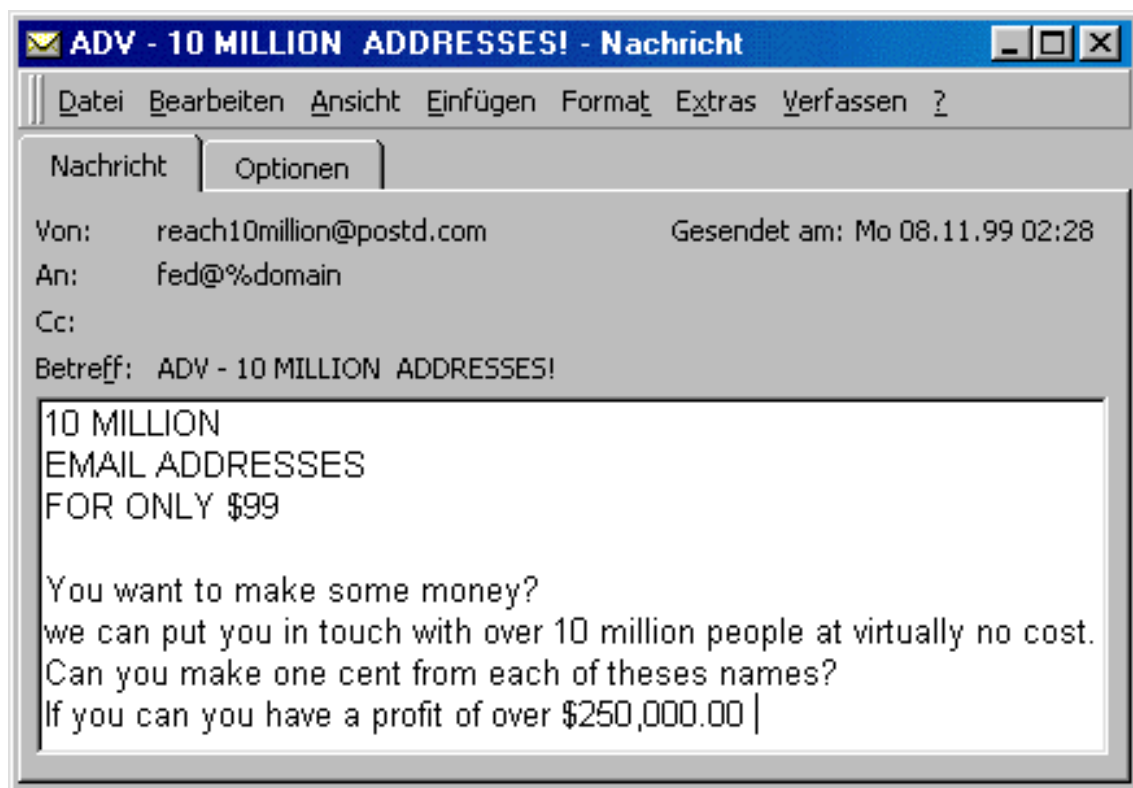


Was tun gegen Email-Grabbing?

1.) Einleitung

Als Email-Grabbing (oder auch Harvesting "Ernte von Email-Adressen") bezeichnet man den Vorgang, wenn Webcrawler (Roboter, Spider) *unaufgefordert* das Internet nach Email-Adressen durchsuchen und bei Erfolg diese in einer Datenbank speichern. Diese Datenbanken werden dann von den Betreibern für Werbezwecke angeboten und verkauft und sind somit Quelle von unerwünschten Werbe-Emails (Spam, Junk) wie z.B. der folgenden:



Ein weiteres Beispiel ist z.B. <http://www.extractor-pro.com>, die bereits beim Laden der Startseite ein JavaScript Fenster öffnen, in dem zu lesen steht:

"Order ANY of our bulk Email Programms and you will receive 25 MILLION Fresh Email Addresses on the web for FREE ..."

Es wird hier natürlich nicht erwähnt, daß diese 25 Millionen "frischen" Email-Adressen gegen den Willen der Besitzer eingesammelt wurden.

Wie aber kann der Konflikt gelöst werden, daß Sie einerseits in Ihrer Homepage Ihre Email-Adresse veröffentlichen möchten, jedoch andererseits diese vor Grabbern schützen können?

Man kann in der Tat seine Email-Adresse durch geschickte Programmierung der HTML-Seiten vor den Grabber-Roboter schützen. Dabei gibt es mehrere Methoden, die wir Ihnen im Folgenden vorstellen möchten. Letztlich liegt es auch an den technischen Möglichkeiten Ihres Providers oder Ihres Speicherpaketes, welche Technik Sie nutzen können oder wollen.

2.) Darstellung der Email mit Leerzeichen

Bauen Sie in jedem Mailto-Link Leerstellen zwischen den einzelnen Zeichen ein:

```
<a href="mailto:?Subject=i n f o @ d o m a i n . d e">Email an uns</a>
```

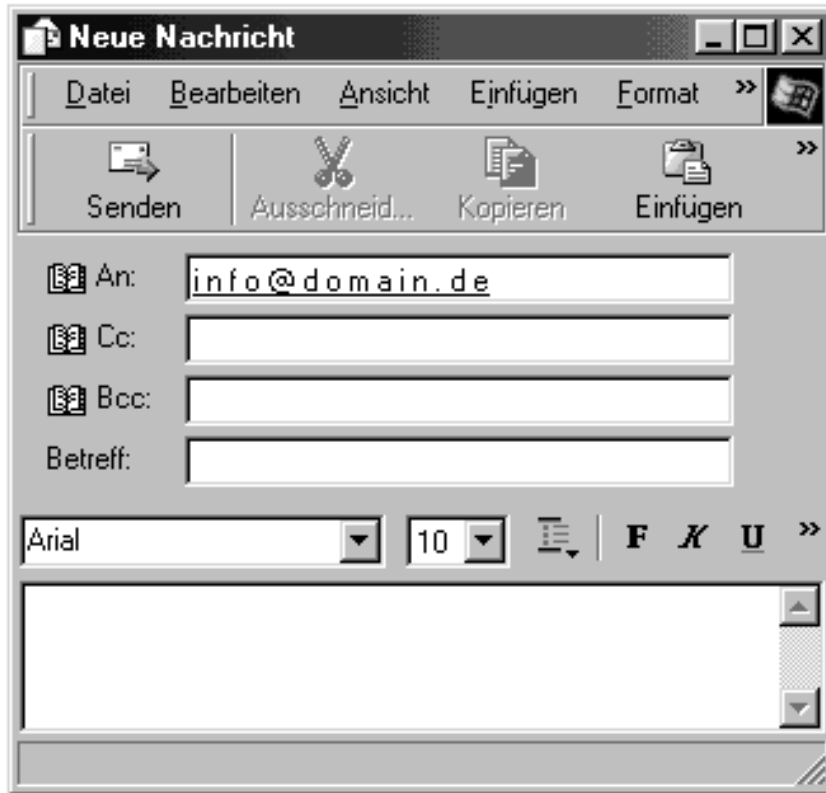
Wir empfehlen Ihnen bei dieser Methode nach dem Doppelpunkt ein "?Subject=" einzufügen, sodass die durch Leerstellen verlängerte Email im *Betreff*-Feld des Fensters erscheint und nicht im *Empfänger*-Feld.

Einige von uns getestete Email-Programme erzeugen sonst folgende Fehler:

- *Netscape Messenger*: Interpretieren jeden Buchstaben als eine neue Email-Adresse und erstellen im Mailfenster für jeden Buchstaben ein neues Empfänger-Feld ("An:") wie im nachfolgenden Bild zu sehen ist.



- *Outlook Express:*
Nach der automatischen Prüfung der Adresse, die trotz Leerzeichen als richtig abgeschlossen wurde (Version 5), wurde die Email unterstrichen sodass die Leerzeichen nicht mehr entfernt werden konnten.



Wenn Sie Ihre Email-Adresse auf diese Weise im Quellcode hinterlegen, sollten Sie jedoch für Ihre Besucher einen Hinweis einbauen, was zu tun ist, um eine funktionierende Email-Adresse zu erhalten, z.B.:

Email an uns
(Spamschutz: Entfernen Sie die Leerzeichen und kopieren Sie die erst dann richtige Adresse in das Empfänger-Feld Ihres Mailprogrammes)

3.) Darstellung der Email mit ASCII

Eine weitere Möglichkeit stellt die Tatsache dar, dass Browser ASCII-Code richtig darstellen, die Grabbing-Roboter diese jedoch nicht interpretieren können:

So wird z.B. aus der Adresse

info@domain.de

der folgende ASCII-Code:

info@domain.de

Wir selbst nutzen diese Möglichkeit, um unsere Email im Footer jeder Seite zu kodieren.

Auf der Seite

<http://www.pro.vider.de/support/workshops/email-grabbing/kodieren2.php>

finden Sie ein Eingabefeld, in dem Sie Ihre Email-Adresse in ASCII-Code umwandeln können, den Sie dann in Ihre Homepage einbauen können (die Idee zu diesem PHP-Script stammt von suchbuch.de).

4.) Darstellung der Email durch JavaScript

Durch ein JavaScript kann eine Email-Adresse ebenfalls kodiert im Quelltext hinterlegt werden. So wird aus der Email-Adresse

info@domain.de wird

das folgende JavaScript:

```
<script language="JavaScript">
  <!--
  var vor_at = "info";
  var nach_at = "domain.de";
  document.write('<a href="mailto:'+vor_at+'@'+nach_at+'">Email</a>');
  // -->
</script>
```

(Ursprung: CDR Software, USA)

5.) Einbindung der Email durch ein PerlScript und SSI

Von der Idee her die wohl beste Methode bietet ein Perl Script von Christoph Ruegg, Siteware Technologies, Schweiz:

Durch die Abfrage der Umgebungsvariablen HTTP_USER_AGENT (wird eine Homepage mit einem normalen Browser aufgerufen, ist diese Variable z.B. "Netscape" oder "IE4.0" etc.) erkennt das Script, ob ein normaler User die Seite aufruft, oder ob ein Grabbing-Roboter die Seite durchsuchen will. Im ersten Fall gibt das Script die richtige Email-Adresse aus, im zweiten Fall jedoch wird die Adresse entweder versteckt oder durch 10.000 falsche Adressen "ersetzt", wodurch eine Verunreinigung der Spammer-Datenbank erzeugt wird.

Im Quellcode wird das Script per SSI-Aufruf eingebaut:

```
<a href="mailto:<!--#exec cgi="/cgi-bin/emp.pl"-->">Email an uns</a>
```

Voraussetzung für die Installation ist ein eigenes cgi-bin Verzeichnis auf Ihrem Server und die Möglichkeit SSI nutzen zu können. Weitere Informationen zu SSI finden Sie unter der URL

<http://www.pro.vider.de/support/ssi>

Das Script ist Freeware kann zusammen mit einer Anleitung zum Konfigurieren und Einbauen unter der URL

<http://www.pro.vider.de/support/workshops/email-grabbing/emp.zip>

downgeloaded werden.

6.) Einbindung der Email durch ein PHP-Script

Vom zuvor besprochenen Perl-Script abgeleitet, kann auch die Scriptsprache PHP herangezogen werden, um dem Browser eine "User_Agent-abhängige" Anzeige auszugeben:

Ist der User_Agent ein Netscape oder IE-Browser wird der Mailto-Link normal (oder z.B. ASCII-codiert) ausgegeben, ist der User_Agent hingegen kein Browser, wird die Adresse versteckt.

Dieses PHP-Fragment kann wie folgt aussehen:

```
<?
if (ereg("Mozilla",$HTTP_USER_AGENT))
{
echo "<a href=\"mailto:info@domain.de\">Email an uns</a>";
}
else
{
echo "Security violation: No Email given!";
}
?>
```

Funktionsweise:

Die Funktion "ereg" durchsucht die Variable "HTTP_User_Agent" nach der Zeichenfolge "Mozilla". Wird diese gefunden wird der Link an den Browser ausgegeben (echo-Befehl).

Ist dies jedoch nicht der Fall, wird die HTML-Seite also von etwas anderem angefordert als einem Browser, wird der "else"-Zweig aktiv, d.h. die Email wird nicht ausgegeben.

Da sich auch Internet-Explorer und Opera mit "Mozilla" melden, reicht die Abfrage nach diesem Namen aus. Nachfolgend einige Browser und Ihre beispielhaften HTTP_USER_AGENT Ausgaben:

```
Netscape 3: Mozilla/3.01 (Win95; I)
Netscape 4: Mozilla/4.7 [de] (Win98; I)
IE 5: Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)
Opera 3.6: Mozilla/4.0 (Windows 3.95;DE) Opera 3.60 [de]
```

Anhang:

Über die pro.vider.de Internetagentur GmbH:

Die Firma wurde im März 2000 von den Brüdern Jürgen und Stefan Müller gegründet und ging aus der Firma "Internet Solutions Jürgen Müller" hervor, die seit 1997 Dienstleistungen im Bereich Internet angeboten hatte.

Zum Leistungsspektrum gehört das Anbieten und die Durchführung von Internet- und Multimediadienstleistungen, insbesondere der Konzeption, Erstellung, Pflege und Marketing von Internet-Präsentationen ("WebDesign"), Datenbanken, eCommerce und dynamische Programmierungen. Weiterhin die Vermietung und Administration von Internet-Servern ("WebSpace") und die Konnektierung von Domains.

Über den Autor Jürgen Müller:

- Studium der Betriebswirtschaft in Johannesburg / Südafrika (nicht beendet ;-)
- Studium der Physikalischen Technik an der FH Lübeck, Dipl.-Ing. (fh)
- Physik-Studium an der University of Bath, England, Master of Philosophy (MPhil)
- Gründung ISJM April 1997
- Gründung pro.vider.de GmbH März 2000



(Jürgen Müller, Geschäftsführender Gesellschafter)

Heidenheim, Juli 2000