

rk7

Geschrieben von Anonymisierung + Statistik - 2010/02/22 03:02

Trotz Anonymisierung will man manchmal ein paar statistische Daten über die Besucher sammeln -- wie z.B. die Anzahl der Besucher oder das Herkunftsland.

Deshalb die Frage, inwieweit die folgenden möglichen Vorgehensweisen mit dem Siegel vereinbar sind, und was ihr dazu meint:

1)

Speicherung nur des letzten Oktetts der IP-Adresse;

z.B. wird statt 85.25.141.154 nur ###.###.###.154 gespeichert.

Die Speicherung dieses IP-Adressen-Teils kann nützlich sein, um die "Visits" abschätzen zu können, bzw. um verschiedene Besucher auseinander halten zu können.

Wäre dies mit dem Siegel vereinbar?

2)

Teilweise ist es für Analyse Zwecke interessant, woher die Nutzer kommen. Eine einfache Möglichkeit dazu ist ein IP-Reverse-Lookup. Um dies zu anonymisieren könnte man hier wieder den vordersten Teil des Hostnamens löschen, und nur den Rest speichern. Die IP wird hierzu lediglich kurzzeitig an ein Skript übergeben aber nie auf der Festplatte gespeichert.

Beispiel:

IP 91.7.239.74 --rDNS--> p5B07EF4A.dip.t-dialin.net;

gespeichert wird dann "#####.dip.t-dialin.net" (oder evtl. nur "#####.t-dialin.net"), die IP-Adresse wird nicht oder nur als ###.###.###.74 gespeichert.

Wäre dies mit dem Siegel vereinbar?

3)

Eine andere Möglichkeit, das Land des Besuchers herauszufinden, ist ein whois-lookup der IP-Adresse. Mit so einem Lookup bekommt man einige Informationen über den "Eigentümer" der IP-Adresse; bei der oben genannten IP wäre das die Deutsche Telekom. Interessante Felder wären hierbei das Land (hier: "DE") und evtl. der Netzname (hier: "DTAG-DIAL22").

3a) Wäre es mit dem Siegel vereinbar, so einen Lookup durchzuführen und nur das Land (z.B. "DE") zu speichern? Die IP-Adresse wird dabei nur als "###.###.###.74" gespeichert.

3b) Wäre es mit dem Siegel vereinbar, zusätzlich den Netznamen (z.B. "DTAG-DIAL22") zu speichern?

4)

Falls (3a) mit dem Siegel vereinbar ist: Solche whois-Lookups dauern meist etwas (z.B. 0.2s); die Anzahl der Lookups könnte man jedoch vermindern, indem ein kleines Skript diese Lookups "cacht".

Dazu müsste der Cache aber die IP-Adressen einige Minuten in einer Variablen halten. Auf der Festplatte gespeichert o.Ä. wird die IP-Adresse dabei weiterhin nicht; um an die IP-Adresse zu kommen müsste man schon während der Laufzeit innerhalb der entsprechenden Zeit den Arbeitsspeicher des Skripts auslesen. Und selbst dann wäre die IP nicht eindeutig einem Seitenaufruf zuordenbar, da bei den Seitenaufrufen nur verkürzte IP-Adressen (a la ###.###.###.74) geloggt werden.

Wäre so ein Cache mit dem Siegel vereinbar?

Re:rk7

Geschrieben von pab - 2010/03/31 20:40

Hallo,

zu 1): Wir verlangen mindestens die Kürzung um zwei Byte.

---

Zu 2): Die Frage ist, ob der Hostname mit deiner Methode trotz Kürzung in einigen Fällen eindeutig identifizierbar bleibt (z.B. wenn ein Hostname eindeutig auf die IP-Adresse einer natürlichen Person schließen lässt). Die Methode dürfte daher nicht zulässig sein.

Zu 3a): Nur das Land zu speichern, ist ok. Nur das letzte Byte der IP-Adresse zu speichern ist ok, wenn sich die ersten drei Bytes nicht aus anderen Informationen konstruieren lassen.

Zu 3b): Der Netname könnte eindeutig auf die IP-Adresse einer natürlichen Person schließen lassen. Daher wäre das nicht ok.

Zu 4): Die temporäre Speicherung vollständiger IP-Adressen ist mit dem Siegel nicht vereinbar.

=====