Auf Spurensuche im Internet

Korrelation personenbezogener Daten für digitale Dossiers

Evren Eren

Jeder, der sich im Internet bewegt, hinterlässt Spuren – sei es durch eine schlichte Suchanfrage oder durch Aktivitäten in Social Networks. Dies birgt erhebliche Gefahren. Ein Großteil der Nutzer bedient gleich mehrere Plattformen mit unterschiedlichen oder mit genau den gleichen Informationen. Rein technisch lassen sich über diese verschiedene digitale Teilidentitäten oder gar umfangreiche Gesamtprofile erstellen.

Laut Statista ist ein Internetnutzer durchschnittlich in mehr als sechs sozialen Netzen aktiv, bei den 16- bis 24-Jährigen sind es sogar sieben [1] und hinterlässt Unmengen an persönlichen Informationen. Eine Studie über Jugendliche im Alter zwischen zwölf und 24 Jahren im Auftrag der Landesanstalt für Medien Nordrhein-Westfalen [2] zeigt, welche persönlichen Informationen preisgegeben werden. Etwa 90 % der Befragten veröffentlichen Vornamen und Geschlecht. 60 bis 80 % geben auch das Geburtsdatum, ein Profilfoto sowie Interessen und Hobbys an. Bei sehr persönlichen Daten sind Jugendliche dagegen zurückhaltender [2].

Verwertung der Daten

Personenbezogene Daten sind zu einer heiß umkämpften Ware geworden. Unternehmen nutzen die Möglichkeit, gigantische neue Reichweiten für Werbemaßnahmen präzise einzugrenzen. Aus angesammelten Daten lassen sich Wahrscheinlichkeiten für künftige Kaufentscheidungen berechnen. Je nach Nutzungsverhalten wird eingestuft, ob Inhalte für den Nutzer relevant sind oder nicht. Dies geschieht maßgeblich in Form von Werbeanzeigen, die spezifisch auf den Nutzer abgestimmt werden.

Neben dieser direkten Beeinflussung der Nutzer kann es jedoch auch zu viel weitreichenderen Folgen kommen. So können sich Unternehmen über potenzielle neue Mitarbeiter ein viel umfassenderes Bild über das Berufs- und Privatleben einholen [3]. Für Personalabteilungen ist es mittlerweile Usus, Bewerber vor dem Bewerbungsgespräch zu googeln. Und Kreditunternehmen prüfen mittels Social Scoring die Vertrauenswürdigkeit ihrer Kunden. Spezielle Scoring-Unternehmen veräußern solche Daten an Arbeitgeber, Vermieter usw. Es werden Profile erstellt, die Aussagen zum Arbeitgeber, psychischen Zustand, Ausgehverhalten, zu Konsumverhalten und Andeutungen von einer Schwangerschaft beinhalten können [4]. Sie gewähren Zugriff auf Social-Media-Accounts – nicht nur auf öffentliche Informationen, sondern auch auf private Bereiche und Nachrichten. Neben der Wirtschaft sind personenbezogene Daten schließlich auch für den Staat und dessen Sicherheitsorgane von Interesse.

Korrelation von Daten

Da die wenigsten in nur einem sozialen Netz angemeldet sind, steigt die Datenfülle mit jeder neuen Plattform. Selbst wenn Nutzername und E-Mail-Adresse nicht in jedem der Netze identisch sind, lässt sich mittels entsprechender Algorithmen nur aufgrund von Übereinstimmungen weniger Merkmale Rückschluss auf ein und dieselbe Identität ziehen. Flankierend kommen Daten aus Portalen wie Onlinemärkten, Suchmaschinen usw. Überflüssig wird eine solche Verknüpfung, wenn sich Nutzer mit einem Account z.B. bei Facebook oder Google in anderen Webseiten anmelden. Diese Art von Single Sign-on führt automatisch dazu, dass Anbieter, die die Verifizierung über die eigenen Login-Daten anbieten, weitere Informationen über Webseitenbesuche außerhalb von Facebook erhalten. Ursprünglich anonyme Daten lassen sich in Verbindung mit tatsächlichen, den Nutzer identifizierenden Daten bringen. Es ist nicht unüblich, dass dabei zusätzlich das Bewegungsverhalten der Maus, Tastaturanschläge, Schlagworte in unverschlüsselten Chats, Bewegungsprofile und Suchprofile aufgezeichnet werden.

Jede Aktivität im Netz bildet dabei eine Teilidentität [5]. Sie wird gebildet aus Profilen in sozialen Netzen, Onlineeinkäufen, Login-Daten von Portalen, Bewertungen usw. Diese Teiliden-

Interessierten NET-Abonnenten steht das umfangreiche Literaturverzeichnis im Heftarchiv 12/17 unter www.NET-im-web.de bereit.

Prof. Dr. Evren Eren ist Professor am Lehrstuhl für IT-Sicherheitsarchitekturen an der Hochschule Bremen

titäten lassen sich korrelieren und zeichnen in ihrer Gesamtheit ein ziemlich detailliertes Bild der Person. So entstehen ganze „digitale Dossiers".

Die größten Datenkraken

Google ist im Ranking der Datenkraken die Nummer Eins. Das Unternehmen hat zwar seit 2011 mit Google+ auch ein eigenes soziales Netz, ist aber mittlerweile mit verschiedensten Webdiensten omnipräsent. Über diese sammelt das Unternehmen demografische Merkmale der Nutzer, Interessen, Gewohnheiten und über Android-Smartphones auch aktuelle Standorte, mit denen Bewegungsprofile möglich sind. Hierdurch ist die Verkettung der im jeweiligen Dienst preisgegebenen Daten der Nutzer ohne Probleme möglich, meist jedoch noch nicht mal notwendig, da Nutzer die Möglichkeit haben, mit nur einem Login sich in allen Diensten anzumelden. Neben den Daten, die das Unternehmen auf seinen eigenen Seiten sammelt, speichert es auch Daten von anderen Seiten. Analog zum Like-Button von Facebook reicht es aus, wenn auf einer Seite ein Youtube-Video oder eine Google Map eingebettet ist, damit Google über Besuche dieser Seiten informiert wird. Auch über Google-Ads geschaltete und in die Webseite eingebundene Werbung liefert Google diese Informationen. Zudem werden mit Google Analytics verschiedene Tools angeboten, die anderen Webseitenbetreibern die Messung und Analyse von Interaktionen auf ihrer Webseite via Tracking ermöglichen.

2015 waren allein 80 % aller deutschen Internetnutzer Mitglied bei Facebook [1]. Facebook sammelt Daten zu sozialen Verbindungen und fordert den Nutzer dafür immer wieder auf, auf das Adressbuch zugreifen zu dürfen. Damit „schenkt" der Nutzer Facebook zusätzliche Daten von Nichtmitgliedern, die dann sogleich ins richtige Umfeld einsortiert werden, wenn sie sich anmelden [6]. Verbindungen zu anderen Profilen werden jedoch auch darüber hergestellt, dass Nutzer Inhalte von anderen Nutzern liken, teilen oder kommentieren. All diese Verknüpfungen speichert Facebook schließlich in einem Social Graph, einem riesigen Netz aus Verbindungen zwischen Nutzern. Neben den sozialen Verknüpfungen werden auch alle Hinweise auf Verknüpfungen zu bestimmten Themen gesammelt. Hierzu wird jede Aktivität minutiös beobachtet und gespeichert. Das Ausklappen von Beiträgen, Betrachten von Fotos über die Einzelansicht, Einblenden verborgener Kommentare reichen dafür aus. Facebook sammelt personenbezogene Daten auch über die eigene Domäne hinaus, z.B. auf der eigenen Webseite mittels eines Like-Buttons durch das Facebook Social Plugin. Weder ist es vonnöten, diesen Button anzuklicken, noch Facebook-Mitglied zu sein, damit personenbezogene Daten an Facebook übermittelt werden.

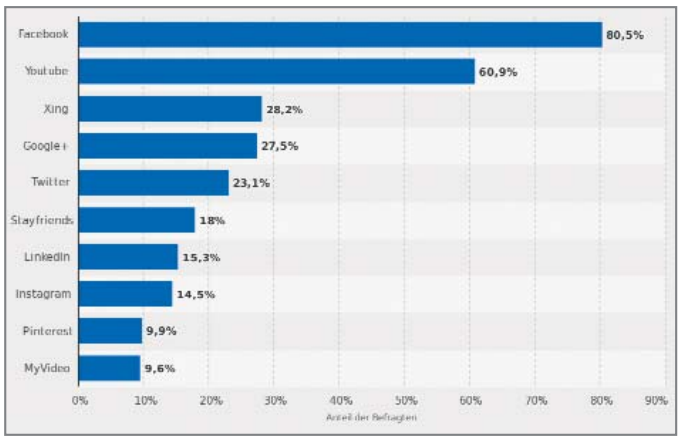

Bild 1: Verbreitung der Nutzung von sozialen Netzen im Jahr 2015 (Quelle: Statista)

Facebook speichert täglich über 600 Tbyte Daten und pflegt weltweit die größte existente biometrische Datenbank [7], [8]. Mit der Übernahme von Whatsapp im Jahr 2014 hat das Unternehmen noch mehr Daten. Künftig werden damit die Metadaten von zwei der weltweit wichtigsten Kommunikationsprogramme von einem einzigen Unternehmen verwaltet [9].

Techniken zur Datensammlung

Cookies

Cookies speichern Informationen zunächst einmal ohne Personenbezug. Es sind kleine Informationshäppchen, die für die Personalisierung der Webseite genutzt werden können [10]. Doch über Korrelation lässt sich der Nutzer identifizieren. Wurden Angaben zum Geschlecht, Postleitzahl und sogar Geburtsdatum gemacht, ist der Rest einfach. Cookies von Drittanbietern sind effizienter und damit gefährlicher. Wurden verschiedene Webseiten mit demselben Werbebanner-Cookie besucht, kann über diesen ermittelt werden, welche Webseiten das wa-ren. Große Werbeanbieter wie Google, die Werbeanzeigen auf tausenden von Webseiten schalten, erhalten über die Verknüpfung der Informationen aus den einzelnen Cookies ziemlich umfangreiche Bewegungsprofile und damit Auskunft über Vorlieben, Gewohnheiten, Interessen bis hin zum psychologischen Befinden des Nutzers. Zudem wird auch der Erfolg der Werbeanzeigen ausgewertet und die Zielgruppe angepasst. Über Cookies von Drittanbietern wird eine Vielzahl von Cookies auf dem Rechner des Nutzers von ganz unterschiedlichen Servern gespeichert, obwohl der Nutzer nur die eine Webseite besucht hat. Über den Browser gibt es zwar die Möglichkeit, das Speichern von Cookies zu verhindern. Jedoch sind mittlerweile viele Webseiten so konfiguriert, dass sie ohne das Einverständnis zu Cookies nicht richtig nutzbar sind.

Webcrawler und Harvester

Webcrawler und Harvester durchsuchen Webseiten und indexieren, filtern und extrahieren Daten wie Webfeeds und E-Mail-Adressen. Auch Facebook-Seiten werden durch solche Crawler durchsucht und Profile als Sucheinträge aufgelistet. So können Facebook-Profile auch von jemandem gesucht werden, der gar nicht bei Facebook angemeldet ist. Diese „Public Search Listings" bestehen aus dem Profilfoto und Links zu den Profilen von weiteren Freunden und könnten in den Privateinstellungen aktiviert sein [11].

Social Plugins

Nahezu jede etwas größere Webseite hat mittlerweile ein Social Plugin inte-

griert, um eigene Inhalte mit sozialen Netzen zu verbinden. Like- oder Share-Button von Facebook, Tweet- oder Follow-Button von Twitter, Pint-it-Button von Pinterest sind Beispiele hierfür. Der Besuch einer Webseite mit eingebundenem Facebook-Like-Button versorgt Facebook mit Informationen über den Seitenaufruf, unabhängig davon, ob der Nutzer selbst Facebook-Mitglied ist oder nicht. Klickt er den Like-Button an, wird diese Aktion auf seinem Profil veröffentlicht [12]. Es können also personenbezogene Daten auch über die eigenen Webseiten hinaus gesammelt werden, wenn ein Social Plugin auf einer beliebigen Webseite integriert ist. Besucht nun ein eingeloggter Facebook-Nutzer die Webseite, wird über diesen Code eine Verbindung zwischen Facebook und dem Browser des Benutzers hergestellt. Über das beim Einloggen des Nutzers abgelegte Cookie werden schließlich der Nutzer und der Login-Status identifiziert.

Bild- und Gesichtserkennung

Hochgeladene Fotos lassen sich durch Bilderkennungsalgorithmen maschinell auswerten. Bekannt ist dies von Facebook durch das Zuweisen von Namen zu ermittelten Gesichtern. Neben Gesichtern können über eine bildinhaltsbasierte Suche auch die umgebenden Bestandteile (Einrichtungsgegenstände, Umgebung, Sehenswürdigkeiten usw.) des Bildes erkannt werden, so dass zusammen mit dem umgebenden Text mehr Metadaten wie z.B. der Aufenthaltsort dem „digitalen Dossier" hinzugefügt werden können [13].

Sowohl Google als auch Facebook haben jeweils eine eigene App auf den Markt gebracht, die mit dem Auswerten von Fotos wirbt. Google Fotos und Facebook Moments analysieren die Fotos auf ihre Inhalte und versehen sie sowohl mit verschiedenen Schlagworten als auch mit den abgebildeten Personen. Die Algorithmen von Google Fotos sind bei der Gesichtserkennung so stark, dass sie angeblich sogar Kinder- und Erwachsenenbilder derselben Person zuordnen können.

Ortung

Ebenfalls werden Ortungsdienste durch

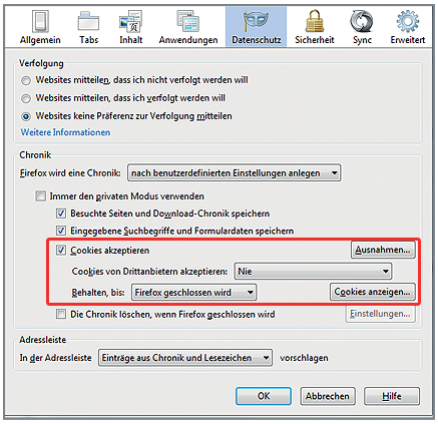

Bild 2: Einstellungen in Firefox

immer mehr Apps genutzt. Auch wenn daraus ortsspezifische Inhalte und Vorschläge geniert werden, ist der Nutzen eher fraglich. Zudem gibt es für Entwickler die Möglichkeit, auf Programmierschnittstellen von sozialen Netzen zuzugreifen. Über diese kann die eigens programmierte Anwendung Nutzerdaten mit dem sozialen Netz austauschen. Onlinespiele können damit z.B. Profilfoto und Namen des Nutzers verwenden [12].

Gegenmaßnahmen

Rein technisch haben Nutzer die Möglichkeit, Einstellungen an ihrem Browser vorzunehmen und diesen mit Addons zu erweitern, um zumindest eine Aktivitätsverfolgung durch bestimmte Tracking-Dienste zu unterbinden. Doch viele Addons schaden mehr als sie nutzen, z.B. die Browser-Erweiterung Web of Trust (WOT). Derzeit kann man das plattformübergreifende und quelloffene Browser-Addon uBlock empfehlen. Es steht für diverse Browser zur Verfügung und blockt mittels Filterlisten Werbung, Tracker, Malwareseiten oder Social Sharing Buttons.

Bild 2 zeigt Datenschutzeinstellungen des Firefox-Browsers gemäß der Empfehlung vom „Unabhängigen Landeszentrum für Datenschutz Schleswig-Holstein". Hiernach sollten Cookies von Drittanbietern niemals akzeptiert und beim Beenden des Browsers gelöscht werden [14].

Des Weiteren eignen sich Tools zur Deanonymisierung wie z.B. TOR (s. NET 9/2015, S. 10). Auch einfache Maßnahmen wie das Nutzen von Pseudonymen in Kombination mit dem Eingrenzen persönlicher Attribute vermindern die Wahrscheinlichkeit zur Korrelationsbildung. Eine weitere Gegenmaßnahme ist das „Drei-Browser-Konzept", bei dem drei verschiedene Browser für unterschiedliche Zwecke Einsatz finden [15].

Empfohlen wird das Ersetzen von Social Plugins durch die Shariff-Lösung, mit der Buttons als HTML-Links in die Webseite eingebaut werden. Im Gegensatz zu den üblichen Share-Buttons wird nicht direkt bei Besuch einer Webseite getrackt, denn der Shariff-Button verbindet den Besucher erst dann mit dem sozialen Netz, wenn er aktiv auf den Share-Button klickt [16].

Die rein technischen Einstellungen sind jedoch nur ein kleiner Schritt in Bezug auf den Datenschutz. Viel wichtiger ist es, dass Nutzer einen verantwortungsvollen und bewussten Umgang mit ihren persönlichen Daten pflegen. Betreiber sozialer Netze haben wenig Interesse daran, Nutzern die Veröffentlichung von personenbezogenen Daten zu erschweren bzw. zu verbieten.

Fazit

In Zukunft werden immer mehr Daten über den einzelnen Menschen von Dritten gesammelt und gespeichert. Alle Daten, die Nutzer im Internet von sich preisgeben, auf welche Art und auf welchen Umwegen auch immer, sind der Erstellung eines umfassenden Nutzerprofils dienlich und daher für die Unternehmen oder auch den Staat interessant. Daten aus den einzelnen sozialen Netzen sind dabei wie Puzzleteilchen, die in ihrer Korrelation ein Gesamtbild des Nutzers darstellen und als „digitales Dossier" einen immensen Marktwert haben. Datenschutz und Privatsphäre leiden darunter, da niemand mehr die völlige Kontrolle über die Weitergabe und Verwendung persönlicher Daten haben wird. Eine Vielzahl technischer Gegenmaßnahmen bietet sich an. Leider greifen auch die politischen Bemühungen in Sachen Datenschutz nicht ausreichend. Deshalb muss man dieser Entwicklung dringend mit Sensibilisierung und Bewusstsein entgegenwirken. *(bk)*