeit für das Vorliegen einer Spam-Mitteilung zu berechnen. Die Bayes'sche Methode würde die Wörter "bares Geld" und "kostenlos" zwar als auffällig einstufen, jedoch gleichzeitig den Namen des Absenders als Geschäftskontakt erkennen und die Mitteilung somit als legitim klassifizieren. Somit wirken mehrere berücksichtigte Wörter und Merkmale als "ausgleichend". Diese Filterung stellt einen wesentlich intelligenteren Ansatz der Spam-Abwehr dar, da alle Merkmale einer Mitteilung untersucht werden – und nicht nur einzelne Stichwörter, deren Verwendung in einer E-Mail bereits ausreicht, diese (fälschlicherweise) als Spam zu klassifizieren.

2. Ein Bayes'scher Filter aktualisiert sich automatisch und kontinuierlich. Der Filter verarbeitet die Merkmale neuartiger Spam-Mitteilungen und neuer, gültiger E-Mails, die verschickt werden, und passt sich somit aktuellen Spam-Methoden und veränderten Korrespondenzgewohnheiten an. Spammer haben früher z. B. die Schreibweise "k-o-s-t-e-n-l-o-s" statt "kostenlos" verwendet, um die Stichwort-Kontrolle zu umgehen. Diese Taktik war solange erfolgreich, bis die neue Methode in der Stichwort-Datenbank aufgenommen wurde. Der Bayes'sche Filter hingegen erkennt diese Manipulationen automatisch. Die veränderte Schreibweise wird von ihm sogar als ein eindeutiges Merkmal für Spam gedeutet, da es unwahrscheinlich ist, dass sie in dieser Form in Ham-Mitteilungen vorkommt. Ein weiteres Beispiel für eine auffällige Schreibweise ist die Verwendung des Wortes "5ex" an Stelle von "Sex". In einer erwünschten Mitteilung wird das Wort "5ex" wohl

kaum verwendet. Die Wahrscheinlichkeit, dass es sich bei einer Mitteilung mit diesem Wort um Spam handelt, steigt somit.

3. Der Bayes'sche Filter passt sich flexibel an die Benutzerkorrespondenz an. Die Korrespondenzgewohnheiten des Unternehmens in der E-Mail-Kommunikation werden vom
   Bayes'schen Filter berücksichtigt. So erkennt der Filter z. B. beim Einsatz im E-Mail-System eines Autohändlers, dass das Wort "Hypothek" in einer E-Mail ein Anzeichen für eine Spam-Mitteilung ist, wohingegen der Filter in einem Finanzunternehmen dieses Wort nicht als verdächtig einstufen würde.

4. Die Bayes'sche Methode kann für mehrere Sprachen und international eingesetzt werden. Da sich der Bayes'sche Anti-Spam-Filter automatisch anpasst, lässt er sich für jede Sprache verwenden. Ein Großteil der Stichwort-Listen ist nur auf Englisch verfügbar und eignet sich daher nicht für die Verwendung in anderen Sprachen. Der Bayes'sche Filter berücksichtigt sogar bestimmte sprachliche Abweichungen oder die unterschiedliche Verwendung von Wörtern in verschiedenen Bereichen. Diese Fähigkeit ermöglicht die Blockierung einer noch größeren Anzahl von Spam-Mitteilungen.

5. Ein Bayes'scher Filter ist schwerer zu überlisten als ein Stichwort-Filter. Versierte Spammer, die einen Bayes'schen Filter täuschen wollen, könnten versuchen, weniger negativ belegte Wörter (Spam-typische wie "kostenlos", "Viagra" etc.) zu verwenden oder mehrere Begriffe, die allgemein auf eine gültige Nachricht hinweisen (z. B. ein gültiger Kontaktname). Letztere Variante kann jedoch nicht realisiert werden, da dem Spammer das E-Mail-Profil eines jeden Empfängers bekannt sein müsste. Für Spam-Versender ist es jedoch so gut wie unmöglich, an diese Information eines jeden potenziellen Empfängers zu gelangen. Bei Verwendung von neutralen Begriffen wie "öffentlich" scheitert dieser Versuch, da diese bei der Endanalyse nicht berücksichtigt werden. Auch die Aufteilung von Spam-typischen Wörtern, z. B. "H-y-p-o-t-h-e-k" an Stelle von "Hypothek" verspricht keinen Erfolg, sondern bewirkt eher das Gegenteil: Wohl kaum ein Benutzer wird die erste Schreibweise in seinen Nachrichten verwenden.

Bayes'sche Filter oder aktualisierte Stichwort-Listen?

Einige Versionen von Anti-Spam-Software laden in regelmäßigen Abständen aktualisierte Listen mit Begriffen herunter, die für Spam-Mitteilungen typisch sind. Obwohl diese Vorgehensweise empfehlenswerter ist, als gar keine Aktualisierung von Stichwort-Listen durchzuführen, handelt es sich hierbei jedoch um einen recht lückenhaften Schutz, der leicht umgangen werden kann. Das System ist bereits vom Prinzip her wesentlich unausgereifter als ein Bayes'scher Filter.

Hinweis zum Einsatz des Bayes'schen Filters

Keine andere Anti-Spam-Technologie ist so effizient in der Abwehr unerwünschter

Werbemitteilungen wie die Bayes'sche Filtertechnologie – eine korrekte Implementierung und Anpassung vorausgesetzt. Es besteht jedoch auf den ersten Blick ein Nachteil, der aber nicht sehr gravierend ist: Ein effizienter Einsatz und eine aussagekräftige Beurteilung des Bayes'schen Filters ist erst möglich, nachdem der Filter mindestens zwei Wochen lang Ihre E-Mail-Korrespondenz analysiert und daraus gelernt hat. Eine weitere Möglichkeit wäre, die Ham- und Spam-Datenbanken manuell zu erstellen. Dies ist jedoch sehr zeitaufwändig und komplex, sodass Sie das Ende der Lernphase und automatischen Konfigurierung des Filters abwarten sollten. Im Laufe der gesamten Einsatzzeit verbessert sich zudem die Effizienz des Bayes'sche Filters, da die Anpassung an die Korrespondenzgewohnheiten immer weiter verfeinert wird.

Bei der Bewertung von Anti-Spam-Software sollte dieser Aspekt daher nicht außer Acht gelassen werden. Kommt die erweiterte, individuelle Bayes'sche Analysemethode zum Einsatz, kann erst nach einiger Zeit über die Zuverlässigkeit ein Urteil abgegeben werden. Anti-Spam-Software mit herkömmlichen Abwehrmethoden kann, verglichen mit dem Bayes'schen Filter, am Anfang des Einsatzes zwar größere Erfolge bei der Blockierung von Spam vorweisen – nach nur wenigen Wochen ist dieser Nachteil jedoch ausgeglichen, und der Bayes'sche Filter zeigt seine überlegene Leistungsfähigkeit mit wesentlich höheren Spam-Erkennungsraten.

Über GFI MailEssentials

GFI MailEssentials for Exchange/SMTP bietet bereits auf Server-Ebene Schutz gegen Spam-Mails und macht eine zusätzliche Installation und Aktualisierung von Anti-Viren-Software auf jedem einzelnen Desktop-Rechner somit überflüssig. GFI MailEssentials ist schnell einzurichten und bietet dank der Bayes'schen Filtertechnologie und anderer effizienter Methoden eine hohe Spam-Erkennungsrate. Es ist keine Konfiguration erforderlich, zudem verringert die automatische Whitelist die Anzahl von Fehlalarmen. GFI MailEssentials passt sich automatisch an das E-Mail-System von Unternehmen an, um stets eine optimale Spam-Erkennung zu garantieren. Zudem lässt sich Spam automatisch in Benutzer-Ordner für Junk-E-Mails verschieben. Zudem erweitert GFI MailEssentials die Funktionalität von E-Mail-Servern mit wichtigen Verwaltungsfunktionen: POP3-Downloader, Server-basierte Auto-Replies, E-Mail-Archivierung, -Überwachung und -Berichterstellung sowie Disclaimer. Weitere Produktinformationen und eine kostenfreie Testversion stehen als Vollprodukt unter http://www.gfisoftware.de/de/mes zum Down-load bereit.

Über GFI

GFI (www.gfisoftware.de) ist ein führender Entwickler und Anbieter von Produkten für Netzwerk- und Inhaltssicherheit sowie von Kommunikationslösungen. Das Produktportfolio von GFI umfasst unter anderem den Fax-Connector GFI FAXmaker für Exchange- und SMTP-Mail-Server, die Sicherheitslösung GFI MailSecurity for Exchange/SMTP zur Überprüfung von E-Mail-Inhalten und zum Schutz vor E-Mail-basierten Exploits und Viren, die Server-basierte Anti-Spam-Software GFI MailEssentials for Exchange/SMTP, die E-Mail-Archivierungslösung GFI MailArchiver, GFI LANguard Network Security Scanner (N.S.S.) für Sicherheits-Scans und Patch-Management, GFI Network Server Monitor zum automatischen Versand von Warnmitteilungen und zur Fehlerbehebung bei Netzwerk- und Server-Problemen, GFI LANguard Security Event Log Monitor (S.E.L.M.) zur Ereignisprotokoll-basierten Eindringlingserkennung und netzwerkweiten Verwaltung von Ereignisprotokollen, GFI LANguard Portable Storage Control (P.S.C.) zur netzwerkweiten Kontrolle wechselbarer Speichermedien sowie GFI WebMonitor zur Überwachung von HTTP/FTP-Verbindungen mit Virenschutz für ISA Server. GFI-Produkte sind im Einsatz bei Microsoft, Telstra, Time Warner Cable, NASA, DHL, Caterpillar, BMW, der US-Steuerbehörde IRS und der USAF. GFI unterhält Niederlassungen in den USA, Großbritannien, Deutschland, Zypern, Rumänien, Australien und Malta und wird von einem weltweiten Netzwerk von Distributoren unterstützt. GFI ist "Microsoft Gold Certified Partner" und erhielt die Auszeichnung "Microsoft Fusion (GEM) Packaged Application of the Year". Weitere Informationen erhalten Sie unter http://www.gfisoftware.de.