Nachlässig programmierte Spambots – eine Anekdote

Julian Fietkau

am 1. Dezember 2011
im KunterBuntenSeminar

Der Anlass

Mein Piwik berichtete mir: ca. 10× pro Monat „Error 404“

- Oh schreck!
- Hmmmm...
- Mal gucken wie es jeweils dazu kommt.

Piwik allein konnte mir das nicht sagen, also begann ich, mit jedem 404er den Referrer zu loggen.

Erkenntnisse

Überraschung: Sämtliche 404er entstanden durch Fragment-Links!

- HTTP GET auf zum Beispiel
  „http://www.julian-fietkau.de/spambots#comments“ → **404**
- WTF?
- Mal schauen, was die IETF dazu sagt...

Was sind Fragment-Links überhaupt?

- Sie verweisen auf einen Teil eines Objekts (z.B. eines Dokuments im Web).
- Im Web werden sie u.A. verwendet, um Browser zu einer bestimmten Stelle auf einer Seite „springen“ zu lassen.

Aus RFC 3986 (URI-Syntax), Kap. 3.5

> (. . .) [T]he fragment identifier is separated from the rest of the URI prior to a dereference, and thus the identifying information within the fragment itself is dereferenced solely by the user agent, regardless of the URI scheme. (. . .) it also serves to prevent information providers from denying reference authors the right to refer to information within a resource selectively.

Siehe `https://tools.ietf.org/html/rfc3986#section-3.5`

Der Client ist schuld

Vielleicht handelt es sich um einen sehr schlechten Browser oder einen besonders dummen Crawler?

- Stellt sich dann so raus, dass die 404er zeitlich nahe an Spam-Kommentar-Einträgen auf meiner Website liegen. D'oh!

User Agents

Nachdem ich wusste, dass sämtliche 404er von Spambots stammten, loggte ich ihre User Agents:

```
Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1)
```

```
Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.5) Gecko/2008120122 Firefox/3.0.5
```

```
Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.10) Gecko/2009042316 Firefox/3.0.10
```

```
Mozilla/5.0 (Windows NT 5.1; U; en) Opera 8.01
```

Was nun?

Ich loggte die IPs der Spammer, sie waren immer verschieden. Vielleicht ein Botnet oder so.

Konsequenzen für mich? Keine Großen. Meine Spam-Erkennung für die Kommentare funktioniert gut genug.

Dumme Spambot-Programmierer leben drei mal hoch, denn sie machen schon die Hälfte meiner Diagnose-Arbeit selbst. :-D

Danke für die Aufmerksamkeit!

http://www.julian-fietkau.de/spambots