Referer-SPAM und „Suchmaschinen-Optimierung“

Viele Webserver werden durch sogenannte „Referer-SPAM“-Attacken immer stärker belastet: Gezielt werden einzelne Webseiten massenhaft aufgerufen.
Eine neue Form des SPAMs greift um sich.

1 Hintergrund

Ziel solcher Referer-SPAM-Attacken ist es, dafür zu sorgen, dass die in der „Referer Information“ übertragene Webadresse in Onlinestatistiken auftaucht.
Diese Statistiken wiederum könnten mit einer gewissen Wahrscheinlichkeit öffentlich sein und so von Suchmaschinen gelesen und analysiert werden.
Dies ist das eigentlich Ziel der Attacken: Viele Suchmaschinen geben einem Webauftritt eine gute Position bei der Ausgabe von Suchergebnissen, wenn viele Links auf diese zeigen.
In anderen Worten:
Referer-Spam ist eine besondere Form des aggressiven Marketings und wird derzeit vor allem von pornografischen Internetangeboten genutzt. Hierbei hinterlässt ein Spambot einen Referer mit dem URL des Angebots auf möglichst vielen Websites. Wenn eine dieser betroffenen Websites ihre Referer veröffentlicht so befinden sich darunter die Spam-Links zu den Pornografie-Seiten.

Dieser Art der Attacken ist leider in letzter Zeit ziemlich stark in Mode gekommen. Es gibt inzwischen eine eigene Branche, die sich selbst als „Suchmaschinen-Optimierer“ bezeichnet und sich dieser (und anderer) Methoden bedienen um Seiten von Auftraggebern zu bewerben.

Im Gegensatz zu E-Mail-SPAM ist diese Form der Werbung noch nicht genauso verpönt und gesellschaftlich gebranntmarkt. Während sich Angebot von Firmen, die E-Mail-SPAM betreiben eher im Hintergrund halten, sind Angebot von „*Suchmaschinen-Optimierer*“ derzeit in jeder Internet-Zeitschrift zu finden.

Die Gefahr durch Referer-SPAM ist jedoch für die Allgemeinheit nicht geringer als wie beim bekannten E-MAIL-SPAM: Die Ergebnisse von Suchmaschinen werden immer mehr durch die Methoden der „*Suchmaschinen-Optimierer*“ beeinflusst. Bei den meisten Suchbegriffen, die eine nahe Verbindung zu einem kommerziellen Umfeld haben, finden sich auf den ersten Plätzen der Ergebnislisten meist nur noch so beworbene Webauftritte; Nicht jedoch solche Seiten, die eine höhere inhaltliche Relevanz haben und die man als Suchender deswegen eigentlich gern auf den ersten Plätzen sehen würde.

Für den Betreiber eines Webauftritts und dessen Dienstleister ergeben sich weitere, nicht minder negative Folgen:
Durch die massenhaften Zugriffe durch Spambots wird der Webserver und die Netzanbindung stark belastet. Zudem werden die Statistiken zu den Zugriffen auf die Webauftritte stark verfälscht und sogar bei der Einzelstatistik der Referer vollkommen unbrauchbar gemacht.
Daneben kann die Netzanbindung und der Webserver so stark belastet werden, dass legitime Zugriffe unter längeren Wartezeiten zu leiden haben.

2 Beispiel einer realen Attacke

Zwischen dem 22. und 24. Juli kam es auf eine Seite eines Webauftritts der FAU zu 198.032 Zugriffen. Hierdurch wurde ein Netzverkehr von insgesamt 31.119.306 kB (29 GB) verursacht. (Bei einem Webauftritt, der bei einem kommerziellen Provider gehostet gewesen wäre, hätte dies übrigens bei den heute marktüblichen Preisen zu Folgekosten von ca. 120 Euro **für den Betreiber des Webauftritts** geführt.)

Bei dieser Referer-SPAM-Attacke wurden 10 verschiedene Attacken gefahren, bei denen mit zwischen 18600 bis 20500 Zugriffen jeweils eine andere kommerzielle Seite beworben wurde. Dies war in der Statistik dann auch entsprechend erfolgreich. Die ersten 10 Plätze in der Einzelstatistik der Referer wurden durch die beworbenen Links belegt.
Andere beworbene Referer und echte „Referer"-Angaben folgten erst danach.

Der Angriff erfolgt im beobachteten Fall von einem Dialin-Provider in New York. Eine strafrechtliche Verfolgung war somit nicht möglich. Zudem zeigten spätere Angriffe, dass hier, wie auch bei E-Mail-SPAM, der Provider selbst nur ein Opfer war, der zum Zeitpunkt der Attacke dummerweise offene Zugänge anbot: Bei Attacken an den folgenden Tagen wurden jeweils andere Provider genutzt.

3 Maßnahmen gegen Referer-SPAM

Selbst bei nicht vorhandener oder nicht öffentlicher Statistik werden die Angriffe unternommen. Wie beim E-Mail-SPAM werden auch hier beliebige offene Rechner oder Provider missbraucht. Für den Angreifer kostet dies in beiden Fällen nichts; Durch Automatisierung nicht einmal Zeit.
Die Webstatistiken zu verbergen ist daher keine effektive Lösung.

Eine properitäre Möglichkeit bietet die Suchmaschinen Google den Betreibern von Webauftritten an: Google bietet die Möglichkeit eines „NOFollow"-Attributs an, mit der Links deklariert werden sollen, die nicht verfolgt werden sollen.
Wird ein Link um diese Angabe ergänzt, wird die Suchmaschine Google bei der Analyse der Seite diesen nicht berücksichtigen oder die mit einem solchen Attribut versehene Seite sogar negativ bewerten.
Leider würde eine automatisierte Setzung dieses Attributs auch andere Links betreffen, bei denen der Betreiber möglicherweise eben eine Berücksichtigung erhofft.

Andere Lösungen gegen Referer-SPAM verwenden ähnliche Verfahren wie bei E-MAIL-SPAM: Zugriffe, bei denen im Referer eine bereits bekannte beworbene Website steht, werden geblockt oder umgeleitet. Ebenso kann verfahren werden, wenn der Zugriff von einem Rechner erfolgt, der bereits mehrfach durch Referer-SPAM-Attacken auffiel.

Um den Webserver anzuweisen, bei bekannten Referer zu reagieren, können Rewrite-Anweisungen in der Datei .htaccess benutzt werden.

Beispiel:

```
RewriteEngine On
RewriteCond %{HTTP_REFERER} ^http://(www\.)?.*adult(-|.).*$ [OR]
RewriteCond %{HTTP_REFERER} ^http://(www\.)?.*anal(-|.).*$ [OR]
RewriteCond %{HTTP_REFERER} ^http://(www\.)?.*tits(-|.).*$ [NC]
RewriteCond %{HTTP_REFERER} ^(.+)$ [NC]

RewriteRule ^(.*)$ %1 [R=301,L]
```

Erklärung:
Zunächst wird die RewriteEngine eingeschaltet. Danach werden mit den Befehl „RewriteCond“ die Bedingungen definiert. Daran abschließend folgt der eigentliche Befehl mit „RewriteRule“.
In diesem Fall wird, wenn in der Referer eine Zeichenfolge wie „adult“, „anal“ oder „tits“ vorkommen der Befehl ausgelöst: Eine Umleitung zu der im Referer stehenden Adresse. Anders gesagt: Wenn jemand Referer-SPAM mit den obigen Domains durchführt, wird er auf die von ihm beworbene Site umgelenkt.

Will man den Zugriff einfach nur blocken und keinen zusätzlichen Traffik verursachen, kann man auch folgende Anweisung nutzen:

```
RewriteRule .* - [F,L]
```

In diesem Fall wird jeder Zugriff lediglich mit einer “403 – Forbidden” Meldung beantwortet.

Beide Lösungen bergen jedoch eine nicht geringe Gefahr für unbeteiligte Dritte:

- Erfolgt der Angriff von einem Rechner oder ein Provider, der selbst nur missbraucht wurde, werden bei einer Blockierung auch echte Zugriffe geblockt.
- Die Referer könnte anstelle für Werbung gezielt dazu missbraucht werden, um Webauftritte für die ansonsten nicht geworben wird, zu diskreditieren. Noch schlimmer: In Verbindung mit einem gesetzten NOFollow-Attribut könnte es so möglich sein, dafür zu sorgen, dass konkurrierende Seiten von den ersten Plätzen der Suchmaschine ausgeschlossen werden würden oder nach unten rückten. Anders ausgedrückt: Es wird **so aggressiv für** ein Produkt der Konkurrenz geworben, dass die potentiellen Kunden es negativ auffassen und sich deswegen entschließen das beworbene Produkt zu ignorieren um anstelle dessen das Produkt einer anderen Firma zu nehmen.

Neben diesen Problemen stellt sich aber auch die Frage des Aufwands für die Verwaltung und Pflege an bekannten Referer und Hosts.

Eine mögliche Lösung, welche ohne die oben beschriebenen Nachteile auskommt, wäre theoretisch denkbar:
Es wäre möglich, im Augenblick des Zugriffes auf eine lokale Seite die Referer auszulesen und zu testen. Wenn die Adresse im Referer nicht bereits in den letzten Minuten abgefragt wurde, wird diese automatisiert geladen und analysiert. Ist dann kein Link zur lokal aufgerufenen Seite zu finden, kann von Referer-SPAM ausgegangen werden, der Zugriff blockiert werden und der verwendete Rechner für eine gewisse Zeitlang abgewiesen werden. Dieses Verfahren hätte den Vorteil, dass Referer-SPAM leicht und relativ sicher zu erkennen wäre und es vollkommen automatisch funktionieren könnte.

Der Nachteil jedoch läge darin, dass die Performance des Webservers für normale Benutzer merkbar sinken könnte. Außerdem wäre auch hier eine absichtliche Diskreditierung anderer Webauftritte oder die gezielte Blockierung von Rechnern möglich.

4 Fazit

Eine verlässliche Methode gegen Referer-SPAM ist Seitens der Betreiber von Webauftritten nur schwer zu erreichen. Zwar ist es mit obigen Rezepten möglich, sich gegen die derzeit üblichen Attacken zu sichern, jedoch ist damit zu rechnen, dass die Verursachen von Referer-SPAM früher oder später auch auf diese effektiv reagieren.
Eine bessere Lösung der Problematik wäre jedoch dann zu erreichen, wenn die Suchmaschinen, die ja auch das eigentliche Ziel sind, geeignete Maßnahmen treffen, so dass Referer-SPAM keine Wirkung mehr zeigt. Ansätze hierzu wären die bessere Analyse der echten Verlinkung der bei den Suchmaschinen bereits vorhandenen Seiten, sowie die Berücksichtigung des Zeitraums einer möglichen massenhaften Verlinkung.
Des weiteren ist es notwendig, dass die negativen Folgen der sogenannten „*Suchmaschinen-Optimierung*“ viel mehr ins Bewusstsein gebracht werden.
„*Suchmaschinen-Optimierung*“ sorgt stark dafür, dass solche Webauftritte eine gute Position bei Suchergebnissen haben, die hierfür am meisten Geld bezahlt haben. Webauftritte dagegen, die Zeit und Geld eher in Qualität der Inhalte setzen bleiben auf der Strecke. Benutzer von Suchmaschinen werden gezielt dahin geleitet wo andere sie haben wollen, nicht aber dorthin von die Benutzer selbst es wollten.

5 Begriffe

Ein **Referer** ist die Internetadresse der Webseite, von der der Benutzer durch Anklicken eines Links zu der aktuellen Seite gekommen ist (to refer = verweisen). Der Referer ist ein Teil der an den Webserver geschickten HTTP-Anfrage. Das RFC 2616 (Hypertext Transfer Protocol HTTP/1.1) erklärt den technischen Hintergrund.
Unter einem **Bot** (angelehnt an *robot*) versteht man eine Klasse von Computerprogrammen, die weitgehend autonom solchen Aufgaben nachgehen, mit denen eine menschlich-interaktiv gesteuerte Software zeit- oder mengenmäßig überfordert wäre.

6 Weitere Informationen zum Thema:

- Wikipedia-Artikel zum Suchmaschinen-Spamming
  http://de.wikipedia.org/wiki/Suchmaschinen-Spamming
- Wikipedia-Artikel zum Thema Referer
  http://de.wikipedia.org/wiki/Referer
- C’t SEO-Wettbewerb „Jagt frei auf die Hommingberger Gepardenforelle“
  http://www.heise.de/ct/SEO-Wettbewerb/
- Beispiel einer längeren htaccess-Anweisung zum Blocken von Referer-SPAM bei Basquiat.de:
  http://www.basquiat.de/exit.php?url_id=4938&entry_id=188