Privacy

Oliver Berthold, Johann-Christoph Freytag
Lehrstuhl für Datenbanken und Informationssysteme (DBIS)
Insitut für Informatik
Humboldt-Universität zu Berlin
{berthold,freytag}@dbis.informatik.hu-berlin.de

Sucht man nach Übersetzungen des Begriffs „Privacy“, so findet man Privatsphäre, Intimsphäre, Ungestörtheit, Abgrenzung und Datenschutz. Privacy meint also das Recht jedes Einzelnen auf einen „privaten, geschützten Raum“, welcher von anderen und von der Gesellschaft als ganzes nur in wohl definierten Ausnahmefällen verletzt werden darf. Der „private Raum“ ist in erster Linie die Wohnung einer Person. Es ist selbstverständlich und auch durch verschiedene Gesetze untermauert, dass die Wohnung einem speziellen Schutz unterliegt. Privatsphäre umfasst aber wesentlich mehr. So umfasst der gesetzlich garantierte Schutz der Privatsphäre oft auch deren Bewegungen und Aktionen im öffentlichen Bereichen und eben auch die Bewegungen und Aktionen in virtuellen Räumen wie dem Internet. Eine Sammlung von Datenschutzgesetzen findet man unter http://www.datenschutzzentrum.de/ldsh/gesetze.htm.

Mit dem Recht auf Privatsphäre soll in erster Linie die Freiheit des Einzelnen sichergestellt werden, da die Freiheit beeinträchtigt wird, wenn andere unkontrollierbar genaue Informationen über einen Menschen haben. Menschenrechte, Verfassungen und Gesetze schreiben mehr oder weniger genau vor, wie weit dieser Schutz der Freiheit geht – und zwar genau so weit, wie nicht die Freiheit eines anderen beeinträchtigt wird. Genau dies geschieht jedoch, wenn über Menschen ohne deren Zustimmung Daten gesammelt und ausgewertet werden, denn diese Informationen gehören im Zweifelsfall immer zum privaten Bereich und unterstehen damit dem Recht auf „Informationelle Selbstbestimmung“, wie das Bundesverfassungsgericht im Volkszählungsurteil vom 15.12.1983 entschieden hat [BVerfGE84].

Ersthaft bedroht wurden und werden diese Rechte im wesentlichen seit Einführung der elektronischen Datenverarbeitung und insbesondere der vernetzten Rechner. Damit war es erstmals möglich, ohne großen Aufwand extrem viele Daten zu speichern und effizient zu verarbeiten. So können insbesondere mit vernetzten Datenbanken oder Data-Warehouses an sich harmlose Datenspuren verknüpft (Datamining, Rasterfahndung) und zu aussagekräftigen Profilen zusammengesetzt werden.

Ein Beispiel für die fortschreitenden Möglichkeiten zur Datensammlung ist die Radio Frequency Identification-Technologie. Damit soll in naher Zukunft jedes Produkt, ja sogar jedes einzelne Objekt eine eindeutige unbemerkt per Funk auslesbare Seriennummer erhalten. Spätestens damit wird Unbeobachtbarkeit im öffentlichen Raum zur Farce: Bei Vernetzung der entsprechenden Datenbanken kann leicht eine Beziehung zwischen Objekt, Produkt, Kaufdatum und ggf. auch Käufer hergestellt werden, wie in einem gemeinsamen Positionspapier einer Reihe internationaler Privacy-Organisationen herausgestellt wurde [Foeb03]. Zusammen mit anderen Daten entstehen so leicht umfassende Profile über die Gewohnheiten und Interessen der betreffenden Person, welche für interessierte Parteien, beispielsweise den Arbeitgeber oder Versicherungen, einen erheblichen Wert besitzen und von diesen gegen die Interessen des einzelnen verwendet werden können. Zwar sind derartige Verknüpfungen in Europa gesetzlich verboten, aber diese Gesetze lassen sich leicht umgehen.

Bei jeder Beantragung einer Kundenkarte stimmen Sie heutzutage einer weitgehenden Auswertung und nahezu beliebigen Verknüpfung Ihrer Daten zu.

Noch relevanter ist das Thema Privacy in weltweiten Datennetzen wie dem Internet, einfach deswegen, weil regionale Datenschutzgesetze in diesem per se internationalen Medium nicht gelten oder zumindest faktisch nicht durchgesetzt werden können. Deshalb spielen beim Thema Privacy technische Mittel, sogenannte „Privacy Enhancing Technologies“, eine wichtige Rolle. Es werden sozusagen technische Gegenmittel gegen technikgestützte Überwachungsmöglichkeiten entwickelt und zunehmend eingesetzt.

Privacy Policies

Ein Ansatz zur Durchsetzung gewisser rechtlicher Rahmen für die Verarbeitung von personenbezogenen Daten im Internet stellen sogenannte Datenschutzerklärungen dar, welche häufig auf Webangeboten zu finden sind. Dies sind freiwillige Zusicherungen des jeweiligen Anbieters über den Umgang mit den Daten der Besucher, welche üblicherweise die Rechte zur Erhebung, Verarbeitung, Nutzung und Weitergabe der personenbezogenen Daten einschränken. Der Vorteil ist, dass durch diese Erklärungen quasi ein Vertrag zwischen Anbieter und Besucher geschlossen wird, an welchen sich der Anbieter halten muss, selbst wenn die Gesetze des jeweiligen Landes weitergehendes erlauben.

Der Nachteil ist, dass kaum ein Nutzer diese Datenschutzerklärungen liest bzw. versteht. Aus diesem Grund wurde vom WWW Consortium ein Standard für Privacy Preferences, kurz P3P (www.w3c.org/p3p/), eine standardisierte, maschinenlesbare Darstellung von Datenschutzerklärungen entwickelt. Der Browser bzw. weitere Software des Nutzers kann das entsprechende P3P-Dokument herunterladen und dem Nutzer den Inhalt der Erklärung in dessen Sprache und in verständlichen Formulierungen anzeigen. Zudem ist eine automatisierte Entscheidung über die Weitergabe von Daten anhand der Inhalte der Erklärung möglich. Der Internet Explorer wertet beispielsweise eine Kurzform der P3P-Policy aus um zu entscheiden, ob eine Webseite Cookies auf dem Nutzerrechner ablegen darf oder nicht.

Cookies

Cookies sind kleine auf dem Nutzerrechner abgelegte Datenpakete, die i.d.R. eine eindeutige Referenz auf ein pseudonymes Nutzerprofil in der Datenbank des Webseitenanbieters enthalten. Diese Referenz wird bei wiederholtem Besuch der Webseite jahrelang automatisch von Browser mitgesendet und identifiziert auf diese Weise den Besucher [RFC2109]. Cookies werden einerseits eingesetzt, um aufeinanderfolgende Zugriffe des Nutzers zu einer Session zusammenzusetzen und so statusbasierte Dienste wie z.B. einen Warenkorb zu ermöglichen. Andererseits verwenden insbesondere sogenannte Werberinge Cookies, um umfassende Interessensprofile der Nutzer zu erstellen. Dies ermöglicht ein einfacher Trick: Normalerweise wird ein Cookie nur an denjenigen Server zurückgesandt, der diesen vorher im Browser abgelegt hat. Werberinge umgehen diese „Schutzmaßnahme“, indem sie Werbeobjekte auf verschiedenen Webangeboten von einem zentralen Server nachladen lassen, wobei natürlich der Cookie dieses Servers jeweils mitgesendet wird. Anhand des angeforderten Objektes und des Cookies wird ermittelt, welches Nutzerprofil welche Webseiten besucht und diese Informationen für angepasste Werbung verwendet. Der Cookie-Mechanismus ist daher eine der wesentlichsten und allgegenwärtigsten Überwachungstechniken im Internet. Die meisten Browser bieten daher heute die Möglichkeit

zwischen normalen Cookies und „Third Party-Cookies“ zu unterscheiden, die von eingebetteten Objekten wie Flash-Animationen oder Bildern gesetzt werden.

Kryptographie

Als eine der wichtigsten Privacy-Techniken ist sicherlich die Kryptographie zu nennen. Damit ist es möglich Nachrichten so zu blenden bzw. zu verschlüsseln, dass nur der berechtigte Empfänger den Inhalt verstehen kann. Kryptographie ist sozusagen die einzige Möglichkeit, so etwas wie das Briefgeheimnis in offenen Datennetzen zu realisieren. Neben dem Verschlüsseln ist es mit Hilfe von kryptographischen Funktionen auch möglich, die Unverändertheit und den Absender einer Nachricht festzustellen und somit ein Äquivalent zur eigenhändigen Unterschrift zu schaffen (Digitale Signatur).

Auf der Spielwiese der Kryptographie gibt es eine umfangreiche Liste an Techniken, die langfristig helfen sollen, verschiedenste Probleme wie z.B. anonymes digitales Geld, Onlinewahlen, Digitales Rechtemanagement usw. zu lösen. Techniken, die dafür genutzt werden, sind beispielsweise blinde Signaturen [Chau83], bei denen der Signierende später nur noch feststellen kann, dass die Signatur gültig ist, aber nicht mehr, um welche der von ihm gegebenen Signaturen es sich handelt – dies Technik wird beispielsweise für anonyme digitale Münzen, Onlinewahlen oder Credentials verwendet. Credentials sind garantierte Eigenschaften einer nicht namentlich bekannten, pseudonymen Person, wie z.B. Volljährigkeit [CaLy00], allerdings mit der Einschränkung, dass verschiedene Verwendungen eines Credentials nicht miteinander verknüpft werden können.

Weitere, der Kryptographie nahe stehende Techniken sind Steganographie und Watermarking. Bei diesen Techniken geht es darum, zusätzliche, geheime Informationen in anderen Daten, beispielsweise Video- oder Audioströmen zu verbergen. Steganographie verfolgt hierbei das Ziel, die Existenz dieser Daten zu verbergen und dadurch geheime Daten beispielsweise über zensurgefährdete Kanäle zu übertragen. Einer der besten Algorithmen wird in [West01] beschrieben. Wichtiges Argument waren und sind derartige Techniken bei gesellschaftlichen Diskussionen um ein Verbot oder eine Einschränkung von Kryptographie: Wenn es möglich ist, Daten unbemerkt und nicht nachweisbar zu übertragen, führt ein Verbot von Kryptographie nicht zur Verhinderung von Kriminalität, sondern sorgt nur für eine weitere Einschränkung der Privatsphäre des „unbescholtenen Bürgers“.

Watermarking stellt prinzipiell die gleiche Technik dar, nur wird hier weniger Wert auf die Nichterkennbarkeit, sondern viel mehr auf die Robustheit der „eingebetteten“ Daten gelegt. So soll es beispielsweise bei Audiodaten, welche einem Urheberrecht unterliegen, nicht möglich sein, das vorhandene Watermarking zu entfernen, ohne die Qualität der eigentlichen Audiodaten hörbar zu verringern. Angedacht und aus Privacy-Sicht nicht unproblematisch ist beispielsweise eine Codierung der Datei auf den jeweiligen Käufer, so dass eine unberechtigte Weitergabe nachgewiesen werden kann (http://www.heise.de/newsticker/meldung/print/46509).

Privacy Enhancing Technologie

Mit Hilfe von Kryptographie werden Inhalte von Nachrichten geschützt. Auf diese Weise nicht schützbar sind jedoch die Kommunikationsumstände: An wen wurde die E-Mail von Person *x* versandt? Auf welche Webseite hat ein Nutzer zugegriffen?

Derartige Informationen können jedoch oft ausreichen, um auf den Inhalt einer Nachricht ermitteln bzw. herauszufinden, nach welchen Informationen ein Nutzer im Internet sucht. Beispielsweise ist der Besuch des Webangebotes der anonymen Alkoholiker sicher aufschlussreich, auch wenn nicht ermittelt werden kann, welche Webseiten konkret aufgerufen wurden. Normalerweise kann diese Information an jeder Zwischenstation, z.B. beim ISP oder vom Systemadministrator in der Firma ermittelt werden, da alle Datenpakete im Internet mit Absender und Ziel-IP-Adresse verschickt werden – die genannten Parteinen kennen zudem die reale Identität des Nutzers hinter einer IP-Adresse.

Anonymisierungsdienste

Anonymisierungsdienste sind Techniken zum Schutz der Kommunikationsbeziehung, indem der Datenstrom über Zwischenstationen umgeleitet und zumindest einen Teil des Weges, verschlüsselt übertragen wird. Leistungsfähige Dienste wie z.B. JAP (www.anon-online.de) oder TOR (http://freehaven.net/tor/) verteilen dabei das Vertrauen auf mehrere Zwischenstationen mit unabhängigen Betreibern, welche von den Nachrichten sequentiell durchlaufen werden. Im Ergebnis ist eine Anfrage nur dann beobachtbar, wenn alle Zwischenstationen gemeinsam gegen den Nutzer agieren. Die zugrundeliegende Idee dieser Anonymisierungstechnik wird als Chaumsche Mixe bezeichnet [Chau81], wobei ein Mix eine Station ist, die eine bestimmte Menge von Nachrichten sammelt, entschlüsselt und in anderer Reihenfolge weiterleitet. Für die spezielle Verschlüsselung für eine Folge von Mixen ist immer ein besonderes Clientprogramm notwendig. Einfache Anonymisierer wie z.B. Anonymizer.com, arbeiten ohne Clientprogramm allein mit Hilfe des Browser. Dafür existiert nur eine Zwischenstation, der man absolut vertrauen muss – auch in der Hinsicht, dass die ein- und ausgehende Kommunikation dieser Station nicht überwacht wird, da selten die Reihenfolge der Nachrichten verändert wird.
Weitere Anwendung erfährt das Mix-Konzept bei dem anonymen Versenden von E-Mails über sogenannte Mixmaster-Remailer (http://www.obscura.com/~loki/)

Private Database Access

Ein anderer, bisher eher theoretischer Ansatz wird als „Private Information Retrieval“ oder als „Private Database Access“ bezeichnet. Hierbei handelt es sich um Konzepte, die verbergen sollen, welchen Datensatz ein Nutzer aus einer Datenbank abgerufen hat. Dabei darf selbst das Datenbankmanagementsystem oder ein Beobachter der physischen Speicherzugriffe dies nicht ermitteln können. Bisher bekannte Konzepte nutzen ähnlich dem Mixkonzept eine Verteilung des Vertrauens auf mehrere Server z.B. [CoBi_95], rein kryptographische Lösungen z.B. [KO97] oder sichere Komponenten [SS00, AsFr02], die im wesentlichen das DBMS enthalten und zusätzliche Techniken wie Verschlüsselung nutzen, um „verräterische“ Speicherzugriffe zu randomisieren. Bisher leiden diese Konzepte jedoch an einen viel zu hohen Berechnungsaufwand auf der Serverseite zur Beantwortung einer Anfrage. In unserer Forschungsgruppe wird gerade ein Verfahren entwickelt, welches verspricht, den Aufwand erheblich zu senken, wenn der Nutzer statt mit perfekter mit probabilistischer Unbeobachtbarkeit zufrieden ist.

Identitätsmanagement

In der Praxis ist ein anonymes Auftreten nicht immer das Ziel. Oft soll Unbeobachtbarkeit nur gegenüber Außenstehenden erreicht werden, vom Kommunikationspartner jedoch identifiziert oder zumindest wiedererkannt werden können. Zum Nachweis der eigenen Identität lassen sich digitale Signaturen verwenden. Mit Pseudonymen kann man hingegen unter einer Art Künstlernamen auftreten. Dabei besteht die Möglichkeit, das Pseudonym zu wechseln oder bezüglich verschiedenen Partnern verschiedene Pseudonyme zu wählen. Dabei wird zwischen verschieden Klassen von Pseudonymen unterschieden, abhängig davon, in welchen Kontexten und über welche Zeiträume diese eingesetzt werden [PfKö01]. Das Ergebnis ist eine aktive Kontrolle über das Wissen anderer bezüglich der eigenen Person: Idealerweise verwendet man ein Pseudonym nur begrenzte Zeit und gegenüber einem oder wenigen Kommunikationspartnern. Wechselt man das Pseudonym, ist das Wissen der Partner verloren, da diese es nur mit dem bisher verwendeten Pseudonym verknüpfen können. Eine unerlaubte Weitergabe personenbezogener Daten ist ausgeschlossen, wenn man konsequent gegenüber jedem Kommunikations- oder Geschäftspartner ein anderes Pseudonym verwendet. Credentials ermöglichen einen noch flexibleren Einsatz von Pseudonymen: So ist es möglich, bestimmte Eigenschaften wie z.B. Volljährigkeit oder Staatsbürgerschaft zu garantieren, ohne dass immer das gleiche Pseudonym verwendet werden müsste. Eine staatliche Stelle kann beispielsweise einer Person das Credential „Volljährigkeit“ austellen, welches diese auf alle ihre Pseudonyme umrechnen kann. In [ClKö01] werden Identitätsmanagementkonzepte auf Basis von Credentials vorgestellt. Die Herausforderung dabei ist, einerseits die „Umrechnung“ möglichst allgemein und ohne erneute Einbeziehung der ausstellenden Stelle zu ermöglichen, andererseits jedoch effektiv die Umrechnung auf Pseudonyme anderer Personen, also die Weitergabe von Eigenschaften an Personen, die diese Eigenschaft möglicherweise nicht besitzen zu verhindern. Ansätze dazu sind z.B. die Aufdeckung der Identität eines Pseudonyms im Missbrauchsfall oder die Kopplung von erheblichen Werten an das Credential. Zweiteres lässt sich beispielsweise so realisieren, dass ein Credential nur von einer anderen Person genutzt werden kann, wenn diese dazu ein zentrales Geheimnis (den geheimen Schlüssel) erfahren müsste, welcher beispielsweise auch den Zugang zum Bankkonto der betreffenden Person eröffnen würde.

Im obigen Beispiel der Werberinge wird ein gemeinsames Pseudonym (ein Cookie) verwendet, um Aktionen auf verschiedenen Webseiten zu verketten. Zwar kann man dies durch geeignete Maßnahmen vermeiden, aber der Webbrowser unterstützt den Nutzer dabei nicht.

Allgemein werden Geräte, Software oder Systeme als „Identitäts-Manager“ bezeichnet, wenn diese den Nutzer aktiv bei dem Verwalten seiner Pseudonyme unterstützen. Der schwierigste Teil der Aufgabe ist die Erkennung der jeweiligen Situation, ein Abgleich mit protokollierten Nutzungsdaten der vorhandenen Pseudonyme und darauf basierend die Auswahl des Pseudonyms. Einige Tools wie z.B. CookieCooker (www.cookiecooker.de) arbeiten mit der normalen Web-Infrastruktur, welche bisher Identitätsmanagement nur sehr begrenzt unterstützt. So verwaltet CookieCooker lokal gespeicherte Cookies und verwendete Anmeldedaten so, dass der Nutzer aktiv entscheiden kann, unter welcher Identität er gegenüber einer Webseite auftreten möchte und welche Daten er über sich selbst offenbaren will. Zusätzlich werden Dienste genutzt, die sogenannte Wegwerf- oder Einweg-Emailadressen (www.trashmail.net, www.spamgourmet.com) bereitstellen. Damit wird eine pseudonyme Kontaktmöglichkeit geschaffen, die zudem effizient gegen Spam schützt.

Mit Systemen, die Identitätmanagement auf Server- und Clientseite aktiv unterstützen kann jedoch für beide Seiten viel mehr erreicht werden. Die Serverseite erhält insbesondere durch Verwendung von Pseudonymen bzw. Credentials mit besonderen Eigenschaften höhere

Sicherheit, beispielsweise beim Abschluss von Geschäften, der Client hingegen kann den Komfort des Nutzers erhöhen, indem notwendige Daten nicht wiederholt eingegeben werden müssen. Durch pseudonymes Auftreten und durch die Möglichkeit zur automatischen Überprüfung des Vorgangs (z.B. der P3P-Policy) kann die Sicherheit des Nutzers und dessen Daten erhöht werden. Dafür sind aber weitere Standardisierungen erforderlich, um Maschinen oder Programmen eine gegenseitige Kommunikation zu ermöglichen, wie dies für viele andere Projekte auch erforderlich ist. Eine gute Übersicht über Forschungspapiere und Projekte zu Identitätsmanagement findet sich unter http://marit.koehntopp.de/pub/idmanage.

Fazit

Der Schutz der Privatsphäre ist ein elementares Recht jedes Einzelnen. Staatliche und kommerzielle Datensammelaktivitäten müssen daher eingeschränkt und überwacht werden. Vorhandene Privacyprobleme können weder allein auf juristischem Wege noch allein durch Privacy-Techniken gelöst werden. Notwendig ist eine Kombination aus technischen und organisatorischen Maßnahmen und geeigneten rechtlichen Rahmenbedingungen. Offene Fragen sind bei der Abwägung gegensätzlicher gesellschaftlicher Interessen zu klären. So ist totale Anonymität oft ebenso problematisch wie vollständige Überwachung. Die oft geforderte Vorratsdatenspeicherung zum Zwecke der Strafverfolgung kann nicht die Lösung sein, schon weil dabei das rechtsstaatliche Prinzip der Unschuldsvermutung in eine generelle Schuldvermutung umgekehrt würde.

Literatur

[Chau81] David Chaum: Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms; Communications of the ACM 24/2 (1981) 65-75.

[Chau83] David Chaum: Blind Signatures for untraceable payments; Crypto ’82, Plenum Press, New York 1983, 199–203.

[BVerfGE84] BVerfGE: Volkszählung; http://www.datenschutz-berlin.de/gesetze/sonstige/volksz.htm, 1984.

[CoBi_95] David A. Cooper, Kenneth P. Birman: Preserving Privacy in a Network of Mobile Computers; IEEE Symposium on Research in Security and Privacy, IEEE Computer Society Press, Los Alamitos 1995, 26-38.

[KO97] E. Kushilevitz, R. Ostrovsky: Replication is not needed: single database, computationally-private information retrieval; FOCS'97, pp. 364-373.

[RFC2109] D. Kristol, L. Montulli: HTTP State Management Mechanism; Internet RFC 2109, 1997, http://www.faqs.org/rfcs/rfc2109.html.

[SS00] Sean W. Smith, Dave Safford: Practical server privacy with secure coprocessors; IBM Systems Journal, 40(3), September 2001.

[CaLy00] Jan Camenisch, Anna Lysyanskaya: Efficient Non-transferable Anonymous Multi-show Credential System with Optional Anonymity Revocation; Research Report RZ 3295 (#93341), IBM Research; November 2000.

[ClKö01] Sebastian Clauß, Marit Köhntopp: Identity Management and Its Support of Multilateral Security; Computer Networks 37 (2001), Special Issue on Electronic Business Systems; Elsevier, North-Holland 2001; 205-219.

[West01] Andreas Westfeld: F5—A Steganographic Algorithm: High Capacity Despite Better Steganalysis; S. 289–302 in Ira S. Moskowitz (Hrsg.): Information Hiding. 4th International Workshop, IH'01, Pittsburgh, USA, April 2001, Proceedings, LNCS 2137, Springer-Verlag Berlin Heidelberg 2001.

[PfKö01] Andreas Pfitzmann, Marit Köhntopp: Anonymity, Unobservability, and Pseudonymity - A Proposal for Terminology; in: Hannes Federrath (Hg.): Designing Privacy Enhancing Technologies; Proc. Workshop on Design Issues in Anonymity and Unobservability; LNCS 2009; 2001.

[AsFr02] Dmitri Asonov; Johann-Christoph Freytag: Almost optimal private information retrieval; Proceedings of 2nd Workshop on Privacy Enhancing Technologies (PET 2003), San Francisco, USA, April 2002.

[Foeb03] Positionspapier über den Gebrauch von RFID auf und in Konsumgütern; http://www.foebud.org/texte/aktion/rfid/positionspapier.html, 2003.