Der Landesbeauftragte für den
DATENSCHUTZ und die
INFORMATIONSFREIHEIT
Rheinland-Pfalz

Hinweise für die Behandlung von Webserverlogdateie

Häufig möchten die Anbieter von Internetseiten das Zugriffsverhalten der Nutzer analysieren. Hierbei können sich jedoch datenschutzrechtliche Bedenken ergeben, wenn die Logdateien der Webserver personenbezogene Daten enthalten und für Analysezwecke die Weitergabe an Dritte vorgesehen ist. Üblicherweise enthalten Log-Dateien der Webserver verschiedene Informationen, bei denen ein unmittelbarer Personenbezug besteht oder durch geeigneten Aufwand hergestellt werden kann.

Folgende Angaben bezüglich eines Zugriffs auf das Angebot des Webservers werden zumeist protokolliert:

- Adresse des aufrufenden Hosts
- Authentifizierungsfelder
- Datum und Uhrzeit des Zugriffs, sowie Zeitzonendifferenz zum UTCNormal
- Zugriffsmethode und Inhalt des HTML-Aufrufs
- Statuscode des Webservers zum Aufruf
- Anzahl der vom Webserver übertragenen Bytes
- Referenzadresse der aufrufenden Seite
- Angaben über den eingesetzten Browser und das Betriebssystem des Clients

Personenbezogene bzw. –beziehbare Daten können sich aus den Informationen der Felder 1, 2, 4 und 7 ergeben. Da Log-Dateien im Regelfall als so genannte CLF-Datei (common log format) vorliegen, ist jedoch eine hinreichende Anonymisierung der Informationen mit vertretbarem Aufwand automatisiert möglich.

Aus Sicht des Landesbeauftragten für den Datenschutz sind insbesondere folgende Maßnahmen vorzusehen:

1. Hostadresse: Da bei Hostadressen grundsätzlich Personenbeziehbarkeit gegeben ist und diese Personenbeziehbarkeit in Einzelfällen mit vergleichsweise geringem Aufwand zu einem direkten Personenbezug führen kann, sollte generell die Hostadresse um den Teil gekürzt werden, der einen unmittelbaren Bezug auf aufrufenden Rechner liefert. Im Falle der nummerischen Notation von IP-Adressen wäre dies – je nach vorliegender Netzklasse – der Hostteil der Adresse (bei ClassA-Netzen die letzten 3 Bytes, bei Class-B-Netzen die letzten beiden Bytes und bei Class-C-Netzen das letzte Byte der IP-Adresse). Da bei Class-A- und Class-BNetzen im Regelfall ein Subnetting vorliegt, d.h. der Hostanteil der Adresse zur teilweisen Adressierung von „Unternetzen“ dient, wäre auch in diesen Fällen ein Verzicht auf das letzte Adressbyte ausreichend, um die Adressangabe hinreichend zu verallgemeinern, so dass ein Personenbezug nicht mehr hergestellt werden kann. In der Praxis wäre dies durch Substitution des letzten Adressbytes durch ein Null-Byte realisierbar. (Beispiel: aus „172.31.47.11“ würde „172.31.47.0“)

2. Bei der DNS-Notation, d.h. der Adressangabe als alphanumerischer Wert, wäre durch Reduktion der Adresse auf Top-Level-Domain und Domain, also durch Abtrennen von eventuellen Subdomains und Hostnamen (vorderer Teil der DNSAdresse), die Anonymisierung ausreichend möglich. (Beispiel: aus „pc_mayer.sales.mydomain.com“ würde „mydomain.com“)
3. Die Authentifizierungsfelder enthalten bei Zugriffen auf geschützte Bereiche des Webangebotes die Benutzerkennungen der aufrufenden Person. Log-Einträge, die auf den Aufruf solcher besonders geschützter Informationen hinweisen, sollten daher in jedem Falle ausgefiltert werden.
4. Die Protokollierung der Zugriffsmethode sowie der aufgerufene HTML-Kontext können, insbesondere im Falle von Formulareingaben, personenbezogene Daten enthalten. Je nach Konfiguration des Webservers und der verwendeten Zugriffsmethode erfolgt hier unter Umständen eine inhaltliche Protokollierung der Formulardaten. Es sollte daher generell eine Filterung dieses Log-Feldes dahingehend erfolgen, dass Formulardaten vollständig entfernt werden. In der Praxis kann dies durch Verkürzen des Feldinhaltes auf den vorderen Teil (Verzeichnis- und Dateiname des Scriptes) erreicht werden. (Beispiel: aus „GET /cgi-bin/search.pl?name=Schmidt&vorname=Paul“ würde „GET /cgi-bin/search.pl“)
5. Im Referenzfeld der Log-Dateien wird der vollständige Universe Resource Identifier (URI) der die Website aufrufenden Stelle angezeigt. Hier gelten die gleichen Aussagen wie unter Pos. 4 genannt. Sofern die vorgenannten Maßnahmen zur Bereinigung der Log-Dateien vorgenommen werden, sind evtl. vorhandene personenbezogene Daten ausreichend anonymisiert. In diesem Fall stehen einer Weitergabe der Log-Dateien keine datenschutzrechtlichen Bedenken entgegen.