IRINA TARANU, SEBASTIAN LABITZKE, HANNES HARTENSTEIN

ZWISCHEN ANONYMITÄT UND PROFILING: EIN TECHNISCHER BLICK AUF DIE PRIVATSPHÄRE IN SOZIALEN NETZWERKEN

1. Einführung

Internetnutzer integrieren Web-2.0-Dienste, insbesondere soziale Netzwerke (engl.: Online Social Networks, OSNs), immer häufiger in ihren Alltag, sei es für private oder auch für berufliche Zwecke. OSNs bieten den Nutzern diverse Dienste zur Kommunikation mit anderen Mitgliedern des Netzwerks (E-Mail, Chat, Pinnwände etc.) sowie eine Plattform zur in Nutzerkreisen gern gesehenen Möglichkeit der Selbstdarstellung (persönliche Informationsseiten, Foto- und Videoalben etc.). Auch die Aggregation von Neuigkeiten und Informationen zu Themen, die durch den Nutzer nach individuellem Interesse ausgewählt werden können, sowie unterhaltungsmediale Angebote in Form von Spielen und Anwendungen (engl. Applications, Apps) werden durch OSNs geboten. In den vergangenen Jahren hat insbesondere das OSN Facebook[1] auch in Deutschland eine Vormachtstellung erlangt.

Für die Nutzer wird die Teilnahme an sozialen Netzwerken als eine Bereicherung des alltäglichen Lebens wahrgenommen, auf die viele Menschen mittlerweile nicht mehr verzichten möchten. Die Teilnahme am OSN und damit das Teilen von Informationen, Fotos, Videos und letztendlich auch oft sensiblen, personenbezogenen Daten führen oft zu Selbstbestätigungen. Das persönliche Online-Netzwerk „belohnt" durch Bestätigungen und Kommentare die durch den Nutzer getätigten Veröffentlichungen von Informationen. Dieses Belohnungsmuster kann aus technischer sowie psychologischer Sicht diskutiert werden.[2] Dabei muss hinterfragt werden, ob dem Drang zur Veröffentlichung von Informationen die Perzeption des Risikos gegenübergestellt werden kann.[3]

---

1 https://www.facebook.com/.

2 Vgl. auch Oliver Günther/Alexander Kovrigin/Aneta Nowobilska/Hanna Krasnova/Thomas Hildebrand, „Why Participate in an Online Social Network: An Empirical Analysis", in: *16th European Conference on Information Systems, ECIS '08*, Galway, 2008.

3 Sebastian Labitzke/Jochen Dinger/Hannes Hartenstein, „How I and Others Can Link My Various Social Network Profiles as a Basis to Reveal My Virtual Appearance", in: *Lecture Notes in Informatics (LNI - Proceedings, GI-Edition)*, 4. DFN Forum Kommunikationstechnologien, Bonn, 2011, S. 123-131.

Gegenüber dem Gewinn für den Nutzer stehen demnach die Gefahren, die eine Teilnahme und insbesondere ein allzu freigiebiger Umgang mit sensiblen Informationen in OSNs mit sich bringen. Vor dem Hintergrund, dass die Privatsphäre im Netz vielfach schlicht aufgegeben wird, hat Daniel J. Solove untersucht, welche Fehlannahmen hinter der oft gehörten „Ich habe nichts zu verbergen"-Einstellung der OSN-Nutzer stecken. Ergebnis der Studie ist, dass letztendlich auch die Veröffentlichung von zunächst wenig prekär scheinenden Informationen zu Nachteilen für den Nutzer führen kann.[4] Generell gilt, je mehr Informationen Dritten über einen Nutzer bekannt sind, desto eher können daraus nachteilige Auswirkungen für die betreffende Person resultieren. Werden Nutzerprofile mit Daten zusammengeführt, die verraten, wie Nutzer das Internet verwenden, welche Inhalte sie abrufen und welche Dienste sie nutzen, verrät das aggregierte Profil Dritten eine weitere Fülle von Informationen. Letztendlich gelangen immer mehr persönliche Details der Nutzer ins Internet bzw. in die Hände Dritter – eine Entwicklung, die nicht zuletzt durch das Nutzungsverhalten und durch die Möglichkeiten, Daten der Nutzer in Beziehung zu setzen, vorangetrieben wird.

Durch die wachsende Integration der Web-2.0-Anwendungen und der zunehmenden Personalisierung von im Internet angebotenen Diensten ergibt sich eine Vielzahl von Herausforderungen. Dabei unterscheidet sich das Verhalten der Nutzer im Internet maßgeblich vom Verhalten in der alltäglichen Face-to-face-Interaktion. Ein durchschnittlicher Nutzer ist im Internet weitaus offener und gibt eine große Menge personenbezogener Daten von sich preis.[5] Informationen, mit denen der Nutzer im realen Leben unter Umständen sehr restriktiv umgeht und nur ausgewählten Personen offenbart, gibt er im Internet zum Teil freiwillig bekannt, ohne sich dem Ausmaß vollständig bewusst zu sein. Durch diesen offenen Umgang der Nutzer mit personenbezogenen Daten entwickeln sich potenzielle Risiken, denen der Nutzer ausgesetzt ist. Klassische Beispiele sind Spamming, Phishing, Identitätsdiebstahl und vieles mehr.[6] Folgen der Preisgabe persönlicher Informationen im Internet sind die personalisierte Schaltung von Werbung durch zum Beispiel Werbebanner, angezeigt beim Besuchen verschiedener Webseiten, das Senden von Newslettern oder die Hervorhebung von Produkten, die für den Nutzer von Interesse sein könnten. Die Daten der Nutzer können selbst von Dienstanbietern, bei denen der Nutzer die Daten hinterlegt, für verschiedene Marketingzwecke genutzt oder aber an Dritte verkauft werden. Ein in diesem Kontext oft unterschätztes, zusätzliches

---

4 Daniel J. Solove, „‚I've Got Nothing to Hide' and Other Misunderstandings of Privacy, in: *San Diego Law Review*, 44 (2007). (GWU Law School Public Law Research Paper No. 289.)

5 Sebastian Labitzke/Jochen Dinger/Hannes Hartenstein, „What Your Friends Tell Others about You: Low Cost Linkability of Online Social Network Profiles", in: *5th International ACM Workshop on Social Network Mining and Analysis*, San Diego, CA, 2011.

6 Giles Hogben, *ENISA Security Issues and Recommendations for Online Social Networks, ENISA Position Paper for W3C Workshop on the Future of Social Networking*, 2007, online unter: http://www.w3.org/2008/09/msnws/papers/Future_of_SN_Giles_Hogben_ENISA.pdf.

potenzielles Risiko, welches sich nicht nur auf Informationen des Nutzers bei *einem* Anbieter beschränkt, sondern Informationen über den Nutzer aus mehreren Quellen einbezieht, ist das *Profiling*. Dabei werden umfassende digitale Profile von Nutzern mit den verfügbaren und aggregierbaren Daten aus dem Internet erstellt. Informationen über einen Nutzer werden dabei aus mehreren Quellen gesammelt und miteinander verknüpft, so dass aus mehreren Teilprofilen eines Nutzers ein Gesamtprofil beziehungsweise ein umfassendes digitales Abbild dieser Person ermittelt werden kann. Hier offenbart sich ein erstes widersprüchliches Moment der Profilbildung im Netz: Einerseits werden die Nutzer durch das mediale Setting dazu gedrängt, möglichst viel über sich preiszugeben und sich damit ihrer Umgebung möglichst ‚authentisch' zu präsentieren. Andererseits können die preisgegebenen Daten durch die technischen Möglichkeiten Dritter aggregiert und zu umfassenden Profilen verrechnet werden, über welche die Nutzer keine Kontrolle mehr haben. Das digitale Abbild basiert also nur zum Teil auf bewusster Selbstdarstellung; ein Großteil des Profilings erfolgt hinter dem Rücken der Nutzer und trägt damit Züge ungeplanter Strukturentstehung im Sinne der Automatismen.[7]

In diesem Beitrag steht die technische Betrachtung der Profiling-Möglichkeiten im Vordergrund. Es werden ausgewählte technische Arbeiten von Autoren vorgestellt, die sich mit der Privatsphäre im Internet und insbesondere in sozialen Netzwerken beschäftigt haben. Ferner werden eigene Untersuchungen vorgestellt, die spezifische, die Privatsphäre bedrohende Risiken aufdecken und aufzeigen, inwiefern diese durch entsprechende Gegenmaßnahmen durch die Nutzer verhindert werden könnten. Abschließend werden reale Szenarien des Nutzer-Profilings großer Dienstanbieter aufgezeigt, dessen Zweck unter anderem die Ermittlung von Informationen über Nutzer und das Aufzeichnen und Analysieren von deren Verhalten im Internet ist.

2. Profiling im Internet

2.1 Profiling und Arten personenbezogener Daten

Profiling beschreibt die Erstellung eines möglichst umfassenden virtuellen Abbilds eines Nutzers auf Basis verfügbarer Daten, die aus Internetdiensten wie etwa OSNs extrahiert und aggregiert werden können. Diese Nutzerdaten werden als personenbezogene Daten der Nutzer definiert und laut dem Bundesdatenschutzgesetz (BDSG) in Angaben zu persönlichen Verhältnissen und Angaben zu sachlichen Verhältnissen der Nutzer kategorisiert.[8] Im Folgenden werden verschiedene Arten personenbezogener Daten vorgestellt. Dabei soll

[7] Vgl. auch den Beitrag von Andreas Weich in diesem Band.

[8] „Personenbezogene Daten sind Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (Betroffener)." (BDSG §3, Absatz 1).

die Vielzahl und die Sensibilität der Daten verdeutlicht werden, die Nutzer im Internet über sich preisgeben.

Direkte Angaben, durch welche Nutzer eindeutig identifiziert werden können, sind zum Beispiel Name, Adresse oder Telefonnummer. Pseudonymisierte Daten (BDSG §3, Absatz 6a), wie zum Beispiel E-Mail-Adressen, IP-Adressen oder Kundennummern, sind ebenfalls Daten, die einem Nutzer eindeutig zuordenbar sind. Allein durch deren Kenntnis, ohne das Wissen anderer Attribute, lässt sich die Identität des Nutzers jedoch nur dann ableiten, wenn beispielsweise der Name eines Nutzers in einem dieser Attribute kodiert ist. Neben diesen Attributen gibt es „hochsensible Daten" über den Nutzer, die prinzipiell nichts oder wenig über die Identität des Nutzers verraten, jedoch in Kombination mit einem anderen personenbezogenen Datum schützenswerte Attribute darstellen. Dies sind zum Beispiel Informationen über die „rassische und ethnische Herkunft, politische Meinungen, religiöse oder philosophische Überzeugungen, Gewerkschaftszugehörigkeit, Gesundheit oder das Sexualleben" (BDSG §3, Absatz 9). Vorlieben und Interessen der Nutzer runden das persönliche Profil ab. Diese Informationen spiegeln die Gedanken und Wünsche der Nutzer wider und geben gegebenenfalls einen Einblick in deren Persönlichkeit.

Die initiale Veröffentlichung der Angaben zu diesen Daten liegt zumeist in der Hand der Nutzer selbst. Sie geben ihre Daten beispielsweise in OSNs, bei Registrierungsprozessen verschiedener Internetportale oder auch bei der Anmeldung zu Veranstaltungen selbsttätig an und tragen so maßgeblich zu deren Verbreitung und öffentlichen Verfügbarkeit bei. Informationen, die Nutzer im Internet bei verschiedenen Dienstanbietern direkt preisgeben, übergeben sie nur mit der obligatorischen Zustimmung, dass der Anbieter das Recht hat, diese Daten für bestimmte Zwecke zu nutzen. Was der Anbieter mit den Daten der Nutzer machen darf, entscheiden lediglich die Datenschutzbestimmungen, die der Anbieter aufsetzt beziehungsweise die Datenschutzgesetze erlauben. In einer Umfrage im Auftrag der Europäischen Kommission wird gezeigt, dass mindestens 42 % der Europäer angeben, Datenschutzbestimmungen im Normalfall nicht durchzulesen.[9]

Zusätzlich können viele Informationen über den Nutzer auf indirektem Weg aus dem Kontext von dessen Angaben interpretiert werden. Beispielsweise könnten sich Dritte aus Meinungsäußerungen in Foren oder Blogs, aus Kommentaren und Statusmeldungen in den OSNs, aus Bewegungsmustern im Internet, durch Einsicht in Buchungen von Reisen oder Events, durch das Lesen von verfassten Artikeln und vielem mehr Informationen erschließen. Darüber hinaus geben Nutzer zudem häufig Informationen über andere Personen preis. Ein klassisches Beispiel hierfür ist die Freigabe des Adressbuchs für einen

[9] Europäische Kommission, *Special Eurobarometer 359: Attitudes on Data Protection and Electronic Identity in the European Union. Report*, Wave 74.3 – TNS Opinion & Social, 2011, online unter: http://ec.europa.eu/public_opinion/archives/ebs/ebs_359_en.pdf.

OSN-Betreiber, das Kontaktdaten enthält, so dass letztendlich dem OSN-Anbieter Informationen über andere Personen, gegebenenfalls sogar über Nichtmitglieder, zur Verfügung gestellt werden. Solche Freigaben erfolgen oft implizit durch die Verwendung von Anwendungen insbesondere auf Smartphones.

Als Ergänzung zu den oben vorgestellten Arten von Daten können zum Beispiel Verkehrsdaten sowie Angaben zu genutzter Hard- und Software der Nutzer erhoben werden. Unter den Begriff der Verkehrsdaten fallen physische Bewegungsdaten wie Informationen über den aktuellen Aufenthaltsort des Nutzers, die beispielsweise auf Smartphones mittels GPS, WLAN oder den verbundenen Basisstationen der Mobilfunkanbieter ermittelt werden können.[10] Andererseits werden darunter auch virtuelle Bewegungsdaten verstanden, die das Bewegungsmuster eines Nutzers im Internet preisgeben. Darin sind oft Protokolldaten wie Zeitpunkt und Dauer des Besuchs diverser Internetseiten, genutzter Dienste oder die Historie der Internetaktivitäten enthalten. Außerdem können Informationen über den Datenverkehr gesammelt werden, die Auskunft über die Inhalte geben, die der Nutzer im Internet hoch- beziehungsweise herunterlädt sowie Inhalte, die bei verschiedenen Cloud-Providern gespeichert und verarbeitet werden. Weitere Verkehrsdaten sind Kommunikationsdaten, die durch Senden und Empfangen von E-Mails, Benutzen von Chats und Telefongesprächen über das Internet oder die Telefonnetze entstehen. Damit wird nicht nur das Kommunikationsverhalten des Nutzers aufgezeichnet, sondern es werden auch sämtliche Kommunikationspartner protokolliert. Dass diese Daten selbst nach gängiger Anonymisierung noch Rückschlüsse auf natürliche Personen erlauben, zeigen etwa Zang und Bolot in ihrer Analyse von Kommunikationsdatensätzen.[11]

Angaben über die vom Nutzer verwendeten Geräte, Systeme und Anwendungen stellen eine weitere Art personenbezogener Daten dar. Anbieter können im Hintergrund Daten über die genutzte Hard- und Software der Nutzer sammeln, um statistische Aussagen über die verwendeten Technologien zu machen. Des Weiteren können diese Daten das Wiedererkennen der Nutzer ermöglichen, beispielsweise sobald sich diese mit dem gleichen Gerät erneut auf einem oder bei verschiedenen zuvor bereits genutzten Internetdiensten anmelden. Auf diese Weise können Informationen über Browsereinstellungen, Betriebs- und Hardwaresysteme sowie Gerätekennungen, Telefonnummern und genutzte Anwendungen (wie zum Beispiel Apps) gesammelt werden. Aufgrund der eindeutigen Kennzeichnung und oft auch eindeutig zuordenbaren Charakteristik eines jeden Geräts und Systems sowie einer jeden Applikation ist es oft leicht Nutzer zu identifizieren.

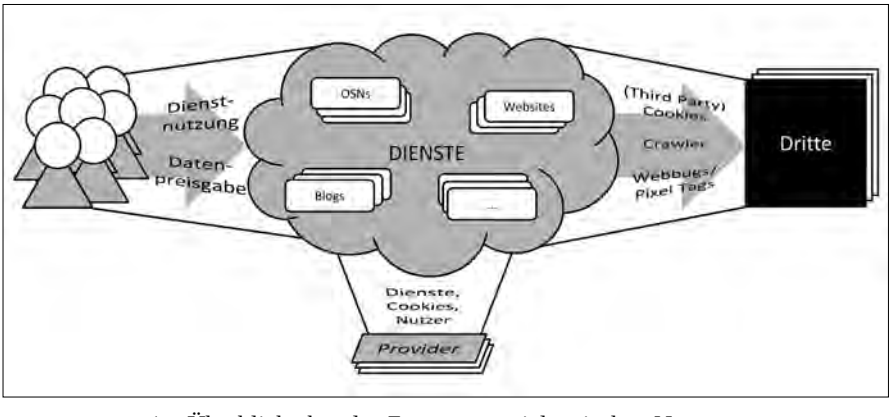
---

10 Vgl. auch den Beitrag von Jens-Martin Loebel in diesem Band.

11 Hui Zang/Jean Bolot, „Anonymization of Location Data Does Not Work: A Large-Scale Measurement Study“, in: *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking, MobiCom '11*, Las Vegas, NV, 2011.

2.2 Datensammlungen im Internet

Es gibt verschiedene Arten, wie Daten der Nutzer im Internet gesammelt werden können. Allgemein werden reaktive und nicht-reaktive Datenerhebungen unterschieden.[12] Bei der reaktiven Datenerhebung ist dem Nutzer bewusst, dass Daten über ihn gesammelt werden, die dieser in den meisten Fällen freiwillig oder auf Nachfrage angibt. Bei der nicht-reaktiven Datenerhebung ist dem Nutzer nicht bewusst, dass Daten über ihn gesammelt werden, seien es Daten über sein Verhalten, seine Vorlieben oder Aktivitäten. Der Nutzer kennt die an Dritte weitergeleiteten Daten nicht, somit könnten dies auch potenziell Daten sein, mit deren Preisgabe der Nutzer nicht einverstanden gewesen wäre.



1 – Überblick über das Zusammenspiel zwischen Nutzern, Anbietern und Dritten im Internet

Im Folgenden werden die verschiedenen Möglichkeiten nicht-reaktiver Datenerhebung skizziert, um einen Einblick in das Spektrum dafür zum Einsatz kommender Techniken zu geben. Abbildung 1 gibt in diesem Zusammenhang einen Überblick über das Zusammenspiel zwischen Nutzern, Anbietern und Dritten im Internet. Nutzer bewegen sich auf verschiedenen Internetdiensten, die sie auf unterschiedliche Art und Weise verwenden. Auf der einen Seite gibt es die OSN-Dienste, wie zum Beispiel Facebook und StudiVZ[13], in denen der Umgang mit personenbezogenen Daten oft freizügiger ist als in einem anderen Umfeld. Grund dafür ist, dass Nutzer selbst Profile von sich erstellen und dadurch aktiv zur Vervollständigung dieser durch personenbezogene Daten beitragen. Somit lassen sich aus OSNs bereits umfangreiche Daten zu einem Nutzer und oft auch eindeutige Identifikatoren der betreffenden Person sammeln (vgl. Abschnitt 3). Wenn Nutzer in verschiedenen OSNs Mitglied

[12] Holger Buxel, „Customer Profiling im Internet: Den Kunden im Visier“, in: *Science Factory*, 1 (2002), S. 1-6.
[13] http://studivz.net/.

sind, könnten zudem deren diverse Profile miteinander verknüpft und so zusätzliche Informationen gewonnen werden. Auf der anderen Seite gibt es Internetseiten wie Einkaufs-, Nachrichten- und Suchportale oder Blogs. Ein Besuch darauf hinterlässt Spuren in Form von Daten, die ebenfalls miteinander verknüpft und potenziell einer natürlichen Person eindeutig zugeordnet werden können. Die Aktivitäten der Nutzer können nachverfolgt werden und so kann eine Sammlung der Kundenprofile und eine Erfassung von Bewegungsmustern stattfinden. Das Sammeln und Verknüpfen von Daten aus OSNs und den restlichen Internetseiten kann Dritten dazu dienen, Informationen über die Nutzer zu gewinnen, um Angebote individuell zuschneiden zu können.

Die Daten der Nutzer, egal ob reaktiv oder nicht-reaktiv erhoben, werden sowohl von Anbietern der Internetportale, die Nutzer besuchen, als auch von Drittanbietern[14] gesammelt. Im Folgenden werden einige der gängigsten technischen Möglichkeiten vorgestellt, durch welche Nutzerdaten aggregiert werden können.

Mechanismen zur Nutzerverfolgung (Tracking) sind technische Implementierungen auf Internetseiten, die das Verhalten der Nutzer im Hintergrund aufzeichnen, um möglicherweise Interessen der Nutzer ableiten zu können. Ein Beispiel solcher Informationsgewinne ist die Ermittlung des Musikgeschmacks und des Alters eines Nutzers, die beispielsweise durch die Analyse aufgerufener Musikvideos im Internet erschlossen werden könnten. Im Hintergrund vieler aufgerufener Webseiten werden Aktionen, Klicks und Verhalten der Nutzer dokumentiert, durch welche Rückschlüsse auf zusätzliche Informationen über den Nutzer gezogen werden können. Hierfür eingesetzte technische Mechanismen sind zum Beispiel Cookies und Web Bugs, auf die im Folgenden eingegangen wird.

Ein Cookie ist eine Profildatei, die einem Webserver erlaubt, auf dem Rechner des Anwenders Informationen zu hinterlegen. Jedes Cookie hat eine ID, mit der ein Nutzer identifiziert werden kann. Besucht ein Nutzer eine Seite erneut, können die in Cookies enthaltenen Informationen genutzt werden, um diesen zu (re-)identifizieren. Somit kann protokolliert werden, welche Internetseiten wie oft besucht wurden.

Sofern nicht ausdrücklich in den Browsereinstellungen verboten, werden beim Aufruf einer Webseite auf der Festplatte des Nutzers Cookies gesetzt, die in First-Party-Cookies und Third-Party-Cookies unterschieden werden. First-Party-Cookies werden von der Domain gesetzt, auf dessen Seiten der Nutzer sich gerade befindet. Third-Party-Cookies werden von anderen Domains und damit Drittanbietern gesetzt, die auf der entsprechenden Webseite zum Beispiel in Form von Werbung präsent sind. Dies dient beispielsweise dazu, Statistiken über die Nutzung der Webseite zu erheben oder um diverse

[14] Drittanbieter sind auf Internetseiten anderer Anbieter zum Beispiel in Form von Werbung präsent und sammeln und/oder analysieren dort Daten und Verhalten der Nutzer, welche die Internetseiten besuchen.

Analysen durchzuführen. Beim Aufruf einer Webseite (durch http(s)-Aufrufe[15]) werden die Inhalte aller Objekte, die sich auf dieser Webseite befinden, von den entsprechenden Servern heruntergeladen. Dadurch werden sowohl First-Party- als auch Third-Party-Cookies gesetzt. Bei erneutem Aufruf der Internetseite werden die Cookies mit zusätzlichen Informationen (zum Beispiel eindeutige Identifikatoren und beim Aufruf der Seite durch einen Link oft auch sogenannte Referrer[16]) zurück an die verschiedenen Server des Anbieters und der Drittanbieter gesendet. Somit wird Dritten unter Umständen mitgeteilt, welche Internetseite der Nutzer gerade besucht und, dank Referrer, zusätzlich, auf welcher Internetseite der Nutzer einem Link zur gerade besuchten Seite gefolgt ist.

Besucht ein Nutzer eine Webseite eines Anbieters können Third-Party-Cookies von Drittanbietern geladen werden. Dabei wird auf den Servern des jeweiligen Drittanbieters eine ID gesetzt, die einen Nutzer eindeutig kennzeichnet. Falls der Nutzer anschließend eine weitere Webseite besucht, auf welcher der gleiche Drittanbieter präsent ist, kann der Nutzer anhand der ID des Cookies auch durch den Drittanbieter (re-)identifiziert werden. Der vorherige Zugriff kann mit dem Zugriff auf die aufgerufene Webseite verknüpft werden. Auf diese Weise können Verhaltensprofile von Nutzern im Internet erstellt werden, die beispielsweise aufzeigen, was für Produkte sich ein Nutzer anschaut, welche Artikel ihn interessieren und vieles mehr. Sobald der gleiche Drittanbieter auf mehreren populären Webseiten präsent ist, könnte dieser ein umfangreiches Profil eines Nutzers erstellen.[17] Durch die Analyse der erstellten Profile können Informationen gewonnen werden, die zum Beispiel für personalisierte Werbung oder Marktforschungszwecke und damit zur gezielten Ausrichtung der Strategie von Unternehmen bezüglich ihrer Produktherstellung verwendet werden könnten.

Web Bugs (auch Pixel Tags genannt) sind Elemente, die, meistens unsichtbar für den Nutzer, auf Internetseiten platziert sind und die protokollieren, welche Aktionen der Nutzer durchführt. Aus technischer Sicht sind Web Bugs Objekte (zum Beispiel kleine Bilder), die nicht unbedingt von der Seite des Anbieters des Seiteninhalts heruntergeladen werden, sondern von dem Server desjenigen, der das Nutzerverhalten analysiert. Um ein Objekt herunterzuladen, wird eine Anfrage an den entsprechenden Server gestellt. In dieser Anfrage befindet sich auch ein Hinweis auf den Teil der Internetseite, auf dem sich der Nutzer derzeit befindet. Ist weiterhin ein Cookie des Web-Bug-Eigners vorhanden, kann dieser die durch die Objektanfrage erhaltenen Informationen mit der Identität des Nutzers verknüpfen. Durch diesen Vorgang hat der Eigner des Web Bugs die Möglichkeit, Informationen zum Verhalten der Nutzer

[15] Vgl. RfC 2616 (http://www.ietf.org/rfc/rfc2616.txt), RfC 2818 (http://www.ietf.org/rfc/rfc2818.txt).

[16] Referrer ist die Adresse der Internetseite, die der Nutzer vor dem Öffnen der aktuellen Internetseite besucht hat.

[17] Bewegungsprofile sind vergleichbar mit Bewegungsprofilen des realen Lebens, welche zum Beispiel durch GPS-Daten bestimmt werden können.

zu erhalten, zum Beispiel, welche Teile der Internetseite diese öfters benutzen oder welche Inhalte sie besonders interessieren.

Eine weitere technische Maßnahme, um an die Daten der Nutzer aus verschiedenen Internetportalen zu gelangen, ist das sogenannte Crawling. Ein Verfahren, mit dessen Hilfe alle verfügbaren Inhalte von Internetseiten gesammelt werden. Beispiele berühmter Web-Crawler sind die Googlebots[18], die Crawler von Google, mit welchen die Internetseiten gesammelt und zur Analyse bereitgestellt werden.

Im Gegensatz zu Crawlern werden durch sogenannte Scraper gezielt Informationen aus Internetseiten extrahiert. Derartige Software befähigt Dritte zum Beispiel Profilinformationen aus OSNs zu sammeln und zu analysieren. Für Dritte stellen OSNs potenziell ein besonders lohnenswertes Angriffsziel im Kontext des Scrapings dar. Hier ist das Datenaufkommen besonders groß und die personenbezogenen Daten der Nutzer stets in geordneter und (technisch gesehen) immer gleicher Form abrufbar.

Um den Informationsgewinn zu maximieren, werden Nutzerdaten mittels Dienstverknüpfungen auch zwischen verschiedenen Internetseiten ausgetauscht. Ein Beispiel dafür ist die „Connect with Facebook"-Schaltfläche[19] oder der „I Like"-Button[20]. Diese werden in immer mehr Internetseiten integriert und erlauben eine Verknüpfung des Facebook-Kontos von Nutzern mit den verschiedenen Internetportalen. Exemplarisches Beispiel einer solchen Verknüpfung und Datenaggregation findet zwischen Amazon[21] und Facebook statt.[22] Amazon-Kunden, die ihr Facebook-Konto mittels der „Connect with Facebook"-Schaltfläche mit Amazon verknüpfen, geben alle ihre Daten und gegebenenfalls auch Daten der eigenen Freunde Amazon preis. Somit hat Amazon Zugriff auf die Vorlieben und Interessen der Nutzer, die diese in ihren Profilen eingetragen haben und die auf der Pinnwand gepostet sind sowie auf deren diverse Facebook-Aktivitäten, wie zum Beispiel der Nutzung des „I Like"-Buttons.

2.3 Ausgewählte Arbeiten zum Profiling im Internet

Krishnamurthy und Wills haben 2008 gezeigt, dass auf populären Internetseiten oft die gleichen Datenaggregatoren beziehungsweise Drittanbieter vorhanden sind.[23] Die Top Ten solcher Drittanbieter sind in über 70 % der populärsten Internetseiten präsent (Stand 2008), wobei Dienste von Google[24] allein

---

18 http://support.google.com/webmasters/bin/answer.py?hl=de&answer=1061943.
19 http://developers.facebook.com/docs/guides/web/.
20 http://developers.facebook.com/docs/reference/plugins/like/.
21 http://www.amazon.com.
22 https://www.amazon.com/gp/facebook/.
23 Balachander Krishnamurthy/Craig E. Wills, „Privacy Diffusion on the Web: A Longitudinal Perspective", in: *InternationalWorld Wide Web Conference*, Madrid, 2009, S. 541-550.
24 http://www.google.com.

auf fast 60 % der Internetseiten existieren. Das bedeutet, dass Googles Tracking- und Analysesysteme (zum Beispiel Google AdSense[25] zum Anzeigen von Werbung und Google Analytics[26] zur Erstellung von Statistiken und Analysen) auf mehr als der Hälfte aller Internetseiten zu finden sind, dort den Nutzern Cookies setzen und so Daten (insbesondere Verkehrsdaten) sammeln. Somit gelangt eine große Datenmenge, samt Bewegungsprofilen und Vorlieben der Nutzer, in die Hände von Dritten. Das Firefox Add-on Collusion[27] visualisiert das diskutierte Tracking von Drittanbietern eindrucksvoll. Auf potenzielle Gefahren im Zusammenhang mit Profiling durch Tracking, wird im Abschnitt 4 detaillierter eingegangen.

Im Jahr 2011 wurde durch eine Studie[28], in der eine Reihe von Internetseiten untersucht wurde (keine OSNs), herausgefunden, dass über die Hälfte der analysierten Internetseiten personenbezogene Informationen über Nutzer an Dritte weitergeben. 56 % der 120 einbezogenen Internetseiten gaben persönliche Informationen der Nutzer an Dritte weiter. Wenn die Weitergabe sogenannter Account-IDs der Nutzer einbezogen wird, erhöht sich dieser Anteil um weitere 19 % auf 75 %. Daher wurde in der Arbeit gezeigt, dass die Daten der Nutzer auch bei vertrauenswürdigen Diensten vor Zugriffen Dritter nicht geschützt sind. Die Untersuchung basierte auf dem jeweiligen Nachweis oben genannter Aggregatoren auf unterschiedlichen Internetseiten. Sie können mittels Cookies einen Nutzer über mehrere Internetseiten verfolgen und so ein anonymes Profil der Nutzer erstellen. Werden zusätzlich Informationen mit der ID der Cookies verknüpft, kann zur Identität des Nutzers auch dokumentiertes Surfverhalten zugeordnet werden.

Die Untersuchung der Autoren wurde auf populären Internetseiten wie bekannten Online-Einkaufsportalen, Seiten von Online-Reiseveranstaltern, Gesundheitsportalen usw. durchgeführt. Es wurde gezeigt, dass sowohl Teile persönlicher Informationen als auch Account-IDs weitergegeben werden. Die offenbarten persönlichen Informationen waren zum Beispiel E-Mail-Adressen, das Geschlecht, die Postleitzahl und/oder Musikinteressen der Nutzer bei verschiedenen Aktivitäten im Internet, zum Beispiel bei Registrierungs- und Anmeldeprozessen sowie dem Anhören von Musik. Zusätzlich offenbarten einige analysierte Seiten Suchanfragen von Nutzern. Dabei wurden auch verschiedene durch Nutzer gesuchte Begriffe, wie zum Beispiel zu Krankheiten, an Dritte weitergegeben. Ferner wurden Informationen über Flugrouten und entsprechende Zeitangaben sowie Reiseziele offenbart.

Aufgrund der Weitergabe der Account-IDs durch diese Internetseiten (Account-ID in Verbindung mit der Internetseite stellen eine eindeutige Kennung

---

[25] http// www.google.de/adsense.
[26] http://www.google.com/intl/de/analytics/.
[27] https://secure.toolness.com/xpi/collusion.html.
[28] Balachander Krishnamurthy/Konstantin Naryshkin/Craig E.Wills, „Privacy Leakage vs. Protection Measures: The Growing Disconnect“, in: *Proceedings of the Web 2.0 Security and Privacy Workshop*, Oakland, CA, 2011, S. 1-10.

des Nutzers dar), können Cookies der Nutzer von unterschiedlichen Arbeitsrechnern (zum Beispiel Arbeitsrechner zu Hause und bei der Arbeit) verknüpft werden. Durch die Verlinkung der Cookie-IDs zwischen den verschiedenen Rechnern und durch die Verknüpfung der Cookies mit der ID der Nutzer aus OSNs sowie mit der ID von verschiedenen Accounts aus anderen Nicht-OSN-Internetseiten, kann ein Nutzer theoretisch durch das ganze Web verfolgt werden.

In einer Arbeit aus dem Jahr 2009[29] wurde gezeigt, dass Informationen über die Identität der Nutzer auf dem gleichen Weg wie Cookies an Drittanbieter gesendet werden, so dass diese die empfangenen Informationen mit den Cookies und den IDs der Nutzer verknüpfen können und somit eine Zuordnung der Identität der Nutzer an den mittels Tracking-Mechanismen erstellten Profilen möglich ist. Dabei wurde herausgefunden, dass externen Applikationen und Drittanbietern der Zugriff auf Daten der Nutzer aus OSNs (zum Teil durch technische Mängel) ermöglicht wird. Die gewonnenen Erkenntnisse ermöglichen Rückschlüsse auf die Identität der Nutzer. Somit ist es Dritten möglich, Identitätsdaten mit Cookies der Nutzer zu verknüpfen. Aufgrund dieser Verknüpfung können Aktivitäten der Nutzer sowohl innerhalb als auch außerhalb der OSNs miteinander in Verbindung gebracht und somit ein Profil des Nutzers erstellt werden.

2010 wurde diese Studie auf mobile OSNs erweitert[30], um weitere Erkenntnisse bezüglich der Weitergabe von Nutzerdaten an Dritte zu gewinnen. Zusätzlich zu personenbezogenen Informationen ist es Dritten hier zum Beispiel möglich, die geografische Position des Nutzers herauszufinden. Diese Information kann potenziell noch mit der Uhrzeit von veröffentlichten Informationen angereichert und so eine Route ermittelt werden, auf der sich ein Nutzer befand bevor, während und nachdem dieser zum Beispiel eine Statusmeldung, Fotos oder Videos veröffentlicht hat.

Profiling in OSNs ist Dritten möglich, da Nutzer veröffentlichte Daten nicht immer mithilfe der Privatsphäre-Einstellungen vor Zugriffen Dritter schützen. Im Folgenden werden Arbeiten vorgestellt, welche die Verfügbarkeit der Nutzerdaten in OSNs in den letzten Jahren untersucht haben.

Im Jahr 2005 wurde von Gross und Acquisti eine empirische Untersuchung mit mehr als 4.000 Profilen von Studenten der Carnegie Mellon Universität durchgeführt.[31] Dabei haben 88 % der Nutzer ihr Geburtsdatum und Ge-

---

[29] Balachander Krishnamurthy/Craig E. Wills, „On the Leakage of Personally Identifiable Information Via Online Social Networks“, in: *Proceedings of the 2nd ACM Workshop on Online Social Networks*, New York, NY, 2009, S. 7-12.

[30] Balachander Krishnamurthy/Craig E. Wills, „Privacy Leakage in Mobile Online Social Networks“, in: *WOSN '10 Proceedings of the 3rd Conference on Online Social Networks*, Berkeley, CA, 2010.

[31] Ralph Gross/Alessandro Acquisti, „Information Revelation and Privacy in Online Social Networks (The Facebook Case)“, in: *ACM Workshop on Privacy in the Electronic Society, WPES '05*, Alexandria, VA, 2005.

schlecht, 77 % ihren Instant Messaging Account, zwischen 40 % und 50 % ihre Adresse, Postleitzahl oder Telefonnummer und über 60 % diverse Interessen angegeben.

Lampe et al. haben im Jahr 2007 mehr als 38.000 Facebook-Nutzerprofile untersucht und festgestellt, wie freizügig Nutzer mit ihren Daten gegenüber unbekannten Dritten umgehen.[32] Im Vergleich zur zuvor genannten Studie, in der nur 0,06 % der Nutzer ihre Daten vor Zugriffen durch Dritte gesperrt hatten, waren in dieser Untersuchung 19 % der analysierten Profile gesperrt. Die Anschrift des Nutzers war lediglich in 13,5 % der untersuchten Nutzerprofile zu finden, wohingegen die restlichen Attribute eine viel höhere Verfügbarkeit aufwiesen: 83,8 % der Nutzer veröffentlichten ihr Geburtsdatum, 92,3 % die E-Mail Adresse, 67,8 % ihren Instant Messaging Account, zwischen 60 % und 80 % verschiedene Interessen.

2008 hat eine Untersuchung von Brown et al. gezeigt, dass aus den 7.919 für diese Studie gesammelten Facebook-Profilen der Universität Michigan nur 68 % offen zugänglich und der Rest für Dritte gesperrt waren.[33] Dennoch gaben über 80 % der Nutzer ihren Geburtstag oder ihre Freundesliste preis.

Wir haben im Jahr 2011 durch eine weitere empirische Untersuchung in vier verschiedenen OSNs und auf Basis von über 180.000 analysierten Nutzerprofilen gezeigt, dass das Bewusstsein der Nutzer bezüglich ihrer eigenen Privatsphäre im Laufe der Zeit gestiegen ist.[34] In den untersuchten OSN-Profilen wurden weitaus mehr Informationen vor dem Zugriff durch Dritte verschlossen und somit restriktiver mit personenbezogenen Daten umgegangen. Es wurde ermittelt, dass in Facebook die Freundesliste das am häufigsten öffentlich verfügbare Attribut war, gefolgt von der Angabe des Geschlechts. Da das Geschlecht oft auch aus dem Namen erkennbar ist, kann unter Umständen aus der Angabe dieses Attributs keine zusätzliche Information gewonnen werden. Alle restlichen Attribute eines Nutzers waren in weniger als 20 % der Fälle für den öffentlichen Zugriff freigegeben, wie in Abbildung 2 verdeutlicht wird. Im Gegensatz dazu sperren jedoch nur lediglich 21,53 % der Nutzer den Zugriff auf sämtliche Informationen ihres OSN-Profils (bis auf den Namen, der nicht verborgen werden kann). Somit zeigt diese Studie, dass immerhin noch eine große Anzahl an Profilen Angaben offenbaren, deren Verfügbarkeit ein Informationsgewinn für Dritte bedeuten kann.

In StudiVZ war die Preisgabe der Informationen der Nutzer viel höher als in Facebook. Nur 7 % der Nutzer hatten ihr Profil vollständig für Fremde gesperrt. Über 64 % der Nutzer gaben ihr Geburtstag und mehr als 50 % ihre

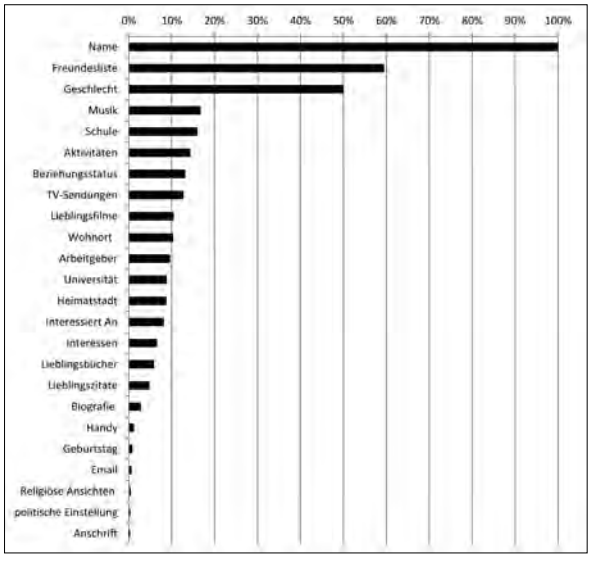
---

[32] Cliff Lampe/Nicole Ellison/Charles Steinfield, „A Familiar Face(book): Profile Elements as Signals in an Online Social Network“, in: *CHI '07 Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, New York, NY, 2007; S. 435-444.

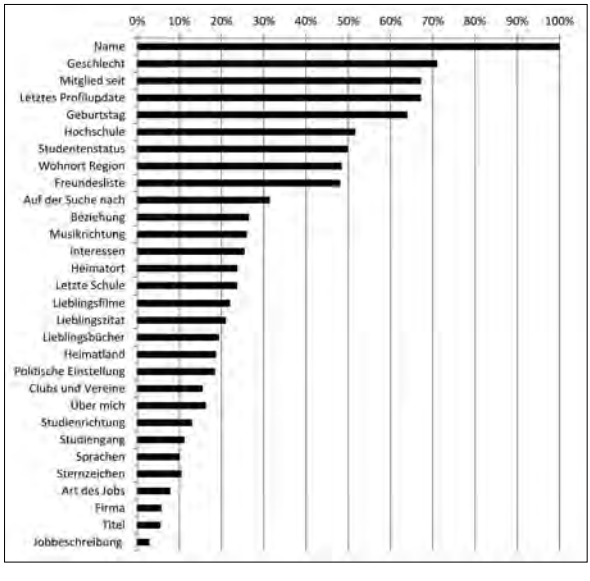
[33] Garrett Brown/Travis Howe/Micheal Ihbe/Atul Prakash/Kevin Borders, „Social Network and Context-Aware Spam“, in: *CSCW*, 2008.

[34] Labitzke/Taranu/Hartenstein (2011), What Your Friends Tell Others about You.

Hochschule an. Die Freundesliste war ebenfalls ein Attribut, das von fast 50 % der Nutzer freigegeben wurde, wie in Abbildung 3 erkennbar ist.



2 – Öffentliche Verfügbarkeit einzelner Attribute bei 110.000 analysierten Facebook-Profilen mit deutschen Vor- und Nachnamen



3 – Öffentliche Verfügbarkeit einzelner Attribute bei 43.000 analysierten StudiVZ-Profilen mit deutschen Vor- und Nachnamen

Nutzer geben demnach oft personenbezogene Daten über sich in OSNs für alle Mitglieder preis. Dies führt zu Möglichkeiten Profilverlinkungen in OSNs durchzuführen, um verteilt vorliegende Daten der Nutzer zu aggregieren. Im Folgenden werden Arbeiten vorgestellt, die sich mit der Verlinkung von Profilen beschäftigt haben.

2007 fand eine compete.com-Studie heraus, in welchem Umfang Nutzer Profile in verschiedenen OSNs erstellen.[35] Haben Dritte die Möglichkeit, diese Profile miteinander zu verknüpfen, können umfassende Profile von Nutzern erstellt werden. Von einer solchen Verknüpfung wären die Nutzer am meisten betroffen, die in einem OSN anonym sind, jedoch in einem anderen OSN mit ihren Daten freizügig umgehen. Durch eine erfolgreiche Verlinkung der Profile wäre die vermeintliche Anonymität des Nutzers nicht mehr gewährleistet.

In einer Untersuchung im Jahr 2009 wurde gezeigt, dass mehr als die Hälfte der Nutzer, die Mitglieder in verschiedenen OSNs sind, auch unterschiedliche Privatsphäre-Einstellungen haben und somit in einem OSN mal mehr und mal weniger Daten freigeben.[36] Außerdem gibt es OSNs, in denen Nutzer in der Vergangenheit Profile erstellt haben, in denen sie aber nicht mehr aktiv sind. Ihre Profile sind zum Großteil jedoch nach wie vor verfügbar und wurden nicht gelöscht. Dadurch haben Dritte potenziell die Möglichkeit, auf diese Daten zuzugreifen und restriktive Profile mit nicht-restriktiven Profilen zu verknüpfen und so ein Gesamtprofil des Nutzers zu erstellen. Außerdem wurde in oben genannter Arbeit ein Konzept entwickelt, um Profile aus verschiedenen OSNs zu identifizieren, die zu einer Person gehören und diese miteinander in Verbindung zu setzen. Dabei wurde versucht, Kriterien zu identifizieren, nach denen ein Nutzer in einem OSNs gesucht und mit einem Profil aus einem anderen OSN verknüpft werden kann.

3. Datenverknüpfung von Angaben aus Profilen mehrerer OSNs

OSNs gewinnen immer mehr an Popularität und die Nutzerzahlen sowie die Nutzungsdauer erhöhen sich jedes Jahr. Die Nutzerzahlen von Facebook sind 2010 weltweit um 69 % gestiegen und die Nutzungsdauer hat sich im Vergleich zu 2009 verdoppelt und erreicht nun im Durchschnitt sechs Stunden pro Monat und Nutzer.[37] Die Nutzerzahlen der deutschen Nutzer haben sich eben-

---

[35] http://blog.compete.com/2007/11/12/connecting-the-social-graph-member-overlap-at-open social-and-facebook/.

[36] Marti Motoyama/George Varghese, „I Seek You: Searching and Matching Individuals in Social Networks“, in: *Proceeding of the 11th International Workshop on Web Information and Data Management, WIDM '09*, New York, NY, 2009, S. 67-75.

[37] Nielsen Pressemeldung, „Starke Nutzerzuwächse für Facebook und Twitter im Vorjahresvergleich“, online unter: http://www.de.nielsen.com/news/NielsenPressemeldung05.05.2010-SocialNetworks.shtml.

falls verändert. Im Jahr 2011[38] wurde gezeigt, dass sich diese im Vergleich zum Vorjahr[39] fast verdoppelt haben, so dass im März 2011 über 22 Millionen Deutsche Facebook bereits einmal oder mehrfach besucht hatten. Die Hälfte aller in Deutschland lebenden Personen und somit 76 % der gesamten deutschen Internetnutzer sind Mitglieder in mindestens einem OSN.[40] Es wurde ferner ermittelt, dass Nutzer gewöhnlich an zwei bis drei verschiedenen OSNs teilnehmen und die Hälfte der Nutzer ihre Profile offen zugänglich belassen, wie es die Standard-Privatsphäre-Einstellungen der OSNs oft vorsehen, und somit in Kauf nehmen, dass Dritte Zugriff auf ihre Daten erhalten.

Durch die Nutzung der OSNs setzen sich die Nutzer Risiken aus. Die OSNs sind zentrale Systeme, die ein beliebtes Angriffsziel darstellen. Nutzer müssen demnach einem OSN-Anbieter in Bezug auf die Sicherheit vertrauen. Die Tatsache, dass der Anbieter selbst Zugriff auf die gesamten eingestellten Daten der Nutzer hat sowie die Rechte auf veröffentlichte Inhalte hält, kann ein Nutzer nicht umgehen. OSNs suggerieren dem Nutzer ferner das Gefühl einer geschlossenen Gemeinschaft. An die Preisgabe von Daten in einem sozialen Kontext knüpfen sich bestimmte Erwartungen an die Konsequenzen, die in diesem spezifischen Kontext entstehen.[41] Die Schnittstellen der OSNs vermitteln das Gefühl, dass Daten nur innerhalb der geschlossenen Gemeinschaft zugänglich sind und daher nicht zusätzlich geschützt werden müssten. Dies könnte ein Grund dafür sein, dass viele Nutzer ihre Privatsphäre-Einstellungen unverändert belassen und damit einen Großteil ihrer in OSNs geteilten Daten frei zugänglich machen. Die Nutzer tragen – im selbsttechnologischen Sinne – die Verantwortung für das eigene Datenmanagement, die durch die Einstellungsmöglichkeiten entstehenden Erwartungen an die Anbieter werden jedoch von diesen nicht eingelöst. Das Aggregieren der Daten im Backend entzieht sich dem Management der Nutzer, zusätzlich bieten OSNs eine Basis für die klassischen Gefahren des Internets, wie zum Beispiel Spamming, Phishing oder Identitätsdiebstahl.[42]

In unserer Arbeit aus dem Jahr 2011 wurde gezeigt, dass Nutzer in verschiedenen OSNs (hier: StudiVZ, Facebook, XING und MySpace) unterschiedliche Daten freigeben und für Zugriffe Dritter zugänglich machen.[43] Für

---

38 Nielsen Pressemeldung, „Deutsche Top-Marken im Internet und Onlinenutzerprofil: März 2011“, online unter: http://www.de.nielsen.com/news/NielsenPressemeldung-OnlineMarz2011.shtml.

39 Nielsen Pressemeldung, „Deutsche Top-Marken im Internet und Onlinenutzerprofil: März 2010“, online unter: http://www.de.nielsen.com/news/NielsenPressemeldung19.04.2010-OnlineMarz.shtml.

40 BITKOM Presseinformation, „Halb Deutschland ist Mitglied in sozialen Netzwerken“, April 2011, online unter: http://www.bitkom.org/de/presse/70864_67667.aspx.

41 Helen Nissenbaum entwickelt unter dem Stichwort „Contextual Integrity“ Modelle, um solche Erwartungen in rechtlicher und informatischer Hinsicht zu formalisieren; vgl. dies., *Privacy in Context: Technology, Policy, and the Integrity of Social Life*, Palo Alto, CA, 2012.

42 Hogben (2007), *ENISA Security Issues and Recommendations for Online Social Networks*.

43 Labitzke/Taranu/Hartenstein (2011), What Your Friends Tell Others about You.

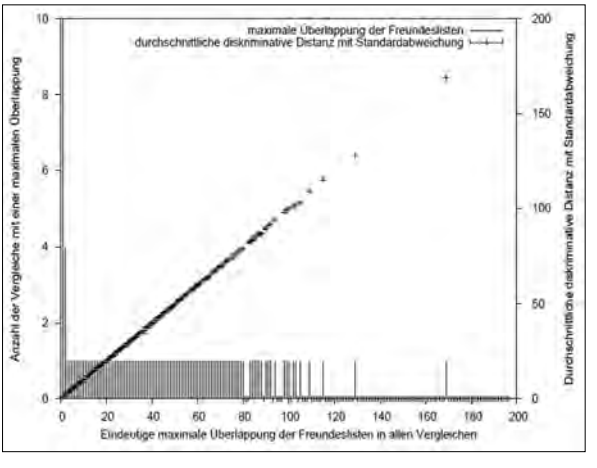
die Untersuchung wurden öffentlich einsehbare Daten aus diesen OSNs analysiert. Die Daten der Nutzer aus den OSNs wurden datenschutzkonform verarbeitet, das gesetzeskonforme Konzept dieser Studie wurde in einer vorangegangenen Publikation vorgestellt.[44] Das Konzept erzwingt, dass in keinem Schritt der Untersuchung Rückschlüsse auf die Identität der Nutzer gezogen werden können und die Auswertung damit anonym durchgeführt werden konnte.

Neben der Ermittlung des Umfangs der Selbstoffenbarung der Nutzer in OSNs wurden mögliche Angriffsszenarien erarbeitet und offengelegt, wie Profile der Nutzer aus verschiedenen OSNs miteinander verknüpft werden könnten. Es wurde gezeigt, dass eine mögliche Profil-Verknüpfung anhand der Namen der Nutzer und deren Freundeslisten durchgeführt werden kann. Wenn Nutzer ihre Profile in unterschiedlichen OSNs auch mit dem gleichen Namen registriert haben, reicht eine Überlappung von mehr als drei Namen aus zwei Freundeslisten, um mit einer hohen Fehlerfreiheit eine korrekte Verknüpfung herzustellen. Dies ist besonders prekär, da gezeigt werden konnte, dass die Freundeslisten je nach OSN in 40 % bis 67 % der Fälle frei verfügbar sind. Das bedeutet, dass, falls Nutzer in verschiedenen OSNs Profile mit gleichem Namen registriert haben und mehr als drei Namen in beiden Freundeslisten zu finden sind, diese Profile mit einer sehr hohen Wahrscheinlichkeit dem gleichen Nutzer gehören.

Der Vergleich der Freunde wurde folgendermaßen durchgeführt: In allen untersuchten OSNs wurde nach Nutzern gesucht, die den gleichen Namen angegeben haben, und deren Freundeslisten wurde betrachtet. Die Freundesliste eines Nutzers aus einem OSN wurde mit allen Freundeslisten der Nutzer aus einem anderen OSN verglichen, die den gleichen Namen haben, und die Anzahl der Überlappungen untersucht. Abbildung 4 zeigt Ergebnisse von insgesamt 7 Millionen einzelner Profilvergleiche, genauer die Überlappung von miteinander verglichenen Freundeslisten der betreffenden Profile. Die Menge von Vergleichen lässt sich in Untermengen unterteilen, die stets Vergleiche eines Profils aus dem OSN StudiVZ mit den Profilen aus Facebook repräsentieren, die potenziell den gleichen Besitzer haben (hier: Profile, die mit demselben Namen registriert sind). Aus diesen Untermengen wurden Histogramme erstellt, auf deren x-Achse die Überlappung zweier verglichener Freundeslisten und auf der y-Achse die Häufigkeit des Auftretens spezifischer Überlappungen aufgetragen sind. Abbildung 4 zeigt nun eine Aggregation dieser Histogramme. Dafür wurde stets das Maximum der einzelnen y-Werte bezogen auf jeweils einen x-Wert aus allen Histogrammen in dieses Diagramm übertragen. Es ist ersichtlich, dass ab einem x-Wert von 4 der y-Wert den Wert 1 nicht überschreitet. Das bedeutet, dass in keiner der Untermengen von Vergleichen mehr als ein Vergleich zu einer gleich hohen Überlappung geführt

[44] Labitzke/Dinger/Hartenstein (2011), How I and Others Can Link My Various Social Network Profiles as a Basis to Reveal My Virtual Appearance.

hat. Ferner ist in diesem Diagramm die sogenannte diskriminative Distanz aufgetragen. Dieser Wert gibt an, welchen durchschnittlichen Abstand eine maximale Überlappung zur nächst kleiner detektierten Überlappung innerhalb einer Untermenge von Vergleichen aufwies. Hieran wird deutlich, dass, wenn eine hohe Überlappung bei einem einzelnen Vergleich ermittelt wurde, diese Überlappung auch mit einem deutlichen Abstand zu sonstigen Überlappungen innerhalb einer Untermenge von Vergleichen lag. Die Ausprägung dieses Ergebnisses deutet darauf hin, dass eine große Überlappung stets als eindeutiger Ausreißer in Vergleichsdaten auftaucht und damit als Merkmal genutzt werden könnte, um Profile einer einzelnen natürlichen Person zu verknüpfen. Dabei kann die Wahrscheinlichkeit als gering angenommen werden, dass zwei Profile, die mit gleichem Namen in einem OSN registriert sind, eine ähnlich große Überlappung mit einem verglichenen Profil aus einem anderen OSN aufweisen.

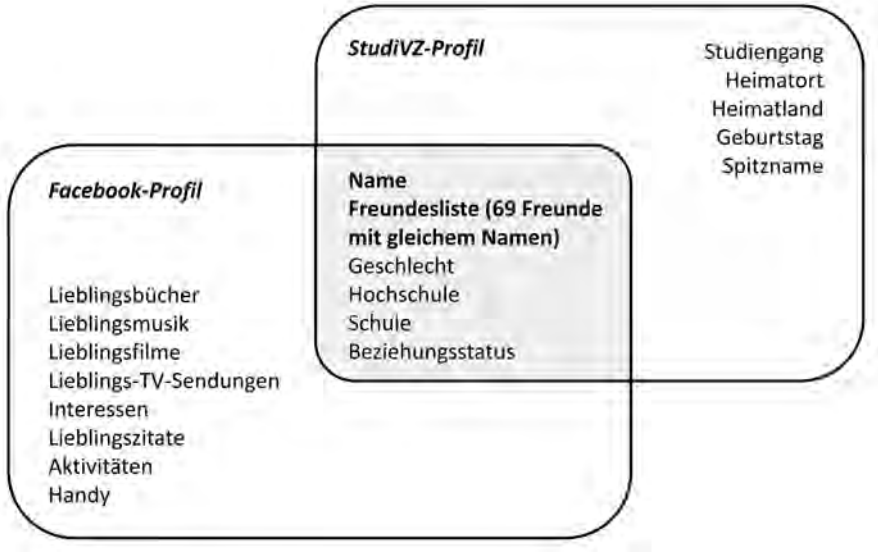


4 – Freundeslistenvergleiche zwischen Facebook und StudiVZ

Damit und auf Basis des Ergebnisses, dass viele und je nach OSN verschiedenartige Informationen öffentlich zur Verfügung gestellt werden, wurde festgestellt, dass das Potenzial von Verlinkungen innerhalb verschiedener OSNs sehr groß ist. Durch die Verknüpfung der Profile können zusätzliche Informationen über Nutzer gewonnen werden.

Im Folgenden wird ein Beispiel zweier Profile gezeigt, die zufällig aus der Menge an Profilen entnommen wurden, die einem einzelnen Nutzer zugeordnet und damit verknüpft werden konnten. Im Beispiel wird ein Profil aus Facebook mit einem Profil aus StudiVZ verglichen. Beide Profile haben den gleichen Namen und 69 Freunde, die in beiden Freundeslisten vertreten sind. Dies entspricht 16,2 % der Facebook-Freundesliste und 22,6 % der StudiVZ-Freundesliste. Anhand des Namens und der Überlappung von 69 gemeinsamen

Freunden wurde die Verknüpfung der Profile durchgeführt. Zusätzlich enthalten beide Profile noch die Attribute Geschlecht, Hochschule, Schule und Beziehungsstatus. Beide Profile stellen jedoch auch Attribute zur Verfügung, die im jeweils anderen Profil nicht auffindbar sind. Das Facebook-Profil enthält zusätzlich noch Angaben zu Lieblingsbüchern, Lieblingsmusik, Lieblingsfilmen, Lieblings-TV-Sendungen, Interessen, Lieblingszitaten, Aktivitäten und die Handynummer des Nutzers. Das StudiVZ-Profil verrät zusätzlich zur Hochschule des Nutzers noch dessen Studiengang, Heimatort, Heimatland, Geburtstag und Spitznamen (siehe Abbildung 5). Durch die erfolgreiche Verknüpfung dieser beiden Profile kann ein umfassendes Profil des Nutzers mit viel mehr Informationen als aus nur einem OSN-Profil erstellt werden. Dieses Beispiel verdeutlicht das Potenzial eines Angriffs auf die Privatsphäre eines Nutzers auf Basis von Profilverknüpfung.



5 – Beispiel einer Profil-Verknüpfung zwischen Facebook und StudiVZ

4. Dienstübergreifende Verknüpfung von Nutzerdaten

In den vorherigen Abschnitten wurden insbesondere *potenzielle* Möglichkeiten zum Profiling im Internet betrachtet. Mit der in Abschnitt 3 vorgestellten Studie wurde beispielsweise ein zunächst hypothetisches Risiko für Nutzer skizziert, das durch die technische Möglichkeit zur Profil-Verknüpfung in OSNs gegeben ist. Im Folgenden soll nun auf konkrete, offensichtlich im Einsatz befindliche Maßnahmen zum dienstübergreifenden Profiling eingegangen werden. Die Frage, ob diese Arten von Profiling ausschließlich ein Risiko für Nut-

zer darstellen oder im Gegenteil eventuell sogar positive Effekte für die Nutzer haben können, soll im Rahmen dieses Beitrags explizit nicht beantwortet werden.

In vielen OSNs ist es möglich, nicht nur die Dienste des eigentlichen OSN-Betreibers zu nutzen, sondern auch sogenannte Apps zu verwenden, die nicht zwingend durch den OSN-Betreiber bereitgestellt werden. Dritte können diese Apps durch definierte Schnittstellen (engl. Application Programming Interface, API) innerhalb der OSNs bereitstellen. Google bietet eine API an, die von einigen OSNs adaptiert wurde und zur Integration von Drittanbieter-Apps genutzt werden kann[45], wie zum Beispiel in StudiVZ und Xing. Facebook bietet eine eigene API (OpenGraph[46]) zur Integration von Apps an. Über diese API wurden laut Facebook bereits mehr als 550.000 Apps integriert.

Den Programmierern einer App wird die Möglichkeit gegeben, Zugriff auf alle oder Teile von Informationen zu erhalten, die innerhalb eines Nutzerprofils zur Verfügung stehen. Dies ist teilweise notwendig, um Funktionen von Apps realisieren zu können. Als Beispiel sei eine App gegeben, die einem Nutzer dessen tägliches Horoskop anzeigt. Diese App wird eine Information über das Geburtsdatum des Nutzers benötigen, um den Horoskop-Dienst zu erbringen. Des Weiteren können zur Verwendung einer App diverse Rechte erbeten werden, die es einer App erlauben bestimmte Aktionen innerhalb des OSNs (teilweise im Namen des Nutzers) durchzuführen. Hierzu zählen beispielsweise das Zugriffsrecht auf private E-Mail-Nachrichten oder auch das Recht, dass eine App im Namen des Nutzers an dessen Facebook-Pinnwand[47] schreiben darf.

Wird einem Drittanbieter durch die Nutzung einer App Zugriff auf Profilinformationen gewährt oder werden diesem Rechte im Sinne der oben genannten Beispiele eingeräumt, bekommen die Nutzer vor der ersten Verwendung dieser App eine Übersicht über die Informationen und Rechte angezeigt, die eine App anfragt. Das Herauslesen der Implikationen, welche die Weitergabe spezifischer Informationen oder diverse Rechte mit sich bringen, obliegt dem Nutzer. Dieser muss vor der ersten Verwendung der App zustimmen, dass die aufgelisteten Daten dem Anbieter der App zur Verfügung gestellt werden.

Eine kürzlich abgeschlossene studentische Arbeit[48] untersuchte, ob die Informationen auf den oben erwähnten Übersichtsseiten mit den Daten und Rechten übereinstimmen, die einer App tatsächlich übermittelt beziehungsweise eingeräumt werden. Ein Ergebnis war, dass zumindest alle Daten und Rechte (in aggregierter Form) aufgelistet werden und vom Nutzer eine entsprechende Zustimmung verlangt wird. Seitens der OSNs wird jedoch nicht geprüft, ob die App tatsächlich alle angeforderten Informationen und Rechte

---

45 OpenSocial: http://code.google.com/intl/de-DE/apis/opensocial/.

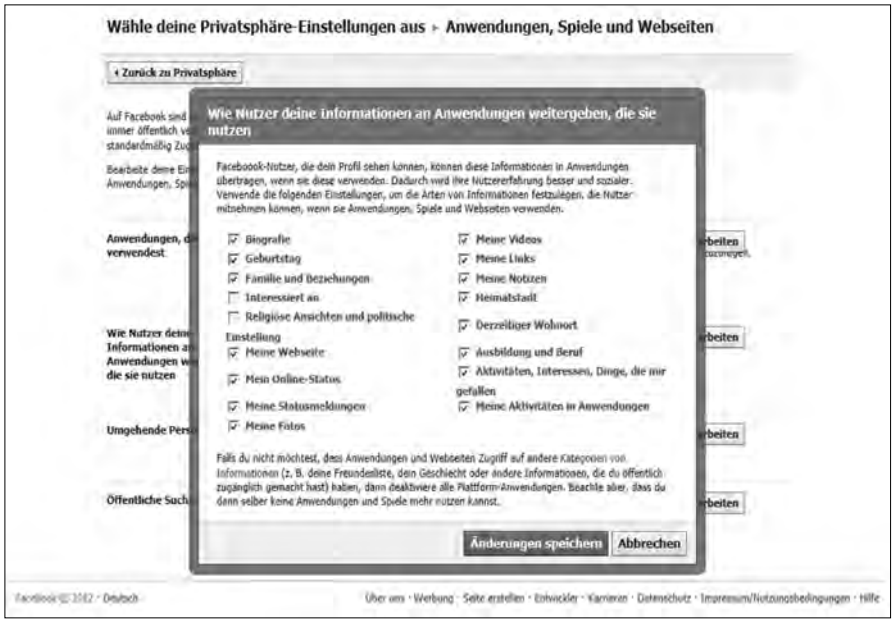
46 Vgl.: https://developers.facebook.com/.

47 Einträge auf der Facebook-Pinnwand sind je nach Privatsphäre-Einstellung für einen bestimmten Nutzerkreis oder aber auch für alle Facebook-Nutzer zugänglich.

48 Heike Hennig, *Analyse von APIs sozialer Netzwerke*, Studienarbeit am KIT, 2011.

benötigt, um den eigentlichen Dienst zu erbringen. Anderseits ist es für den Nutzer nicht möglich die App zu nutzen, ohne der Gesamtheit an weiterzugebenden Informationen und zu erteilenden Rechten zuzustimmen. Eine Auswahl der Informationen und Rechte, mit deren Weitergabe beziehungsweise Einräumung der Nutzer einverstanden wäre, ist nicht möglich.

Ein weiteres Ergebnis der Studie zeigt, dass zum Teil auch Daten von OSN-Freunden eines Nutzers übermittelt werden, wenn dieser eine App benutzt, ohne dass die Freunde dieser Weitergabe explizit zustimmen müssen. Abbildung 6 zeigt einen Screenshot eines Bereichs der Facebook-Privatsphäre-Einstellungen. Innerhalb des gezeigten Dialogfensters können Nutzer konfigurieren, welche ihrer Daten Drittanbietern von Facebook-Anwendungen (Apps) zur Verfügung gestellt werden, wenn Personen diese Anwendungen nutzen, deren Profile mit den Profilen der Nutzer verknüpft sind (OSN-Freunde). Die Voreinstellungen dieser Konfigurationsmöglichkeit erlauben, dass eine Vielzahl von Daten eines Nutzers an App-Anbieter gesendet wird, wenn einer von dessen Facebook-Freunden eine App mit entsprechenden Zugriffsrechten nutzt.



6 – Voreinstellung der Privatsphäre-Einstellungen bezüglich der Weitergabe der eigenen Daten an Betreiber von Anwendungen, welche die eigenen Facebook-Freunde nutzen.

Allgemein kann geschlussfolgert werden, dass OSNs wie Facebook Daten von App-Nutzern an die Anbieter von Apps weitergeben und Zugriffsrechte ein-

räumen, wenn die Nutzer dieser Apps diesen explizit zustimmen. Ferner werden Daten von Personen weitergegeben, die einerseits mit einem App-Nutzer befreundet sind, jedoch anderseits der Weitergabe lediglich implizit und ohne die zum Beispiel vom deutschen Datenschutz geforderte Zweckbindung zugestimmt haben. Grund hierfür sind entsprechende Privatsphäre-Einstellungen, die gegebenenfalls nicht verändert und vom Nutzer in der Voreinstellung belassen wurden.

Vor dem Hintergrund des Profiling zeigt der skizzierte Umgang mit Apps in OSNs, dass Dritten ausreichend Möglichkeiten gegeben werden, um Daten von Nutzern zu sammeln und zu analysieren. Es ist offensichtlich, dass sich potenziell umfassende Profile von Nutzern und deren OSN-Freunden gewinnen lassen, wenn diese Datensammlungen über längere Zeit und mit entsprechender Tiefe durchgeführt werden. Nicht immer besteht jedoch lediglich die Möglichkeit, dass Dritten die Gelegenheit eingeräumt wird Profiling zu betreiben. Große Unternehmen, insbesondere diejenigen, die selbst ein breites Spektrum an Internetdiensten anbieten, haben selbst die Möglichkeit über Informationen, die sich über die Dienste hinweg von Nutzern sammeln lassen, umfassende Datenprofile der Nutzer anzulegen.

Das Unternehmen Google hat als Anbieter zahlreicherer Dienste (wie zum Beispiel Google Mail, YouTube, Picasa, Google Suche und viele mehr) Zugriff auf alle preisgegebenen personenbezogenen Informationen der Nutzer in den jeweiligen Diensten. Im März 2012 hat google offiziell eingeräumt, dass Informationen von Nutzern dienstübergreifend ausgewertet werden und damit Informationen aus verschiedenen Diensten miteinander verknüpft werden. Das Ziel ist, dass der Nutzer über alle Dienste hinweg als ein einziger, eigenständiger Nutzer behandelt wird. Informationen, die Google aus einem Dienst gewonnen hat, werden genutzt, um das Angebot eines anderen internen Dienstes zu verbessern und auf den Nutzer personalisierter einzugehen.

Es gibt verschiedene Quellen, durch welche Google Zugang zu Informationen über Nutzer erhalten kann. Google ist es beispielsweise möglich, ein internes Profil über den Nutzer zu erstellen, indem die Daten des Nutzers aus den verschiedenen Google-Diensten miteinander verknüpft werden. Durch einen Google-Account (z. B. bei Google Mail oder dem OSN Google+[49]) erstellt der Nutzer selbst ein Profil von sich. Durch die Nutzung weiterer Dienste können weitere Informationen mit dem Profil des Nutzers verknüpft werden. Zum Beispiel werden durch die Nutzung der Google Suchmaschine Suchanfragen samt IP-Adressen und Cookies gespeichert. Durch die Nutzung des OSN Google+ erhält Google Zugriff auf Freundeslisten, Fotos und Neuigkeiten der Nutzer. Durch das Einstellen, Anschauen und Kommentieren von YouTube-Videos können weitere Informationen hinzugefügt werden. Auch durch Informationen aus dem Dienst Google Maps kann ein Profil einer Person mit Informationen angereichert werden, die hier preisgeben, an welchem physischen

[49] https://plus.google.com.

Ort dieser sich befindet oder wohin er sich bewegen möchte. Mit Informationen über genutzte Hard- und Software der Nutzer ergibt sich ein sehr umfangreiches Profil des Nutzers.

Zusätzlich dazu kann Google Informationen über einen Nutzer aus weiteren Quellen beziehen. Durch die große Präsenz von Google auf vielen Internetseiten und durch das Setzen von Cookies ist Google in der Lage umfassende Profile der Nutzer zu erstellen beziehungsweise diese zu erweitern. In oben vorgestellten Arbeiten wurde gezeigt, dass mittels Cookies auch personenbezogene Informationen übertragen werden. Dadurch und durch die anderweitig gesammelten Informationen könnte Google zum Beispiel die Identität eines Nutzers in Erfahrung bringen.

5. Fazit

Zusammenfassend zeigt sich, dass personenbezogene Daten ein schützenswertes Gut sind, auch wenn ungeklärt bleibt, ob Profiling tatsächlich nur negative oder auch positive Konsequenzen für Nutzer nach sich ziehen kann. Die Risiken lassen sich mit Blick auf die umfangreichen Möglichkeiten zum Profiling leicht herausstellen. Gehen Nutzer nicht bedacht mit der Preisgabe ihrer Daten um, besteht die Möglichkeit, dass diese für weitere Zwecke Verwendung finden. Anbieter von Internetdiensten sowie Dritte können (gewollt oder ungewollt) Zugriff auf die Daten erlangen. Andererseits zeigt eine Befragung[50], dass 74 % der Europäer der Meinung sind, dass die Informationspreisgabe ein Teil des modernen Lebens ist und sie damit einverstanden sind, dass Daten wie Namen, Fotos, Adressen oder Telefonnummern in verschiedenen Diensten anzugeben sind, um diese Dienste in Anspruch nehmen zu können. Jedoch haben die Nutzer auch Misstrauen gegenüber den Anbietern geäußert. Lediglich 22 % der Nutzer vertrauen Diensten wie OSNs, Suchmaschinen oder E-Mail-Dienstleistern. Viele haben den Eindruck, dass die Daten, die Anbieter über sie sammeln, nicht nur für die angegeben Zwecke genutzt werden.

Allgemein wurde gezeigt, dass, wenn Nutzer einen Internetdienst nutzen, oft ein Stück der Privatsphäre mit Dritten geteilt wird. Dieser Beitrag zeigt anhand der vorgestellten technischen Möglichkeiten zum Profiling, dass es sinnvoll ist, ein Bewusstsein der Nutzer für das Thema Profiling zu etablieren, indem demonstriert wird, wie protokollierbar Aktivitäten im Internet tatsächlich sind und wie umfassend Profiling durch Dritte möglich ist. Ferner sollte Nutzern bewusst werden, dass sie im virtuellen Leben die gleiche Identität annehmen, wie im realen Leben und dass ein Verstecken hinter Teilidentitäten im Internet oft nicht (mehr) möglich ist. Auch wenn es nicht sinnvoll erscheint, Nutzer in ihrem Handeln im Internet einzuschränken, hat dieser Beitrag deutlich gemacht, dass eine unbedachte Datenpreisgabe und gegebenenfalls an-

[50] Europäische Kommission (2011), *Special Eurobarometer 359.*

gewandtes Profiling umfangreiche Konsequenzen für Nutzer nach sich ziehen kann. Die Bewertung, ob diese Konsequenzen positiver oder negativer Natur sind, obliegt weiterhin dem Nutzer selbst.

Allerdings ist auch deutlich geworden, dass die Entwicklungen der technischen Infrastruktur der OSNs einen selbstbestimmten Umgang mit der digitalen Identität zunehmend erschweren. Aus Nutzersicht scheint der Reiz der OSNs gerade darin zu liegen, dass die Preisgabe von Informationen nicht bis ins letzte Detail gesteuert wird, denn erst so können unerwartete Verbindungen zu anderen Nutzern und damit ungeplante Strukturen auf der kommunikativen Ebene entstehen. Diese Dynamik wird durch die Nutzer-Schnittstellen der OSNs aktiv vorangetrieben, sie fordern das Ausfüllen der Formularfelder im Hinblick auf eine möglichst umfassende Selbstdarstellung geradezu ein. Man könnte also einerseits von Selbst-Technologien im Sinne eines bewusst gesteuerten Selbstmanagements sprechen: Das eigene Profil als Lebenslauf, der für unterschiedliche Netzwerke und Kontexte präzise zugeschnitten und angepasst wird. Genau dieses aktive Spiel mit unterschiedlichen digitalen Teilidentitäten wird jedoch durch die technischen Möglichkeiten der Profilbildung unterlaufen. Damit erhält der Begriff Selbst-Technologien eine zweite Bedeutung: Es geht um Technologien zur Herstellung einer digitalen Identität bzw. eines digitalen Abbilds jenseits der Kontrolle durch die Nutzer. Auf welche Quellen diese Technologien zugreifen, welche Gewichtungen sie vornehmen, wie der Prozess der Aggregation verläuft und wie die entstehenden Profile wiederum in die Inhalte zurückgespeist werden, entzieht sich der Kenntnis der Nutzer. Selbst-Technologien auf der Ebene der Nutzer und Selbst-Technologien im technischen Sinn verflechten sich damit auf eine immer diffusere Weise, kausale Zuordnungen zwischen Handlungen an einer Stelle und Konsequenzen an anderer Stelle sind vor dem Hintergrund des Profiling kaum noch möglich.

Literatur

BITKOM Presseinformation, „Halb Deutschland ist Mitglied in sozialen Netzwerken“, April 2011, online unter: http://www.bitkom.org/de/presse/70864_67667.aspx.

Brown, Garrett/Howe, Travis/Ihbe, Micheal/Prakash, Atul/Borders, Kevin, „Social Network and Context-Aware Spam“, in: *CSCW*, 2008.

Buxel, Holger, „Customer Profiling im Internet: Den Kunden im Visier“, in: *Science Factory*, 1 (2002), S. 1-6.

Europäische Kommission, *Special Eurobarometer 359: Attitudes on Data Protection and Electronic Identity in the European Union. Report*, Wave 74.3 – TNS Opinion & Social, 2011, online unter: http://ec.europa.eu/publicopinion/archives/ebs/ebs359en.pdf.

Gross, Ralph/Acquisti, Alessandro, „Information Revelation and Privacy in Online Social Networks (The Facebook Case)“, in: *ACM Workshop on Privacy in the Electronic Society, WPES '05*, Alexandria, VA, 2005.

Günther, Oliver/Kovrigin, Alexander/Nowobilska, Aneta/Krasnova, Hanna/Hildebrand, Thomas, „Why Participate in an Online Social Network: An Empirical Analysis“, in: *16th European Conference on Information Systems, ECIS '08*, Galway, 2008.

Hennig, Heike, *Analyse von APIs sozialer Netzwerke*, Studienarbeit am KIT, 2011.

Hogben, Giles, *ENISA Security Issues and Recommendations for Online Social Networks, ENISA Position Paper for W3C Workshop on the Future of Social Networking*, 2007, online unter: http://www.w3.org/2008/09/msnws/papers/Future_of_SN_Giles_Hogben_ENISA.pdf.

Krishnamurthy, Balachander/Naryshkin, Konstantin/Wills, Craig E., „Privacy Leakage vs. Protection Measures: The Growing Disconnect“, in: *Proceedings of the Web 2.0 Security and Privacy Workshop*, Oakland, CA, 2011, S. 1-10.

Krishnamurthy, Balachander/Wills, Craig E., „Privacy Diffusion on the Web: A Longitudinal Perspective“, in: *InternationalWorld Wide Web Conference*, Madrid, 2009, S. 541-550.

Dies., „On the Leakage of Personally Identifiable Information Via Online Social Networks“, in: *Proceedings of the 2nd ACM Workshop on Online Social Networks*, New York, NY, 2009, S. 7-12.

Dies., „Privacy Leakage in Mobile Online Social Networks“, in: *WOSN '10 Proceedings of the 3rd Conference on Online Social Networks*,Berkeley, CA, 2010.

Labitzke, Sebastian/Dinger, Jochen/Hartenstein, Hannes, „How I and Others Can Link My Various Social Network Profiles as a Basis to Reveal My Virtual Appearance“, in: *Lecture Notes in Informatics (LNI - Proceedings, GI-Edition)*, 4. DFN Forum Kommunikationstechnologien, Bonn, 2011, S. 123-131.

Labitzke, Sebastian/Taranu, Irina/Hartenstein, Hannes, „What Your Friends Tell Others about You: Low Cost Linkability of Online Social Network Profiles“, in: *5th International ACM Workshop on Social Network Mining and Analysis*, San Diego, CA, 2011.

Lampe, Cliff/Ellison, Nicole/Steinfield, Charles, „A Familiar Face(book): Profile Elements as Signals in an Online Social Network“, in: *CHI '07 Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, New York, NY, 2007; S. 435-444.

Motoyama, Marti/Varghese, George, „I Seek You: Searching and Matching Individuals in Social Networks“, in: *Proceeding of the 11th International Workshop on Web Information and Data Management, WIDM '09*, New York, NY, 2009, S. 67-75,

Nielsen Pressemeldung, „Deutsche Top-Marken im Internet und Onlinenutzerprofil: März 2010“, online unter: http://www.de.nielsen.com/news/NielsenPressemeldung19.04.2010-OnlineMarz.shtml.

Dies., „Deutsche Top-Marken im Internet und Onlinenutzerprofil: März 2011“, online unter: http://www.de.nielsen.com/news/NielsenPressemeldung-OnlineMarz2011.shtml.

Dies., „Starke Nutzerzuwächse für Facebook und Twitter im Vorjahresvergleich“, online unter: http://www.de.nielsen.com/news/NielsenPressemeldung05.05.2010-SocialNetworks.shtml.

Nissenbaum, Helen, *Privacy in Context: Technology, Policy, and the Integrity of Social Life*, Palo Alto, CA, 2012.

Solove, Daniel J., „‚I've Got Nothing to Hide' and Other Misunderstandings of Privacy, in: *San Diego Law Review*, 44 (2007). (GWU Law School Public Law Research Paper No. 289.)

Zang, Hui/Bolot, Jean, „Anonymization of Location Data Does Not Work: A Large-Scale Measurement Study", in: *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking, MobiCom '11*, Las Vegas, NV, 2011.

Internetquellen

http://blog.compete.com/2007/11/12/connecting-the-social-graph-member-overlap-at-opensocial-and-facebook/
http://code.google.com/intl/de-DE/apis/opensocial/
http://developers.facebook.com/docs/guides/web/
http://developers.facebook.com/docs/reference/plugins/like/
http://studivz.net/
http://support.google.com/webmasters/bin/answer.py?hl=de&answer=1061943
http://www.amazon.com
http://www.google.com
http://www.google.com/intl/de/analytics/
http://www.google.de/adsense
https://developers.facebook.com/
https://plus.google.com
https://secure.toolness.com/xpi/collusion.html
https://www.amazon.com/gp/facebook/
https://www.facebook.com/