Eine neue Dimension des Trackings

Wie Webseiten per Session-Replay ihre Nutzer verfolgen

Mehrere Analysefirmen betten eine Technik auf Webseiten ein, mit der Seitenbetreiber so gut wie alles über ihre Besucher herausfinden können. Die Skripte überwachen jeden Tastendruck und jegliche Mausbewegung.

Von Fabian A. Scherschel

Viele Internetnutzer sind sich darüber im Klaren, dass sie auf Webseiten nicht nur vom Seitenbetreiber und dessen Admins, sondern auch von Drittanbietern verfolgt werden können. Dass Seitenaufrufe und Sucheingaben von Tracking- und Werbefirmen analysiert werden, ist mittlerweile Allgemeinwissen. Was allerdings bisher nur die wenigsten auf dem Schirm haben, ist eine mächtige Technik namens Session-Replay. Damit lassen sich alle Tastatureingaben in Echtzeit erfassen und jede Mausbewegung des Besuchers wird nachvollziehbar.

Vor allem Webentwickler wissen, dass moderne Web-Browser jede Eingabe und jede Mausbewegung eines Anwenders bis ins kleinste Detail tracken können. Neuerdings gibt es Analytics-Firmen, die sich auf das Erheben und Auswerten genau dieser Daten spezialisiert haben. Seitenbetreiber, die den JavaScript-Code dieser Firmen in ihre Webseiten eingebettet haben, können so ihre Besucher von Seite zu Seite ihres Webangebotes verfolgen und sehen jegliche Aktionen des Nutzers: Tastatureingaben, Mausbewegungen und alle aufgerufenen Inhalte. Sie können so den gesamten Besuch eines Nutzers im Detail Revue passieren lassen, die Web-Session also quasi wiederholen – daher der Name Session-Replay.

Für Seitenbetreiber hat Session-Replay klare Vorteile: Sie können sehr genau analysieren, was die Besucher auf ihrer Seite treiben. Welche Inhalte finden sie interessant, wo verweilen sie gerne? Welche Eingabefelder sind verwirrend, welche Teile des Webauftritts müssen verbessert werden? Solche und ähnliche Fragen lassen sich mit Session-Replay so gut wie mit keiner anderen Tracking-Technik beantworten.

Dafür opfern sie aber fast jegliche Privatheit ihrer Benutzer. Über Session-Replay lässt sich nämlich nicht nur pixelgenau nachverfolgen, was der Nutzer klickt und wohin er wann scrollt, sondern die Skripte erfassen auch jegliche Tastatureingaben des Anwenders. Dabei ist es völlig egal, ob die Eingaben vom Nutzer an die Seite geschickt werden. Eingaben werden in dem Moment übermittelt, in dem sie passieren – ganz egal ob zum Beispiel der Inhalt eines Textfeldes vor der Übermittlung wieder gelöscht wird.

Datenschutz kommt regelmäßig zu kurz

Die Anbieter solcher Analyseskripte geben an, sensible persönliche Daten vor der Übermittlung an die eigenen Server herauszuredigieren. Laut einer Untersuchung von Forschern der Princeton-Universität in den USA (siehe ct.de/y4by) stimmt das allerdings nur bedingt. Ohne Einblick in die Web-App auf dem Server, auf dem das Analyse-Skript eingesetzt wird, wissen diese Firmen oft viel zu wenig über die erhobenen Daten, um personenbezogene Informationen herauszufiltern.

Passwort-Felder sind leicht an HTML-Tags zu erkennen und werden von den meisten Session-Replay-Skripten ausgenommen. Bei vielen anderen Input-Feldern sind die Analysefirmen allerdings darauf angewiesen, dass die Seitenbetreiber diese im Quellcode der Seite manuell mit Tags versehen. Das ist viel Arbeit und bietet eine Menge Spielraum für Fehler. In der Praxis werden die meisten dieser Daten deswegen wohl früher oder später irgendwo an die Analyse-Server übermittelt. Und die Seiteninhalte, die dem Nutzer angezeigt wurden, werden sowieso immer mitgeschickt. Schon diese enthalten oft sensible Informationen wie Kreditkarteninformationen oder Adressen, bei manchen Seiten sogar Gesundheitsdaten.

Oft scheint die Verschleierung personenbezogener Daten nur pro forma zu erfolgen. Manche der Firmen bieten ihren Kunden sogar explizit an, die Besucher-Sessions mit bekannten E-Mail-Adressen oder Klarnamen der Besucher zu verknüpfen. Alles, damit Webseitenbetreiber möglichst viel über ihre Nutzer herausfinden können. Es liegt schließlich nicht im Interesse der Analysefirmen, die privaten Daten der Web-Nutzer zu schützen. Bei der Menge und Art der über diese Technik erhobenen Daten kann ohnehin niemand ernsthaft davon ausgehen, dass die Privatsphäre des Nutzers gewahrt bleibt.

Verschlüsselungs-Fail

Aber es kommt noch schlimmer: Einige der Firmen übermitteln diese Daten unverschlüsselt – selbst wenn es sich dabei um Informationen handelt, die zwischen Seitenbesucher und Server der Webseite TLS-verschlüsselt waren. Sie exfiltrieren diese Daten also über den eigenen Dienst und eröffnen so die Möglichkeit der passiven Web-Spionage eines Besuchers, der nach eigenem Wissen komplett sicher mit einer Webseite kommuniziert hat.

Im Rahmen ihrer Untersuchung konnten die Princeton-Forscher mehrere Anbieter davon überzeugen, ihren Traffic in Zukunft zu verschlüsseln. Das ändert allerdings nichts an dem massiven Vertrauensbruch, den diese an unwissenden

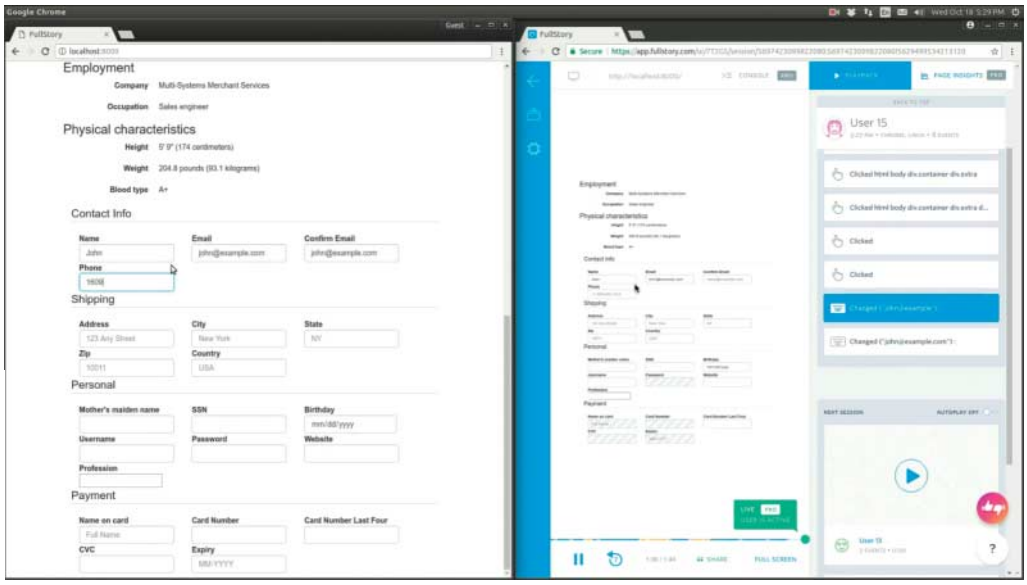

Bild: CITP, Princeton University

Session-Replay bei einem Web-Nutzer in Echtzeit: Links sieht man die Eingaben des Benutzers, rechts das Web-Interface von FullStory, in dem der Seitenbetreiber diese live verfolgen kann.

Nutzern begangen haben. Und genauso wenig geht es das Problem an, dass Session-Replay-Anbieter Unmengen an privaten Daten horten und somit ein lohnendes Ziel für kriminelle Hacker oder Geheimdiensteingriffe sind.

Session-Replay wird immer beliebter

Noch ist Session-Replay nicht weitläufig im Einsatz, aber die Princeton-Forscher warnen vor einem Aufwärtstrend. Bei ihrer systematischen Untersuchung zehntausender der am meisten besuchten Webseiten fanden sie 482 Sites, vor allem aus den USA, welche die Skripte der sieben beliebtesten Session-Replay-Anbieter einsetzen. Sie testeten auf JavaScript-Code der Firmen Clicktale, FullStory, Hotjar, UserReplay, SessionCam, Smartlook und der russischen Suchmaschine Yandex. Die Forscher vermuten allerdings eine hohe Dunkelziffer an Seiten, die sie zwar getestet haben, die ihnen allerdings nicht ins Netz gingen. Das liegt vor allem daran, dass viele Webseiten Session-Replay nicht bei jedem Besucher aktivieren und die Skripte selektiv in ihre Seiten einbetten.

Aus der Sicht von Webseitenbetreibern ist es leicht nachzuvollziehen, warum diese Technik immer beliebter wird. Session-Replay ist ein mächtiges Analysewerkzeug, das es Seitenbetreibern, Admins und Webdesignern ermöglicht, die Schwachstellen bei Inhalten und Design ihres Webangebotes genau zu identifizieren und dann gegenzusteuern. Softwarehersteller zahlen normalerweise gutes Geld, ihre Entwicklungen in Fokusgruppen einer Auswahl an Nutzern zu präsentieren, denen sie dann bei der Benutzung über die Schulter schauen können. Mit Session-Replay bekommen Webentwickler den gleichen Effekt viel billiger und sie haben weit mehr Nutzer-Sessions zur Verfügung, die sie nach Belieben analysieren können.

Nutzer, die Praktiken wie Session-Replay einen Riegel vorschieben wollen, können dies in der Theorie recht einfach bewerkstelligen. Die Skripte der verschiedenen Anbieter ließen sich mit diversen Skript-Blockern an der Ausführung hindern. Im Regelfall beeinflusst das den Rest der Webseite überhaupt nicht. In der Praxis scheitert das allerdings daran, dass dazu die verschiedenen Web-Domains der Anbieter bekannt sein müssten: Bisher existiert keine umfassende Liste, die es mit einem Klick ermöglicht, alle bekannten Dienste zu blocken.

Eine andere Lösung besteht darin, Skripte von Drittanbietern pauschal zu unterdrücken und nur selektiv zu erlauben. Ein solcher Whitelisting-Ansatz erfordert allerdings eine Menge Verwaltungsaufwand vom Nutzer und führt anfangs dazu, dass viele Webseiten nicht oder nur sehr eingeschränkt benutzbar sind. Früher oder später sollte das aktuelle Medieninteresse an Session-Replay allerdings auch dazu führen, dass Nutzer und Entwickler von Skript-Blockern die Server der Anbieter in bekannte Blocklisten aufnehmen. Vor allem wenn sich die Voraussage der Princeton-Forscher bewahrheitet und Session-Replay in Zukunft immer breiter zum Einsatz kommt.

(fab@ct.de) c't

Princeton-Studie: ***ct.de/y4by***