Fingerabdrücke im Internet

Juri Schulte[1]

28. Juni 2013

Zusammenfassung

Informationen sind die wichtigsten Ressourcen der neuen Medien. Können verschiedenste Informationen in einen festen Kontext gebracht werden, steigert dies den Wissensgewinn teils erheblich. Geschieht eine solche Zusammenführung von Informationen mit personenbezogenen Daten, können Verletzungen der Privatsphäre die Folge sein. In dieser Ausarbeitung werden mögliche Profilbildungsansätze im Rahmen von HTTP[2] erörtert und einige Schutzmaßnahmen vorgestellt und bewertet.

1 Einleitung

Für den Aufruf einer Webseite sind heutzutage nur wenige Mausklicks nötig. Falls die geforderte Webseite nicht bereits beim Start des Browsers[3] ausgewählt werden kann, reicht die Eingabe weniger Zeichen in die Adressleiste, um die gewünschten Webressourcen auf den Bildschirm zu holen. Damit wird allerdings nur ein Bruchteil der Informationen gezeigt die der Browser übermittelt. Neben der Adresse der gewünschten Webseite werden zahlreiche Parameter versendet und empfangen, die unter Anderem die Verbindung aufrecht erhalten und die Qualität der Ressourcen beeinflussen. Die teils große Anzahl an Parametern mit höchst divergenter Belegung kann so einzigartig sein, dass sie zur Identifikation eines Besuchers, auch als Profilbildung bezeichnet, verwendet werden kann.
In Kapitel 2 werden beispielhaft drei Identifikationsmöglichkeiten beschrieben. Da dies nur Beispiele sind, werden in Kapitel 3 weitere Möglichkeiten angeschnitten und den beschriebenen Beispielen gegenüber gestellt.

[1] `juri.schulte@uni-ulm.de`, Matrikelnummer: `677063`, Studiengang: `Ma Informatik`, Universität Ulm, Wissenschaftliche Nutzung der Medien

[2] Hypertext Transfer Protocol: Protokoll zur Übertragung von Daten über ein Netzwerk

[3] Computerprogramm zur Interaktion mit Webseiten und deren Darstellung

2 Profilbildung

Profilbildungsmechanismen werden besonders häufig für gezielte digitale Werbung verwendet. Damit können beispielsweise Produkte, die von einem Besucher in einem Onlineshop angesehen wurden, über verschiedene Webseiten hinweg durch Werbeanzeigen angepriesen werden. Durch die Sicherheit, dass der potenzielle Kunde den Artikel bereits einmal freiwillig ausgewählt hat, kann die Werbung auf die Artikelgruppe oder ähnliche Artikel ausgeweitet werden. Die nötigen Technologien zur Verfolgung von Benutzern sind u. A. Cookies, HTTP und Javascript, die im Folgenden näher beschrieben werden.

2.1 Cookies

Cookies sind kleine Textdateien, die der Browser für eine Webseite anlegen kann. Diese Textdateien können beliebig gefüllt werden - meist mit persönlichen Einstellungen, wie bspw. der bevorzugten Sprache und Schriftgröße in der eine Webseite angezeigt werden soll. Es wird dann auch von *Persistent Cookies* gesprochen. Die Webseite erkennt den angelegten Cookie bei einem erneuten Besuch wieder und kann die Daten auslesen. Normalerweise erhalten solche Cookies einen Zeitraum, in dem sie gültig sind und werden nach Ablauf gelöscht. Allerdings wird der Zeitraum von der Webseite selbst festgelegt und kann damit unter Anderem ganze Dekaden umfassen. Bezogen auf das Beispiel der gezielten Werbung, könnte nun beim Aufruf eines Produkts ein Cookie mit der Produktbezeichnung gesetzt und beim nächsten Besuch ausgelesen werden. Um diese Informationen über Webseiten hinweg zugänglich zu machen, werden so genannte *Retargeting pixel* verwendet, die letztendlich einer weiteren Webseite gestatten Cookies auf der ursprünglichen Webseite zu lesen und zu setzen.
Cookies generell zu verbieten ist zwar eine Möglichkeit diese Geschäftspraktiken zu unterbinden, haben aber den Nachteil, dass vor allem Webseiten mit Benutzeranmeldung nicht mehr korrekt funktionieren. Für eine Sitzung[1] werden *Session Cookies* benötigt die bei einem Cookieverbot ebenfalls unterbunden werden. Eine Alternative bietet die manuelle Bestätigung von Cookies (die in den meisten Browsern angeboten wird).

[1] Zusammenhängende Zugriffe auf eine Webseite für die nur eine einmalige Anmeldung erforderlich ist

2.2 Hypertext Transfer Protocol

Im Gegensatz zu Cookies ist die Verwendung von HTTP heute im Internet zwingend erforderlich. Protkolle, wie HTTP, steuern und erhalten die Verbindung zur Webseite. Eine HTTP-Nachricht enthält immer einen *User Agent* - das verwendete Betriebssystem mit dem verwendeten Browser des Seitenaufrufers - und *HTTP_ACCEPT Headers* - gewünschte Sprachen und gewünschte Dateiformate. Allein diese beiden Einstellungen bieten genug Informationen um den Besucher eindeutiger identifizieren zu können (mit einem gängigen Browser erhöht sich die Wahrscheinlichkeit um ca. das 100.000-fache) ([Bir02, Pet10]). [Ele13] stellt einen Online-Test zur Verfügung, der die Einzigartigkeit des eigenen Browsers berechnet. In [Moz13] werden Vorschläge genannt, wie durch Anpassungen der HTTP-Standards der Informationsgehalt dieser beiden HTTP-Parameter reduziert werden kann. Besonders hinsichtlich der Anzahl der im User Agent enthaltenen Informationen könnten Einsparungen getroffen werden - so ist es nicht erforderlich das genaue Betriebssystem mitzuschicken, eine grobe Einordnung reicht aus. Diese Änderung würde den Wertebereich des Parameters reduzieren und damit den Informationsgehalt.

2.3 Javascript

Javascript wird von der Webseite an den Browser übertragen und auf dem Computer des Aufrufers ausgeführt. Durch spezielle Javascript-Funktionen können zusätzliche Informationen über den Besucher gesammelt werden. Zu den Parametern aus 2.2 fügen sich dadurch Informationen über installierte Schriftarten, installierte Plugins, Zeitzone, Bildschirmauflösung und Farbtiefe sowie Fenstergröße hinzu. Wie bereits in 2.2 gezeigt genügt schon die Existenz von zwei Parametern um eine Besucheridentifikation signifikant zu erleichtern. Durch die Parameter, die Javascript liefert, erhöht sich dieser Faktor so stark, dass der Besucher in ca. 90%[1] aller Fälle eindeutig identifizierbar wird ([[Pet10]]).

Javascript einfach zu deaktivieren hindert viele Webseiten daran sich richtig darzustellen und den vollen Funktionsumfang zu gewähren. Auch wenn die Deaktivierung von Javascript eine mögliche Schutzmaßnahme darstellt, so sollten andere Möglichkeiten in Betracht gezogen werden. Das Tor Project ([PI13]) reduziert in ihrem Browser-Bundle bspw. die Identifikationsmöglichkeiten, indem die Fenstergröße auf einen festen Wert gesetzt, Flash deaktiviert, das Auslesen von Schriftarten verhindert und die Zeit-

[1] falls Java oder Flash verwendet wurden

zone auf einen festen Wert gesetzt wird. Desweiteren werden die HTTP-Parameter auf die nötigsten Informationen gekürzt (statt mehreren bevorzugten Sprachen wird bspw. nur eine fest angegeben).

3 Zusammenfassung und Ausblick

Profilbildungen bedarf kein hoher Aufwand. Mit einfachsten Mitteln können Webseiten so präpariert werden, dass Benutzer wiedererkannt und verfolgt werden. Die einfachen Schutzmaßnahmen hingegen schränken den Benutzer oftmals zu stark in der Nutzung des Internets ein. Cookies 2.1 und Skripte 2.3 jedes mal manuell zu erlauben wird auf Dauer ermüdend und gestaltet sich für nicht fachkundige Benutzer schwierig. Zwar bietet unter Anderem das Tor Project einen gewissen Grad an Sicherheit für nicht erfahrene Nutzer, allerdings bilden die hier genannten drei Profilbildungsmechanismen nur einen Auszug aus dem existierenden Repertoire. Besonders IP-Adressen, DOM Cookies und Flash Cookies ergänzen Informationen über Besucher und können sogar Profiländerungen ausgleichen.
Außer Acht wurde dabei bisher die gesetzliche Grundlage gelassen. So einfach Profilbildungsmaßnahmen auch sein mögen, so illegal sind sie meist auch. Personenbezogene Daten dürfen in Deutschland nur zu einem festgelegten Zweck gesammelt werden und auch nur unter strikten Vorlagen. Und zu diesem Zweck muss der Benutzer erst einmal zustimmen.

Literatur

[Bir02] Birgit Pfitzmann, Michael Waidner. *Proceedings of the 2002 ACM workshop on Privacy in the Electronic Society: Privacy in Browser-Based Attribute Exchange.* ACM, New York and NY, 2002.

[Ele13] Electronic Frontier Foundation. Panopticlick: How unique and trackable is your browser?, zuletzt besucht am 24.06.2013. Erreichbar unter https://panopticlick.eff.org/.

[Moz13] MozillaWiki. Fingerprinting, zuletzt besucht am 17.06.2013. Erreichbar unter https://wiki.mozilla.org/Fingerprinting.

[Pet10] Peter Eckersley. *Proceedings of the Privacy Enhancing Technologies Symposium: How Unique Is Your Web Browser?* 2010.

[PI13] The Tor Project and Inc. Tor browser bundle, zuletzt besucht am 23.06.2013. Erreichbar unter https://www.torproject.org/ projects/torbrowser.html.en.