

Datenschutz beim Website-Tracking

Informatiker des KIT und der TU Dresden untersuchen, wie gut das Generalisieren von Trackingdaten unsere Spuren im Internet verwischt



Beim Surfen im Internet sammeln Unternehmen nicht nur Daten über besuchte Webseiten, sondern auch über den Zeitpunkt des Abrufes oder Ortsinformationen. (Bild: Amadeus Bramsiepe, Markus Breig ,KIT)

Das Tracking des Surfverhaltens gehört zum Alltag der Internetnutzung. Unternehmen verwenden es beispielsweise, um Werbeanzeigen auf die persönlichen Belange der potenziellen Kunden zuzuschneiden oder ihre Reichweite zu messen. Viele Anbieter von Trackingdiensten werben mit sicherem Datenschutz, indem sie die Datensätze generalisieren und so anonymisieren. Wie sicher dieses Verfahren ist, haben Informatikerinnen und Informatiker des Karlsruher Instituts für Technologie (KIT) und der Technischen Universität Dresden (TUD) nun untersucht und anlässlich der IEEE Security and Privacy-Konferenz in einem wissenschaftlichen Paper veröffentlicht.

Trackingdienste sammeln große Datenmengen der Internetnutzerinnen und -nutzer. Darunter fallen neben den besuchten Webseiten beispielsweise auch Informationen zu den verwendeten Endgeräten, der Zeitpunkt des Abrufes (Zeitstempel) oder Ortsinformationen. „Da diese Daten sehr sensibel sind und einen

Monika Landgraf
Leiterin Gesamtkommunikation
Pressesprecherin

Kaiserstraße 12
76131 Karlsruhe
Tel.: +49 721 608-21105
E-Mail: presse@kit.edu

Weiterer Pressekontakt:

Sandra Wiebe
Redakteurin/Pressereferentin
Tel.: +49 721 608-21172
E-Mail: sandra.wiebe@kit.edu

hohen Personenbezug haben, nutzen viele Unternehmen die Generalisierung, um sie scheinbar zu anonymisieren und damit Datenschutzregelungen zu umgehen“, sagt Professor Thorsten Strufe, Leiter der Forschungsgruppe „Praktische IT-Sicherheit“ am KIT. Bei einer Generalisierung wird der Detailgrad der Informationen reduziert, sodass eine Identifizierung von Einzelpersonen nicht mehr möglich sein soll. So werden beispielsweise die Ortsinformationen auf die Region beschränkt, die Abrufzeit auf den Tag oder die IP-Adressen um einige Zahlen gekürzt. Ob so wirklich keine Rückschlüsse mehr auf das Individuum gezogen werden können, hat Strufe gemeinsam mit seiner Forschungsgruppe und Kolleginnen und Kollegen der TU Dresden untersucht.

Mithilfe einer Vielzahl an Metadaten deutscher Webseiten mit etwa 66 Millionen Nutzern und über 2 Milliarden Seitenaufrufen konnten die Informatikerinnen und Informatiker nicht nur Rückschlüsse auf die aufgerufenen Seiten, sondern auch auf die Verkettung der einzelnen Seitenaufrufe, sogenannten „Click Traces“, ziehen. Die Daten stellte ihnen die INFOnline GmbH, eine Institution für Reichweitenmessung in Deutschland, zur Verfügung.

Der Verlauf von Seitenaufrufen hat große Aussagekraft

„Um die Wirksamkeit der Generalisierung zu testen, haben wir zwei unterschiedliche Anwendungsszenarien betrachtet“, sagt Strufe. „Zum einen haben wir die gesamten Click Traces auf ihre Eindeutigkeit untersucht. Denn ist ein Click Trace, also der Verlauf vieler aufeinanderfolgender Seitenaufrufe, klar von anderen abgrenzbar, so ist er nicht mehr anonym.“ Dabei zeigte sich, dass Informationen zur besuchten Webseite und benutztem Browser komplett aus den Daten entfernt werden müssen, um Rückschlüsse auf Personen zu vermeiden. „Die Daten werden erst dann anonym, wenn die Sequenzen entweder zu einzelnen Klicks verkürzt, also völlig ohne Zusammenhang gespeichert werden, oder alle Informationen mit Ausnahme des Zeitstempels entfernt werden“, so Strufe. „Selbst wenn lediglich die Domain, die Themenzuordnung, wie „Politik“ oder „Sport“, und die Zeit nur tagesgenau gespeichert werden, können 35 bis 40 Prozent der Daten individuellen Personen zugeordnet werden.“ In diesem Szenario konnten die Forscherinnen und Forscher zeigen, dass der Ansatz der Generalisierung nicht der Definition der Anonymität entspricht.

Wenige Beobachtungen reichen, um Nutzerprofile zu identifizieren

Die Wissenschaftlerinnen und Wissenschaftler haben außerdem untersucht, wie man auch nur mit Teilmengen eines Click Traces Rückschlüsse auf ein Individuum ziehen kann. „Wir haben die

generalisierten Informationen aus der Datenbank mit weiteren Beobachtungen wie auf den Sozialen Medien oder in Chats geteilten Links verbunden. Wird beispielsweise die Zeit auf eine Minute genau generalisiert, reicht mit dieser Methode eine Beobachtung, um über 20 Prozent der Click Traces eindeutig einer Person zuzuordnen“, sagt Clemens Deusser, der als Doktorand in Strufes Arbeitsgruppe maßgeblich an der Studie beteiligt war. „Zwei weitere Beobachtungen steigern diesen Erfolg auf über 50 Prozent. In der Datenbank kann dann einfach abgelesen werden, welche anderen Webseiten die Person noch besucht und welche Inhalte sie betrachtet hat.“ Selbst wenn der Zeitstempel nur tagesgenau gespeichert werde, benötige man für die Personenerkennung nur fünf weitere Beobachtungen.

„Unsere Ergebnisse zeigen, dass einfache Generalisierung nicht geeignet ist, um Webtrackingdaten wirksam zu anonymisieren. Die Daten bleiben personenscharf und die Anonymisierung ineffektiv. Um einen effektiven Datenschutz zu erreichen, müssten Verfahren angewandt werden, die darüber hinausgehen, wie beispielsweise eine Verrauschung durch zufälliges Einfügen kleiner Fehlbeobachtungen in die Daten“, so Strufes Empfehlung.

Das Forscherteam um Strufe hat seine Ergebnisse innerhalb der „IEEE Security and Privacy“ vom 18. bis 20. Mai 2020 vorgestellt. Dabei handelt es sich um die führende internationale Konferenz im Bereich der IT-Sicherheit und versammelt bereits seit 1980 international renommierte Spitzenforscherinnen und -forscher.

Erläuterung der Ergebnisse im Video:

<https://www.youtube.com/watch?v=IhYbNBnMYgE>

Als „Die Forschungsuniversität in der Helmholtz-Gemeinschaft“ schafft und vermittelt das KIT Wissen für Gesellschaft und Umwelt. Ziel ist es, zu den globalen Herausforderungen maßgebliche Beiträge in den Feldern Energie, Mobilität und Information zu leisten. Dazu arbeiten rund 9 300 Mitarbeiterinnen und Mitarbeiter auf einer breiten disziplinären Basis in Natur-, Ingenieur-, Wirtschafts- sowie Geistes- und Sozialwissenschaften zusammen. Seine 24 400 Studierenden bereitet das KIT durch ein forschungsorientiertes universitäres Studium auf verantwortungsvolle Aufgaben in Gesellschaft, Wirtschaft und Wissenschaft vor. Die Innovationstätigkeit am KIT schlägt die Brücke zwischen Erkenntnis und Anwendung zum gesellschaftlichen Nutzen, wirtschaftlichen Wohlstand und Erhalt unserer

natürlichen Lebensgrundlagen. Das KIT ist eine der deutschen Exzellenzuniversitäten.

Diese Presseinformation ist im Internet abrufbar unter:
www.sek.kit.edu/presse.php

Das Foto steht in der höchsten uns vorliegenden Qualität auf www.kit.edu zum Download bereit und kann angefordert werden unter: presse@kit.edu oder +49 721 608-21105. Die Verwendung des Bildes ist ausschließlich in dem oben genannten Zusammenhang gestattet.