bley – intelligentes Greylisting ohne Verzögerung

Evgeni Golov

Lehrstuhl für Rechnernetze, Institut für Informatik,
Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf
evgeni.golov@uni-duesseldorf.de
http://bley.mx

Zusammenfassung. Die derzeitigen Lösungen zur Klassifikation von E-Mail als Spam *vor* der Annahme durch den Mail-Server lassen sich in zwei Gruppen einordnen: Greylisting- und Blacklisting-Lösungen. Greylisting verzögert die Annahme von E-Mails, um dadurch den Kosten/Nutzen-Faktor für Spammer zu minimieren – und nimmt dafür eine verzögerte Zustellung auch regulärer E-Mails in Kauf. Blacklisting verweigert die Annahme ganz, sofern der Absender auf einer bekannten schwarzen Liste steht – mit entsprechend negativen Folgen fehlerhafter Einträge in diesen schwarzen Listen. Mit intelligentem Greylisting gehen wir einen dritten Weg, der zwischen diesen Extremen angesiedelt ist. Hierfür erweitern wir Greylisting um eine selektive Komponente, um E-Mails nur dann zu verzögern, wenn sich der Absender nicht standardkonform verhält oder wenn er in einer schwarzen Liste gefunden wurde. Dies verhindert Verzögerungen bei normalen E-Mails und schützt gleichzeitig Versender, die nur versehentlich in eine schwarzen Liste eingetragen wurden. Wir stellen eine Implementation von intelligentem Greylisting – bley – als Policy Daemon für den häufig eingesetzten Mail-Server Postfix vor. Evaluationsergebnisse aus dem Realwelt-Betrieb auf einem großen Mail-Server belegen die Leistungsfähigkeit des Ansatzes.

Schlüsselwörter: E-Mail, Spam, Greylisting, Blacklisting

1 Einführung

Spam ist ein bekanntes und großes Problem. Spam-E-Mails lassen Postfächer volllaufen und verhindern produktives Arbeiten. Gute Spam-Filter werden deshalb zunehmend unverzichtbar. Dabei kann man inhaltsbasierte Filter (z.B. SpamAssassin [1]) verwenden, die zwar effektiv und genau sein können, die aber auch viele Ressourcen benötigen. Alternativ – oder zusätzlich – kann man auch bereits auf Protokollebene Absender herausfiltern, denen man nicht vertraut oder die sich nicht standardkonform verhalten. Dies ist mit sehr viel geringerem Aufwand verbunden als eine inhaltsbasierte Filterung.

Unabhängig von der eingesetzten Filtertechnik sollten aber keinesfalls normale E-Mails verloren gehen oder auch nur verzögert werden. Ersteres ist ein immer wieder auftretendes Problem beim Einsatz schwarzer Listen (Blacklists),

letzteres eine Begleiterscheinung sogenannter Greylisting-Verfahren. Mit intelligentem Greylisting – implementiert im Postfix-Policy-Daemon bley – stellen wir einen Ansatz vor, der Eigenschaften von Blacklisting und Greylisting so kombiniert, dass eine effektive Filterung von Spam-E-Mails auf Protokollebene ohne unnötige Verzögerungen und ohne fehlerhaft abgewiesene E-Mails möglich wird.

2 Bestehende Probleme bei Black- und Greylisting

2.1 Blacklisting

Unter Blacklisting versteht man das Filtern von E-Mails anhand des Absenders, welcher in der Regel anhand seiner IP-Adresse identifiziert wird. Wird die IP-Adresse des Absenders in einer schwarzen Liste gefunden, lehnt der Mail-Server die Annahme der E-Mail mit einem Fehler ab, ohne dass der Empfänger davon in Kenntnis gesetzt wird.

Dies ist kein Problem, solange schwarze Listen nur echte Spam-Versender enthalten, die niemals normale E-Mails verschicken. Es kann aber durchaus vorkommen, dass normale Mail-Server dort eingetragen werden. Gründe dafür können beispielsweise eine kurzzeitige Fehlkonfiguration, Spam-Versender im selben IP-Subnetz (die Verwalter schwarzer Listen blocken gerne ganze Subnetze statt einzelner Adressen) oder Fehlverhalten von Blacklist-Betreibern [2,3] sein. Auch eine Auswertung mehrerer schwarzer Listen vor der Ablehnung einer E-Mail kann nicht gänzlich vor diesem Problem schützen, denn im Normalfall genügt es, wenn die Adresse eines Mail-Servers in zwei schwarzen Listen eingetragen ist, um beispielsweise von `policyd-weight`, einer Blacklisting-Implementierung für den Mail-Server Postfix, abgelehnt zu werden [4].

2.2 Greylisting

Das Problem der verlorenen E-Mails existiert bei Greylisting nicht. Dieses verlässt sich nicht auf externe Informationen, sondern ausschließlich darauf, dass der sendende Mail-Server sich standardkonform verhält und eine erneute Zustellung der E-Mail versucht, wenn der empfangende Mail-Server die Annahme für eine gewisse Zeitspanne (i.d.R. 10–30 Minuten) mit einem temporärem Fehler verweigert. Standardkonforme Mail-Server werden dann nach einer Wartezeit einen erneuten Zustellversuch unternehmen, der dann akzeptiert wird. Greylisting geht von der Annahme aus, dass Spam-Versender den erneuten Zustellversuch entweder gar nicht oder nur einmal nach sehr kurzer Zeit unternehmen, da der Vorgang zu viele Ressourcen verbraucht, die sonst anderweitig genutzt werden könnten [5].

Leider bedeutet dieser Ansatz auch, dass alle E-Mails verzögert im Postfach des Empfängers landen. Zwar versuchen die meisten Greylisting-Implementationen diese Verzögerung mittels Caching bereits früher positiv evaluierter Absenderadressen zu minimieren, dennoch bleibt die Wartezeit zumindest bei der ersten E-Mail eines unbekannten Absenders bestehen.

3 Intelligentes Greylisting mit bley

Um zugleich die unerwünschten Verzögerungen von Greylisting und auch die inhärenten Schwierigkeiten von durch Dritte zusammengestellten schwarzen Listen zu vermeiden, haben wir *intelligentes Greylisting* entwickelt. Anstatt jeden unbekannten Absender zunächst temporär abzuweisen und seine Reaktion abzuwarten, wird der erste Zustellversuch zunächst analysiert. Anhand dieser Analyse wird entschieden, ob die E-Mail sofort angenommen werden soll; andernfalls – und nur dann – kommt Greylisting zum Einsatz und die E-Mail wird mit einem temporären Fehler zunächst abgelehnt.

Die Analyse erfolgt dabei in mehreren Schritten:

1. Ist der Absender in einer bekannten Whitelist, dann nimm die E-Mail sofort an.
2. Ist der Absender in einer bekannten Blacklist, dann setze Greylisting ein.
3. Verwendet der Absender einen nicht standardkonformen Namen im SMTP-HELO, sind Absender und Empfänger Adresse identisch oder verbindet sich der Absender aus einem DialUp-Netzwerk, dann setze Greylisting ein.
4. Schlägt der Sender-Policy-Framework-Check [6] fehl, dann setze Greylisting ein.
5. Falls keiner der vorhergehenden Schritte zum Greylisting geführt hat, dann nimm die E-Mail sofort an.

Mit *bley* haben wir ein Software-Paket entwickelt, das den oben skizzierten Algorithmus implementiert. Bley verwendet die Policy-Daemon-Schnittstelle des bekannten SMTP-Servers Postfix [7] und ist somit leicht in existierende Umgebungen integrierbar. Bley ist unter der Adresse http://bley.mx zum Download verfügbar und wird bereits erfolgreich eingesetzt.

Bley kommuniziert über einen Socket mit dem Mail-Server Postfix. Über diesen Socket signalisiert Postfix eine eingehende SMTP-Verbindung, nachdem der Sender HELO, MAIL FROM und RCPT TO ausgeführt hat – also noch vor der Übertragung der eigentlichen Nachricht. Postfix erwartet dann eine Anweisung, was mit der offenen Verbindung geschehen soll. An dieser Stelle wird der Sender anhand des vorher beschriebenen Verfahrens analysiert und die E-Mail wird entweder abgelehnt (DEFER_IF_PERMIT) oder nicht (DUNNO). Im Falle der Nicht-Ablehnung prüft Postfix, ob die E-Mail an ein existierendes Konto geht und stellt sie ggf. zu.

4 Evaluation

Um die praktische Leistungsfähigkeit von bley zu untersuchen, haben wir das System über die Dauer von vier Wochen auf einem Mail-Server evaluiert, der pro Tag ca. 20.000 E-Mails (davon ca. 99% Spam) verarbeiten muss. Von Interesse war dabei insbesondere die Frage, wie effektiv bley Spam filtern kann, ohne dabei normale E-Mails zu verzögern oder gar zu verlieren.

Vor der Installation wurde auf dem Server `policyd-weight` eingesetzt. Mit dieser Lösung wurden ca. 97,5% der E-Mails als Spam erkannt, gelegentlich traten jedoch auch irrtümlich als Spam klassifizierte E-Mails auf; solche False Positives sind auf jeden Fall zu vermeiden.

Im Vergleich dazu hat `bley` zwar mit 97% nur einen geringfügig kleineren Teil der E-Mails als Spam erkannt, dabei jedoch im gesamten Versuchszeitraum auch nur zwei normale E-Mails verzögert zugestellt und keine einzige fälschlich nicht zugestellt. In Tabelle 1 werden die Ergebnisse der Analyse eines 24-Stunden-Zeitraumes aufgelistet. In diesem Zeitraum gab es 18.843 Zustellversuche, von denen 18.304 (97%) durch `bley` abgewiesen wurden.

%	Grund und ausgeführte Handlung
81.96%	Sender in Blacklist gefunden, Greylisting gestartet
11.56%	Sender hat ein nicht standardkonformes HELO gesendet, Greylisting gestartet
2.63%	Greylisting bereits aktiv, Sender hat nicht ausreichend gewartet, Greylisting fortgesetzt
2.24%	Unbekannter Sender, aber kein Analyse-Schritt verlief negativ, E-Mail angenommen
0.75%	Sender kommt aus einem DialUp Netzwerk, Greylisting gestartet
0.37%	Sender in Whitelist gefunden, E-Mail angenommen
0.23%	SPF-Check fehlgeschlagen, Greylisting gestartet
0.14%	Sender bereits als gutartig bekannt, E-Mail angenommen
0.11%	Sender hat ausreichend gewartet, E-Mail angenommen

Tabelle 1. Ergebnisse eines 24 Stunden Zeitfensters

Dabei ist besonders aufgefallen, dass nur 0,23% der Ablehnungen durch SPF zustande kommen, woraus man schließen kann, dass man den (relativ aufwendigen) SPF-Check ohne wesentliche Verluste einsparen könnte.

5 Zusammenfassung

Die effiziente Filterung von Spam ohne das Risiko des Verlustes erwünschter E-Mails stellt nach wie vor eine Herausforderung dar. Mit bley, unserer Implementation eines intelligenten Greylisting-Verfahrens, stellen wir eine praxistaugliche Lösung vor, die Eigenschaften von Blacklisting und Greylisting kombiniert und dadurch sowohl eine zügige und zuverlässige Zustellung standardkonformer E-Mails als auch eine effektive Filterung von Spam erreicht.

Für den praktischen Einsatz schlagen wir die Kombination von bley mit einer inhaltsbasierten Spam-Erkennung (beispielsweise mit SpamAssassin) als zweite Stufe vor. Durch den hohen Anteil an Spam-E-Mails, die bley bereits mit minimalem Rechenaufwand auf Protokollebene abweisen kann, müssen rechenaufwändige inhaltsbasierte Verfahren dann nur noch auf eine sehr viel geringere Zahl von E-Mails angewendet werden.

Literatur

1. SpamAssassin. http://spamassassin.apache.org/
2. Daniel AJ Sokolov, Peter-Michael Ziegler: Spamhaus.org setzt Österreichs Domainverwaltung unter Druck. http://www.heise.de/newsticker/meldung/91417 (2007)
3. Wolf-Dieter Mergenthaler: Dunkle Mächte. Linux-Magazin 06/10. S. 92–95 (2010)
4. Evgeni Golov: Intelligent Greylisting. Bachelor's thesis, Heinrich-Heine-Universität Düsseldorf. (2009)
5. Evan Harris: The Next Step in the Spam Control War: Greylisting. http://www.greylisting.org/articles/whitepaper.shtml (2003)
6. M. Wong and W. Schlitt: Sender Policy Framework (SPF) for Authorizing Use of Domains in E-Mail, Version 1. RFC 4408 (Experimental). Internet Engineering Task Force. (2006)
7. Postfix SMTP Access Policy Delegation. http://www.postfix.org/SMTPD_POLICY_README.html