Phishing, Pharming und Phraud: Warum es Phreude macht, Spam zu bekämpfen

Von René Wienholtz, CTO STRATO AG, Berlin

„Man lebt nur einmal – probieren Sie es aus!“ Zahlreiche originelle Fundstücke aus dem Anti-Spam-Alltag machen den Umgang mit den lästigen Werbemails zu einem erfüllenden Forschungsobjekt. Dabei ist die Lage ernst: Allein bei STRATO kommen an manchen Tagen bis zu einer Milliarde Mails an – die meisten davon Spam. Steigt diese Flut dramatisch an, so könnte dies schon bald weltweit das komplette E-Mail-System an den Rand des Zusammenbruchs bringen. Dabei ist es aber der falsche Weg, mehr Kapazitäten zur Spam-Filterung, also einfach mehr Rechenleistung bereitzustellen, da diese Kapazitäten von den Spammern einfach sofort ausgelastet würden. Mehr Rechenleistung führt demnach direkt zu mehr Spam. Für einen langfristigen Erfolg geht es also darum, nicht größere Filterkapazitäten aufzubauen, sondern intelligentere Filter zu entwickeln.

Kooperation zwischen Wissenschaft und Industrie

STRATO entwickelt seit 2005 gemeinsam mit der Humboldt-Universität zu Berlin, dem Max-Planck-Institut für Informatik und der Universität Potsdam einen modularen Spam-Filter. Oberstes Ziel ist es dabei, die erwünschten E-Mails zuverlässig zuzustellen und gleichzeitig eine möglichst hohe Spam-Erkennungsrate zu erreichen. Denn es ist für jeden E-Mail-Nutzer einfacher, ab und zu vereinzelte Spam-E-Mails zu löschen, als täglich einen umfangreichen Ordner mit hunderten E-Mails nach versehentlich aussortierten E-Mails zu durchsuchen. Wir wissen: Wer seinem Spam-Filter nicht vertraut, verbringt mehr Zeit mit seinem Junk-Mail-Ordner als mit seinem Posteingang. Deshalb gilt es, die erwünschten E-Mails so sicher zuzustellen, dass der E-Mail-Nutzer den Spam-Ordner gar nicht mehr sehen möchte. Um dies zu erreichen, hat STRATO in Kooperation mit den beiden wissenschaftlichen Instituten verschiedene Module entwickelt, die auch die Analyse von erwünschten E-Mails einschließt.

Freunde schicken sich keinen Spam

Ein Modul analysiert dabei die Beziehungen zwischen den E-Mail-Adressen. Die Grundannahme lautet: Freunde schicken sich keinen Spam. Dazu wird zum Beispiel der Kommunikationsvorgang „mail@wunschname.de schreibt an info@strato.de“ in Form eines anonymen Zahlenwerts gespeichert. Tritt dieser Vorgang erneut auf, wird dies in einer Datenbank vermerkt. Es bildet sich ein „sozialer Graph“. Auch Antwort-E-Mails werden dabei berücksichtigt. So erhält STRATO einen zuverlässigen Hinweis, dass es sich hierbei um erwünschte Kommunikation handelt – auch wenn dabei typische Spam-Begriffe auftauchen.

Der digitale Fingerabdruck verrät den Spammer

Ein weiteres Modul ist in der Lage, Bilderspam zu erkennen. Dabei hängen die Spammer einzeln generierte Bilder mit Textbotschaften an die E-Mails. Da sich alle Bilder innerhalb einer Spam-Kampagne leicht unterscheiden, können sie nicht als identisch erkannt werden.

Deshalb extrahiert das Modul „Fingerprinting“ Gemeinsamkeiten aus einer Bilder-Spam-Kampagne, etwa Farbverteilung oder Aufbau. Im Gegensatz zu einer sehr

rechenaufwendigen Texterkennung aller einzelnen Bilder ist Fingerprinting nicht nur deutlich effizienter, sondern auch zuverlässiger. In einem weiteren Schritt haben STRATO und die Wissenschaftler die Fingerprinting-Funktion auf weitere Dateitypen ausgeweitet. Der Filter erkennt auch zuverlässig Excel-, PDF- und sogar MP3- und Video-Spam.

Bündelweise E-Mails

Die Erweiterung des Fingerprinting ist die „Batch-Erkennung“. Batch heißt auf Deutsch „Bündel“ und geht auf die Versandart ein, denn Spam-Mails werden immer bündelweise verschickt. Wenn der Spam-Filter also eine gewisse Anzahl ähnlicher Mails entdeckt, ist es sehr wahrscheinlich, dass weitere ähnliche Mails zum gleichen Bündel gehören und daher auch Spam sind. Die Batch-Erkennung arbeitet unabhängig vom Spam-Typ und liefert so ein starkes Kriterium für unerwünschte Mails.

Mailserver auf Bewährung

Zahlreiche Statistiken erhöhen die Genauigkeit zusätzlich. Dazu zählt die Ergänzung des „Blacklistings“ von Spam-Servern durch ein Scoring-System. Das Scoring erlaubt eine feinere Abstufung, da nur eine Wahrscheinlichkeit ausgedrückt wird, dass der sendende Server Spam verschickt. Weitere statistische Werte liefert eine umfangreiche Analyse der Mail-Header und anderer Informationen aus dem Internet-Protokoll, die bei der Übertragung der E-Mail anfallen wie beispielsweise die Zahl der abgelehnten Zustellversuche („Bounces“), der Mailserver-Verbindungen und Varianzen in Betreffzeilen. Auch hier sucht der Spam-Filter nach Gemeinsamkeiten, die darauf hindeuten, dass eine Mail Teil einer Spam-Kampagne ist. Insgesamt identifiziert diese so genannte „Serverklassifizierungsmaschine“ rund 20 Prozent mehr Spam-Server als ein konventioneller Blacklist-Provider.

Die dunkle Seite des Netzes analysieren

Spam stammt heutzutage zu einem großen Teil aus Botnetzen, also Zusammenschlüssen gekaperter Rechner. Dank der Batch-Erkennung lässt sich leicht nachvollziehen, woher die E-Mails eines Spam-Bündels kommen. Dabei lässt sich sogar präzise beurteilen, aus welchem Botnetz welcher Spam stammt. Daraus kann dann nicht nur eine dynamische Echtzeit-Blacklist erstellt werden, sondern auch eine Botnet-Karte, die die aktuelle Verteilung und Aktivität von Botnetzen zeigt. Mit dem Wissen über die aktuellen Botnetze ist es dann einfach, auch neue Spam-Kampagnen direkt zu erkennen und deren Spam-Mails abzulehnen. Gleichzeitig bietet die Botnetz-Kartographie ein gutes Kriterium für erwünschte E-Mails, die aber auch bündelweise verschickt werden – zum Beispiel Newsletter. Denn der Mailserver, von dem aus die Newsletter versendet werden, ist nicht auf der Botnetz-Karte verzeichnet. Darüber hinaus lassen sich neue Botnetze und Ausbreitungsprozesse schnell auskundschaften und nachverfolgen.

Link-Spam: Ziele prüfen

Der so genannte „Linked Content Checker“ prüft die Anzahl identischer Links in eingehenden E-Mails. Kommt ein Link verräterisch oft vor, wird das Linkziel automatisch aufgerufen und geprüft. Handelt es sich um Glücksspielseiten, Medikamentenversand oder billige Markenuhren, hat man ein starkes Kriterium für Spam. Da die Spammer sich jedoch auch an die Gegenmaßnahmen anpassen, haben sie den Server, der die Links prüft, auf eine

Blacklist gesetzt. Aber auch wir haben Gegenmaßnahmen: Dieser Server befindet sich mittlerweile hinter dynamischen IP-Ranges – er kann also nicht von einem erwünschten Besucher unterschieden werden.

Zusammenspiel der einzelnen Module

Die Gewichtung der einzelnen Module bestimmt maßgeblich die Filterqualität. Das Modul der sozialen Graphen liefert zum Beispiel ein Positivkriterium für erwünschte E-Mails. Es eignet sich aber nicht dafür, Spam direkt zu erkennen. Auf der anderen Seite spricht eine hohe Anzahl von ähnlichen Bildern in vielen verschiedenen Mails stark für Spam, unterschiedliche Bilder sind aber kein Zeichen für erwünschte E-Mails. Deshalb justiert STRATO die Filtermodule und deren Gewichtungen regelmäßig nach, um auf neue Spam-Entwicklungen zu reagieren.

Spieltheoretische Ansätze

Jeder Spam-Filter muss regelmäßig mit echten Daten trainiert werden, da sonst die Filterqualität mit der Zeit stark nachlässt. Der STRATO Filter muss nur in relativ langen Intervallen trainiert werden, da bereits Ansätze aus der so genannten Spieltheorie integriert sind. Dabei erlauben die Trainingsparameter eine gewisse Unschärfe, um sich verändernden Spam trotzdem zuverlässig erkennen zu können.

Ausblick in die Zukunft

Gemeinsam mit den Wissenschaftlern arbeitet STRATO daran, dass sich der Filter in Zukunft selbst verbessert. Die Ansätze aus der Spieltheorie sollen so erweitert werden, dass der Filter seine Parameter automatisch an neue Spam-Formen anpasst. Gelänge dies, würde sich der manuelle Trainingsaufwand weiter drastisch reduzieren. Ein weiteres wichtiges Thema ist ein personalisiertes Filtertraining, damit sich Urologen über Viagra, Sammler über Rolex-Uhren und Finanzprofis über Pennystocks austauschen können, ohne dabei auf einen Spam-Filter Rücksicht nehmen zu müssen. **Grund zur Phreude**

„Diese Nachricht ist durch alle Ihre Filter durchgekommen“, steht es hämisch in einer Spam-Nachricht. Nein, ist sie nicht. Ich habe sie aus meiner Sammlung bizarrer Spam-Nachrichten. So etwas freut mich.

Über STRATO: STRATO ist der Hosting-Anbieter mit dem besten Preis-Leistungs-Verhältnis: Als eines der weltgrößten Hosting-Unternehmen bietet STRATO Profi-Qualität zum günstigen Preis an. Die Produktpalette reicht von Domains, E-Mail- und Homepage-Paketen, Online-Speicher, Webshops und Servern bis hin zu High-End-Lösungen. STRATO hostet vier Millionen Domains aus sechs Ländern und betreibt zwei TÜV-zertifizierte Rechenzentren. STRATO ist ein Unternehmen der Deutschen Telekom AG.

Pressekontakt: Christina Witt, Pressesprecherin, STRATO AG, Pascalstraße 10, 10587 Berlin, Telefon: 030/88615-262, Telefax: 030/88615-263, presse@strato.de, www.strato.de/presse, http://twitter.com/strato_ag