Ausgleich von Datenschutz und Überwachung mit technischer Zweckbindung am Beispiel eines Pseudonymisierers *

Joachim Biskup
Universität Dortmund
D-44221 Dortmund
biskup@ls6.cs.uni-dortmund.de

Ulrich Flegel
Universität Dortmund
D-44221 Dortmund
ulrich.flegel@udo.edu

Abstract:
Es wird am Beispiel des Datenschutzes und der Überwachung gezeigt, daß gegenläufige Interessen mit Hilfe technischer Maßnahmen zum Ausgleich kommen können. Entsprechende Technologien werden vorgestellt und diskutiert. Anhand eines Pseudonymisierungsverfahrens wird gezeigt, daß durch die technische Durchsetzung der Zweckbindung bei der Pseudonymaufdeckung ein fairer Interessenausgleich zwischen Datenschutz und Überwachung im laufenden Betrieb eines IT-Systems automatisierbar ist. Diese Konzepte wurden implementiert und einer Leistungsanalyse unterzogen. Dabei zeigt sich, daß die Konzepte für Praxisanwendungen tragfähig sind. Die Implementierung wird im Internet frei verfügbar gemacht.

1 Einleitung

Mit den zunehmenden Möglichkeiten der Informations- und Kommunikationstechnologie (IuKT) nimmt auch die Unterstützung und Durchdringung von Vorgängen des gesellschaftlichen Lebens durch IuKT zu. Die Digitalisierung von Abläufen schafft stark erweiterte Potentiale, die sowohl Chancen als auch Risiken mit sich bringen. Durch entsprechende Technikgestaltung sind bereits heute Überwachungsmöglichkeiten als auch anonyme Vorgangsabläufe realisierbar, die weit über die konventionellen Möglichkeiten hinausgehen und dabei zudem geringere Kosten verursachen. Dabei sind sowohl die Überwachungsmöglichkeiten als auch die anonymen Vorgangsabläufe jeweils mit Chancen und Risiken behaftet. Dementsprechend gewinnen Schutzansprüche ebenfalls an Bedeutung. Viele Internet-Nutzer verzichten auf die Nutzung von Diensten, die zwecks Registrierung persönliche Daten erheben [Ros98]. Diverse Studien [CDT02, LHA99, Ros98] zeigen die Wertschätzung anonymer Zugriffsmöglichkeiten auf Dienstangebote im Internet.

Dienstanbieter mögen mehr oder weniger bemüht sein, diesen Nutzererwartungen zu entsprechen. In jedem Fall haben sie nach geltendem Recht datenschutzrechtliche Anforderungen umzusetzen. Eine Umsetzung in der digitalen Welt erfordert spezielle technische

*Die beschriebenen Arbeiten werden derzeit zum Teil von der Deutschen Forschungsgemeinschaft gefördert unter Bi 311/10-2.

Maßnahmen. Denn sind persönliche Daten erst einmal erhoben worden, ist deren Mißbrauch ohne technische Hilfe idR. schwer bis gar nicht nachvollziehbar bzw. verhinderbar.

2 Datenschutzfördernde Technologien

Datenschutzfördernde Technologien, zu neudeutsch privacy enhancing technologies (PETs), finden im Internet bereits Anwendung. In einigen Internet-Browsern ist bereits das P3P-Framework integriert [CLM+01], welches den Austausch personenbezogener Daten in der Anwendungsschicht, nicht jedoch in tieferen Netzwerkprotokoll-Schichten schützt. Hierfür geeignete Anonymisierungsdienste sind z.B. *Anonymizer*, *Anonymouse*, *Proxymate*, *Crowds*, *Onion Routing*, *JAP* sowie *Cypherpunk*- und *Mixmaster*-Remailer [EP01, FH01, BFK00]. Mit Hilfe dieser Technologien kann der Nutzer nach eigener Wahl Dienstangebote anonym in Anspruch nehmen.
Die genannten Technologien werden allein vom Nutzer und ggf. von ihm vertrauten Dritten kontrolliert. Diese Eigenschaft kann Vor- und Nachteile haben. Um Anonymität zu erlangen, muß der Nutzer die entsprechenden Technologien kennen und bedienen können. Datenschutzrechtliche Vorgaben sind von Dienstanbietern jedoch unabhängig von der technischen Versiertheit der Nutzer umzusetzen. Mithin ist bereits anbieterseitig die Anwendung von PETs notwendig.

Die meisten modernen anbieterseitig eingesetzten Betriebssysteme erzeugen Audit- bzw. Log-Daten, die Personenbezüge aufweisen. Dies gilt insbesondere für Betriebssysteme, die im Hinblick auf Konformität mit den Trusted Computer System Evaluation Criteria (TCSEC) [Cen85, Cen87] (`C2` oder höher) oder den Common Criteria (CC) [Boa99] (`FAU_GEN1.2` und `FAU_GEN2`) entworfen wurden. Meist ist es nicht möglich, das vorliegende System dergestalt zu konfigurieren, daß es zwar alle betriebsnotwendigen Meldungen generiert, aber dabei keine Personenbezüge in Audit- bzw. Log-Daten nachhält. Einen Lösungsansatz hierfür bildet die Anonymisierung oder Pseudonymisierung personenbezogener Audit- bzw. Log-Daten.
Der *Anonymouse log file anonymizer* [EP01] vergröbert personenbezogene Daten, oder er ersetzt sie durch Standardwerte, so daß die restlichen Log-Daten auswertbar bleiben. Allerdings können die ursprünglichen Personenbezüge nicht wiederhergestellt werden.
Dies ist nicht nur ein Nachteil des *Anonymouse log file anonymizer*, sondern auch der anderen oben beschriebenen Anonymisierungsdienste. Einerseits braucht deren Nutzer hinsichtlich seiner Anonymität den Anbietern anonym genutzter Dienste kein Vertrauen entgegenbringen. Andererseits besteht auch bei mißbräuchlicher Nutzung für den Anbieter oder für Ermittlungsbehörden nahezu keine Möglichkeit, die Anonymität aufzuheben, um Zurechenbarkeit herzustellen. Der von uns entwickelte Pseudonymisierer schließt diese Lücke.

Technologien für mehrseitige Sicherheit: Es besteht ein Spannungsfeld zwischen dem Interesse einzelner Nutzer an Datenschutz und Anonymität einerseits und der Zurechenbarkeit andererseits, um im Mißbrauchsfall die Interessen anderer beteiligter Parteien schützen zu können. Wie etwa die Diskussion in [Fie01, Roß02, Fie02] deutlich macht, kann eine für die beteiligten Parteien zufriedenstellende Lösung i.a. nicht darin bestehen, eine

der beiden Anforderungen zugunsten der anderen vollständig aufzugeben.
Vielmehr scheint ein fairer Ausgleich der Interessen aller beteiligter Parteien unter Berücksichtigung der jeweiligen Anwendungssituation erstrebenswert. IuKT-Technologien, die anwendungsspezifisch eine ebensolche Balance widerstreitender Sicherheits-Interessen erreichen, bezeichnen wir als Technologien zur Herstellung *mehrseitiger Sicherheit* [RPM99].

Intrusion Detection Systeme befinden sich von Grund auf in dem oben beschriebenen Interessenkonflikt (siehe auch [Fie01]). Ansätze für mehrseitig sichere Intrusion Detection findet man etwa bei den Systemen IDA, AID, und ANIDA [FH01, BK99], welche pseudonymisierte Audit-Daten analysieren und bei Bedarf Personenbezüge aufdecken. Wünschenswert ist die Pseudonymaufdeckung im Zusammenhang mit einem hinreichenden Anfangsverdacht für einen Angriff bzw. Angriffsversuch, um (weiteren) Schaden abwehren zu können. Der Zweck der Pseudonym-Aufdeckung ist hier also mit dem Vorliegen von Anfangsverdachten verbunden. Obige Systeme bieten keinen technischen Schutz gegen eine Pseudonym-Aufdeckung in beliebigen anderen Situationen für andere, auch unrechtmäßige Zwecke. Dieses Problem löst unser Pseudonymisierer durch die technische Durchsetzung der Zweckbindung bei der Pseudonymaufdeckung.

Technische Zweckbindung: Ein wesentliches Merkmal unseres Pseudonymisierers ist die unverzügliche Pseudonym-Aufdeckbarkeit im Fall eines vorliegenden Anfangsverdachts. Dies geschieht unter der Annahme, daß bei Vorliegen eines hinreichenden, a priori zu definierenden Anfangsverdachts die Herstellung von Zurechenbarkeit rechtmäßig ist. Da die Zurechenbarkeit ausschließlich im Fall eines Anfangsverdachts hergestellt werden kann und auf die mit dem Verdacht assoziierten Personenbezüge beschränkt bleibt, besteht ein fairer Ausgleich mit dem Nutzerinteresse an Pseudonymität.

Ähnliche Annahmen und Konzepte findet man bei fairen elektronischen offline-Zahlungssystemen. Diese erlauben anonyme Zahlungsvorgänge, solange das Zahlungsmittel, z.B. eine elektronische Münze, nicht öfter als erlaubt eingelöst wird. Versucht jedoch jemand, eine Münze mehrfach einzulösen, ist die Anonymität dieser Zahlungsvorgänge aufhebbar [Pet97, DFTY97]. Dem Themenkomplex der elektronischen offline-Zahlungsmittel verwandt sind anonyme Credentials und anonyme Authentisierungsprotokolle, die Zurechenbarkeit gewährleisten [Bra00, CL01]. Allerdings wird dabei die Zweckbindung idR. nicht technisch, sondern organisatorisch mit Hilfe vertrauenswürdiger Dritter durchgesetzt.
Sowohl die erwähnten Zahlungssysteme als auch die Systeme für anonyme Authentisierung bieten dem Nutzer sehr vorteilhafte Vertrauensmodelle. Jedoch setzen sie auf der anderen Seite eine Infrastruktur voraus, um Nutzer zu registrieren und um eine breite Einsetzbarkeit der Zahlungsmittel bzw. Credentials zu gewährleisten. Da noch nicht absehbar ist, welche Systeme sich am Markt durchsetzen werden, ist die derzeitige Zurückhaltung beim Aufbau einer solchen kostenintensiven Infrastruktur verständlich. In Ermangelung einer breitenwirksamen Infrastruktur können Dienstanbieter im Internet auf Pseudonymisierer zurückgreifen, wie hier vorgeschlagen, um unabhängig von nutzerseitigen Maßnahmen datenschutzrechtliche Vorgaben im Sinne mehrseitiger Sicherheit zu erfüllen.

3 Der Pseudonymisierer

Die Pseudonymisierung mit Aufdeckbarkeit bei technischer Durchsetzung der Zweckbindung basiert auf der Annahme, daß es legitim ist, diejenigen Pseudonyme offen zu legen, die an einem hinreichenden Anfangsverdacht für einen Angriff auf das IT-System beteiligt sind. Ein hinreichender Anfangsverdacht ist dabei definiert als ein Schwellenwert über gewichteten Beobachtungen potentiell angriffsbezogener Aktivitäten. Überschreiten die beobachteten Aktivitäten den Schwellenwert, so sollen die enthaltenen Pseudonyme aufdeckbar sein, sonst aber nicht.

Grundlage für die technische Realisierung derartiger Pseudonyme ist in unserem Ansatz das Shamir'sche Schwellenwert-Schema zur informationstheoretisch sicheren Geheimnisteilung [Sha79]. Ein in einer beobachteten Aktivität eingebettetes und zu pseudonymisierendes personenbezogenes Datum wird zunächst verschlüsselt, um es zu verbergen. Dann wird das Shamir'sche Schema leicht modifiziert eingesetzt, um den kryptographischen Dechiffrier-Schlüssel des verschlüsselten personenbezogenen Datums kryptographisch in Anteile aufzuteilen. Jedes Vorkommnis des verschlüsselten personenbezogenen Datums, das einem bestimmten Anfangsverdacht zugeordnet ist, erhält einen eigenen Anteil des Dechiffrier-Schlüssels. Überschreitet die Anzahl der vorliegenden Anteile den Schwellenwert des Anfangsverdachts, so kann der Dechiffrier-Schlüssel per Lagrange-Interpolation zurückgewonnen und das personenbezogene Datum entschlüsselt werden. Detaillierte technische Ausführungen des Verfahrens wurden in [BF00] veröffentlicht.

Die Implementierung des Pseudonymisierers ist für Unix-Systeme ausgelegt und wurde unter folgenden Betriebssystemen getestet: Solaris, OpenBSD und Linux. Der Pseudonymisierer kann benutzt werden, um Log-Daten im *Syslog*-Format zu pseudonymisieren, z.B. auch Web-Server Log-Daten. *Syslog* ist auf allen wichtigen modernen Unix-Systemen verfügbar. Auch Windows-Systeme und sehr viele Netzwerk-Komponenten können in eine Syslog-Infrastruktur eingebunden werden. Dementsprechend hoch ist die erreichbare Abdeckung anfallender Log-Daten durch den Pseudonymisierer. Der Installationsaufwand des Werkzeugs ist sehr gering.
Der *Syslog*-Dienst wird so konfiguriert, daß ausschließlich betriebsnotwendige Log-Daten erhoben werden. Diese werden direkt dem lokalen Pseudonymisierer zugeführt, welcher sie inspiziert und personenbezogene Daten unter Erhaltung des Log-Daten-Formats und gemäß obiger Ausführung durch Pseudonyme ersetzt. Bei diesem Ablauf werden die personenbezogenen Daten nie offen in einer Datei gespeichert. Die pseudonymisierten Log-Daten werden zwecks Auswertung an eine zentrale Sammelstelle weitergeleitet. Dort können die pseudonymisierten Log-Daten analysiert werden. Im Bedarfsfall können nur diejenigen Pseudonyme mit Hilfe unseres Reidentifizierers aufgedeckt werden, die im Zusammenhang mit hinreichenden Anfangsverdachten auftreten.

Vertrauensmodell: Generell sind *Syslog*-Log-Daten integer, solange kein Angreifer hinreichende Privilegien auf der erhebenden Maschine erlangt hat, welche es ihm erlauben, Log-Daten zu löschen oder zu korrumpieren. Entsprechendes gilt für die auf derselben Maschine erzeugten Pseudonyme.

Der Pseudonymisierer bringt die Schutzinteressen Pseudonymität und Zurechenbarkeit

zum Ausgleich. Auf der einen Seite soll der Datenschutz gewahrt bleiben, und die Nutzer möchten einen Dienst unerkannt in Anspruch nehmen. Auf der anderen Seite ist Zurechenbarkeit in Mißbrauchsfällen notwendig, um die Interessen anderer Parteien zu wahren. Die Herstellung von Zurechenbarkeit in Mißbrauchsfällen fällt meist in den Aufgabenbereich der Sicherheitsadministratoren des Dienstanbieters. Die Herstellung von Zurechenbarkeit für andere Zwecke, etwa für Direktmarketing, bleibt hier unberücksichtigt.
Da sich die genannten Interessen der Nutzer und der Sicherheitsadministratoren diametral gegenüberstehen, darf im Sinne eines fairen Interessenausgleichs die Erhebung und die Pseudonymisierung nicht unter der Kontrolle einer dieser beiden Parteien stehen. In unserem Ansatz vertrauen beide Parteien ihre Interessen dem Datenschutzbeauftragten des Anbieters an. Dieser modelliert, möglicherweise in Zusammenarbeit mit dem Betriebsrat und im Sinne der Nutzerinteressen a priori das Wissen des Pseudonymisierers hinsichtlich zu pseudonymisierender personenbezogener Daten. In Zusammenarbeit mit den Sicherheitsadministratoren modelliert der Datenschutzbeauftragte a priori die Anfangsverdachte, die eine Pseudonymaufdeckung rechtfertigen. Dieses Wissen stellt der Datenschutzbeauftragte dem von ihm kontrollierten *Syslog*-Dienst und dem von ihm kontrollierten Pseudonymisierer in Form von Konfigurationsdaten zur Verfügung.
Diese Vorgehensweise ist geeignet, um die widerstreitenden Interessen im Vorfeld detailliert zu erfassen und dann zu implementieren, wenn eine Einigung erzielt wurde. Die Durchsetzung der so im vorhinein genau spezifizierten Interessen erfolgt im laufenden Betrieb automatisch und sicher durch den Pseudonymisierer, sofern die oben spezifizierten Kontrollverhältnisse wirksam sind.

Laufzeitverhalten: Der Pseudonymisierer läuft ununterbrochen im System mit und verarbeitet die Log-Daten sofort nach ihrer Erhebung. Dabei ist es wichtig, daß er hinreichend schnell arbeitet und ein Daten-Rückstau lediglich temporärer Natur ist. Der Reidentifizierer hingegen muß keine so strengen Laufzeitanforderungen erfüllen, da er nur vereinzelt und gezielt zum Einsatz kommt.

Um ein realistisches Maß für übliche Log-Datenvolumina zu erhalten, wurden die *Syslog*- und Web-Server-Log-Daten eines Zentralen Servers am Zentrum für Kommunikation und Informationsverarbeitung der Universität Dortmund ausgewertet. Die betrachtete Solaris SUN Ultra Enterprise 4000 Maschine verfügt über sechs Ultra SPARC 168MHz CPUs, 3GB RAM und drei Platten-Arrays mit insgesamt 396GB sowie eine 100Mbps Netzwerkkarte. Während der Arbeitszeit sind von den 1050 registrierten Nutzern durchschnittlich 25 Nutzer auf der Maschine aktiv tätig. Die Maschine trägt 37 weltweit erreichbare Web-Server, einen FTP-Server mit monatlich 112000 Transfers und einem Transfervolumen von 12GB, sowie Email-Dienste im Umfang von monatlich 45000 Emails. Die beobachteten maximalen stündlichen Log-Datenaufkommen betragen 33506 Datensätze/h (9.31 Datensätze/s) für die Gesamtheit der Web-Server und 4956 Datensätze/h (1.38 Datensätze/s) für *Syslog*.

Laufzeitmessungen des Pseudonymisierers wurden bei einer Schlüssel- und Zahlenlänge von 128 Bit auf einem OpenBSD-Rechner mit einer Pentium III 650MHz CPU mit 256MB RAM und einer 100Mbps Netzwerkkarte durchgeführt. Umfangreiche Makrobenchmarks des gesamten Pseudonymisierers zeigen eine Verarbeitungsgeschwindigkeit zwischen 1060 und 70 Log-Datensätzen pro Sekunde. Die Verarbeitungsgeschwindigkeit

ist von verschiedenen Parametern abhängig, z.B. der Anzahl der zu pseudonymisierenden Personenbezüge je Datensatz. Es wird jedoch deutlich, daß selbst für ungünstige Parameterwerte der Durchsatz des Pseudonymisierers deutlich über den maximalen anfallenden Log-Datenvolumina liegt.

4 Schlußwort

Es wurde gezeigt, daß die widerstreitenden Interessen der Nutzer an Datenschutz einerseits und andererseits an der Überwachung zur Wahrung der Interessen anderer Parteien – im Fall eines Mißbrauchs durch Nutzer – gewahrt werden können. Der Interessenausgleich beruht darauf, daß Nutzer pseudonym und damit als Individuum unbeobachtet agieren können. Besteht jedoch ein hinreichender Verdacht, daß Interessen anderer in Form eines Angriffs auf das IT-System gefährdet sind, lassen sich die betreffenden Pseudonyme aufdecken und Zurechenbarkeit herstellen. Die Überwachung hinsichtlich Anfangsverdachten basiert auf pseudonymisierten Log-Daten.

Wir haben für diese Konzepte eine Implementierung angegeben und gezeigt, daß ihr Laufzeitverhalten für die praktische Anwendung geeignet ist. Die Implementierung gestattet es Dienstanbietern, den Ausgleich von Datenschutz und Überwachung durchzuführen, ohne daß die Dienstnutzer dafür zusätzliche Maßnahmen treffen müssen. Die Implementierung wird im Internet frei verfügbar gemacht.

Literaturverzeichnis

[BF00] Joachim Biskup and Ulrich Flegel. Threshold-based Identity Recovery for Privacy Enhanced Applications. In Sushil Jajodia and Pierangela Samarati, editors, *Proceedings of the 7th ACM Conference on Computer and Communications Security*, pages 71–79, Athens, Greece, November 2000. ACM SIGSAC, ACM Press.

[BFK00] Oliver Berthold, Hannes Federrath, and Marit Köhntopp. Project "Anonymity and Unobservability in the Internet". In *Workshop on Freedom and Privacy by Design / Conference on Freedom and Privacy*, pages 57–65, Toronto, Canada, April 2000. ACM.

[BK99] Roland Büschkes and Dogan Kesdogan. Privacy Enhanced Intrusion Detection. In Günter Müller and Kai Rannenberg, editors, *Multilateral Security in Communications*, Information Security, pages 187–204. Addison Wesley, 1999.

[Boa99] Common Criteria Implementation Board, editor. *Common Criteria for Information Technology Security Evaluation — Part 2: Security functional requirements, Version 2.1*. Number CCIMB-99-032. National Institute of Standards and Technology, August 1999. `http://csrc.ncsl.nist.gov/cc/ccv20/p2-v21.pdf`.

[Bra00] Stefan A. Brands. *Rethinking Public Key Infrastructures and Digital Certificates: Building in Privacy*. MIT Press, Cambridge, Massachusetts, 2000.

[CDT02] Privacy Survey Results, January 2002. `http://www.cdt.org/privacy/survey/findings/`.

[Cen85] National Computer Security Center. US DoD Standard: Department of Defense Trusted Computer System Evaluation Criteria. DOD 5200.28-STD, Supercedes CSC-STD-001-83, dtd 15 Aug 83, Library No. S225,711, December 1985. http://csrc.ncsl.nist.gov/secpubs/rainbow/std001.txt.

[Cen87] National Computer Security Center. Audit in Trusted Systems. NCSC-TG-001, Library No. S-228,470, July 1987. http://csrc.ncsl.nist.gov/secpubs/rainbow/tg001.txt.

[CL01] Jan Camenisch and Anna Lysyanskaya. An Efficient System for Non-transferable Anonymous Credentials with Optional Anonymity Revocation. In B. Pfitzmann, editor, *Advances in Cryptology – EUROCRYPT 2001*, number 2045 in Lecture Notes in Computer Science, pages 93–118, Austria, May 2001. Springer.

[CLM+01] Lorrie Cranor, Marc Langheinrich, Massimo Marchiori, Martin Presler-Marshall, and Joseph Reagle. The Platform for Privacy Preferences 1.0 (P3P1.0) Specification, September 2001. http://www.w3.org/TR/2001/WD-P3P-20010928/.

[DFTY97] George Davida, Yair Frankel, Yiannis Tsiounis, and Moti Yung. Anonymity Control in E-Cash Systems. In R. Hirschfeld, editor, *Proceedings of the First International Conference on Financial Cryptography (FC'97)*, number 1318 in Lecture Notes in Computer Science, pages 1–16, Anguilla, British West Indies, February 1997. Springer.

[EP01] Claudia Eckert and Alexander Pircher. Internet Anonymity: Problems and Solutions. In Michel Dupuy and Pierre Paradinas, editors, *Proceedings of the IFIP TC11 16th International Conference on Information Security (IFIP/Sec'01)*, pages 35–50, Paris, France, June 2001. IFIP, Kluwer Academic Publishers.

[FH01] Simone Fischer-Hübner. *IT-Security and Privacy: Design and Use of Privacy-Enhancing Security Mechanisms*. Number 1958 in Lecture Notes in Computer Science. Springer, 2001.

[Fie01] Herbert Fiedler. Der Staat im Cyberspace. *Informatik Spektrum*, 24(5):309–314, 2001.

[Fie02] Herbert Fiedler. Cyber-libertär. *Informatik Spektrum*, 25(3):215–219, 2002.

[LHA99] Inc. Louis Harris & Associates. IBM Multi-National Consumer Privacy Survey. Technical Report 938568, IBM Global Services, 1999.

[Pet97] Holger Petersen. Faires elektronisches Geld (in German). In *Mit Sicherheit in die Informationsgesellschaft*, pages 427–444, Bonn, Germany, April 1997. Bundesamt für Sicherheit in der Informationstechnik, SecuMedia Verlag, Ingelheim.

[Roß02] Alexander Roßnagel. Freiheit im Cyberspace. *Informatik Spektrum*, 25(1):33–38, 2002.

[Ros98] Jarek et al. Rossignac. GVU's 10th WWW User Survey, December 1998. http://www.cc.gatech.edu/gvu/user_surveys/survey-1998-10/graphs/graphs.%html#privacy.

[RPM99] Kai Rannenberg, Andreas Pfitzmann, and Günther Müller. IT Security and Multilateral Security. In Günter Müller and Kai Rannenberg, editors, *Multilateral Security in Communications*, Information Security, pages 21–29. Addison Wesley, 1999.

[Sha79] A. Shamir. How to Share a Secret. *Communications of the ACM*, 22:612–613, 1979.