

Aktuelle Rechtsfragen von IT und Internet  
DGRI-Herbstakademie 2005

Call for Papers

Themenkomplex: Verbraucherschutz im E-Commerce

Abstract

Tobias Eggendorfer

**Rechtliche Konsequenzen aus dem Einsatz von Teergruben zur  
Blockade von Harvestern**

# 1. Harvester

Die Versender von unerwünschten Werbe-E-Mails sammeln die E-Mail-Adressen ihrer Opfer mit speziellen Programmen automatisch von Webseiten, wie Studien belegen<sup>1</sup>. Diese Programme nennt man „Harvester“.

Dabei arbeiten Harvester prinzipiell nach dem gleichen Verfahren wie die sogenannten Spider oder Bots<sup>2</sup> von Suchmaschinen: Ausgehend von einer Startseite werden sämtliche Links auf der Seite verfolgt und relevante Seiteninhalte ausgelesen. Harvester suchen dabei mit Regular Expressions<sup>3</sup> nach Zeichenmustern, die mit „mailto:“<sup>4</sup> beginnen, das einen Link auf eine Mailadresse einleitet, oder nach Zeichenmustern, die ein @-Symbol und mindestens einen Punkt enthalten.

Mit den typischer Weise auf Unix-Systemen verfügbaren Befehlen *wget*, *sed*, *tr*, *sort* und *uniq*<sup>5</sup> läßt sich ein Minimalharvester in zwei Befehlszeilen<sup>6</sup> schreiben, der allerdings sehr zurückhaltend sammelt. Da *wget* jedoch Open-Source ist, kann dieser Minimalharvester jederzeit angepaßt und damit effektiver werden.

Auch Programmiersprachen wie Perl<sup>7</sup> oder PHP<sup>8</sup> stellen notwendige Grundfunktionen ebenfalls zur Verfügung und ermöglichen damit auch eine schnelle Entwicklung ausgereifterer Harvester, die programmiertechnisch wenig anspruchsvoll ist.

## 2. Teergrube

Damit ergibt sich neben dem Ansatz, E-Mail-Adressen für Harvester unlesbar auf Webseiten anzugeben, die Idee, den Harvester selbst oder die gesammelten Adressen durch sogenannte Teergruben nutzlos zu machen.

### 2.1. Begriff der Teergrube

Dabei ist eine Teergrube ein Programm, das Verbindungsversuche maximal verlangsamt beantwortet und damit den Lauf des verbindenden Programmes deutlich verzögert, es bildlich gesprochen wie auf warmen Teer ankleben läßt. Zudem reduziert jede blockierte Verbindung eines Harvesters oder auch Bulkmailers, eines Programmes zum massenhaften Versand von E-Mails, die Möglichkeit dieser Software, zeitgleich weitere Verbindungen zu öffnen, da jeder Rechner nur über ein – wenn auch großes – beschränktes Kontingent von Ports für ausgehende Verbindungen verfügt. Die notwendigen Verzögerungen lassen sich technisch mit verschiedenen Verfahren realisieren.

---

1 Siehe dazu mit weiteren Quellen [CDT03], [EGG05a], [EGG05b]

2 Abgeleitet von „Robot“

3 Siehe dazu als praktische Einführung [FRI02], als theoretische Fundierung z.B. [AST02], [HOP02]

4 Siehe dazu z.B. [MÜN02]

5 Erläuterungen zu den genannten Befehlen z.B. in [WIE99], [BARR04] und [TANS00]

6 Siehe dazu u. a.: [EGG05a]

7 Siehe dazu u. a.: [HAJ98], [SIE00] oder auch <http://www.perl.com>

8 Siehe [KRA00] oder auch <http://www.php.net>

## **2.2. Teergruben gegen Harvester**

So habe ich eine Teergrube vorgestellt<sup>9</sup>, die Harvester unmittelbar beim Sammeln von E-Mail-Adressen auf Webseiten ausbremsen soll. Diese Teergrube habe ich später verfeinert<sup>10</sup> und im praktischen Einsatz mit gutem Erfolg auf ihre Wirksamkeit getestet<sup>11</sup>.

Die Teergrube funktioniert dabei nach einem einfachen Prinzip: Sie gibt eine zufällig generierte Liste von Links aus, von denen anzunehmen ist, daß ein Spider jeden einzelnen verfolgt. Diese Ausgabe erfolgt massiv verzögert – in meiner Testteergrube benötigte eine Seite mit 20 Links ca. 16 Sekunden zur Ausgabe. Da jede Seite die gleiche Zahl an neuen Links enthält, wächst die Liste der noch zu besuchenden Links des Harvesters exponentiell und enthält einen laufend wachsenden Anteil von Links auf die Teergrube selbst. Der Harvester wird so in einer Endlosschleife gefangen und effektiv daran gehindert, andere Seiten nach Mailadressen zu durchsuchen.

Allerdings würden so auch Spider von Suchmaschinen, wie zum Beispiel der GoogleBot, blockiert. Das könnte eine Computersabotage (§303b Abs. 1 Alt. 2 StGB) darstellen. Allerdings gibt es den sogenannten robots.txt-Standard<sup>12</sup>, der es ermöglicht, Spidern und theoretisch auch Harvestern zu verbieten, bestimmte Teile einer Webseite zu besuchen.

Dazu wird eine Datei mit dem Name robots.txt im Hauptverzeichnis des Webservers abgelegt, in der in einem maschinenlesbaren, standardisierten Format angegeben wird, welche Robots welche Seitenteile besuchen oder auch nicht besuchen können. Die Spider von Suchmaschinen werten diese Datei üblicher Weise aus, Harvester nicht – denn das würde ihre Ausbeute an E-Mail-Adressen voraussichtlich deutlich reduzieren.

Damit stellt sich die Frage, ob die Betreiber von Spidern, die sich nicht an den de-facto-Standard der robots.txt halten und in die Teergrube fallen, Unterlassungs- und Beseitigungsansprüche aus §§1004 analog, 823 Abs. 1 BGB i. V. m. dem Recht auf den eingerichteten und ausgeübten Gewerbebetrieb bzw. §§1004 analog, 823 Abs. 2 BGB i. V. m. §303b Abs. 1 Alt. 2 StGB als Schutzgesetz gegen den Betreiber einer Teerfalle geltend machen können und ob ein Strafverfahren zu einer Verurteilung führen könnte, oder ob durch den standardisierten Hinweis mittels der robots.txt hinreichend vor der drohenden Gefahr gewarnt wurde.

## **2.3. Teergruben gegen Bulkmailer**

Andere Autoren<sup>13</sup> schlagen vor, die von Harvestern gesammelten Adressbestände gezielt zu verseuchen. Sie publizieren dazu auf Webseiten E-Mail-Adressen, deren zugehöriger Mailserver eine SMTP-Teergrube ist.

Diese nimmt eingehende Mailverbindungen deutlich verlangsamt entgegen und versucht so, das

---

9 In [EGG04a]

10 Siehe: [EGG05a]

11 Ebenfalls: [EGG05a]

12 Siehe z.B. [W3CAPPB] und [HEM03]

13 So z.B. [DON04], [DON04a] und [REHWWWa]

zum Massenversand von Spam verwendete Programm auszubremsen.

Auch hier stellt sich die Frage, ob Computersabotage (§303b Abs. 1 Alt. 2 StGB) vorliegen könnte. Die Betreiber solcher Teerfallen verneinen das, denn nach ihrer Ansicht könne nur ein Harvester in den Besitz dieser Adressen gelangt sein und sie träfen somit nur Spammer, jedoch keine legitimen Absender. Zudem würde die Verzögerung Versender von einzelnen Mails kaum treffen, sondern nur Absender, die an zahlreiche Adressen unter diesen Domains ihre Werbung schicken würden, da im Einzelfall die Verzögerung problemlos zu tolerieren sei.

Sieht man davon ab, daß wohl kaum ein Spammer seine häufig aufwendig konstruierte Tarnung aufgäbe, um gegen einen Spam-Gegner vorzugehen<sup>14</sup>, wirft dieses Vorgehen auch die Frage auf, inwieweit nicht unbeteiligte Dritte, deren Infrastruktur, zum Beispiel in Form von Open Proxys, von Spammern zum Versand von Massenmails mißbraucht wird und solchermaßen lahmgelegt wird, ihrerseits Ansprüche gegen den Betreiber einer SMTP-Teergrube geltend machen könnten.

---

14 Gegenbeispiele nennt z.B. [McW05]

## 3. Literatur

### Literaturverzeichnis

- [AST02] Asteroth, Alexander; Baier, Christel, Theoretische Informatik, Pearson Studium, München, 2002
- [BARR04] Barret, Daniel J., Wilhelm, Torsten (Übersetzer), Linux. Kurz & Gut., O'Reilly, Köln, 2004
- [CDT03] Center for Democracy and Technology, Why am I getting all this spam?, 2003, <http://www.cdt.org/speech/spam/030319spamreport.pdf>
- [DON04] Donnerhacke, Lutz, Teergruben FAQ, 2004, <http://www.iks-jena.de/mitarb/lutz/usenet/teergrube.html>
- [DON04a] Donnerhacke, Lutz, Teergrubing Wrapper, 2004, <http://www.iks-jena.de/mitarb/lutz/usenet/antispam.html>
- [EGG04a] Eggendorfer, Tobias, Ernte - nein danke. E-Mail-Adressenjägern auf Webseiten eine Falle stellen in: Linux Magazin, Linux New Media, München, 2004
- [EGG05a] Eggendorfer, Tobias, Methoden der präventiven Spambekämpfung im Internet, Diplomarbeit, Fernuniversität in Hagen, 2005
- [EGG05b] Eggendorfer, Tobias, Möglichkeiten der Prävention der Zusendung unerwünschter Werbe-E-Mails in: Lecture Notes in Informatik, Köllen Verlag, Bonn, 2005
- [FRI02] Friedl, Jeffrey E. F., Mastering Regular Expressions, O'Reilly, Sebastopol, 2002
- [HAJ98] Hajji, Farid, Perl: Einführung, Anwendungen, Referenz, Addison-Wesley, Bonn, 1998
- [HEM03] Hemenway, Kevin, Calishain, Tara, Spidering Hacks. 100 Industrial-Strength Tips & Tools, O'Reilly, Sebastopol, 2003
- [HOP02] John E. Hopcroft, Rajeev Motwani, Jeffrey D. Ullman, Einführung in die Automatentheorie, Formale Sprachen und Komplexitätstheorie, Pearson Studium, München, 2002
- [KRA00] Krause, Jörg, PHP. Grundlagen und Lösungen., Carl Hanser Verlag, München, 200
- [McW05] McWilliams, Brian, Spam Kings. The Real Story Behind the High-Rolling Hucksters pushing porn, pills, and @\*#?% Enlargements, O'Reilly, Sebastopol, 2005
- [MÜN02] Münz, Stefan, HTML & Web-Publishing Handbuch (Band 1). HTML - JavaScript - CSS - DHTML, Franzis, Poing, 2002
- [REHWWa] Rehbein, Daniel, Gift für Harvester - Aus meiner Entwicklung, 2003, <http://www.daniel-rehbein.de/spamgift.html>
- [SIE00] Siever, Ellen; Spainhour, Stephen; Patwardhan, Nathan, Perl in a Nutshell, O'Reilly, Köln, 2000
- [TANS00] Tansley, David, Linux & Unix Shell Programming, Addison-Wesley, Harlow, 2000
- [W3CAPPB] W3C, W3C Recommendations. Appendix B: Performance, Implementation and Design, o. A., <http://w3.org/TR/REC-html40/appendix/notes.html>

[WIE99]

Wielsch, Michael; Prahm, Jens; Eßer, Hans-Georg, Linux intern. Technik, Administration und Programmierung, Data Becker, Düsseldorf, 1999