Individuell überwacht*

Dominik Herrmann

28. Februar 2015

Löschen Sie regelmäßig Ihre Cookies, um sich vor unliebsamer Überwachung durch Werbetreibende zu schützen? Das bringt weniger als bislang gedacht. Gewohnheiten und Interessen verraten uns nämlich bei unseren Streifzügen durch das Internet.

Wir alle werden im Internet auf Schritt und Tritt überwacht – und zwar nicht nur von Nachrichtendiensten. Auch die Werbewirtschaft unternimmt enorme Anstrengungen, um herauszufinden, wofür wir uns interessieren. „Im Prinzip schaut uns beim Surfen dauernd jemand über die Schulter", sagt Dr. Dominik Herrmann von der Universität Hamburg. Dort entwickelt er in der Gruppe „Sicherheit in verteilten Systemen" neue Möglichkeiten zur maschinellen Überwachung von Menschen – gewissermaßen arbeitet er den Spähern also zu. Was paradox erscheint, hat durchaus seinen Sinn, denn „nur wenn wir wissen, mit welchen Waffen die Gegenseite kämpft, können wir uns davor schützen", erläutert er. Herrmanns jüngste Untersuchung verschiebt die Frontlinie wieder einmal zugunsten der Überwacher.

Praktisch alle großen Webseiten finanzieren sich durch die Vermarktung von Werbung. Hier kommen die sogenannten Werbenetze ins Spiel. Sie sind darauf spezialisiert, jedem Besucher genau die Anzeigen zu präsentieren, die zu seinen Interessen passen. Jeder Nutzer wird dazu mit einer eindeutigen Nummer markiert, die in einem Cookie gespeichert wird. Cookies sind nicht die einzige Möglichkeit, Nutzer im Internet zu verfolgen. Inzwischen ziehen einige Werbenetze auch weitere Merkmale heran, etwa die Liste der installierten Schriftarten, die einen Rechner häufig einzigartig identifizieren (Browser-Fingerabdruck). Je länger ein Nutzer auf seinen Streifzügen durch das Internet verfolgt werden kann, desto detaillierter ist das resultierende Interessenprofil. Mit dem nötigen Handwerkszeug kann man diesem fragwürdigen Treiben allerdings ein Ende setzen. Die Späher sind zur Überwachung nämlich bislang auf die Mithilfe des Browsers angewiesen.

Und genau in diesem Punkt unterscheidet sich die Überwachungstechnik, die Herrmann in seiner Dissertation vorstellt. Sie benötigt im Gegensatz zu anderen Techniken nämlich keine Unterstützung durch den Brow-

*Dieser Text (http://dhgo.to/diss4) gibt ausgewählte Ergebnisse der Dissertation „Beobachtungsmöglichkeiten im Domain Name System: Angriffe auf die Privatsphäre und Techniken zum Selbstdatenschutz" (http://d-nb.info/1051341248) allgemein verständlich wieder. Eine ausführlichere Zusammenfassung ist unter http://dhgo.to/diss10 abrufbar.

ser. Nicht einmal Experten können daher erkennen, ob eine Vermarktungsfirma sie einsetzt oder nicht. Der Clou: nicht der Browser wird anhand eines eindeutigen Fingerabdrucks wiedererkannt, sondern gleich der Nutzer selbst. Wohlgemerkt geht es dabei nicht um die einzigartigen Muster auf unseren Fingerkuppen, sondern um einen Fingerabdruck im übertragenen Sinn. Herrmann hat nämlich herausgefunden, dass das Surfverhalten vieler Nutzer einzigartige Verhaltensmuster enthält, die eine spätere Wiedererkennung möglich machen. „Wir leben heute in einer Gesellschaft von Individualisten. Die meisten von uns haben Interessen, die in dieser Kombination ziemlich einzigartig sind – und unseren Interessen gehen wir heute immer öfter im Internet nach", führt er aus.

Seine Technik, die er „verhaltensbasiertes Tracking" nennt, funktioniert folgendermaßen: Die Späher sammeln kontinuierlich die Adressen (Domains) aller Webseiten, die jeder Nutzer (identifiziert durch seine aktuelle IP-Adresse) innerhalb einer Internetsitzung abruft. Dazu sind Geheimdienste und Werbenetze immer in der Lage – auch ohne Cookies und ohne Browser-Fingerprinting. Die besuchten Webseiten und die Anzahl der Anfragen, die auf jede Webseite entfallen, werden abgespeichert. Sie stellen den Fingerabdruck des Nutzers dar, also das, was auf unseren Fingerkuppen Schleifen, Wirbel und Minuzien sind. Ein Auszug aus einem solchen Fingerabdruck könnte etwa lauten: „Nutzer ‚17338' war am 27.02.2015 insgesamt 18 Mal auf www.google.de, sieben Mal auf www.welt.de, zwei Mal auf intranet.lufthansa.com und ein Mal auf www.tirol.at".

Nehmen wir an, dass Nutzer 17338 in seiner nächsten Internetsitzung am 28.02.2015 unter einer anderen IP-Adresse auftaucht, er zwischenzeitlich alle Tracking-Cookies gelöscht hat, und dass sein Browser keinen einzigartigen Fingerabdruck besitzt. Eigentlich sollte ein Werbenetz in diesem Fall keine Verbindung zwischen der aktuellen Sitzung und seiner vorherigen Sitzung herstellen können. Beim verhaltensbasierten Tracking gelingt das allerdings trotzdem. Dazu vergleicht das Werbenetz die Webseiten und Anfragehäufigkeiten aus der aktuellen Sitzung des Nutzers mit allen Fingerabdrücken in der Datenbank. Für den Vergleich kommen bewährte Algorithmen aus dem Data-Mining zum Einsatz. Der von Herrmann verwendete Naive-Bayes-Klassifikator wird etwa auch in Spam-Filtern verwendet, um vollautomatisch Werbung auszusortieren. Was ursprünglich für einen guten Zweck ersonnen worden ist, wird jetzt also gegen uns verwendet.

Damit der Klassifikator Nutzer 17338 anhand seines Verhaltens wiedererkennen kann, muss sich seine Sitzung von den Sitzungen der anderen Nutzer ausreichend stark unterscheiden. Das allein reicht jedoch nicht: „Ich war mir von Anfang an zwar ziemlich sicher, dass die meisten Nutzer ein charakteristisches Surfverhalten an den Tag legen. Jeder von uns hat doch gewisse Lieblingsseiten im Internet. Die spannende Frage war, ob wir unsere Lieblingsseiten auch so regelmäßig besuchen, dass wir daran jeden Tag wiedererkannt werden können", erinnert sich Herrmann.

Diese Frage lässt sich nicht im Labor beantworten. Man benötigt einen Datensatz, in dem das Internetverhalten möglichst vieler Menschen enthalten ist. Herrmann bat deutsche Vermarktungsfirmen um Unterstützung. Die gaben sich allerdings zugeknöpft – „Datenschutz, Sie verstehen schon", lautete die einhellige Antwort. Also alles nur graue Theorie? Nein. Zu verdanken ist dies vor allem Martin Wimmer, dem Leiter des Rechenzentrums der Universität Regensburg. Er ließ sich für

Herrmanns Idee begeistern und stellte ihm einen umfangreichen Datensatz mit den DNS-Anfragen aller Studenten und Mitarbeiter zur Verfügung. Darin sind die Domains der besuchten Webseiten enthalten, also genau die Daten, die für die Erzeugung der verhaltensbasierten Fingerabdrücke erforderlich sind.

Mit den gesammelten Daten konnte Herrmann seine Vermutungen überprüfen. Ist das Surfverhalten von Internetnutzern so charakteristisch, dass sich verschiedene Nutzer auseinanderhalten lassen? Treten die Verhaltensmuster regelmäßig genug auf, um Nutzer daran täglich wiederzuerkennen? Herrmann hat zahlreiche Simulationen durchgeführt, um diese Fragen zu beantworten. „Es klappt nicht immer, aber überraschend oft", fasst er seine Ergebnisse zusammen. In einem Experiment sollte der Klassifikator beispielsweise die Sitzungen von mehr als 3800 Studenten über einen Zeitraum von zwei Monaten von einem Tag auf den nächsten verknüpfen. Im Mittel gelang ihm dies in 86% der Fälle. Bei knapp 14% der Studenten wurden sogar alle Sitzungen korrekt verbunden. Auch bei größeren Gruppen gelang das Tracking noch erstaunlich gut: Bei mehr als 12.000 Nutzern wurde noch eine Genauigkeit von 76% erreicht.

„Wie gut das Verfahren bei einem Werbenetz funktioniert, lässt sich anhand dieser Ergebnisse natürlich nicht genau vorhersagen", ergänzt Herrmann. Werbenetze sehen schließlich nur die Anfragen für Webseiten, auf denen sie Werbung vermarkten. Die Großen der Branche, zu denen etwa Googles Vermarkter „Doubleclick" gehört, sind allerdings auf nahezu jeder beliebten Webseite vertreten – und das könnte bereits ausreichen, um Nutzer wiederzuerkennen. So fällt die Genauigkeit nur um sechs Prozentpunkte, wenn man dem Klassifikator nur die 500 beliebtesten Domains zur Bildung der Fingerabdrücke zur Verfügung stellt. Erkennbar sind also auch diejenigen, die nur im Mainstream schwimmen.

Wir Menschen sind nun einmal Gewohnheitstiere – und genau das wird uns beim verhaltensbasierten Tracking zum Verhängnis. Ist verhaltensbasiertes Tracking also die neue Wunderwaffe, gegen die jeglicher Widerstand zwecklos ist? Wie verteidigen wir unsere Privatsphäre gegen dermaßen übermächtige Überwacher? Zunächst denkt man an die bereits bekannten Techniken zum Selbstschutz. Diese haben aber unangenehme Nebenwirkungen. So taucht man zwar in der Masse unter, wenn man einen Anonymisierer wie Tor und einen restriktiv konfigurierten Browser verwendet; allerdings funktionieren einige Webseiten dann nicht mehr ordnungsgemäß und die Ladezeigen steigen erheblich.

Herrmann arbeitet daher an Schutztechniken, die so leichtgewichtig sind, dass man sie gar nicht bemerkt. Aus seinen Untersuchungen weiß er, dass das verhaltensbasierte Tracking nur dann funktioniert, wenn die Späher innerhalb einer Sitzung ausreichend viele Webseiten sammeln können. Dauern die Sitzungen hingegen immer nur wenige Minuten und wechselt man danach konsequent die eigene IP-Adresse, verliert sich die Spur relativ schnell im Sande. Mit Unterstützung des Internetproviders könnte der Adresswechsel allerdings automatisiert werden und ohne Zutun des Nutzers im Hintergrund stattfinden. Herrmann ist derzeit in Gesprächen mit einem Anbieter aus dem Saarland, der sich gut vorstellen kann, seinen DSL-Kunden diese Technik anzubieten.