

---

## Ideen für Intelligentes logging im Webserver

Geschrieben von Hanno - 2007/10/04 00:34

---

Hi,

Als kleiner Webserver-Betreiber diskutieren wir seit ein paar Tagen intensiv über die Auswirkungen des Urteils und die Kampagne. Unser Hauptproblem dabei ist, dass unsere Kunden häufig webalizer oder ähnliche Spalten haben wollen, und deren Aussagekraft gegen Null geht wenn man keine IP-Adressen hat.

Meine Ideallösung würde in etwa in die Richtung gehen: Man bräuchte ein Logging-System im Webserver, welches für jeden neuen IP-Zugriff einen temporären Alias erzeugt (einen zufallswert, 32 Bit oder so sollten genügen, kann auch ne fake-ip sein).

Kommt von der IP noch ein Zugriff, wird der Alias wieder vorgekramt. Kommt ne zeitlang (Stunde oder so) keiner wird der alias gelöscht.

Gewährleistet müsste halt sein, dass die Tabelle ip -> alias nur im RAM existiert.

Imho müsste man sowas als apache-modul coden können. Erfordert aber sicher ein bisschen Fähigkeiten in algorithmen und so, damit das effizient läuft und nicht den server in die knie zwingt.

Thoughts?

=====

## Re:Ideen für Intelligentes logging im Webserver

Geschrieben von reaper - 2007/10/04 14:34

---

Ich habe auch schon darüber nachgedacht, ich setze selbst Webalizer ein und ganz ehrlich ich möchte schon ganz gern wissen wieviele Leute so auf meinen Seiten rumtanzen ;)

Deine Idee ist nicht schlecht aber solche Hashing-Tabellen finde ich nur bedingt sinnvoll. Diese Tabellen sind wieder abgreifbar, wenn sie nur im RAM liegen ist das nur halb so wild aber wer sagt denn das die Tanten vom \*KA den Rechner runterfahren und sich nicht per Bundestrojaner Zugang verschaffen :( Dazu kommt das eine Stunde vielleicht zu knapp ist. Ein RSS Reader kommt vielleicht alle 2 Stunden wieder und eigentlich willst du den ja nicht zweimal loggen... Das ganze ist nicht ganz trivial, wenns um einen selbst geht alles kein Problem aber wenn es um Kunden geht sehe ich mit der Anonymisierung durch mod-removeip echt ein Kommunikationsproblem :(

In einer längeren Diskussion zu diesem Thema kam bei uns die folgende Idee auf:

Das gute alte HTTP Protokoll liefert noch massig Informationen an den Webserver die nirgends gelogged werden die aber den Rechner des Clients recht gut beschreiben. In Verbindung mit der IP Adresse sogar für die Dauer der Verbindung sehr eindeutig. Zumindest eindeutig genug für Statistiken :) Unser Gedanke ist nun der, diese Daten zu nutzen um die IP Adresse zu überlagern. Dazu verknüpft man einfach die IP mit den nicht geloggen Daten und speichert dann die überlagerte Adresse in den Logs. Dort finden sich dann wieder IP Adressen die aber nichts mit der ursprünglichen zu tun haben. Man nutzt also die Geschwindigkeit des Protokolls um die Daten zu anonymisieren. Voraussetzung ist natürlich das die Streudaten nicht mitgelogged werden.

=====

## Re:Ideen für Intelligentes logging im Webserver

Geschrieben von anonym - 2007/10/04 18:40

---

Man könnte zur Berechnung eines eindeutigen Hash-Schlüssels z.B. einen substr von md5sum(IP + aktuelle Stunde) nutzen und den zur Referenzierung benutzen. Dabei muss der substring natürlich deutlich länger sein als der Adressraum von IPv4 es zulässt. Dann könnte man natürlich zwei verschiedene Besucher als einen erkennen, aber das ist auch bei einem kleinen Hash-Schlüssel-Raum eher unwahrscheinlich.

So habe ich meine Logfiles anonymisiert bis ich auf ein Access-Logging komplett ohne IP-Adresse umgestiegen bin. Wenn ich wirklich mal grob meine Visits abschätzen will, dann zahle ich halt die Zugriffe auf mein CSS. ;-)

Das könnte man IMHO (evtl. optional) auch mit der von reaper vorgeschlagenen Überlagerung durch verschiedene HTTP-Werte kombinieren.

---

---

## Re:Ideen für Intelligentes logging im Webserver

Geschrieben von Matthias Mansfeld - 2007/10/04 19:08

---

Nebenbemerkung - Wenn ein Bundestrojaner für Webserver möglich ist, wird der auch das IP-Logging wieder einschalten können. DEN Aspekt kann man in der Diskussion insofern mal fallenlassen...

Und allzuviel Verknüpfung von "Ersatzdaten" statt der IP könnte IMHO zu viel mehr personalisierten Daten führen als die IP. De facto bekomme ich als Webserverbetreiber, egal ob als Hostingkunde oder mit eigenem Server, die Zuordnung IP zu Person nie oder nur über den Staatsanwalt mit Hilfe des Zugangsproviders.

Wenn ich diverse andere Daten (meinetwegen noch Cookies usw.) schlau verwurstle, weiß ich ohne IP mehr über den Surfer, da würde selbst Patrick Breyer die nackte IP aus dem Apache-Log mit Kußhand akzeptieren.

Aus Sicht des Websurfers versteh ich die Intention der Aktion absolut: Mit dem Standard.Apache-Log incl. IP alleine kann mich der normale Serverbetreiber de facto kaum identifizieren, und wenn ich weiß, daß die Webseite meine IP mitschreibt (spätestens wenn ich die Datenschutzerklärung nach TMG gelesen habe), hätte ich gegenüber dem Serverbetreiber kein Problem damit, ich kann ihm vertrauen oder es lassen (soviel Vertrauen muß sein). Aber eine (momentan nur dem Staat mögliche) Verknüpfung von Verkehrsdaten (VDS!! - beim Zugangsprovider) und Serverlogs (Staatsanwalt fragt artig beim Webseitenbetreiber) bringt die Verknüpfung, welche Person wann welche Dateien abgerufen hat, und das will ich als Surfer eigentlich auch nicht.

Insofern wäre ein sauberes Verbot des IP-Loggings ganz ohne Ausnahme sogar ein Schutz für den Seitenbetreiber, weil er achselzuckend und wahrheitsgemäß sagen kann, sorry, ich kann keine Logs haben.

Der konkrete Anlaßfall mit der Justizministeriumswebseite zu den mg ist für mich erst mal ein spektakulärer Einzelfall/Vertrauensbruch - da macht der Webseitenbetreiber etwas, OHNE mich zu informieren (im Gegensatz zu denen, die eine ehrliche Datenschutzerklärung geben) UND der Webseitenbetreiber gleichzeitig selbst die Instanz, die am kurzen Dienstweg die Verkehrsdaten zur Verknüpfung beschaffen kann.....

Aus Sicht des Server- oder Webseitenbetreibers ist das ganze halt ein Riesenproblem: Was ist mit den IMHO genauso legitimen Interessen (Statistik - wie weit ist das Benutzerbespitzelung; Mißbrauchsverhinderung - gibt es praktische Alternativen). Und, wenn er auf die IP gerne verzichten will, hat er auf die Serverkonfiguration den nötigen Einfluß? Die Breyerschen Musterklagen als Damoklesschwert überm Kopf machen uns das Leben nicht leichter, und setzen uns bei der Suche nach sinnvollen Lösungen unnötig unter Druck.

mfG Matthias Mansfeld

---

---

## Re:Ideen für Intelligentes logging im Webserver

Geschrieben von anonym - 2007/10/05 04:41

---

Hab mal mod\_removeip geändert, so dass die IP-Adresse gegen eine zufällig genierte ausgetauscht wird.

Aktuelle Probleme/Bugs:

\* In der Version wird der Hash nie geleert. ;-)

\* Benutzt man eine MPM mit Prozessen hat natürlich jeder Prozess seine eigene Hash-Table.

<http://rafb.net/p/MbmeEa36.html>

Vielleicht kann jemand das noch ausbauen oder ein Projekt draus machen. Makefile kann man von removeip nehmen und einfach anpassen.

---

---

## Re:Ideen für Intelligentes logging im Webserver

Geschrieben von CeBe - 2007/10/08 11:28

---

Ich finde ein Hash aus IP+UserAgent, der für die Statistik bis zu 12 Stunden gespeichert wird ist anonym genug.

---

Niemand kann damit, wie in einem anderen Thread befürchtet wieder die original IP herstellen. Vielmehr wird er für die Statistik sogar genauer, da vllt. in einem Haus mehrere Benutzer über eine IP ins netz gehen oder aus nem Inet-Caffee mehrere Nutzer auf die Seite zugreifen. Sicher wird ein solcher hash, ja dadurch, dass der zum erzeugen eingegebene Text länger als der Hash selbst ist. Dachte hierbei an MD5.

> Hab mal mod\_removeip geändert, so  
> dass die IP-Adresse gegen eine  
> zufällig genierte ausgetauscht  
wird.

Halte ich für wenig sinnvoll, was sollte man mit einer Zufalls-IP wollen? das wäre nur Datenschrott, wovon am Tag beim Access-Log auf ner Großen Seite schon einiges zusammenkommen würde.

Mfg Carsten

---

## Re:Ideen für Intelligentes logging im Webserver

Geschrieben von Christian - 2008/02/24 12:58

anonym schrieb:

Hab mal mod\_removeip geändert, so dass die IP-Adresse gegen eine zufällig genierte ausgetauscht wird.

Aktuelle Probleme/Bugs:

\* In der Version wird der Hash nie geleert. ;-)

\* Benutzt man eine MPM mit Prozessen hat natürlich jeder Prozess seine eigene Hash-Table.

<http://rafb.net/p/MbmeEa36.html>

Vielleicht kann jemand das noch ausbauen oder ein Projekt draus machen. Makefile kann man von removeip nehmen und einfach anpassen.

Ich habe auf meinen Seiten folgendes eingebaut:

Jeder Session wird eine fortlaufende IP zugewiesen (127.x.y.z), diese IP wird ins access-log geschrieben. Benutzer die Cookies blocken, wird über URL-Rewriting eine Session angehängt. Die Lebenszeit der Daten im Speicher ist folglich identisch mit der der Session, ein Rückschluss auf die echte IP des Nutzers ist nicht möglich, da es hier keinerlei Mapping gibt.

Externe Verlinkungen auf Ressourcen meiner Server bei geblockten Cookies dann halt viele, viele neue Sessions, aber damit kann ich leben.

Einziger Haken: Der Filter funktioniert nur in einem Tomcat. :)

---

## Re:Ideen für Intelligentes logging im Webserver

Geschrieben von vfvf - 2010/02/04 05:43

Matthias Mansfeld schrieb:

Nebenbemerkung - Wenn ein Bundestrojaner für Webserver möglich ist, wird der auch das IP-Logging wieder einschalten können. DEN Aspekt kann man in der Diskussion insofern mal fallenlassen...

Und allzuviel Verknüpfung von "Ersatzdaten" statt der IP könnte IMHO zu viel mehr personalisierten Daten führen als die IP. De facto bekomme ich als Webserverbetreiber, egal ob als Hostingkunde oder mit eigenem Server, die Zuordnung IP zu Person nie oder nur über den Staatsanwalt mit Hilfe des Zugangsproviders.

Wenn ich diverse andere Daten (meinetwegen noch Cookies usw.) schlau verwurstle, weiß ich ohne IP mehr über den Surfer, da würde selbst Patrick Breyer die nackte IP aus dem Apache-Log mit Kußhand akzeptieren.

Aus Sicht des Websurfers versteh ich die Intention der Aktion absolut: Mit dem Standard.Apache-Log incl. IP alleine kann mich der normale Serverbetreiber de facto kaum identifizieren, und wenn ich weiß, daß die Webseite meine IP mitschreibt (spätestens wenn ich die Datenschutzerklärung nach TMG gelesen habe), hätte ich gegenüber dem Serverbetreiber kein Problem damit, ich kann ihm vertrauen oder es lassen (soviel Vertrauen muß sein). Aber eine (momentan nur dem Staat mögliche) Verknüpfung von Verkehrsdaten (VDS!! - beim Zugangsprovider) und Serverlogs

---

(Staatsanwalt fragt artig beim Webseitenbetreiber) bringt die Verknüpfung, welche Person wann welche Dateien abgerufen hat, und das will ich als Surfer eigentlich auch nicht.

Insofern wäre ein sauberes Verbot des IP-Loggings ganz ohne Ausnahme sogar ein Schutz für den Seitenbetreiber, weil er achselzuckend und wahrheitsgemäß sagen kann, sorry, ich kann keine Logs haben.

Der konkrete Anlaßfall mit der Justizministeriumswebseite zu den mg ist für mich erst mal ein spektakulärer Einzelfall/Vertrauensbruch - da macht der Webseitenbetreiber etwas, OHNE mich zu informieren (im Gegensatz zu denen, die eine ehrliche Datenschutzerklärung geben) UND der Webseitenbetreiber gleichzeitig selbst die Instanz, die am kurzen Dienstweg die Verkehrsdaten zur Verknüpfung beschaffen kann.....

Aus Sicht des Server- oder Webseitenbetreibers ist das ganze halt ein Riesenproblem: Was ist mit den IMHO genauso legitimen Interessen (Statistik - wie weit ist das Benutzerbespitzelung; Mißbrauchsverhinderung - gibt es praktische Alternativen). Und, wenn er auf die IP gerne verzichten will, hat er auf die Serverkonfiguration den nötigen Einfluß? Die Breyerschen Musterklagen als Damoklesschwert überm Kopf machen uns das Leben nicht leichter, und setzen uns bei der Suche nach sinnvollen Lösungen unnötig unter Druck.

mfG Matthias Mansfeld

=====