Preprint. Edited version published as: Langheinrich, Karjoth: Online-Marketing – Eine Balance zwischen Nutzen und Schutz. digma – Zeitschrift fuer Datenrecht und Informationssicherheit, Vol. 12, No. 3, Schulthess, October 2012, pp. 116-123

Eine Balance zwischen Nutzen und Schutz

Wie Internetnutzer sich vor der Beobachtung durch Werbetreibende schützen können.

Lead

Online-Marketing ist die grosse Einnahmequelle für viele Dienste im Internet. Wie kann sich der Einzelne vor zu viel Beobachtung schützen?

Viele Inhalte und Dienste des Internets sind kostenlos, d.h. durch Werbung finanziert. In diesem Geschäftsmodell wird Wissen über Interessen der Konsumenten in Nutzungsprofile umgewandelt, um passende Werbeanzeigen zu platzieren. Je genauer das Profil, desto höher sind in der Regel die erzielten Werbeeinnahmen. Es ist ein schmaler Grat zwischen Kundenanpassung und Nachstellen (Stalking). Denn nicht jeder möchte die Annehmlichkeit personalisierter Angebote oder verweigert sich der unausgesprochenen Vereinbarung, dass Konsumenten kostenlos Webangebote durchstöbern können und im Gegenzug persönliche Daten abtreten.

Betrachten wir eine Website als nicht vertrauenswürdig, können wir deren Besuch vermeiden. Doch wenn wir im Zeitalter von Web 2.0 eine Webseite laden, stammt deren Inhalt nicht nur von dem Betreiber der Website, sondern es werden auch Teile der Seite – z.B. Bilder oder andere Inhalte – von externen Quellen bezogen, sogenannten Fremdanbietern. Um diese Inhalte zu laden, kontaktiert der Browser die Server der Fremdanbieter. Sind diese Fremdanbieter mit mehreren vom Nutzer besuchten Webseiten verlinkt, kann der Nutzer auf seinem Weg durchs Internet nachverfolgt und ein Nutzerprofil erstellt werden. Da die Existenz dieser zahlreichen Fremdanbieter beim Betrachten der Webseite nicht sichtbar ist, können Nutzer nur schwer eine bewusste Auswahl treffen. Wer z.B. versucht, ausschliesslich auf schweizer oder europäischen Websites zu bleiben, wird dennoch in vielen Fällen von Fremdanbietern aus den USA dabei beobachtet werden.

Obwohl beim Web-Tracking im Allgemeinen keine personenbezogenen Daten erfasst werden, sondern vielmehr „nur“ statistische Daten über die Nutzung der Websites, haben doch viele die Befürchtung, dass durch Zusammenführung mit anderen Daten ein Personenbezug hergestellt werden kann, was die Angst vor Missbrauch schürt. Auch ist es in vielen Fällen bereits problematisch genug, zwar nicht identifziert, sondern lediglich kategorisiert zu werden. Der Soziologe David Lyon nennt dies „social sorting“ – die Einteilung von Menschen ohne ihr Wissen (und daher auch ohne die Möglichkeit der Einflussnahme) in verschiedene Klassen, deren Zugehörigkeit dann über die Zuteilung oder die Wegnahme von Resourcen und Angeboten entscheiden – den Rabatt auf ein bestimmtes Buch oder auch die Vergabe eines Kredits. Es ist daher verständlich, dass viele Nutzer den Drang nach personalisierter Werbung im Web durchaus unterstützen, dennoch aber eine angemessene Balance bei der Akquise personenbezogener Daten herstellen wollen.

Immer ausgeklügelteren Tracking-Techniken auf der einen Seite stehen inzwischen eine grosse Anzahl von Software Werkzeugen – meistens Browser-Plug-Ins – gegenüber, mit denen der Einzelne das Tracking im Web zum Teil kontrollieren kann. Der Gebrauch dieser Tools und ihr Zusammenspiel in der Praxis ist allerdings komplex, da einfache Lösungen typischerweise auch signifikante Einschränkungen bezüglich der Funktionalität eines Webdienstes mit sich bringen. Wie sehen Techniken aus, die eine Balance zwischen Funktionalität und Datenschutz im Web ermöglichen? Dieser Artikel befasst sich mit dem Sammeln von Daten, die man mit Hilfe dieser Tracking-Techniken gewinnen kann. Missbrauch von

Sicherheitsschwachstellen oder der Einsatz zusätzliche Massnahmen wie Registrierung für einen Dienst oder die Teilnahme an einem Quiz werden nicht berücksichtigt.

Ökosystem Web

In den letzten Jahren hat sich rund um das Anbieten von Inhalten im Web eine feinmaschige Infrastruktur gebildet, die auch Auswirkungen auf die Werbewirtschaft hat.[i] Gab es früher lediglichen den Webseitenbetreiber (Erstanbieter von Online-Inhalten), der die Anzeigen eines Werbekunden direkt auf seinen Seiten schaltete, ist die Situation heute ungleich komplexer. Werbevermarkter schliessen sich in Werbenetzwerken (ad networks) zusammen, um die koordinierte Platzierung von Werbung über mehrere Websites hinweg vorzunehmen; Analytics-Dienste erfassen möglichst umfassende Daten über die Nutzer und ihre Nutzung einer Webseite; Content Provider stellen weiteres Material wie Videos, News, und Wetterberichte zur Verfügung; und Hosting Platforms helfen dem Webseitenbetreiber bei der Verteilung ihres Inhaltes. Desweiteren entstehen *Werbebörsen* (ad exchanges), in denen Anzeigenplätze von verschiedenen Werbenetzwerken versteigert werden.

Als Folge dieser Entwicklung sind inzwischen viele unabhängig voneinander betriebenen Websites indirekt durch Dienste von Fremdanbietern vernetzt.[ii] Diesen Fremdanbietern kommt dabei eine immere zentralere Rolle beim Tracking von Webnutzern zu, da sie oft in der Lage sind, ein und denselben Nutzer beim Besuchen höchst unterschiedlichen Webseiten zu verfolgen. Mit der Browser-Erweiterung Collusion lässt sich diese Vernetzung sichtbar machen. Es zeigt die besuchten Websites als Knoten in einen Graphen dar. Wählt man einen Knoten aus, zeigen Pfeile an, welche anderen Knoten ebenfalls im Rahmen des ursprünglichen Besuches Inhalte oder Funktionalitäten lieferten. So wird schnell ersichtlich, in welcher Verbindung beispielsweise die im Laufe eines Tags besuchten Websites stehen. Hat man etwa das Webangebot von NZZ und Tagesanzeiger besucht, sieht man, dass diese über zwei Werbefirmen (AdTech und Net-Metrix) in Beziehung stehen.

Profilbildung

Ein Vorteil der digitalen Medien für die Werbewirtschaft ist, dass die Reaktion der angesprochenen Verbraucher über Klickrate, Konversionsrate oder Anzahl Seitenabrufe unmittelbar gemessen werden kann. Wurde früher Werbung gemäss der Ausrichtung der Website oder dem Inhalt der Webseite geschaltet, zieht man heute das Verhalten des Nutzers mit in Betracht. In dieser personalisierten Werbung erscheint der Hinweis für einen Blumen-Shop nur, wenn jemand tatsächlich nach „Rosen“ oder „Sträussen“ sucht. Man nennt das auch verhaltensorientierte Werbung. Behavioral Advertising beruht auf dem Verfolgen (Tracken) der Online- Aktivitäten eines Nutzers über die Zeit – einschliesslich der Suchen, die der Nutzer durchgeführt hat, die besuchten Webseiten und der betrachtete Inhalt – um Werbung zielgerichtet auf die Interessen des einzelnen Nutzers liefern zu können. Dabei können die Informationen über das Surfverhalten eines Nutzers prinzipiell auch mit Daten, die ausserhalb des Internets von einer Drittpartei gesammelt wurden, verbunden werden. Oft reichen jedoch wenige zusätzliche Informationen um Personen in anonymisierten Daten zu erkennen.

Daten, welche für das Behavioral Advertising gesammelt werden, sind in der Regel keine personenbezogenen Daten, weil sie Personen nicht identifizieren, sondern in Interessenkategorien einteilen – Werbung für Motorradfahrer sollen nur Nutzer sehen, die sich auch für Motorräder interessieren. Tracking liefert die Parameter, mit denen die Nutzer automatisiert Zielgruppen zugewiesen werden (Targeting). Zum Beispiel schalten Online-Werbefirmen auf Basis von Geotargeting nationale oder sogar regional differenzierte Werbung. Besucher sehen Anzeigen, die – unabhängig vom Standort der aufgerufenen Seite – ihren derzeitigen Aufenthaltsort als Zielmarkt ansprechen. Dazu unterstützen aktuelle Browser inzwischen das Geolokalisierungsprotokoll[iii] des W3Cs, welches es dem Browser ermöglicht, aufgrund der momentan sichtbaren drahtlosen Netze (WiFi) den Aufenthaltsort des Nutzers an die momentan besuchte Website mitzuteilen.

Wie gut solch eine anonyme Klassifikation von Besuchern funktioniert zeigte die „What They Know“-Serie[iv] des Wall Street Journals

bereits im August 2010: mit nur einem einzigen Aufruf der Homepage einer Kreditkartenfirma konnte die auf die anonyme Nutzerklassifkation spezialisierte [x+1] Inc. sechs Testnutzer mit unheimlicher Präzision in verschiedenste werberelevante Klassen einteilen (z.B. „Detroit Home Owner“ oder „White-collar worker“). Zwar gab es auch Fehlklassifikationen, doch würden diese durch das weiterführende Tracking nach und nach verbessert.

Tracking

Ziel des Tracking ist es, Besucher einer Webseite über mehrere Besuche hinweg zu (re-)identifizieren um so die Bildung eines Nutzerprofils zu unterstützen. Mittels eines solchen Profils kann dann eine Klassifikation des Besuchers vorgenommen werden, um so zielgerichtet Werbung bzw. Produktangebote unterbreiten zu können. Kann eine solche Nutzerverfolgung über mehrere verschiedene Websites durchgeführt werden, erhöht dies typischerweise die Qualität der erhebbaren Daten. Dies kann z.B. durch die Zusammenarbeit eines Webseitenbetreibers mit einem Fremdanbieter, oder durch den Austausch von Daten mit anderen Betreibern erreicht werden.

Tracking-Technologien

Explizite Trackingmechanismen verwenden an Benutzer zugewiesene eindeutige Identifikatoren. Tracker können diese Identifikatoren (Pseudonyme) auf der Maschine des Benutzers speichern und sie bei einem späteren Besuch der gleichen Website wieder auslesen. Implizite Trackingmechanismen ermöglichen eine (probabilistische) Re-identifikation auch ohne vorher platzierte Pseudonyme. Beim Tracken durch Fremdanbieter bzw. in Kooperation mit anderen Website-betreibern kann das Verhalten des Nutzers über mehrere verschiedene Websites hinweg miteinander zu einem einzigen Profil verknüpft werden.

Vier grundlegende Techniken erlauben solch eine Nutzerverfolgung: Cookies, eingebettetes Bildmaterial, Protokolldaten und Skripte.

Cookies

Cookies wurden Mitte der 90er Jahre von Netscape entwickelt, dem damaligen Platzhirsch der Browseranbieter, um komplexere Webseiten zu unterstützen. David Whalen, Autor der ersten „Unofficial Cookie FAQ“[vii], erklärt das Prinzip von Cookies mit dem Besuch einer Textilreinigung: Beim ersten Besuch gibt man Wäsche ab und erhält ein Ticket. Dieses muss beim zweiten Besuch vorgelegt werden, da sonst der Verkäufer weder erkennt, dass man schon einmal da war, noch welche Kleidungsstücke man abholen will. Da HTTP – das Protokoll für die Kommunikation zwischen Browser und Website – ein sogenanntes „zustandsloses“ Protokoll ist, wird jeder Besuch, jeder Klick auf einer Website als „neuer“ Besuch gesehen. Um diese individuellen Klicks und Seitenaufrufe miteinander zu verbinden, konnten Browser nun Tickets – Cookies – verwenden, die jeder Seitenanfrage beigelegt wurden und es so der Website erlauben, all diese Klicks zu einer scheinbar nahtlosen Interaktion zu verbinden.

Ein Cookie ist nichts anderes als ein Stück Text in Form eines „Key-Value Pairs“ (Schlüssel-Wert-Paar), welches der Server als Teil der vom Browser angeforderten Webseite zurück schickt. Der Browser speichert typischerweise diese Informationen ungefragt auf dem Computer des Nutzers und schickt sie nun bei jeder erneuten Seitenanfrage an diesen Server mit. Cookies werden grundsätzlich beim Schliessen des Browsers gelöscht (sogenannte „Session-Cookies“). Ein Server kann seinen Cookies allerdings auch ein explizites „Verfallsdatum“ geben (sogenannte „Permanent Cookies“) – diese werden vom Browser erst zum angegebenen Datum gelöscht und überleben so sowohl einen Neustart des Browsers als auch den Neustart des ganzen Systems. Das Verfallsdatum kann natürlich auch 100 Jahre betragen. Cookies haben zwar eine Maximalgrösse von 4096 Zeichen, doch kann eine Website weitere Daten in ihrer eigenen Datenbank speichern und bei einem späteren Besuch mittels des (kurzen) Pseudonyms nachschlagen.

Da das Speichern der Cookies Sache des Browsers ist, kann dieser – bei entsprechender Konfiguration – dies natürlich auch unterlassen. In vielen Fällen wird dadurch aber das Verwenden einer Website praktisch verunmöglicht: ganz ohne Cookies funktioniert kaum ein eCommerce System, keine Wettersuche, Kartennavigation, oder Forenbesuch. Die vollständige Abschaltung

von Cookies in einem Browser ist daher kaum praktikabel.

Solche „guten“ Cookies, die die Funktionalität einer Webseite erst ermöglichen, von potentiell „schlechten“, d.h. nur zu Trackingzwecken eingesetzten Cookies zu unterscheiden, ist oft nur durch langwieriges Probieren herauszufinden. Zum Beispiel werden allein beim Besuch der NZZ-Homepage (www.nzz.ch) 36 verschiedene Cookies,gesetzt. Neben sieben Cookies vom nzz.ch Server selbst (sogenannte „Erstanbieter-Cookies“ – *first-party cookie*) sind dies beispielsweise auch neun von Google und elf von Twitter (sogenannte „Fremdanbieter-Cookies“ - *third-party cookie*), fast alle mit kryptischen Namen („BEAT“, „s_cc“) und ebenso kryptischen Inhalt. Zwar begrenzt die Cookie-Spezifikation das Auslesen von Cookies auf diejenige Website, die sie auch gesetzt hat (so kann beispielsweise Google keine der von nzz.ch gesetzten Cookies lesen), doch ist eine gewollte Kooperation verschiedener Webdienste durch geschickte Koordination der einzelnen Elemente einer Webseite durch den Website-Betreiber leicht möglich.

Super-, Cache- und Zombie-Cookies

Das generelle Prinzip von Cookies hat inzwischen auch in verschiedenen Browser-Erweiterungen (sog. Plug-ins) Anwendungen gefunden. Das wohl mit Abstand populärste Plugin ist das von der Firma Adobe entwickelte Flash Plug-in, welches interaktive Grafiken, Anwendungen und Spiele in Browsern ermöglicht. Ein Website-Betreiber kann nun beispielsweise ein Cookie nicht über den herkömmlichen Weg wie oben beschrieben setzen (mit den damit einhergehenden Beschränkungen), sondern innerhalb eines vom Flash Plug-in ausgeführten Programms. Das Cookie wird so also nicht gemeinsam mit den anderen Cookies vom Browser verwaltet, sondern ist nun quasi versteckt als Teil einer Flash Anwendung. Dies macht es unmöglich, diese Art von „Super-Cookies“ mit herkömmlichen Browser-Werkzeugen zu erkennen bzw. zu löschen. Auch gelten die oben beschriebenen Beschränkungen für herkömmliche Cookies bzgl. des Auslesens nicht, d.h. prinzipiell kann eine Flash-Applikation so das Nutzerverhalten auf beliebigen Websites ausspähen.

Auch das neu vorgestellte HTML5 Format erlaubt nun Webseiten, mittels „Client-side storage“ beliebige Informationen auf der Festplatte des Besuchers abzulegen. Diese haben, im Gegensatz zu herkömmlichen Cookies, eine mehr als 1000fache Maximalgrösse und kennen keinerlei Ablaufdatum.

Cache-Cookies (nach dem dazugehörigen Protokollkopfeintrag auch „ETag-Cookie“ oder „ETags“ genannt) machen sich den vom Browser verwendeten Zwischenspeicher zunutze, der dazu dient, das Herunterlanden von Webseiten zu beschleunigen. Dabei hält der Browser Webseiten oder Elemente von Webseiten (z.B. Bilder) lokal vor und „fragt“ bei einem Neubesuch der Website zunächst den Server, ob an den bereits heruntergeladenen Elementen Änderungen vorliegen. Falls nicht, kann der Browser die lokale Kopie verwenden und muss diese nicht erneut herunterladen. Ein Cache-Cookie ist nichts anderes als beispielsweise ein Bild, welches vom Server für jeden Nutzer einen anderen „ETag“-Wert im Protokoll erhält. Besucht der Nutzer diese Seite später wieder, schickt sein Browser unter anderem auch den ETag-Wert für das entsprechende Bild mit und gibt dabei unabsichtlich auch das vorher vom Server vergebene Pseudonym des Nutzers preis.

Mit Hilfe solcher Super- und Cache-Cookies können schliesslich sogenannte „Zombie-Cookies“ erzeugt werden, d.h. es können vom Nutzer explizit gelöschte „herkömmliche“ Cookies heimlich wieder hergestellt werden. Im Falle des oben beschriebenen Cache-Cookies würde der Server das vom Browser übermittelte Pseudonym im Anschluss neu als herkömmliches Cookie setzen. Selbst wenn der Nutzer dieses bewusst löschen sollte, würde es beim nächsten Besuch – Cache-Cookie sei Dank – wieder „auferstehen“.

Zählpixel und Fingerprints

Zählpixel sind in die Webseite eingebettete Bilddateien, die typischerweise nur aus einem einzigen Pixel (in Hintergrundfarbe bzw. transparent) bestehen und von Servern eines Fremdanbieters zur Verfügung gestellt werden. Für den Besucher einer Webseite sind solche Miniatur-„Bilder“ praktisch unsichtbar. Durch das Laden dieser Datei kann der Fremdanbieter eigene

Cookies setzen bzw. lesen, sowie den Browser-Typ, das Betriebssystem und die IP-Adresse des Nutzers der Originalwebsite feststellen.

Letztere Informationen sind Bestandteil des Protokollkopfs und werden vom Browser bei jeder Webanfrage mitgeschickt. Peter Eckersley von der US-amerikanischen Electronic Frontier Foundation (EFF) bemerkte als einer der Ersten, dass die Kombination der scheinbar banalen Protokollkopfdaten – hauptsächlich Informationen zur verwendeten Browsersoftware und -version, dem Betriebssystem und den installierten Plug-ins im Browser – infolge der Heterogenität der Installationen eine Art Fingerabdruck ergeben. Durch den Einsatz von Skripten bzw. Flash Programmen können weitere „harmlose" Informationen (z.B. installierte Schriftarten, Zeitzone) abgerufen werden. In einem ersten Experiment war Eckersley in der Lage, beinahe 95% der Besucher einer Testseite[viii] eindeutig aus diesen Informationen zu identifizieren. Firmen wie BlueCava[ix] vertreiben diese Technik inzwischen kommerziell, um Website die Identifikation von Besuchern ohne den Einsatz von Cookies zu ermöglichen.

Skripte

Moderne Websites lassen die Grenze zwischen Webseiten und installierten Anwendungen verschwimmen. So hat das von Google propagierte „Chromebook" das Konzept der lokal installierten Anwendungen aufgegeben und bietet lediglich einen Web-Browser, in dem dann alle Anwendungen (z.B. Office-Applikationen) als Web-Apps laufen. Ermöglicht wird dies durch den Einsatz von Skripten (Programmen), welche sich in Webseiten einbetten lassen. Lädt ein Browser solch eine Seite herunter, führt er gleichzeitig auch die darin enthaltenen Skripte aus.

Mithilfe solcher Skripte können nicht nur komplexe Applikationen geschaffen werden. Auf „normale" Webseiten können dadurch beispielsweise detaillierte Interaktionsmessungen vorgenommen werden, insbesondere die Bewegung der Maus, das Aufklappen von Menüeinträgen, oder das Seitenblättern mit dem Scroll-Wheel. Firmen wie PicNet[x] bieten umfassende Analysetools, mit denen für jeden einzelnen Besucher jegliche Mausbewegungen innerhalb des Browserfensters festgehalten und analysiert werden können. Prinzipiell liesen sich solche Messungen natürlich auch für die Klassifikation bzw. Identifikation von Besuchern nutzen.

Für den Einsatz von Skripten gelten an sich strenge Regeln, die – ähnlich wie bei den „normalen" Cookies – die Kommunikation zwischen dem Browser des Nutzers und den Servern einschränken: ein Skript kann lediglich mit dem Server kommunizieren, auf dessen Seite es platziert wurde. Ein Skript auf der Homepage der NZZ kann also nur mit Web-Servern der NZZ (unter nzz.ch) kommunizieren. Auch können Skripte nicht auf Inhalte von etwaig geöffneten anderen Browser-Fenstern bzw. –Tabs zugreifen. Dennoch gibt es eine Vielzahl von Sicherheitslücken, die es Angreifern erlauben, Skripte so zu platzieren, dass sie die auf einer vermeintlich vertrauenswürdigen und sicheren Website eingegebenen Daten (z.B. beim Online Shopping oder E-Banking) auf Fremdserver kopieren bzw. umleiten. Solche sogenannten „Cross-Site Scripting (XSS)" Angriffe sind sehr schwer zu erkennen und können daher unbemerkt für längere Zeit aktiv sein. Auch kann hier der Nutzer kaum durch „sicheres" Verhalten vorbeugen, da weder Anti-Virus Software noch das ausschliessliche Besuchen vertrauenswürdiger Webseiten einen Schutz bieten kann.

Wie für Cookies gilt auch für Skripte: Zwar lassen diese sich komplett abschalten, doch wird dann kaum noch eine Webseite für den Nutzer sinnvoll besuchbar sein.

Schutzmechanismen

Für Nutzer modernen Web-Browser ergeben sich drei grundlegende Möglichkeiten, das ungewollte Tracking in Web zu minimieren: Das stetige Löschen von ungewollten Cache-, Cookie-, und Local Storage-Einträgen; das explizite Setzen von Opt-Out Markern; sowie das aktive Blockieren ungewollter Inhalte.

Cookie- und Local Data Management

Im Browser-Menü können (Erstanbieter-) Cookies sowohl für spezifische Websites als auch prinzipiell blockiert werden. Dies führt aber dazu, dass praktisch keine Website mehr funktioniert. Praktikabler mag da die

Einschränkung solch einer Blockade auf Fremd-Cookies sein, da diese oft von externen Anbietern zwecks Tracking verwendet werden. Doch auch hier unerwünschte Seiteneffekte auftreten können, z.B. bei oft durch Fremdanbieter durchgeführte Zahlungsprozesse. Eine Alternative ist, statt einer generellen Blockierung lediglich die „Lebensdauer“ von Cookies auf die aktuelle Browsing-Session zu beschränken. Bei einem Neustart des Browsers sind so alle Cookies gelöscht. Browser Plug-Ins wie das populäre „Cookie-Monster“ (Firefox) erlauben die Definition komplexer Regeln und Ausnahmen.

Ebenfalls mit Bordmitteln lassen sich der Browser-Cache und die HTML5-basierten lokalen Datensätze löschen; doch auch hier bieten Plug-Ins wie „BetterPrivacy“ (Firefox) eine komfortablere Bedienung. Hartnäckige Cookies wie Flash-Cookies können oft nur mit solchen Plug-Ins gelöscht werden.

Moderne Browser bieten inzwischen einen „Private Browsing“ Modus an, welcher mit einer einfachen Tastaturkombination (z.B. Strg-P) aktiviert wird und jegliche Aktivitäten (Cache, Lokale Speicherung, Cookies) innerhalb dieser Session nach Beendigung löscht.

Opt-out Cookies und Do-Not-Track Protokoll

Weitaus weniger Probleme gibt es bei dem praktisch umgekehrten Ansatz, durch das Setzen von speziellen Cookies bzw. Protokollkopf-Informationen im Browser Websites mitzuteilen, dass Tracking unerwünscht ist. Zwar hilft dies nur auf solchen Websites, die diese Informationen auch verwenden (und dann tatsächlich keine Tracking Cookies setzen), aber dafür wird die Verwendung von Websites in keiner Weise eingeschränkt. Verschiedene Interessenverbände der Online-Werbeindustrie (z.B. die Network Advertising Initiative, NAI) bieten Listen von „Opt-Out“ Cookies für jedes ihrer Mitglieder an, welche interessierte Nutzer auf ihrem Browser installieren können. Wird dann eine Website besucht, deren Online-Werbung von einem dieser Mitglieder geliefert wird, schickt der Browser automatisch das passende Opt-Out Cookie. Da es allerdings Dutzende, wenn nicht Hunderte von Mitgliedern gibt, werden so entsprechend viele zusätzliche Cookies nötig, die es womöglich umso schwerer machen, sich manuell einen Überblick über seine Cookies zu verschaffen. Auch würde das Löschen aller Cookies solche Opt-Out Cookies löschen und damit das Tracking wieder ermöglichen.

Hoffnung verspricht hier der sich noch in der Entwicklung befindliche „Do Not Track“ Standard, welcher gegenwärtig von einem Konsortium bestehend aus namhaften Industrievertretern sowie Datenschützern unter der Regie des World Wide Web Consortium (W3C) vorangetrieben wird[xi]. Erste Implementationen finden sich allerdings bereits in den aktuellsten Version der bekanntesten Browsern (Firefox, Chrome, IE10). Das Prinzip ist simpel: ein Protokollkopf namens „DNT“ kann entweder den Wert „1“ haben (bitte nicht tracken), „0“ (tracken erlaubt), oder nicht definiert sein (keine Angabe). Microsoft erregte grosses Aufsehen als es sich 2011 entschied, bei der Entwicklung seiner neuesten Browser-Version (IE10) per Voreinstellung DTN auf „1“ zu setzen (typischerweise wäre der Eintrag undefiniert bis der Nutzer explizit einen Wert setzt). Der DoNotTrack-Ansatz macht es sicherlich leichter, explizit das Tracking zu unterbinden, doch gibt es bisher erst wenige Werbeanbieter, die sich bereiterklärt haben, einen solchen Protokolleintrag zu honorieren. Auch fehlt, wie bei den Opt-Out Cookies, in den meisten Ländern eine rechtliche Grundlage, auf derer sich diese Angaben erzwingen liessen. Daher kommt dem W3C „Tracking Compliance and Scope“-Dokument[xii] einer wichtige Rolle zu, da es definiert, was es bedeutet, dieses DNT zu erfüllen.

Blocking

Den grösstmöglichen Schutz bieten immer noch individuelle Blocking-Werkzeuge, welche – ähnlich wie die oben beschriebenen Cookie-Management Plug-Ins – bereits im Ansatz die Annahme von Cookies und anderen Daten, das Herunterladen von Zählpixeln bzw. das Ausführen von Skripten unterbinden. Hier kommt es vor allem darauf an, eine komfortable und flexible Bedienung zu ermöglichen, welche das Definieren von Regeln und Ausnahmen mit einem für den Nutzer akzeptablen Aufwand erlauben.

Ghostery[xiii] ist ein kostenloses Plugin für die gängigsten Browser, welches die wohl umfangreichste Abdeckung gegen die oben

beschriebenen Tracking-Techniken bietet. Auf jeder besuchten Website wird eingeblendet, welche bekannten Tracker auf der Seite aktiv sind und wie Informationen über den Nutzer bezogen werden (z.B. über Zählpixel oder Cookies). Aus dieser Übersicht heraus kann der Nutzer dann geziehlt einzelne oder alle diese Fremdanbieter blockieren bzw. sich weitere Informationen zu jedem der Trackinganbieter anzeigen lassen. Ghostery blockiert versteckte Skripte also nicht kategorisch, sondern informiert den Nutzer lediglich über deren Existenz. Der Nutzer hat dann die Möglichkeit, das Skript zu blockieren und/oder sich über dessen Urheber und Datenschutzrichtlinien zu informieren.

Weitaus aktiver im Schutz sind explizite Script- bzw. Ad-Blocker, wie etwa die bekannten Erweiterungen AdBlock Plus und NoScript/NotScript. Auf der Grundlage stetig aktualisierter Blacklists blockieren beide Plug-Ins direkt nach der Installation eine sehr grosse Anzahl von Trackern und anderen, sicherheitsrelevanten Problemen. Durch individuelle Ausnahmelisten bzw. angepassten Regeln können dann nicht funktionierende aber vertrauenswürdige Websites vom Blocking ausgenommen werden.

Kategorien von Trackern

Nicht alle Tracker sind gleich. Roesner et al (2012) teilt Tracker in fünf unterschiedliche Klassen ein, mit jeweils unterschiedlichen Fähigkeiten des Trackings und damit unterschiedlichen Anforderungen an die vom Nutzer eingeleiteten Schutzmassnahmen.

Single-Site Tracking – Beispiel Webanalyse

Für den technischen und inhaltlichen Betrieb einer Website ist es heute unabdingbar, die täglichen Besucherströme zu verstehen: Wann kommen die meisten Besucher? Welche Seiten sind am populärsten? Zwar finden sich Antworten auf all diese Fragen in den vom Webserver aufgezeichneten lokalen Log-Dateien, doch bieten inzwischen zahlreiche Fremdanbieter komfortable Analysetools an, die es dem Betreiber ermöglichen, ohne grossen Aufwand detaillierte Statistiken über sein Angebot zu erhalten. Platzhirsch ist dabei, wie so häufig, Google mit seinem „Google Analytics“ Angebot. Die Verwendung ist denkbar einfach: Die besuchte Webseite enthält eine Bibliothek (in der Form eines Skripts), welches vom Webanalyse-Anbieter zur Vefügung gestellt wird. Sobald der Browser die Seite darstellt wird auch das Skript ausgeführt und um Besucher zu tracken. Um wiederkehrende Besucher zu erkennen, setzt das Skript im Browser des Benutzers ein eindeutiges Cookie. Das Skript schickt diesen Identifikator an an die auswertende Stelle, zusammen mit Informationen über die einbindende Webseite und das verwendete System.

Prinzipiell ist es möglich, die Identifikatoren so zu koordinieren, dass der Webanalyse-Anbieter Nutzer über praktisch alle teilnehmenden Websites hinweg tracken kann. Doch im Falle von Google Analytics wird der Identifikator „lokal“ gesetzt, d.h. ohne etwaige Koordination mit einem zentralen Server. Daher ist Google Analytics ein „single-site tracker“. Will man als Anbieter mehrerer Websites diese gemeinsam in einer einzigen Analyse zusammenfassen, muss man manuell Skriptcode einfügen, der diese verschiedenen Identifikatoren zusammenführt.

Der Schutz vor solchen Trackern ist beispielsweise durch das Blockieren aller Skripte des Webanalyse-Anbieters leicht möglich (z.B. mit dem NoScript Plug-In), meist ohne die Funktionalität der Webseite merklich einzuschränken. Allerdings wird so dem Website-Betreiber die durchaus legitime Analyse seines Webangebots erschwert.

Cross-Site Tracking I – Beispiel Werbung

Was bei der Webanalyse nicht nötig ist (und von Anbietern wie Google Analytics, auch wenn es möglich wäre, nicht gemacht wird), ist in Werbenetzwerken die „raison d’etre“: das Tracken über mehrere Websites hinweg. Das zum Google Konzern gehörende Ad-Network *Doubleclick* funktioniert daher weitaus invasiver als Googles Webanalysedienst. Bindet ein Anbieter ein Werbebanner von Doubleclick in seine Webseite ein, so kontaktiert das dazugehörige Skript die Doubleclick-Server auf der Suche nach einem passenden Werbebanner. Dabei wird dem Nutzer ein eindeutiger Identifikator in einem Cookie gesetzt, mithilfe dessen sich der Nutzer auf allen teilnehmenden Websites verfolgen lässt.

Da der Besucher nicht direkt die Seite von Doubleclick besucht, sondern die des

ursprünglichen Webanbieters, gilt das so von Doubleclick gesetzte Cookie als „Fremdanbieter-Cookie“. Hat der Nutzer pauschal die Annahme von Fremdanbieter-Cookies deaktiviert, würde daher auch das Doubleclick Cookie nicht angenommen. Scheut man diese pauschale Deaktivierung aller Fremdanbieter Cookies kann man mittels Ad-Blocker (z.b. AdBlock Plus) die Server von Werbeanbietern blockieren, bzw. via Cookie Manager (z.b. Cookie-Monster) die Cookies von Doubleclick ablehnen.

Cross-Site Tracking II – Beispiel Popup Ads

In den meisten Browsern kontrolliert das Fremdanbieter-Blocking nur das Setzen, aber nicht das Senden von Cookies (ausser Firefox).[xiv] Falls der Tracker also in der Lage ist, sein Cookie mindestens einmal als Erstanbieter zu setzen, umgeht er damit womöglich das Fremdanbieter Cookie Blocking. Eine Möglichkeit, dies zu erreichen, ist das Öffnen eines Popup-Fensters. Dort kann der Tracker ein Erstanbieter Cookie setzen, welches in den meisten Browsern selbst bei deaktivierten Fremdanbieter-Cookies nun bei allen folgenden Abrufen – auch von eingebetteten Werbebannern – gesendet wird.

Viele Browser haben das Öffnen von Popup-Fenstern per Voreinstellung bereits deaktiviert. Solange solche Popup-Blocker vom Nutzer nicht deaktiviert wurden, kann kein Erstanbieter-Cookie gesetzt werden. Auch wird bei der Verwendung von AdBlock Plug-Ins bzw. Cookie Managern das Setzen des Erstanbieter Cookies bekannter Werbeanbietern bereits komplett unterbunden.

Cross-Site Tracking III – Beispiel Ad-Network

Nicht alle Werbeanbieter haben die Grösse von Doubleclick, einem der grössten digitalen Werbeanbieter weltweit. Um dennoch aussagekräftige Profile bilden zu können, schliessen sich viele Anbieter in Netzwerken zusammen, um so Nutzer über alle von ihren Mitgliedern unterstützten Websites hinweg tracken zu können. Das Werbebanner wird hierbei von einem zentralen Server angefordert (z.b. admeld.com), welcher die Anfrage je nach Anbieter-ID an den eigentlichen Werbeanbieter weiterleitet. Während die Werbung also vom Werbeanbieter kommt, setzt der zentrale Server ein Fremdanbieter Cookie und kann so im Auftrag aller teilnehmenden Werbeanbieter den Nutzer verfolgen.

Während Ad-Networks technisch einen anderen Informationsfluss aufweisen als einzelne Werbeanbieter sind für den Nutzer die Schutzmassnahmen identisch: entweder pauschal die zentralen Server des Netzwerks blockieren (z.B. via AdBlock Plug-In) oder mit einem Cookie Manager die Annahme des zentralen Cookies verweigern.

Cross-Site Tracking IV – Social Widgets

In immer mehr Webseiten sind Schaltflächen (Buttons) von Sozialen Netzwerken eingebunden: Facebook „Likes“, Google’s „+1“, Twitter’s „Tweets“, etc. Wird eine Webseite aufgerufen, die beispielsweise einen Facebook-Button enthält, baut der Browser eine direkte Verbindung mit den Facebook-Servern auf und der Button wird von dort geladen. Dabei wird die Information an Facebook übermittelt, dass die entsprechende Webseite aufgerufen wurde.

Ist der Nutzer bereits Mitglied bei dem eingebetten Sozialen Netzwerks, ist nicht nur das pseudonyme Tracken sondern die exakte Zuordnung des Seitenabrufs zu seinem Profil möglich. So kann das Soziale Netzwerk beim betätigen („liken“, „empfehlen“) der Schaltfläche diese Aktion direkt im Profil des Nutzers veröffentlichen. Facebook kann so gegebenenfalls weitere Nutzungsdaten erheben und speichern. Doch auch ohne einen einzigen Klick wird natürlich zumindest der Seitenbesuch bereits bei dem Sozialen Netzwerk registriert. So können bei Facebook Nutzerprofile entstehen, die über das hinausgehen, was man selbst bei Facebook preisgegeben hat.

Im Gegensatz zu Werbeanbietern wie Doubleclick (deren Webseiten Nutzer typischerweise nie direkt besuchen) haben Soziale Netzwerke wie Facebook in vielen Fällen bereits ein Erstanbieter-Cookie beim Nutzer platziert – bei deren direkten Besuch von facebook.com. Will man weiterhin sein Profil in einem Sozialen Netzwerk pflegen, beim Besuch anderer Seiten jedoch nicht von deren Widgets beobachtet werden, können spezielle Plug-Ins helfen (z.B. Facebook Blocker für Firefox bzw. Facebook Disconnect oder WidgetBlock für Chrome). Auch kann man mit frei verfügbaren Konfigurationslisten

die oben beschriebenen AdBlocker wie AdBlock Plus dazu verwenden.

Fazit

Viele Webseiten tracken ihre Besucher um mehr über die Wirkung ihrer Inhalte zu erfahren. Die gesammelten Daten erfüllen aber nicht nur legitime Funktionen im Interesse der Besucher, sondern werden auch für die Platzierung von Werbung verwendet. Möchte man nicht verfolgt werden, hilft in vielen Fällen bereits das regelmässige Löschen der Cookies im Browser, da dann der Tracker jeweils neue Identitäten vergeben muss.

Schliessen sich Tracker zu Netzwerken zusammen, können sie nicht nur das "Surfen" auf einer Webseite beobachten, sondern den Nutzer in ihrem Netzwerk verfolgen. Fremdanbieter-Cookie-Blocking schützt vor diesen Trackern, solange man deren Seiten nie direkt besucht. Gelingt es aber dem Tracker, den Nutzer auf seine Seite umzulenken, kann er ein Erstanbieter-Cookie setzen, welches bei manchen Browsern das Tracken trotz blockierter Fremdanbieter-Cookies ermöglicht. Mitglieder von sozialen Netzwerken sollten sich deshalb bewusst sein, dass beim Besuch ihres Netzwerkes dieser in ihrem Browser ein Cookie setzt. Social Widgets, von populären sozialen Netzwerken auf Millionen von Webseiten plaziert, hinterlassen auf den Rechnern der Mitglieder Tracking-Information, die ihr Bewegungen protokollieren.

Tracking benützt aber nicht nur HTTP Cookies, sondern es gibt daneben zahlreiche zustandsbasierte („Supercookie“) wie auch zustandslose („Fingerprinting“) Technologien, mit denen Webaktivitäten korreliert werden können. Auch wenn die verfügbaren Blocker-Tools eine hohe Effektivität besitzen, braucht es langfristig Techniken, die ein "Data Leakage" innerhalb des Browsers verhindern.[xv]

Dienstanbieter und Werber wie auch der Konsument brauchen „gleichlange Spiesse“. Dies bedingt, dass man die Motive beider Seiten versteht und Techniken sowie Verhaltensregeln entwickelt, die den Austausch von Gütern und Daten, insbesondere unter dem Aspekt des Datenschutzes, gerecht werden.

Kurz & bündig

Web-Tracking ist heute gang und gebe. Aus Sicht der Webseiten-Besitzer und der Tracker stellt es wünschenswerte Funktionalität wie Personalisierung, Site Analytics und gezielte Werbung zur Verfügung. Je grösser das Benutzerprofil ist, welches ein Tracker anlegen kann, umso besser ist der Dienst, den er seinen Kunden (den Webseiten-Besitzern) und dem Benutzer selbst (Personalisierung) anbieten kann. Aus der Sicht der Benutzer sind grössere Benutzerprofile problematisch, da sie Beziehungen enthalten können, die nicht an Unbekannte gelangen sollten. Dies schliesst insbesondere Third-Parties ein, mit denen kein direkter Kontakt aufgenommen wird, aber das Tracking über mehrere Websites erlaubt. Wie funktioniert Web-Tracking, welche Daten können gesammelt werden und welche Möglichkeiten gibt es, sich dagegen zu schützen?

Literatur

- DeGroef, W., Devriese, D., Nikiforakis, N. und Piessens, F., FlowFox: a Web Browser with Flexible and Precise Information Flow Control, ACM Conference on Computer and Communications Security, 2012
- Eckersley, P., How Unique Is Your Web Browser? Privacy Enhancing Technologies, Lecture Notes in Computer Science 6205, 1-18. Springer, 2010
- Krishnamurthy, B., Naryshkin, K., & Wills, C. E., Privacy leakage vs. Protection measures: the growing disconnect. Web 2.0 Security and Privacy 2011
- Lyon, D., Surveillance as Social Sorting: Privacy, Risk and Automated Discrimination. Taylor & Francis, 2002
- Mayer, J. und Mitchell, J.C., Third-Party Web Tracking: Policy and Technology. IEEE Security & Privacy, 2012
- Roesner, F., Kohno, T. und Wetherall, D., Detecting and Defending Against Third-Party Tracking on the Web. 9th USENIX Symposium on Networked Systems Design and Implementation, 2012
- Solove, D.J. und Hoofnagle, C.J., A Model Regime of Privacy Protection, University of Illinois Law Review 2006(2)

- Soltani, A., Canty, S., Mayo, Q., Thomas, L., und Hoofnagle, C., Flash Cookies and Privacy II: Now with HTML5 and ETag respawning. AAAI Spring Symposium on Intelligent Information Privacy Management 2010.
- Tene, O. und Polonetsky, J., To Track or "Do Not Track": Advancing Transparency and Individual Control in Online Behavioral Advertising, Minnesota Journal of Law, Science & Technology, 2012(1)

Autoren

Marc Langheinrich, Prof. Dr., Università della Svizzera italiana. langheinrich@acm.org

Günter Karjoth, Dr., IBM Research – Zurich, Rüschlikon. karjoth@acm.org

[i] Siehe auch Mayer und Mitchell (2012)
[ii] Krishnamurthy et al (2011)
[iii] http://dev.w3.org/geo/api/
[iv] http://online.wsj.com/public/page/what-they-know-digital-privacy.html
[vii] http://www.cookiecentral.com/faq/
[viii] https://panopticlick.eff.org/; siehe auch Eckersley (2010)
[ix] http://www.bluecava.com
[x] http://www.picnet.com.au
[xi] http://www.w3.org/2011/tracking-protection/
[xii] http://www.w3.org/TR/tracking-compliance/
[xiii] http://www.ghostery.com/
[xiv] Siehe Roesner et al (2012)
[xv] DeGroef et al (2012)