

Ronald Petrlc
Christoph Sorge

Datenschutz

Einführung in technischen Datenschutz,
Datenschutzrecht und angewandte
Kryptographie



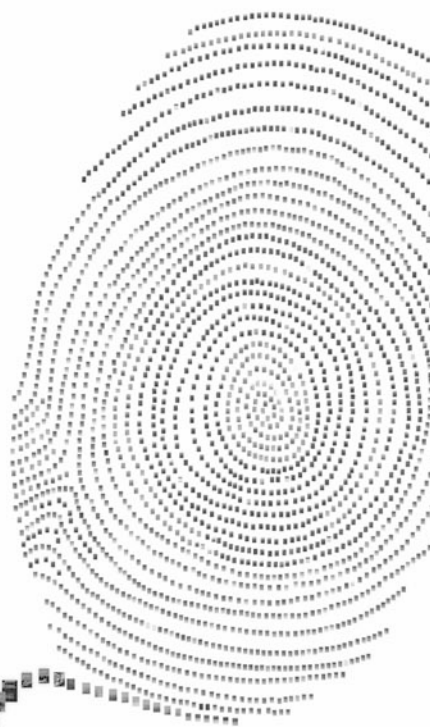
Springer Vieweg

Datenschutz

Lizenz zum Wissen.




Sichern Sie sich umfassendes Technikwissen mit Sofortzugriff auf tausende Fachbücher und Fachzeitschriften aus den Bereichen: Automobiltechnik, Maschinenbau, Energie + Umwelt, E-Technik, Informatik + IT und Bauwesen.

Exklusiv für Leser von Springer-Fachbüchern: Testen Sie Springer für Professionals 30 Tage unverbindlich. Nutzen Sie dazu im Bestellverlauf Ihren persönlichen Aktionscode **C0005406** auf www.springerprofessional.de/buchaktion/



**Jetzt
30 Tage
testen!**

Springer für Professionals.
Digitale Fachbibliothek. Themen-Scout. Knowledge-Manager.

-  Zugriff auf tausende von Fachbüchern und Fachzeitschriften
-  Selektion, Komprimierung und Verknüpfung relevanter Themen durch Fachredaktionen
-  Tools zur persönlichen Wissensorganisation und Vernetzung

www.entschieden-intelligenter.de

Springer für Professionals

 Springer

Ronald Petrlc • Christoph Sorge

Datenschutz

Einführung in technischen Datenschutz,
Datenschutzrecht und angewandte
Kryptographie

Ronald Petrlc
Der Landesbeauftragte für den Datenschutz,
Baden-Württemberg
Stuttgart, Deutschland

Christoph Sorge
Universität des Saarlandes
Saarbrücken, Deutschland

ISBN: 978-3-658-16838-4
DOI 10.1007/978-3-658-16839-1

ISBN: 978-3-658-16839-1 (eBook)

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer Vieweg

© Springer Fachmedien Wiesbaden GmbH 2017

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften. Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag, noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer Vieweg ist Teil von Springer Nature

Die eingetragene Gesellschaft ist Springer Fachmedien Wiesbaden GmbH

Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Strasse 46, 65189 Wiesbaden, Germany

Vorwort

Seit vielen Jahren erforschen wir datenschutzfördernde Technologien – sogenannte „Privacy-Enhancing Technologies“. Wir sind der Meinung, dass eine datenschutzfreundliche Technikgestaltung ein vielversprechender Ansatz zur Gewährleistung des informationellen Selbstbestimmungsrechts der Nutzer ist. Aus diesem Grund lehren wir zum Thema Datenschutz, ebenfalls seit einigen Jahren. Und dies mit großem Erfolg. Unsere Vorlesungen zum Datenschutz erfreuen sich größter Beliebtheit – nicht erst seit den Enthüllungen von EDWARD SNOWDEN. Die Studierenden haben großes Interesse daran, zu erfahren, wie der Datenschutz in heutigen Systemen auf vielerlei Hinsicht ausgehöhlt wird, welche Maßnahmen zum Selbstschutz und welche Verfahren zur datenschutzgerechten Technikgestaltung existieren. Da bei der Technikgestaltung auch die gesetzlichen Vorgaben berücksichtigt werden müssen, beschäftigen wir uns in unseren Vorlesungen auch mit dem (komplexen) Thema Datenschutzrecht – aufbereitet in einer Form, wie es für „Techniker“ verständlich wird.

Mit der neuen EU-Datenschutzgrundverordnung gewinnt „Privacy by Design“ – also der „Datenschutz durch Technik“ – erstmals auch aus gesetzlicher Sicht enorm an Bedeutung. Datenschutz wird nicht mehr überwiegend ein rein juristisches Thema sein, sondern auch die Entwickler werden sich mit dem Thema beschäftigen müssen.

Umso erstaunlicher ist für uns die Tatsache, dass es bisher keine Lehrbücher zum Technischen Datenschutz gibt. Diese Lücke möchten wir nun mit diesem Lehrbuch schließen. Wir hoffen, dass sich das Thema Datenschutz zukünftig in mehr Lehrplänen von technischen Studiengängen an Universitäten und Hochschulen wiederfindet, als dies heute noch der Fall ist. Die Absolventen sollten über das nötige technische Know-How verfügen, um die gesetzlichen Vorgaben bei der Entwicklung neuer Technologien berücksichtigen zu können. Wir hoffen, dass dieses Lehrbuch Sie dabei unterstützt.

November 2016

Dr. Ronald Petrlic
Stuttgart, Deutschland
Prof. Dr. Christoph Sorge
Saarbrücken, Deutschland

Abkürzungsverzeichnis

ABC	Attribute-Based Credentials
ACS	Anonymous Credential System
ADV	Auftragsdatenverarbeitung
AES	Advanced Encryption Standard
BAC	Basic Access Control
BDSG	Bundesdatenschutzgesetz
BSI	Bundesamt für Sicherheit in der Informationstechnik
CA	Certificate Authority
CAN	Card Access Number
CVCA	Country Verifying CA
DAA	Direct Anonymous Attestation
DES	Data Encryption Standard
DH	Diffie-Hellman
DNT	Do Not Track
DVCA	Document Verifying CA
EAC	Extended Access Control
ECDH	Elliptic curve Diffie-Hellman
EFF	Electronic Frontier Foundation
ENISA	Europäische Agentur für Netz- und Informationssicherheit
ESP	Encapsulating Security Payload
GG	Grundgesetz
HIPPA	Health Insurance Portability and Accountability Act
HKDF	HMAC-based Extract-and-Expand Key Derivation Function
HTTP	Hypertext Transfer Protocol
IdM	Identitätsmanagement
IdP	Identitätsprovider
IM	Instant Messaging
ISP	Internet Service Provider
MAC	Message Authentication Code
MITM	Man-in-the-Middle
MPC	Secure Multiparty Computation

OTR	Off-the-Record
PACE	Password Authenticated Connection Establishment
PbD	Privacy by Design
PFS	Perfect Forward Secrecy
PET	Privacy-Enhancing Technology
PGP	Pretty Good Privacy
PIN	Personal Identification Number
PIR	Private Information Retrieval
PKI	Public Key Infrastructure
PSI	Private Set Intersection
PT	Payment Token
RFC	Request for Comments
S/MIME	Secure/Multipurpose Internet Mail Extensions
SMP	Socialist Millionaire's Protocol
SOP	Same Origin Policy
SRTP	Secure Real-Time Transport Protocol
SSL	Secure Sockets Layer
SSO	Single-Sign-On
TMG	Telemediengesetz
TKG	Telekommunikationsgesetz
TLS	Transport Layer Security
TPM	Trusted Platform Module
TTP	Trusted Third Party
URL	Uniform Resource Locator
VPN	Virtual Private Network
ZKP	Zero-Knowledge Proof

Inhaltsverzeichnis

- 1 Einführung** 1
 - 1.1 Haben wir etwas zu verbergen?..... 2
 - 1.2 Säulen des Datenschutzes 4
 - 1.3 Themen dieses Buchs und Lernziele 5
 - 1.4 Danksagung..... 7
 - Literatur 7

- 2 Einführung in den Technischen Datenschutz** 9
 - 2.1 Schutzziele..... 9
 - 2.1.1 „Klassische“ IT-Sicherheits-Schutzziele 10
 - 2.1.2 „Neue“ Datenschutz-Schutzziele 10
 - 2.2 Begriffsbestimmungen 11
 - 2.2.1 Begriff des Datenschutzes..... 11
 - 2.2.2 Begriffe zum technischen Datenschutz 12
 - 2.3 Grundlegende kryptographische Verfahren 14
 - 2.3.1 Verschlüsselung..... 14
 - 2.3.2 Digitale Signatur..... 17
 - 2.3.3 Blinde Signatur 18
 - 2.3.4 Kryptographische Hashfunktion 19
 - 2.3.5 Diffie-Hellman-Verfahren 20
 - 2.4 Grundlegende Verfahren aus der IT-Sicherheit 22
 - 2.4.1 Transport Layer Security 22
 - 2.4.2 Virtual Private Networks 24
 - 2.5 Fazit 26
 - 2.6 Übungsaufgaben..... 26
 - Literatur 26

- 3 Anonymitätsmaße** 27
 - 3.1 Überblick 27
 - 3.1.1 Anonymitäts-Modelle 28
 - 3.1.2 Quasi-Identifikatoren 30

3.2	k-Anonymität	32
3.2.1	Generalisierung von Daten	33
3.2.2	Angriffe auf k-Anonymität	34
3.2.3	l-Diversität	37
3.3	Differential Privacy	38
3.4	Anonymisierung in der Praxis	40
3.5	Fazit	40
3.6	Übungsaufgaben	41
	Literatur	43
4	Anonymität im Internet	45
4.1	Verkehrsflussanalyse	46
4.1.1	Angreiferklassifikation	46
4.1.2	Beispiel: Ablauf der Ticketbestellung	47
4.1.3	Beispiel: Mögliche Gegenmaßnahme	48
4.2	Mixes	49
4.2.1	Verfahren	50
4.2.2	Analyse	50
4.3	Mix-Kaskaden	51
4.3.1	Verfahren	51
4.3.2	Analyse	53
4.3.3	Antwort-Nachrichten	53
4.4	Onion Routing / Tor	55
4.4.1	Grundkonzept von Tor	55
4.4.2	Tor-Zellen	56
4.4.3	Aufbau eines Circuits	56
4.4.4	Leaky Pipe	59
4.4.5	Missbrauch von Tor	59
4.4.6	Hidden Services	60
4.4.7	Angriffe auf Tor	61
4.4.8	Zensurresistenz mit Tor	63
4.5	Fazit	63
4.6	Übungsaufgaben	64
	Literatur	64
5	Identitätsmanagement	67
5.1	Überblick	68
5.1.1	Schwerpunkte und Sichtweisen im Identitätsmanagement	68
5.2	OpenID	69
5.2.1	Ablauf der Authentifizierung	70
5.2.2	Analyse	70

5.3	OAuth	71
5.3.1	Verfahren	72
5.3.2	Analyse	73
5.4	OpenID Connect	74
5.5	Fazit	74
5.6	Übungsaufgaben	75
	Literatur	75
6	Anonymes Bezahlen	77
6.1	Anforderungen an ein anonymes Bezahlverfahren	77
6.2	Anonymes Bezahlen nach Chaum	78
6.2.1	Verfahren im Überblick	79
6.2.2	Bewertung	81
6.3	Bitcoin	82
6.3.1	Anonymität von Bitcoin	84
6.4	Anonymes Bezahlen in der Praxis	85
6.4.1	Geldkarte	85
6.4.2	Prepaid-Karten	86
6.5	Fazit	86
6.6	Übungsaufgaben	87
	Literatur	87
7	Datenschutz im World Wide Web	89
7.1	Tracking im Web	89
7.1.1	Cookies	90
7.1.2	Tracking-Pixel	93
7.1.3	Device Fingerprinting	94
7.1.4	History Hijacking	94
7.1.5	P3P	95
7.2	Social Plugins	95
7.3	Fazit	96
7.4	Übungsaufgaben	96
8	Instant Messaging	97
8.1	Abgrenzung des Instant Messagings von E-Mail	97
8.1.1	Schutzziele bei der E-Mail-Sicherheit	98
8.1.2	Schutzziele beim Instant Messaging	98
8.2	Off-the-record Messaging	98
8.2.1	Protokoll	99
8.2.2	Implementierung	101
8.2.3	Angriffe auf OTR Messaging	101
8.2.4	SIGMA-Protokoll	102

8.3	WhatsApp.....	103
8.3.1	Signal-Protokoll	104
8.3.2	Medien-Verschlüsselung	105
8.3.3	Sichere Telefonie	106
8.3.4	Schlüssel-Verifikation	106
8.3.5	Datenschutzrechtliche Probleme	106
8.4	Fazit	106
8.5	Übungsaufgaben	107
	Literatur	108
9	Elektronische Ausweisdokumente	109
9.1	Elektronischer Reisepass	110
9.1.1	Passive Authentication	110
9.1.2	Basic Access Control	111
9.1.3	Extended Access Control	113
9.2	Elektronischer Personalausweis	116
9.2.1	PACE	117
9.2.2	Extended Access Control Version 2	118
9.2.3	Restricted Identification	120
9.2.4	Weitere Anwendungen	122
9.2.5	Exkurs: Elektronische Signaturen	124
9.2.6	Administrative Aspekte	125
9.3	Fazit	126
9.4	Übungsaufgaben	127
	Literatur	128
10	Weitere kryptographische Verfahren für PETs	129
10.1	Weitere Signaturverfahren	129
10.1.1	Gruppensignatur	130
10.1.2	Ringsignatur	131
10.2	Secure Multiparty Computation	131
10.2.1	Klassische MPC-Protokolle	132
10.2.2	Anwendungen der Secure Multiparty Computation	133
10.3	Zero-Knowledge Proof	134
10.4	Anonyme Berechtigungsnachweise	135
10.4.1	Probleme	135
10.4.2	Verfahren	136
10.5	Fazit	137
10.6	Übungsaufgaben	137
	Literatur	137

11	Datenschutzrecht	139
11.1	Geschichte des Datenschutzes	140
11.2	Datenschutz im Grundgesetz	142
11.3	Bundesdatenschutzgesetz	144
11.3.1	Personenbezogene Daten	145
11.3.2	Zweck und Anwendungsbereich	146
11.3.3	Weitere Begriffsbestimmungen	148
11.3.4	Grundkonzept des deutschen Datenschutzrechts	149
11.3.5	Auskunftsanspruch	151
11.4	Bereichsspezifischer Datenschutz	152
11.4.1	Anwendungsbereich	152
11.4.2	Telemediengesetz	153
11.4.3	Telekommunikationsgesetz	158
11.5	Datenschutz-Grundverordnung	161
11.6	Datenschutzrechtliche Einzelfragen	163
11.6.1	Personenbezug bei IP-Adressen?	163
11.6.2	Einwilligung	164
11.6.3	Anwendungsbereich des TMG	165
11.6.4	Anwendung TMG bei E-Mail	166
11.6.5	Herausgabe personenbezogener Daten	167
11.6.6	Auskunft über Datei-Downloads	167
11.6.7	Ausweitung der Protokollierung	168
11.6.8	Rufnummernunterdrückung	168
11.7	Datenschutzrechtliche Betrachtung von Tracking im Web	169
11.7.1	Cookies	169
11.7.2	Google Analytics	169
11.7.3	Device Fingerprinting	170
11.7.4	Social Plugins	170
11.8	Fazit	171
11.9	Übungsaufgaben	171
	Literatur	172
12	Zusammenfassung und Ausblick	173
	Sachverzeichnis	175

Abbildungsverzeichnis

Abb. 1.1	Panoptikum: Presidio Modelo, Isla De la Juventud, Kuba.....	2
Abb. 1.2	Drei Säulen des Datenschutzes	4
Abb. 1.3	Gliederung des Lehrbuchs	6
Abb. 2.1	Senderanonymität	13
Abb. 2.2	Symmetrische Verschlüsselungsverfahren	15
Abb. 2.3	Asymmetrische Verschlüsselungsverfahren	16
Abb. 2.4	Digitale Signatur eines Dokuments	20
Abb. 2.5	VPN: Tunnel-Modus bei IPsec	25
Abb. 4.1	Schutz vor Verkehrsflussanalysen mittels VPN	48
Abb. 4.2	Verstecken der Kommunikationsbeziehungen mittels Mix	49
Abb. 4.3	Mix-Kaskade	52
Abb. 4.4	Reply Onion	54
Abb. 4.5	Aufbau eines Tor-Circuits	57
Abb. 4.6	Tor Hidden Services.....	60
Abb. 5.1	Identitätsmanagement im Überblick	68
Abb. 5.2	Autorisierung mit OAuth	73
Abb. 6.1	Das Grundkonzept der Blockchain	83
Abb. 8.1	AES im Counter Mode	99
Abb. 8.2	DH-Austausch beim OTR Messaging	100
Abb. 8.3	Authentifizierung des DH-Austauschs beim OTR Messaging	100
Abb. 8.4	Kompletter Protokollablauf des OTR Messagings	101
Abb. 8.5	Identity Misbinding bei OTR Messaging	102
Abb. 8.6	Verwendung des SIGMA-Protokolls für das OTR Messaging	103
Abb. 9.1	Signatur-PKI	111
Abb. 9.2	EAC: Chip Authentication	113
Abb. 9.3	EAC: Terminal Authentication	114
Abb. 9.4	Verifikations-PKI	115
Abb. 9.5	Elektronischer Personalausweis (ePA)	116

Abb. 9.6	PACE-Protokoll	117
Abb. 9.7	EAC Terminal Authentication Version 2	119
Abb. 9.8	EAC Chip Authentication Version 2	119
Abb. 9.9	Restricted Identification	120
Abb. 9.10	Vergabe der öffentlichen DH-Werte für die Bereiche	121
Abb. 9.11	Online-Einsatz des elektronischen Personalausweises	124
Abb. 11.1	Anwendungsbereich des TMG	152

Zusammenfassung

Warum ist Datenschutz eigentlich wichtig? Und was hat es damit auf sich? Und ist der Datenschutz relevant, wenn wir doch nichts zu verbergen haben? Dies sind Fragen, denen wir in diesem Kapitel auf den Grund gehen werden. Nachdem wir geklärt haben, dass Datenschutz relevant ist, auch wenn wir nichts zu verbergen haben, behandeln wir die Themen und die Lernziele dieses Lehrbuchs.

Jeder von uns hat bestimmte Assoziationen zum Thema *Datenschutz*. Dies rührt daher, dass jeder von uns tagtäglich mit modernen Kommunikationstechnologien zu tun hat. Wir kaufen im Internet ein, führen unsere Bankgeschäfte online, nutzen wie selbstverständlich immer und überall unsere Smartphones, um mit anderen in Kontakt zu bleiben. Moderne, elektronische Assistenten helfen uns in jeder Lebenslage. Sie wissen ganz genau, wann wir für gewöhnlich aufstehen, wie lange wir zur Arbeit fahren, welche Termine anstehen. Erkennen sie Probleme, etwa einen voraussichtlichen Stau aufgrund eines Unfalls auf der Strecke zur Arbeit, wecken sie uns ein wenig früher, damit wir es rechtzeitig zur Arbeit schaffen. Oder sie lassen uns länger schlafen, weil wir die letzte Nacht nicht gut geschlafen haben. Auch diese Information haben sie. Wir liefern mit „Wearables“ nicht nur unsere Schlafgewohnheiten, sondern alle möglichen Daten über unseren (Gesundheits-)Zustand [1]. Doch wem liefern wir diese Daten eigentlich? Hier setzt bei vielen Menschen ein Gefühl des Unwohlseins ein. Wenn wir darüber nachdenken, wie stark wir vernetzt sind und wo wir überall Daten „produzieren“, fühlen wir uns „gläsern“. Kennt uns der Staat, kennen uns Unternehmen vielleicht besser als wir uns selbst? Kann uns hier nicht vielleicht der Datenschutz helfen, eine „Erfindung“ aus längst vergangener Zeit, in der von PCs, Smartphones, Wearables etc. überhaupt noch keine Rede war – oder ist dieser Datenschutz überhaupt nicht mehr zeitgemäß?

Dieses Buch ist so vielfältig wie die skizzierten Themen. Wir werden uns mit unterschiedlichen Technologien auseinandersetzen und untersuchen, wie wir zeitgemäßen Datenschutz umsetzen können. Doch was ist Datenschutz nun eigentlich genau? Geht es um den Schutz von Daten? Auf den ersten Blick ja; aber eigentlich geht es um viel mehr: um den Schutz von Menschen! – vor Missbrauch von Daten.

1.1 Haben wir etwas zu verbergen?

Die Frage, die sich viele von uns stellen ist: „Sind meine Daten schützenswert?“ Daran schließt sich die Frage an: „Haben andere Interesse an meinen Daten?“

Leben wir in einer total überwachten Welt? Finden Totalüberwachung der Telekommunikation und ein ständiges Verfolgen des Aufenthaltsortes, wie in der Welt aus „1984“ von GEORGE ORWELL, wirklich statt? Sind wir im täglichen Leben genauso überwacht wie die Gefängnis-Insassen im Panoptikum (in Abb. 1.1 dargestellt)? Wir werden auf diese Fragen im Laufe des Buches immer wieder zurückkommen und Antworten liefern.

Das gezeigte Panoptikum ist ein Gefängnis, in dem die Zellen so angeordnet sind, dass ein Wächter in alle Zellen hineinschauen kann. Der Wächter selbst ist (im Dunkeln) nur schwer zu erkennen, so dass kein Gefangener weiß, ob er gerade beobachtet wird. Die Idee stammt aus dem späten 18./ frühen 19. Jahrhundert. Auch Fabriken sollten so gebaut werden.

Neben der Frage, *ob* andere Interesse an unseren Daten haben, müssen wir uns natürlich auch fragen, *wer* Interesse daran hat. In den Anfangstagen des „modernen“



Abb. 1.1 Panoptikum: Presidio Modelo, Isla De la Juventud, Kuba. (Quelle: <http://de.wikipedia.org/w/index.php?title=Datei:Presidio-modelo2.JPG&filetimestamp=20070714204645>, Friman/Wiki media Commons, Zugriffen am 01.08.2016)

Datenschutzes stand vor allem der Staat als Interessent an erster Stelle. Die Bürger hatten Angst vor der totalen Überwachung durch den Staat und gingen damals, um etwa gegen die Volkszählung – mit der wir uns in Abschn. 11.2 noch näher beschäftigen werden – zu protestieren, auf die Straße. Danach hat die Wirtschaft den Staat als „Gefährder“ für die informationelle Selbstbestimmung der Bürger, also der potentiellen Kunden, abgelöst. Im Vordergrund stand und steht die zielgerichtete Werbung und sonstige Instrumente zur Verhaltenssteuerung. Die Bürger haben das Gefühl, zum „gläsernen Kunden“ zu werden. Unternehmen kennen ihre Kunden oftmals besser als sie sich selbst. Für Aufsehen hat bspw. ein Fall in den USA gesorgt, als eine junge Frau von einer Supermarktkette plötzlich Werbung für Babyprodukte erhielt. Zu diesem Zeitpunkt wusste ihr Vater noch nicht, dass sie schwanger war. Die Supermarktkette hingegen „wusste“ es bereits¹: dank Auswertung ihres Einkaufsverhaltens und entsprechender „Big Data“-Analyse. Durch die Enthüllungen von EDWARD SNOWDEN rund um die Geheimdienstaktivitäten der NSA und des britischen Geheimdienstes (GCHQ) rückt in letzter Zeit auch der Staat als Überwacher wieder vermehrt in das Bewusstsein der Bürger. Daneben hat sich eine Reihe von Kriminellen – nicht nur im Internet – breit gemacht, die Interesse an unseren Daten haben. Erpressung, Identitätsdiebstahl und Vorbereitung anderer Straftaten sind nur einige Delikte in diesem Zusammenhang. Am Rande sei erwähnt, dass auch „Freunde“, Verwandte und Ehepartner Interesse an unseren Daten haben können.

Dass die „Big Player“ im Daten-Geschäft, etwa *Google* oder *Facebook*, Interesse an noch mehr persönlichen Daten aller Nutzer haben, ist kein Geheimnis. Daten werden als das „neue Öl“ der Wirtschaft gesehen, die neue Geschäftsmodelle entstehen lassen. ERIC SCHMIDT, CEO von Google, hat auf der IFA 2010 einen kleinen Vorgeschmack auf die Zukunft gegeben:

„Ultimately, search is not just the web but literally all of your information—your email, the things you care about, with your permission—this is personal search, for you and only for you. [...]

We can suggest what you should do next, what you care about. Imagine: *We know where you are, we know what you like.* [...]

A near-term future in which you don't forget anything, because the computer remembers. You're never lost.“

Inzwischen sind wir, wenn wir uns überlegen, wie viele Dienste wir von Google im täglichen Leben nutzen, wahrscheinlich schon in der vorausgesagten Zukunft angekommen.

¹ <http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#45c7595534c6> (Zugegriffen am 01.08.2016).

1.2 Säulen des Datenschutzes

Der Schutz der Privatsphäre basiert auf drei Säulen, die in Abb. 1.2 dargestellt sind.

Zum ersten kann Privatsphäre durch *Regulierungen*, wie Gesetze, Richtlinien oder Verordnungen, geschützt werden. In Deutschland wird die Privatsphäre z. B. durch Datenschutzgesetze geschützt, die genau festlegen, welche personenbezogenen Daten zu welchem Zweck erhoben und verarbeitet werden dürfen. Da das Internet weltumspannend ist, ist es schwierig, durch Regulierung einen Schutz der Privatsphäre zu erreichen, da Regulierungen meist nationale Gültigkeit haben. Zudem ist Regulierung ein relativ langwieriger Prozess, der eventuell nicht mit dem Fortschritt der Technologie einhergeht.

Weiterhin kann auch *Selbstregulierung* die Privatsphäre effektiv schützen. Bei Selbstregulierung verpflichtet sich ein Diensteanbieter (z. B. Betreiber eines Online-Shops) dazu, gewisse Maßnahmen zum Schutz der Privatsphäre einzuhalten. Meist geht die Verpflichtung mit der Verleihung eines Zertifikats einher, das publikumswirksam auf den erhöhten Schutz hinweisen soll. In Deutschland trifft man Selbstregulierung zum Schutz der Privatsphäre meist im Rahmen von Selbstverpflichtungen von Online-Shop-Betreibern an.

Schließlich und endlich kann ein Benutzer auch zum *Selbstschutz* greifen und mithilfe von technischen Maßnahmen seine Privatsphäre schützen. Ebenso kann der Betreiber den Datenschutz durch technische Maßnahmen in sein System integrieren. Man spricht in diesem Fall von *Privacy by Design (PbD)*. Sowohl beim Selbstschutz als auch bei *PbD* kommen *Privacy-Enhancing Technologies (PETs)* zum Einsatz.

Im Laufe dieses Buchs werden wir uns hauptsächlich mit der ersten und der dritten Säule befassen.

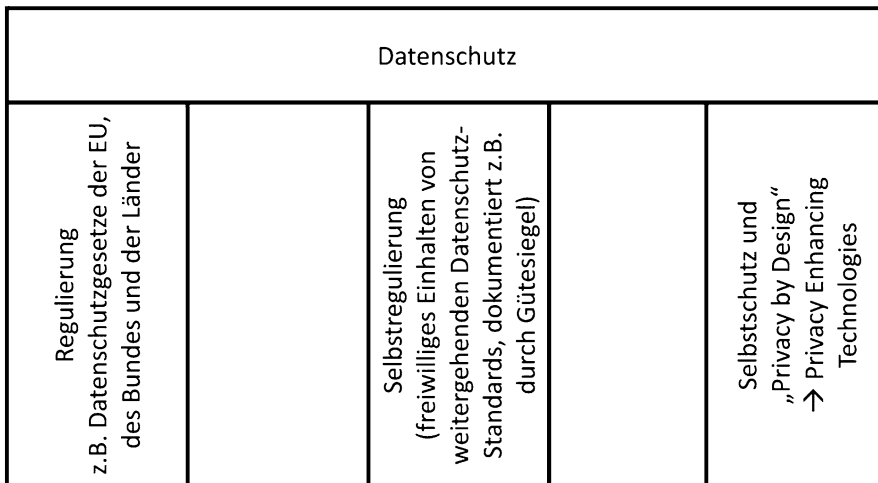


Abb. 1.2 Drei Säulen des Datenschutzes

1.3 Themen dieses Buchs und Lernziele

Der Fokus dieses Lehrbuchs liegt auf dem *Datenschutz durch Technik* – mit einem Schwerpunkt auf Telekommunikation und vernetzte Systeme. Sie werden Techniken kennenlernen, die speziell für den Schutz personenbezogener Daten entwickelt wurden. Die vorgestellten Techniken gehen über Verfahren hinaus, die lediglich Vertraulichkeit der Kommunikation erreichen und eher im Bereich der „IT-Sicherheit“ angesiedelt werden können, etwa SSL/TLS, IPsec, PGP etc. Daneben werden wir die Grundzüge des Datenschutzrechts behandeln. Das Datenschutzrecht bildet einerseits die rechtliche Grundlage für die vorgestellten technischen Verfahren; andererseits werden wir sehen, dass auch neue (Datenschutz-)Technologien wiederum Anpassungen im Datenschutzrecht nötig machen können.

Wir möchten an dieser Stelle auch darauf hinweisen, welche Themen wir in diesem Buch *nicht* behandeln werden. Zum einen sind dies Fragen zum Prozessrecht und Administratives, d. h. Fragen, die sich um die Durchsetzung von Datenschutzrecht drehen. Zum anderen sind dies Fragen rund um das Thema Arbeitnehmerdatenschutz. Außerdem werden wir in diesem Lehrbuch nicht auf Datenschutzbeauftragte und deren vielfältige Aufgaben eingehen können. Nichtsdestotrotz kann dieses Lehrbuch auch für Datenschutzbeauftragte, gerade im Hinblick auf *Privacy by Design (PbD)*, interessant sein, um sich einen Überblick über neuartige, datenschutzfördernde Technologien zu verschaffen.

Was die *Lernziele* betrifft, so sollen Sie nach der Lektüre dieses Lehrbuchs

- verschiedene Mechanismen im Bereich des technischen Datenschutzes im Detail verstanden haben,
- für konkrete Beispiele entscheiden können, mit welchen Mechanismen Ziele des Datenschutzes erreicht werden können,
- ausgewählte aktuelle Forschungsarbeiten zum technischen Datenschutz kennen und verstanden haben,
- die Struktur des deutschen Datenschutzrechts kennen,
- Gesetze im Bereich des Datenschutzes verstehen und einfache Sachverhalte darunter subsumieren können,
- technische und juristische Sachverhalte zueinander in Beziehung setzen können,
- erkennen können, wann für datenschutzrechtliche Fragestellungen rechtlicher Beistand nötig ist.

Zur Erreichung der Lernziele arbeiten wir sehr praxisorientiert. Wir werden uns im Bereich der Technik mit Technologien beschäftigen, mit denen wir täglich zu tun haben. Wir werden diese Technologien analysieren, Probleme aus Sicht des Datenschutzes aufzeigen und mögliche Lösungen im Detail besprechen und diskutieren. Im Bereich des Datenschutzrechts werden wir datenschutzrechtliche Fälle (aus der Praxis) bearbeiten.

Die Gliederung des Lehrbuchs ist mit einer Auswahl der behandelten Themen in Abb. 1.3 übersichtlich dargestellt.

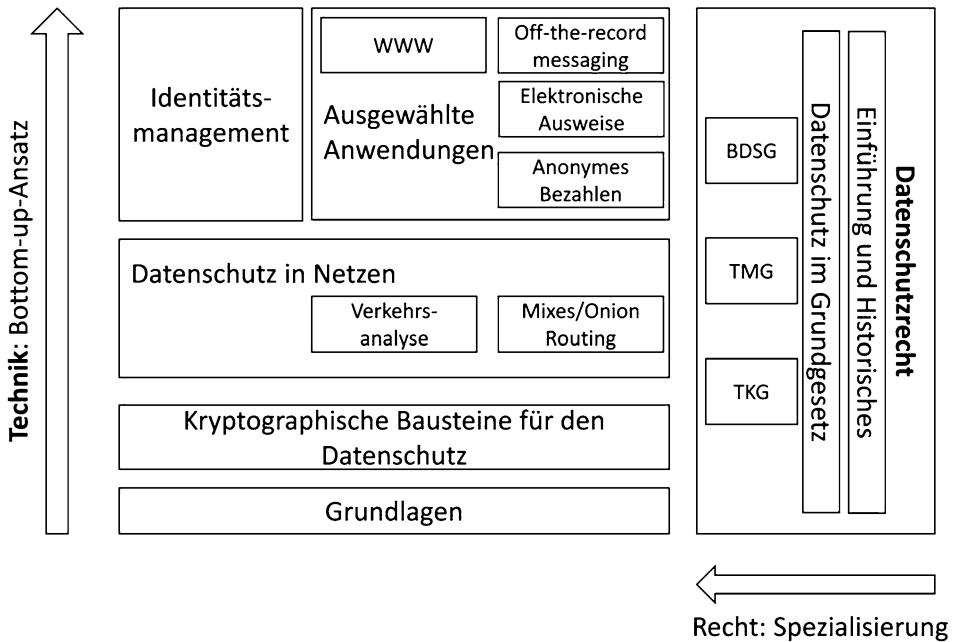


Abb. 1.3 Gliederung des Lehrbuchs

Zunächst werden wir in Kap. 2 in die Welt des technischen Datenschutzes einstiegen und uns dabei mit den Datenschutz-Schutzziele, Begriffsbestimmungen und den kryptographischen Bausteinen vertraut machen, mit denen wir im weiteren Verlauf des Lehrbuchs arbeiten werden. In Kap. 3 widmen wir uns den Anonymitätsmaßen, die eine Klassifizierung der Güte von Anonymisierungsmaßnahmen erlauben. Im „bottom-up“-Ansatz arbeiten wir uns dann im technischen Bereich weiter vor. In Kap. 4 beschäftigen wir uns zunächst allgemein mit dem Datenschutz auf Netzebene. Wir behandeln in diesem Kapitel Anonymisierungsdienste wie Onion Routing/Tor. Als nächstes steht in Kap. 5 das Thema Identitätsmanagement (IdM) auf der Tagesordnung. Wir untersuchen die im Internet gängigen Verfahren OpenID, OAuth und OpenID Connect auf ihre Datenschutzzeigenschaften hin. Danach widmen wir uns ausgewählten Anwendungen. So beschäftigen wir uns in Kap. 6 mit dem anonymen Bezahlen. Unter anderem widmen wir uns dem Thema Bitcoin. Dem Datenschutz im World Wide Web wenden wir uns in Kap. 7 zu. Das Thema Instant Messaging, das in den vergangenen Jahren enorm an Bedeutung gewonnen hat, steht in Kap. 8 im Vordergrund. In Kap. 9 lernen Sie das Sicherheits- und Datenschutzkonzept von elektronischem Reisepass und elektronischem Personalausweis kennen. Anschließend beschäftigen wir uns in Kap. 10 mit weiterführenden kryptographischen Verfahren rund um das Thema Privacy-Enhancing Technologies (PETs). Schließlich werden wir in Kap. 11 eine Einführung in das (deutsche) Datenschutzrecht geben – uns dabei insbesondere mit dem Bundesdatenschutzgesetz (BDSG) und

den Datenschutzvorgaben im Telemediengesetz (TMG) und Telekommunikationsgesetz (TKG) befassen – und im Anschluss datenschutzrechtliche Fälle bearbeiten.

Der Datenschutz ist nicht nur für IT-Verantwortliche relevant, sondern auch für Systemadministratoren, Manager, Webmaster, Web-Surfer etc. Stellen Sie sich vor, Sie betreiben einen öffentlichen Webserver für ein Unternehmen. Die Nichtbeachtung des Datenschutzrechts kann zu Abmahnungen, Unterlassungsklagen und/oder Bußgeldbescheiden führen. Beispiele für Fragen, die Sie im Anschluss an die Lektüre dieses Lehrbuchs beantworten können sollen, sind:

- Darf ich Logfiles führen?
- Darf ich Drittanbieter mit dem Führen von Statistiken beauftragen?
- Wie können sich Nutzer mit neuem Personalausweis bei einem Diensteanbieter authentifizieren?
- Wie können Nutzer ihre Identität vor einem böartigen Betreiber schützen?
 - Was kann der böartige Betreiber tun, um Nutzer dennoch zu identifizieren?

Wir wünschen Ihnen viel Spaß beim Eintauchen in die spannende Welt des Datenschutzes!

1.4 Danksagung

Die konstruktiven Diskussionen mit Studierenden und Kollegen prägen die Entwicklung eines Lehrbuchs. Unser Dank gilt diesen Menschen! Auch für die Unterstützung unserer Lektorinnen, DR. SABINE KATHE und SYBILLE THELEN vom Springer-Verlag, möchten wir uns bedanken. Bei ANNA PETRLIC bedanken wir uns für das Korrekturlesen.

Es ist uns ein Anliegen, unser Lehrbuch stets weiter zu entwickeln. Deshalb möchten wir Sie an dieser Stelle ermuntern, uns Verbesserungsvorschläge per E-Mail an auto-ren@datenschutzlehrbuch.de mitzuteilen. Verbesserungen werden wir unter <http://www.datenschutzlehrbuch.de> für Sie bereitstellen.

Literatur

1. Ronald Petrlc. Das vermessene Selbst. *Datenschutz und Datensicherheit – DuD*, 40(2):94–97, 2016.

Zusammenfassung

Für das Verständnis des *Technischen Datenschutzes* sind zunächst einige Grundlagen wichtig. Deshalb wenden wir uns in diesem Kapitel zuerst in Abschn. 2.1 den grundlegenden Schutzziele im Bereich der IT-Sicherheit und des Datenschutzes zu. Im Anschluss daran werden wir uns in Abschn. 2.2 mit den Begriffen des technischen Datenschutzes vertraut machen. Danach tauchen wir in Abschn. 2.3 in die Welt der Kryptographie ein, bevor wir uns schließlich in Abschn. 2.4 mit Maßnahmen aus dem Bereich der IT-Sicherheit auseinandersetzen, die, wie wir später sehen werden, auch dem Datenschutz dienlich sind.

Lernziele

Am Ende dieses Kapitels sollten Sie die Grundbegriffe des technischen Datenschutzes kennen und verstehen. Sie sollten die Schutzziele kennen, die aus Sicht des Datenschutzes die Ziele (und gleichzeitig die Bewertungskriterien) für die technischen Verfahren vorgeben, die wir im weiteren Verlauf des Buchs behandeln werden. Außerdem sollten Sie mit den grundlegenden kryptographischen Verfahren und den Maßnahmen aus dem Bereich der IT-Sicherheit vertraut sein – diese dienen als Grundlage für die später behandelten *Privacy-Enhancing Technologies (PETs)*.

2.1 Schutzziele

Betrachten wir zum Einstieg das folgende Beispiel. Sie betreiben einen öffentlichen Webserver für ein Unternehmen. Die Kommunikation soll „sicher“ und „datenschutzgerecht“ sein. Welche Schutzziele möchten Sie erreichen?

Zunächst können wir zwischen den „klassischen“ Schutzzielen, mit denen im Bereich der IT-Sicherheit schon lange gearbeitet wird und den „neuen“ Datenschutz-Schutzzielen unterscheiden, die erst in den letzten Jahren vermehrt betrachtet werden.

2.1.1 „Klassische“ IT-Sicherheits-Schutzziele

Wir stützen uns hier auf die (leicht modifizierten) Definitionen nach ECKERT [2].

► **Authentizität** *Authentizität* eines Subjekts oder Objekts ist die Echtheit und Glaubwürdigkeit des Subjekts oder Objekts, die anhand einer eindeutigen Identität und charakteristischen Eigenschaften überprüfbar ist.

Die allgemeine Definition ist sehr schwammig, der Begriff wird in verschiedenen Kontexten genutzt. *Datenauthentizität* (oder genauer Datenursprungsauthentizität) ist gegeben, wenn Daten vom behaupteten bzw. erwarteten Absender stammen. *Authentizität einer Person* (z. B. als Kommunikationspartner) ist gegeben, wenn die tatsächliche Identität mit der behaupteten oder erwarteten Identität übereinstimmt.

► **Integrität** Ein System gewährleistet *Datenintegrität*, wenn es Subjekten nicht möglich ist, die zu schützenden Daten unautorisiert und unbemerkt zu verändern. Integrität ist dabei eng verknüpft mit der Authentizität. Außerdem ist ein Verfahren zur Festlegung der Rechte an Daten notwendig, d. h. es muss klar sein, wer eigentlich autorisiert ist.

► **Vertraulichkeit** Ein System gewährleistet die *Informationsvertraulichkeit*, wenn es keine unautorisierte Informationsgewinnung ermöglicht. Hierbei ist wiederum wichtig zu klären, wer eigentlich autorisiert ist.

► **Verfügbarkeit** Ein System gewährleistet die *Verfügbarkeit*, wenn authentifizierte und autorisierte Subjekte in der Wahrnehmung ihrer Berechtigungen nicht unautorisiert beeinträchtigt werden können.

► **Verbindlichkeit** Ein System gewährleistet die *Verbindlichkeit* (auch Zurechenbarkeit, Zuordenbarkeit, Urhebernachweis, oder Nichtabstreitbarkeit genannt), wenn die für ein bestimmtes Ereignis verantwortliche Instanz unabstreitbar identifizierbar ist.

2.1.2 „Neue“ Datenschutz-Schutzziele

Die *Konferenz der Datenschutzbeauftragten des Bundes und der Länder* hat 2010 ein Eckpunktpapier für „Ein modernes Datenschutzrecht für das 21. Jahrhundert“ [4] verabschiedet. Darin wurde die Aufnahme der folgenden Datenschutz-Schutzziele in die Liste der zuvor beschriebenen Schutzziele beschlossen. Die Definitionen in diesem

Unterabschnitt richten sich nach dem Landesdatenschutzgesetz Schleswig-Holstein, in das die Datenschutz-Schutzziele bereits Einzug gehalten haben.

► **Transparenz** *Transparenz* ist gegeben, wenn die Verarbeitung von personenbezogenen Daten mit zumutbarem Aufwand nachvollzogen, überprüft und bewertet werden kann.

Transparenz stellt damit die Voraussetzung für die Steuerung und Regulation technisch-organisatorischer Prozesse sowie zur Abwägung bezüglich des Zwecks der Datenverarbeitung sowie der Erforderlichkeit dar [6].

► **Nicht-Verkettbarkeit** *Nicht-Verkettbarkeit* ist gegeben, wenn personenbezogene Daten nicht oder nur mit unverhältnismäßig hohem Aufwand für einen anderen als den ausgewiesenen Zweck erhoben, verarbeitet und genutzt werden können.

Nicht-Verkettbarkeit sorgt also für die Zweckbindung – eine der wesentlichen Grundpfeiler im Datenschutzrecht.

► **Intervenierbarkeit** *Intervenierbarkeit* ist gegeben, wenn die Verfahren so gestaltet werden, dass sie den Betroffenen die Ausübung der ihnen zustehenden Rechte wirksam ermöglichen.

2.2 Begriffsbestimmungen

2.2.1 Begriff des Datenschutzes

Bevor wir uns im Detail mit den Begriffen zum technischen Datenschutz auseinandersetzen, möchten wir uns zuerst noch allgemein mit dem Begriff „Datenschutz“ beschäftigen. Im Deutschen bezeichnet der Datenschutz nicht etwa den Schutz von Daten, sondern vielmehr den

„Schutz des Einzelnen vor Beeinträchtigung in seinem Persönlichkeitsrecht durch Umgang mit seinen personenbezogenen Daten“.

Diese Definition folgt § 1 Abs. 1 Bundesdatenschutzgesetz (BDSG).

Die genaue englische Übersetzung von „Datenschutz“ lautet „Data Protection“. Dieser Begriff ist allerdings eher unter Juristen und weniger unter Technikern gängig. Außerdem wird der Begriff eher in Europa als in den USA verwendet.

Der englische Begriff „Privacy“ wird oft als Entsprechung zu „Datenschutz“ gesehen. Ihm kommt aber eine breitere Verwendung, nämlich allgemein der Privatsphäre, zu. In Informatik-Veröffentlichungen wird eher der Begriff „Privacy“ verwendet. Technische Hilfsmittel, um Datenschutz zu erreichen, werden (oft auch in der deutschen Sprache) als „Privacy-Enhancing Technologies (PETs)“ bezeichnet.

2.2.2 Begriffe zum technischen Datenschutz

Das Modell und die Definitionen, die wir in diesem Abschnitt beleuchten werden, richten sich nach dem Arbeitspapier zur Terminologie beim Datenschutz nach PFITZMANN und HANSEN [5], den Vorreitern des technischen Datenschutzes in Deutschland. Wir möchten an dieser Stelle auch darauf hinweisen, dass der Gebrauch der Begriffe keineswegs einheitlich ist, so dass je nach Kontext und sogar je nach individueller Quelle kleine Bedeutungsunterschiede oder sogar grundlegend andere Modelle verwendet werden. Deshalb Vorsicht bei der Verwendung der Begriffe!

Identität

Eine *Identität* eines Subjekts ist eine Menge von Attributwerten, die das Subjekt in einer Menge von Subjekten identifizierbar macht, d. h. von den anderen Subjekten unterscheidet. Die Menge von Subjekten kann unterschiedlich sein. Beispielsweise könnten alle Personen in einem Hörsaal oder alle Professoren an der Universität gemeint sein. Was aus der Definition einer Identität auch deutlich wird: Jede Person kann viele Identitäten haben. Dies ist keine technische Abstraktion. Für manche Leute, mit denen Sie umgehen, sind Sie vielleicht „der Nachbar aus dem 2. Stock“, „der Student, der immer in der letzten Reihe sitzt“ o.ä., ohne dass dabei ihr Name eine Rolle spielen muss.

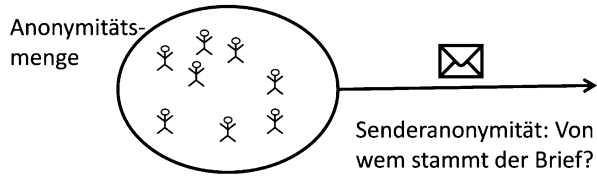
Ein Subjekt ist *identifizierbar*, wenn es innerhalb einer Menge von Subjekten (der Identifizierbarkeitsmenge) durch einen Angreifer hinreichend von anderen Subjekten unterscheidbar ist. Der Angreifer ist derjenige, vor dem man sich schützen möchte – beispielsweise der Betreiber einer Website, der Ihnen personenbezogene Werbung zukommen lassen möchte.

Man unterscheidet ferner noch zwischen der *vollständigen Identität* und der *partiellen Identität*. Die vollständige Identität vereint alle Attributwerte der Identitäten eines Subjekts. Auf die vollständige Identität einer Person werden i.d.R. nur sehr wenige Personen Zugriff haben. Normalerweise tritt man unter vielen partiellen Identitäten auf. Eine solche partielle Identität ist dabei eine echte Teilmenge der vollständigen Identität, die durch ein *Pseudonym* identifiziert werden kann.

In Kap. 5 werden wir uns mit dem Thema *Identitätsmanagement (IdM)* beschäftigen. Dabei geht es gerade um die Verwaltung mehrerer Identitäten von Subjekten, d. h. die Verwaltung von Attributwerten und die Auswahl einer zu verwendenden Identität in einem spezifischen Kontext, einschließlich der Authentifizierung unter einer ausgewählten Identität. Wir werden sehen, dass IdM dem Datenschutz dient, wenn sichergestellt wird, dass mehrere Identitäten eines Subjekts nicht miteinander verknüpft werden können.

Anonymität

Ein Subjekt ist *anonym* gegenüber einem Angreifer, wenn der Angreifer das Subjekt in einer Menge von Subjekten (der Anonymitätsmenge) nicht hinreichend identifizieren kann. Identifizieren heißt hier, dass der Angreifer das Subjekt nicht von anderen Subjekten unterscheiden kann. Die Formulierung „nicht hinreichend“ bedeutet wiederum, dass der

Abb. 2.1 Senderanonymität

Begriff der Anonymität auch vom Szenario abhängt und es verschiedene Abstufungen der Anonymität geben kann.

Betrachtet man Kommunikationsnetze – wie wir es in Kap. 4 tun werden – so kann man zwischen *Senderanonymität* und *Empfängeranonymität* unterscheiden. Die Senderanonymität, dargestellt in Abb. 2.1, bezeichnet dabei die Anonymität des Absenders einer Nachricht in einer Menge möglicher Absender. Die Empfängeranonymität hingegen bezeichnet die Anonymität des Empfängers einer Nachricht in einer Menge möglicher Empfänger.

In diesem Zusammenhang spricht man auch von der sogenannten *Anonymitätsdifferenz* (oder dem Anonymitätsdelta). Diese bezeichnet die Differenz aus der Anonymität eines Subjekts nach einem Angriff (a-posteriori-Wissen des Angreifers) und der Anonymität des Subjekts vor dem Angriff (a-priori-Wissen des Angreifers). Dies setzt eine Messung/Quantifizierung von Anonymität voraus, für die es allerdings keine universell gültige Metrik gibt; eine derartige Messung ist vielmehr nur in gewissen (oft idealisierten) Szenarien möglich.

Anonymisierung bezeichnet schließlich den Prozess, um Anonymität zu erreichen. Es geht darum, personenbezogene Daten derart zu verändern, dass Einzelangaben nicht mehr oder nur sehr schwierig einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können.

Pseudonymität

Den Begriff des „Pseudonyms“ haben wir im Zusammenhang mit partiellen Identitäten bereits ins Spiel gebracht. Ein *Pseudonym* ist nichts anderes als ein Identifikator für ein Subjekt ungleich dem realen Namen¹ des Subjekts. Beispiele für Pseudonyme sind der Nutzernamen in einer Online-Community, das Pseudonym eines Romanautors, die Kundennummer bei einem Versandhändler etc.

Pseudonymität ist erreicht, wenn das Subjekt ein Pseudonym verwendet und ein Angreifer nicht hinreichend auf den realen Namen des Subjekts schließen kann.

Pseudonymisierung schließlich bezeichnet wieder den Prozess, um Pseudonymität zu erreichen. Es geht um die Veränderung von personenbezogenen Daten mittels einer Zuordnungsvorschrift derart, dass Einzelangaben ohne Kenntnis oder Nutzung der Zuordnungsvorschrift nicht mehr einer natürlichen Person zugeordnet werden können.

¹Der reale Name meint den Namen, unter dem eine Person allgemein bekannt ist.

Im einfachsten Fall besteht die Pseudonymisierung darin, dass man in einer Datenbank jedes Auftreten eines Namens durch eine (für diesen Namen jeweils gleichbleibende) Nummer ersetzt. Die Zuordnungsvorschrift könnte dann einfach jeder Nummer wieder den dazu passenden Namen zuordnen.

Unverkettbarkeit

Eine Ausprägung des Konzepts der *Unverkettbarkeit* haben wir in Form der „Nicht-verkettbarkeit“ als eines der „neuen“ Datenschutz-Schutzziele in Abschn. 2.1.2 bereits kennengelernt.

Allgemein versteht man unter Unverkettbarkeit von zwei (oder mehreren) betrachteten Objekten² (wie Nachrichten, Personen etc.) in einem System, dass ein Angreifer nicht hinreichend sicher feststellen kann, ob zwischen diesen Objekten eine Beziehung besteht oder nicht. Verkettbarkeit ist analog definiert. Als Beispiel können zwei Nachrichten dienen, bei denen ein Angreifer wissen möchte, ob diese von der selben Person verschickt wurden. Unverkettbarkeit wird auch als „Unverknüpfbarkeit“ (im Englischen „Unlinkability“) bezeichnet.

Unentdeckbarkeit

Unentdeckbarkeit (im Englischen „Undetectability“) stellt die stärkste Eigenschaft dar. Unentdeckbarkeit eines betrachteten Objekts (bspw. einer Nachricht) ist gegeben, wenn ein Angreifer nicht hinreichend entscheiden kann, ob das Objekt existiert oder nicht. Der Angreifer kann also bspw. nicht sagen, ob eine Nachricht verschickt wurde. Entdeckbarkeit ist analog definiert.

2.3 Grundlegende kryptographische Verfahren

In diesem Abschnitt beschäftigen wir uns mit den für uns im weiteren Verlauf relevanten kryptographischen Verfahren zur *Verschlüsselung*, dem *Schlüsselaustausch* und *Digitalen Signaturen*.

2.3.1 Verschlüsselung

Verschlüsselung dient der Wahrung des Schutzziels *Vertraulichkeit*. Nachrichten, die mit einem „sicheren“ Verschlüsselungsverfahren verschlüsselt werden, sollen also nicht durch unautorisierte Dritte gelesen werden können.

Zunächst unterscheidet man zwischen *symmetrischen* und *asymmetrischen* Verfahren. Bei den symmetrischen Verfahren erfolgt die Ver- und Entschlüsselung mit dem *gleichen*

²In der Originaldefinition wird das „betrachtete Objekt“ als „item of interest“ bezeichnet.

Schlüssel. Der Schlüssel muss zwischen den Kommunikationspartnern ausgetauscht werden. Symmetrische Verfahren sind meist wenig rechenaufwendig. Bei asymmetrischen Verfahren erfolgt die Ver- und Entschlüsselung hingegen mit *unterschiedlichen Schlüsseln*. Asymmetrische Verfahren sind meist sehr rechenaufwendig.

In der Praxis werden oft symmetrische und asymmetrische Verfahren kombiniert. Zum Beispiel werden beim Versand verschlüsselter E-Mails in der Regel die eigentlichen Daten mit einem symmetrischen Verfahren verschlüsselt; dafür wird ein (im Vergleich zu den Daten in der Regel kurzer) Sitzungsschlüssel verwendet. Dieser Sitzungsschlüssel wird wiederum mit einem asymmetrischen Verfahren verschlüsselt.

Symmetrische Verschlüsselung

Sehen wir uns symmetrische Verfahren in Abb. 2.2 etwas genauer an.

Zunächst lernen wir noch Alice und Bob kennen. Sie sind die wohl bekanntesten Charaktere in der Kryptographie. Im klassischen Szenario möchte Alice Bob eine Nachricht schicken und Eve (von „to eavesdrop“) möchte diese mithören. Der Ablauf der Kommunikation sieht folgendermaßen aus. Zunächst vereinbaren Alice und Bob einen gemeinsamen, geheimen Schlüssel $K_E = K_D = K_{A,B}$. K_E bezeichnet dabei den „Encryption Key“ und K_D den „Decryption Key“. Um einen Klartext M („Message“) von Alice an Bob zu versenden, verschlüsselt Alice den Text mittels $K_{A,B}$, also $C = E(M, K_{A,B})$ und sendet den „Ciphertext“ C an Bob. Bob entschlüsselt C mittels $K_{A,B}$, also: $M = D(C, K_{A,B}) = D(E(M, K_{A,B}), K_{A,B})$. E bezeichnet hierbei die Verschlüsselungsfunktion und D die Entschlüsselungsfunktion. Eines der heutzutage als sicher geltendes und weit verbreitetes symmetrisches Verschlüsselungsverfahren ist der *Advanced Encryption Standard (AES)*. Der Schlüsselaustausch stellt bei symmetrischen Verfahren ein Problem dar; dieser muss über einen sicheren Kanal erfolgen, ansonsten könnte eine Lauscherin Eve den Schlüssel mithören und damit selbst die verschlüsselte Nachricht entschlüsseln und somit an den Klartext gelangen.

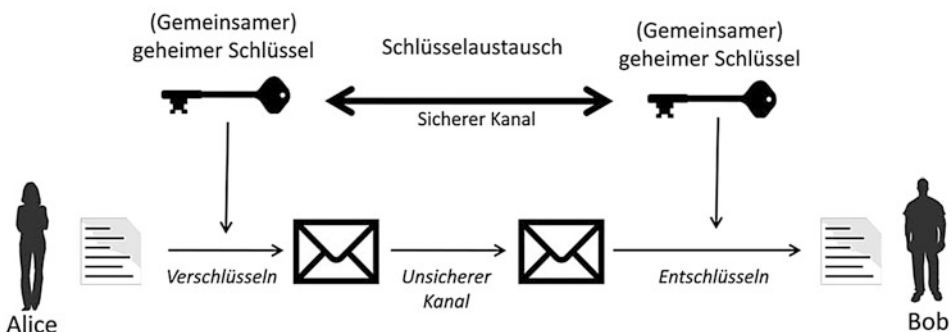


Abb. 2.2 Symmetrische Verschlüsselungsverfahren

Asymmetrische Verschlüsselung

Asymmetrische Verfahren (häufig auch *Public Key*-Verfahren genannt) umgehen das zuvor genannte Problem des sicheren Schlüsselaustauschs. Wir erinnern uns, dass bei asymmetrischen Verfahren die Schlüssel zum Ver- und Entschlüsseln unterschiedlich sind. Der Schlüssel zum Verschlüsseln kann veröffentlicht werden; er wird deshalb auch „öffentlicher Schlüssel“ genannt. Der Schlüssel zum Entschlüsseln hingegen ist der „private Schlüssel“ – dieser darf unter keinen Umständen weitergegeben werden. Einen Überblick über asymmetrische Verfahren gibt Abb. 2.3.

Wie wir in der Abbildung sehen, wird der Schlüsselaustausch bei asymmetrischen Verfahren einfacher. Bob kann seinen öffentlichen Schlüssel veröffentlichen, bspw. auf seiner Webseite; er muss ihn nicht mehr auf einem sicheren Kanal an Alice übertragen. Dass der Schlüsselaustausch bei asymmetrischen Verfahren überhaupt kein Problem darstellt, wäre allerdings nur die halbe Wahrheit. In der Praxis ist es für Alice schwer, die Authentizität des öffentlichen Schlüssels von Bob zu überprüfen. Für diesen Zweck werden *Public Key Infrastrukturen (PKIs)* eingesetzt. PKIs sind nicht Thema dieses Lehrbuchs. Daher verweisen wir an dieser Stelle auf Lehrbücher zur IT-Sicherheit.

Bei asymmetrischen Verfahren haben wir es mit Schlüsselpaaren zu tun. K_E bezeichnet dabei den Schlüssel zum Verschlüsseln, den öffentlichen Schlüssel, der bekannt gegeben werden kann. K_D hingegen bezeichnet den Schlüssel zum Entschlüsseln, den privaten Schlüssel. Für die Ver- und Entschlüsselungsfunktion E und D muss gelten, dass $D(E(M, K_E), K_D) = M$, d. h. die Entschlüsselung einer verschlüsselten Nachricht mit zum öffentlichen Schlüssel passenden privaten Schlüssel muss wieder die ursprüngliche Nachricht ergeben.

Eines der bekanntesten asymmetrischen Verfahren ist *RSA*, benannt nach seinen Entwicklern RIVEST, SHAMIR und ADLEMAN. *RSA* basiert auf dem Faktorisierungsproblem. Es nutzt aus, dass die Multiplikation großer Zahlen sehr viel einfacher ist als die Faktorisierung. Es lässt sich allgemein sagen, dass die Sicherheit bei asymmetrischen Verfahren schwer kalkulierbar ist. Die Angreifbarkeit hängt vom algorithmischen Fortschritt bei der Lösung der den Verfahren zugrundeliegenden mathematischen Problemen ab.

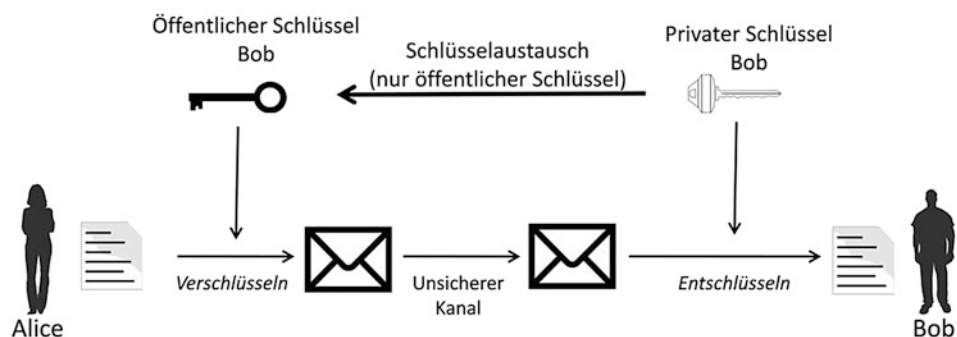


Abb. 2.3 Asymmetrische Verschlüsselungsverfahren

Da wir in den nachfolgenden Kapiteln immer wieder auf das RSA-Verfahren zurückgreifen werden, möchten wir es an dieser Stelle näher betrachten:

RSA-Verfahren:

1. Wähle zwei große Primzahlen p und q , berechne $n = pq$.
2. Wähle $0 < d < n - 1$, so dass gilt: d ist relativ prim zu $(p - 1)(q - 1)$, d. h. $\text{ggT}((p - 1)(q - 1), d) = 1$. Mögliche Kandidaten: Primzahlen d mit $\max(p, q) < d < (p - 1)(q - 1)$
3. Wähle $0 < e < n - 1$ mit $ed = 1 \pmod{(p - 1)(q - 1)}$
4. (e, n) ist der öffentliche Schlüssel
5. (d, n) ist der private Schlüssel
6. Verschlüsseln des Klartexts M : $C = E(M) = M^e \pmod n$
7. Entschlüsseln des Chiffretexts C : $D(C) = C^d \pmod n = M$

Es gilt für e, d, n wie oben definiert: $M^{ed} \pmod n = M^{de} \pmod n = M$

Die Sicherheit des Verfahrens beruht auf der Geheimhaltung der Werte p, q und $(p - 1)(q - 1)$.

Im weiteren Verlauf werden wir vor allem die Punkte 4 bis 7 benötigen. Den Rest des Verfahrens nehmen wir als gegeben hin. Nähere Informationen und Beweise finden sich in nahezu jedem Lehrbuch zur Kryptographie bzw. zur IT-Sicherheit, etwa bei SORGE et al. [7].

2.3.2 Digitale Signatur

Asymmetrische Kryptographie erlaubt neben der Verschlüsselung eine weitere Anwendung: die *Digitale Signatur*.

Die Anforderungen an eine digitale Signatur sind wie folgt:

- Die Signaturerstellung ist nur durch die berechtigte Entität möglich,
- Die Signaturprüfung ist durch beliebige Dritte möglich,
- Die Signatur ist an ein Dokument gebunden,
- Der Signaturersteller kann nicht abstreiten, eine Signatur erstellt zu haben.

In der Praxis erfolgt die Signaturerstellung durch eine Entität mit dem passenden *privaten* Schlüssel. Die Signaturprüfung kann durch jeden erfolgen, der Zugriff auf den zugehörigen *öffentlichen* Schlüssel hat. Eine elektronische Signatur eines Dokuments durch Alice liefert den Beweis, dass Alice *genau dieses* Dokument signiert hat. Niemand sonst

kann die Signatur im Namen von Alice erzeugen (außer, er hat ihren privaten Schlüssel). Die Signatur ist auch für kein anderes Dokument gültig. Damit wird das Schutzziel *Verbindlichkeit* erreicht.

Das zuvor besprochene RSA-Verfahren eignet sich auch als digitales Signaturverfahren. Nehmen wir an, Alice möchte ein Dokument signieren. Dazu wendet sie die Verschlüsselungsfunktion von RSA an. Allerdings verwendet sie nun ihren eigenen privaten Schlüssel. Sie berechnet also $\text{sig}_{\text{Alice}}(M) = M^d \bmod n$. Bob, in Besitz des öffentlichen Schlüssels von Alice, kann nun die Signatur folgendermaßen prüfen: Ist $\text{sig}_{\text{Alice}}(M)^e \bmod n = M$?

2.3.3 Blinde Signatur

Einer der zentralen kryptographischen Bausteine in einer Reihe von Privacy-Enhancing Technologies (PETs) ist die *blinde Signatur*. Das Konzept wurde 1983 von DAVID CHAUM entwickelt, einem Kryptographen der, wie wir noch sehen werden, für zahlreiche Entwicklungen im Bereich der PETs verantwortlich zeichnet [1]. Die Grundidee ist, dass der Signierende nicht sieht, was er signiert. Das Äquivalent in der analogen Welt sieht folgendermaßen aus: Zunächst wird Kohlepapier auf ein Dokument gelegt. Das Dokument und das Kohlepapier werden in einen Umschlag gesteckt. Der Signierende signiert den Umschlag. Die Unterschrift drückt sich dabei auf das Dokument durch.

Soll nun Alice ein von Charly vorbereitetes Dokument M mittels RSA blind signieren, so führen die beiden folgendes Protokoll durch:

Blinde Signatur mittels RSA:

Der öffentliche Schlüssel von Alice ist (n, e) .

- Charly wählt eine Zufallszahl r und berechnet $M' = Mr^e \bmod n$
- Alice erhält und signiert M' : $\text{sig}_{\text{Alice}}(M') = \text{sig}_{\text{Alice}}(Mr^e \bmod n) = M^d r^{ed} \bmod n = M^d r \bmod n$
- Charly erhält $\text{sig}_{\text{Alice}}(M')$ und berechnet $M^d r r^{-1} \bmod n = M^d \bmod n = \text{sig}_{\text{Alice}}(M)$

Charly erhält als Resultat also eine Signatur von Alice über die eigentliche Nachricht M , die Alice nicht gesehen hat.

Anwendung findet die blinde Signatur bspw. bei elektronischen Bezahlverfahren. Der Aussteller einer „digitalen Münze“ kennt den Inhalt nicht, kann ihn aber signieren. Im Detail beschäftigen wir uns damit in Abschn. 6.2.

2.3.4 Kryptographische Hashfunktion

Hash-Funktionen sind uns bereits aus anderen Bereichen als der Kryptographie bekannt, bspw. zur Realisierung der Datenstruktur „Hash Table“. Die Grundidee einer Hash-Funktion besteht im wesentlichen darin, eine Eingabe beliebiger Länge auf eine Ausgabe fester Länge abzubilden, z.B. eine Textdatei auf 128 Bit. Die Funktion sollte dabei effizient berechenbar sein.

Eine *kryptographische* Hash-Funktion $h(x)$ ist eine Hash-Funktion mit den folgenden Eigenschaften:

- Zu einem gegebenen $h(x)$ kann kein passendes x effizient berechnet werden (Resistenz gegenüber Urbild-Angriff),
- Zu einem gegebenen x_1 kann kein x_2 effizient berechnet werden, so dass $h(x_1) = h(x_2)$ (Resistenz gegenüber „Zweites-Urbild-Angriff“),
- Es ist nicht effizient möglich, ein Paar (x_1, x_2) zu finden, so dass $h(x_1) = h(x_2)$.

Die zweite und dritte Eigenschaft werden auch schwache Kollisionsresistenz und starke Kollisionsresistenz genannt. Der Unterschied ist, dass im ersten Fall einer der beiden Werte, x_1 , fest vorgegeben ist.

Mit „nicht effizient“ meinen wir hier: Nicht schneller als durch systematisches Ausprobieren aller Möglichkeiten („brute force“).

Anwendung finden kryptographische Hash-Funktionen u. A. zur (symmetrischen) Datenauthentifizierung. Nehmen wir wieder an, dass Alice und Bob miteinander kommunizieren und einen gemeinsamen Schlüssel haben. Niemand soll die ausgetauschten Daten unbemerkt modifizieren können. Die Lösung besteht darin, einen sogenannten *Message Authentication Code* (MAC) an die Daten anzuhängen. Der MAC könnte bspw. folgendermaßen berechnet werden: $MAC = Hash(\text{Daten} || \text{Schlüssel})$. Diese Variante der Berechnung wird auch als „keyed hash“ bezeichnet.

Auf den ersten Blick verfolgen MACs und digitale Signaturen ähnliche Ziele: beide gewährleisten *Integrität* und *Authentizität*. Da es sich bei MACs allerdings um symmetrische Verfahren handelt, ist für die Datenauthentifizierung ein gemeinsames Geheimnis (Schlüssel) mit dem Kommunikationspartner vonnöten. Bei digitalen Signaturen, als Vertreter von asymmetrischen Verfahren, ist für die Prüfung der Datenauthentizität nur der öffentliche Schlüssel des Kommunikationspartners nötig. Neben den Unterschieden hinsichtlich der Performanz (MACs weisen einen geringen Rechenaufwand auf, während Signaturen sehr rechenaufwändig sind), gibt es noch einen zentralen Unterschied: Digitale Signaturen erfüllen das Schutzziel *Verbindlichkeit*. Der Nachweis der Authentizität erfolgt gegenüber jedem Inhaber des öffentlichen Schlüssels. Bei MACs hingegen ist für Dritte nicht nachvollziehbar, ob nun Alice oder Bob – die beide in Besitz des gemeinsamen Geheimnisses sind – eine Nachricht authentifiziert hat.

Kommen wir noch einmal auf den hohen Rechenaufwand bei der Erstellung digitaler Signaturen zu sprechen. Der Rechenaufwand zum „direkten“ Signieren eines langen

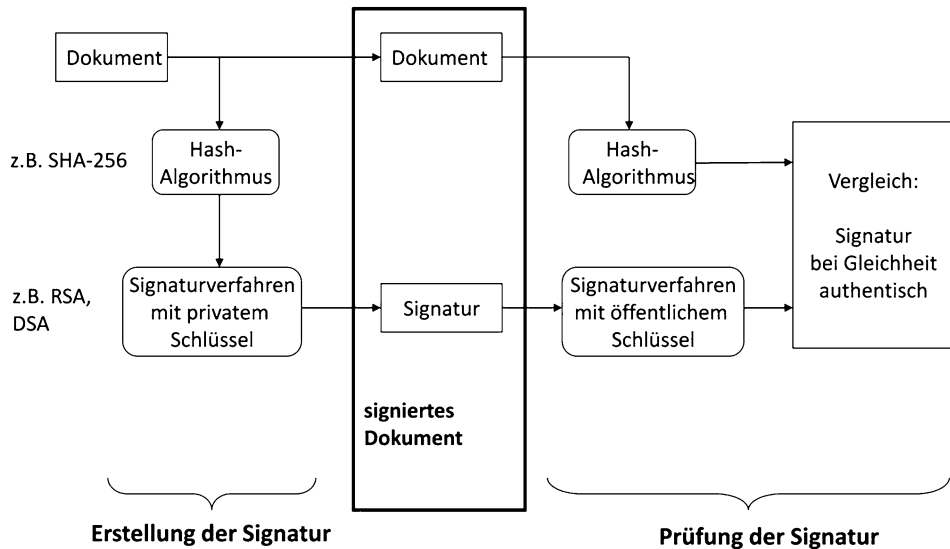


Abb. 2.4 Dokument-Signatur: Zusammenspiel von kryptographischen Hash-Funktionen und Digitalen Signaturen

Dokuments mit einem asymmetrischen Verfahren wäre für Alice zu hoch. Stattdessen greift sie auf eine kryptographische Hashfunktion zurück. Sie berechnet, wie in Abb. 2.4 dargestellt, zunächst den Hashwert des Dokuments und signiert dann diesen kurzen Hashwert mit dem asymmetrischen Verfahren. Bob wendet zum Prüfen der Signatur dieselbe Hash-Funktion auf das Dokument an und prüft, ob die Signaturverifikation mit dem öffentlichen Schlüssel von Alice denselben selbst berechneten Hash-Wert ergibt – ist dies der Fall, kann sich Bob sicher sein, dass Alice das Dokument signiert hat und dass das Dokument nicht manipuliert wurde.

2.3.5 Diffie-Hellman-Verfahren

Für die Ver- und Entschlüsselung von Nachrichten mit einem symmetrischen Verfahren benötigen Alice und Bob, wie wir zuvor gesehen haben, einen gemeinsamen Schlüssel. Zuvor hatten wir gesagt, dass der Schlüsselaustausch über einen sicheren Kanal erfolgen muss. Der *Diffie-Hellman (DH)-Schlüsselaustausch*, benannt nach seinen Entwicklern DIFFIE und HELLMAN, erlaubt einen derartigen Schlüsselaustausch auch über einen *unsicheren* Kanal. Das Ziel ist es, einen symmetrischen Schlüssel zu erzeugen, der nicht im Klartext übertragen wird. Bei dem DH-Verfahren handelt es sich um ein asymmetrisches Verfahren, das auf dem Problem des diskreten Logarithmus basiert.

Der Ablauf des DH-Schlüsselaustauschs gestaltet sich wie folgt.

DH-Schlüsselaustausch-Protokoll:

- Zunächst vereinbaren Alice und Bob eine große Primzahl p und einen Generator g . Diese Informationen können öffentlich sein.
- Alice wählt eine Zufallszahl a (mit $1 < a < p - 1$) und schickt $g^a \bmod p$ an Bob.
- Bob wählt eine Zufallszahl b (mit $1 < b < p - 1$) und schickt $g^b \bmod p$ an Alice.
- Beide können nun $K = (g^a \bmod p)^b \bmod p = (g^b \bmod p)^a \bmod p = g^{ab} \bmod p$ berechnen.

Der Generator g hat die Eigenschaft, dass alle Zahlen $1, \dots, p - 1$ als Potenz $g^i \bmod p$ dargestellt werden können. Die Kenntnis von $g^a \bmod p$ und $g^b \bmod p$ reicht einer Lauscherin Eve nicht, um K zu berechnen.

Dieses Verfahren ist nach heutigem Stand sicher gegen Angreifer, die nur mithören (passive Angreifer). Es ist anfällig gegen Man-in-the-Middle (MITM)-Angriffe, bei denen der Angreifer je einen DH-Austausch mit Alice und mit Bob durchführt. Als Gegenmaßnahme käme bspw. eine Authentifizierung mittels digitaler Signatur in Betracht.

Der DH-Schlüsselaustausch liefert zudem die für unsere späteren Verfahren schöne Eigenschaft der sogenannten *Perfect Forward Secrecy* (PFS). PFS bedeutet, dass es kein langlebiges Geheimnis gibt, dessen Kompromittierung die gesamte, aufgezeichnete Kommunikation der Vergangenheit aufdecken würde. Sofern für jede übermittelte Nachricht ein neuer DH-Schlüssel (*Sitzungsschlüssel*) ausgetauscht wird, kann ein Angreifer, dem es gelingt einen einzigen DH-Schlüssel K zu erfahren, nicht die Inhalte der anderen Nachrichten erlangen.

Wenn wir uns vorstellen, dass Alice und Bob hingegen einen Schlüsselaustausch mittels RSA durchführen würden – wie dies in der Praxis auch häufig gemacht wird – wäre PFS nicht gewährleistet. Nehmen wir an, Alice verschlüsselt mit RSA vor jeder Nachrichtenübermittlung einen neuen, zufällig gewählten symmetrischen Sitzungsschlüssel mit dem öffentlichen Schlüssel von Bob und sendet ihm das Chifftrat. Des Weiteren verschlüsselt sie die (lange) Nachricht mit einem symmetrischen Verfahren und mit dem zuvor erstellten Sitzungsschlüssel und sendet dieses Chifftrat auch an Bob. Bob würde nun im ersten Schritt mit seinem privaten Schlüssel (und dem asymmetrischen Verfahren) den Sitzungsschlüssel entschlüsseln. Im nächsten Schritt würde er mit diesem Sitzungsschlüssel (und dem symmetrischen Verfahren) die Nachricht entschlüsseln. Wenn es nun einem Angreifer, der die gesamte verschlüsselte Kommunikation zwischen Alice und Bob aufzeichnet, irgendwann in der Zukunft gelingen würde, den privaten RSA-Schlüssel von Bob zu erfahren, so könnte er alle zuvor ausgetauschten Nachrichten zwischen Alice und Bob entschlüsseln.

Als nächstes können wir uns nun grundlegenden Verfahren aus der IT-Sicherheit zuwenden, die allesamt auf den kennengelernten kryptographischen Verfahren aufbauen.

2.4 Grundlegende Verfahren aus der IT-Sicherheit

In diesem Abschnitt möchten wir einen kurzen Überblick über aktuelle Verfahren geben, die heute in der Praxis häufig zum Einsatz kommen, wenn es um die Erfüllung der „klassischen“ IT-Sicherheits-Schutzziele aus Abschn. 2.1.1 geht. Diese Verfahren, die im Hinblick auf die in der Anlage zu § 9 Bundesdatenschutzgesetz (BDSG) geforderte Maßnahme zur „Weitergabekontrolle“ dem *Stand der Technik* entsprechen, werden klassischerweise dem Bereich der *IT-Sicherheit* zugeordnet. Allen voran das gebotene Schutzziel *Vertraulichkeit* ist aber auch dem Datenschutz sehr dienlich, weshalb viele Privacy-Enhancing Technologies (PETs) auch auf diesen etablierten „Standard“-Mechanismen aufbauen. Wir möchten darauf hinweisen, dass die *IT-Sicherheit* und der *Datenschutz* auf den ersten Blick unterschiedliche Ziele verfolgen. Bei der IT-Sicherheit steht typischerweise tatsächlich der „Schutz von Daten“ und Datenverarbeitungsanlagen im Vordergrund. Hier geht es etwa darum sicherzustellen, dass Daten nicht unberechtigt gelesen werden können, dass Systeme nicht unberechtigt genutzt werden können etc. Beim Datenschutz hingegen steht, wie bereits zuvor erwähnt, der Schutz von Personen im Vordergrund. Es geht u. A. darum zu kontrollieren, wie mit personenbezogenen Daten umgegangen wird. Bei genauerer Betrachtung wird schnell klar, dass in vielen Fällen dieselben Maßnahmen, etwa kryptographische Verfahren, angewendet werden können, um sowohl die Schutzziele der IT-Sicherheit als auch die des Datenschutzes zu erfüllen. Dies gilt insbesondere für die beiden Verfahren, die wir als nächstes genauer betrachten.

2.4.1 Transport Layer Security

Das *Transport Layer Security (TLS)*-Protokoll deckt die Schutzziele *Authentizität*, *Integrität*, sowie *Vertraulichkeit* bei der Übermittlung von Daten auf der Transport-Ebene ab. Bei TLS handelt es sich um ein weit verbreitetes Sicherheits-Protokoll, das als Grundlage zur Absicherung einer Vielzahl an „darüber liegenden“ Diensten/Anwendungen dient. So sorgt TLS bspw. für die Sicherheit für einzelne Anwendungen im Web, wie dem Online-Banking oder dem Online-Shopping. Erkennbar ist der Einsatz von TLS dabei an dem Präfix „https“ in der Adresszeile des Browsers. Im E-Mail-Bereich gibt es mit SMTPS die über TLS abgesicherte Version von SMTP. Die aktuelle Version ist TLS 1.2, spezifiziert in RFC 5246. Secure Sockets Layer (SSL), die Vorgängerversion von TLS, entspricht nicht mehr dem Stand der Technik und sollte aus Sicherheitsgründen entsprechend nicht mehr eingesetzt werden. Vielen Nutzern ist das Protokoll immer noch unter „SSL“ bekannt und so werben auch einige Certificate Authoritys (CAs) noch mit der Ausstellung von „SSL-Zertifikaten“.

Verbindungsaufbau

Beim Verbindungsaufbau, dem sogenannten „TLS Handshake“, erfolgt einerseits die *Authentifizierung der Kommunikations-Partner* (in der Regel des Servers gegenüber dem Client), sowie die *Aushandlung des Sitzungsschlüssels* für die folgende Kommunikation.

Es gibt unterschiedliche Varianten des Verbindungsaufbaus. Um die Langzeitsicherheit, die sogenannte *Perfect Forward Secrecy (PFS)*, von übertragenen Daten sicherzustellen, sollte der Schlüsselaustausch nach Diffie-Hellman (DH), wie in Abschn. 2.3.5 beschrieben, erfolgen.

Die Authentifizierung des Servers gegenüber dem Client erfolgt dabei dadurch, dass der Server seinen für den DH-Austausch verwendeten, öffentlichen Parameter³ mit seinem privaten (Langzeit-)Schlüssel – in den meisten Fällen ein RSA-Schlüssel – signiert. Zusätzlich übermittelt der Server dem Client ein von einer vertrauenswürdigen Zertifizierungsstelle (einer *Certificate Authority (CA)*) ausgestelltes X.509-Zertifikat, in dem die CA mit ihrer digitalen Signatur bestätigt, dass der öffentliche (Langzeit-)RSA-Schlüssel im Zertifikat tatsächlich diesem Server gehört. Der Client kann nach Prüfung des Zertifikats⁴ und Entnahme des öffentlichen Schlüssels aus dem Zertifikat die Signatur (über den öffentlichen DH-Parameter) des Servers überprüfen und damit sicher sein, dass er mit dem „erwarteten“ Server kommuniziert; nur der Server ist in Besitz des zum im Zertifikat enthaltenen öffentlichen Schlüssel zugehörigen privaten Schlüssels, der für die Signaturerstellung verwendet wurde.

Im Anschluss an den Schlüsselaustausch etablieren Client und Server den „sicheren Kanal“, über den sie nun kommunizieren. Zur Verschlüsselung der Daten wird ein symmetrisches Verschlüsselungsverfahren verwendet, das den zuvor per DH-Verfahren ausgehandelten Sitzungsschlüssel – der genau für diese eine Kommunikationsbeziehung gilt⁵ – nutzt. Die Integrität und Authentizität der Nachrichten wird auch über ein symmetrisches Verfahren, das ebenso das zuvor ausgehandelte Schlüsselmateriale nutzt, gewährleistet. In der Praxis authentifiziert sich der Client gegenüber dem Server innerhalb des sicheren Kanals; im Web bspw. durch Eingabe von Benutzername und Passwort.

Schwachstellen

In den letzten Jahren wurden einige Schwachstellen bei der Implementierung des SSL/TLS-Protokolls aufgedeckt. Eine der gravierendsten war der sogenannte „Heartbleed-

³Für jede Kommunikationsverbindung werden neue DH-Parameter gewählt, man spricht in diesem Kontext auch von „ephemeral“ DH-Schlüsseln.

⁴Hierzu gehört u. A. die Prüfung der Zertifikatskette, der Signatur der CA, des Gültigkeitszeitraums etc.

⁵Nach Beendigung der Kommunikationsbeziehung muss der Sitzungsschlüssel gelöscht werden; nur so kann PFS erreicht werden.

Bug“, der im April 2014 entdeckt wurde und von dem zahlreiche Server betroffen waren, die die OpenSSL Kryptographie-Bibliothek nutzten – der Fehler bestand dort seit Ende 2011.

Daneben gab es zahlreiche Vorfälle bei den Zertifizierungsstellen. So wurden bspw. Fälle von „Einbrüchen“ bei CAs bekannt, bei denen sich die Täter etwa (im Namen der CA) selbst ein Zertifikat für *.google.com ausgestellt haben. Damit waren Man-in-the-Middle (MITM)-Angriffe auf Nutzer von Google-Diensten möglich. Den Nutzern, die der ausstellenden CA vertrauten, konnte das Zertifikat untergeschoben werden und die Angreifer konnten sich auf diese Art als Betreiber von Google-Diensten ausgeben. Die Nutzer hatten also eine „sichere Verbindung“ über TLS aufgebaut – allerdings nicht mit Google sondern mit den Angreifern. In den Medien wurde berichtet, dass iranische Täter hinter einem dieser Einbrüche vermutet wurden und der Google-Dienst „GMail“ im Zentrum des Angriffs stand. Eine Regierung, die die Netz-Infrastruktur eines Landes kontrolliert und es dann noch schafft, „richtige“ Zertifikate für von den Bürgern dieses Landes genutzte (ausländische) Dienste zu erlangen, ist in der Lage, sehr weit in die Privatsphäre der Bürger einzudringen.

Aus heutiger Sicht sollte das veraltete SSL-Protokoll nicht mehr zum Einsatz kommen, hier bestehen zu viele Sicherheitslücken. Es sollte auf die neue TLS-Version 1.2 in einer aktuellen Implementierung zurückgegriffen werden. Das Bundesamt für Sicherheit in der Informationstechnik (BSI) rät von der Verwendung von TLS 1.0, dem heute noch am weit verbreitetsten TLS-Protokoll, ab. TLS in der Version 1.1 sollte aus Sicht des BSI ebenfalls nur noch als Übergangslösung zum Einsatz kommen [3]. Eine Untersuchung von ca. 40.000 Websites von Unternehmen in Baden-Württemberg durch den Landesbeauftragten für den Datenschutz Baden-Württemberg im Jahr 2016 hat allerdings ergeben, dass bislang keine Webserver-Betreiber die strengen BSI-Vorgaben einhalten [8].

2.4.2 Virtual Private Networks

Virtual Private Networks (VPNs) werden vor allem im Unternehmensumfeld häufig eingesetzt, um einen „sicheren Kanal“ zwischen zwei Standorten, oder auch für die Anbindung von Außendienstmitarbeitern an das Unternehmensnetzwerk zu schaffen. Im Gegensatz zum zuvor vorgestellten TLS-Protokoll, das auf der Transportebene im ISO/OSI-Schichtenmodell⁶ angesiedelt ist, wird ein VPN typischerweise eine Schicht darunter, auf der Netzwerkschicht, angesiedelt. Der sichere Kanal umfasst somit die gesamte IP-Kommunikation zwischen den beteiligten Instanzen und nicht, wie im TLS-Protokoll, „nur“ die Kommunikation zwischen zwei Anwendungen.

⁶Die Details zum ISO/OSI-Schichtenmodell können Sie in Lehrbüchern zu Kommunikationsnetzen nachschlagen.



Abb. 2.5 VPN: Tunnel-Modus bei IPsec

Für die Etablierung eines VPNs kann *IPsec* zum Einsatz kommen; ein Protokoll, das bei IPv6 ebenfalls Anwendung findet. IPsec ist in den RFCs 4301–4309 spezifiziert.⁷ IPsec bietet unterschiedliche Modi und Protokolle. Für uns sind für die weitere Betrachtung lediglich der *Tunnelmodus* und das *Encapsulating Security Payload (ESP)*-Protokoll relevant.

Der Tunnelmodus sorgt, wie in Abb. 2.5 dargestellt, dafür, dass zum eigentlichen IP-Paket ein zusätzlicher, äußerer IP-Header hinzugefügt wird. Die IPsec-Daten sowie der innere IP-Header werden als „Payload“ des äußeren IP-Pakets transportiert.

Das ESP-Protokoll erlaubt die Verschlüsselung des gesamten inneren IP-Pakets. Damit wird das Schutzziel *Vertraulichkeit* bei IPsec erfüllt. Außerdem sorgt ESP für die *Integrität* und *Authentizität* der übertragenen IP-Pakete. Im Tunnelmodus ist zusätzlich auch der oben dargestellte IPsec-Header integritätsgeschützt.

Sehen wir uns nun den Ablauf bei IPsec bei Verwendung des Tunnelmodus und ESP-Protokolls am Beispiel der Anbindung eines Außendienstmitarbeiters an sein Unternehmensnetzwerk genauer an. Der Mitarbeiter sendet das in der Abbildung dargestellte IP-Paket an das IPsec-Gateway seines Unternehmens. Das Gateway prüft die Integrität und Authentizität und entschlüsselt das innere, eigentliche IP-Paket (innerer IP-Header und Payload). Erst jetzt liegt dem Gateway die Adresse des Zielrechners, die sich im inneren IP-Header befindet, im Klartext vor. Das Gateway sendet das IP-Paket nun an die gewünschte Station weiter.

Damit wird klar, dass ein VPN auch einen (limitierten) Schutz vor Verkehrsflussanalyse bietet. Ein Außenstehender sieht lediglich, dass der Mitarbeiter mit seinem Unternehmen kommuniziert, aber nicht, mit welchem Zielsystem. Genau so wenig erfährt ein Außenstehender, mit welchem Zielsystem ein Student kommuniziert, der eine VPN-Verbindung zu seiner Universität aufgebaut hat und diese für seine Aktivitäten nutzt. Mit dem Thema *Verkehrsflussanalyse* werden wir uns in Abschn. 4.1 noch näher befassen.

⁷RFC 4305 ist inzwischen durch RFC 4835 ersetzt worden.

2.5 Fazit

In diesem Kapitel haben wir die technischen Grundlagen für die weiteren Kapitel gelegt. Sie sollten die wichtigsten kryptographischen Verfahren und IT-Sicherheits-Protokolle kennen und wissen, welche Schutzziele diese erfüllen. Außerdem sollten sie wissen, was unter Anonymität und Pseudonymität zu verstehen ist. In den nächsten Kapiteln werden wir immer wieder auf die hier vorgestellten Verfahren zurückgreifen und zeigen, wie diese in bestimmten Szenarien zur Erfüllung von Datenschutz-Schutzzielen angewandt und kombiniert werden können.

2.6 Übungsaufgaben

Aufgabe 1

„Klassische“ Schutzziele der IT-Sicherheit stehen gelegentlich im Widerspruch zum Datenschutz – zeigen Sie einen solchen Zielkonflikt an einem *konkreten* Beispiel auf!

Aufgabe 2

Geben Sie ein konkretes Beispiel eines Angriffs an, durch den ein Angreifer eine positive Anonymitätsdifferenz erzielen kann, ohne dabei ein Subjekt eindeutig zu identifizieren! Wie können Sie die Anonymität in Ihrem Beispiel messen?

Literatur

1. David Chaum. Blind signatures for untraceable payments. In *Advances in Cryptology – Crypto '82*, pages 199–203. Springer, 1982.
2. Claudia Eckert. *IT-Sicherheit*, 8 Auflage. Oldenbourg-Verlag, 2013.
3. Bundesamt für Sicherheit in der Informationstechnik. Kryptographische Verfahren: Empfehlungen und Schlüssellängen. Teil 2 – Verwendung von Transport Layer Security (TLS), Version 2016–01. Technische Richtlinie TR-02102-2, Jan. 2016.
4. Konferenz der Datenschutzbeauftragten des Bundes und der Länder. Ein modernes Datenschutzrecht für das 21. Jahrhundert. Technical report, 2010. Herausgegeben durch den Vorsitzenden der Konferenz im Jahr 2010: Der Landesbeauftragte für den Datenschutz Baden-Württemberg.
5. Andreas Pfitzmann and Marit Hansen. A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management. Technical Report Version v0.33, Apr. 2010. http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.33.pdf.
6. Martin Rost und Kirsten Bock. Privacy By Design und die Neuen Schutzziele. *Datenschutz und Datensicherheit – DuD*, 35(1):30–35, 2011.
7. Christoph Sorge, Nils Gruschka, und Luigi Lo Iacono. *Sicherheit in Kommunikationsnetzen*. Oldenbourg Wissenschaftsverlag, 2013.
8. Ronald Petrlc und Klaus Manny. Wie sicher ist der Zugriff auf Websites im Internet? *Datenschutz und Datensicherheit – DuD*, 41(2):88–92, 2017.

Zusammenfassung

Eine besondere Herausforderung für die Praxis ist das *Anonymisieren* von Daten. Es existieren zahlreiche Anonymisierungs-Techniken – einige davon greifen wir in diesem Kapitel auf. So beschäftigen wir uns in Abschn. 3.1 zunächst mit den allgemeinen Anforderungen an Anonymisierungs-Techniken. Danach lernen wir in Abschn. 3.2 mit dem Konzept der *k*-Anonymität einen der grundlegendsten und bekanntesten Ansätze zur Anonymisierung von Daten kennen. Danach betrachten wir in Abschn. 3.3 mit *Differential Privacy* ein neueres Konzept zur Anonymisierung von Daten, das vor allem von Seiten der Forschung als zur Zeit aktuellstes Verfahren in diesem Bereich gesehen wird. Schließlich gehen wir in Abschn. 3.4 auf die Praxis im Hinblick auf Anonymisierung ein.

Lernziele

Am Ende dieses Kapitels sollten Sie Verfahren zur Anonymisierung von Daten kennen. Sie sollten auch die Schwachstellen der unterschiedlichen Ansätze kennen und die erreichten Anonymitätsmaße bewerten können, um zu entscheiden, ob Daten hinreichend gut anonymisiert wurden.

3.1 Überblick

Die Motivation für die Anonymisierung von Daten besteht darin, dass eine Teilmenge von Daten aus einer (Datenbank-)Tabelle, die personenbezogene Daten enthält, veröffentlicht werden soll. Eine Identifikation einzelner Personen soll dabei allerdings nicht möglich sein.

Beispiel

Häufig wird in diesem Zusammenhang das Veröffentlichen von Krankheitsdaten für Forschungszwecke genannt. Forscher interessieren sich für die statistischen Zusammenhänge bei bestimmten Krankheiten um z.B. zu untersuchen, welchen Einfluss bestimmte Faktoren wie das Umfeld/der Wohnort, das Alter, oder die Ernährung auf einen Krankheitsverlauf haben. Krankenhäuser verfügen in der Regel über solche Daten – dürfen diese aber nur anonymisiert weitergeben, d. h. einzelne Patienten dürfen nicht identifizierbar sein. Wie wir später sehen werden, ist es dabei nicht damit getan, einfach die offensichtlichen identifizierenden Daten wie Name oder Anschrift der Patienten zu entfernen, um eine korrekte Anonymisierung der Daten durchzuführen, die keine Rückschlüsse auf die Personen (d. h. eine Re-Identifizierung) zulassen.

3.1.1 Anonymitäts-Modelle

Szenarien

Wenn wir uns das oben genannte Beispiel ansehen, dann gehen wir davon aus, dass in solch einem Fall die weiterzugebenden Daten typischerweise bereits vorliegen und vor der Weitergabe („am Stück“) anonymisiert werden sollen. Man spricht in diesem Fall auch von einem *nicht-interaktiven* Szenario der Datenweitergabe. Bei diesem Szenario, bei dem die einzelnen (anonymisierten) Datensätze weitergegeben werden, geht man davon aus, dass diese Datensätze größtmögliche „Nutzbarkeit“, etwa für die Forschung, aufweisen. Vor der Weitergabe muss genau überlegt werden, wie die Anonymisierung durchzuführen ist, so dass niemand in der Lage ist, die Daten zu de-anonymisieren. Man muss davon ausgehen, dass der anonymisierte Datensatz öffentlich bekannt wird. Eine Methode der Anonymisierung besteht dabei in der Sicherstellung der *k-Anonymität*, mit der wir uns in Abschn. 3.2 genauer beschäftigen werden. Daneben existiert das sogenannte *interaktive* Szenario der Datenweitergabe. Hier geht man davon aus, dass nicht der komplette Datensatz (anonymisiert) weitergegeben wird, sondern dass Interessierte einzelne Anfragen stellen können und anonymisierte Antworten – in der Regel nur aggregierte Werte über die Datensätze – erhalten. Die Nutzbarkeit der Daten ist dabei eingeschränkter als bei der Weitergabe einzelner Datensätze – dies hängt jedoch stark vom Szenario und den Fragestellungen ab. In Abschn. 3.3 beschäftigen wir uns mit *Differential Privacy*, einem Vertreter eines Anonymisierungsansatzes, der zum interaktiven Szenario gezählt werden kann. Dabei werden wir sehen, dass zur Anonymisierung nicht nur eine Aggregation der Daten stattfindet, sondern zusätzlich noch ein „Rauschen“ hinzugefügt wird, um ein möglichst hohes Maß an Anonymität sicherzustellen.

Bedrohungen

In der Literatur¹ wird zwischen unterschiedlichen Typen von Attributen (von personen-bezogenen Daten) unterschieden: (*direkten*) *Identifikatoren*, *Quasi-Identifikatoren* und *sensiblen Werten*. Erstgenannte Attribute (bspw. Name, (E-Mail-)Adresse, Personalausweisnummer etc.) erlauben eine direkte Re-Identifizierung von natürlichen Personen. Quasi-Identifikatoren, mit denen wir uns in Abschn. 3.1.2 genauer befassen, erlauben erst in Form von Attribut-Kombinationen (bspw. Postleitzahl in Verbindung mit Geschlecht und Geburtsdatum) eine Re-Identifizierung. Sensible Werte sind jene Werte, mit denen natürliche Personen nicht in Verbindung gebracht werden möchten – dies können bspw. Diagnosen sein. Ausgehend von diesen Attribut-Typen können wir nun die folgenden Bedrohungen feststellen:

- *Aufdecken der Identität (Re-Identifizierung)*: Bei dieser Bedrohung geht es darum, dass ein Angreifer einen anonymisiert veröffentlichten Datensatz einer natürlichen Person zuordnen kann.
- *Aufdecken der Zugehörigkeit*: Bei dieser Bedrohung geht es darum, dass ein Angreifer mit hoher Wahrscheinlichkeit sicher sein kann, dass der Datensatz einer bestimmten natürlichen Person in dem anonymisiert veröffentlichten Datensatz enthalten ist.
- *Aufdecken eines Attributs (sensible Information)*: Bei dieser Bedrohung geht es darum, dass ein Angreifer eine natürliche Person mit einer sensiblen Information in Verbindung bringen kann.

Die erste Bedrohung ist die „klassische“ Bedrohung, wenn es um das Thema Re-Identifizierung (auch De-Anonymisierung genannt) geht. SWEENEY [8] hat eindrucksvoll bewiesen, dass die Re-Identifizierung von anonymisiert geglaubten Daten möglich ist, wie wir in Abschn. 3.2 sehen werden, und hat mit dem Konzept der *k-Anonymität* ein Verfahren vorgestellt, um Re-Identifizierung zu erschweren. Die zweite Bedrohung kann selbst Daten treffen, die gegen das Aufdecken der Identität „gesichert“ sind. Als Beispiel kann hier eine Datenbank dienen, die Daten über HIV-positiv-Erkrankte enthält. Wenn es einem Angreifer gelingt, festzustellen, dass der Datensatz einer bestimmten Person in der Datenbank enthalten ist, hat er damit die Information gewonnen, dass der Patient HIV-positiv ist – ohne den entsprechenden Datensatz dem Patienten zuordnen zu müssen. Bei der dritten Bedrohung besteht das Problem darin, dass ein Angreifer zusätzliche Informationen über eine Person gewinnen kann. In diesem Kontext ist die *Netflix-De-Anonymisierung* einzuordnen. Netflix hat im Jahr 2006 hundert Millionen Datensätze zu Film-Bewertungen von einer halben Million seiner Nutzer „anonymisiert“ veröffentlicht. Tatsächlich wurden die Daten nicht anonymisiert sondern pseudonymisiert, d. h. die Nutzernamen wurden durch ein gleichbleibendes Pseudonym ersetzt. Wissenschaftler [6]

¹Anonymisierung spielt vor allem im Gesundheitsbereich eine wesentliche Rolle. Die folgende Darstellung orientiert sich an [3].

konnten nachweisen, dass schon wenige Informationen über Nutzer (bewertete Filme) reichen, um die pseudonymen Netflix-Datensätze diesen Nutzern zuzuordnen. Dass bei der *Internet Movie Database* (IMDb) die Film-Bewertungen inklusive Nutzernamen veröffentlicht werden, konnte zusätzlich ausgenutzt werden, um IMDb-Nutzer mit den vermeintlich anonymisiert veröffentlichten Netflix-Bewertungen (wo auch „sensible“ Filme enthalten sein könnten, die Nutzer nicht mit ihrem richtigen Namen bewerten würden) in Verbindung zu bringen. Laut GKoulalas-Divanis et al. [3] führten die zweite und dritte Bedrohung bisher noch zu keinen dokumentierten Datenschutz-Verstößen im Gesundheitsbereich.

Modelle

Gkoulalas-Divanis et al. [3] haben für den Gesundheitsbereich die Anonymitäts-Modelle herausgearbeitet, die für die unterschiedlichen Szenarien nach Bewertung der Bedrohungslage sinnvoll sind. Beispielhaft möchten wir hier im Kontext der ersten Bedrohung *k*-Anonymität zum Schutz von demographischen Daten und k^m -Anonymität zum Schutz von sensiblen Werten nennen. In Bezug auf die zweite Bedrohung kann zum Schutz der demographischen Daten etwa *Differential Privacy* oder δ -Presence zum Einsatz kommen. Für den Schutz demographischer Daten in Bezug auf die dritte Bedrohung können *l*-Diversity und *t*-Closeness zum Einsatz kommen und für den Schutz sensibler Werte *p*-Uncertainty. In diesem Buch beschäftigen wir uns lediglich mit *k*-Anonymität und *Differential Privacy* im Detail. Interessierte Leser möchten wir auf die bei Gkoulalas-Divanis et al. [3] genannte Literatur verweisen.

3.1.2 Quasi-Identifikatoren

Bevor wir uns den Verfahren zur Anonymisierung von Daten zuwenden, müssen wir uns noch mit dem Begriff des *Quasi-Identifikators* beschäftigen.

► **Definition** *Quasi-Identifikatoren* sind Attributkombinationen, die in Verbindung mit extern verfügbaren Informationen die eindeutige Identifikation einer Person ermöglichen.

Das Problem bei der Anonymisierung besteht im Wesentlichen darin, dass eine Identifikation einer einzelnen Person nicht nur durch „klassische“ Identitätsmerkmale (z. B. Name, Geburtsdatum, Geburtsort) möglich ist.

Beispiel

Die Kombination (Beruf, Wohnort) ist für die meisten Menschen wahrscheinlich nicht eindeutig – aber was, wenn der Beruf „Bürgermeister“ ist?

Dass eine Re-Identifizierung von eigentlich anonymisiert geglaubten Daten möglich sein kann, wird von vielen nicht vorher erwartet und es hat in der Vergangenheit schon unzählige Fälle gegeben, in denen die Verantwortlichen nicht damit rechneten, wie im Fall

der *Massachusetts Group Insurance Commission*, der in diesem Zusammenhang häufig als lehrreiches Beispiel genannt wird [8].

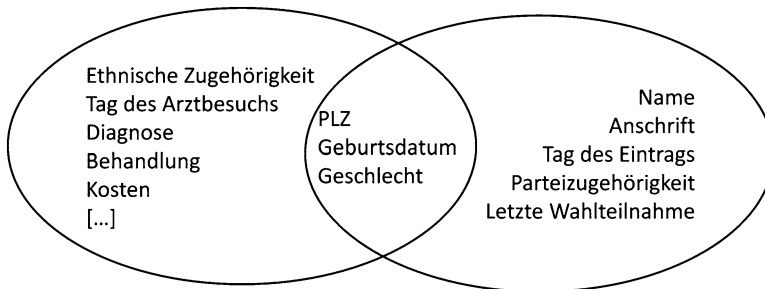
Beispiel

Die Massachusetts Group Insurance Commission ist zuständig für die Krankenversicherung von 135.000 Staatsbediensteten und deren Familien. Sie sammelt die Gesundheitsdaten (einschließlich der Diagnosen) sowie Postleitzahl (ZIP Code), Geburtsdatum, Geschlecht und ethnische Zugehörigkeit.

Die Weitergabe der „anonymisierten“ Daten (ohne Name/Anschrift) wurde als unproblematisch betrachtet. Aus diesem Grund wurden die Daten an die Industrie verkauft und an Forscher weitergegeben.

Nun kommen die externen Informationen, von denen zuvor bei der Definition von Quasi-Identifikatoren die Rede war, ins Spiel. Zur Kontrolle der ordnungsgemäßen Wahlvorgänge gibt es bspw. für Cambridge in Massachusetts ein öffentliches Wählerverzeichnis, das für 20 US-Dollar zu erwerben ist. Dieses Wählerverzeichnis enthält u. A. Name, Anschrift, Postleitzahl, Geburtsdatum und Geschlecht aller registrierten Wähler.

Legt man nun die Daten der Versicherungs-Datenbank und des öffentlichen Wählerverzeichnisses übereinander, so wird schnell klar, dass es sich bei der Schnittmenge, also der Attributkombination (PLZ, Geburtsdatum, Geschlecht) um einen Quasi-Identifikator in der Versicherungs-Datenbank handelt:



Dieser Quasi-Identifikator ermöglicht zusammen mit der extern verfügbaren Information aus dem öffentlichen Wählerverzeichnis eine Identifikation von Personen. So konnte bspw. der Gouverneur des Staats identifiziert werden; es gab nur eine männliche Person mit seinem Geburtsdatum und seiner Postleitzahl.

Zusätzliche Information

Eine Studie von 2006 [4] kommt zu dem Ergebnis, dass:

- 63,3 % der amerikanischen Bevölkerung eindeutig identifizierbar ist durch die Attributkombination (Geschlecht, Geburtsdatum, 5-stellige Postleitzahl),

- 14,8 % identifizierbar ist durch (Geschlecht, Geburtsdatum, County),
- 0,2 % identifizierbar ist durch (Geschlecht, Geburtsjahr, 5-stellige Postleitzahl).

Dieses Ergebnis deutet schon einen möglichen Lösungsweg zur Anonymisierung an, nämlich das Einführen von Ungenauigkeit. Widmen wir uns als nächstes also den Anonymisierungs-Techniken.

3.2 k-Anonymität

Das Konzept *k*-Anonymität wurde 2002 von Sweeney [8] vorgestellt:

► **Definition** Eine Tabelle (Relation) erfüllt *k*-Anonymität, wenn mindestens *k* Zeilen (Tupel) in allen zum Quasi-Identifikator gehörenden Spalten (Attributen) identische Werte besitzen.

Anders gesagt: Zu jedem Tupel existieren mindestens *k*-1 andere, davon auch mit externen Informationen nicht unterscheidbare.

Beispiel

Die Tabelle weist eine *k*-Anonymität mit *k* = 2 („2-Anonymität“) auf; je zwei (hier: aufeinanderfolgende) Zeilen haben die gleichen Werte für alle Attribute des Quasi-Identifikators (bestehend aus den Feldern (Geburtsjahr, Postleitzahl, Geschlecht)):

Geburtsjahr	PLZ	Geschlecht	Diagnose
1982	33098	Männlich	Migräne
1982	33098	Männlich	Erkältung
1983	33098	Männlich	Rheuma
1983	33098	Männlich	Depression
1985	33100	Weiblich	Heuschnupfen
1985	33100	Weiblich	Hypochondrie
1983	33098	Weiblich	Übergewicht
1983	33098	Weiblich	Migräne

Das Herstellen von *k*-Anonymität ist allgemein nicht ohne Informationsverlust möglich. Mögliche Ansätze sind:

- Hinzufügen von Rauschen,
- Hinzufügen oder Löschen von Tupeln,
- Generalisierung von Daten.

Hinzufügen von Rauschen bedeutet, dass eine zufällige Veränderung einzelner Werte vorgenommen wird, bspw. durch Addition von zufällig (gemäß einer bekannten Verteilung) gezogenen Werten.

Für das Hinzufügen oder Löschen von Tupeln gilt: Gibt es eine bestimmte Kombination von Werten für Attribute des Quasi-Identifikators nur $k-1$ mal, so können diese $k-1$ Tupel gelöscht werden oder noch ein k -tes Tupel hinzugefügt werden. Die Werte der Attribute, die nicht zum Quasi-Identifikator gehören, können dabei z. B. zufällig gewählt werden.

Bei der Generalisierung von Daten wird die Genauigkeit von Werten reduziert; die Werte werden also verallgemeinert – und zwar so lange, bis die k -Anonymität erreicht ist.

Tendenziell gilt, dass das *Löschen* von Tupeln geeignet ist, wenn *genaue Angaben über wenige Subjekte* benötigt werden, wohingegen die *Generalisierung* geeigneter ist, wenn *ungenau Angaben über viele Subjekte* benötigt werden. Es hängt vom Szenario ab, welche Methode am vielversprechendsten ist, d. h. dass am Ende mit den anonymisierten Daten noch statistisch signifikante Aussagen gemacht werden können. Bei den Übungsaufgaben können Sie für ein konkretes Szenario die beiden Methoden „durchspielen“ und selbst herausfinden, welcher Ansatz vielversprechender ist.

3.2.1 Generalisierung von Daten

Sehen wir uns die Generalisierung von Daten etwas genauer an. Eine minimale Generalisierung durchzuführen, bei der möglichst viele Informationen erhalten bleiben, ist wahrscheinlich NP-schwer – für einige Spezialfälle wurde dies auch bereits bewiesen.

Eine Generalisierung führt man in der Regel abhängig vom Attribut durch. So lässt sich etwa ein Geburtsdatum „01.02.1970“ zu „02/1970“ oder noch einen Schritt weiter zu „1970“ generalisieren.

Eine Generalisierung kann dabei auf eine ganze Spalte angewendet werden oder auch nur auf einzelne Werte.

Beispiel

Für die folgende Tabelle soll eine k -Anonymität mit $k = 4$ erreicht werden:

Name	Geburtsjahr	PLZ	Geschlecht	Diagnose
John Doe	1982	33098	Männlich	Migräne
Thomas Muster	1982	33098	Männlich	Erkältung
Max Maier	1983	33098	Männlich	Rheuma
Otto Normal	1983	33098	Männlich	Depression
Jane Doe	1985	33100	Weiblich	Heuschnupfen
Lieschen Müller	1985	33100	Weiblich	Hypochondrie
Erika Musterfrau	1983	33098	Weiblich	Übergewicht
Jane Average	1983	33098	Weiblich	Migräne

Im ersten Schritt entfernen wir den Namen, der einen eindeutigen Schlüssel darstellt und bereits alleine ein Quasi-Identifikator ist:

Geburtsjahr	PLZ	Geschlecht	Diagnose
1982	33098	Männlich	Migräne
1982	33098	Männlich	Erkältung
1983	33098	Männlich	Rheuma
1983	33098	Männlich	Depression
1985	33100	Weiblich	Heuschnupfen
1985	33100	Weiblich	Hypochondrie
1983	33098	Weiblich	Übergewicht
1983	33098	Weiblich	Migräne

Im zweiten Schritt müssen wir noch weitere Attribute generalisieren. Wir möchten in diesem Beispiel das Geschlecht beibehalten, deshalb generalisieren wir die beiden anderen Attribute des Quasi-Identifikators:

Geburtsjahr	PLZ	Geschlecht	Diagnose
1982–1983	33098	Männlich	Migräne
1982–1983	33098	Männlich	Erkältung
1982–1983	33098	Männlich	Rheuma
1982–1983	33098	Männlich	Depression
1983–1985	33*	Weiblich	Heuschnupfen
1983–1985	33*	Weiblich	Hypochondrie
1983–1985	33*	Weiblich	Übergewicht
1983–1985	33*	Weiblich	Migräne

Damit haben wir die gewünschte k -Anonymität mit $k = 4$ erreicht.

3.2.2 Angriffe auf k -Anonymität

In den letzten Jahren wurden einige Angriffe auf k -Anonymität publiziert [5]. Eine Auswahl davon werden wir uns nun genauer ansehen.

Unsorted Matching-Angriff

In der Theorie haben Tupel einer relationalen Datenbank keine Reihenfolge. In der Praxis wird die Reihenfolge hingegen oft erhalten. Dieses Problem tritt bei der Veröffentlichung von Teilen einer Datenbank auf, die nicht miteinander verknüpft werden sollen.

Beispiel

Veröffentlichung zweier (eigentlich unkritischer) Teile der obigen Beispieltabelle mit $k = 2$:

Name	Geburtsjahr	PLZ	Geschlecht	Diagnose
John Doe	1982	33098	Männlich	Migräne
Thomas Muster	1982	33098	Männlich	Erkältung
Max Maier	1983	33098	Männlich	Rheuma
Otto Normal	1983	33098	Männlich	Depression
Jane Doe	1985	33100	Weiblich	Heuschnupfen
Lieschen Müller	1985	33100	Weiblich	Hypochondrie
Erika Musterfrau	1983	33098	Weiblich	Übergewicht
Jane Average	1983	33098	Weiblich	Migräne

Komplementärveröffentlichung

Eine Komplementärveröffentlichung ist auch bei zufälliger Reihenfolge problematisch. Damit kann die Kombination zweier k-anonymer Tabellen aus der gleichen Ursprungstabelle möglich werden.

Beispiel

Aus der folgenden Ausgangstabelle werden zwei Komplementärveröffentlichungen durchgeführt.

Geburtsjahr	PLZ	Geschlecht	Diagnose
1982	33098	Männlich	Migräne
1982	33100	Männlich	Erkältung
1982	33098	Weiblich	Rheuma
1983	33098	Weiblich	Depression
1983	33098	Männlich	Heuschnupfen
1983	33098	Weiblich	Migräne

Veröffentlichung 1:

Geburtsjahr	PLZ	Geschlecht	Diagnose
1983	33098	*	Depression
1982	33098	*	Migräne
1983	33098	*	Heuschnupfen
1982	33*	*	Erkältung
1983	33098	*	Migräne
1982	33*	*	Rheuma

Veröffentlichung 2:

Geburtsjahr	PLZ	Geschlecht	Diagnose
198*	33*	Männlich	Migräne
198*	33*	Männlich	Erkältung
198*	33*	Weiblich	Rheuma
198*	33*	Weiblich	Depression
198*	33*	Männlich	Heuschnupfen
198*	33*	Weiblich	Migräne

Wie wir sehen, lassen sich über die Diagnose die beiden 2-anonymen Tabellen doch wieder verknüpfen, obwohl die Tupel in den Veröffentlichungen unsortiert veröffentlicht wurden. Fehlt bspw. beim Datensatz mit der Diagnose „Heuschnupfen“ in der ersten Veröffentlichung das Geschlecht, so lässt sich das Geschlecht in der zweiten Veröffentlichung über dieselbe Diagnose ermitteln.

Die Diagnose ist eigentlich nicht Teil des Quasi-Identifikators. Dennoch wird sie in dem Beispiel verwendet, um eine Verknüpfung der (für sich allein genommenen jeweils k-anonymen) Tabellen herzustellen.

Als Lösung für das Problem der Komplementärveröffentlichung müssen bereits veröffentlichte Tabellen als öffentliche Informationen betrachtet werden. Es kommt also zu einer Erweiterung des Quasi-Identifikators. Alternativ können alle Veröffentlichungen auch aus *einer* bereits k-anonymisierten Tabelle zusammengestellt werden.

Der sogenannte „temporale Angriff“ basiert auf dem gleichen Prinzip. Hier werden mehrere Veröffentlichungen basierend auf verschiedenen Versionen der Daten für den Angriff ausgenutzt.

Homogenitätsangriff

Beim Homogenitätsangriff wird ausgenutzt, dass alle Mitglieder einer Gruppe eine Eigenschaft teilen. Insbesondere für kleine k und große Relationen wird dies mit hoher Wahrscheinlichkeit passieren.

Beispiel

Nehmen wir an, John Doe und Thomas Muster sind beide 1983 geborene Männer mit der PLZ 33098. Die 2-anonyme Tabelle erlaubt nicht, herauszufinden, welche Zeile John Doe entspricht. Es lässt sich allerdings leicht sehen, dass sich die Diagnose von John Doe herausfinden lässt: er hat eine Depression. Dazu muss die Zeile von John Doe nicht identifiziert werden; es ist egal, ob er der erste oder der zweite der 1983 geborenen Männer in der Tabelle ist – beide haben Depression.

Geburtsjahr	PLZ	Geschlecht	Diagnose
1982	33098	Männlich	Migräne
1982	33098	Männlich	Erkältung
1983	33098	Männlich	Depression
1983	33098	Männlich	Depression

Angriff mit Hintergrundwissen

Auch Hintergrundwissen kann genutzt werden, um eine Re-Identifizierung vorzunehmen. So kann etwa der statistische Zusammenhang verschiedener Attribute betrachtet werden.

Beispiel

In der 2-anonymen Tabelle ist die Attributkombination (Geburtsjahr, PLZ, Geschlecht) der Quasi-Identifikator. Das Attribut Geschlecht wurde generalisiert.

Geburtsjahr	PLZ	Geschlecht	Diagnose
1982	33098	*	Migräne
1983	33098	*	Schwangerschaftsbeschwerden
1982	33098	*	Rheuma
1983	33098	*	Depression

Der Algorithmus, der k-Anonymität herstellt, sieht nur, dass k-Anonymität erreicht ist. Dass Schwangerschaftsbeschwerden nur bei Frauen auftreten und somit das Geschlecht der Person in der zweiten Zeile doch bekannt ist, wird nicht automatisch erkannt. Das Beispiel ist ein Extrembeispiel – es gibt auch Fälle, in denen nur eine Wahrscheinlichkeitsaussage möglich ist (z. B. ist Brustkrebs bei Männern selten, kommt aber vor).

3.2.3 l-Diversität

l-Diversität stellt eine Weiterentwicklung der k-Anonymität dar. Dabei wird eine Unterteilung der Datenbank-Relation in nicht sensible und ein sensibles Attribut vorgenommen. Das sensible Attribut ist dabei nicht Teil des Quasi-Identifikators. Man betrachtet nun Blöcke mit identischen Attributwerten des Quasi-Identifikators. Wenn in jedem Block mindestens *l* verschiedene Werte des sensiblen Attributs vorhanden sind, ist *l-Diversität* gegeben.

Wir können an dieser Stelle festhalten, dass k-Anonymität ein fundiertes Konzept zur Beurteilung der Anonymität veröffentlichter Daten ist. Das Konzept hat große Aufmerksamkeit erfahren, es gab zahlreiche Veröffentlichungen (z. B. Algorithmen zur Herstellung von k-Anonymität). Allerdings ist k-Anonymität in den meisten Fällen in der

Praxis nicht ausreichend. Aus diesem Grund beschäftigen wir uns als nächstes mit einem weitergehenden Ansatz zur Anonymisierung von Daten.

3.3 Differential Privacy

Eines der wesentlichen Probleme der k -Anonymität und verwandter Ansätze ist, dass Angreifer Hintergrundwissen haben können. Wenn man also ein Modell schaffen will, in dem die Anonymität gegen alle denkbaren Angreifer gewahrt werden kann, muss man auch alles denkbare Hintergrundwissen berücksichtigen – im obigen Beispiel etwa, dass Männer keine Schwangerschaftsbeschwerden haben. Die Bandbreite denkbaren Hintergrundwissens ist aber sehr groß, so dass eine vollständige Modellierung in praktischen Anwendungen nur selten möglich sein wird.

Aus diesem Grund wurde das alternative Konzept der Differential Privacy entwickelt. Es ist für Datenbanken gedacht, die für statistische Zwecke benötigt werden (also z. B. in der Forschung). Man betrachtet hier nicht nur die Datenbank als solche. Vielmehr geht es darum, dass eine Funktion κ Anfragen an die Datenbank beantwortet und dabei sicherstellt, dass der Datenschutz nicht verletzt wird. Dazu können Daten zufällig verändert oder nur partiell freigegeben werden.

Die Grundidee ist nun, dass die personenbezogenen Daten einer einzelnen Person² keinen Unterschied bei der Antwort herbeiführen dürfen. Wenn das Ergebnis einer Anfrage an die Datenbank unabhängig davon ist, ob die Daten der Person enthalten sind oder nicht, dann ist der Datenschutz dieser Person gewährleistet. Ob ein Angreifer Hintergrundwissen hat oder nicht, ist unerheblich.

Natürlich ist es nicht möglich, zu garantieren, dass die Daten einer Person in keinem Fall einen Unterschied für das Ergebnis der Anfrage ausmachen. Differential Privacy betrachtet nur statistische Garantien. Die gängige Definition von Differential Privacy geht auf CYNTHIA DWORK [1] zurück:

► **Definition** Eine randomisierte Funktion κ bietet ϵ -differential privacy, wenn

- für alle Datensätze D_1 und D_2 , die sich nur in höchstens einem Element unterscheiden (also einer der beiden Datensätze höchstens ein Element mehr enthält und die anderen identisch sind)
- für alle Teilmengen S des Wertebereichs W von κ (also $S \subset W(\kappa)$)

gilt:

$$P[\kappa(D_1) \in S] \leq e^\epsilon \times P[\kappa(D_2) \in S]$$

²In der Modellierung abstrahiert man von den „Daten über eine Person“ und betrachtet Tupel bzw. Zeilen in einer Tabelle.

Die Wahrscheinlichkeit, dass die Funktion einen bestimmten Wert annimmt, soll also (für alle möglichen Werte) zwischen den beiden Datensätzen nur geringfügig unterschiedlich sein. Lassen Sie sich nicht dadurch täuschen, dass in der Definition nur „kleiner oder gleich“ gefordert ist: Die Wahrscheinlichkeit für eine bestimmte Ausgabe soll bei der Eingabe D_1 tatsächlich kleiner (oder gleich) sein als bei der Eingabe D_2 , aber das gilt für alle D_1 und D_2 , die die geforderte Bedingung erfüllen. D_1 und D_2 sind also austauschbar. Für S können beliebige Teilmengen des Wertebereichs eingesetzt werden, also auch einzelne Werte.

Beispiel

Betrachten Sie die Tabelle von Seite 32! Nehmen wir an, die Funktion κ erlaubt es, die Tupel zu zählen, die eine bestimmte Bedingung erfüllen – etwa die Anzahl der Tupel, bei denen das Attribut „Geschlecht“ einen bestimmten Wert hat. Wird κ so definiert, ist Differential Privacy zunächst nicht erreicht: Sobald man beispielsweise Daten über einen weiteren Mann in die Relation einfügt, ändert sich der Funktionswert bei Anfragen, die die Anzahl von Männern in der Datenbank feststellen sollen.

Wenn solche Anfragen erlaubt werden sollen, kann also nicht deterministisch die korrekte Anzahl zurückgegeben werden. Erlaubt man, dass ein Mann mit einer gewissen Wahrscheinlichkeit nicht als Mann gezählt wird, lässt sich dies erreichen: Werden Daten über einen weiteren Mann aufgenommen, lässt sich die Wahrscheinlichkeit berechnen, dass dies nichts am Ergebnis der Zählfrage ändert. Man spricht von einem „Randomized Response“-Mechanismus. Das Konzept wird auch bei Befragungen im Rahmen sozialwissenschaftlicher Studien verwendet, um Rückschlüsse auf einzelne Personen möglichst zu verhindern.

Wie Sie erkennen können, hängt die Definition von einem Parameter ϵ ab. Anders als der Parameter k der k -Anonymität hat ϵ keine intuitive Bedeutung für konkrete Anwendungen. Je größer ϵ ist, desto schwächer ist die Garantie für den Datenschutz. Gleichzeitig ist anzunehmen, dass die meisten Anwendungen von einem großen ϵ profitieren, da genauere Ausgaben möglich sind. In der Literatur werden teils kleine (z. B. $\epsilon = 0,01$), teils aber auch sehr große ϵ -Werte (auch $\epsilon > 1$) verwendet. Dies zeigt, dass die Wahl des Parameters anwendungsabhängig ist.

Die Beschreibung des Modells, in dem Differential Privacy eingesetzt werden kann, legt nahe, lediglich an *interaktive* Mechanismen zu denken. In der Tat ist die Wahl der Funktion κ in diesem Fall flexibler. Beispielsweise können Anfragen dort so lange mit hoher Genauigkeit beantwortet werden, bis weitere Ausgaben die Definition von Differential Privacy verletzen würden (da der Angreifer aus ihrer Kombination Rückschlüsse auf einzelne enthaltene Elemente ziehen könnte). Es ist aber auch möglich, eine vollständige Datenbank offenzulegen. Diese muss aber mit anderen Mechanismen geschützt werden – üblicherweise durch das „Verrauschen“ der Daten, beispielsweise, indem zusätzliche Datensätze eingefügt werden.

3.4 Anonymisierung in der Praxis

In der europäischen Datenschutz-Richtlinie bzw. dem Bundesdatenschutzgesetz (BDSG) werden keine konkreten Maßnahmen zur Anonymisierung genannt. Es werden auch keine Anonymitätsmaße angegeben, bei deren Erfüllung von einer ausreichend durchgeführten Anonymisierung ausgegangen werden darf. *Anonymisierung* wird in § 3 Abs. 6 Bundesdatenschutzgesetz (BDSG) lediglich als „Verändern personenbezogener Daten derart [definiert], dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können.“

Der US-Gesetzgeber geht bei der *Health Insurance Portability and Accountability Act (HIPPA) Privacy Rule* von 1996 dann von anonymisierten Daten aus, wenn ein akzeptierbar geringes Risiko bzw. sehr geringes Risiko der Re-Identifizierung besteht. Außerdem gibt er zwei Maßnahmen der Anonymisierung von Gesundheitsdaten vor. Zum einen ist eine sogenannte „Expert Determination“ vorgesehen, bei der ein Experte mithilfe von statistischen und wissenschaftlichen Methoden das Risiko einer möglichen Re-Identifizierung auf Basis der anonymisierten Daten untersucht. Bei der *Safe Harbour-Methode*³ werden 18 Typen von identifizierenden Daten bzgl. einer Person, seiner Verwandten, seines Arbeitgebers etc. genannt, die entfernt werden müssen, um von anonymisierten Daten auszugehen.

Einige Datenschützer vertreten zudem die Auffassung, dass *Anonymisierung* eine *Verarbeitung* personenbezogener Daten darstellt. Demnach wäre nach § 4a Abs. 1 BDSG eine Einwilligung der Betroffenen erforderlich.

3.5 Fazit

Eine sehr gute Übersicht über die Vielzahl an unterschiedlichen Anonymisierungsverfahren – mit einem Schwerpunkt auf Gesundheitsdaten – findet sich bei Gkoulalas-Divanis et al. [3]. Im Allgemeinen lässt sich sagen, dass *richtige* Anonymisierung von Daten wohl eines der schwierigsten Probleme im Datenschutz darstellt. Es gibt nicht *das* Verfahren, das möglichst gut brauchbare, anonymisierte Daten liefert. Stattdessen muss je Szenario eine Einzelfallentscheidung getroffen werden, um zu bewerten, welches Verfahren sinnvollerweise angewendet werden sollte. Die Artikel 29-Datenschutzgruppe hat mit ihrer „Opinion 05/2014 on Anonymization Techniques“ [7] ein Dokument vorgelegt, in dem unterschiedliche Anonymisierungsverfahren bewertet werden. Einige Forscher kommen zum Schluss, dass man sich von den starken Anonymitäts-Eigenschaften, die diese Verfahren versprechen, lösen muss, da diese nicht praktikabel sind. Anstatt davon auszugehen, dass *überhaupt kein Risiko* bestehen darf, dass Daten de-anonymisiert

³Diese Methode hat nichts mit dem vom EuGH im Jahr 2015 gekippten Abkommen zu tun, das die USA als „sicheren Hafen“ für personenbezogene Daten von EU-Bürgern ansah.

werden können, sollte nach deren Auffassung Anonymisierung stattdessen als Risiko-Management-Aufgabe verstanden werden, bei der sowohl die Folgen für die Betroffenen bei einer De-Anonymisierung als auch Annahmen über Motive der „Angreifer“ (also derjenigen, die versuchen, aufgrund einer De-Anonymisierung der Daten Gewinn zu schlagen) in Betracht gezogen werden [2]. Gerade im Hinblick auf Themen wie *Big Data* – bei denen zahlreiche personenbezogene Daten im Spiel sind – erscheint es unerlässlich, dass weiter an verbesserten Anonymisierungsverfahren geforscht wird und praktikable Tools entwickelt werden. Gerade bei *Big Data* zeigt sich jedoch, dass auch Anonymisierung kein „Allheilmittel“ darstellt. So kann es sein, dass aufgrund *anonymisierter* Daten ein Modell gebildet wird, das später auf Personen angewendet wird, die gar nicht in dem ursprünglichen Datensatz enthalten waren. Aus datenschutzrechtlicher Sicht mag in solchen Fällen kein Verstoß begangen worden sein, obwohl ein Eingriff in das Persönlichkeitsrecht der Betroffenen stattfindet.

Apple hat im Juni 2016 auf der *Apple Worldwide Developers Conference (WWDC)*⁴ die Einführung von *Differential Privacy* mit dem neuen Apple-Betriebssystem *iOS 10* angekündigt. *Differential Privacy* soll Apple eine *datenschutzgerechte Analyse der Nutzungsmuster* einer Vielzahl an Nutzern ermöglichen um allgemeine Muster erkennen zu können und damit die Benutzererfahrung zu verbessern. Nutzungsmuster sollen durch das Hinzufügen von Rauschen keinen einzelnen Nutzern zugeordnet werden können. Als Anwendungsbeispiele wurden die Verbesserungen von Wort-Vorschlägen (*QuickType*) und Emoji-Vorschlägen genannt. Dieses Beispiel zeigt, dass Unternehmen erkannt haben, dass mit datenschutzfreundlicher Technik geworben werden kann. Die ersten Reaktionen in der Presse auf diesen Vorstoß von Apple waren überwiegend positiv. Für eine Bewertung ist es freilich noch zu früh, vor allem da zum jetzigen Zeitpunkt noch keine technischen Details zur konkreten Umsetzung vorliegen.

3.6 Übungsaufgaben

Aufgabe 1

Geben Sie für die dargestellte Tabelle die Attribute an, die zu einem Quasi-Identifikator gehören! Wie könnten Sie vorgehen, um *k*-Anonymität a) durch Löschen von Tupeln und b) durch Generalisierung herzustellen? (Nehmen Sie zum Beispiel $k = 3$ an).

⁴<http://www.apple.com/de/apple-events/june-2016/> (Zugegriffen am 01.08.2016)

Vorname	Nachname	Matrikelnummer	Punktzahl Datenschutz-Übung	Geschlecht	Alter	Größe
Elias	Bauer	100012	101	m	25	163
Maximilian	Becker	100008	20	m	25	175
Finn	Fischer	100004	87	m	25	165
Amelie	Fuchs	100041	90	w	35	177
Sarah	Herrmann	100036	87	w	22	169
Felix	Hoffmann	100010	95	m	23	185
Fabian	Hofmann	100021	95	m	23	196
Laura	Huber	100034	105	w	24	180
Philip	Klein	100015	165	m	24	180
Max	Koch	100013	102	m	21	172
Sofie	König	100039	109	w	27	170
Hannah	Krüger	100025	112	w	23	180
Lina	Lehmann	100033	110	w	23	164
Emilie	Maier	100032	99	w	22	165
Marie	Mayer	100035	99	w	25	165
Leonie	Meier	100030	95	w	24	167
Paul	Meyer	100006	130	m	27	200
Leon	Müller	100001	110	m	22	190
Moritz	Neumann	100018	165	m	28	163
Noah	Richter	100014	95	m	23	185
Tim	Schäfer	100011	122	m	24	188
Lena	Schmid	100026	130	w	24	168
Lucas	Schmidt	100002	95	m	23	185
Simon	Schmitt	100023	130	m	25	180
Anna	Schmitz	100029	112	w	23	165
Ben	Schneider	100003	87	m	24	163
Julian	Schröder	100017	155	m	26	177
Luka	Schulz	100009	110	m	22	190
Maja	Schulze	100040	90	w	33	170
Jan	Schwarz	100019	88	m	32	167
Louis	Wagner	100007	120	m	24	172
Jonas	Weber	100005	110	m	22	190
Lea	Werner	100027	17	w	25	182
Niclas	Wolf	100016	167	m	25	180
Alexander	Zimmermann	100022	105	m	24	195

Literatur

1. Cynthia Dwork. *Differential Privacy*. In Automata, Languages and Programming. Volume 4052 of the series Lecture Notes in Computer Science, pages 1–12. Springer, Berlin, Heidelberg, 2006.
2. Khaled El Emam and Cecilia Álvarez. A critical appraisal of the Article 29 Working Party Opinion 05/2014 on data anonymization techniques. *International Data Privacy Law*, 2014.
3. Aris Gkoulalas-Divanis, Grigorios Loukides, and Jimeng Sun. Publishing data from electronic health records while preserving privacy: A survey of algorithms. *Journal of Biomedical Informatics*, 50:4–19, 2014. Special Issue on Informatics Methods in Medical Privacy.
4. Philippe Golle. Revisiting the uniqueness of simple demographics in the US population. In *Proceedings of the 5th ACM workshop on Privacy in electronic society*, pages 77–80. ACM, 2006.
5. Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data*, 1(1), Mar. 2007.
6. Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, SP '08, pages 111–125, Washington, DC, USA, 2008. IEEE Computer Society.
7. Article 29 Data Protection Working Party. Opinion 05/2014 on anonymisation techniques, Apr. 2014. 0829/14/EN WP216.
8. Latanya Sweeney. k-anonymity: a model for protecting privacy. *International Journal of Uncertainty*, 10(5):557–570, Oct. 2002.

Zusammenfassung

In Abschn. 2.3 haben wir mit der *Verschlüsselung* eine Maßnahme kennengelernt, um personenbezogene Daten *vertraulich* zu übertragen. Allerdings reicht Verschlüsselung alleine nicht aus, um Anonymität zu erreichen. Verschlüsselung verbirgt nicht die Tatsache, *dass* Kommunikation stattfindet – und insbesondere auch nicht, zwischen welchen Kommunikationspartnern.

In diesem Kapitel gehen wir der Frage nach, wie sich Kommunikationsbeziehungen im Netz verstecken lassen. Dazu beschäftigen wir uns zunächst in Abschn. 4.1 mit dem Thema *Verkehrsflussanalyse*. Danach lernen wir mit *Mixes* (Abschn. 4.2), *Mix-Kaskaden* (Abschn. 4.3) und schließlich *Onion Routing/Tor* (Abschn. 4.4) unterschiedliche Konzepte kennen, die Anonymität – bzw. das Verstecken von Kommunikationsbeziehungen – im Netz versprechen.

Das folgende Beispiel soll uns als Motivation für dieses Kapitel dienen.

Beispiel

Alice möchte Bob zum Geburtstag mit einem Konzertticket seiner Lieblingsband überraschen. Die Tickets bekommt sie im Online-Ticket-Shop der Band. Alice weiß, dass Bob sehr neugierig ist und hin und wieder die aufgerufenen Webseiten (von Alice) überprüft – etwa in dem er das lokale Netzwerk belauscht. Wie kann Alice den Online-Ticket-Shop besuchen, ohne dass Bob dies bemerkt?

Lernziele

Am Ende dieses Kapitels sollten Sie verstehen, wie Anonymisierungsdienste funktionieren und welche Schutzziele damit erreicht werden können.

4.1 Verkehrsflussanalyse

Bei der Verkehrsflussanalyse geht es um die Auswertung von Metadaten und den nicht verschlüsselten Informationen in übermittelten Nachrichten. Daraus sollen weitergehende Informationen abgeleitet werden. Ausgewertet werden beispielsweise die Absender- und Empfängeradressen (E-Mail-Adressen, IP-Adressen etc.), die immer im Klartext übermittelt werden – auch wenn die Nachrichten selbst verschlüsselt sind. Selbst bei der verschlüsselten Sprachtelefonie lassen sich aufgrund der Größe und Anzahl der übertragenen verschlüsselten Blöcke im zeitlichen Verlauf Rückschlüsse auf die gesprochenen Worte ziehen.

4.1.1 Angreiferklassifikation

Zuallererst müssen wir untersuchen, mit welcher Art von Angreifern wir es zu tun haben, d. h. welche Fähigkeiten der Angreifer besitzt, um eine Verkehrsflussanalyse durchzuführen. Wir gehen in diesem Kapitel von einem *passiven Angreifer* aus, der Kommunikation nur mithört. Wir beschäftigen uns in diesem Kapitel nicht mit Angreifern, die versuchen oder in der Lage sind, ein zugrundeliegendes Verschlüsselungsverfahren zu brechen. Wir unterscheiden zwei Angreifer-Typen mit jeweils unterschiedlichen Fähigkeiten: *Globaler Angreifer* und *Lokaler Angreifer*. Weitergehende Überlegungen zur Angreiferklassifikation finden sich bei PANCHENKO et al. [5].

Globaler Angreifer

Einem globalen Angreifer wird die Fähigkeit unterstellt, sämtliche Kommunikation im Netz zu sehen. Auch wenn diese sehr starke Annahme in der Praxis eher unrealistisch ist, so gibt es eine Reihe von abgeschwächten Varianten dieses Angreifers. In Frage kommen beispielsweise Angreifer, die viele oder alle zentralen Router in einem Land kontrollieren und damit alle über diese Router ausgetauschten Daten der Bürger dieses Landes einsehen können. Polizeibehörden und Geheimdienste, die sowohl an verschiedenen Stellen mithören können als auch auf Logfiles zugreifen, können als abgeschwächte globale Angreifer angesehen werden.

Lokaler Angreifer

Die Fähigkeiten von lokalen Angreifern gegenüber globalen Angreifern sind hingegen deutlich eingeschränkt. Sie haben in der Regel keine Möglichkeit das gesamte Netz zu überwachen. Nichtsdestotrotz können sie die Kommunikationsbeziehungen einzelner Nutzer (in ihrem Umfeld) analysieren. Aus Sicht eines Nutzers sind die folgenden lokalen Angreifer relevant:

- Kommunikationspartner (z. B. Webserver-Betreiber),
- Angreifer im eigenen lokalen Netz (z. B. Systemadministrator),
- Angreifer auf einem einzelnen Zwischensystem (z. B. Internet Service Provider).

4.1.2 Beispiel: Ablauf der Ticketbestellung

Wenden wir uns nun wieder unserem Beispiel von vorhin zu. Zunächst können wir sagen, dass Alice mit Bob ein lokaler Angreifer gegenübersteht: Der Router steht unter der Kontrolle von Bob. Um zu verstehen, wie Bob nun herausfinden kann, welches Geschenk er von Alice erwarten kann, müssen wir uns den Ablauf der Ticket-Bestellung im Detail ansehen. Alice gibt die URL des Ticket-Online-Shops in ihren Browser ein. Was passiert aus der Sicht von *Alice*?

- 1. Schritt:** DNS-Anfrage nach der eingegebenen URL
- 2. Schritt:** TLS-gesicherte Kommunikation zwischen Alices Browser und dem Shop-Webserver

Was passiert aus der Sicht von *Bob*?

- 1. Schritt:** DNS sieht keine Verschlüsselung der Anfrage bzw. Antwort vor. Bob sieht beide Nachrichten und kann Rückschlüsse ziehen.

Um eine Analyse der DNS-Anfrage auszuschließen, könnte sich Alice die IP-Adresse des Shops von einem Freund durchgeben lassen. Sie würde dann den Webserver direkt über die IP-Adresse ansprechen und TLS-gesichert mit diesem kommunizieren.

- 2. Schritt:** Bob könnte zwar den Inhalt der Nachrichten nicht mitlesen, würde aber die IP-Adresse des Webserver erfahren.

Die IP-Adresse reicht in diesem Beispiel für Bob aus, um Rückschlüsse auf sein Geschenk ziehen zu können: Er weiß nun, dass Alice den Online-Ticket-Shop seiner Lieblingsband aufgerufen hat. Ob sie tatsächlich Tickets gekauft hat, erfährt er aufgrund der TLS-gesicherten Kommunikation allerdings nicht.

- **Metadaten** Dieses Beispiel macht deutlich, dass nicht nur Inhaltsdaten bei der Kommunikation schützenswert sind, sondern auch die Metadaten, die Auskunft über den Zeitpunkt, den Ort und den Kommunikationspartner geben.

Die Bedeutung von Metadaten wurde im Zuge der Enthüllungen der NSA-Überwachungsmaßnahmen im *Guardian*¹ eindrucksvoll dargelegt. Demnach erlaubt gerade die Strukturiertheit und Maschinenlesbarkeit von Metadaten eine sehr gute Auswertung – im Gegensatz etwa zum Abhören von Telefongesprächen. Außerdem sagen Browsing-Logs und Telefon-Logs sehr viel darüber aus, wer welche Rolle in der Gesellschaft hat und wo die Interessen der Personen liegen.

¹<https://www.theguardian.com/technology/2013/jun/21/nsa-surveillance-metadata-content-obama>
(Zugegriffen am 01.08.2016)

Eindrucksvoll lässt sich im Internet² etwa ein halbes Jahr im Leben von MALTE SPITZ, einem Politiker der die Herausgabe seiner Telefondaten von der Deutschen Telekom eingeklagt hatte, rekonstruieren – allein aufgrund der Metadaten (die auch Teil der Vorratsdatenspeicherung sind, wie wir in Abschn. 11.4.3 sehen werden).

Laut Aussage von MICHAEL HAYDEN, ehemaliger Chef der NSA und der CIA, werden Menschen auf Basis von Metadaten getötet.³

4.1.3 Beispiel: Mögliche Gegenmaßnahme

Eine mögliche Gegenmaßnahme zur Verkehrsflussanalyse besteht in der Verwendung eines *Virtual Private Networks (VPNs)*. Dabei leitet Alice ihre gesamte Kommunikation über das VPN-Netz, beispielsweise wie in Abb. 4.1 dargestellt, über ihr Uni-Netz um. Alice sendet ihre Anfrage nicht direkt an den Shop-Webserver sondern – unter Verwendung des Tunnelmodus und IPsec-ESP als VPN-Protokoll – an das VPN-Gateway ihrer Universität. Das ESP-Protokoll sorgt dafür, dass die IP-Kommunikation zwischen Alice und dem VPN-Gateway authentifiziert und verschlüsselt erfolgt. Der Tunnelmodus erlaubt das Verstecken der Adressinformationen. Bob sieht also weder den Inhalt der Kommunikation, noch die IP-Adresse des angefragten Webservers (diese verbirgt sich in dem in der Abbildung als verschlüsselt dargestellten IP-Kopf). Bob sieht nur, dass Alice mit ihrem Uni-Netz kommuniziert. Gegenüber dem Shop-Webserver tritt das VPN-Gateway sozusagen als „Proxy“ für Alice auf, d. h. der Webserver sieht nicht die Heim-IP-Adresse von Alice sondern nur die IP-Adresse, die Alice für ihre VPN-Sitzung

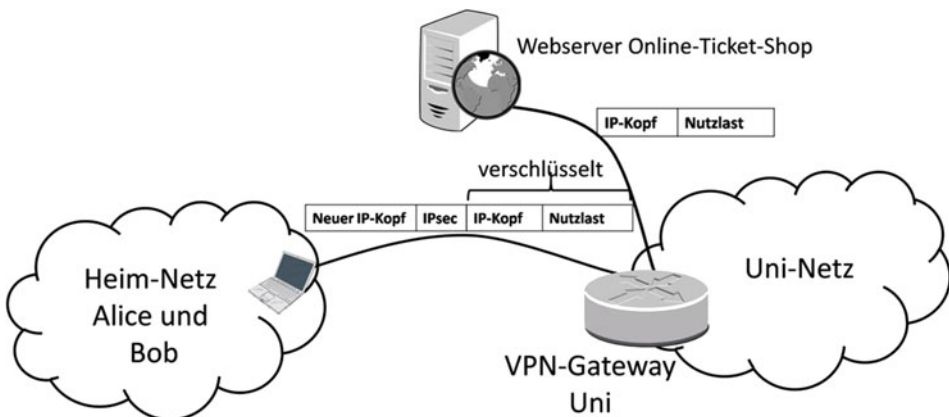


Abb. 4.1 Schutz vor Verkehrsflussanalysen mittels VPN

²<http://www.zeit.de/datenschutz/malte-spitz-vorratsdaten> (Zugegriffen am 01.08.2016)

³<http://www.heise.de/newsticker/meldung/Ex-NSA-Chef-Wir-toeten-auf-Basis-von-Metadaten-2187510.html> (Zugegriffen am 01.08.2016)

von ihrer Universität zugewiesen bekommen hat. Die Antwortpakete des Webserver gehen auf dem gleichen Weg (über das Uni-Netz) zurück zu Alice.

Auf zwei verbleibende Probleme bei dieser Lösung muss noch hingewiesen werden. Zum einen sind Rückschlüsse auf den Inhalt der Kommunikation durch Analysen der Paketgrößen und der Anzahl möglich. Einer derartigen Analyse lässt sich durch Einfügen von Dummy-Paketen entgegenwirken. Zum anderen ist die dargestellte Lösung nur eine Problemverlagerung: Bob sieht die Kommunikationsbeziehung zwar nicht mehr, aber nun kann eventuell ein Angreifer im Uni-Netz eine Verkehrsflussanalyse durchführen. Im Allgemeinen ist es fraglich, ob einem einzelnen VPN-Anbieter vertraut werden kann, um sich gegen Verkehrsflussanalysen zu schützen.

Eine Möglichkeit der besseren Absicherung liegt in der Verkettung mehrerer Zwischenstationen. Dies lässt sich grundsätzlich auch mit IPsec im Tunnelmodus erreichen, wird in der Praxis aus Gründen der Performance allerdings so gut wie nicht durchgeführt. In den nächsten Abschnitten werden wir ein derartiges Verkettungskonzept, das bei Mix-Kaskaden und Onion Routing/Tor zum Einsatz kommt, genauer untersuchen.

4.2 Mixes

Den Grundstein zur anonymen Kommunikation in Netzen legte DAVID CHAUM 1981 mit seiner Arbeit „Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms“ [2]. Darin schlägt er das Konzept eines *Mixes* vor, der eingehende E-Mails sammelt und dann stapelweise verschickt. Der Mix soll die Kommunikationsbeziehungen zwischen Absendern und Empfängern von E-Mails verstecken. Das Konzept ist in Abb. 4.2 dargestellt.

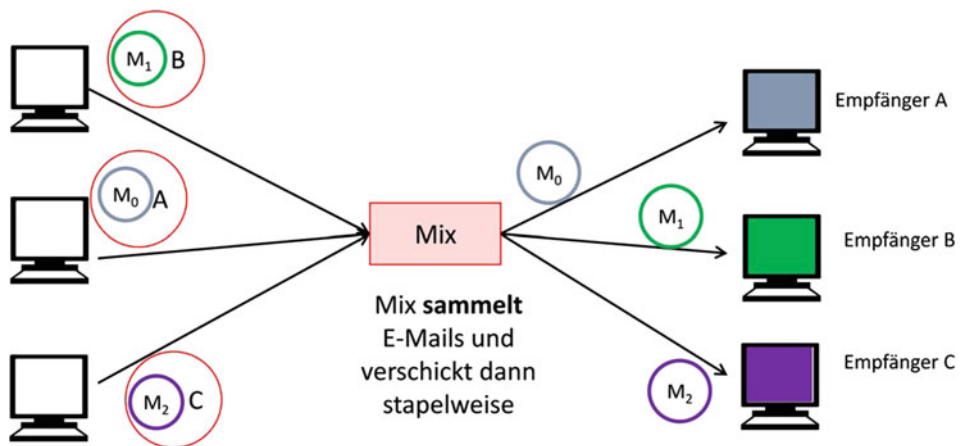


Abb. 4.2 Verstecken der Kommunikationsbeziehungen mittels Mix

Die bunten Kreise um eine Nachricht in der Abbildung bedeuten, dass die Verschlüsselung mit dem öffentlichen Schlüssel des Empfängers erfolgt, der diese Farbe trägt.

4.2.1 Verfahren

Zunächst werden alle Nachrichten mit dem öffentlichen Schlüssel des Empfängers verschlüsselt. Anschließend werden die verschlüsselten Nachrichten mit dem öffentlichen Schlüssel des Mix verschlüsselt. In die Verschlüsselung fließt jeweils noch eine Zufallszahl mit ein, d. h. wenn E die Verschlüsselungsfunktion ist, dann ist $E(r||M)$ die verschlüsselte Nachricht, wobei r zufällig gewählt wird und M die zu verschlüsselnde Nachricht bezeichnet. So wird verhindert, dass zwei identische Nachrichten einander zugeordnet werden können, wie es bei der Verwendung der klassischen RSA-Verschlüsselung der Fall wäre.

Die Ausgabe des Mixes kann öffentlich zugänglich sein. Chaum sieht sogar vor, dass der Mix seine Ausgabe (die nicht nur aus einer einzelnen E-Mail, sondern einem „Stapel“ besteht) signiert und jeweils öffentlich zur Verfügung stellt, damit Absender prüfen können, ob ihre Nachricht dabei ist. Dieses Vorgehen macht die Verschlüsselung des Nachrichteninhalts (für den jeweiligen Empfänger) notwendig.

4.2.2 Analyse

Auf den ersten Blick bietet der Mix einen guten Schutz der Kommunikationsbeziehungen, d. h. die Information, wer mit wem kommuniziert, wird vor Angreifern verborgen. Die Annahme ist, dass der Angreifer nicht den Mix selbst kontrolliert. Wenn der Mix nicht vertrauenswürdig ist, besteht überhaupt kein Schutz. Der Mix:

- kennt keine Kommunikationsinhalte,
- kennt die Kommunikationsbeziehungen.

Ein globaler Angreifer, der alle Nachrichten mithören kann – wie wir ihn in Abschn. 4.1.1 charakterisiert haben:

- kennt keine Kommunikationsinhalte,
- kennt keine Kommunikationsbeziehungen.

Bei genauerer Betrachtung des Mix-Konzepts treten jedoch einige Probleme ans Tageslicht. Zum einen gibt es eine – aus Sicherheitsgründen erforderliche – Verzögerung der Nachrichten durch die Sammelfunktion des Mixes. Gäbe es diese Sammelfunktion nicht, so könnte ein globaler Angreifer die Kommunikationsbeziehungen zwischen Absendern und Empfängern einfach ermitteln indem er die beim Mix eingehende Nachricht mit der vom Mix ausgehenden Nachricht in Verbindung bringt. Durch die Verzögerung in der

Kommunikation eignet sich dieses Konzept nicht für Dienste die zeitkritisch sind, wie bspw. das Web.

Ein weiterer möglicher Angriffspunkt besteht darin, Rückschlüsse aus der Größe von E-Mails zu ziehen. Eine Gegenmaßnahme könnte hier im „Auffüllen“ der E-Mails auf Einheitsgröße bestehen – dies erhöht natürlich den Overhead.

Zudem können Rückschlüsse bei der Beobachtung über mehrere Perioden und viele E-Mails gezogen werden, indem eine Korrelation zwischen der Anwesenheit von Absendern und Empfängern bei langlebigen Kommunikationsbeziehungen durchgeführt wird.

Ist hingegen der Empfänger der Nachrichten der Angreifer, so besteht kein Schutz vor Analyse der übertragenen Daten in der Anwendungsschicht (z. B. im E-Mail-Header).

Der sogenannte $(n - 1)$ -Angriff ist ein aktiver Angriff, also ein Angriff, der über das Lauschen hinausgeht und in die Kommunikation eingreift. Aufgrund der Sammelfunktion des Mixes sammelt dieser zunächst n Nachrichten, bevor er sie weiterschickt. Diese Tatsache macht sich der Angreifer zunutze: Nach der ersten echten Nachricht (des Opfers) schickt er $(n - 1)$ Nachrichten an einen bekannten Empfänger. Damit ist er in der Lage, die Kommunikationsbeziehung des Opfers offenzulegen, indem er prüft, an welchen Empfänger die erste Nachricht (des Opfers) weitergeleitet wurde.

4.3 Mix-Kaskaden

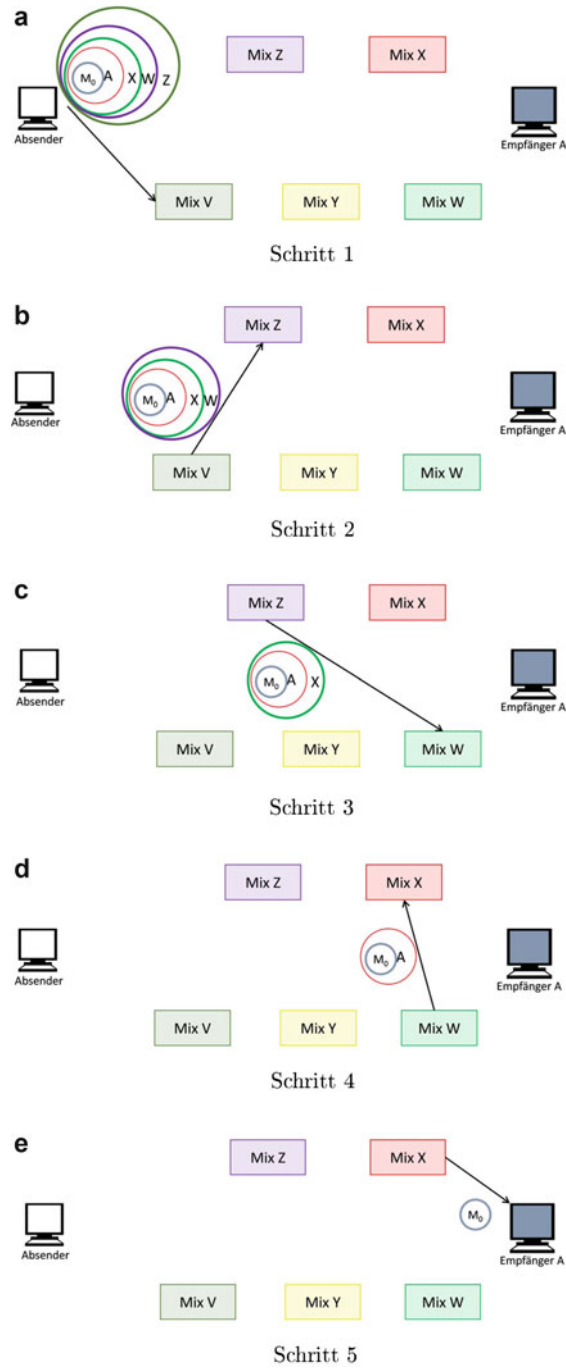
Zuvor haben wir gesehen, dass wir die *Vertrauenswürdigkeit des Mixes* unterstellen müssen, um die gewünschte Anonymität zu erreichen. Die Fragen, die wir uns stellen können sind: Welchem Betreiber eines Mixes würden wir vollständig vertrauen? Würden wir einem Mix vertrauen, der unter staatlicher Kontrolle steht? Oder doch lieber einer NGO oder einer Universität? Vermutlich müssten wir lange suchen, um eine solch voll vertrauenswürdige Stelle zu finden. Diese Überlegung führt uns zum Ansatz der Mix-Kaskaden. Die Idee ist, dass einfach mehrere Mixes (die von unterschiedlichen Stellen betrieben werden) hintereinandergeschaltet werden. Solange *mindestens ein Mix vertrauenswürdig* ist, bleibt die Kommunikationsbeziehung zwischen Absender und Empfänger verborgen.

4.3.1 Verfahren

Das Verfahren ist in Abb. 4.3 dargestellt. Wie zuvor, bedeutet ein bunter Kreis um eine Nachricht, dass die Verschlüsselung mit dem öffentlichen Schlüssel des Empfängers erfolgt, der diese Farbe trägt.

Der Absender legt die zu verwendenden Mixes und die gewünschte Reihenfolge zu Beginn fest. In Abb. 4.3a sehen wir, dass der Absender die Nachricht M_0 zunächst mit dem öffentlichen Schlüssel für den Empfänger A verschlüsselt (dem „blauen Schlüssel“). Die verschlüsselte Nachricht verschlüsselt er im nächsten Schritt gemeinsam mit der

Abb. 4.3 Mix-Kaskade



Empfängeradresse „A“ mit dem öffentlichen Schlüssel von Mix *X*. Dieser Mix wird später der letzte Mix auf dem Weg zum Empfänger *A* sein. Danach verschlüsselt der Absender das Resultat des vorherigen Schrittes inklusive der Adressinformation „X“ mit dem öffentlichen Schlüssel von Mix *W*. Diesen Vorgang wiederholt der Absender solange, bis er die gewünschte Anzahl an Mixes in seinen Pfad aufgenommen hat. Man kann sich diesen Aufbau wie eine Zwiebel vorstellen, die aus mehreren verschlüsselten Schichten (für die jeweiligen Mixes) besteht. In unserem Beispiel verschlüsselt der Absender im letzten Schritt für Mix *V* und sendet diesem Mix die „Zwiebel“ (Onion) zu.

Der Mix *V* (als erster Mix in der Kaskade) entschlüsselt die Nachricht und entnimmt die Adressinformation „Z“ sowie die innere (verschlüsselte) Nachricht, die für *Z* bestimmt ist. *V* entfernt also eine Zwiebelschale und leitet die Zwiebel weiter an *Z* (Abb. 4.3b).

Die weiteren Schritte funktionieren analog. So kann *Z* nach Entschlüsselung die Adressinformation „W“ lesen und leitet die Nachricht entsprechend weiter (Abb. 4.3c). *W* verfährt ebenso (Abb. 4.3d). *X* erfährt im letzten Schritt (Abb. 4.3e) die Adressinformation „A“ des endgültigen Empfängers und leitet die Nachricht an diesen weiter. Die Nachricht ist im Beispiel immer noch mit dem öffentlichen Schlüssel von *A* verschlüsselt, so dass *X* den Inhalt der Kommunikation nicht lesen kann.

4.3.2 Analyse

Als größter Vorteil der Mix-Kaskaden gegenüber einem einfachen Mix gilt die größere Toleranz gegenüber Angreifern, die die Mixe kontrollieren. Solange mindestens ein Mix vertrauenswürdig ist, bleibt die Beziehung zwischen Absender und Empfänger verborgen.

Auf der anderen Seite gibt es auch eine Reihe von Nachteilen:

- Noch größere Verzögerung als bei einem einzelnen Mix,
- hoher Rechenaufwand für mehrfache Ver-/Entschlüsselungen,
- großer Kommunikations-Overhead,
- erhöhte Wahrscheinlichkeit für einen Ausfall mit wachsender Länge der Kaskade.

4.3.3 Antwort-Nachrichten

Bisher sind wir davon ausgegangen, dass ein Absender eine Nachricht an einen Empfänger schickt und damit die Kommunikation beendet ist. In der Praxis möchte man natürlich auch einen Rückweg haben: der Empfänger soll dem Absender auch antworten können. Der Knackpunkt dabei ist, dass der Empfänger die Absenderadresse nicht sieht. An wen soll er also die Antwort senden?

Hierzu gibt es zwei mögliche Lösungsansätze: die Zustandshaltung auf den Mixes und die Festlegung des Antwortpfads durch den Absender.

Zustandshaltung auf Mixes

Eine Möglichkeit besteht in der Einführung von Identifikatoren für Kommunikationsbeziehungen. Die Mixes merken sich den jeweiligen Vorgänger-Mix für jede Kommunikationsbeziehung. Für das Beispiel von vorhin bedeutet dies:

- *V* merkt sich, dass die Ausgangsnachricht vom Absender kommt,
- *Z* merkt sich, dass sein Vorgänger *V* ist,
- *W* merkt sich, dass sein Vorgänger *Z* ist,
- *X* merkt sich, dass sein Vorgänger *W* ist,
- *A* merkt sich, dass er die Nachricht von *X* erhalten hat,
- *A* schickt die Antwort an *X*, *X* an *W*, *W* an *Z*, *Z* an *V* und *V* schließlich an den Absender.

Der Vorteil dieser Lösung ist, dass damit kein zusätzlicher Kommunikations-Overhead entsteht.

Auf der anderen Seite steigt mit der Wartezeit auf eine Antwort die Wahrscheinlichkeit für einen Ausfall bzw. einen Zustandsverlust eines Mixes in der Kaskade. Bei E-Mails kann eine Antwort bspw. erst nach Tagen kommen.

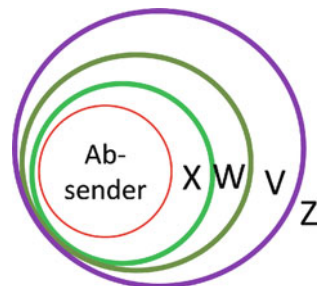
Durch die Zustandshaltung auf den Mixes werden diese prinzipiell auch anfällig für Denial-of-Service-Angriffe. Bei diesen Angriffen versuchen die Angreifer, durch eine Erschöpfung der Ressourcen der Opfer (in diesem Fall der Mixes), diese in der Erbringung ihrer Dienste (für Nutzer) zu behindern.

Festlegung des Antwortpfads durch Absender

Eine zweite Möglichkeit besteht in der Verwendung eines sogenannten „Reply Onions“ („Onion“ bezieht sich hier wieder auf die zuvor beschriebene Zwiebelstruktur). Der Reply Onion, über den der Absender den Antwortpfad festlegt, ist eine Datenstruktur (wie in Abb. 4.4 dargestellt), die der Absender an den Empfänger übermittelt.

Der Reply Onion wird wiederum durch die Mix-Kaskade geroutet – analog zur Ursprungsnachricht. Die verschiedenen Ringe stehen wieder für die Verschlüsselung. Zuerst wird mit dem öffentlichen Schlüssel des letzten Zwischenknotens auf dem Rückweg verschlüsselt, dann mit dem des vorletzten usw. Ganz innen steht die Adresse des ursprünglichen Absenders. Der Antwortpfad kann mit dem Pfad der Ursprungsnachricht (in umgekehrter Richtung durchlaufen) übereinstimmen, muss es aber nicht.

Abb. 4.4 Reply Onion



In dem Beispiel wird die Antwort vom ursprünglichen Empfänger (A), gemeinsam mit dem Reply Onion erst an Z geschickt. Z entnimmt eine Schicht des Reply Onions und entnimmt die Adressinformation „V“. Er leitet beides an V weiter. V verfährt analog, W ebenso. X entschlüsselt schließlich die letzte Schicht des Reply Onions und entnimmt die Adressinformation des Absenders.

Der ursprüngliche Absender sollte einen temporären öffentlichen Schlüssel mitschicken, den der Empfänger verwenden kann, um die Antwort zu verschlüsseln. Als Ergänzung ist empfehlenswert, dass jeder Mix auf dem Weg die Nachricht mit dem öffentlichen Schlüssel seines Nachfolgers im Antwortpfad verschlüsselt.

Das Problem der ausfallenden Knoten besteht auch bei diesem Ansatz, allerdings ist der „Zustandsverlust“ von Zwischenknoten unproblematisch, solange Zwischenknoten ihre Schlüssel noch kennen.

4.4 Onion Routing / Tor

Onion Routing, das auf dem Konzept der Mixes (hier „Onion Router“ genannt) basiert, wurde 1996 von GOLDSCHLAG et al. [4] als Architektur zum Schutz IP-basierter Kommunikation vor Verkehrsanalyse vorgestellt. Die Veröffentlichung weist explizit darauf hin, dass Onion Routing nicht für den Schutz der Anonymität von Absender und Empfänger gedacht ist, sondern lediglich für den Schutz vor Verkehrsanalyse. Im praktischen Einsatz steht die Anonymität aber oft im Vordergrund. Ein wesentliches Entwurfsziel bestand darin, dass geringe Latenzen eine Priorität gegenüber hoher Sicherheit haben.

Im Jahre 2004 erfolgte die Weiterentwicklung zu *Tor* [3], der bekanntesten Umsetzung von Onion Routing – die wir uns im weiteren Verlauf genauer ansehen werden. Bei Tor steht auch die Praktikabilität im Vordergrund. Eine weitere Entwurfsentscheidung besteht in der Anforderung nach *Perfect Forward Secrecy*: Die Verschlüsselung einzelner Sitzungen soll unabhängig von einem Langzeitgeheimnis sein; eine Kompromittierung eines Sitzungsschlüssels soll somit auch nur einen kleinen Ausschnitt der Kommunikation offenlegen.

4.4.1 Grundkonzept von Tor

Bei Tor handelt es sich um ein vollvermaschtes Netz von Onion Routern (auch „Tor Relays“ genannt), die sich untereinander „kennen“ – selbst wenn keine permanenten Verbindungen zwischen ihnen bestehen. Verbindungen können jederzeit aufgebaut werden: Wenn ein Nutzer (bzw. der Proxy des Nutzers) eine Verbindung mit einem Kommunikationspartner aufbauen möchte, sucht er sich einen Pfad über mehrere Onion Router aus und baut diese auf. Eine derart aufgebaute Verbindung wird (*Virtual*) *Circuit* genannt. Dabei speichert jeder Onion Router im Pfad für die Dauer der Verbindung einen Zustand: er merkt sich jeweils seinen Vorgänger und Nachfolger.

4.4.2 Tor-Zellen

Bevor wir uns den Aufbau eines Circuits genauer ansehen, müssen wir uns noch mit den *Tor-Zellen* vertraut machen. Eine Zelle ist bei Tor die zu transportierende Basisdateneinheit – diese ist 512 *Byte* groß. Man unterscheidet zwischen zwei Arten von Zellen: *Kontrollzellen* und *Relay-Zellen*. Die Kontrollzellen werden durch die empfangenden Onion Router immer ausgewertet und nicht weitergeleitet. Die Relay-Zellen hingegen enthalten den Ende-zu-Ende-Datenstrom.

Eine Kontrollzelle hat das folgende Format:

CircID	CMD	DATA
--------	-----	------

Die 2 *Byte* große Circuit-ID *CircID* gibt dabei den verbindungs-spezifischen Identifikator für einen Circuit an; sie wird also für die zuvor angesprochene Zustandshaltung verwendet. Die Circuit-ID ist nicht für alle Zwischenknoten identisch, sondern nur lokal – d. h. zwischen jeweils zwei Onion Routern – gültig. Der 1 *Byte* große Command-Befehl *CMD* gibt den Befehl an, bspw. „create“, „created“, „destroy“. Das Daten-Feld *DATA* enthält schließlich 509 *Bytes* an Nutzdaten.

Eine Relay-Zelle hat folgendes Format:

CircID	CMD	StreamID	Digest	Len	RelayCMD	Data
--------	-----	----------	--------	-----	----------	------

CMD ist dabei auf „relay“ gesetzt. Ab diesem Feld werden alle weiteren Felder mittels AES (128 *Bit* im Counter-Modus) verschlüsselt übertragen. Die *StreamID* (2 *Bytes*) identifiziert den Datenstrom: Es können mehrere TCP-Verbindungen auf einen Circuit gemultiplext werden. Die 6 *Byte* lange Prüfsumme für den Integritätsschutz ist im Feld *Digest* enthalten. Das 2 *Byte* lange Feld *Len* gibt die Länge des Feldes *Data* an, das die Nutzlast enthält. *RelayCMD* gibt an, was als nächstes passieren soll, ob etwa der Circuit erweitert werden soll („extend“), eine Kommunikationsverbindung mit einem Partner aufgebaut werden soll („begin“) oder Daten weitergeleitet werden sollen („data“) etc.

4.4.3 Aufbau eines Circuits

Bevor wir uns dem Aufbau des Circuits zuwenden, wollen wir noch klären, welche Schlüssel die Onion Router dabei nutzen. Die „Schlüssel“ sind eigentlich Schlüsselpaare, werden aber lediglich als Schlüssel bezeichnet:

- *Identity Key*: Hierbei handelt es sich um einen langlebigen Schlüssel, zu dem der Onion Router ein selbstsigniertes Zertifikat hält. Der öffentliche Schlüssel wird in einem Verzeichnis (dem sogenannten „Directory Server“) veröffentlicht.

- *Onion Key*: Dies ist ein kurzlebiger Schlüssel für den Aufbau der „Onions“. Er wird vom Onion Proxy für die Verschlüsselung verwendet. Der öffentliche Schlüssel wird ebenfalls im Directory Server veröffentlicht.
- *Verbindungsschlüssel*: Der Onion Router hält ein Zertifikat zu diesem Schlüssel – signiert mit dem Identity Key. Dieser Schlüssel kann jederzeit ausgetauscht werden, mindestens jedoch einmal am Tag. Der Schlüssel wird im TLS-Handshake verwendet.

Sehen wir uns nun den Aufbau eines Circuits anhand der Abb. 4.5 im Detail an.

Alice möchte eine Verbindung mit ihrem Kommunikationspartner Dave aufbauen. Dazu wählt sie als Zwischenknoten die Onion Router „Bob“ und „Charly“ aus und baut die Verbindung über diese auf. Zunächst wählt sie eine neue Circuit-ID C_{AB} , die zwischen ihr und Bob gültig ist. Danach wählt sie einen neuen Diffie-Hellman (DH)-Wert DH_{A1} . Diesen verschlüsselt sie mit Bobs Onion Key, den Alice aus dem Directory Server ausliest. Alice sendet nun (mittels Kontrollzelle) den „Create“-Befehl, der die Circuit-ID und den verschlüsselten DH-Wert beinhaltet, an Bob. Bob entschlüsselt Alices DH-Wert, wählt seinerseits einen DH-Wert DH_{B1} und berechnet aus beiden Werten einen gemeinsamen, symmetrischen Sitzungsschlüssel Key_{AB} nach dem Diffie-Hellman-Verfahren (siehe Abschn. 2.3.5). Bob berechnet daraufhin den kryptographischen Hash-Wert des berechneten Sitzungsschlüssels, $H(Key_{AB})$, und sendet diesen, gemeinsam mit seinem DH-Wert

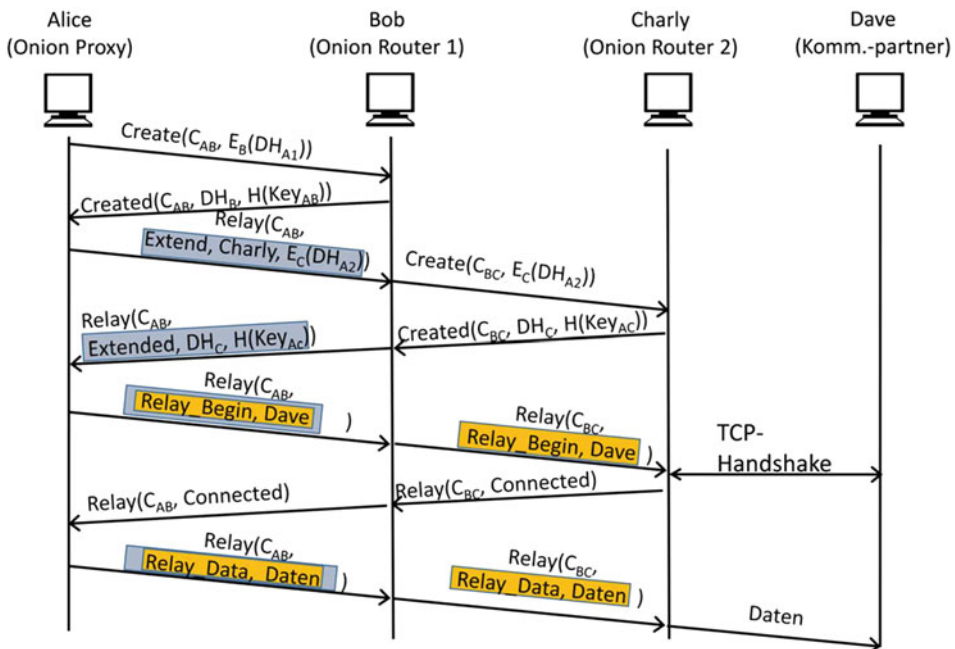


Abb. 4.5 Aufbau eines Tor-Circuits nach [3]

DH_{B1} unter der gemeinsamen Circuit-ID C_{AB} mit dem Befehl „Created“ zurück an Alice. Alice berechnet ihrerseits den gemeinsamen Sitzungsschlüssel Key_{AB} aus den beiden DH-Werten und prüft, ob der Hash-Wert des berechneten Sitzungsschlüssels dem Wert $H(Key_{AB})$ von Bob entspricht. Ist dies der Fall ist der erste Teil des Circuits erfolgreich aufgebaut.

Durch die Überprüfung des Hash-Werts authentifiziert sich Bob implizit gegenüber Alice. Es gilt zu beachten, dass nur Bob in der Lage ist, den verschlüsselten DH-Wert von Alice zu entschlüsseln. Indem er die Entschlüsselung erfolgreich durchführt, aus dem gewonnenen Wert den gemeinsamen Sitzungsschlüssel berechnet und Alice den Hash-Wert über diesen Sitzungsschlüssel mitteilt, beweist er ihr, dass er den zugehörigen privaten Schlüssel (zum öffentlichen Onion Key, den Alice vom Directory Server bezogen hat) kennt. Nach demselben Prinzip authentifizieren sich auch alle anderen Onion Router während des Circuit-Aufbaus gegenüber Alice.

Im nächsten Schritt möchte Alice den aufgebauten Circuit um Charly erweitern. Dazu sendet sie Bob eine Relay-Zelle unter der zwischen den beiden bereits etablierten Circuit-ID C_{AB} . Alice teilt Bob in dieser Nachricht mit, dass sie den Circuit um „Charly“ erweitern („Extend“⁴) möchte. Des Weiteren sendet sie einen neuen, unter dem Onion Key von Charly verschlüsselten DH-Wert mit: $E_C(DH_{A2})$. Diese gesamte Nachricht ist mit dem gemeinsamen, zuvor ausgehandelten symmetrischen Sitzungsschlüssel zwischen Alice und Bob verschlüsselt (dargestellt in der Abbildung als blaue Box).

Bob entschlüsselt die Nachricht und kommt Alices Wunsch nach der Erweiterung des Circuits um Charly nach, indem er seinerseits einen „Create“-Befehl (Kommando-Zelle) an Charly sendet und den verschlüsselten DH-Wert $E_C(DH_{A2})$ weiterleitet. Für diese Nachricht wählt Bob eine neue Circuit-ID C_{BC} . Diese Circuit-ID wird er später wieder verwenden, wenn er eine Relay-Nachricht von Alice erhält, die in dem Circuit an Charly weitergeleitet werden soll. Am Ende dieses Durchlaufs ist der Circuit erfolgreich um Charly erweitert und Alice und Charly verfügen über einen gemeinsamen Sitzungsschlüssel K_{AC} . Nachrichten, die danach von Alice für Charly unter diesem Schlüssel verschlüsselt werden, sind in der Abbildung in einer gelben Box dargestellt.

Schließlich empfängt Charly in einer Relay-Zelle das *Relay-CMD* „Begin“. Dies veranlasst ihn, eine TCP-Verbindung mit Dave aufzubauen.

Der Datenaustausch zwischen Alice und dem Kommunikationspartner Dave erfolgt danach über diesen Circuit, wobei Charly in diesem Beispiel der letzte Knoten, der sogenannte „Exit Node“ ist, der die Daten (im Klartext) an Dave überträgt (in einem sogenannten „Stream“) und von diesem empfängt und über den Circuit an Alice weiterleitet.

Der dargestellte Aufbau des Circuits ist ziemlich aufwendig – aufgrund der vielfachen Verschlüsselungen. Aus „Sicherheitsgründen“, d. h. um besser vor Verkehrsflussanalysen

⁴Zur Verdeutlichung ist hier nur das in Abschn. 4.4.2 erwähnte *Relay-CMD* „extend“ dargestellt und keine weiteren Daten aus dem Header.

zu schützen, wird trotzdem alle 10 Minuten die Route geändert, also ein neuer Circuit aufgebaut. Die Verwendung des Diffie-Hellman-Schlüsselaustauschs während des Circuit-Aufbaus sorgt für *Perfect Forward Secrecy*.

Zu Beginn dieses Abschnitts hatten wir den *Verbindungsschlüssel* erwähnt und gesagt, dass dieser für den TLS-Handshake verwendet wird. Beim Aufbau des Circuits werden alle Verbindungen zwischen den Onion Routern über TLS gesichert, d. h. zwischen Alice und Bob besteht eine TLS-gesicherte Verbindung und zwischen Bob und Charly ebenso. TLS sorgt für eine einseitige (pro Hop) Authentifizierung, d. h. Bob authentifiziert sich gegenüber Alice und Charly authentifiziert sich gegenüber Bob.

4.4.4 Leaky Pipe

Zuvor hatten wir gesagt, dass der Datenaustausch zwischen Alice und Dave über den aufgebauten Circuit vonstatten geht. Alice hat aber auch die Möglichkeit, nur einen Teil des Circuits zu nutzen. Der Vorteil dieses als „Leaky Pipe“ benannten Konzepts ist es, dass Angriffe durch Beobachtung des „Endes“ des Circuits – also des Exit Nodes – erschwert werden. Beobachtet werden können bspw. die Anzahl oder das zeitliche Muster von Paketen. Im Beispiel von vorhin könnte Alice also bereits Bob die Daten an Dave weitergeben lassen – der Stream verlässt den Circuit also bei Bob.

Bob erkennt den Wunsch nach einer Leaky Pipe anhand des oben erwähnten *DIGEST*-Feldes, also anhand der Prüfsumme. Nur im Falle, dass Bob als Exit Node der Leaky Pipe fungieren soll, wird die Prüfsumme der Zelle korrekt sein. In allen anderen Fällen wird die Prüfsumme aufgrund der Mehrfach-Verschlüsselung mit sehr hoher Wahrscheinlichkeit „falsch“ sein. Hierzu muss man sich noch einmal das Konzept von Tor vor Augen führen. Die Zellen werden mehrfach verschlüsselt, wobei sich die Gesamtlänge der Zelle nicht ändert. Es wird auch nicht etwa eine Zelle als verschlüsselte Nutzlast in einer äußeren Zelle mitgeführt, sondern lediglich die Verschlüsselungsfunktion mehrfach angewendet. Wenn das *DIGEST*-Feld die richtige Prüfsumme über die Zelle enthält, ist die Zelle mit hoher Wahrscheinlichkeit vollständig entschlüsselt, d. h. es muss keine weitere „Zwiebelschale“ mehr entfernt werden. Bob erkennt damit, dass er als Exit Node fungieren soll und schickt den Stream direkt aus dem Circuit zu Dave.

4.4.5 Missbrauch von Tor

Eines der Probleme von Tor besteht im Missbrauch der Anonymität zum Schutz vor der Aufdeckung rechtswidrigen Verhaltens, bspw. Urheberrechtsverletzungen, Phishing und anderen Betrugsversuchen. Die Gefahr für die Nutzer bei der Verwendung von Tor liegt darin, dass auch Betreiber eines Onion Routers unter Verdacht geraten können.

Als Gegenmaßnahme wird in der Literatur eine Kennzeichnung als Anonymisierungsdienst gegenüber dem Kommunikationspartner vorgeschlagen. Eine solche Kennzeich-

nung kann entweder anwendungsabhängig geschehen, z. B. durch einen Zusatzheader für HTTP, oder über einen Reverse-DNS-Eintrag. Es steht jedoch zu bezweifeln, ob diese Kennzeichnung überhaupt ausgewertet wird – und falls sie ausgewertet wird, welche Schlüsse daraus gezogen werden. Eine weitere Gegenmaßnahme besteht im Filtern besonders „anfälliger“ Dienste, wie etwa dem Peer-to-Peer-Filesharing. Eine Risiko-Reduktion für einzelne Betreiber kann auch dadurch herbeigeführt werden, dass ein Onion Router nur als „middleman node“ fungiert, Daten also nur an andere Tor-Knoten weitergibt.

4.4.6 Hidden Services

Der Ausgangspunkt von *Hidden Services* besteht darin, dass der Anbieter eines Dienstes (bspw. ein Webserver) seine Identität nicht preisgeben will. In Abb. 4.6 möchte Alice einen Hidden Service nutzen, der von Bob angeboten wird.

Der Diensteanbieter wählt Onion Router als „Introduction Points“ und baut Circuits zu diesen auf. Danach veröffentlicht er seine Dienstbeschreibung und die Liste der Introduction Points in einer (verteilten) Datenbank. Der Nutzer bezieht die Dienstbeschreibung und die Liste und wählt seinerseits einen sogenannten „Rendezvous Point“, zu dem er einen Circuit aufbaut und den Diensteanbieter (über den Introduction Point) darüber informiert. Der Diensteanbieter baut den Circuit zum Rendezvous Point auf und der Rendezvous Point verbindet die Circuits.

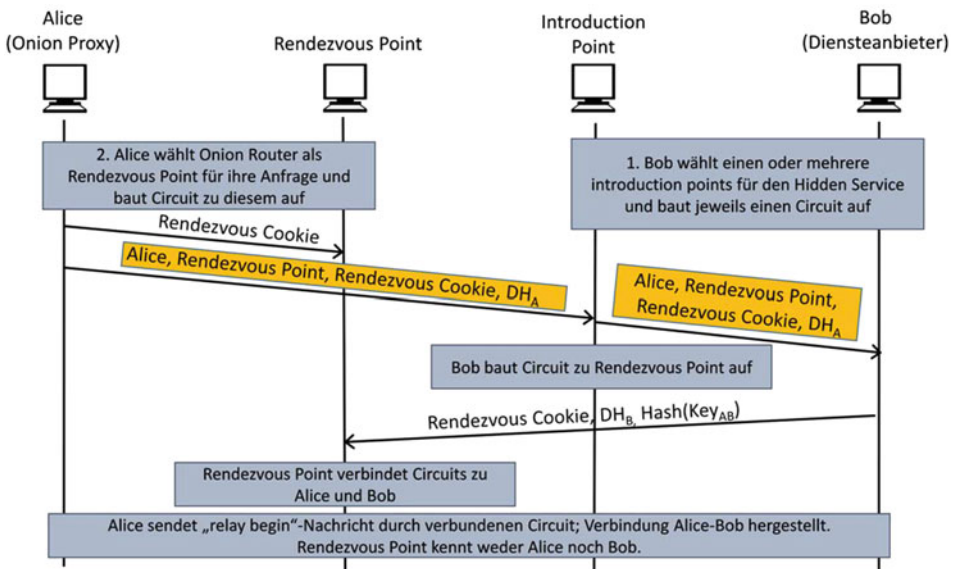


Abb. 4.6 Tor Hidden Services

Die gelbe Box in der Abbildung steht für die Verschlüsselung unter Bobs öffentlichem Schlüssel. Dieser öffentliche Schlüssel muss zusammen mit der Liste der Introduction Points veröffentlicht werden. Das „Rendezvous Cookie“ in der Abbildung ist ein Zufalls-wert.

Diese Lösung bietet Anonymität sowohl für Anbieter als auch für Nutzer.

4.4.7 Angriffe auf Tor

Die Angriffe auf Onion Routing sind verwandt zu jenen, die wir bereits im Zusammenhang mit Mixes kennengelernt haben. Wir hatten gesagt, dass kein Schutz besteht, wenn ein Mix nicht vertrauenswürdig ist. Dies gilt bei Mix-Kaskaden und Onion Routing/Tor zunächst nur, wenn *kein* Mix/Onion Router auf dem Pfad vertrauenswürdig ist.

Das Ziehen von Rückschlüssen aus der Größe von Paketen gilt bei Tor analog zu Mixes.

Der (n-1)-Angriff, den wir bei Mixes kennengelernt haben, funktioniert bei Tor nicht, da es hier keine Sammelfunktion gibt. Die lokale Beobachtung nur eines Onion Routers ist für einen Angreifer auch nicht ausreichend, selbst wenn aufgrund der fehlenden Sammelfunktion eine Korrelation von eingehendem und ausgehendem Verkehr an einem Knoten möglich ist.

Wenn der Empfänger als Angreifer auftritt, gibt es bei Tor genauso wie bei Mixes keinen Schutz vor der Analyse übertragener Daten in der Anwendungsschicht, z. B. im E-Mail-Header.

Wir hatten bereits gesagt, dass der Exit Node, nachdem er die Zelle vollständig entschlüsselt hat – also alle Zwiebelschalen entfernt hat –, die Daten im Klartext vorliegen hat. Möchte man sich vor einem neugierigen Exit Node schützen, so muss man als Nutzer dafür sorgen, dass man auf Transport- bzw. Anwendungsebene (über bspw. TLS oder verschlüsselte E-Mails) zusätzliche Sicherungsmaßnahmen ergreift.

Ein Angreifer, der sowohl den eingehenden als auch den ausgehenden Onion Router kontrolliert, ist in der Lage, die Anonymität zu brechen.

Beispiel für einen Angriff

Im Jahr 2007 haben BAUER et al. [1] einen Angriff auf Tor veröffentlicht, der auf der Vortäuschung einer größeren Bandbreite der Onion Router des Angreifers basiert. Bei Tor wird zwischen der Auswahl der Entry Router und anderer (Non-Entry) Router unterschieden. Als Entry Router werden solche gewählt, die von den Verzeichnis-Servern als „schnell“ und „stabil“ bezeichnet werden (d.h. Bandbreite bzw. Uptime liegen über dem jeweiligen Median). Die Auswahl der anderen Router geschieht in zwei Schritten. Im ersten Schritt werden nur als „stabil“ markierte Onion Router ausgewählt. Im zweiten Schritt wird aus diesen dann einer zufällig ausgewählt, wobei die Auswahlwahrscheinlichkeit proportional zur, durch den Router angegebenen, Bandbreite ist. Die Auswahlwahrscheinlichkeit für Onion Router i : $p_i = b_i / (\sum_{j=1}^N b_j)$, wobei b_i die

angegebene Bandbreite des Onion Routers i , und N die Anzahl zur Auswahl stehender Onion Router angeben. Der Angriff läuft in zwei Phasen ab:

Phase 1: Der Angreifer betreibt selbst Onion Router oder kompromittiert existierende. Um die Wahrscheinlichkeit zu erhöhen, dass seine Router ausgewählt werden, wird eine unbeschränkte Exit Policy angegeben (d. h. die Onion Router sind bereit, Daten an beliebige Zwischen- oder Endknoten auszuliefern). Die Onion Router teilen zudem größere Ressourcen (insbesondere Bandbreiten) mit, als ihnen tatsächlich zur Verfügung stehen. Auf diese Weise wird die Wahrscheinlichkeit für die Auswahl dieser Onion Router erhöht. Ist in einem Pfad mindestens ein Onion Router des Angreifers enthalten, ohne dass der Angreifer gleichzeitig Entry- und Exit-Router kontrolliert, unterbricht er die Kommunikation, so dass ein neuer Pfad (womöglich mit besserer Ausgangssituation des Angreifers) etabliert werden muss.

Phase 2: Die Onion Router des Angreifers kommunizieren untereinander, erkennen somit ihre Position im Pfad und können durch Timing-Analysen beide Kommunikationspartner bestimmen, falls sowohl Entry- als auch Exit-Router unter Kontrolle des Angreifers stehen.

Es besteht ein Unterschied zwischen früheren analytischen Modellen und der experimentell ermittelten Erfolgswahrscheinlichkeit. Der Unterschied lässt sich dadurch erklären, dass:

- der Angreifer über vorhandene Ressourcen lügen kann,
- die analytischen Modelle davon ausgingen, die Ressourcen seien gleichmäßig über die Onion Router verteilt,
- der Angreifer Pfade unterbrechen und somit das Generieren neuer Pfade erzwingen kann,
- Onion Router in der verwendeten Implementierung pro Circuit nur einmal verwendet werden können.

Damit wurde diesem Angriff eine hohe praktische Relevanz zugesprochen. Inzwischen wurde in Form von „Bandwidth Authorities“ eine Gegenmaßnahme implementiert.

Man muss festhalten, dass es neben den in diesem Abschnitt dargestellten Angriffen noch weitere gibt. Tor ist das vermutlich bekannteste Anonymisierungsnetz und deshalb beliebtes „Angriffsziel“. In mehreren Presseberichten der letzten Jahre wurde dargelegt, dass die NSA versucht, die Anonymität von Tor-Nutzern zu brechen. Die Berichte deuten aber darauf hin, dass das Tor-System, das ursprünglich durch das *United States Naval Research Laboratory* – dem gemeinsamen Forschungslabor der Navy und der Marine Corps – unterstützt wurde, selbst durch die NSA nicht vollständig gebrochen werden kann. Offenbar wurden aber Schwachstellen in den Endsystemen (konkret beispielsweise in Firefox) ausgenutzt, um diese zu einer direkten Datenübermittlung – unter Umgehung der Anonymisierung durch Tor – zu bewegen.

4.4.8 Zensurreistenz mit Tor

Tor kann genutzt werden, um Zugriffssperren zu umgehen, die in manchen Ländern zur Zensur verwendet werden. Damit können über Tor Dienste genutzt werden, zu denen der Zugang eigentlich gesperrt ist.

Die Liste der Onion Router ist öffentlich zugänglich – und damit ist es für Zensur-Behörden einfach möglich, den Zugang zu Tor zu sperren, indem Verbindungen zu diesen bekannten Onion Routern unterbunden werden. Die Lösung für dieses Problem sind sogenannte *Bridges*, d.h. spezielle Knoten, die eine Überbrückung zwischen gesperrten Nutzern und dem Tor-Netz herstellen. Jeder Tor-Teilnehmer kann diese Bridging-Funktionalität zur Verfügung stellen. Die Adresse wird dabei den gesperrten Nutzern direkt mitgeteilt oder kann bei einer vertrauenswürdigen Partei („Bridge Authority“) hinterlegt werden. Bei der Bridge Authority wird die Adresse in einer von drei Adress-Pools hinterlegt. Je nachdem in welchem Pool sich die Adresse befindet, findet eine unterschiedliche Verteilung statt.

- Pool 1: Verteilung über Webseite
- Pool 2: Verteilung über E-Mail
- Pool 3: Verteilung über Instant Messaging, soziale Netzwerke etc.

Um ein erschöpfendes Abgreifen aller vorhandenen Bridges durch die Zensur-Behörde zu verhindern, wird je Anfrage-IP-Adresse, bzw. je E-Mail immer nur ein bestimmter (gleichbleibender) Bereich von Adressen zurückgeliefert. Allerdings schließt dieser Ansatz einen Angriff (vor allem von Zensur-Behörden mit vielen Ressourcen) nicht vollständig aus – es gibt bekannte Fälle von „erfolgreichen“ Bridge-Suchstrategien.

4.5 Fazit

In diesem Kapitel haben wir unterschiedliche Ansätze zur Verschleierung der Kommunikationsbeziehungen im Internet kennengelernt. Mixes und Mix-Kaskaden bilden die Grundlage zu Onion Routing/Tor, das wir aufgrund seiner Beliebtheit näher untersucht haben. Laut aktuellem Tor-Statusreport <http://torstatus.blutmagie.de/> standen im Januar 2016 über 7000 Onion Router zur Verfügung. Genutzt werden diese von etwa zwei Millionen Nutzern.⁵ Tor lebt zu einem großen Teil von der aktiven Mitarbeit der Forscher-Community und Freiwilligen, die den Dienst und die Implementierung ständig auf ihre Sicherheit hin analysieren und verbessern. Die Bedienbarkeit von Tor ist ein weiterer, wesentlicher Faktor für dessen Verbreitung, auch unter Laien. Mit dem „Tor

⁵<http://www.wired.com/2016/01/david-chaum-father-of-online-anonymity-plan-to-end-the-crypto-wars/>

Browser Bundle“ steht eine vorkonfigurierte, einfach zu verwendende portable Lösung zur Verfügung. Sie besteht aus einem modifizierten Firefox Browser und dem Tor Proxy.

Neben Tor existierten und existieren weitere Anonymisierungsdienste, wie etwa *Crowds* oder *JAP*. Keinem dieser Anonymisierungsnetze ist es allerdings gelungen, eine derart hohe Nutzerzahl wie Tor zu erreichen.

Wie sich Anonymisierungsdienste weiter entwickeln werden und wie die Nachfrage danach in Zukunft aussehen wird ist noch nicht absehbar. Durch die Veröffentlichungen von EDWARD SNOWDEN haben diese Dienste, allen voran Tor, einen starken Zulauf verzeichnen können. Für Whistleblower, Dissidenten, Regimekritiker etc. werden Anonymisierungsdienste auch in Zukunft von enormer Bedeutung sein. Gleichzeitig werden Anonymisierungsdienste von vielen in ein schlechtes Licht gerückt, da mit deren Hilfe rechtswidriges Verhalten möglich wird, dem schwer bis gar nicht nachgegangen werden kann. In diesem Zusammenhang ist auch vom „Dark Net“ die Rede, einem Teil des Internets, das von Drogen- bis Waffenhandel alles mögliche an illegalen Diensten bereithält. Die Plattform *Silk Road* war ein derartiger als Hidden Service getarnter Schwarzmarkt, der 2014 von den Behörden abgeschaltet wurde. Bezahlt wurde auf diesem Schwarzmarkt mit der anonymen Krypto-Währung *Bitcoin*, die wir im übernächsten Kapitel genauer kennenlernen werden.

4.6 Übungsaufgaben

Aufgabe 1

Warum wird die Circuit-ID nicht global für den gesamten Circuit gewählt? Warum werden stattdessen Circuit-IDs gewählt, die jeweils nur zwischen zwei Knoten gelten?

Aufgabe 2

Warum ist es bei Tor kein Problem, dass Bob während des Verbindungsaufbaus den Hash-Wert des berechneten Sitzungsschlüssels an Alice überträgt?

Literatur

1. Kevin Bauer, Damon McCoy, Dirk Grunwald, Tadayoshi Kohno, and Douglas Sicker. Low-resource routing attacks against Tor. In *Proceedings of the 2007 ACM Workshop on Privacy in Electronic Society*, WPES '07, pages 11–20, New York, NY, USA, 2007. ACM.
2. David L. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 24(2):84–90, February 1981.
3. Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. In *Proceedings of the 13th Conference on USENIX Security Symposium – Volume 13*, SSYM'04, pages 21–21, Berkeley, CA, USA, 2004. USENIX Association.

4. David M. Goldschlag, Michael G. Reed, and Paul F. Syverson. Hiding routing information. In *Proceedings of the First International Workshop on Information Hiding*, pages 137–150, London, UK, 1996. Springer-Verlag.
5. Andriy Panchenko and Lexi Pimenidis. Towards practical attacker classification for risk analysis in anonymous communication. In Herbert Leitold and Evangelos P. Markatos, editors, *Communications and Multimedia Security*, volume 4237 of *Lecture Notes in Computer Science*, pages 240–251. Springer Berlin Heidelberg, 2006.

Zusammenfassung

Identitätsmanagement (IdM) ist die Verwaltung mehrerer partieller Identitäten von Subjekten (Personen). Es geht dabei also um eine Verwaltung von Attributwerten und die Auswahl einer zu verwendenden Identität in einem spezifischen Kontext. Dabei spielt außerdem die Authentifizierung unter der ausgewählten Identität eine wichtige Rolle.

Wir werden in diesem Kapitel sehen, dass Identitätsmanagement dem Datenschutz dient, wenn sichergestellt wird, dass mehrere Identitäten eines Subjekts nicht miteinander verknüpft werden können.

In Abschn. 5.1 verschaffen wir uns zunächst einen Überblick über das umfangreiche Thema Identitätsmanagement. Danach lernen wir in Abschn. 5.2 *OpenID*, ein im Internet sehr häufig genutztes IdM-System, kennen. Danach betrachten wir mit *OAuth* in Abschn. 5.3 ein populäres Protokoll, das überwiegend zur Autorisierung im Internet verwendet wird. In Abschn. 5.4 lernen wir schließlich *OpenID Connect* kennen – ein neuartiges Protokoll, das die Konzepte von OpenID und OAuth vereint.

Lernziele

Am Ende dieses Kapitels sollten Sie mit ausgewählten Verfahren aus dem Bereich des Identitätsmanagements, die heute in der Praxis häufig zum Einsatz kommen, vertraut sein. Sie sollten die unterschiedlichen Eigenschaften die diese Verfahren bieten – vor allem im Hinblick auf den Datenschutz – kennen und bewerten können, welches Verfahren für welchen Zweck eingesetzt werden kann.

5.1 Überblick

Im Identitätsmanagement (IdM) haben wir es häufig mit den folgenden Entitäten zu tun: *Subjekt/Nutzer*, *Identitätsprovider (IdP)* und *Dienstanbieter*. Das Zusammenspiel zwischen diesen Identitäten ist in Abb. 5.1 dargestellt.

Wir werden in den nachfolgenden Abschnitten unterschiedliche Verfahren kennenlernen, die auf dieser Basis arbeiten.

5.1.1 Schwerpunkte und Sichtweisen im Identitätsmanagement

IdM spielt in (größeren) Unternehmen und Organisationen, bspw. an Universitäten, eine wesentliche Rolle. Nutzer, an einer Universität etwa Mitarbeiter, Studierende und Gäste, verwenden zahlreiche Dienste. Von der Immatrikulation über die Nutzung der Bibliotheksdienste und der Buchung von Sportkursen bis hin zum Prüfungsmanagement muss an unterschiedlichen Stellen eine Identifizierung (sowie Authentifizierung und Autorisierung) der Nutzer erfolgen. Dabei werden jeweils (partielle) Identitäten der Nutzer benötigt. Ein IdM-System sollte also einen geordneten Umgang mit partiellen Identitäten ermöglichen.

Zum einen muss eine einheitliche Authentifizierung gewährleistet werden. Dabei spielt das sogenannte *Single-Sign-On (SSO)* eine wesentliche Rolle. SSO ermöglicht es einem Nutzer, mit einem Login-Vorgang verschiedene Dienste zu nutzen. Dabei erhalten die Dienstanbieter (im Idealfall) lediglich die Informationen (Attribute), die sie für die Erbringung ihrer Dienste benötigen.

Zum anderen muss das Management der partiellen Identitäten über den vollständigen Lebenszyklus gewährleistet werden. Zwischen Eintritt eines Studierenden bzw. Mitar-

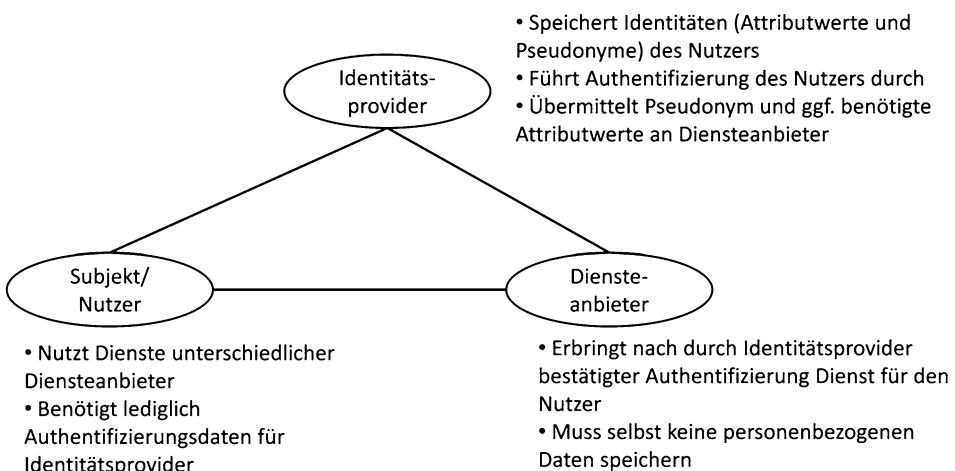


Abb. 5.1 Identitätsmanagement im Überblick

beiters und Ausscheiden aus der Universität können eine Reihe weiterer Veränderungen der Rollen eintreten und diese muss das IdM-System entsprechend berücksichtigen können. Selbst nach dem Ausscheiden kann das IdM-System unter Umständen noch als „Zulieferer“ für ein Alumni-System dienlich sein.

In der Praxis gibt es zahlreiche IdM-Systeme, angefangen von Active Directory-Lösungen bis hin zu SAP-Systemen, die diese Aufgaben erfüllen. Bei all diesen Lösungen steht die *Sicht der Organisation* im Vordergrund.

Daneben existieren eine Reihe von IdM-Lösungen für web-basierte Dienste. Diese in den letzten Jahren populärer werdenden Systeme haben das Ziel, den „planlosen“ Umgang mit Identitäten bei der Verwendung von unterschiedlichen Diensten im Internet zu beseitigen. Häufig ist es so, dass Nutzer dieselben Zugangsdaten (Nutzername und Passwort) für verschiedene Dienste nutzen. Im schlimmsten Fall kann die Kompromittierung eines Diensteanbieters die Kompromittierung aller Zugänge nach sich ziehen. Außerdem können verschiedene Diensteanbieter die personenbezogenen Daten der Nutzer zusammenführen. IdM-Systeme für web-basierte Dienste versprechen, diesem Problem u. A. mit Single-Sign-On (SSO) entgegenzutreten zu können. Die Idee ist, dass Nutzer sich nur *einen* Nutzernamen und *ein* starkes Passwort merken müssen, um sich bei einem Identitätsprovider (IdP) zu authentifizieren. Für die einzelnen Diensteanbieter müssen die Nutzer dann nur noch festlegen, unter welcher partiellen Identität sie dort auftreten möchten, d. h. welche personenbezogenen Daten sie dort preisgeben möchten. Im Gegensatz zu den vorhin genannten IdM-Ansätzen, die die Diensteanbieter-Sicht (d. h. Organisations-Sicht) in den Vordergrund stellen, steht hier die Nutzer-Sicht im Vordergrund. Man spricht in diesem Zusammenhang auch häufig von einem „*user-centric identity management*“, also einem Nutzer-zentrierten IdM. Wir werden in diesem Kapitel allerdings auch noch sehen, dass nicht alle Nutzer-zentrierten IdM-Systeme automatisch ein hohes Maß an Datenschutz gewährleisten.

5.2 OpenID

OpenID ist ein Framework für die *Nutzerauthentifizierung* im Web. Dabei ist das System, das aus *OpenID-Providern* (Identitäts Providern) und *Relying Parties* (Diensteanbietern) besteht, dezentral organisiert. Es gibt keine zentrale Stelle, die die vielzähligen Identitätsprovider und Diensteanbieter überprüft oder bestätigt. Weiterentwickelt wird OpenID von der *OpenID Foundation*, einer internationalen Non-Profit-Standardisierungsorganisation, der sich Unternehmen und Einzelentwickler angeschlossen haben. OpenID in der aktuellen Version 2.0 [3] wird von zahlreichen Unternehmen unterstützt, darunter Google, Microsoft, Facebook, PayPal etc.

Ein zentrales Element bei OpenID ist die *Authentifizierung*. Ein Nutzer, der sich bei einem OpenID-Provider – den er selbst wählt – anmeldet, kann sich mithilfe seines *Identifiers* bei den Relying Parties anmelden. Die Authentifizierung gegenüber dem OpenID-Provider erfolgt mittels Nutzername und Passwort. Gegenüber den Relying

Parties muss der Nutzer seinen Nutzernamen und sein Passwort nicht preisgeben. Der Identifier ist in diesem Kontext ein Uniform Resource Locator (URL), bspw. könnte ein Nutzer die Identität <http://example.com/user123> haben.¹

5.2.1 Ablauf der Authentifizierung

Der Nutzer besucht eine Website, die OpenID-Authentifizierung als Diensteanbieter (Relying Party) unterstützt. Um die Authentifizierung zu starten, gibt der Nutzer seine Identität, also etwa seine URL <http://example.com/user123> an. Der Diensteanbieter führt eine Auflösung der URL durch. Die URL verweist auf ein Dokument, das die URL des zuständigen Identitätsproviders (OpenID-Providers) des Nutzers enthält. Nachdem der Diensteanbieter und der Identitätsprovider (IdP) ein gemeinsames Geheimnis nach Diffie-Hellman aushandeln, das für die spätere Authentifizierung der Nachrichten verwendet wird, wird der Nutzer zu seinem IdP weitergeleitet. Die Weiterleitung ist dabei entweder ein HTTP Redirect, oder es wird ein Formular eingebunden, das an den IdP geschickt wird.

Der Nutzer authentifiziert sich nun mittels Nutzernamen und Passwort gegenüber dem IdP. Der IdP leitet den Nutzer schließlich an den Diensteanbieter weiter. Dabei gibt er dem Nutzer eine Authentifizierungsantwort mit, die mit dem zuvor abgeleiteten Geheimnis zwischen IdP und Diensteanbieter authentifiziert wird.

5.2.2 Analyse

OpenID verwendet HTTP zum Transport von Nachrichten. Dies ist aus Sicherheitssicht ein Problem, da der IdP nicht authentifiziert wird. Aus diesem Grund empfiehlt der Standard auch die Verwendung von HTTPS, also der Kombination von HTTP und Transport Layer Security (TLS), zur Kommunikation. HTTPS garantiert sowohl die Authentizität² und Integrität als auch die Vertraulichkeit der ausgetauschten Nachrichten.

Trotz der Verwendung von TLS stellt die Authentifizierung des Identitätsproviders gegenüber dem Nutzer ein praktisches Problem dar. Die Zieladresse für die Eingabe der Authentifizierungsdaten kommt vom Diensteanbieter. Der Nutzer müsste das Zertifikat des (vermeintlichen) Identitätsproviders überprüfen, um sicherzugehen, dass er seine Authentifizierungsdaten an den richtigen IdP sendet und nicht Opfer eines Phishing-Angriffs (vom Diensteanbieter) wurde. Gerade bei der Verwendung von HTML-Formularen ist dies aber nicht einfach sicherzustellen.

¹ Kritiker sprechen in diesem Kontext auch von einer „Entmenschlichung“, da sich Nutzer im Internet auf die selbe Art und Weise wie Webseiten identifizieren: mit einer URL.

² Der OpenID-Standard spricht bei der Authentifizierung von Nachrichten mittels Message Authentication Code (MAC) von „Signaturen“. Eine eigentliche Signaturprüfung findet aber bei OpenID nicht statt.

Aus Datenschutzsicht gibt es bei OpenID zwei Probleme. Zum einen „recyclen“ große Identitätsprovider Nutzer-Identitäten von inaktiven Accounts. Dadurch können die neuen Besitzer einer Identität Zugriff auf Daten vom alten Besitzer erlangen, sofern der Diensteanbieter vom Wechsel nichts weiß. Zum anderen ist der IdP in der Lage, sehr detaillierte Nutzer-Profile zu bilden. Der IdP erfährt, welche Dienste die Nutzer in Anspruch nehmen. Hierzu heißt es unter www.openidexplained.com: „OpenID is not Big Brother–It doesn’t keep track of what you do on those websites; that is still controlled by the websites“.

Zu guter Letzt gilt es noch zu beachten, dass man als Nutzer ein starkes Passwort für seinen OpenID-Account wählen sollte. Ein Angreifer, der den Account eines Nutzers kapert, kann sich damit bei allen Diensteanbietern als legitimer Nutzer anmelden. Dies wird unter www.openidexplained.com auch klar kommuniziert:

„OpenID is no less (or more) secure than what you use right now. It’s true that if someone gets your OpenID’s username and password, they can usurp your online identity. But, that’s already possible. Most websites offer a service to e-mail you your password (or a new password) if you’ve forgotten it, which means that if someone breaks into your e-mail account, they can do just as much as they can if they get your OpenID’s username and password.“

5.3 OAuth

OAuth 2.0 (in weiterer Folge OAuth) ist ein Framework zur Unterstützung der Entwicklung von Authentifizierungs- und Autorisierungsprotokollen für Desktop-, Web- und Mobil-Anwendungen. OAuth basiert auf standardisierten Nachrichten, die mittels JSON und HTTP ausgetauscht werden. Standardisiert ist OAuth in Request for Comments (RFC) 6749 [1] und RFC 6750 [2]. Da es sich um ein Framework handelt, sind keine genauen Implementierungsvorgaben enthalten.

Der Fokus von OAuth liegt auf der *Autorisierung*. OAuth erlaubt es Nutzern, einer Anwendung Zugriff auf seine Daten zu geben, die von einer anderen Anwendung verwaltet werden. Dabei muss der Nutzer seine Authentifizierungsdaten nicht an diese Anwendung weitergeben.

Praxis

Facebook setzt beim „Login-Button“ auf eine (eigene) Implementierung von OAuth. Wird der Login-Button von einem Diensteanbieter auf einer Website eingesetzt, so können sich die Nutzer auf dieser Website mit ihrem Facebook-Account (über Facebook) „einloggen“. Eine Registrierung auf dieser Website ist nicht mehr nötig. Der Nutzer kann nach dem Einloggen entscheiden, auf welche Daten – die bei Facebook, also dem Identitätsprovider (IdP), hinterlegt sind – er dem Diensteanbieter Zugriff gewährt. Tatsächlich ist es allerdings so, dass der Nutzer, wenn er den Facebook Login-Button nutzen möchte, jene Daten freigeben muss, die vom Diensteanbieter verlangt werden. Typischerweise sind dies die öffentlichen Profilinformationen sowie die Freundesliste.

Analog verhält es sich bei Apps, die das Einloggen über Facebook erlauben. Hier kommt ebenfalls OAuth zum Einsatz.

Auch andere große Betreiber von sozialen Netzwerken, etwa *Google*, *Twitter* oder *LinkedIn* treten als Identitätsprovider auf und erlauben ähnliche Dienste mittels OAuth.

5.3.1 Verfahren

Bevor wir das Verfahren im Detail betrachten, machen wir uns noch mit den beteiligten Parteien vertraut. Um in der Nomenklatur in diesem Kapitel konsistent zu bleiben, sprechen wir auch weiterhin vom Nutzer, dem Identitätsprovider (**IdP**) und dem Diensteanbieter. Wir nennen zum bessern Verständnis beim Nachschlagen in weiteren Quellen auch die Bezeichnungen der Parteien, wie sie in den RFCs genannt werden.

Der *Identitätsprovider (IdP)* hält die Daten über Nutzer („Protected Resources“) für die Nutzung der Diensteanbieter vor. Außerdem stellt der IdP die Implementierung von OAuth bereit. Der IdP trägt im OAuth-Kontext auch die Bezeichnung „Resource Server“. Der IdP führt in seiner Rolle als „Authorization Server“ auch die Autorisierung durch, d. h. er gewährt dem Diensteanbieter nach erfolgreicher Authentifizierung durch den Nutzer Zugriff auf die gespeicherten Daten. Dazu stellt der IdP sogenannte „Access Tokens“ aus; dies sind Zeichenketten aus Buchstaben und Zahlen, die schwer zu erraten sind.

Der *Nutzer* (auch „Resource Owner“ genannt) ist der „Eigentümer“ seiner Daten, die beim IdP gespeichert sind. Er behält mittels OAuth die Kontrolle darüber, welche seiner Daten an Diensteanbieter weitergegeben werden.

Der *Diensteanbieter* (auch „Consumer“ oder „Client“ genannt) ist im OAuth-Kontext strenggenommen eine Anwendung, die Zugriff auf die Nutzer-Daten erfragt.

Sehen wir uns nun den Ablauf, der in Abb. 5.2 dargestellt ist, im Detail an.

Zur einfacheren Lesbarkeit haben wir den Nutzer und den Browser des Nutzers, der außer der Bestätigung des Zugriffs (dem „Granting“) alle Schritte vornimmt, zusammengefasst.

Der Nutzer startet die Dienstnutzung bei einem Diensteanbieter und gibt dort an, dass sich der Diensteanbieter die Daten über den Nutzer bei einem bestimmten **IdP** abholen soll. Hierzu stellt der Dienst zunächst einen Authorization Request. Dieser Authorization Request veranlasst im Idealfall, bei Verwendung des sogenannten „Authorization Code“ als „Grant Type“, wie es auch in der Abbildung dargestellt ist, einen Redirect des Nutzers zum IdP. Der Authorization Request enthält neben der ID des Dienstes unter anderem auch den „Scope“ sowie eine „Redirect URI“, also eine Adresse zu der der Nutzer nach der Authentifizierung und Autorisierung geleitet wird. Der Scope definiert die Berechtigungen für die Nutzung der Daten, die vom Nutzer erfragt werden.

Der Nutzer authentifiziert sich also gegenüber seinem IdP (klassischerweise mittels Nutzernamen und Passwort) und gewährt dem Diensteanbieter Zugriff auf seine Daten. Der IdP bestätigt diese Autorisierung des Nutzers, also die Einwilligung zur Datenweitergabe, über ein „Authorization Grant Credential“ (auch „Authorization Code“ genannt).

Dieses Authorization Grant Credential wird dem Diensteanbieter zugeleitet. Der Diensteanbieter authentifiziert sich im Rahmen des „Access Token Requests“ gegenüber

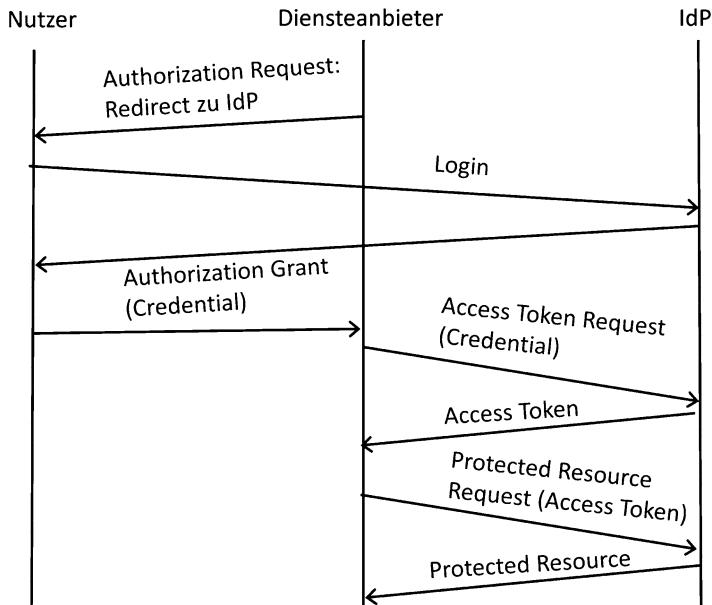


Abb. 5.2 Autorisierung mit OAuth

dem IdP und sendet diesem auch das Credential. Nach erfolgreicher Authentifizierung des Diensteanbieters und Überprüfung des Credentials stellt der IdP dem Diensteanbieter ein „Access Token“ aus, das den Zugriff auf die Daten über den zuvor festgelegten und akzeptierten Scope für eine gewisse Zeitspanne erlaubt. Die Authentifizierung des Diensteanbieters erfolgt in der Regel über ein Passwort, das dem Diensteanbieter bei der Registrierung beim IdP mitgeteilt wird.

Schließlich erhält der Diensteanbieter nach Vorlage des Access Tokens beim IdP Zugriff auf die beim IdP hinterlegten Daten des Nutzers.

5.3.2 Analyse

Die Sicherheit von OAuth basiert im Wesentlichen auf der Verwendung des Transport Layer Security (TLS)-Protokolls. OAuth bietet von sich aus keine Maßnahmen, um Schutzziele wie etwa Vertraulichkeit oder Authentizität zu bieten.

Daneben kommt es noch auf die Wahl des „Grant Types“ an. Der Grant Type „Authorization Code“, den wir in der Beschreibung des Ablaufs gewählt haben, bietet das höchste Maß an Sicherheit und garantiert damit auch das höchste Maß an Datenschutz. Bei dieser Art der Autorisierung gelangen die Authentifizierungsdaten der Nutzer nicht zum Diensteanbieter. Die Authentifizierung des Diensteanbieters gegenüber dem IdP während des Access Token Requests schützt vor einer Impersonation des Diensteanbieters durch

einen betrügerischen Diensteanbieter. Als weiterer Vorteil bei dieser Art der Autorisierung kann die Tatsache gesehen werden, dass der IdP dem Diensteanbieter das Access Token direkt übermittelt – ohne den Umweg über den Browser des Nutzers zu gehen; und damit einen gewissen Schutz vor der Aufdeckung dieses Tokens gegenüber Dritten bietet.

Beim Grant Type „Implicit“ hingegen findet keine Authentifizierung des Diensteanbieters statt und das Access Token nimmt den Umweg über den Browser des Nutzers. Die Ausstellung des Authorization Grant Credentials entfällt bei dieser Methode aus Performance-Gründen; stattdessen wird direkt ein Access Token ausgestellt.

Aus Datenschutzsicht am bedenklichsten ist die Verwendung des Grant Types „Resource Owner Password Credentials“. Hierbei stellen die Login-Daten des Nutzers zur Authentifizierung beim IdP die Authorization Grant Credentials dar. Damit gelangt auch der Diensteanbieter an die Login-Daten (Nutzername und Passwort) des Nutzers.

5.4 OpenID Connect

OpenID Connect ist die neueste Entwicklung der OpenID Foundation. Der Standard wird unter anderem von Google für den Dienst „Google Sign-In“ verwendet, der verspricht, Nutzer damit schnell und einfach – und ohne viel Entwicklungsarbeit – Zugang zu Websites, Apps und Geräten (bspw. TVs) zu verschaffen. OpenID Connect ist zudem die Grundlage für die Integration von Google-Diensten bei Diensten von Drittanbietern.

OpenID Connect (in der Version 1.0) ist im Wesentlichen nichts anderes als OAuth 2.0, mit einem zusätzlichen Identitäts-„Layer“ (basierend auf OpenID 2.0) „on top“. Damit erlaubt OpenID Connect den Diensten einerseits die Identifizierung und Authentifizierung der Nutzer. Andererseits sorgt das OAuth-Protokoll auch für eine Autorisierung und Bereitstellung von beim IdP gespeicherten personenbezogenen Daten (bspw. Profilinformationen, Freundschaftslisten etc.); dieser Austausch findet über REST („Representational State Transfer“, ein Programmierparadigma für Webservices) statt.³ OpenID Connect unterstützt dabei unterschiedliche Dienste, von web-basierten über mobile bis hin zu JavaScript-Diensten. Die Spezifikation ist dahingehend erweiterbar, dass Dienste bspw. zusätzlich eine Verschlüsselung der Identitätsdaten bzw. das Auffinden von OpenID-Providern einführen können.

5.5 Fazit

In der Praxis basiert eine Vielzahl von Diensten auf der Verwendung von *OpenID*, *OAuth* und *OpenID Connect*. Eine Bewertung, ob diese IdM-Verfahren datenschutzfördernd sind, fällt nicht leicht. Auf der einen Seite ist der Ansatz, dass Nutzer sich ihre Identitätsprovider

³Es kommt eine „RESTful HTTP-API“ mit JSON als Daten-Format zum Einsatz.

(IdP) – bei denen die personenbezogenen Daten gespeichert sind – selbst auswählen können, vielversprechend. Die Idee, dass Nutzer dann für jeden Dienst entscheiden können, welche personenbezogenen Daten sie diesem überlassen möchten, entspricht geradezu dem in Abschn. 2.2 kennen gelernten Konzept der *partiellen Identität*, mit der man unterschiedlichen Diensteanbietern gegenübertritt. Allerdings steht und fällt das ganze mit der Auswahl des Identitätsproviders. Entscheiden sich die Nutzer für den „einfachen“ Weg und wählen als IdP einen Anbieter eines großen sozialen Netzwerks aus, so darf bezweifelt werden, ob dies dem Datenschutz dienlich ist. Warum tritt bspw. Facebook als großer IdP auf? Weil Facebook dadurch erfährt, welche Dienste seine Nutzer außerhalb des sozialen Netzwerks sonst noch verwenden – eine Diskussion, die wir im Zusammenhang mit dem „Like Button“ auch in Abschn. 7.2 noch einmal führen werden. Facebook erfährt durch jeden Login-Vorgang auch die Intensität der Dienst-Nutzungen der Nutzer. Diese Informationen sind übrigens nicht nur für die Nutzer-Analysen wertvoll, sondern auch zur Bewertung von Diensten: Facebook weiß, wie populär Dienste sind und wie sie sich entwickeln.

5.6 Übungsaufgaben

Aufgabe 1

Welche Dienste an Ihrer Universität/in Ihrem Unternehmen benötigen (Teil-) Identitäten von Studierenden/Mitarbeitern? Unter welchen Teil-Identitäten treten Sie gegenüber diesen Diensten auf? Wie authentifizieren Sie sich? Wie gelangen Sie an diese Teil-Identitäten? Berücksichtigen Sie auch Dienste, die nicht von der Universität/von dem Unternehmen selbst angeboten werden. Welche Eigenschaften sollte eine Identitätsmanagement-Lösung haben? Ist eine einheitliche Lösung für alle Dienste denkbar?

Literatur

1. Internet Engineering Task Force (IETF). The OAuth 2.0 Authorization Framework, Oct. 2012. Request for Comments (RFC) 6749.
2. Internet Engineering Task Force (IETF). The OAuth 2.0 Authorization Framework: Bearer Token Usage, Oct. 2012. Request for Comments (RFC) 6750.
3. OpenID. OpenID Authentication 2.0 - Final. http://openid.net/specs/openid-authentication-2_0.html, Dec. 2007. Final Specification.

Zusammenfassung

Nicht erst durch die Popularität von „Bitcoin“ – als vermeintlich anonymen Bezahlungssystem – in der jüngsten Vergangenheit steht das Thema *Anonymes Bezahlen* immer wieder auch im Fokus gesellschaftlicher Debatten, zuletzt im Rahmen der Diskussion um die Abschaffung des Bargeldes, der von Vielen als letzten Möglichkeit betrachteten Möglichkeit des anonymen Bezahlens überhaupt. Vielmehr bildeten anonyme Bezahlungssysteme, allen voran des von DAVID CHAUM entwickelten Ansatzes, die Grundlage für Privacy-Enhancing Technologies (PETs). Wir beschäftigen uns mit diesem „Meilenstein“ der PETs-Forschung in Abschn. 6.2, nachdem wir in Abschn. 6.1 allgemeine Anforderungen an anonyme Bezahlverfahren formulieren. Als nächstes wenden wir uns in Abschn. 6.3 dem Thema *Bitcoin* zu und gehen dabei vor allem auf die Frage ein, wie anonym Bitcoin wirklich ist. Schließlich betrachten wir in Abschn. 6.4 anonyme Bezahlverfahren in der Praxis.

Lernziele

Am Ende dieses Kapitels sollten Sie die Grundzüge anonymer Bezahlverfahren und die zugrundeliegenden kryptographischen Protokolle – die auch in anderen Anwendungen zum Einsatz kommen – kennen.

6.1 Anforderungen an ein anonymes Bezahlverfahren

Gerade in Bezug auf die unterschiedlichen Anforderungen, die sich nicht zuletzt durch die unterschiedlichen Beteiligten ergeben, ist das Thema *Anonymes Bezahlen* so spannend. Hier wird ganz besonders deutlich, dass neben der *Anonymität* auch das Thema *Sicherheit*

eine wichtige Rolle spielt. Anonymität ohne Sicherheit ist beim anonymen Bezahlen nicht möglich. Dies ist zwar auch in den meisten anderen Bereichen rund um den Datenschutz der Fall, wird aber von vielen gerne vergessen.

Anonymität gegenüber Bank

Zunächst stellt man an ein anonymes Bezahlverfahren die Anforderung, dass Nutzer gegenüber der Bank (oder dem Betreiber des Systems) „anonym“ agieren können. Dies heißt in diesem Fall, dass die Bank nicht erfahren soll, welcher Kunde Transaktionen mit bestimmten Händlern durchführt. Diese Forderung deckt sich mit unserer Definition von Anonymität aus Abschn. 2.2.2.

Zudem knüpft sich an diese Forderung die Anforderung nach *Unverkettbarkeit* verschiedener Transaktionen durch die Bank an. Die Bank soll also nicht erfahren, dass bestimmte Transaktionen – von der sie nicht weiß, wer daran beteiligt ist – von ein und demselben Kunden durchgeführt werden. Auch an dieser Stelle sei wieder an die Definition von „Unverkettbarkeit“ in Abschn. 2.2.2 erinnert.

Anonymität gegenüber Händlern

Zum einen wird gefordert, dass der Händler durch den Bezahlvorgang keine Identität des Kunden erfährt.

Zum anderen spielt auch hier wieder die Unverkettbarkeit der Transaktionen eines (anonymen) Kunden gegenüber verschiedenen Händlern eine Rolle.

Offline-Funktionalität

Ein anonymes Bezahlverfahren sollte keine ständige Verbindung zwischen Banken/Betreibern und Händlern erfordern.

Sicherheit

Es soll für Kunden nicht möglich sein, unter dem Deckmantel der Anonymität (digitales) Geld zu kopieren, d. h. mehrfach auszugeben (sogenanntes „Double Spending“).

Hierbei lässt sich ferner unterscheiden zwischen dem *Verhindern* von Betrug und dem nachträglichen *Aufdecken* eines Betrugs. Beim Verhindern geht es darum, dass technisch sichergestellt wird, dass ein mehrmaliges Ausgeben von digitalem Geld nicht möglich ist. Beim Aufdecken wird ein mehrmaliges Ausgeben zwar nicht verhindert, es soll aber sichergestellt werden, dass der Betrug nachträglich erkannt wird und der Betrüger mit hoher Wahrscheinlichkeit identifiziert werden kann.

6.2 Anonymes Bezahlen nach Chaum

Der 1985 von David Chaum präsentierte Ansatz [1] zum anonymen Bezahlen mit „elektronischem Geld“ ist der Klassiker und darf auch als Meilenstein für die weitere Entwicklung datenschutzfördernder Technologien gesehen werden.

Das Prinzip, das in leicht abgewandelten Formen auch bei anderen Verfahren zum Einsatz kommt, ist recht simpel. Der Nutzer generiert zunächst lokal elektronische

Münzen (Zufallszahlen). Er bereitet die elektronischen Münzen für die „blinde Signatur“ (siehe Abschn. 2.3.3) – ein Konzept das von Chaum entwickelt wurde und in einer Vielzahl von **PETs** zum Einsatz kommt – vor. Als Analogie für die blinde Signatur können wir uns vorstellen, dass der Nutzer einen Scheck und Kohlepapier in einen Briefumschlag legt. Danach signiert die Bank die elektronischen Münzen ohne Kenntnis des Inhalts (Analogie: Bank unterschreibt auf dem Umschlag). Der Nutzer extrahiert im Anschluss die Signatur der Münze (Analogie: Nutzer nimmt Scheck mit durchgedruckter Unterschrift aus dem Umschlag).

6.2.1 Verfahren im Überblick

Sehen wir uns nun das Verfahren, das von Chaum et al. [2] um ein Offline-Verfahren zur Erkennung von „Double Spending“ erweitert wurde, im Detail an.

Involvierte Parteien

Als involvierte Parteien treten *Nutzer*, eine *Bank* und *Händler* auf.

Setup-Phase

Während der Initialisierung des Systems generiert die Bank RSA-Schlüsselpaare

$$(sk_{bankwert} = (d_{wert}, N), pk_{bankwert} = (e_{wert}, N)).$$

Der jeweilige Geldwert, der durch dieses Schlüsselpaar repräsentiert wird, wird durch *wert* angegeben.

Außerdem werden zwei kollisionsresistente, kryptographische Hash-Funktionen (wie in Abschn. 2.3.4 beschrieben) f und g gewählt. Die Kontonummer des Nutzers bezeichnen wir mit *konto* und einen Zählwert mit v . Sowohl *konto* als auch v sind dem Nutzer und der Bank bekannt. Ein Wert k gibt den Parameter zur Offenlegung von Betrüger-Identitäten an. Dieser bestimmt, wie viele „Kandidaten“ später berechnet werden müssen.

Beziehen von Payment Token

Um ein *Payment Token* (**PT**) von seiner Bank zu beziehen, generiert der Nutzer zufällige Werte $a_i, c_i, d_i, r_i \pmod{N}$ für $1 \leq i \leq k$. Mit diesen Werten berechnet der Nutzer k „Kandidaten“ B_i , die alle seine Kontonummer *konto* enthalten:

$$B_i = r_i^{e_{wert}} \cdot f(x_i, y_i) \pmod{N}, \quad 1 \leq i \leq k, \text{ wobei}$$

$$x_i = g(a_i, c_i), \quad y_i = g(a_i \oplus (konto \parallel (v + i)), d_i).$$

Die Kandidaten werden „geblendet“, indem sie mit $r_i^{e_{wert}}$ multipliziert werden, so dass die Bank die eigentlichen Werte, die sie später blind signiert, nicht sehen kann. Der Nutzer übermittelt alle k Kandidaten zur Bank. Um der Bank zu beweisen, dass tatsächlich alle

Kandidaten seine Kontonummer *konto* enthalten, führen die beiden ein sogenanntes „Cut-and-Choose“-Protokoll durch.

Bei diesem Cut-and-Choose-Protokoll wählt die Bank zufällig $k/2$ Indizes $R \subseteq \{1, \dots, k\}$ aus, für die der Nutzer die zugehörigen Werte a_i , c_i , d_i und r_i offenlegen muss.¹ Da die Bank die Auswahl zufällig vornimmt, ist es unwahrscheinlich, dass der Nutzer betrügen kann, indem er nur genau die Hälfte der (von der Bank ausgewählten) Kandidaten mit seiner Kontonummer ausstattet und die andere Hälfte nicht. Wenn alle $k/2$ der von der Bank ausgewählten Kandidaten den Test bestehen (also tatsächlich die Kontonummer des Nutzers enthalten), signiert die Bank (blind) die anderen $k/2$ Kandidaten, die vom Nutzer nicht offengelegt wurden:

$$\begin{aligned} TPT'_{wert} &= \prod_{i \notin R} B_i^{d_{wert}} \pmod{N} = \prod_{i=1}^{k/2} (r_i^{e_{wert}} \cdot f(x_i, y_i))^{d_{wert}} \pmod{N} \\ &= \prod_{i=1}^{k/2} r_i \cdot f(x_i, y_i)^{d_{wert}} \pmod{N} \end{aligned}$$

Schließlich wird der Betrag *wert* vom Konto des Nutzers abgebucht und der Zählwert v erhöht. Die Bank sendet das temporäre Payment Token TPT' zum Nutzer, der die Blendfaktoren r_i , $1 \leq i \leq k/2$ entfernt, um das temporäre Payment Token TPT zu erhalten:

$$TPT_{wert} = TPT'_{wert} / \prod_{i=1}^{k/2} r_i \pmod{N} = \prod_{i=1}^{k/2} f(x_i, y_i)^{d_{wert}} \pmod{N}$$

Außerdem erhöht auch der Nutzer seinen Zählwert v und generiert und speichert das „finale“ Payment Token $PT_{wert} = (TPT_{wert}; a_i, c_i, d_i, \text{ für } 1 \leq i \leq k)$. TPT_{val} dient als eindeutiger Identifikator für ein PT.

Bezahlen mit einem Payment Token

Um einen Händler zu bezahlen, sendet der Nutzer TPT_{wert} (als Teil von PT_{wert}) an den Händler. Der Händler generiert daraufhin einen zufälligen Bitstring $z_1, z_2, \dots, z_{k/2}$ als „Challenge“ und sendet diesen zum Nutzer. Abhängig vom jeweiligen Wert von z_i , $1 \leq i \leq k/2$, antwortet der Nutzer auf diese Challenge folgendermaßen²:

- $z_i = 0$: In diesem Fall sendet der Nutzer x_i , $a_i \oplus (\text{konto} \parallel (v + i))$ und d_i zum Händler
- $z_i = 1$: In diesem Fall sendet der Nutzer a_i , c_i und y_i zum Händler.

¹Um die Notation einfach zu halten, nehmen wir an, dass die Bank $R = \{k/2 + 1, \dots, k\}$ auswählt.

²Sowohl die Challenge, als auch die Response werden in einem *Payment Transcript* gespeichert.

In beiden Fällen kann der Händler $f(x_i, y_i) \pmod{N}$ berechnen. Damit kann der Händler, sobald er die Antwort („Response“) vom Nutzer erhalten hat, die Authentizität von TPT_{wert} prüfen:

$$\prod_{i=1}^{k/2} f(x_i, y_i) \pmod{N} \stackrel{?}{=} TPT_{wert}^{e_{wert}} \pmod{N}. \quad (6.1)$$

Wenn die Gl. (6.1) gültig ist, ist TPT authentisch und der Händler akzeptiert die Bezahlung.

Einzahlen eines Payment Tokens

Um den Geldbetrag von der Bank ausbezahlt zu bekommen, für den ein Nutzer zuvor beim Händler bezahlt hat, muss der Händler sowohl das temporäre Payment Token TPT_{wert} als auch das Transkript (also die Mitschrift des zuvor durchgeführten Challenge-Response-Protokolls) an die Bank übermitteln. Die Bank überprüft nun, wie zuvor gezeigt (6.1), die Authentizität von TPT_{wert} . Wenn die Überprüfung in Ordnung ist, wird der Betrag $wert$ aus dem PT dem Konto des Händlers gutgeschrieben. Wurde das PT hingegen bereits zuvor schon einmal eingereicht, so ist die Bank nun in der Lage, den Betrüger aufzudecken.

Aufdecken eines Betrügers

Die Bank speichert jedes bei ihr eingereichte TPT_{wert} , sowie jedes Transkript (Abschrift des Challenge-Response-Protokolls). Sofern ein TPT_{wert} einlangt, das bereits zuvor eingereicht wurde, kennt die Bank nun zwei Mengen an Challenges und Responses. Da die Challenges von den Händlern zufällig gewählt werden, ist es sehr wahrscheinlich, dass ein $z_i \neq z'_i$, existiert, für das z_i zur Challenge des bereits gespeicherten Transkripts und z'_i zur Challenge des neuen Transkripts gehört. Nehmen wir an, $z_i = 0$ und $z'_i = 1$, dann kennt die Bank a_i, c_i, d_i, x_i, y_i und $a_i \oplus (konto \parallel (v + i))$. Nun ist die Bank in der Lage, die folgende Berechnung durchzuführen:

$$a_i \oplus (a_i \oplus (konto \parallel (v + i))) = (konto \parallel (v + i)).$$

Damit erhält die Bank die Kontonummer des Betrügers: *konto*.

6.2.2 Bewertung

Ein Betrüger kann bei dem Verfahren nach Chaum mit einer Wahrscheinlichkeit von $p_{Identifizierung} = (1 - (\frac{1}{2})^{k/2})$ identifiziert werden, da sich in diesem Fall die eingereichten Challenges von einem zweimal eingereichten Payment Token an mindestens einer Stelle unterscheiden. Eine Identifizierung des Betrügers durch die Offenlegung seiner Kontonummer ist bereits bei einer Challenge-Länge von 2 zu 75 % möglich.

Sofern der Nutzer Payment Tokens nur einmal ausgibt bleibt seine Anonymität sowohl gegenüber dem Händler als auch gegenüber der Bank gewahrt.

6.3 Bitcoin

In den Medien wird immer wieder über die „anonyme Währung“ Bitcoin berichtet.³ Um zu verstehen, wie anonym Bitcoin wirklich ist, müssen wir zunächst verstehen, wie Bitcoin funktioniert. Tatsächlich handelt es sich bei Bitcoin im Kern um ein Verfahren zur Abwicklung von Überweisungen. Diese Überweisungen lauten allerdings nicht auf Euro oder eine sonstige Währung, sondern verwenden ein eigenes Wertmaß (eine eigene *Rechnungseinheit*). Wie das Verfahren selbst heißt das Wertmaß ebenfalls Bitcoin oder abgekürzt BTC. Bitcoin verzichtet – zumindest seinem Grundkonzept nach – vollständig auf zentrale oder besonders vertrauenswürdige Instanzen.

Wer Bitcoin verwenden will, generiert sich zunächst ein Schlüsselpaar für ein digitales Signaturverfahren. Ein Hashwert des öffentlichen Schlüssels wird zur Bitcoin-Adresse, die die Funktion einer Kontonummer hat. Wer von seinem eigenen Konto eine Überweisung auf ein anderes Konto tätigen möchte, veröffentlicht die Überweisungsdaten (den Betrag, die Zieladresse und den öffentlichen Schlüssel, der zur Absenderadresse gehört) zusammen mit einer digitalen Signatur. Für die Erstellung der Signatur verwendet er den privaten Schlüssel, der zur Absenderadresse gehört.

Die Grundidee hinter Bitcoin ist also sehr einfach. Um aus dieser Grundidee ein praktikables Bezahlssystem zu machen, müssen aber noch zwei Probleme gelöst werden: Erstens muss der Empfänger einer Überweisung sicher sein, dass vorher überhaupt genug auf das Konto des Überweisenden eingezahlt wurde. Zweitens muss sichergestellt sein, dass der Betrag nicht schon vorher wieder ausgegeben wurde – also kein Geldbetrag zweimal ausgegeben wird. Zusammengefasst geht es also darum, die Deckung des Kontos sicherzustellen.

Das erste Problem wird in Bitcoin wiederum sehr einfach gelöst. Alle Transaktionen sind öffentlich: Wer eine Bitcoin-Transaktion durchführen will, nimmt an einem Peer-to-Peer-Netz teil. Er muss einige andere Teilnehmer des Bitcoin-Systems kennen und schickt diesen die Transaktion; diese leiten sie wiederum an weitere Teilnehmer weiter. Eine neue Transaktion enthält Verweise auf die zugehörigen eingehenden Transaktionen, die ebenfalls signiert sind und überprüft werden können. Schwieriger ist jedoch, sicherzustellen, dass das „Geld“ nicht mittlerweile wieder ausgegeben wurde. Eine Signatur alleine kann das nicht leisten. Vielmehr muss sichergestellt werden, dass eine Transaktion, mit der Geld ausgegeben wurde, nicht einfach versteckt werden kann. Es genügt also nicht, wenn alle

³Tatsächlich ist Bitcoin – wie wir sehen werden – nicht im technischen Sinne anonym; juristisch betrachtet handelt es sich auch nicht um eine Währung [3].

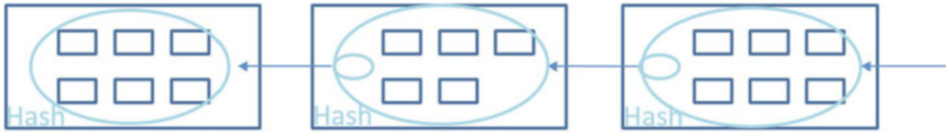


Abb. 6.1 Das Grundkonzept der Blockchain

Transaktionen öffentlich sind – vielmehr müssen sich die Teilnehmer des Bitcoin-Systems auf eine einzige Liste aller Transaktionen in einer festgelegten Reihenfolge *einigen*.

Um dies zu erreichen, wird das Konzept der Blockchain eingesetzt. Transaktionen werden in *Blöcken* gesammelt, die an eine Kette (die *Blockchain*) angehängt werden. Jeder Block enthält einen Hashwert des vorhergehenden Blockes (Abb. 6.1). Das bedeutet, dass jede Änderung eines Blocks alle folgenden Blöcke ungültig macht. Will ein Angreifer eine alte Transaktion verstecken, mit der er „Geld“ ausgegeben hat, muss er diese aus einem alten Block entfernen und alle folgenden Blöcke anpassen. Damit er das nicht tun kann, hat Bitcoin eine weitere Hürde eingebaut: Ein Block ist nur gültig, wenn ein *kryptographischer Arbeitsbeweis* erbracht wurde. Der Beweis ist so ausgestaltet, dass er nur mit Kenntnis aller Transaktionen des aktuellen Blocks sowie des Hashwerts des vorherigen Blocks sowie durch die Investition von sehr viel Rechenleistung erbracht werden kann. Wer einen Arbeitsbeweis erfolgreich erbracht hat, darf sich neben einer Transaktionsgebühr auch einen gewissen Bitcoin-Betrag „aus dem nichts“ gutschreiben.

Zusätzliche Information

Wie kann man einen solchen Arbeitsbeweis nun realisieren? Bitcoin verwendet das Konzept eines *Hash-basierten Arbeitsbeweises*. Dahinter steckt folgende Überlegung: Wenn Sie mit einer gegebenen kryptographischen Hashfunktion den Hashwert eines zufälligen Wertes berechnen, steht in der Binärdarstellung an der ersten Stelle mit einer Wahrscheinlichkeit von 0,5 die Ziffer 0. Für jede weitere Ziffer gilt das Gleiche (wobei der Wert jeder Stelle als Ergebnis eines unabhängigen Zufallsexperiments betrachtet werden kann). Wäre es anders, ließen sich beispielsweise Kollisionen effizienter finden, als dies bei einer kryptographischen Hashfunktion erlaubt ist. Die Wahrscheinlichkeit, dass die ersten n Stellen 0 sind, ist $0,5^n$.

Will Alice nun einen kryptographischen Arbeitsbeweis erbringen und zugleich die Kenntnis gewisser Daten (bei Bitcoin der Hashwert des vorherigen Blocks sowie alle Transaktionen für den neuen Block) nachweisen, kann sie wie folgt vorgehen: Sie nimmt die bekannten Daten bzw. deren Hashwert und hängt eine weitere, selbstgewählte Zeichenkette c an. Dann berechnet sie den Hashwert über alles. Sind die ersten n Stellen 0, gilt der Arbeitsbeweis als erbracht. Andernfalls ersetzt Alice die Zeichenkette c durch eine neu gewählte c' und versucht die Berechnung erneut. Dies wiederholt sie so oft, bis die ersten n Stellen 0 sind. In der Praxis wird sie außerdem abbrechen, falls ein anderer Teilnehmer den Arbeitsbeweis erbracht hat. Wenn sie eine neue Transaktion empfängt, kann sie diese problemlos mit aufnehmen und danach auf dem geänderten Datensatz weiterrechnen. Die Erfolgswahrscheinlichkeit ist bei jedem einzelnen Versuch gleich groß – unabhängig davon, wie viele Versuche vorher auf den gleichen Daten durchgeführt wurden.

Die Schwierigkeit des Arbeitsbeweises lässt sich durch die Wahl von n steuern. Bei Bitcoin passiert diese Steuerung regelmäßig und zielt darauf ab, dass ein Arbeitsbeweis für einen neuen Block im Durchschnitt ca. alle 10 Minuten gefunden wird – unabhängig davon, wie viel Rechenleistung gerade insgesamt investiert wird. Auch bei großem n ist es theoretisch denkbar –

nur sehr unwahrscheinlich –, dass ein Teilnehmer bereits im ersten Versuch den Arbeitsbeweis erbringt. Der Arbeitsbeweis ist letztlich eine Art Lotterie, bei der sich die Anzahl der Lose aus der Rechenleistung ergibt. Wie bei Lotterien gibt es auch Tippgemeinschaften (sogenannte Mining Pools): Die geringe Wahrscheinlichkeit für den Einzelnen, einen hohen Gewinn zu erzielen, wird dabei eingetauscht gegen eine hohe Wahrscheinlichkeit (da die Auszahlung erfolgt, sobald einer der Teilnehmer einen Arbeitsbeweis erbracht hat) auf einen kleinen Gewinn (da der Gewinn auf alle Teilnehmer aufgeteilt wird). Mining Pools sind auch ein Risiko, da sie bei zentraler Koordination die Dezentralitätsannahme von Bitcoin verletzen; ein einzelner Mining Pool kann genug Rechenleistung auf sich vereinen, um theoretisch alleine die Transaktionsgeschichte ändern zu können.

Natürlich kann ein Angreifer dennoch eine Manipulation versuchen. Stehen verschiedene Varianten der Blockchain zur Auswahl, gilt die längste als korrekt. Wenn der Angreifer also schneller Arbeitsbeweise erbringt als alle anderen Teilnehmer des Bitcoin-Systems zusammengenommen, kann er seine eigene Blockchain durchsetzen. Will er außerdem alte Blöcke manipulieren, muss er zusätzlich die Rechenleistung für alle seitdem erbrachten Arbeitsbeweise aufbringen.

In der praktischen Umsetzung sind noch einige Ergänzungen und Abweichungen von den beschriebenen Grundprinzipien hinzugekommen, die zu erläutern an dieser Stelle aber zu weit führen würde.

6.3.1 Anonymität von Bitcoin

Bezüglich der mit Bitcoin erreichten Anonymität lassen sich drei Ebenen unterscheiden: Das abstrakte Konzept von Bitcoin, die Umsetzung in einem Peer-to-Peer-System und die Schnittstellen nach außen.

Bereits im Bitcoin-Konzept wird deutlich, dass das System nicht mit dem Ziel der Anonymität entworfen wurde. Jede Transaktion ist dauerhaft öffentlich. Es lässt sich also nachverfolgen, wann welcher Betrag an welche Bitcoin-Adresse überwiesen wurde. Jeder Bitcoin-Teilnehmer kann sich beliebig viele Adressen zulegen. Sobald er aber Transaktionen zwischen diesen Adressen vornimmt, wird eine Verbindung auch für Dritte sichtbar. Das gleiche gilt, wenn er in einer Transaktion verschiedene Adressen als Absenderkonten verwendet, weil die auf einzelnen Konten verfügbaren Beträge nicht ausreichen. Dass Bitcoin als anonym bezeichnet wird, liegt lediglich daran, dass Bitcoin selbst ausschließlich mit den Adressen, nicht aber den zugehörigen Namen arbeitet. Wer aber, etwa als Empfänger von Spenden, eine Bitcoin-Adresse öffentlich bekanntgibt, gibt damit auch diese beschränkte Anonymität preis.

Mit der Umsetzung in einem Peer-to-Peer-System kommt ein weiteres Problem hinzu: Wer eine Transaktion ausführt, muss diese an andere Teilnehmer des Bitcoin-Systems propagieren. Sie wird dann weitergeleitet, bis sie im gesamten Netz bekannt ist. Wer den Ausgangspunkt dieser Weiterleitungskaskade identifiziert, gelangt also an den Initiator der Transaktion bzw. zumindest an dessen IP-Adresse. Damit dies mit hoher Erfolgswahrscheinlichkeit gelingt, müsste ein Angreifer einen erheblichen Anteil des

gesamten Peer-to-Peer-Systems beobachten können. Da die Größe des Netzes und die Anzahl der Transaktionen in Bitcoin aber im Verhältnis zu den Möglichkeiten aktueller Rechner sehr überschaubar sind, ist dies auch schon mit einem geringen Budget möglich.

Wer Bitcoin möglichst anonym einsetzen möchte, sollte schließlich auch die Schnittstellen nach „außen“, letztlich also in die Realwirtschaft, bedenken. Nur selten ist es möglich, Waren oder Dienstleistungen direkt mit Bitcoin zu bezahlen. Es gibt Anbieter, die Bitcoin in Währungen wie US-Dollar oder Euro umtauschen; wenn diese den entsprechenden Betrag auf ein Bankkonto überweisen, geht auch damit letztlich die Anonymität verloren. Es gab diverse spektakuläre Kriminalfälle, in denen (in Euro umgerechnet) Millionenbeträge in Bitcoin erbeutet wurden. Es dürfte den Tätern besonders schwer fallen, diese Beträge unauffällig und anonym umzutauschen.

6.4 Anonymes Bezahlen in der Praxis

Bargeld bietet in der Praxis immer noch das höchste Maß an Anonymität. Durch die Verlagerung von Geschäften in die „Online-Welt“ wurde der Ruf nach Anonymität auch für das elektronische Bezahlen laut. Allerdings können die heute in der Praxis angebotenen elektronischen Bezahlverfahren diese Erwartungen nicht erfüllen.

David Chaum hat 1990 ein Unternehmen namens *DigiCash* gegründet, das das in Abschn. 6.2 vorgestellte Verfahren in der Praxis umsetzen sollte. Einige Banken haben das Verfahren zeitweise auch angeboten. Allerdings stellte sich kein wirklicher Erfolg in der Praxis ein und so musste DigiCash 1998 Insolvenz anmelden.

Bitcoin, das wir in Abschn. 6.3 im Detail untersucht haben, ist derzeit das einzige elektronische Bezahlverfahren mit einem gewissen Grad an Anonymität, das den Weg heraus aus der Forschung in die Praxis gefunden hat und sich langsam etabliert.

Am häufigsten verbreitet als (mehr oder weniger) anonyme Bezahlverfahren neben Bargeld sind (in Deutschland) die Geldkarte, sowie Prepaid-Karten.

6.4.1 Geldkarte

Die meisten heutzutage in Deutschland von den Banken ausgegebenen Bankkarten („girocards“) sind mit der „GeldKarte“-Funktion ausgestattet, die ein bargeldloses Bezahlen von Kleinbeträgen an Automaten erlaubt. Die Geldkarte wird gelegentlich als anonymes Bezahlverfahren bezeichnet. Geldkarten können kontenungebunden ausgegeben werden; die meisten Banken binden die Geldkarte allerdings an ein Girokonto. Die Geldkarte kann sowohl an Geldautomaten, als auch über das Internet aufgeladen werden. Dabei wird der aufgeladene Geldbetrag auf der Bankkarte – einer Smartcard – gespeichert. Bei jedem Bezahlvorgang, der in der Regel offline durchgeführt wird, wird der gespeicherte Betrag auf der Geldkarte verringert. Alle Transaktionen werden (i. d. R. zeitverzögert) bei der Bank auf einem Schattenkonto nachvollzogen. Dies dient der

Missbrauchserkennung. Die Anonymität wird durch eine organisatorische Trennung von Schattenkonto und Inhaber-Identität gewährleistet. Die Schattenkontendaten bleiben über Jahre hinweg gespeichert. Die größtmögliche Anonymität wird bei der Verwendung einer kontenungebundenen Geldkarte erreicht, die mittels Barzahlung bei der Bank aufgeladen wird, so die Bremer Landesbeauftragte für den Datenschutz.⁴

6.4.2 Prepaid-Karten

In den letzten Jahren sind Prepaid-Karten wie die *Paysafecard* und Karten in Form von Gutscheinen, die bspw. in Supermärkten verkauft werden, sehr populär geworden. Kunden können solche Karten anonym mit Bargeld kaufen. Auf der Karte ist eine Seriennummer aufgedruckt. Über diese Seriennummer ist die Karte beim Betreiber mit einem Konto verknüpft. Während des Bezahlvorgangs gibt der Kunde beim Händler die Seriennummer an. Der Händler gibt die Seriennummer an den Betreiber und der Betreiber bucht den angefragten Betrag vom Konto ab. Der Restbetrag kann vom Kunden bei anderen Händlern ausgegeben werden.

Prepaid-Karten stellen ein einfaches, praxistaugliches Online-Verfahren zur anonymen Bezahlung dar. Voraussetzung ist, dass die Händler vertrauenswürdig sind.

6.5 Fazit

Insgesamt sind die derzeit praktisch angebotenen digitalen Bezahlverfahren wesentlich weniger elegant als etwa das Verfahren nach CHAUM, das wir in diesem Kapitel betrachtet haben. Offline-Verfügbarkeit und Anonymität sind Eigenschaften, die meist nicht geboten werden.

Die Kryptowährung *Bitcoin* hingegen erfreut sich großer Beliebtheit – allen großen Kursschwankungen in den letzten Jahren zum Trotz. Neben zahlreichen (Online-) Shops akzeptiert seit Juli 2016 auch die Stadt Zug in der Schweiz das Bezahlen von Gebühren mittels Bitcoins – vorerst in einem Pilotprojekt bis Ende 2016.⁵ Stand die Finanzbranche der neuen Kryptowährung zunächst skeptisch gegenüber, gibt es heute wohl kaum eine Bank die sich nicht mit dem Thema beschäftigt. Vor allem der Bitcoin zugrundeliegenden *Blockchain*-Technologie wird eine große Zukunft vorausgesagt: von der Anwendung im Internet der Dinge, über die Nutzung zur Führung von digitalen Grundbüchern bis hin

⁴<https://ssl.bremen.de/datenschutz/sixcms/detail.php?gsid=bremen236.c.3881.de> (Zugegriffen am 05.11.2016)

⁵<http://www.nzz.ch/schweiz/crypto-valley-zukunftsmodell-oder-marketing-gag-ld.22911> (Zugegriffen am 05.11.2016)

zur Unterstützung der Aushandlung und Abwicklung von Verträgen, sogenannter „Smart Contracts“, ist die Rede.⁶

6.6 Übungsaufgaben

Aufgabe 1

Bei dem vorgestellten Bezahlverfahren nach Chaum ist es möglich, dass ein Kunde mit einem Händler kollaboriert, um zu betrügen. Der Kunde wickelt dabei eine Bezahlung mit einem Händler ordnungsgemäß ab. Danach gibt der Kunde die Challenge dieses Händlers an den zweiten Händler und bezahlt mit der gleichen Münze. Kann die Bank den betrügenden Kunden identifizieren? Kann die Bank den betrügenden Händler identifizieren? Welcher Lösungsansatz könnte dieses Problem verhindern?

Literatur

1. David Chaum. Security without identification: transaction systems to make big brother obsolete. *Communications of the ACM*, 28(10):1030–1044, Oct. 1985.
2. David Chaum, Amos Fiat, and Moni Naor. *Advances in Cryptology — CRYPTO' 88: Proceedings*, chapter Untraceable Electronic Cash, pages 319–327. Springer New York, New York, NY, 1990.
3. Christoph Sorge und Artus Krohn-Grimberghe. Bitcoin: Eine erste Einordnung. *Datenschutz und Datensicherheit*, 36(7):479–484, 2012.

⁶http://diepresse.com/home/wirtschaft/economist/wertsachen/5112939/Blockchain_Die-zweite-Phase-der-digitalen-Revolution (Zugegriffen am 05.11.2016)

Zusammenfassung

Das Thema Datenschutz verbinden viele Menschen zunächst mit dem *World Wide Web*. Täglich nutzen wir Social Media-Dienste und kaufen in Online-Shops ein. Zielgerichtete Werbung ist ein ständiger Begleiter. In diesem Kapitel beschäftigen wir uns mit dem Thema Datenschutz im World Wide Web. Wir werden in Abschn. 7.1 sehen, dass das *Tracking* im Web mittels „Cookies“ und „Tracking-Pixel“ fast allgegenwärtig zu sein scheint. Ebenso dienen *Social Plugins*, denen wir uns in Abschn. 7.2 zuwenden, dem Verfolgen von Nutzeraktivitäten im Web.

Lernziele

Am Ende dieses Kapitels sollten Sie die heutzutage verwendeten Technologien zum „Tracking“ von Nutzern im World Wide Web kennen.

7.1 Tracking im Web

Tracking bezeichnet das Verfolgen von Nutzeraktivitäten (im Web) und das Bilden von Nutzungsprofilen. Die Ziele sind u. A. das Bilden von Statistiken über Webseitenzugriffe und das Anbieten zielgerichteter Werbung. Gerade das Anbieten zielgerichteter Werbung erfordert eine genaue Kenntnis der Nutzer-Interessen. Diensteanbieter können umso genauere Nutzer-Profile bilden, je weiter sie Nutzer im Web verfolgen können. Längst erfolgt das Verfolgen von Nutzern über Website-Grenzen (von unterschiedlichen Diensteanbietern) hinweg. Verantwortlich dafür sind spezielle Werbe-Dienstleister, die mit den unterschiedlichen Diensteanbietern kooperieren. Nur so ist es möglich, dass ein in einer Online-Gebrauchtwagenbörse gesuchtes Auto einen Nutzer weiter begleitet, etwa

in Form einer Werbeeinblendung auf einer Nachrichten-Seite. Für das Tracking stehen unterschiedliche Verfahren zur Verfügung, die wir uns im Folgenden genauer ansehen werden.

7.1.1 Cookies

Cookies stellen die derzeit (noch) am häufigsten verwendete Methode der Wiedererkennung von Nutzern im Web dar. Zunächst müssen wir festhalten, dass es sich bei *Hypertext Transfer Protocol (HTTP)*, dem Protokoll, das zur Übertragung von Webseiten verwendet wird, um ein zustandsloses Protokoll auf Anwendungsebene handelt. Zustandslos bedeutet, dass für den Webserver jede Anfrage eines Nutzers eine eigene Sitzung darstellt, d. h. es ist nicht vorgesehen, dass der Webserver Nutzer wiedererkennen kann. Eine Anfrage eines Browsers an den Webserver (sogenannter „HTTP-Request“) enthält u. A. folgende Informationen:

- GET/Host: Welche Webseite von welchem Server abgerufen werden soll,
- User Agent: Angaben zum verwendeten Browser,
- Accept-Language: Bevorzugte Sprache(n),
- Referer: Welche Seite der Benutzer zuletzt besucht hat.

Wie wir später in Abschn. 7.1.3 noch sehen werden, stellen ungewöhnliche Kombinationen dieser Informationen einen *Quasi-Identifikator* dar.

Zusätzlich zu den genannten Informationen ist dem Webserver zusätzlich die IP-Adresse des Nutzers und der Anfragezeitpunkt bekannt.

Um die Zustandslosigkeit des HTTP-Protokolls zu umgehen, wurden HTTP-Cookies entwickelt. Hierbei handelt es sich um Textdateien,¹ die vom Server erzeugt werden und in der „HTTP-Response“ (als Antwort auf einen HTTP-Request) über den HTTP-Befehl „Set-Cookie“ an den Client übermittelt werden. Beim erneuten Aufruf der (selben) Website wird das Cookie vom Client wieder an den Server übermittelt. Das Cookie dient der Zustandshaltung auf Clientseite. Es ermöglicht so, Sitzungen zu realisieren. Nur so ist es erst möglich, dass bspw. ein Online-Shop realisiert werden kann. Ohne ein Sitzungsmanagement wären etwa die Schritte „Produkt auswählen“, „Produkt in den Warenkorb geben“ und „Produkt bezahlen“ für den Webserver nicht einem einzelnen Nutzer zuordenbar, da es sich für ihn um unabhängige HTTP-Requests handelt. Cookies sind in RFC 2109 (veraltet, aber noch in Gebrauch) und RFC 2965 spezifiziert. Eine Obergrenze für die Anzahl akzeptierter Cookies ist nicht vorgeschrieben. RFC 2965 empfiehlt aber, dass ein Client mindestens 300 Cookies speichern können soll, darunter

¹Es handelt sich nicht um ausführbaren Code, weshalb Cookies auch nicht zum Infizieren eines Clients mit Schadsoftware missbraucht werden können, wie häufig fälschlicherweise behauptet wird.

mindestens 20 pro Host bzw. pro Domain. Ein Cookie sollte dabei mindestens 4096 Byte groß sein können.

Datenschutz-Problematik

Aus Datenschutz-Sicht stellen Cookies aufgrund ihrer Verwendung zum Tracking, d. h. zum Nachverfolgen der Nutzeraktivitäten im Web, ein „Problem“ dar. Zunächst müssen wir unterscheiden zwischen einem Tracking innerhalb einer Website, wie es etwa beim zuvor angesprochenen Online-Shop-Beispiel genutzt wird und einem Tracking über mehrere Websites (von unterschiedlichen Diensteanbietern) hinweg.

Beim Tracking innerhalb einer Website handelt es sich, wie wir zuvor gesehen haben, durchaus um eine nützliche Funktion. Das Unterbinden des Setzens von (Erstanbieter)-Cookies im Browser kann also dazu führen, dass viele Webanwendungen nicht mehr funktionieren.

Das Tracking von Nutzern über Website-Grenzen hinweg birgt jedoch das Potential, ein vollständiges Nutzerprofil bilden zu können. Grundsätzlich ist ein solches Tracking eigentlich nicht vorgesehen. Die sogenannte *Same Origin Policy* (*SOP*) sorgt dafür, dass Cookies nur an den Ursprungsserver zurückgeschickt werden, also zu jenem Server, der das Cookie auch gesetzt hat. Diese SOP-Policy wird in der Praxis jedoch dadurch umgangen, dass Webseiten externe Inhalte von Dritten (bspw. von Werbedienstleistern) einbinden und dabei die externen Server anstoßen, Cookies zu setzen und zu lesen.

Werbedienstleister werden von einer Vielzahl von unterschiedlichen Diensteanbietern genutzt. Bei jedem Seitenaufruf eines Nutzers wird ein Cookie, das zuvor von einem Werbedienstleister beim Besuch einer anderen Webseite gesetzt wurde und einen Identifikator enthält, an den Werbedienstleister übermittelt. Der Werbedienstleister erhält dadurch die Information, welcher Nutzer (repräsentiert durch den Identifikator im Cookie) sich für die aufgerufene Webseite interessiert. Die aufgerufene Webseite „sagt“ einiges über die Interessen des Nutzers aus und somit ist der Werbedienstleister in der Lage, den Nutzer, je länger er ihn im Web beim Surfen „begleitet“, besser „kennenzulernen“, d. h. ein Nutzerprofil zu bilden. Ruft ein Nutzer häufig die Fußballseiten von Nachrichtenportalen, sowie Automobil-Portale auf, kann der Werbedienstleister daraus schließen, dass es sich mit großer Wahrscheinlichkeit um einen männlichen Nutzer handelt. Diese Information nutzt der Werbedienstleister später für *zielgerichtete Werbung* aus. Er wird den Diensteanbietern, die ebenfalls mit dem Werbedienstleister kooperieren, jene Werbebanner ausliefern – die dort dann auf den Webseiten für diesen Nutzer erscheinen – die für den Nutzer höchstwahrscheinlich interessant sind. Mitunter kann es sich dabei auch um Werbung für Produkte handeln, die sich der Nutzer zuvor in einem Online-Shop schon einmal angesehen hat, die er aber noch nicht gekauft hat.

Eine Analyse der 100 meist-besuchten Websites hat ergeben, dass viele Diensteanbieter jeweils mit über 100 dieser Werbedienstleister zusammenarbeiten. DOUBLE CLICK – mittlerweile in Besitz von Google – ist einer der größten Werbedienstleister weltweit.

Schutz vor Cookies

Das Nachladen von Server-fremden (Drittanbieter-)Cookies kann in den Browser-Einstellungen verhindert werden. Außerdem bieten die meisten modernen Browser eine *Do Not Track (DNT)*-Funktion. Ist diese Funktion aktiviert, so meldet der Browser beim Besuchen einer Website dem Diensteanbieter (über einen HTTP-Befehl), dass er nicht verfolgt werden möchte. Der Diensteanbieter muss diesem Wunsch nachkommen. Diese Einstellung bewirkt übrigens nicht, dass etwa keine Werbung mehr angezeigt wird, sondern lediglich, dass die angezeigte Werbung nicht mehr auf den Informationen basierend auf dem Surf-Verhalten basiert. Daneben gibt es bei vielen Browsern die Möglichkeit, im sogenannten „privaten Modus“ zu surfen. Dabei werden Cookies von Drittanbietern blockiert. Außerdem werden im privaten Modus keine Chronik und keine Anmeldeinformationen gespeichert. Zudem können Nutzer bei (seriösen) Werbedienstleistern selbst bestimmen, ob sie zielgerichtete Werbung bekommen möchten. Problematisch dabei ist, dass in einigen Fällen dieses „Opt-out“ durch das Setzen eines Opt-out-Cookies umgesetzt wird. Beim Löschen aller Cookies ist das Opt-out damit nicht mehr erkennbar. Beim Werbedienstleister Google lässt sich das Opt-out in den „Google Ads Settings“ unter <https://www.google.de/settings/ads> vornehmen.

Die *Electronic Frontier Foundation (EFF)*, eine Nichtregierungsorganisation in den USA die sich für Grundrechte im Netz stark macht, hat mit „Privacy Badger“ ein Browser-Add-On bereitgestellt, der alle „Tracker“ auf einer Webseite blockiert. Nicht-zielgerichtete Werbung wird weiterhin angezeigt. Zudem lohnt sich ein Blick auf eine Webseite mit dem Browser-Add-On „Ghostery“. Es zeigt, wie viele Tracker sich auf einer Webseite befinden, d. h. an wie viele Werbedienstleister das Surf-Verhalten übermittelt wird. Gerade bei Nachrichten-Portalen wird die Liste dabei sehr lang.

Flash-Cookies

Flash-Cookies (auch „Local Shared Objects“ genannt und an der Dateiendung .sol zu erkennen) ähneln HTTP-Cookies. Gesetzt werden Flash-Cookies von Servern von Websites mit eingebundenen Flash-Inhalten. Für die Verwaltung ist nicht der Browser des Nutzers zuständig, sondern die Flash-Player-Anwendung. Damit sind Flash-Cookies browserunabhängig und können auch nicht über den Browser gelöscht werden. Das Nutzen unterschiedlicher Browser für unterschiedliche Websites, um Website-übergreifendes Tracking einzuschränken, greift damit ins Leere. Egal mit welchem Browser die Website aufgerufen wird, die Flash-Player-Anwendung sendet das gespeicherte Cookie mit. Flash-Cookies haben kein Verfallsdatum und keine Größenbegrenzung von 4 KB. Aus Datenschutzsicht sind Flash-Cookies also weitaus problematischer als „normale“ HTTP-Cookies. Die abnehmende Verwendung von Flash im Web² sorgt aber dafür, dass Flash-Cookies in Zukunft eine eher untergeordnete Rolle spielen werden.

²Google hat 2015 angekündigt, bei YOUTUBE in Zukunft auf HTML5 anstatt auf Flash zu setzen. Grund dafür sind die zahlreichen Sicherheitslücken im Flash Player.

7.1.2 Tracking-Pixel

Eine weitere, gängige Methode zum Verfolgen von Nutzern im Web stellen sogenannten *Tracking-Pixel* (auch „Web Bugs“ genannt) dar. Hierbei handelt es sich um 1x1 Pixel große, nicht sichtbare, Abbildungen auf einer Webseite. Beim Aufruf der Webseite wird das Tracking-Pixel vom (externen) Server nachgeladen. Der externe Server, etwa von einem Werbedienstleister betrieben, loggt den Abruf. Heute wird meist ein JavaScript-Code verwendet. Dies hat (aus Sicht des Trackers) den Vorteil, dass zusätzliche Informationen wie z. B. die Bildschirmauflösung des Clients in die URL, die beim Abruf des Bildes verwendet wird, mit eingebettet werden können.

Tracking Pixel finden in zahlreichen Szenarien Anwendung. Bei der „Clickstream“-Analyse werden die Seitenaufrufe eines Nutzers verfolgt. Bei bezahlter Werbung kann mittels Tracking Pixel geprüft werden, über welche Seite ein Nutzer ein Produkt gekauft hat. Auch bei E-Mails kommen Tracking Pixel zum Einsatz. So können Spammer prüfen, ob eine E-Mail-Adresse gültig ist. Werbedienstleister können prüfen, ob ein Nutzer eine E-Mail tatsächlich gelesen hat. Das automatische Nachladen von Abbildungen durch E-Mail-Clients wird heute aber in den meisten Fällen unterbunden. Auch für ein „singles“ Zählen der Aufrufe einer Webseite werden Tracking Pixel verwendet.

Google Analytics

Einer der bekanntesten Dienste, der Tracking Pixel verwendet ist *Google Analytics*. Google Analytics wird zur Zugriffsanalyse verwendet. Es erlaubt den Betreibern einer Website etwa zu erfahren, woher Nutzer kommen, welche Browser-Versionen sie verwenden, wie lange sie auf welchen Webseiten verweilen etc. Dazu binden die Betreiber einen von Google bereitgestellten JavaScript-Code, der eine spezifische Website-Betreiber-ID enthält, in ihre Webseiten ein. Beim Laden der Webseite durch einen Nutzer wird ein Tracking-Pixel (.gif-Datei) von Google nachgeladen. Der Aufruf sieht (verkürzt) folgendermaßen aus: http://www.google-analytics.com/__utm.gif?utmwv=4&utmn=769876874&utmhn=example.com&utmcs=ISO-8859-1&utmsr=1280x1024&utmsc=32-bit&utmcl=en-us&utmje=1&utmfl=9.0%20%20r115&utmcn=1&utmdt=GATC012%20setting%20variables&utmhid=2059107202&utmr=0&utmp=/auto/GATC012.html?utm_source=www.gatc012.org&utm_campaign=campaign+gatc012&utm_term=keywords+gatc012&utm_content=content+gatc012&utm_medium=medium+gatc012&utmcc=__utma%3D97315849.1774621898.1207701397.1207701397.1207701397.1%3B...³ Die gesammelten Daten (Browser-Informationen, System-Informationen, Cookies, zuvor besuchte Webseite etc.) werden über diesen Aufruf der GIF-Datei als Parameter an Google übermittelt und von Google gespeichert. Bei Online-Shops fließen zudem die Informationen mit ein, welches Produkt

³Die Beispiel-URL stammt von <https://developers.google.com/analytics/resources/concepts/gaConceptsTrackingOverview#howAnalyticsGetsData>.

(Stückzahl, Preis, Größe, Farbe etc.) ein Nutzer in den Warenkorb gegeben hat bzw. auch tatsächlich gekauft hat. Nutzt ein Betreiber *Remarketing* basierend auf Google Analytics, werden auch DoubleClick Cookies als 3rd-Party Cookies mit an Google Analytics übermittelt und gesetzt. Remarketing erlaubt die zielgruppenorientierte Werbung für Produkte basierend auf dem Surf-Profil der Nutzer.

Aus Datenschutzsicht ist der Einsatz von Google Analytics problematisch. Google ist in der Lage, sehr detaillierte Nutzungsprofile – auf Basis der IP-Adressen der Nutzer – anzulegen. Nutzen Nutzer einen Google Account und sind dort mit ihren tatsächlichen Namen registriert, so lassen sich die Nutzungsprofile eindeutig einer bestimmten Person zuordnen. Viele Nutzer sind sich nicht bewusst, dass beim Besuch einer Website personenbezogene Daten an einen Dritten, in diesem Fall Google, übermittelt werden.

7.1.3 Device Fingerprinting

Device Fingerprinting (bzw. hier im Spezialfall Browser Fingerprinting) bezeichnet eine Methode, mittels derer ein Website-Betreiber bzw. Werbedienstleister versucht, seine Nutzer zu identifizieren, d. h. in diesem Kontext, sie von anderen zu unterscheiden.

Die Idee von Device Fingerprinting besteht darin, die Browser der Nutzer aufgrund ihrer unterschiedlichen Konfigurationen zu unterscheiden; möglichst jedem Browser einen eindeutigen Fingerabdruck zuzuweisen, der beständig ist, sich im Lauf der Zeit also nicht all zu stark verändert. Unterschiedliche Studien gehen davon aus, dass Browser-Konfigurationen zu mehr als 80 % unterscheidbar sind, also eine über einen längeren Zeitraum einzigartige Konfiguration aufweisen. Die Electronic Frontier Foundation (EFF) stellt unter <https://panopticklick.eff.org> einen Online-Test zur Verfügung, bei dem Nutzer ihren Browser auf Einzigartigkeit hin überprüfen können. In das dort durchgeführte Device Fingerprinting fließen Informationen zur Bildschirmauflösung und Farbtiefe, installierten Plugins, der Farbtiefe, den installierten Schriftarten, der Sprache und Zeitzone, des Betriebssystems und der Browserversion etc. mit ein. Die Informationen werden mittels JavaScript erhoben. Die Attributkombinationen stellen einen *Quasi-Identifikator* dar.

7.1.4 History Hijacking

Daneben existieren Ansätze zum *History Hijacking* (auch *History Sniffing* genannt), bei denen Server-Betreiber die Liste der vom Besucher zuvor besuchten Webseiten (Browser History) auslesen möchten. Ein (heute nicht mehr nutzbares) Verfahren basiert etwa in der Auswertung der Farbe der auf der Webseite dargestellten Links. Links zu bereits besuchten Websites werden im Browser typischerweise in einer anderen Farbe dargestellt. Mittels JavaScript lässt sich diese Information auslesen und an den Betreiber senden. Ein weiterer Ansatz basiert auf dem *Cache-Timing*. Hierbei bindet der Diensteanbieter in einer Webseite eine Datei aus einer anderen Website ein. Hat ein Nutzer diese andere Website zuvor besucht, befindet sich die Datei noch im Cache und die Ladezeit, die wiederum

mittels JavaScript gemessen und an den Betreiber gesendet wird, ist kürzer als wenn die Datei zum ersten Mal geladen werden muss.

7.1.5 P3P

Platform for Privacy Preferences (P3P) ist ein durch das WORLD WIDE WEB CONSORTIUM (W3C) spezifizierter Standard zum Austausch von Datenschutzinformationen. Auf Webserver-Seite gibt es eine P3P-Policy in Form einer XML-Datei⁴, die festlegt, welche personenbezogenen Daten auf Server-Seite zu welchem Zweck und wie lange gespeichert werden. Der Nutzer hat seinerseits seine P3P-Präferenzen (im Browser) angelegt, in denen er festlegt, für welche Zwecke er welche personenbezogenen Daten preiszugeben bereit ist. Beim Besuch einer Website findet (im Browser) ein automatischer Abgleich der Nutzer-Präferenzen mit der Policy des Webservers statt. Das Ergebnis wird dem Nutzer angezeigt und er wird in die Lage versetzt, selbst zu entscheiden, ob er unter diesen Umständen bereit ist, die Website weiter zu nutzen. P3P schafft also Transparenz hinsichtlich der Datenschutzbestimmungen einer Website. Nutzer müssen keine umfangreichen Datenschutzerklärungen durchlesen sondern es findet eine automatische Überprüfung statt. Selbst fremdsprachige Datenschutzerklärungen lassen sich mittels P3P einfach analysieren. P3P hat sich in der Praxis allerdings nicht durchgesetzt; lediglich der Internet Explorer unterstützt P3P noch.

7.2 Social Plugins

Social Plugins sind Schaltflächen, bereitgestellt u. A. von sozialen Netzwerken wie *Facebook*, *Twitter*, *LinkedIn* etc., die auf Websites in Form eines iFrames eingebunden werden und die es somit den Nutzern ermöglichen, die Websites über das soziale Netzwerk weiterzuempfehlen. Der „Like-Button“ von Facebook ist ein solcher Vertreter eines Social Plugins. Website-Betreiber binden diese Social Plugins auf ihren Websites ein, weil sie sich erhoffen, dass ihr Angebot dadurch einer größeren Nutzerzahl – etwa den Facebook-Freunden desjenigen Nutzers, der die Website „geliked“ hat, was sich auf seiner „Timeline“ widerspiegelt, bekannt gemacht wird. Die Nutzer sozialer Netzwerke nutzen diese Social Plugins sehr intensiv. Facebook-Nutzern ist der Like-Button genauso bekannt wie Twitter-Nutzern die „Retweet“-Funktion. Diese Funktionalitäten waren von Anfang an Teil der sozialen Netzwerke. Durch den „Export“ dieser Funktionalitäten aus den sozialen Netzwerken hinaus in das „offene Web“ haben sich allerdings ganz neue Möglichkeiten nicht nur für die Nutzer, sondern insbesondere für die sozialen Netzwerke, ergeben. Durch die Social Plugins sind die sozialen Netzwerke in der Lage, ihre Nutzer nicht nur innerhalb des sozialen Netzwerks zu verfolgen, sondern auch auf all jenen externen Websites, auf denen die Social Plugins zum Einsatz kommen.

⁴Die Datei liegt typischerweise unter /w3c/p3p.xml

Jedes mal, wenn ein Nutzer eine Website aufruft, die ein Social Plugin beinhaltet, werden Daten an das soziale Netzwerk gesendet. Dies geschieht direkt beim Aufruf der Website und erfordert keinerlei Interaktion des Nutzers – der Nutzer muss also nicht etwa den Like-Button anklicken. Im Falle von Facebook wird beim Besuch einer Website ein Cookie mit einer eindeutigen ID an Facebook übermittelt. Sofern der Besucher der Website bei Facebook angemeldet ist, kann Facebook damit die Information, welche externe Website besucht wird, direkt dem Nutzer zuordnen und somit das Nutzer-Profil weiter ausbauen. Sofern der Nutzer kein Facebook-Mitglied ist, ist Facebook immerhin noch in der Lage, ein Nutzungsprofil unter Pseudonym zu erstellen. Entscheidet sich der Nutzer zu einem späteren Zeitpunkt, Facebook beizutreten, kann das bereits angelegte Profil direkt mit dem neuen Nutzer in Verbindung gebracht werden.

7.3 Fazit

Insgesamt können wir festhalten, dass im Web eine Reihe von Datenschutzproblemen lauert. Nutzer werden ungewollt identifiziert und verfolgt und es kommt zu unerwünschten und von vielen Nutzern nicht erwarteten Informationsflüssen zu Dritten. Leider fehlt es an vielen Stellen an Transparenz – ausufernde Datenschutzerklärungen werden von den wenigsten Nutzern gelesen. Als gängigste Gegenmaßnahme (aus Nutzer-Sicht) haben wir in diesem Kapitel das Löschen von Informationen (Cookies, History, Cache etc.) kennengelernt. Die meisten Browser bieten einen entsprechenden *privaten Modus*, der diese Maßnahmen automatisch ergreift. Werkzeuge wie P3P würden für verbesserte Transparenz sorgen, haben sich allerdings bisher nicht durchgesetzt. In Zukunft werden wir eine zunehmende Nutzung von Tracking-Techniken, ähnlich den hier vorgestellten, auch in der „Offline-Welt“ sehen. Google hat bereits erste Ansätze entwickelt, beide Welten zu verschmelzen. Nutzer, die sich im Web für ein Produkt interessiert haben und anschließend im Geschäft in der Stadt kaufen, sollen keine Werbung mehr (im Web) für dieses Produkt erhalten. Genauso verhält es sich anders herum: Auch die Information, dass Offline-Werbung zum Online-Kauf animiert hat bzw. dass ein im Geschäft getestetes Produkt im Anschluss online bestellt wird, ist für Händler interessant.

Im Kap. 11 werden wir eine datenschutzrechtliche Betrachtung der in diesem Kapitel kennengelernten Tracking-Techniken vornehmen. Außerdem werden wir aufzeigen, wie ein datenschutzkonformer Einsatz von Google Analytics und Social Plugins möglich ist.

7.4 Übungsaufgaben

Aufgabe 1

Ist der Einsatz von „Device Fingerprinting“ auf einer Website für Nutzer erkennbar?

Zusammenfassung

Beim *Instant Messaging* (*IM*) unterhalten sich („chatten“) Kommunikations-Teilnehmer klassischerweise mittels Textnachrichten in Echtzeit. *ICQ* gilt als eines der ersten IM-Dienste im Internet. Es erfreute sich vor allem in den 90er-Jahren großer Beliebtheit, mit mehr als 100 Millionen Nutzern um die Jahrhundert-Wende. Inzwischen gibt es zahlreiche IM-Dienste, die neben Textnachrichten auch einen Austausch von Fotos, Videos etc. erlauben. Hierzu zählt etwa *WhatsApp*. In diesem Kapitel grenzen wir in Abschn. 8.1 zunächst das Thema Instant Messaging von E-Mail-Sicherheit ab. Wir werden sehen, dass jeweils unterschiedliche Schutzziele zum Tragen kommen. Danach lernen wir in Abschn. 8.2 mit *Off-the-Record* (*OTR*) *Messaging* einen wichtigen Vertreter eines sicheren IM-Protokolls kennen, bevor wir uns in Abschn. 8.3 schließlich der Ende-zu-Ende-Verschlüsselung von *WhatsApp* widmen.

Lernziele

Am Ende dieses Kapitels sollten Sie mit Off-the-record Messaging eines der grundlegendsten datenschutzfreundlichen Instant Messaging-Protokolle im Detail kennen. Außerdem sollten Sie verstehen, wie die Ende-zu-Ende-Verschlüsselung bei *WhatsApp* funktioniert.

8.1 Abgrenzung des Instant Messagings von E-Mail

Bevor wir uns in diesem Kapitel im Detail mit unterschiedlichen IM-Protokollen beschäftigen, grenzen wir IM zunächst von E-Mail ab und zeigen die unterschiedlichen Sicherheits- und Datenschutzschutzziele auf.

8.1.1 Schutzziele bei der E-Mail-Sicherheit

Bei der klassischen Ende-zu-Ende E-Mailverschlüsselung und -Signatur stehen die Schutzziele *Authentizität und Integrität*, *Vertraulichkeit*, sowie *Verbindlichkeit* im Vordergrund. Zum Erreichen der Authentizität/Integrität sowie der Verbindlichkeit wird eine digitale Signatur verwendet. Eine Prüfung der Authentizität/Integrität und Verbindlichkeit ist auch durch Dritte möglich. Dies ist vor allem im geschäftlichen Umfeld eine wichtige Eigenschaft. Um Vertraulichkeit zu gewährleisten, wird Verschlüsselung eingesetzt. Hierbei kommt ein hybrides Verfahren zum Einsatz: Die Nachricht wird vom Sender mit einem zufällig gewählten Sitzungsschlüssel verschlüsselt. Der Sitzungsschlüssel wiederum wird mit dem (langlebigen) öffentlichen Schlüssel des Empfängers verschlüsselt. Beide Teile werden als E-Mail zum Empfänger gesendet, der zunächst den Sitzungsschlüssel mit seinem privaten Schlüssel entschlüsselt und mit dem Sitzungsschlüssel im Anschluss die eigentliche Nachricht entschlüsselt. *Perfect Forward Secrecy (PFS)* wird bei diesem Ansatz nicht erreicht. Sollte es einem Angreifer also gelingen, das langlebige Geheimnis zu kompromittieren, ist er in der Lage, die gesamte vorangegangene, verschlüsselte und von ihm aufgezeichnete E-Mail-Kommunikation offenzulegen. Umgesetzt wird die Ende-zu-Ende E-Mail-Sicherheit typischerweise durch *Pretty Good Privacy (PGP)* und *Secure/Multipurpose Internet Mail Extensions (S/MIME)*.

8.1.2 Schutzziele beim Instant Messaging

Bei der privaten Kommunikation¹ per Instant Messaging kommen weitere Schutzziele in Betracht:

Abstreitbarkeit

Eine dritte Partei soll den Ursprung der Nachricht nicht verifizieren können, d. h. das Schutzziel Verbindlichkeit ist dabei gerade *nicht* erwünscht.

Perfect Forward Secrecy

Eine zukünftige Kompromittierung des (langlebigen) privaten Schlüssels soll die aufgezeichneten Nachrichten nicht gefährden.

8.2 Off-the-record Messaging

Sehen wir uns nun ein Protokoll an, das die zuvor geforderten Schutzziele an das Instant Messaging erfüllt: *Off-the-Record (OTR) Messaging*. OTR Messaging wurde 2004 von BORISOV et al. [1] entwickelt. Der Titel der Arbeit gibt bereits den Hinweis, dass klassi-

¹ „Privat“ im Sinne von „nicht geschäftlich“; Tratsch zwischen Freunden.

sche E-Mail-Verschlüsselung beim IM nicht sinnvoll ist: „Off-the-Record Communication, or, Why Not To Use PGP“.

OTR Messaging bietet sowohl *Authentizität* gegenüber dem Kommunikationspartner, als auch *Abstreitbarkeit* gegenüber Dritten. Authentizität und Abstreitbarkeit sind auf den ersten Blick gegensätzliche Forderungen. Wenn Alice mit Bob kommuniziert, soll Bob prüfen können, ob die Nachricht tatsächlich von Alice kommt – allerdings soll eine dritte Partei nicht prüfen können, ob die Nachricht von Alice stammt. Nach der Kommunikation soll es niemandem (auch nicht Alice und Bob) möglich sein, ein Transkript der Kommunikation zu erstellen.

Außerdem bietet OTR Messaging Vertraulichkeit und PFS.

8.2.1 Protokoll

Zunächst betrachten wir die einzelnen zum Einsatz kommenden Bausteine, die die unterschiedlichen Schutzziele erfüllen sollen.

Zum Schutz der *Vertraulichkeit* kommt ein Verschlüsselungsverfahren zum Einsatz, das „Verfälschbarkeit“ („malleability“) von Nachrichten bietet. Der Kryptotext soll dabei ohne Kenntnis des Schlüssels so verändert werden können, dass der Klartext nach der Entschlüsselung wiederum Sinn ergibt. Diese Eigenschaft ist normalerweise nicht erwünscht, hier nutzt sie dem Datenschutz. Verwendet wird hierfür das *Advanced Encryption Standard (AES)*-Verschlüsselungsverfahren im „Counter Mode“. Bei diesem Modus wird ein Zählerwert verschlüsselt und das Ergebnis mit dem Klartext XOR-verknüpft. Dadurch ergibt sich eine Verwendung von AES als „Stromchiffre“. Die Ver- und Entschlüsselung ist in Abb. 8.1 dargestellt.

Der Schlüsselaustausch erfolgt nach Diffie-Hellman (DH) und bietet, wie bereits in Abschn. 2.3.5 besprochen, PFS. Der Austausch der kurzlebigen Sitzungsschlüssel nach

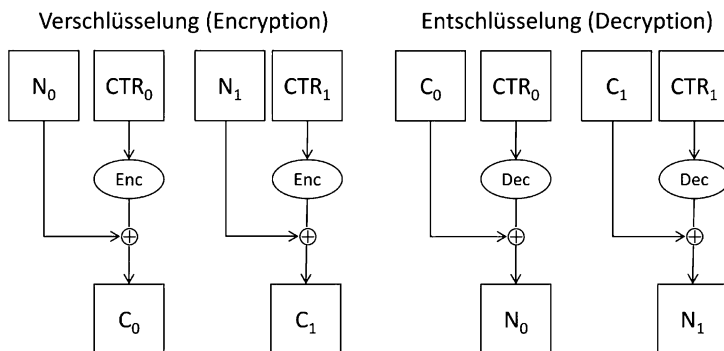


Abb. 8.1 AES im Counter Mode

Abb. 8.2 DH-Austausch beim OTR Messaging

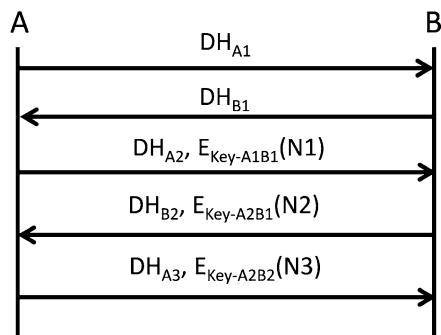
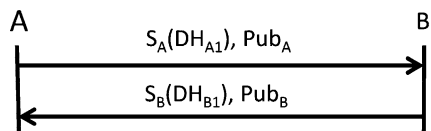


Abb. 8.3 Authentifizierung des DH-Austauschs beim OTR Messaging

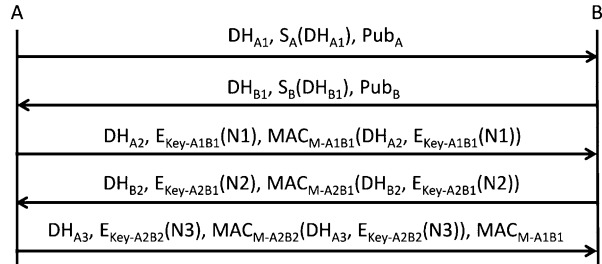


DH ist in Abb. 8.2 dargestellt. Die Erneuerung wird dabei mit dem Nachrichtenaustausch kombiniert. *Key* bezeichnet den aus dem DH-Schlüsselaustausch abgeleiteten, gemeinsamen Schlüssel zwischen A und B, bzw. genauer, den 128 Bit SHA-1-Hashwert daraus. Nx bezeichnet die Nachricht x .

Die *Teilnehmerauthentifizierung* erfolgt über eine digitale Signatur, wie in Abb. 8.3 dargestellt. Die Signatur wird nur für den initialen Schlüsselaustausch benötigt, der pro Sitzung erfolgt, d. h. bis sich ein Teilnehmer abmeldet bzw. nach einer inaktiven Periode wird die Sitzung ebenfalls beendet. Die Teilnehmerauthentifizierung über die digitale Signatur ist nötig, da der DH-Schlüsselaustausch anfällig ist gegenüber MITM-Angriffen. $S_x(DH_x)$ bezeichnet dabei die Signatur von Partei x über den öffentlichen DH-Parameter von x ; Pub_x bezeichnet den öffentlichen Schlüssel von x .

Zur Authentifizierung der Nachrichten und zum *Integritätsschutz* der Nachrichten kommt ein *Message Authentication Code (MAC)* zum Einsatz. Der Schlüssel wird dabei vom Verschlüsselungsschlüssel durch das Anwenden einer Hash-Funktion auf den Verschlüsselungsschlüssel abgeleitet. Die Authentizität der „neuen“ Schlüssel ergibt sich durch die Authentizität des initialen Schlüssels (der per digitaler Signatur authentifiziert wurde). Durch die Ableitung aus dem Verschlüsselungsschlüssel ergibt sich auch, dass jeder, der die Nachricht lesen kann, diese auch modifizieren und den MAC-Wert entsprechend anpassen kann. Zudem wird der MAC-Schlüssel in der nächsten Nachricht veröffentlicht. Dies führt dazu, dass *jeder* mit diesem Schlüssel beliebige Nachrichten authentifizieren kann. Die Nachricht kann also keinem Autor mehr zugeordnet werden, sie ist also *abstreitbar*.

Abb. 8.4 Kompletter
Protokollablauf des OTR
Messagings



Wie passen diese Bausteine nun zusammen? Der komplette Protokollablauf ist in Abb. 8.4 dargestellt. M_x bezeichnet dabei den aus Key_x , wie zuvor besprochen abgeleiteten MAC-Schlüssel.

8.2.2 Implementierung

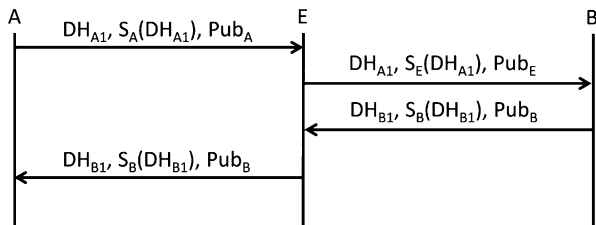
OTR Messaging ist als Plugin implementiert und steht u. a. für die Instant Messenger *Pidgin*, *Miranda* und *Trillian* zur Verfügung. Es verwendet die bestehenden IM-Protokolle zur Nachrichtenübermittlung. Die Nachricht wird vom Plugin verschlüsselt und authentifiziert und das Ergebnis wird als Text kodiert und als IM-Nachricht versendet. Das Plugin hält fest, welcher Kommunikationspartner auch OTR Messaging verwendet. Die erste Nachricht wird unverschlüsselt versendet. Dabei wird ein Identifikator für die Plugin-Verwendung angehängt. Falls der Partner auch das Plugin nutzt, wird der DH-Austausch eingeleitet. Der „Fingerprint“ (ein Hash-Wert) des öffentlichen Schlüssels des Kommunikationspartners wird dem Nutzer angezeigt. Der Nutzer kann den (kurzen) Fingerprint „out-of-band“, d. h. über einen anderen Kanal, überprüfen, um festzustellen, ob der öffentliche Schlüssel tatsächlich seinem Kommunikationspartner gehört. Der öffentliche Schlüssel wird danach gespeichert und bei der nachfolgenden Kommunikation wird geprüft, ob derselbe Schlüssel verwendet wird. Dieser Ansatz wird bspw. auch bei SSH verfolgt.

8.2.3 Angriffe auf OTR Messaging

DI RAIMONDO ET AL. haben in ihrer Arbeit „Secure Off-the-Record Messaging“ [2] Angriffe auf das OTR Messaging vorgestellt und vorgeschlagen, das SIGMA-Protokoll [4] zu nutzen, um diesen Angriffen zu begegnen.

Das zuvor vorgestellte OTR Messaging-Protokoll hat ein „Identity Misbinding“-Problem. Dabei wird der ausgehandelte Schlüssel der falschen Identität zugeschrieben,

Abb. 8.5 Identity Misbinding
bei OTR Messaging



wie in Abb. 8.5 dargestellt. Bob denkt, er kommuniziert mit Eve und Alice denkt, sie kommuniziert mit Bob. Stattdessen kommunizieren beide über die MITM-Angreiferin Eve.

Außerdem weist das OTR Messaging-Protokoll das Problem der „Freshness Impersonation“ auf, das für einen Angriff ausgenutzt werden kann. Die Signatur über den DH-Wert enthält keinen Schutz vor Wiedereinspielung, also einem Wiederholungsangriff. Die Voraussetzung für die Ausnutzung ist, dass der Angreifer den privaten DH-Wert kennt.² Ist dies der Fall, kann er eine Nachricht $DH_{A1}, S_A(DH_{A1})$ wieder einspielen. Der Angreifer kann sich die Sitzungsschlüssel für jede beliebige Antwort DH_{B1} ableiten, ohne ein langlebiges Geheimnis kennen zu müssen.

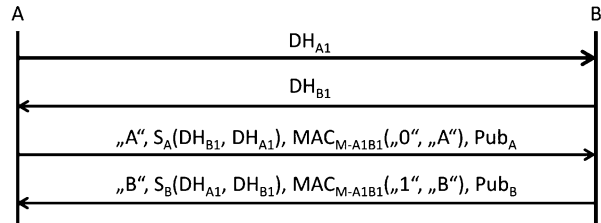
8.2.4 SIGMA-Protokoll

Die Lösung für die beiden Angriffe besteht in der Verwendung des SIGMA-Protokolls. Das Verfahren ist in Abb. 8.6 dargestellt. Die Verwendung des MACs in der dargestellten Form, unter Einbeziehung der Identitäten der Kommunikationspartner, verhindert das Identity Misbinding. Der Wiederholungs-Angriff wird dadurch verhindert, dass die Signatur nicht nur über den eigenen DH-Wert erzeugt wird, sondern auch über den Wert des Kommunikationspartners. Zusätzlich können die Identitäten geschützt werden, indem ab Nachricht 3 bereits verschlüsselt kommuniziert wird. Anders als beim zuvor vorgestellten OTR Messaging-Protokoll wird hier zuerst ein nicht-authentifizierter DH-Schlüsselaustausch durchgeführt. Im Anschluss wird ein gesicherter Kanal aufgebaut und die Authentifizierung erfolgt schließlich über den gesicherten Kanal.

Das vorgestellte Protokoll kommt in dieser Form für den authentifizierten Schlüsselaustausch bei Version 2 des OTR Messaging-Protokolls zum Einsatz. Anstatt der Authentifizierung mittels Prüfung der Fingerprints, wie dies beim „Basis“-OTR Messaging-Protokoll vorgesehen ist, wird hier eine Authentifizierung auf Grundlage des

²Die Annahme bei diesem Angriff ist, dass die privaten DH-Werte nicht so sicher behandelt werden wie das langlebige Geheimnis, das bspw. auf einer Smartcard gespeichert wird.

Abb. 8.6 Verwendung des SIGMA-Protokolls für das OTR Messaging



Socialist Millionaire's Protocol (SMP),³ einer Abwandlung von *Yao's Millionaires' Problem*,⁴ das wir in Abschn. 10.2.1 kennen lernen werden, ermöglicht. Ziel ist es, MITM-Angriffe zu erkennen. Dabei haben Alice und Bob eine gemeinsame, geheime Information x und y . Alice und Bob möchten nun prüfen, ob $x == y$. Das SMP erlaubt die Prüfung, ohne dass irgendwelche Informationen preisgegeben werden müssen. Hierfür kommt ein *Zero-Knowledge Proof (ZKP)* (siehe Abschn. 10.3) zum Einsatz. Wenn $x == y$, dann haben beide dieselbe, geheime Information eingegeben.

OTR Messaging erfreut sich in der Praxis großer Beliebtheit und ist auch die Grundlage für das Signal-Protokoll, das bei der Ende-zu-Ende-Verschlüsselung bei WhatsApp zum Einsatz kommt.

8.3 WhatsApp

WhatsApp wurde im Jahr 2014 für 19 Milliarden US-Dollar von Facebook aufgekauft. Damals hatte WhatsApp weltweit 450 Millionen Nutzer. Anfang 2016 wurde die Marke von 1 Milliarde Nutzern überschritten.

Im April 2016 kündigte WhatsApp an, dass mit der neuen Version von WhatsApp (ab 31.03.2016) sämtliche Kommunikation (also nicht nur Textnachrichten zwischen jeweils zwei Teilnehmern sondern auch Gruppenchats, der Versand von Multimedia-Dateien und Telefonie über WhatsApp) auf allen Plattformen Ende-zu-Ende-verschlüsselt wird. Durch die Ende-zu-Ende-Verschlüsselung hat selbst WhatsApp keinen Zugriff auf die Daten im Klartext. Experten lobten diesen Schritt von WhatsApp und begrüßten auch die einfache Handhabung. Die *Electronic Frontier Foundation (EFF)* nahm WhatsApp in ihre Liste der Krypto-Messenger mit auf, die von OTR Messengern, SilentText, TextSecure etc. angeführt wird.

³Zwei Millionäre wollen dabei feststellen, ob sie gleich reich sind, ohne dem jeweils anderen etwas über ihren eigenen Reichtum zu verraten.

⁴Zwei Millionäre wollen dabei feststellen, wer von ihnen reicher ist, ohne dem jeweils anderen ihren eigenen Reichtum offenlegen zu müssen.

8.3.1 Signal-Protokoll

Zur Ende-zu-Ende-Verschlüsselung bei WhatsApp kommt das *Signal*-Protokoll⁵ zum Einsatz, das bis März 2016 „Axolotl“-Protokoll genannt wurde und von *Open Whisper Systems* entwickelt wurde.⁶ Das Signal-Protokoll basiert auf OTR Messaging, u. a. die Erneuerung der Sitzungsschlüssel durch wiederholten DH-Schlüsselaustausch ist daran angelehnt, und dem Silent Circle Instant Messaging Protocol (SCIMP).

Die nachfolgende Beschreibung des Signal-Protokolls orientiert sich an dem „WhatsApp Security Whitepaper“ von April 2016 [5].

Initialisierung

Während der Installation von WhatsApp werden folgende Schlüssel lokal auf dem Smartphone von der WhatsApp-App generiert:

- *Identity Key*: Langzeitschlüssel, der während der Installation von WhatsApp generiert wird.
- *Signed Pre Key*: Schlüssel der ebenfalls während der Installation generiert wird und vom Identity Key signiert wird. Dieser Schlüssel wird regelmäßig erneuert.
- *One-Time Pre Key*: Einmalschlüssel. Davon werden während der Installation mehrere generiert, nach Bedarf werden während der Nutzung weitere Schlüssel generiert.

Registrierung

Bei der Registrierung übermittelt die App den zuvor generierten Identity Key sowie den Signed Pre Key und eine Reihe von One-Time Pre Keys an den WhatsApp-Server. Der WhatsApp-Server speichert diese (öffentlichen)⁷ Schlüssel unter der Identität des Nutzers.

Sitzungs-Initialisierung

Damit Nutzer verschlüsselt miteinander kommunizieren können, muss zunächst eine Sitzung etabliert werden. Diese Sitzung ist solange gültig, bis einer der beteiligten Kommunikationspartner etwa die App neu installiert.

Der Initiator einer Sitzung bezieht dafür in einem ersten Schritt den Identity Key, Signed Pre Key, sowie einen One-Time Pre Key des Empfängers. Der Server schickt die Keys zurück und löscht seinerseits den nunmehr verwendeten Einmalschlüssel (One-Time Pre Key) des Nutzers. Bei Bedarf fordert er neue Einmalschlüssel vom Nutzer an.

Der Initiator erzeugt nun einen kurzlebigen Schlüssel und generiert ein *Master Secret*, das von diesem kurzlebigen Schlüssel, seinem eigenen Identity Key, dem Identity Key

⁵Das Signal-Protokoll kommt auch beim gleichnamigen Krypto-Messenger „Signal“ zum Einsatz, der von *Open Whisper Systems* entwickelt wurde.

⁶Die bei WhatsApp verwendete Signal-Protokoll-Bibliothek ist unter <https://github.com/whispersystems/libsignal-protocol-java/> als Open Source-Projekt abrufbar.

⁷Der WhatsApp-Server erhält keinen Zugriff auf die zugehörigen privaten Schlüssel.

und dem Signed Pre Key des Empfängers sowie dem One-Time Pre Key des Empfängers abhängt. Diese Schlüsselableitung wird mittels DH-Verfahren auf elliptischen Kurven (Elliptic curve Diffie-Hellman (ECDH)) durchgeführt.

Aus diesem Master Secret werden wiederum ein *Root Key* sowie *Chain Keys* abgeleitet.⁸

Von nun an ist eine Langzeit-Sitzung aufgebaut und der Sender kann bereits verschlüsselte Nachrichten an den Empfänger senden – selbst wenn der Empfänger zu diesem Zeitpunkt offline ist. Dies stellt einen Vorteil gegenüber dem zuvor vorgestellten OTR Messaging-Protokoll dar, bei dem für die Etablierung einer Sitzung die Kommunikationspartner gleichzeitig online sein müssen.

Sitzungs-Entgegennahme

Der Empfänger erhält vom Sender neben der verschlüsselten, eigentlichen Nachricht zusätzliche Informationen zum Sitzungs-Aufbau: den zuvor vom Sender erzeugten, kurzlebigen Schlüssel sowie den Identity Key des Senders. Aus diesen (öffentlichen) Schlüsseln des Senders und seinen eigenen, privaten Schlüsseln kann der Empfänger seinerseits nach DH das Master Secret berechnen. Davon leitet der Empfänger ebenfalls den zugehörigen Root Key sowie Chain Keys ab. Damit ist der Sitzungsaufbau zwischen den Kommunikationspartnern abgeschlossen.

Nachrichten-Austausch

Die Nachrichten selbst werden mittels eines *Message Keys* verschlüsselt und authentifiziert. Diese Message Keys werden von den zuvor abgeleiteten Chain Keys abgeleitet. Zur Verschlüsselung kommt AES256 im CBC-Modus und zur Authentifizierung HMAC-SHA256 zum Einsatz. Der Message Key wird für jede Nachricht erneuert, d. h. erneut vom Chain Key abgeleitet.⁹ Außerdem wird mit jedem gegenseitigen Nachrichten-Austausch ein neuer Chain Key nach DH abgeleitet. Diese Vorgehensweise haben wir bereits bei OTR Messaging kennengelernt. Damit wird Perfect Forward Secrecy (PFS) gewährleistet.

8.3.2 Medien-Verschlüsselung

Größere Anhänge, wie etwa Fotos oder Videos, werden ebenfalls Ende-zu-Ende-verschlüsselt. Der Sender erzeugt hierfür kurzlebige Schlüssel zur Verschlüsselung und Authentifizierung. Danach verschlüsselt er die Datei mittels AES256 im CBC-Modus und authentifiziert sie mittels HMAC-SHA256. Die verschlüsselte und authentifizierte Datei lädt er auf den WhatsApp-Server. Die verwendeten Sitzungsschlüssel übermittelt er nach

⁸Diese Ableitung erfolgt nach der HMAC-based Extract-and-Expand Key Derivation Function (HKDF) [3].

⁹Dieser Ansatz wird „hash ratchet“ genannt.

dem zuvor beschriebenen Verfahren, gemeinsam mit einem Link zur verschlüsselten Datei, an den Empfänger. Der Empfänger entschlüsselt die Nachricht, bezieht die verschlüsselte und authentifizierte Datei vom Server, entschlüsselt sie und prüft die Authentizität.

8.3.3 Sichere Telefonie

Zur Ende-zu-Ende-Verschlüsselung von Telefongesprächen über WhatsApp kommt das *Secure Real-Time Transport Protocol* ([SRTP](#)) zum Einsatz, das ein gängiges Verfahren für sichere VoIP-Telefonie darstellt. Der Schlüsselaustausch hierfür geschieht analog zum vorgestellten Austausch für IM-Nachrichten.

8.3.4 Schlüssel-Verifikation

Beim OTR Messaging sind wir bereits auf die Prüfung der Authentizität der öffentlichen Schlüssel der Kommunikationspartner eingegangen. Bei WhatsApp kann diese Prüfung über einen QR-Code, der u. A. den Identity Key enthält, vorgenommen werden. Dazu wird der QR-Code vom Kommunikationspartner etwa direkt von seiner WhatsApp-App gescannt und der Inhalt (also insbesondere der Identity Key) gegen den vom WhatsApp-Server gelieferten Inhalt geprüft. Alternativ dazu kann auch ein 60-stelliger Code zur Prüfung verwendet werden.

8.3.5 Datenschutzrechtliche Probleme

Trotz der Ende-zu-Ende-Verschlüsselung bei WhatsApp gibt es nach wie vor datenschutzrechtliche Bedenken. So überträgt die WhatsApp-App nach der Installation am Smartphone das gesamte Telefonbuch an den WhatsApp-Server. Dies dient dem Abgleich der eigenen Kontakte gegen die so immer weiter wachsende Datenbank von WhatsApp, um Kontakte zu ermitteln, die ebenfalls WhatsApp nutzen. Dabei werden personenbezogene Daten von Dritten (die Telefonnummern) an WhatsApp-Server in den USA übertragen. Diese Übertragung findet in der Regel ohne Einwilligung der Betroffenen statt.

Außerdem gelangt WhatsApp an die Metadaten der stattfindenden Kommunikation und erfährt damit, wer mit wem wie häufig in Kontakt steht.

8.4 Fazit

Mit der Ende-zu-Ende-Verschlüsselung bei WhatsApp haben wir ein positives Beispiel für sicheres Instant Messaging kennengelernt. Im Gegensatz zu E-Mail, wo sich Ende-

zu-Ende-Verschlüsselung bislang nicht durchsetzen konnte, ist sichere Kommunikation für viele Nutzer möglich geworden – ohne dafür großen Aufwand betreiben zu müssen. Nichtsdestotrotz darf nicht verschwiegen werden, dass WhatsApp auf die Metadaten der stattfindenden Kommunikation Zugriff hat. Außerdem wird nach wie vor, was aus datenschutzrechtlicher Sicht problematisch ist, das Telefonbuch des Nutzers an WhatsApp übermittelt, um eigene Kontakte mit registrierten Nutzern abzugleichen. In dieser Hinsicht besteht nach wie vor Forschungsbedarf. Es ist noch nicht klar, wie eine datenschutzfreundliche Suche nach gemeinsamen Kontakten („Private Contact Discovery“) in großem Maßstab funktionieren könnte. Darauf weist der Entwickler der WhatsApp-Verschlüsselung in einem Blog-Eintrag im Jahr 2014 hin.¹⁰ *Private Set Intersection (PSI)* stellt einen möglichen Ansatz dafür dar. In diesem Bereich wurde in den letzten Jahren intensiv geforscht und es wurden effiziente Protokolle vorgestellt. Wir werden uns damit in Abschn. 10.2.2 beschäftigen.

Facebook hat im Sommer 2016 angekündigt, auch beim Facebook Messenger Ende-zu-Ende-Verschlüsselung einzuführen. Im Gegensatz zur Verschlüsselung bei WhatsApp soll eine private, verschlüsselte Konversation beim Messenger allerdings explizit aktiviert werden müssen – nicht alle Nachrichten werden also standardmäßig immer verschlüsselt. Als Anwendungsszenario für die verschlüsselte Kommunikation wurde die Übertragung von Gesundheits- und Finanzdaten über den Messenger genannt. Zur Verschlüsselung kommt beim Messenger ebenfalls das in diesem Kapitel beschriebene Signal-Protokoll zum Einsatz.¹¹ Ende September wurden die „Geheimen Unterhaltungen“, die ausschließlich reine Textnachrichten unterstützen, beim Facebook Messenger eingeführt.¹² Wie beim Instant Messenger *Snapchat* lässt sich bei einer geheimen Unterhaltung ein Haltbarkeitsdatum vergeben, so dass die Nachricht nach einer bestimmten Zeit automatisch gelöscht wird.

8.5 Übungsaufgaben

Aufgabe 1

Kann OTR Messaging auch für die E-Mail-Kommunikation verwendet werden?

¹⁰<https://whispersystems.org/blog/contact-discovery/>

¹¹https://fbnewsroom.us.files.wordpress.com/2016/07/secret_conversations_whitepaper-1.pdf (Zugegriffen am 01.08.2016)

¹²<http://t3n.de/news/facebook-messenger-geheime-unterhaltungen-ende-zu-ende-verschlueselung-724213/> (Zugegriffen am 15.10.2016)

Literatur

1. Nikita Borisov, Ian Goldberg, and Eric Brewer. Off-the-record Communication, or, Why Not to Use PGP. In *Proceedings of the 2004 ACM Workshop on Privacy in the Electronic Society*, WPES '04, pages 77–84, New York, NY, USA, 2004. ACM.
2. Mario Di Raimondo, Rosario Gennaro, and Hugo Krawczyk. Secure off-the-record messaging. In *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society*, WPES '05, pages 81–89, New York, NY, USA, 2005. ACM.
3. Internet Engineering Task Force. Hmac-based extract-and-expand key derivation function (hkdf). RFC 5869, May 2010.
4. Hugo Krawczyk. *Advances in Cryptology – CRYPTO 2003: 23rd Annual International Cryptology Conference, Santa Barbara, California, USA, August 17–21, 2003. Proceedings*, chapter SIGMA: The ‚SIGn-and-MAC‘ Approach to Authenticated Diffie-Hellman and Its Use in the IKE Protocols, pages 400–425. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.
5. WhatsApp. Whatsapp encryption overview. Technical white paper, Apr. 2016. <https://www.whatsapp.com/security/WhatsApp-Security-Whitepaper.pdf>.

Zusammenfassung

Ausweisdokumente dienen dem Nachweis der Identität einer natürlichen Person. Sie enthalten Attribute der Person. Dies können Attribute zur eindeutigen Identifizierung, wie z. B. Lichtbild, Angaben über Größe, Augenfarbe etc. sein, oder auch Attribute, die durch den Ausweis nachgewiesen werden, etwa das Geburtsdatum zur Prüfung des Alters. Elektronische Ausweise ermöglichen die Identifizierung/Authentifizierung unter Zuhilfenahme lokal auf dem Ausweis elektronisch gespeicherter Daten. Dabei kann eine Identitätsfeststellung entweder vor Ort stattfinden oder die Identitätsfeststellung (oder der Nachweis von Attributen) erfolgt über ein Kommunikationsnetz.

Der *elektronische Reisepass* und der *elektronische Personalausweis* sind zwei Vertreter von elektronischen Ausweisdokumenten, mit denen wir uns in diesem Kapitel im Detail beschäftigen werden. Zweifellos handelt es sich bei den in den Ausweisdokumenten gespeicherten Daten um *personenbezogene Daten*, die die meisten Menschen wohl als besonders schutzwürdig einstufen würden. Nicht jeder würde sie zur Verfügung stellen. In diesem Kapitel werden wir uns zunächst in Abschn. 9.1 damit beschäftigen, wie die im elektronischen Reisepass auf einem Chip gespeicherten und per Funk übertragenen Daten vor dem unberechtigten Auslesen geschützt werden. Als nächstes beschäftigen wir uns in Abschn. 9.2 mit dem Datenschutzkonzept des elektronischen Personalausweises – einem Lehrstück für *Privacy by Design*.

Lernziele

Am Ende dieses Kapitels sollten Sie die Sicherheits- und Datenschutzkonzepte des elektronischen Reisepasses und des elektronischen Personalausweises kennen. Sie

(Fortsetzung)

sollten die Designentscheidungen nachvollziehen können, die kryptographischen Lösungen verstehen und die erreichten Sicherheits- und Datenschutzniveaus der elektronischen Ausweisdokumente bewerten können.

9.1 Elektronischer Reisepass

Der elektronische Reisepass (im Weiteren auch „ePass“ genannt) wurde am 1. November 2005 eingeführt. Die Speicherung des Gesichtsbilds und der Fingerabdrücke¹ des Besitzers erfolgt dabei auf einem Chip mit Funkschnittstelle („RF-Chip“). Die Grundlage dafür bilden die Standards der *International Civil Aviation Association* (ICAO), einer Sonderorganisation der Vereinten Nationen, die für die zivile Luftfahrt zuständig ist. Die Gesichtsbilder und Fingerabdrücke werden einfach als Bilder gespeichert. Eine Vorverarbeitung und Speicherung sogenannter Templates wäre zwar wünschenswert gewesen, doch gibt es dafür keinen weltweit anerkannten Standard. Die Entscheidung gegen ein kontaktbehaftetes Auslesen des Chips wurde aufgrund der Verschleißanfälligkeit getroffen. Eine kontaktbehaftete Lösung hätte zudem ein anderes Pass-Format erfordert.

9.1.1 Passive Authentication

Das Ziel des elektronischen Reisepasses ist es, Fälschungen zu erschweren. Dies erfordert eine Authentifizierung der gespeicherten Daten. Aus diesem Grund wird bereits bei der Herstellung des Passes eine digitale Signatur über die gespeicherten Daten erstellt und im Chip gespeichert.

Die Frage, die sich stellt ist: wie kann diese digitale Signatur überprüft werden? Es muss überprüft werden können, ob der Pass tatsächlich von einem legitimen Ausweishersteller erstellt, und die Daten sowie die digitale Signatur von diesem angebracht wurden. An dieser Stelle kommt eine Public Key Infrastructure (PKI) ins Spiel.

PKI für elektronische Ausweisdokumente: Signaturen

Die Signatur-PKI mit der Berechtigung Ausweisdaten für die Passive Authentication zu signieren ist in Abb. 9.1 dargestellt.

Verschiedene Länder tauschen ihre CSCA-Zertifikate aus, aber es findet keine echte Kreuzzertifizierung statt.

¹Die Speicherung von Fingerabdrücken wurde erst später eingeführt.

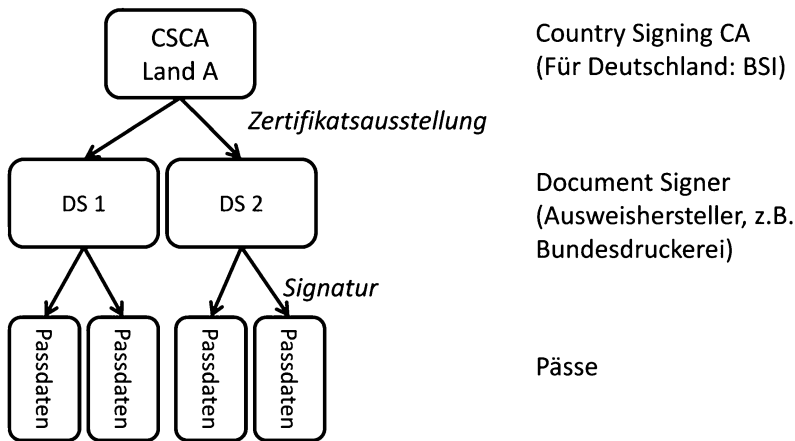


Abb. 9.1 Signatur-PKI

9.1.2 Basic Access Control

Bei der Übertragung von Daten über die Funkschnittstelle ist problematisch, dass diese mithörbar ist – selbst wenn die Pässe nur auf Reichweiten von einigen Zentimetern ausgelegt sind.² Damit besteht die Gefahr des unberechtigten Auslesens beim Mitführen des Passes. Die *Basic Access Control (BAC)* soll diesem Problem entgegenwirken. Die Entwurfsziele des Protokolls sind:

- Zugriffskontrolle,
- Kryptographische Absicherung der Datenübertragung,
- Einfachheit,
- Kompatibilität zu bestehenden Formaten und Verfahren.

Das Verfahren generiert das Schlüsselmaterial aus Informationen in der bestehenden maschinenlesbaren Zone des Reisepasses. Die enthaltenen Daten sind:

- Ausstellungsland,
- Art des Dokuments (z. B. Reisepass, vorläufiger Reisepass),
- Name,
- Seriennummer (in Deutschland besteht diese aus Behörden- und Passnummer),
- Nationalität,
- Geburtsdatum,
- Geschlecht,

²Mit entsprechender Antennentechnik sind auch größere Abstände möglich.

- Ablaufdatum,
- Personenkennziffer (bei deutschen Pässen nicht enthalten).

Die Idee hinter der Verwendung der maschinenlesbaren Zone ist, dass der Reisepass dazu vorgelegt werden muss. Wer nun mit Basic Access Control auf die im Chip gespeicherten Daten zugreift, erhält keine weiteren Informationen als die ohnehin auf der gleichen Seite des Passes sichtbaren (einschließlich des Gesichtsbildes).

Die Verschlüsselung der zu übertragenden Daten erfolgt im Anschluss über 3DES³ mit zwei Schlüsseln.

Angriff auf BAC

Bei frühen elektronischen Reisepässen bestand das Problem darin, dass der Schlüsselraum sehr klein war. Es wurden nur die Seriennummer, das Geburtsdatum und das Ablaufdatum (jeweils mit Prüfziffer) für die Schlüsselgenerierung verwendet. Die Seriennummer wurde früher innerhalb einer Behörde einfach aufsteigend vergeben. Das Ablaufdatum liegt innerhalb eines bekannten Zeitraums. Dies war anfangs besonders gravierend, da die Gültigkeitsdauer fest ist (derzeit 6 Jahre bei Kindern und 10 Jahre bei Erwachsenen ab Ausstellung). Das Geburtsdatum kann von einem Angreifer, der den Passinhaber sieht ungefähr erraten werden. Je nach Schätzung bietet die BAC damit nur ca. 30 bis 40 *Bits* an effektiver Sicherheit.

Schutz gegen Angriffe

Man muss zwischen zwei Arten von Angriffen auf BAC unterscheiden: *Aktive Angriffe* und *Passive Angriffe*.

Bei einem aktiven Angriff versucht der Angreifer einen direkten Zugriff auf den elektronischen Reisepass. Die benötigte Zeit pro Zugriffsversuch beträgt etwa 1 Sekunde [1]. Damit gewähren auch kurze Schlüssellängen hinreichenden Schutz.

Bei einem passiven Angriff hört der Angreifer die Kommunikation mit und probiert offline die Schlüssel durch. Dieses Offline-Durchprobieren ist möglich, da für den Schlüsselaustausch nicht Diffie-Hellman eingesetzt wird (vermutlich sollte das Verfahren auch auf leistungsschwachen Prozessoren einsetzbar sein, was Diffie-Hellman ausschließt). Stattdessen wird im Wesentlichen ein Hashverfahren auf Zufallswerte von beiden Seiten angewendet.

Um gegen beide Angriffe einen verbesserten Schutz zu gewährleisten, wurde eine zufällige Vergabe der Seriennummern sowie die Verwendung von Buchstaben zusätzlich zu Ziffern bei der Seriennummer eingeführt. Die zufällige Vergabe der Seriennummern erhöht zwar die Sicherheit. In Deutschland bleibt es aber dabei, dass die ersten vier

³Bei Triple-DES wird der Data Encryption Standard (DES)-Verschlüsselungsalgorithmus 3-fach angewendet um die Sicherheit (des schwachen DES-Algorithmus) zu erhöhen.

Abb. 9.2 EAC: Chip Authentication



Stellen der neunstelligen Seriennummer die Meldebehörde angeben; nur die restlichen fünf Stellen werden zufällig vergeben.

9.1.3 Extended Access Control

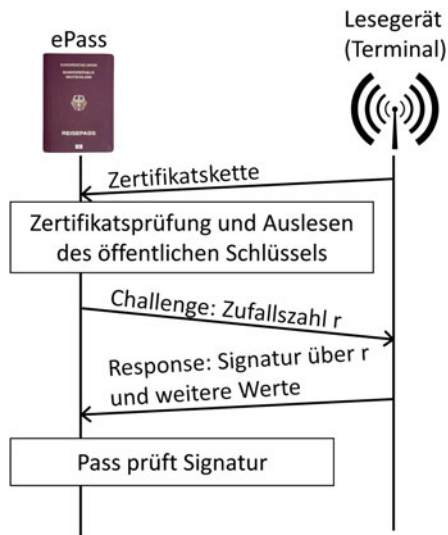
Neben der Basic Access Control als Basisschutz für das Gesichtsbild und andere im Pass sichtbar abgedruckte Daten bietet die *Extended Access Control (EAC)* einen weitergehenden Zugriffsschutz für weitere (sensiblere) Daten, wie etwa die Fingerabdrücke. Zunächst war die EAC kein ICAO-Standard, sondern wurde durch die EU getrieben. EAC spezifiziert⁴ zwei Protokolle für die beidseitige Authentifizierung zwischen Chip (im ePass) und Terminal: *Chip Authentication* und *Terminal Authentication*. Diese Protokolle sind hier jeweils in Version 1 beschrieben, wie sie in den elektronischen Pässen der EU verwendet werden. Für den deutschen elektronischen Personalausweis kommt jeweils Version 2 zum Einsatz. Die Kommunikation der EAC-Protokolle wird zunächst mit Schlüsselmateriale aus der BAC geschützt.

Chip Authentication

Das Protokoll zur Authentifizierung des Chips im ePass gegenüber dem Terminal, die sogenannte „Chip Authentication“, ist in Abb. 9.2 dargestellt.

⁴Die Spezifikation der zugehörigen Technischen Richtlinie 03110, Version 2.03 stammt vom Bundesamt für Sicherheit in der Informationstechnik (BSI).

Abb. 9.3 EAC: Terminal Authentication



Die Chip Authentication soll dafür sorgen, dass der Chip als authentisch erkannt wird und dass ein Schlüssel zur Absicherung der gemeinsamen Kommunikation abgeleitet wird.

Die Voraussetzung für die (implizite) Authentifizierung des Chips ist, dass der statische DH-Wert des Passes geprüft werden kann. Dazu wird die Passive Authentication verwendet.

Die Fälschung eines Passes wird durch die Ausführung des Protokolls nicht verhindert. Ein Fälscher könnte die Kommunikation mit einem echten Pass mithören und dessen (öffentlichen) DH-Wert einfach ebenfalls verwenden. Die weitere Kommunikation schlägt dann aber fehl, denn dem Fälscher fehlt der zugehörige geheime DH-Wert, der für die Schlüsselableitung nötig ist.

Terminal Authentication

Das Protokoll zur Authentifizierung des Terminals gegenüber dem Chip im ePass, die sogenannte „Terminal Authentication“, ist in Abb. 9.3 dargestellt.

Die Durchführung der Terminal Authentication erfolgt nach der Chip Authentication. Alle Nachrichten sind mit dem zuvor abgeleiteten Schlüssel verschlüsselt und authentifiziert. Das Terminal besitzt ein Zertifikat und einen zugehörigen privaten Schlüssel. Der Pass schickt, wie in der Abbildung dargestellt, eine Challenge r an das Terminal, das diese Challenge (sowie die Identität des Passes – eine Seriennummer) mit seinem privaten Schlüssel signiert und zum Chip zurück sendet. Der Chip prüft schließlich die Signatur – mit Hilfe des öffentlichen Schlüssels aus dem Zertifikat des Terminals.

Die Fähigkeit zur Durchführung eines DH-Schlüsselaustauschs und zur Prüfung von Zertifikaten bzw. Zertifikatsketten erfordert aufwendige Chips. Dies ist vermutlich auch der Grund, warum zunächst nur die BAC eingeführt wurde und die später eingeführte EAC auch nur in europäischen Pässen verwendet wurde.

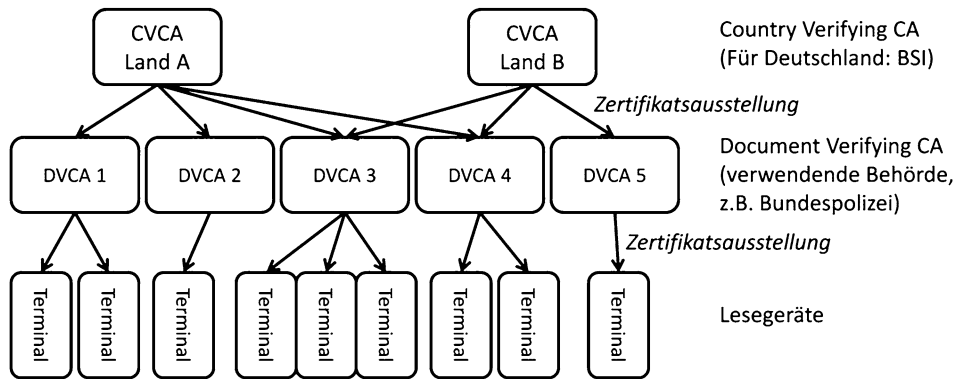


Abb. 9.4 Verifikations-PKI

Bei der hier beschriebenen Version 1 von EAC gibt es noch eine theoretische Lücke, für die allerdings keine praxisrelevanten Angriffsszenarien bekannt sind: Sowohl die Chip Authentication, als auch die Terminal Authentication gelten als sicher; es gibt allerdings keine kryptographische Bindung der beiden Protokolle. Diese Lücke wurde mit Version 2, die in deutschen Personalausweisen zum Einsatz kommt, behoben.

PKI für elektronische Ausweisdokumente: Verifikation

Zuvor hatten wir uns bereits die Signatur-PKI für elektronische Ausweisdokumente angesehen. Für die Überprüfung des Terminalzertifikats, im Zusammenhang mit der EAC Terminal Authentication, müssen wir uns nun noch mit der Verifikations-PKI vertraut machen. Hierbei geht es um die Prüfung der Berechtigung, Ausweise zu lesen. Die Verifikations-PKI ist in Abb. 9.4 dargestellt.

Ein elektronischer Reisepass hat nur das Country Verifying CA (CVCA)-Zertifikat seines Aussteller-Landes gespeichert. Trotzdem kann es sein, dass die CVCA des Landes A (wie im Beispiel dargestellt) nicht alle Document Verifying CAs (DVCAs) in allen Ländern zertifiziert. Es kann Fälle geben, in denen die CVCA bestimmten Ländern (oder nur einzelnen Behörden) nicht genügend vertraut um diese zu zertifizieren. Lesegeräte, die von diesen Behörden zertifiziert wurden, können dann nicht auf die sensiblen Daten, die im Pass gespeichert sind, zugreifen.

Die Zertifikats-Gültigkeit in Deutschland gestaltet sich wie folgt:

- CVCA: 26 Monate; Verwendung: $21\frac{1}{2}$ Monate,
- DVCA: $2\frac{1}{2}$ Monate; Verwendung: 2 Monate,
- Terminal-Zertifikate: 36 Stunden; Verwendung: 24 Stunden.

Ein Widerruf, der insbesondere bei den Terminals nötig werden könnte, ist praktisch nur sehr schwer zu realisieren. Der Widerruf müsste auf den Ausweisdokumenten geprüft

werden. Deshalb wählt man einen kurzen Gültigkeitszeitraum. Wird ein Terminal z. B. gestohlen, kann damit ein Schaden nur in einem sehr kurzen Zeitraum angerichtet werden.

Die Zertifikatsprüfung durch den ePass erfordert das aktuelle Datum – der ePass hat jedoch keine eingebaute Uhr. Deshalb hat man folgende Lösung gewählt: Initial wird im ePass das Produktionsdatum als aktuelles Datum gesetzt. Nach jeder Prüfung eines neuen (Terminal-)Zertifikats wird das aktuelle Datum im ePass auf den Gültigkeitsbeginn dieses Zertifikats gesetzt.

Die Gültigkeit von Reisepässen beträgt bis zu zehn Jahre. In dieser Zeit könnten sich die Zertifikate ändern. Aus diesem Grund wird das aktuelle CVCA-Zertifikat bei der Produktion im ePass gespeichert. Bei der Änderung des CVCA-Schlüssels erhält der Pass ein Link-Zertifikat, wobei der neue öffentliche Schlüssel mit dem alten Schlüsselpaar signiert wird. Nach erfolgreicher Überprüfung tauscht der ePass das alte CVCA-Zertifikat gegen das neue aus.

9.2 Elektronischer Personalausweis

Der *elektronische Personalausweis (ePA)* wurde am 1. November 2010 in Deutschland eingeführt. Die Ziele des ePAs sind:

- Verbesserte Fälschungssicherheit,
- Möglichkeit zur Authentifizierung für Netzanwendungen,
- Vereinfachte Handhabung,
- Integration qualifizierter elektronischer Signaturen.

Der ePA ist ein verkleinerter Ausweis im „Scheckkartenformat“, wie in Abb. 9.5 dargestellt.

Der elektronische Personalausweis enthält, ähnlich wie der elektronische Reisepass, einen kontaktlosen Chip. Der Chip bietet im Gegensatz zum Chip im ePass Zusatzfunktionen, etwa zum Altersnachweis und zur *Restricted Identification*, die wir später genauer betrachten werden. Der Chip enthält ebenfalls biometrische Daten wie das Gesichtsbild

Abb. 9.5 Elektronischer Personalausweis (ePA)



Abb. 9.6 PACE-Protokoll

und die Fingerabdrücke – sofern diese freiwillig hinterlegt werden. Auf Protokollebene werden neue Extended Access Control (EAC)-Verfahren sowie das Password Authenticated Connection Establishment (PACE)-Protokoll als Ersatz für die Basic Access Control (BAC) unterstützt.

9.2.1 PACE

Das Password Authenticated Connection Establishment (PACE)-Protokoll ist der Nachfolger der Basic Access Control (BAC). Entwickelt wurde PACE vom Bundesamt für Sicherheit in der Informationstechnik (BSI) für den elektronischen Personalausweis. Mittlerweile ist eine internationale Standardisierung im Gange.

Die Authentifizierung erfolgt bei PACE durch ein gemeinsames Geheimnis (Passwort) zwischen ePA und Terminal. Das Passwort kann eine geheime Benutzer-PIN, die am ePA aufgedruckte Card Access Number (CAN) oder ein aus der maschinenlesbaren Zone abgeleiteter Schlüssel sein. Die unterschiedlichen, möglichen Passworte werden nicht für den gleichen Zweck verwendet. Die aufgedruckte CAN und der aus der maschinenlesbaren Zone abgeleitete Schlüssel sind nur für hoheitliche Anwendungen (z. B. Grenzkontrolle) gedacht und erfordern eine erfolgreiche Terminal Authentication mit einem für hoheitliche Anwendungen vorgesehenen Zertifikat.

Das Protokoll ist in Abb. 9.6 dargestellt.

Die gewählte Zufallszahl s ist 128 Bit lang. Diese wird vom ePA mit K_p , dem aus dem gemeinsamen Geheimnis abgeleiteten Schlüssel, verschlüsselt und gemeinsam mit den verschlüsselten statischen DH-Parametern (dem Generator und dem Modulus) zum

Terminal übertragen. Der dynamische DH-Parameter, genauer der Generator, wird aus dem statischen Generator, laut BSI-TR 03110, wie folgt abgeleitet: $g_{neu} = g^s \times h$ mit h (aus der DH-Gruppe) wird so gewählt, dass $\log_g h$ unbekannt ist; h soll durch einen DH-Austausch bestimmt werden. Dieser Schritt sorgt dafür, dass einerseits der DH-Austausch von s abhängt, so dass er nur erfolgreich ist, wenn beide Seiten s kennen; andererseits kann aus dem DH-Austausch aber auch nicht auf s geschlossen werden. Bei diesem Verfahren ist kein Offline-Angriff möglich: Der Angreifer kann einen geratenen K_p nicht offline verifizieren.

Ein Problem bei PACE sind die vergebenen Benutzer-Personal Identification Numbers (PINs), die zu kurz sind. Ein Durchprobieren ist eigentlich in kurzer Zeit möglich. Deshalb wird nach zwei Fehlversuchen die PIN nicht mehr akzeptiert. Ein weiterer Versuch zur Freischaltung der PIN ist erst nach der Eingabe der aufgedruckten CAN möglich. Die anschließende Freigabe ist schließlich nur noch durch die Eingabe des langen PIN Unblock Keys möglich.

9.2.2 Extended Access Control Version 2

Bei der EAC im ePA kommen die Chip Authentication und Terminal Authentication in Version 2 zum Einsatz. Im Gegensatz zum ePass erfolgt nun erst die Terminal Authentication und danach erst die Chip Authentication. Außerdem besteht eine kryptographische Bindung zwischen der Chip Authentication und der Terminal Authentication.

Terminal Authentication

Die Terminal Authentication in Version 2 ist in Abb. 9.7 dargestellt.

Der Sitzungsschlüssel wird bei der Terminal Authentication nun durch PACE etabliert – alle Nachrichten werden damit verschlüsselt und authentifiziert. Die Authentifizierung des Terminals gegenüber dem Chip im ePA erfolgt ähnlich zur Authentifizierung wie wir sie beim ePass gesehen haben; signiert wird hier folgendes:

Identität des Ausweises || r || hash(DH-Wert des Terminals) || weitere Daten. Beim ePA ist die Identität des Ausweises der Hash-Wert des öffentlichen DH-Werts des Ausweises aus dem PACE-Protokoll. Wie wir bereits beim ePass festgestellt haben, erfordert die Durchführung des DH-Schlüsselaustauschs und die Prüfung von Zertifikaten bzw. Zertifikatsketten recht aufwendige Chips.

Chip Authentication

Die Chip Authentication in Version 2 ist in Abb. 9.8 dargestellt.

Die Chip Authentication soll sicherstellen, dass der Chip als authentisch erkannt wird und einen Schlüssel zur Absicherung der gemeinsamen Kommunikation ableiten, der den zuvor verwendeten ersetzt.

Die Voraussetzung zur Authentifizierung ist, wie beim ePass, dass der statische DH-Wert geprüft werden kann. Dazu wird wieder die Passive Authentication verwendet.

Abb. 9.7 EAC Terminal
Authentication Version 2



Abb. 9.8 EAC Chip
Authentication Version 2



Das Terminal verwendet seinen DH-Wert aus der Terminal Authentication wieder. Vor der Berechnung des gemeinsamen Schlüssels K prüft der ePA, ob der DH-Wert des Terminals zu dem Hash des DH-Werts passt, den das Terminal während der Terminal Authentication gesendet hat. Dies sorgt für die zuvor bereits angesprochene kryptographische Bindung zwischen der Terminal Authentication und der Chip Authentication.

Im Gegensatz zur EAC-Chip Authentication beim ePass findet hier eine explizite Chip-Authentication statt.

9.2.3 Restricted Identification

Die *Restricted Identification* des elektronischen Personalausweises bietet eine datenschutzfreundliche Identifizierung gegenüber Diensteanbietern. Dabei wird ein „bereichsspezifischer Identifikator“ generiert, unter dem eine Person in einem Bereich – von einem Diensteanbieter – wiedererkannt werden kann. Der bereichsspezifische Identifikator eines Bereichs ist dabei nicht aus den bereichsspezifischen Identifikatoren anderer Bereiche ableitbar. Damit ist ein Tracking einer Person über verschiedene Bereiche hinweg nicht möglich. Die Identifikation erfolgt dabei auch ohne Kenntnis der realen Identität des Ausweisinhabers.

Das Protokoll ist in Abb. 9.9 dargestellt.

Der ePA hat einen geheimen DH-Wert SK_{ID} . Der ePA wendet den DH-Schlüsselaustausch an, der eigentlich einen gemeinsamen DH-Schlüssel generiert. Allerdings wird dieser nicht als Schlüssel verwendet, sondern der Hash-Wert davon stellt den bereichsspezifischen Identifikator dar. Der ePA berechnet also $hash(PK_{sector}^{SK_{ID}} \bmod N)$. Der Vorteil dieses Verfahrens besteht darin, dass eine Aufdeckung nur bei Zusammenarbeit möglich ist. Es ist kein zentrales Melderegister nötig, wo alle Identifikatoren hinterlegt sein müssen.

Der Ausweis muss prüfen können, ob der öffentliche DH-Wert des Bereichs korrekt ist. Dies geschieht dadurch, dass der Wert auch im Terminal-Zertifikat enthalten sein muss.

Das Problem bei der Restricted Identification besteht darin, dass das Protokoll über einen sicheren Kanal ausgeführt werden muss. Das Terminal muss wissen, dass der bereichsspezifische Identifikator von einem echten Ausweis berechnet wird; ansonsten wären beliebige Manipulationen möglich. Dazu ist eine vorhergehende Chip Authentication nötig. Die Chip Authentication ermöglicht allerdings ein Tracking aufgrund der

Abb. 9.9 Restricted Identification

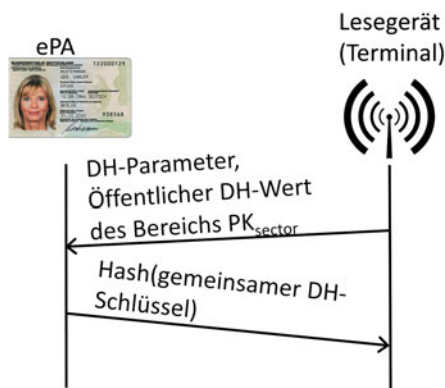
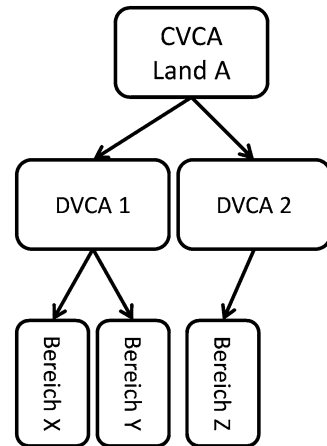


Abb. 9.10 Vergabe der öffentlichen DH-Werte für die Bereiche



Verwendung ausweisspezifischer Schlüssel.⁵ Aus diesem Grund ist die Lösung des BSI die Verwendung des gleichen Schlüsselpaars für die Chip Authentication in einer ganzen Charge von Ausweisen. Der Schlüssel für die Chip Authentication wird etwa alle drei Monate geändert. Die Lösung ist nicht problemlos. Gelingt es jemandem, den privaten Schlüssel für die Chip Authentication aus seinem Ausweis zu extrahieren, so wird der Widerruf selbst dann problematisch, wenn der erfolgreiche Angriff entdeckt wird. Es müssten dann (das Bestehen einer funktionierenden Widerrufslösung vorausgesetzt) alle Ausweise der entsprechenden Charge (Generation) zurückgerufen werden. Das wird teuer, wenn die Chargen mit identischen Schlüsselpaaren groß sind. Sind sie klein, besteht wiederum eine erhöhte Wahrscheinlichkeit des Trackings. Eine theoretisch bessere Lösung wäre der Einsatz von Gruppensignaturen, wie wir sie in Abschn. 10.1.1 kennenlernen werden. Allerdings ist der Einsatz von Gruppensignaturen auf kontaktlosen Chips derzeit noch nicht praxisreif.

Widerruf

Der bereits angesprochene Widerruf bei der Restricted Identification gestaltet sich wie folgt. Für die Vergabe der öffentlichen DH-Werte für einen Bereich sind, wie in Abb. 9.10 dargestellt, die Document Verifying CAs (DVCA) zuständig.

Schritt 1 (Initiierung CVCA): Die CVCA generiert einen privaten DH-Wert $SK_{Revocation}$ und berechnet den zugehörigen öffentlichen DH-Wert

$$PK_{Revocation} = g^{SK_{Revocation}} \mod N.$$

⁵Für die hoheitliche Anwendung wird bei der Chip Authentication ein ausweis-individuelles Schlüsselpaar verwendet.

$PK_{Revocation}$ und die DH-Parameter werden veröffentlicht.

Schritt 2 (Initiierung DVCA): Die DVCA wählt pro Bereich einen privaten DH-Wert SK_{sector} und berechnet PK_{sector} nach dem DH-Schlüsselaustausch aus:

$$PK_{sector} = PK_{revocation}^{SK_{sector}} \mod N.$$

SK_{sector} darf nicht weitergegeben werden, insbesondere nicht an die Terminals.

Schritt 3 (Sperrung CVCA): Die Sperrung läuft nun folgendermaßen ab. Die CVCA erhält einen öffentlichen DH-Wert PK_{ID} eines zu sperrenden Chips. Die CVCA berechnet PK_{ID-rev} nach dem DH-Schlüsselaustausch:

$$PK_{ID-rev} = PK_{ID}^{SK_{revocation}} \mod N = (g^{SK_{ID}})^{SK_{revocation}}.$$

Danach leitet die CVCA PK_{ID-rev} an alle DVCAs weiter.

Schritt 4 (Sperrung DVCA): Die DVCAs berechnen für jeden Bereich den zugehörigen bereichsspezifischen Identifikator nach dem DH-Schlüsselaustausch:

$$\text{hash}(PK_{ID-rev}^{SK_{sector}} \mod N) = \text{hash}(((g^{SK_{ID}})^{SK_{revocation}})^{SK_{sector}} \mod N).$$

Zur Erinnerung: Der ePA berechnet $\text{hash}(PK_{sector}^{SK_{ID}} \mod N)$. Aufgrund der Berechnung von PK_{sector} ist das:

$$\text{hash}((PK_{revocation}^{SK_{sector}})^{SK_{ID}} \mod N) = \text{hash}(((g^{SK_{revocation}})^{SK_{sector}})^{SK_{ID}} \mod N).$$

Dieser Wert ist identisch zu dem Wert, den die DVCAs berechnen.

Der Vorteil dieses Verfahrens ist, dass die CVCA und die DVCAs zusammenarbeiten müssen, um die Anonymität zu brechen – eine Instanz alleine kann die Anonymität also nicht brechen.

9.2.4 Weitere Anwendungen

Auslesen von Attributen

Das Auslesen von Attributen aus dem ePA ist nur mit einem passenden Berechtigungszertifikat möglich. Ein Jugendschutzsystem benötigt bspw. nur das Alter; ein Auslesen der Anschrift ist nicht möglich. Die Vergabestelle für Berechtigungszertifikate für Anwendungen der Privatwirtschaft ist das Bundesverwaltungsamt. Die zugehörige Vorschrift, § 21 Abs. 2 und 3 Personalausweisgesetz, lautet wie folgt:

(2) Die Berechtigung nach Absatz 1 ist zu erteilen, wenn

1. der angegebene Zweck nicht rechtswidrig ist,
2. der Zweck nicht in der geschäftsmäßigen Übermittlung der Daten besteht und keine Anhaltspunkte für die geschäftsmäßige oder unberechtigte Übermittlung der Daten vorliegen,
3. der antragstellende Diensteanbieter die Erforderlichkeit der zu übermittelnden Angaben für den beschriebenen Zweck nachgewiesen hat,
4. die Anforderungen, insbesondere an Datenschutz und Datensicherheit, gemäß der Rechtsverordnung nach § 34 Nr. 7 erfüllt sind und
5. keine Anhaltspunkte für eine missbräuchliche Verwendung der Berechtigung vorliegen.

Der Diensteanbieter hat durch Selbstverpflichtung die Anforderungen nach Nummer 4 schriftlich zu bestätigen und auf Anforderung nachzuweisen.

(3) Die Berechtigung ist zu befristen. Die Gültigkeitsdauer darf einen Zeitraum von drei Jahren nicht überschreiten. Die Berechtigung darf nur von dem im Berechtigungszertifikat angegebenen Diensteanbieter und nur zu dem darin vorgesehenen Zweck verwendet werden. Die Berechtigung kann mit Nebenbestimmungen versehen und auf entsprechenden Antrag wiederholt erteilt werden.

Die Anwendungssoftware soll die Attributsfreigabe durch den Nutzer bestätigen lassen. Die beim Auslesen der Attribute übermittelten Daten werden nicht signiert. Dies soll den Nutzen für den Datenhandel reduzieren. Eine Authentifizierung der Daten erfolgt zumindest implizit durch die Chip Authentication.

Altersverifikation

Das Ziel der Altersverifikation mit dem ePA ist die Prüfung, ob ein Nutzer ein Mindestalter erreicht hat und damit bspw. berechtigt ist, eine Webseite mit nicht jugendfreien Inhalten abzurufen. Das genaue Geburtsdatum wird dabei nicht preisgegeben.

Nach der erfolgreichen Chip Authentication schickt das Terminal ein Datum an den Ausweis. Der Ausweis antwortet – einmal pro PACE-Authentifizierung –, ob das Geburtsdatum vor dem genannten Datum liegt. Dass der Ausweis nur einmal pro Authentifizierung antwortet, ist wichtig. Sonst könnten einfach verschiedene Geburtsdaten durchprobiert werden. In der implementierten Lösung müsste der Nutzer hier jedes Mal sein Passwort eingeben.

Ablauf des Online-Einsatzes

Abb. 9.11 fasst den Online-Einsatz des elektronischen Personalausweises, bspw. zum Auslesen von Attributen bzw. zur Durchführung der Altersverifikation, noch einmal (vereinfacht) zusammen. Die Abbildung soll deutlich machen, wie die unterschiedlichen Protokolle, die wir in diesem Abschnitt kennengelernt haben, zusammenwirken.

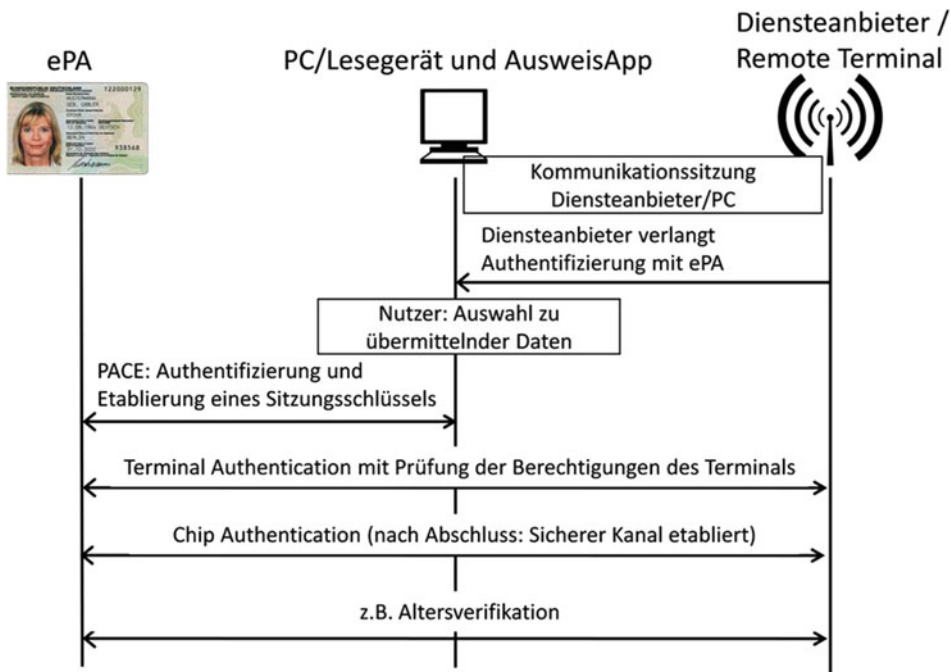


Abb. 9.11 Online-Einsatz des elektronischen Personalausweises

9.2.5 Exkurs: Elektronische Signaturen

„Elektronische Signaturen“ ist ein juristischer Begriff für Signaturen über digitale Daten. Das „Signaturgesetz“⁶ kennt einfache, fortgeschrittene und qualifizierte Signaturen. Das (deutsche) Signaturgesetz wurde am 01.07.2016 durch die EU-Verordnung Nr. 910/2014 über elektronische Identifizierung und Vertrauensdienste für elektronische Transaktionen im Binnenmarkt (eIDAS-Verordnung) abgelöst. Die unterschiedlichen Arten elektronischer Signaturen haben sich nicht geändert.

Einfache Signaturen sind Daten, die der Authentifizierung dienen. Hier werden keine Sicherheitsvorkehrungen erfordert. Damit sind auch keine Rechtsfolgen an das Vorliegen solcher einfachen Signaturen geknüpft. Bei fortgeschrittene Signaturen kommen zusätzlich kryptographische Verfahren zum Einsatz. Qualifizierte Signaturen erfordern eine sichere Speicherung des (privaten) Signaturschlüssels und eine 2-Faktor-Authentifizierung für die Signierung. Man spricht auch davon, dass bei qualifizierten Signaturen ein „qualifiziertes Zertifikat“ verwendet wird. Ein solches qualifiziertes Zertifikat muss von

⁶Tatsächlich liegen elektronischen Signaturen die *Europäische Signaturrichtlinie*, das *Deutsche Signaturgesetz*, die *Signaturverordnung*, sowie der *Algorithmenkatalog der Bundesnetzagentur* zugrunde.

einer Zertifizierungsstelle stammen, die bestimmte Voraussetzungen erfüllt, u. a. eine Deckungsvorsorge für den Fall von Pflichtverletzungen. Darüber hinaus gibt es noch qualifizierte Signaturen mit Anbieterakkreditierung – diese unterscheiden sich von den „normalen“ qualifizierten Signaturen allerdings nur durch administrative Unterschiede. Qualifizierte Signaturen erfüllen den „Anscheinsbeweis“ für die Echtheit des signierten Dokuments – damit geht ein besonderer Beweiswert im Zivilprozess einher (§ 371a ZPO (Zivilprozessordnung)). Dokumente mit qualifizierter Signatur können die „elektronische From“ (§ 126a BGB (Bürgerliches Gesetzbuch)) erfüllen und damit als Ersatz für die Schriftform gelten.

Der elektronische Personalausweis als „sichere Signaturerstellungseinheit“ unterstützt die qualifizierte Signatur. Allerdings ist von Haus aus kein Zertifikat enthalten. Die Zertifikatsausstellung erfolgt durch (private) Zertifizierungsdienstanbieter (Certificate Authoritys (CAs)). Dafür ist eine vorhergehende Authentifizierung, bspw. über die Authentifizierungsfunktion des ePAs, nötig. Gegenüber dem ePA authentifiziert sich der Signierende beim Signieren eines Dokuments schließlich mittels PACE.

9.2.6 Administrative Aspekte

Der Antragsprozess für einen elektronischen Personalausweis gestaltet sich grundsätzlich wie bisher; allerdings fällt die Gebühr mit 28, 80 € für Bürger ab 24 Jahren deutlich höher aus als früher. Neu ist hingegen, dass die Bundesdruckerei einen Brief mit Sperrkennwort, Start-PIN und PUK an die Meldeanschrift versendet. Der Ausweis hingegen geht an die Meldebehörde. Die Freischaltung der Online-Authentifizierung ohne Start-PIN ist bei der Meldebehörde möglich. Bei einer späteren Freischaltung fallen Gebühren an.

Bei Verlust des Ausweises ist eine Sperrung der Online-Authentifizierung unter der Angabe des Sperrkennworts nötig (telefonisch unter 0180-1-33 33 33). Das Sperrkennwort kann (persönlich) bei der Meldebehörde erfragt werden. Außerdem ist die Meldebehörde über den Verlust zu informieren. Sofern die Signaturfunktion genutzt wird, also ein qualifiziertes Zertifikat am ePA vorhanden ist, muss dieses separat beim Zertifizierungsdienstanbieter widerrufen werden.

Außerdem gibt es seit der Einführung des ePAs eine Einschränkung der Nutzung des Ausweises als „Pfand“. So heißt es in § 1 Abs. 1 Satz 3 PAuswG:

„Vom Ausweisinhaber darf nicht verlangt werden, den Personalausweis zu hinterlegen oder in sonstiger Weise den Gewahrsam aufzugeben.“

ePA-basierte Authentifizierung in eigenen Anwendungen

Zusätzliche Information

Die Nutzung der Nutzer-Authentifizierung mittels ePA in eigenen Anwendungen erfordert ein Berechtigungszertifikat. Die Vergabestelle für Berechtigungszertifikate verlangt dabei:

- eine Begründung für die enthaltenen Datenfelder,
- die Erfüllung funktionaler Anforderungen gemäß den BSI-Richtlinien (BSI-TR-03130 sowie BSI-TR-03112),
- die Erfüllung der Datenschutz- und Datensicherheitsanforderungen.

Die Gebühren für Berechtigungszertifikate finden sich in § 3 PAuswGebV. So schlägt die Erteilung einer Berechtigung mit 102 € zu Buche; die Versagung einer Berechtigung mit 80 €. Für die Rücknahme bzw. den Widerruf einer Berechtigung werden 115 € fällig.

Hinsichtlich der Umsetzung der ePA-basierten Authentifizierung in eigenen Anwendungen bestehen mehrere Möglichkeiten. So kann entweder ein lokaler eID-Server oder ein selbst betriebener, über das Internet mit dem Webserver verbundener, eID-Server zum Einsatz kommen. Eine weitere Möglichkeit besteht im Outsourcing. Dabei handelt es sich um eine Auftragsdatenverarbeitung gemäß § 11 Bundesdatenschutzgesetz (BDSG). Der durchführende Anbieter muss der Vergabestelle für Berechtigungszertifikate bekanntgegeben werden.

Die geforderten Sicherheitsanforderungen beim eID-Server-Betrieb sind in der BSI-TR-03130 zu finden – im Wesentlichen wird die Einhaltung des Stands der Technik gefordert. Wesentliche Gefährdungen müssen dabei identifiziert und adressiert werden. Des Weiteren müssen die BSI-Richtlinien zu Schlüssellängen etc. eingehalten werden. Die *Richtlinie Technische und organisatorische Anforderungen zur Nutzung von Berechtigungszertifikaten vom 17. Mai 2011*, die auf die BSI-Richtlinien Bezug nimmt, wurde durch das Bundesverwaltungsamt im eBundesanzeiger veröffentlicht.

9.3 Fazit

In diesem Kapitel haben wir uns mit elektronischen Ausweisdokumenten beschäftigt, die die meisten von uns besitzen und die wir mehr oder weniger häufig verwenden.

Wir haben zu Beginn die Sicherheitsmaßnahmen des elektronischen Reisepasses kennengelernt. Dabei haben wir gesehen, wie die personenbezogenen Daten vor unberechtigtem Auslesen bzw. Mitlesen von Dritten geschützt werden. Vordergründig waren dies klassische IT-Sicherheits-Protokolle, die in diesem Kontext auch dem Datenschutz dienlich sind. Im Zusammenhang mit dem elektronischen Personalausweis haben wir darüber hinaus mit der *Restricted Identification*, der *Funktion zum Auslesen von Attributen* und der *Funktion zur Altersverifikation* Protokolle kennengelernt, die als vordergründiges Ziel den Datenschutz haben – und sich damit aus unserer Sicht besonders gut eignen, um diese in diesem Buch genauer zu betrachten. Mit dem Verständnis dieser Protokolle können Sie sich selbst ein Bild davon machen, inwieweit Sie den (gerade zu Beginn der Einführung des elektronischen Personalausweises) häufig genannten Kritikpunkten zu den Gefahren des Trackings und dem unberechtigten Auslesen der Daten folgen. Unbestritten ist sicherlich die mangelnde Sicherheit bei der Verwendung einfacher Lesegeräte ohne eigene Tastatur. Die Zukunft wird zeigen, inwieweit der elektronische Personalausweis, und hierbei insbesondere die Online-Funktionen, von den Bürgern genutzt werden wird und welche Dienste es geben wird, die eine Verwendung des ePAs unterstützen. Durch die

EU-Verordnung Nr. 910/2014 über elektronische Identifizierung und Vertrauensdienste für elektronische Transaktionen im Binnenmarkt (eIDAS-Verordnung) könnte die eID-Funktion des elektronischen Personalausweises neuen Auftrieb erhalten. Die Verordnung ermöglicht seit 1. Juli 2016 die Verwendung von nationalen Identifizierungsmitteln auch für die Nutzung von Diensten der öffentlichen Verwaltung in anderen EU-Staaten sowie im Europäischen Wirtschaftsraum (EWR). Die eID-Funktion im elektronischen Personalausweis stellt ein derartiges Identifizierungsmittel (mit hohem Vertrauensniveau) für deutsche Bürger dar, das von den anderen Staaten ab September 2018 anerkannt werden muss.

Im nächsten Kapitel werden wir uns mit weitergehenden datenschutzfördernden Technologien, sogenannten Privacy-Enhancing Technologies (PETs), beschäftigen. Im Zusammenhang mit der in diesem Kapitel vorgestellten Restricted Identification werden wir mit Gruppensignaturen ein kryptographisches Primitiv kennenlernen, das in zukünftigen Anwendungen sehr vielversprechend zu sein scheint.

9.4 Übungsaufgaben

Aufgabe 1

Die *Basic Access Control* verwendet Seriennummer, Geburtsdatum und Ablaufdatum (jeweils mit Prüfziffer) als Eingaben zur Erzeugung des Schlüssels. Nehmen Sie an, der Inhaber eines „E-Passes“ stehe Ihnen gegenüber, und Sie wollten heimlich das Gesichtsbild aus dem Pass in dessen Jackentasche auslesen. Ihr Lesegerät hat die dafür benötigte Reichweite. Wie sollten Sie bei einem Angriff vorgehen? Nutzen Sie möglichst viel (vorhandenes oder recherchierbares) Wissen! Geben Sie eine begründete Schätzung ab, wie viele benötigte Versuche Sie erwarten (und wie lange das dauern würde)

- bei serieller Vergabe der (nur aus Ziffern bestehenden) Seriennummern, entsprechend den bis Oktober 2007 gültigen Regeln!
- bei Vergabe der Seriennummern, entsprechend den derzeit gültigen Regeln (informieren Sie sich, wie viele Zeichen zulässig sind)!

Sollten Sie benötigte Informationen nicht finden, so treffen Sie plausible Annahmen!

Aufgabe 2

Warum kann bei der *Extended Access Control* die Terminal Authentication auch mit einem lange abgelaufenen Zertifikat des Terminals erfolgreich verlaufen? Wie kann man diesem Problem entgegenwirken?

Aufgabe 3

Wie wird bei der *Restricted Identification* sichergestellt, dass verschiedene Diensteanbieter auch verschiedene Pseudonyme erhalten, und wieso können die Diensteanbieter dies nicht umgehen?

Literatur

1. Dennis Kügler und Ingo Naumann. Sicherheitsmechanismen für kontaktlose Chips im deutschen Reisepass. *Datenschutz und Datensicherheit*, 31(3):176–180, 2007.

Zusammenfassung

Bisher haben wir in diesem Buch datenschutzfördernde Verfahren vorgestellt, die bereits Einzug in die Praxis erhalten haben. In diesem Kapitel beschäftigen wir uns nun mit weiteren kryptographischen Ansätzen, die sich heute überwiegend noch im „Forschungsstadium“ befinden, in Zukunft aber Einzug in die Praxis der Privacy-Enhancing Technologies (PETs) finden können. Zunächst betrachten wir in Abschn. 10.1 weitere Signaturverfahren, die dem Datenschutz dienlich sind. Im Anschluss daran lernen wir in Abschn. 10.2 die mächtigen Werkzeuge rund um die Secure Multiparty Computation (MPC) kennen. In Abschn. 10.3 und Abschn. 10.4 betrachten wir schließlich Zero-Knowledge Proofs (ZKPs) bzw. anonyme Berechtigungsnachweise.

Lernziele

Am Ende dieses Kapitels sollten Sie weitere kryptographische Verfahren kennen, die in Zukunft als Grundlage für Privacy-Enhancing Technologies (PETs) dienen werden.

10.1 Weitere Signaturverfahren

Digitale Signaturen haben wir bereits in Abschn. 2.3.2 kennengelernt. In Abschn. 2.3.3 haben wir mit der *blinden Signatur* eine Abwandlung eines „normalen“ Signaturverfahrens kennengelernt, die sich als Grundlage für datenschutzfreundliche Protokolle, wie etwa das anonyme Bezahlen nach Chaum (Abschn. 6.2), eignet. In diesem Abschnitt lernen wir mit der *Gruppensignatur* und der *Ringsignatur* zwei weitere Signaturverfahren kennen, die bestimmte Datenschutz-Eigenschaften aufweisen.

10.1.1 Gruppensignatur

Digitale Signaturen erlauben den Nachweis der Integrität und Authentizität eines Dokuments unter der *Identität eines Unterzeichners* sowie Verbindlichkeit (Nicht-Abstreitbarkeit).

Eine Gruppensignatur ist im Wesentlichen eine „normale“, digitale Signatur, die im Gegensatz zu dieser allerdings nur den Nachweis erbringt, dass ein *Teilnehmer aus einer bestimmten Gruppe ein Dokument signiert* hat. Die Gruppen werden durch einen sogenannten Gruppenmanager verwaltet, der in der Regel die Identität des Signierenden im Nachhinein aufdecken und die Gruppenmitgliedschaft widerrufen kann.

Algorithmen

Die für ein Gruppensignaturverfahren im Allgemeinen verwendeten Algorithmen sind die folgenden. M bezeichnet den Gruppenmanager und P ist die Menge der Gruppenmitglieder P_1, \dots, P_n .

- *Setup* (zwischen M und P)
 - Eingabe: Sicherheitsparameter
 - Ausgabe:
 - * Privater Schlüssel x_i für jedes Gruppenmitglied P_i
 - * Öffentlicher Schlüssel für die Gruppe
 - * Offenlegungsschlüssel für den Gruppenmanager
- *Sign* (durch P_i)
 - Eingabe: Nachricht m , privater Schlüssel x_i
 - Ausgabe: Signatur von m
- *Verify* (durch beliebige Partei)
 - Eingabe: Nachricht m , Signatur von m , öffentlicher Schlüssel für die Gruppe
 - Ausgabe: Wahrheitswert (Signatur gültig/nicht gültig)
- *Open* (durch Gruppenmanager)
 - Eingabe: Nachricht m , Signatur von m , Offenlegungsschlüssel
 - Ausgabe: Identität des Signierenden i oder Fehlausgabe

Anwendung findet ein modifiziertes Gruppensignaturverfahren bspw. in der „Direct Anonymous Attestation“, die den Nachweis der Authentizität eines Trusted Platform Module (TPM) ohne Preisgabe der vollständigen Identität und ohne Tracking-Möglichkeit erlaubt. Die Direct Anonymous Attestation wurde durch die Trusted Computing Group standardisiert. Nähere Informationen zu dem Verfahren finden sich bei BRICKELL et al. [3].

Daneben gibt es einige Vorschläge zur Verwendung im Identitätsmanagement, z. B. zum Nachweis von Berechtigungen. Bei Fahrzeug-Ad-hoc-Netzen könnte die Gruppensignatur etwa für den Nachweis der Authentizität einer Nachricht sorgen, ohne dass ein Tracking möglich ist.

10.1.2 Ringsignatur

Die *Ringsignatur* bietet ähnliche Eigenschaften wie die Gruppensignatur. Allerdings ist hier kein Gruppenmanagement nötig; die Gruppenbildung kann „ad-hoc“ beim Signieren erfolgen. Damit ist kein Gruppenmanager vorhanden der einen Widerruf der Anonymität ermöglichen könnte. Der Rechenaufwand ist bei der Ringsignatur deutlich geringer als bei der Gruppensignatur, allerdings ist die praktische Nutzbarkeit aufgrund der mangelnden Widerrufsmöglichkeit eingeschränkt.

Ein möglicher Anwendungsfall für die Ringsignatur, der bei RIVEST et al. [10] genannt wird, besteht im anonymen Zuspätschicken („Leaking“) von Geheimnissen:

Beispiel

Bob ist Minister in Lower Kryptonia. Er möchte pikante Informationen über die Eskapaden des Premierministers an die Presse geben. Dabei möchte er anonym bleiben. Ein Journalist soll sich allerdings darauf verlassen können, dass Bob tatsächlich Mitglied des Kabinetts ist. Die Gruppensignatur reicht hierfür nicht: eine Gruppe müsste erst etabliert werden und Bobs Identität könnte offengelegt werden. Deshalb verwendet Bob eine Ringsignatur, wobei der Ring alle Minister enthält.

10.2 Secure Multiparty Computation

Bei der *Secure Multiparty Computation* (*MPC*) haben wir es mit einer Reihe von Parteien P_1, \dots, P_n zu tun, die jeweils in Besitz von „privaten“ Informationen x_1, \dots, x_n sind. Nun möchten sie eine Funktion $y = f(x_1, \dots, x_n)$ über ihre Daten ausführen, um ein gemeinsames Ergebnis in Abhängigkeit ihrer Eingaben – der privaten Informationen – zu berechnen, ohne dass sie dabei ihre jeweiligen Eingaben den anderen Parteien mitteilen. *MPC* ist im Bereich der Kryptographie schon seit Jahrzehnten ein Forschungsthema. Im Wesentlichen geht es dabei darum, kryptographische Protokolle zu entwickeln, die eine „private“ Ausführung der Funktion f erlauben, d. h. die Beteiligten lernen durch die Ausführung der Funktion (außer dem Ergebnis) keine zusätzlichen Informationen über die Eingaben der anderen Beteiligten. Selbst Betrüger, die an der Ausführung der Funktion beteiligt sind lernen bei einigen Protokollen nichts über die Eingaben der ehrlichen Beteiligten. Das Ziel der (theoretischen) Forschung war es lange Zeit, möglichst „allgemeine“ Protokolle zu entwickeln, die für eine Vielzahl an unterschiedlichen Aufgabenstellungen eingesetzt werden können. Der Rechenaufwand ist bei den meisten dieser Protokolle sehr hoch.

10.2.1 Klassische MPC-Protokolle

ANDREW YAO gilt als einer der Begründer des Forschungsgebiets rund um das Thema *Secure Multiparty Computation* (MPC). Das 1982 von ihm vorgestellte und nach ihm benannte „Millionärsproblem“ (*Yao's Millionaires' Problem*) gilt als das erste sichere „two-party computation“-Protokoll. 1986 hat er mit dem nach ihm benannten *Yao's Garbled Circuit* einen weiteren Meilenstein in der Forschung in diesem Gebiet gelegt.

Yao's Millionaires' Problem

In Abschn. 8.2.4 hatten wir bereits die Anwendung einer Abwandlung von Yao's Millionaires' Problem in Form des *Socialist Millionaire's Protocol* (SMP) beim SIGMA-Protokoll bei OTR Messaging kennengelernt.

Der Name des Protokolls rührt daher, dass es beim „klassischen“ Yao's Millionaires' Problem darum geht, dass zwei Millionäre wissen möchten, wer reicher ist, ohne dabei dem jeweils anderen den genauen Reichtum zu offenbaren. Abstrakt geht es bei diesem Protokoll um einen Vergleich von Daten, ohne die Daten offenzulegen. Eine vertrauenswürdige Partei, die den Vergleich durchführt, gibt es gerade nicht.

Nehmen wir an, Alice hat das Vermögen V_A und Bob das Vermögen V_B – beide Vermögen seien Elemente der Menge $\{1, \dots, 10\}$ (bspw. in Millionen). Des Weiteren sei k eine bijektive Trapdoor-Einwegfunktion (d. h. eine Funktion, die sich nur dann effizient umkehren lässt, wenn man eine Zusatzinformation kennt)¹ auf den Zahlen mit der Länge N Bit. Nur Alice besitzt den privaten Schlüssel, d. h. nur sie kann die Umkehrung der Einwegfunktion (also k^{-1}) effizient berechnen.

Im ersten Schritt wählt Bob nun eine zufällige Zahl X der Länge N Bit und übermittelt $k(X) - V_B$ an Alice. Alice berechnet Y_1, \dots, Y_{10} mit $Y_i = k^{-1}(k(X) - B + i)$ für $i \in \{1, \dots, 10\}$. Alice wählt außerdem eine zufällige Primzahl P der Länge $\frac{N}{2}$ Bit, so dass $|Z_i - Z_j| \geq 2$ für alle Paare aus der Folge Z_1, \dots, Z_{10} mit $Z_i = Y_i \bmod P$ und $i, j \in \{1, \dots, 10\}$. Alice sendet $Z_1, \dots, Z_{V_A}, Z_{V_A+1} + 1, \dots, Z_{10} + 1$ und P an Bob. Wenn der B -te Eintrag aus der Liste gleich $X \bmod P$ ist, dann ist $A \geq B$, andernfalls ist $A < B$. Bob informiert Alice über das Resultat.

Leibenger et al. [9] schlagen etwa die Verwendung des Protokolls bei Anwendungen des „Quantified Self“ vor. Das Protokoll eignet sich für einen datenschutzfreundlichen Vergleich von personenbezogenen Daten unterschiedlicher Nutzer.

Yao's Garbled Circuit

Yao's Garbled Circuit ist eines der bekanntesten MPC-Protokolle, das die Evaluation einer Booleschen Schaltung erlaubt. Es bildet die Grundlage für eine Reihe von weiteren MPC-Protokollen.

¹RSA (Abschn. 2.3.1) basiert auf der Trapdoor-Einwegfunktion der Multiplikation zweier großer Primzahlen; die Umkehrung (also die Primfaktorzerlegung) ist ohne Zusatzwissen nicht effizient durchführbar.

Es gibt einen Konstrukteur einer Schaltung, Alice, und einen Evaluator Bob. Alice verschlüsselt (bzw. aus dem Englischen „garble“ übersetzt: „macht unkenntlich“) die boolesche Schaltung und sendet sie an Bob. Bob erhält dann von Alice einen Schlüssel X_a , der von Alices Input a abhängt. Bob benötigt seinerseits einen Schlüssel Y_b , der von seinem Input b abhängt; Alice soll den Input b allerdings nicht erfahren. Deshalb führen die beiden einen *Oblivious Transfer* (wie in Abschn. 10.2.2 beschrieben) durch, bei dem Alice Y_0 und Y_1 als Eingabe verwendet und Bob als Eingabe b verwendet und als Ergebnis der Ausführung Y_b als Ausgabe erhält; Alice lernt bei der Protokollausführung nichts. Die Verschlüsselung der Schaltung erfolgt beispielhaft für ein *AND*-Gatter folgendermaßen: Alice berechnet 4 Ciphertexte $C_{00} = E_{X_0, Y_0}(Z_0)$, $C_{01} = E_{X_0, Y_1}(Z_0)$, $C_{10} = E_{X_1, Y_0}(Z_0)$, $C_{11} = E_{X_1, Y_1}(Z_1)$, vertauscht sie zufällig und sendet sie an Bob. Bob kann nur denjenigen Ciphertext entschlüsseln, der mithilfe von X_a und Y_b verschlüsselt wurde; er lernt also Z_{ab} .

Yao's Garbled Circuit-Protokoll bietet Sicherheit lediglich gegen passive Angreifer. Um eine betrügerische Partei zu überführen, die eine andere boolesche Schaltung als die vereinbarte übermittelt, kann ein *cut-and-choose*-Ansatz ähnlich wie wir ihn in Abschn. 6.2 kennengelernt haben, zum Einsatz kommen.

10.2.2 Anwendungen der Secure Multiparty Computation

Erst in den letzten Jahren haben Forscher Protokolle entwickelt, die als Lösung für spezielle Aufgabenstellungen in der Praxis dienen sollen. So berichten BOGETOFT et al. [2] über die erste größere Anwendung von MPC im Jahr 2008: in Dänemark wurde damit eine landesweite „Double Auction“ für den Zuckerrübenmarkt realisiert, bei der die Farmer die Angaben über ihre Gebote – die Aussagen über ihre wirtschaftliche Lage zulassen – nicht gegenüber dem Auktionär preisgeben mussten und dieser trotzdem ein Ergebnis berechnen konnte. Wir werden darauf später zurückkommen, wenn wir über Private Set Intersection (PSI) sprechen. Als weitere Beispiele für Anwendungen von MPC-Protokollen können datenschutzgerechte Intrusion Detection, datenschutzgerechtes Data Mining, datenschutzgerechte statistische Analyse etc. genannt werden.

Private Information Retrieval

Private Information Retrieval (PIR) ist ein spezielles MPC-Problem, für das in den letzten Jahren einige Lösungen erforscht wurden. Mittels PIR-Protokoll lässt sich die Zugriffsstruktur eines Clients bei einem (Datenbank)-Betreiber (dem Server) verstecken. Der Client möchte das i -te Bit aus einer Bit-Sequenz, die beim Server abgespeichert ist, abfragen, ohne dass der Server erfährt, auf welches Bit der Client zugreifen möchte. In zahlreichen Forschungsarbeiten wurden PIR-Protokolle entwickelt, die vor allem das Ziel haben, möglichst wenig Kommunikations-Overhead aufzuweisen, d. h. dass etwa nicht die komplette Bitsequenz übertragen werden muss.

Oblivious Transfer

Oblivious Transfer verfolgt ein ähnliches Ziel wie *PIR*. Der Sender überträgt hierbei einen bestimmten Teil einer Information an den Empfänger. Er erfährt dabei nicht, um welchen Teil es sich handelt.

Private Set Intersection

Private Set Intersection (PSI) erlaubt es zwei Parteien, die Schnittmenge ihrer jeweils (geheimen) Mengen zu berechnen, ohne Informationen über jene Elemente dem jeweils anderen zu offenbaren, die nicht in der Schnittmenge liegen. PSI ist eines der am besten erforschten Anwendungen der *MPC*. Es gibt zahlreiche Protokolle und eine Reihe von Anwendungen, für die der Einsatz von PSI vorgeschlagen und umgesetzt wurde. Als mögliche Einsatzfelder für PSI wurden etwa der Abgleich von Geheimdienstinformationen zwischen Behörden unterschiedlicher Länder [6] bzw. DNA-Vaterschaftstests [1] vorgeschlagen. PSI hat auch Anwendung beim datenschutzgerechten Abgleich von Freunden/Kontakten bei sozialen Netzwerken gefunden [7].

10.3 Zero-Knowledge Proof

Zero-Knowledge Proofs (ZKPs) ermöglichen den Beweis der Kenntnis eines Geheimnisses, ohne irgendetwas über das Geheimnis selbst preiszugeben.

Ein bekanntes ZKP-Protokoll ist das *Fiat-Shamir-Protokoll* [8], bei der die Kenntnis einer Quadratwurzel einer bekannten Quadratzahl bewiesen wird, ohne die Quadratwurzel zu offenbaren.²

ZKPs, einschließlich der nicht-interaktiven Varianten, dienen als Basis zahlreicher kryptographischer Protokolle. Sie kommen bspw. bei elektronischen Wahlverfahren, Gruppensignaturverfahren und anonymen Berechtigungsnachweisen zum Einsatz. In der Praxis haben sich ZKPs bisher allerdings aufgrund der nötigen Interaktion (dem Austausch vieler Nachrichten) nicht durchsetzen können.

Beispiel

Alice kennt zwei (große) zueinander isomorphe Graphen G_1 und G_2 , sowie den zugehörigen Isomorphismus i und möchte die Kenntnis des Isomorphismus gegenüber Bob, der nur die Graphen kennt, beweisen.

Alice wählt zunächst zufällig ein $r \in \{1, 2\}$ und berechnet einen zufälligen Isomorphismus von G_1 , d. h. sie benennt die Knoten um. Der entstehende Graph heißt H . Alice schickt H an Bob.

Bob wählt im nächsten Schritt zufällig ein $s \in \{1, 2\}$ und verlangt von Alice, einen Isomorphismus zwischen H und G_s vorzulegen.

²Die Sicherheit beruht auf der Schwierigkeit, Quadratwurzeln im Restklassenring \mathbb{Z}_n zu berechnen.

Alice legt den Isomorphismus vor und Bob überprüft ihn.

Dieser Prozess wird n mal wiederholt.

Die Berechnung des von Bob angeforderten Isomorphismus ist einfach. Falls $r = s$, handelt es sich tatsächlich um den von Alice gewählten Isomorphismus. Nur in diesem Fall könnte ein Betrüger auch ohne Kenntnis des ursprünglichen Isomorphismus richtig antworten. Der Prozess wird aus diesem Grund n mal wiederholt. Andernfalls handelt es sich um eine Verkettung des ursprünglichen mit dem von Alice gewählten Isomorphismus. Die Überprüfung der Antwort ist ebenfalls einfach.

Bob hat nach Ausführung des ZKPs keine Informationen über den Isomorphismus i gelernt, die er im Anschluss nutzen könnte, um gegenüber einem Dritten zu beweisen zu versuchen, dass er den Isomorphismus kennt.

10.4 Anonyme Berechtigungsnachweise

Anonyme Berechtigungsnachweise (Anonymous Credentials) erlauben den Nachweis einer Berechtigung, ohne dabei die eigene Identität preiszugeben.

Das Grundprinzip von anonymen Berechtigungsnachweisen ist, dass ein Aussteller für einen Nutzer einen Berechtigungsnachweis ausstellt. Der Nutzer kann gegenüber einem Prüfer den Besitz eines Berechtigungsnachweises vom Aussteller beweisen. Der Prüfer erfährt sonst nichts über den Nutzer, insbesondere keine Identität. Außerdem können Berechtigungsnachweise gegenüber verschiedenen Prüfern nicht miteinander in Verbindung gebracht werden. Es existieren Varianten von Berechtigungsnachweisen, die bspw. eine einmalige oder mehrmalige Benutzung erlauben.

10.4.1 Probleme

Durch die gewährleistete Anonymität bei anonymen Berechtigungsnachweisen ergeben sich eine Reihe von Problemen.

So ist etwa der Widerruf von Berechtigungen schwierig. Hierfür wurden unterschiedliche Protokolle entwickelt, die einen Widerruf bspw. global oder nur gegenüber einem einzelnen Prüfer erlauben.

Zudem ist die unerlaubte Weitergabe von Berechtigungen schwer nachvollziehbar. Lösungsansätze hierfür bestehen in der Verknüpfung mit „wertvollen“ Informationen, die Berechtigungsinhaber ungern weitergeben, bzw. in der Bindung an Hardware (Smartcards o. Ä.).

Der Rechenaufwand bei anonymen Berechtigungsnachweisen ist durch die Verwendung einer Vielzahl an „komplexen“ kryptographischen Protokollen, wie etwa *Zero-Knowledge Proofs* (siehe Abschn. 10.3) typischerweise höher als bei „klassischen“

IdM-Verfahren. Dies mag auch der Grund dafür sein, warum sich anonyme Berechtigungsnachweise, mit Ausnahme weniger Beispiele, in der Praxis bis heute kaum durchgesetzt haben.

10.4.2 Verfahren

DAVID CHAUM hat 1985 das Konzept der *Anonymous Credentials* entwickelt [5]. Es bildet die Grundlage für moderne, nutzerzentrierte Identitätsmanagement (IdM)-Systeme. Ein Nutzer erhält von einer Trusted Third Party (TTP) ein „Credential“, das bestimmte (geprüfte) Aussagen (Attribute) über den Nutzer enthält. Dieses Credential ist an einen privaten Schlüssel gebunden, den nur der Nutzer kennt. Möchte der Nutzer gegenüber Dritten bestimmte Attribute aus dem Credential offenbaren und beweisen, so kann er dies unter einem Pseudonym datensparsam tun, ohne dabei die anderen Attribute offenbaren zu müssen. Außerdem ist es kollaborierenden Dritten nicht möglich, die Informationen über die verwendeten Pseudonyme miteinander in Verbindung zu bringen – *Unverkettbarkeit* ist also gegeben. Das Verfahren basiert auf der Verwendung von *blinden Signaturen*, wie wir sie in Abschn. 2.3.3 kennen gelernt haben. Außerdem hat Chaum vorgeschlagen, dass dieser Ansatz, basierend auf *Einmal-Pseudonymen*, auch für anonymes Bezahlen verwendet werden kann. Dieses Verfahren haben wir in Abschn. 6.2 betrachtet.

CAMENISCH et al. [4] haben Chaums Ideen aufgegriffen und ein *Anonymous Credential System* (ACS) entwickelt, das sowohl eine *globale*, als auch eine *lokale* Aufdeckung der Anonymität erlaubt. Bei einer globalen Aufdeckung wird die Identität eines betrügenden Nutzers aufgedeckt, so dass jedermann die Identität des Betrügers erfährt. Bei der lokalen Aufdeckung hingegen wird nur das Pseudonym des Nutzers offengelegt.

Das *Identity Mixer*-System von IBM, das im Rahmen des von der EU geförderten *PrimeLife*-Projekt entwickelt wurde, und das *U-Prove*-System von Microsoft sind weitere Vertreter von ACSs. Mittels Zero-Knowledge-Beweisen lassen sich einzelne Attribute eines Nutzers gegenüber einem Dritten beweisen. So lässt sich etwa nachweisen, dass ein Nutzer ein bestimmtes Alter erreicht hat, ohne das genaue Geburtsdatum preiszugeben.³ Unverkettbarkeit der verwendeten Pseudonyme gegenüber Dritten ist auch bei diesen Ansätzen gegeben. Das von der EU geförderte Projekt *ABC4Trust*, dessen Ziel die Entwicklung von *Attribute-Based Credentials* (ABC) ist, ist sowohl mit dem Identity Mixer-System als auch mit dem U-Prove-System kompatibel.

Anonyme Berechtigungsnachweise lassen sich auch mittels Gruppensignatur umsetzen. Der wesentliche Unterschied zwischen Gruppensignatur und anonymem Berechtigungsnachweis ist, dass die Gruppensignatur das Signieren eines Dokuments erlaubt (womit Integrität und Authentizität des Dokuments garantiert wird – Authentizität wird dabei in dem Sinn verstanden, dass garantiert wird, dass ein Mitglied der Gruppe das Dokument

³Man spricht in diesem Zusammenhang auch von einem „Selective Disclosure“.

signiert hat). Verbindlichkeit kann bei der Gruppensignatur im Prinzip auch garantiert werden, allerdings lässt sich die Signatur nicht auf einen Einzelnen zurückführen. Ein anonymer Berechtigungsnachweis ist lediglich der Beweis, dass der Nutzer eine Information besitzt, die die Berechtigung für eine bestimmte Handlung nachweist.

10.5 Fazit

In diesem Kapitel haben wir weitergehende *Privacy-Enhancing Technologies* (*PETs*) kennengelernt, die zur Zeit zu einem großen Teil noch Gegenstand der Forschung sind, vereinzelt aber bereits Einzug in Anwendungen gefunden haben. Wir dürfen damit rechnen, dass einige der Protokolle, sobald sie ihre Praxis-Tauglichkeit bewiesen haben, in Zukunft zum „Stand der Technik“ gehören werden. Die Europäische Agentur für Netz- und Informationssicherheit (*ENISA*) hat in ihrer Publikation „Privacy and Data Protection by Design—from policy to engineering“ aus dem Jahr 2014 bereits einige der in diesem Kapitel angesprochenen Protokolle als mögliche Maßnahmen im Kontext *Privacy by Design* (*PbD*) genannt.

10.6 Übungsaufgaben

Aufgabe 1

Ein Unternehmen möchte die ausgehenden E-Mails seiner Abteilungen signieren. Die Kunden sollen den Signaturen nicht ansehen können, welcher Sachbearbeiter sie erstellt hat; außerdem sollen sie nicht sehen können, ob zwei Signaturen durch denselben Sachbearbeiter erstellt wurden. Aus Sicherheitsgründen sollen aber nicht alle Mitarbeiter Zugriff auf den gleichen kryptographischen Schlüssel erhalten. Welche kryptographischen Bausteine eignen sich für diese Anwendung?

Literatur

1. Pierre Baldi, Roberta Baronio, Emiliano De Cristofaro, Paolo Gasti, and Gene Tsudik. Counterring gattaca: Efficient and secure testing of fully-sequenced human genomes. In *Proceedings of the 18th ACM Conference on Computer and Communications Security, CCS '11*, pages 691–702, New York, NY, USA, 2011. ACM.
2. Peter Bogetoft, Dan Lund Christensen, Ivan Damgård, Martin Geisler, Thomas Jakobsen, Mikkel Krøigaard, Janus Dam Nielsen, Jesper Buus Nielsen, Kurt Nielsen, Jakob Pagter, Michael Schwartzbach, and Tomas Toft. *Financial Cryptography and Data Security: 13th International Conference, FC 2009, Accra Beach, Barbados, February 23–26, 2009. Revised Selected Papers*, chapter Secure Multiparty Computation Goes Live, pages 325–343. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

3. Ernie Brickell, Jan Camenisch, and Liqun Chen. Direct anonymous attestation. In *Proceedings of the 11th ACM Conference on Computer and Communications security (CCS '04)*, pages 132–145. ACM, 2004.
4. Jan Camenisch and Anna Lysyanskaya. *An Efficient System for Non-transferable Anonymous Credentials with Optional Anonymity Revocation*, pages 93–118. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.
5. David Chaum. Security without identification: transaction systems to make big brother obsolete. *Communications of the ACM*, 28(10):1030–1044, Oct. 1985.
6. Emiliano De Cristofaro and Gene Tsudik. Practical private set intersection protocols with linear complexity. In *Financial Cryptography and Data Security, 14th International Conference, FC 2010, Tenerife, Canary Islands, January 25–28, 2010, Revised Selected Papers*, pages 143–159, 2010.
7. Sky Faber, Ronald Petrlic, and Gene Tsudik. Unlinked: Private proximity-based off-line OSN interaction. In *Proceedings of the 14th ACM Workshop on Privacy in the Electronic Society, WPES '15*, pages 121–131, New York, NY, USA, 2015. ACM.
8. Amos Fiat and Adi Shamir. *Advances in Cryptology — CRYPTO' 86: Proceedings*, chapter How To Prove Yourself: Practical Solutions to Identification and Signature Problems, pages 186–194. Springer Berlin Heidelberg, Berlin, Heidelberg, 1987.
9. Dominik Leibenger, Frederik Möllers, Anna Petrlic, Ronald Petrlic, and Christoph Sorge. Privacy Challenges in the Quantified Self Movement - An EU Perspective. *Proceedings on Privacy Enhancing Technologies*, 2016(4), 2016. Conference Presentation at PETS 2016.
10. Ronald Rivest, Adi Shamir, and Yael Tauman. How to leak a secret. In *Advances in Cryptology - ASIACRYPT 2001*, volume 2248 of *LNCS*, pages 552–565, 2001.

Zusammenfassung

In diesem Kapitel beschäftigen wir uns zu guter Letzt mit dem Datenschutzrecht – mit dem Fokus auf Deutschland und der Perspektive auf ein zukünftiges, gesamteuropäisches, gemeinsames Datenschutzrecht. Doch zunächst geben wir in Abschn. 11.1 einen Überblick über die historische Entwicklung des Datenschutzrechts. In Abschn. 11.2 werden wir uns damit beschäftigen, was der Datenschutz mit dem Grundgesetz zu tun hat. Danach werden wir in Abschn. 11.3 das Bundesdatenschutzgesetz (BDSG) als zentrales Datenschutzgesetz in Deutschland kennen lernen. In Abschn. 11.4 werden wir uns mit dem bereichsspezifischen Datenschutz im Telemediengesetz (TMG) und Telekommunikationsgesetz (TKG) auseinandersetzen. In Abschn. 11.5 geben wir einen kurzen Ausblick auf die kommende Datenschutz-Grundverordnung. Schließlich beschäftigen wir uns in Abschn. 11.6 mit der Lösung von datenschutzrechtlichen Einzelfragen und in Abschn. 11.7 mit einer datenschutzrechtlichen Betrachtung von Tracking im Web.

Lernziele

Am Ende dieses Kapitels sollten Sie einen Überblick über das Datenschutzrecht (in Deutschland) haben. Sie sollten Gesetze im Bereich des Datenschutzes verstehen und einfache Sachverhalte darunter subsumieren können – und erkennen können, wann für datenschutzrechtliche Fragestellungen rechtlicher Beistand nötig ist. Dieses Kapitel soll Ihnen die Grundlage dafür bieten, technische und juristische Sachverhalte in Beziehung setzen zu können. Außerdem sollten Sie die datenschutzrechtlichen Probleme beim Einsatz von Google Analytics und Social Plugins kennen und wissen, wie ein datenschutzkonformer Einsatz möglich ist.

11.1 Geschichte des Datenschutzes

Die Idee eines umfassenden, systematisierten Datenschutzrechts ist vergleichsweise neu. Das heißt aber nicht, dass vor dem 20. Jahrhundert keinerlei Privatsphärenschutz existierte. So enthält bspw. der *Hippokratische Eid* (von ungefähr 400 v.Chr.) Regelungen zur ärztlichen Schweigepflicht. Das Beichtgeheimnis für Geistliche ist ebenfalls eine sehr alte Norm.

Die Bedeutung des Schutzes der Privatsphäre wurde unter Juristen im 19. Jahrhundert verstärkt diskutiert. 1890 erschien in den USA ein Beitrag von SAMUEL WARREN und LOUIS BRANDEIS unter dem Titel „The Right to Privacy“ [4]. „Privacy“ wird dort als „the right to be let alone“ definiert. Die beiden Autoren beschreiben eine bereits damals stattfindende Entwicklung hin zu einem Recht auf Privatsphäre, bedingt u. a. durch technische Entwicklungen (z. B. Erfindung der Fotografie), aber auch z. B. aufgrund von Verletzungen der Privatsphäre durch Zeitungsreporter.

Das Konzept des Privatsphärenschutzes gewann im 20. Jahrhundert weiter an Bedeutung. Im Jahr 1948 nahm die Generalversammlung der Vereinten Nationen die „Allgemeine Erklärung der Menschenrechte“ (auch „UN-Menschenrechtscharta“ genannt) an. In Artikel 12 der Erklärung heißt es:

„Niemand darf willkürlichen Eingriffen in sein Privatleben, seine Familie, seine Wohnung und seinen Schriftverkehr oder Beeinträchtigungen seiner Ehre und seines Rufes ausgesetzt werden. Jeder hat Anspruch auf rechtlichen Schutz gegen solche Eingriffe oder Beeinträchtigungen.“

Eine vergleichbare Norm findet sich auch in Artikel 8 der Europäischen Menschenrechtskonvention, die im Rahmen des Europarats ausgearbeitet und 1950 durch dessen Mitgliedsstaaten unterzeichnet wurde. Sie trat 1953 in Kraft. Alle – derzeit 47 – Mitgliedsstaaten des Europarats sind auch Unterzeichner der Europäischen Menschenrechtskonvention.

Auch, wenn die bislang genannten Entwicklungen einen Bezug zum Datenschutz haben: Gefährdungen der Privat- und ggf. der Intimsphäre lagen damals nicht in massenhafter Datenverarbeitung, sondern beispielsweise in unberechtigten Wohnungsdurchsuchungen, der unerwünschten Einsichtnahme in ein Tagebuch oder dem Abhören eines Telefonanschlusses. Das änderte sich in den 1960er-Jahren. Durch die fortschreitende Entwicklung der Computertechnik entstand auch ein wachsendes Bewusstsein für die Gefährdung der Privatsphäre durch die Datenverarbeitung. Insbesondere in den USA führte der Plan zu massiven Protesten, ein „National Data Center“ einzuführen, in dem Daten verschiedener Behörden zusammengeführt worden wären und somit eine umfassende Datenbank über amerikanische Bürger entstanden wäre. ALAN WESTIN hat in diesem Kontext im Jahr 1967 „Privacy“ als „the desire of people to choose freely under what circumstances and to what extent they will expose themselves, their attitude and their behavior to others“ definiert. [5] Westins Definition wird im amerikanischen Raum

oft zitiert; seine Arbeit leistete einen wesentlichen Beitrag in der Debatte, die zum U.S. Privacy Act von 1974 führte.

Auch in Deutschland wuchs die Sensibilität für den Datenschutz und mündete in gesetzgeberischen Aktivitäten. Im Jahr 1970 trat das Hessische Datenschutzgesetz – als weltweit erstes – in Kraft. Sieben Jahre später folgte das Bundesdatenschutzgesetz (BDSG). Dass Datenschutz auch verfassungsrechtlich verankert ist, wurde 1983 deutlich: Das *Volkszählungsurteil* des Bundesverfassungsgerichts aus diesem Jahr markiert einen wichtigen Meilenstein für den Datenschutz in Deutschland. 1990 wurde das BDSG – unter Berücksichtigung des Volkszählungsurteils – neu gefasst. In Abschn. 11.2 werden wir uns damit noch intensiver befassen.

In Zeiten der Globalisierung und internationaler Warenflüsse kann ein wirksamer Datenschutz aber nicht mehr nur auf rein nationaler Ebene erreicht werden. Die *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data* (*OECD-Richtlinien über Datenschutz und grenzüberschreitende Ströme personenbezogener Daten*), deren Grundsätze zahlreiche Datenschutzgesetze beeinflussten, wurden 1980 verabschiedet. Sie sind allerdings allgemein gehalten und haben reinen Empfehlungscharakter. Um innerhalb der EU den Austausch personenbezogener Daten zu ermöglichen, aber dabei ein ausreichendes Schutzniveau zu gewährleisten, wurde eine stärkere Verbindlichkeit von Datenschutzregeln gebraucht. Deshalb wurde 1995 die EG-Datenschutzrichtlinie 95/46/EG verabschiedet. Sie führte zur Anpassung bzw. Schaffung nationaler Datenschutzgesetze, denn die Mitgliedsstaaten sind zur Umsetzung europäischer Richtlinien verpflichtet.

Die *Charta der Grundrechte* der EU wurde 1999/2000 erarbeitet, aber aufgrund der Wirrungen um den Europäischen Verfassungsvertrag erst 2009 – durch Verweis aus dem Lissaboner Vertrag – für fast alle EU-Staaten rechtlich bindend. Das Besondere an dieser Grundrechte-Charta ist, dass sie in Artikel 8 ein *Datenschutz-Grundrecht* enthält.

Trotz der Grundrechtecharta und der Richtlinie von 1995 wurde allerdings offenbar, dass keineswegs ein auch nur annähernd gleiches Datenschutzniveau in den Mitgliedsstaaten der EU erreicht worden war. Die Europäische Kommission legte daher 2012 den Entwurf einer neuen Datenschutzverordnung vor. Im Gegensatz zur Datenschutzrichtlinie aus dem Jahr 1995, die von den einzelnen Mitgliedsstaaten in nationales Recht umgesetzt wurde, wird die *Datenschutz-Grundverordnung* unmittelbar in allen EU-Mitgliedsstaaten gelten. Durch die Vereinheitlichung des Datenschutzniveaus innerhalb der EU soll es nicht mehr möglich sein, dass einzelne Länder die nationalen Regelungen abschwächen und somit von den großen, datenverarbeitenden Konzernen als Standort bevorzugt werden. Allerdings sind diverse Öffnungsklauseln enthalten, die in einzelnen Bereichen Spielraum für nationale Regelungen lassen. Nach langen Verhandlungen ist die Datenschutz-Grundverordnung schließlich im Frühjahr 2016 in Kraft getreten. Sie wird aber erst ab Mai 2018 angewendet. Der Zeitraum kann von den Mitgliedsstaaten für die Anpassung nationaler Gesetze und von Unternehmen für die Überprüfung ihrer Datenverarbeitungsprozesse genutzt werden.

Die Erarbeitung der Datenschutz-Grundverordnung ist vor dem Hintergrund der Veröffentlichungen zu den Überwachungsprogrammen von EDWARD SNOWDEN zu sehen. Durch diese Offenlegung ist das Thema Datenschutz wieder stärker in den Fokus der öffentlichen Diskussion gerückt. Zur jüngeren Geschichte des Datenschutzes gehören andererseits aber auch zahlreiche – unterschiedlich erfolgreiche – Initiativen, die zur Gewährleistung effektiver Strafverfolgung und Gefahrenabwehr den Datenschutz einschränken wollten. Das wohl prominenteste Beispiel ist die *Vorratsdatenspeicherung*, die bereits eine wechselvolle Geschichte hinter sich hat. 2006 trat die europäische Richtlinie 2006/24/EG zur Vorratsdatenspeicherung in Kraft. Sie sah unter anderem vor, dass Telekommunikationsanbieter sowohl für Telefonanrufe als auch für E-Mails Verbindungsdaten (aber keine Inhalte) speichern müssen; die Speicherfrist konnte durch nationales Recht auf 6 bis 24 Monate festgelegt werden. Die deutsche Umsetzung der Regelungen zur Vorratsdatenspeicherung, die 2008 in Kraft getreten sind, wurde 2010 durch das Bundesverfassungsgericht aufgehoben. In der Begründung führte das Gericht unter anderem aus, dass die anlasslose Speicherung von Verbindungsdaten geeignet sei, „ein diffus bedrohliches Gefühl des Beobachtetseins hervorzurufen“. Der Europäische Gerichtshof hat schließlich 2014 die EG-Richtlinie zur Vorratsdatenspeicherung gekippt. Seit Herbst 2015 ist in Deutschland ein neues Gesetz zur Vorratsdatenspeicherung in Kraft; die darin enthaltenen Speicherpflichten sind spätestens ab 1. Juli 2017 zu erfüllen. Wir werden in Abschn. 11.4.3 noch einmal auf die Vorratsdatenspeicherung zu sprechen kommen.

Im Jahr 2015 hat der Europäische Gerichtshof das *Safe Harbor*-Abkommen, das die bisherige Grundlage für die Übermittlung personenbezogener Daten aus EU-Ländern in die USA darstellte, gekippt. Die Nachfolgeregelung *Privacy Shield* wurde 2016 ausverhandelt.

11.2 Datenschutz im Grundgesetz

Im gesamten Grundgesetz kommt das Wort „Datenschutz“ nicht vor. Dennoch ist der Datenschutz im Grundgesetz verankert. Um zu verstehen, wie man durch Auslegung des Grundgesetzes zu einem Datenschutz-Grundrecht kommt, müssen wir zunächst ein allgemeineres Konzept anschauen, nämlich das sogenannte *Allgemeine Persönlichkeitsrecht*.

In Artikel 1 Absatz 1 des Grundgesetz heißt es:

„Die Würde des Menschen ist unantastbar. Sie zu achten und zu schützen ist Verpflichtung aller staatlichen Gewalt.“

Artikel 2 Absatz 1 des Grundgesetz besagt:

„Jeder hat das Recht auf die freie Entfaltung seiner Persönlichkeit, soweit er nicht die Rechte anderer verletzt und nicht gegen die verfassungsmäßige Ordnung oder das Sittengesetz verstößt.“

Die Rechtsprechung hat aus diesen beiden Normen ein umfassendes Recht auf Achtung und Entfaltung der Persönlichkeit hergeleitet. Dazu gehört der Schutz der Intim- und Privatsphäre sowie der weiter gefassten sogenannten Individualsphäre. Letztere umfasst das öffentliche Leben einer Person und genießt einen schwächeren Schutz als das Privatleben, das nur einer beschränkten Personenzahl zugänglich ist. Wenn durch eine staatliche Maßnahme etwa das Tagebuch einer Person ausgewertet wird, ist dies ein Eingriff in das Allgemeine Persönlichkeitsrecht.

Was hat das Allgemeine Persönlichkeitsrecht nun mit Datenschutz zu tun? Für das Jahr 1983 war aufgrund eines entsprechenden Gesetzes eine Volkszählung in der damaligen Bundesrepublik Deutschland geplant. Die Befragung sollte durch Beamte und ehrenamtliche „Zähler“ erfolgen. Ziel war die Korrektur der Melderegister und die Erhebung zahlreicher Daten. Das Bundesverfassungsgericht untersagte das Vorhaben zunächst mittels einer einstweiligen Anordnung. Im Dezember 1983 erklärte es in seinem richtungweisenden *Volkszählungsurteil* Teile des Gesetzes für nichtig, so dass die Volkszählung auch zu einem späteren Zeitpunkt nicht im geplanten Umfang stattfinden konnte. Die Volkszählung wurde daraufhin angepasst und 1987 durchgeführt.¹ Die Anpassung der Volkszählung beinhaltete insbesondere die Trennung erhobener statistischer Daten von identifizierenden Daten. Bedeutend für den Datenschutz war das Volkszählungsurteil deswegen, weil das Bundesverfassungsgericht darin das *Informationelle Selbstbestimmungsrecht* als Spezialfall des Allgemeinen Persönlichkeitsrechts definiert hat. Beim Informationellen Selbstbestimmungsrecht handelt es sich um die

„Befugnis des Einzelnen, grundsätzlich selbst über die Preisgabe und Verwendung seiner persönlichen Daten zu bestimmen.“

Das Bundesverfassungsgericht begründet seine Entscheidung folgendermaßen (hier nur ein kurzer Auszug):

„Wer nicht mit hinreichender Sicherheit überschauen kann, welche ihn betreffende Informationen in bestimmten Bereichen seiner sozialen Umwelt bekannt sind, und wer das Wissen möglicher Kommunikationspartner nicht einigermaßen abzuschätzen vermag, kann in seiner Freiheit wesentlich gehemmt werden, aus eigener Selbstbestimmung zu planen oder zu entscheiden. Mit dem Recht auf informationelle Selbstbestimmung wären eine Gesellschaftsordnung und eine diese ermöglichende Rechtsordnung nicht vereinbar, in der Bürger nicht mehr wissen können, wer was wann und bei welcher Gelegenheit über sie weiß. Wer unsicher ist, ob abweichende Verhaltensweisen jederzeit notiert und als Information dauerhaft gespeichert, verwendet oder weitergegeben werden, wird versuchen, nicht durch solche Verhaltensweisen aufzufallen. [...] Hieraus folgt: Freie Entfaltung der Persönlichkeit setzt unter den modernen Bedingungen der Datenverarbeitung den Schutz des Einzelnen gegen unbegrenzte Erhebung, Speicherung, Verwendung und Weitergabe seiner persönlichen Daten voraus. Dieser Schutz ist daher von dem Grundrecht des Art. 2 Abs. 1 in Verbindung mit Art. 1 Abs. 1 GG umfaßt.“

¹Zwischen 1987 (Westdeutschland) bzw. 1981 (Ostdeutschland) und 2011 gab es keine Volkszählung.

Im Kern stellt das Bundesdatenschutzgesetz, mit dem wir uns im nächsten Abschnitt genauer befassen werden, eine Umsetzung und Konkretisierung der Entscheidung dar.

Bei der Interpretation der Entscheidung könnte leicht ein Missverständnis entstehen: das Wort „grundsätzlich“ wird umgangssprachlich oft mit „ausnahmslos“ gleichgesetzt. Wenn Juristen allerdings den Begriff verwenden, meinen sie die wörtliche Bedeutung „im Grundsatz“ – es kann also auch Ausnahmen geben. Auch Grundrechte sind also nicht als absolute und schrankenlose Rechte zu verstehen. Meist ist dies auch gar nicht möglich. Die Freiheit eines Einzelnen bedeutet in der Regel auch Einschränkungen eines anderen. Die Schranken des informationellen Selbstbestimmungsrechts werden im Urteil des Bundesverfassungsgerichts explizit erwähnt. Einschränkungen sind etwa „im überwiegenden Allgemeininteresse“ möglich. Dass ein Staat personenbezogene Daten auch ohne Einwilligung der Betroffenen verarbeiten muss, wird beispielsweise im Bereich der Steuerverwaltung offensichtlich: Wer steuerpflichtig ist, kann nicht mit Verweis auf den Personenbezug der Daten die Anfertigung einer Steuererklärung ablehnen.

Die Voraussetzungen für Einschränkungen des Informationellen Selbstbestimmungsrechts sind laut dem Volkszählungsurteil:

- eine „gesetzliche Grundlage, aus der sich die Voraussetzungen und der Umfang der Beschränkungen klar und für den Bürger erkennbar ergeben“ (Normenklarheit),
- die Einhaltung des Grundsatzes der Verhältnismäßigkeit.

Verhältnismäßigkeit bedeutet, dass die Maßnahme zur Erreichung des (legitimen) Zwecks geeignet ist. Die Maßnahme muss ferner erforderlich zur Erreichung des Zwecks sein – es ist also keine „mildere“ Maßnahme möglich – und die Maßnahme muss angemessen sein, das Verhältnis der Vor- und Nachteile muss also stimmen.

11.3 Bundesdatenschutzgesetz

Wie bereits dargestellt, hat der Gesetzgeber einen gewissen Spielraum bei der konkreten Umsetzung des Informationellen Selbstbestimmungsrechts. Diese Ausgestaltung und Konkretisierung hat der Gesetzgeber im Bundesdatenschutzgesetz (BDSG), in speziellen Gesetzen für einzelne Bereiche (etwa für den Bereich der Telekommunikation im Telekommunikationsgesetz) sowie in den Datenschutzgesetzen der Länder vorgenommen. Das BDSG regelt dabei nicht nur die Datenverarbeitung durch den Staat, sondern explizit auch diejenige durch Wirtschaftsunternehmen.

Wir werden in diesem Abschnitt auch eine Reihe von datenschutzrechtlichen Fragen erörtern.

Beispiel

Sie möchten von einer Auskunft wissen, welche Daten dort über Sie und Ihre Kreditwürdigkeit gespeichert sind – haben Sie ein Recht darauf?

Beispiel

Sie betreiben einen öffentlichen Webserver für ein Unternehmen.

- Ist Datenschutz für Sie relevant?
- Sind Daten, mit denen Sie umgehen, für das Datenschutzrecht „schützenswert“?
- Welche Daten sind vom Datenschutzrecht eigentlich betroffen?

Viele Gesetze sind nach einem simplen Schema aufgebaut: Allgemeine Bestimmungen – beispielsweise Begriffsdefinitionen, die später verwendet werden – finden sich am Anfang; es folgen speziellere Regelungen sowie ganz am Ende des Gesetzes Übergangsvorschriften, Regelungen zum Inkrafttreten u. ä. Auch das BDSG folgt diesem Aufbau:

- Erster Abschnitt: Allgemeine und gemeinsame Bestimmungen,
- Zweiter Abschnitt: Datenverarbeitung der öffentlichen Stellen,
- Dritter Abschnitt: Datenverarbeitung nicht-öffentlicher Stellen und öffentlich-rechtlicher Wettbewerbsunternehmen,
- Weitere Abschnitte: Sonder-, Schluss-, Übergangsvorschriften.

11.3.1 Personenbezogene Daten

Daten sind – aus Sicht des Datenschutzrechts – genau dann schützenswert, wenn sie personenbezogen sind. Deshalb sehen wir uns zunächst die Definition von *Personenbezogenen Daten* in § 3 Abs. 1 BDSG genauer an:

„Personenbezogene Daten sind Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (Betroffener).“

Einzelangaben sind Daten mit einem Bezug zu einer einzelnen Person. Aggregierte oder anonymisierte Daten fallen also nicht in die Kategorie personenbezogener Daten. Allerdings gilt es zu beachten, dass auch Daten über eine Gruppe durch Zuordnung einer Person zu Einzelangaben werden können – selbst bei statistischen Aussagen. Ein Beispiel dafür ist: „Herr Müller gehört zur Kaufkraftgruppe B“.

Persönliche Verhältnisse sind Angaben über den Betroffenen, zum Beispiel der Name, die Adresse, Fingerabdrücke, Größe, Alter etc. *Sachliche Verhältnisse* sind Angaben über Sachverhalte, die mit dem Betroffenen in Verbindung stehen, etwa „X hat am 01.12.2016 um 12:30 mit Y telefoniert“. Dies sind sachliche Verhältnisse von X und Y.

Ein weiterer wesentlicher Aspekt der Definition ist die Bestimmtheit bzw. Bestimmbarkeit. Nach dem Gesetzeskommentar zum BDSG von GOLA und SCHOMERUS ist eine Person *bestimmt*, wenn sich der Bezug zur Person unmittelbar aus den Daten ergibt. Dies ist bspw. bei der Speicherung mit Name und Anschrift der Fall. *Bestimmbar* ist eine Person dann, wenn der Bezug zur Person ohne unverhältnismäßigen Aufwand herstellbar ist. Der

Aufwand hängt dabei von den Möglichkeiten der verarbeitenden Stelle ab (*relativer* oder *subjektiver* Personenbezugsbegriff). [1] Es gibt auch Stimmen in der Literatur, die davon ausgehen, Daten seien bereits dann personenbezogen, wenn *irgendjemand* die Person bestimmen kann (statt lediglich die verantwortliche Stelle mit Zugriff auf die Daten zu betrachten). Man spricht auch vom *absoluten* oder *objektiven* Personenbezugsbegriff.

Schließlich verlangt die Definition einen Bezug zu einer *natürlichen Person*. Daraus folgt, dass juristische Personen, z. B. Aktiengesellschaften, nicht vom BDSG geschützt werden. Allerdings gilt es auch hier wieder zu beachten, dass auch Angaben über juristische Personen natürliche Personen betreffen können, wie dies bei einer Einpersonen-GmbH (also einer GmbH mit nur einem Gesellschafter) der Fall sein kann.

An dieser Stelle sei die Definition der Anonymisierung aus § 3 Abs. 6 BDSG erwähnt:

Anonymisieren ist das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können.

Anonymisierte Daten sind also nicht mehr personenbezogen. Damit Daten aber als anonymisiert betrachtet werden können, reichen simple Maßnahmen wie das Entfernen von Namen und Adressen nicht unbedingt aus. Zur Beurteilung, ob Daten anonymisiert sind, müssen vielmehr auch Informatikmethoden herangezogen werden, wie sie im vorliegenden Buch dargestellt sind.

11.3.2 Zweck und Anwendungsbereich

Der Zweck des BDSG ergibt sich aus § 1 Abs. 1 und hat einen direkten Bezug zum Informationellen Selbstbestimmungsrecht, wie es im Volkszählungsurteil definiert ist:

„Zweck dieses Gesetzes ist es, den Einzelnen davor zu schützen, dass er durch den Umgang mit seinen personenbezogenen Daten in seinem Persönlichkeitsrecht beeinträchtigt wird.“

Der Anwendungsbereich ist in § 1 Abs. 2 gegeben:

„Dieses Gesetz gilt für die Erhebung, Verarbeitung und Nutzung personenbezogener Daten durch

1. öffentliche Stellen des Bundes,
2. öffentliche Stellen der Länder, soweit der Datenschutz nicht durch Landesgesetz geregelt ist und soweit sie
 - a) Bundesrecht ausführen oder
 - b) als Organe der Rechtspflege tätig werden und es sich nicht um Verwaltungsangelegenheiten handelt,

3. nicht-öffentliche Stellen, soweit sie die Daten unter Einsatz von Datenverarbeitungsanlagen verarbeiten, nutzen oder dafür erheben oder die Daten in oder aus nicht automatisierten Dateien verarbeiten, nutzen oder dafür erheben, es sei denn, die Erhebung, Verarbeitung oder Nutzung der Daten erfolgt ausschließlich für persönliche oder familiäre Tätigkeiten.“

Ferner gilt laut § 1 BDSG das *Subsidiaritätsprinzip*. Das BDSG ist also ein „Auffanggesetz“. Speziellere Vorschriften zum Datenschutz, wie wir sie noch in Abschn. 11.4.2 und Abschn. 11.4.3 kennen lernen werden, haben Vorrang. Außerdem wird in § 1 BDSG der *räumliche Anwendungsbereich* festgelegt. Das BDSG findet demnach keine Anwendung bei der Erhebung, Verarbeitung oder Nutzung personenbezogener Daten in Deutschland durch eine verantwortliche Stelle in einem anderen EU/EWR-Land. Eine Ausnahme besteht lediglich dann, wenn es eine Niederlassung in Deutschland gibt. Sofern die verantwortliche Stelle außerhalb der EU/des EWR-Raums ansässig ist, findet das BDSG Anwendung.

Beispiel: Fall 1

Das französische Unternehmen X schickt von Straßburg aus Mitarbeiter nach Deutschland, die dort Häuser fotografieren und die Bewohner nach dem Zustand der Häuser in ihrer Kreditwürdigkeit beurteilen. Richtet sich die Zulässigkeit des Vorgehens nach dem BDSG?

Beispiel: Fall 2

Das US-amerikanische Unternehmen Y schickt von New York aus Mitarbeiter nach Deutschland, die dort Häuser fotografieren und die Bewohner nach dem Zustand der Häuser in ihrer Kreditwürdigkeit beurteilen. Richtet sich die Zulässigkeit des Vorgehens nach dem BDSG?

Auf den ersten Blick mag es verwundern, dass im ersten Fall für das französische Unternehmen das BDSG nicht zur Anwendung kommt, für das US-amerikanische Unternehmen im zweiten Fall aber schon. Der Grund für diese Regelung ist das Ziel der Schaffung eines gemeinsamen europäischen Wirtschaftsraums. Ein französisches Unternehmen muss sich nicht mit den Datenschutzgesetzen der anderen europäischen Länder beschäftigen, es soll sich stattdessen an die (bekannten) datenschutzrechtlichen Gesetze im eigenen Land halten. Da sowohl das deutsche als auch das französische Datenschutzrecht auf der gleichen europäischen Richtlinie beruhen, wird angenommen, dass das letztendlich erreichte Datenschutzniveau in beiden Fällen vergleichbar ist. In der Praxis haben sich freilich deutliche Unterschiede in der Auslegung des Datenschutzrechts in verschiedenen Ländern der Europäischen Union ergeben.

11.3.3 Weitere Begriffsbestimmungen

In § 3a BDSG werden *Datenvermeidung* und *Datensparsamkeit* eingefordert:

„Die Erhebung, Verarbeitung und Nutzung personenbezogener Daten und die Auswahl und Gestaltung von Datenverarbeitungssystemen sind an dem Ziel auszurichten, so wenig personenbezogene Daten wie möglich zu erheben, zu verarbeiten oder zu nutzen. Insbesondere sind personenbezogene Daten zu anonymisieren oder zu pseudonymisieren, soweit dies nach dem Verwendungszweck möglich ist und keinen im Verhältnis zu dem angestrebten Schutzzweck unverhältnismäßigen Aufwand erfordert.“

Die Besonderheit dieser Regelung ist, dass der Gesetzgeber auf die Gestaltung der Technik Einfluss nehmen will (und nicht erst auf deren Verwendung). Dadurch, dass die Regelung (notwendigerweise) sehr unkonkret ist, spielt sie in der Praxis aber kaum eine Rolle.

Wir haben gesehen, dass im BDSG an mehreren Stellen von „Erheben“, „Verarbeiten“ und „Nutzung“ personenbezogener Daten die Rede ist. Auch diese Vorgänge sind in den weiteren Begriffsbestimmungen in § 3 genau definiert.

In § 3 Abs. 3 BDSG finden wir die Definition für das Erheben:

„Erheben ist das Beschaffen von Daten über den Betroffenen.“

Dabei spielt es keine Rolle, was später mit den Daten geschieht. Eine Erhebung personenbezogener Daten ist nur dann gegeben, wenn es sich um eine zielgerichtete Beschaffung handelt.

Beispiel

Ein Student schreibt seinem Dozenten eine E-Mail, wodurch dieser die E-Mail-Adresse des Studenten erhält. Hat der Dozent nun ein personenbezogenes Datum erhoben?

Der Dozent hat kein personenbezogenes Datum erhoben, denn er hat sich nicht zielgerichtet Daten beschafft.

In § 3 Abs. 4 BDSG finden wir die Definition für die Verarbeitung:

„Verarbeiten ist das Speichern, Verändern, Übermitteln, Sperren und Löschen personenbezogener Daten.“

Die einzelnen Begriffe sind ähnlich zu ihrer umgangssprachlichen Bedeutung definiert, doch empfiehlt sich trotzdem ein Blick in das BDSG, zu finden u. A. unter https://www.gesetze-im-internet.de/bdsg_1990/ (Zugegriffen am 01.10.2016). Der Begriff „Sperren“ mag auf den ersten Blick überraschen. Doch dem Gesetzgeber war bewusst, dass das Löschen personenbezogener Daten nicht immer möglich ist. Im Geschäftsleben könnten gesetzlich vorgegebene Aufbewahrungsfristen dem entgegenstehen, aber auch technische Gründe können das Löschen zumindest sehr aufwendig machen. Um in solchen Fällen

dennoch den Interessen der Betroffenen gerecht zu werden, ist das Konzept der Sperrung vorgesehen: Daten werden gekennzeichnet, um eine spätere Verarbeitung oder Nutzung auszuschließen.

Schließlich finden wir die Definition der Nutzung, als sogenannter „Auffangtatbestand“ in § 3 Abs. 5 BDSG:

„Nutzen ist jede Verwendung personenbezogener Daten, soweit es sich nicht um Verarbeitung handelt.“

11.3.4 Grundkonzept des deutschen Datenschutzrechts

Das deutsche Datenschutzrecht sieht in § 4 BDSG ein *Verbot mit Erlaubnisvorbehalt* vor. Die Erhebung, Verarbeitung oder Nutzung personenbezogener Daten ist grundsätzlich verboten, es sei denn es gibt

- eine Erlaubnis oder Anordnung im BDSG,
- eine Erlaubnis oder Anordnung in einer anderen Rechtsvorschrift,
- oder eine Einwilligung des Betroffenen.

Somit erscheint das Datenschutzrecht auf den ersten Blick übertrieben restriktiv. Es gibt aber zahlreiche Erlaubnisnormen, und insbesondere diejenigen im BDSG sind recht weit gefasst. Sie sind jeweils spezifisch für öffentliche (zweiter Abschnitt des BDSG) bzw. nichtöffentliche Stellen (dritter Abschnitt). Wir betrachten hier nur die Normen für *nichtöffentliche Stellen*. Die Erlaubnisnormen des BDSG sind überwiegend wiederum eingeschränkt; dort finden sich also zwar Erlaubnisse für die Datenverarbeitung, diese sind aber an Bedingungen geknüpft.

Die wichtigste Norm ist § 28 BDSG unter dem Titel „Datenerhebung und -speicherung für eigene Geschäftszwecke“. Das bedeutet, dass § 28 den Fall regelt, dass die Datenerhebung bzw. -speicherung nicht selbst der Geschäftszweck ist, sondern nur als Mittel verwendet wird, um einen anderen Geschäftszweck zu erfüllen. Wenn also ein Versandhändler Adressen speichert, um Waren zustellen zu können, ist dies ein Fall von § 28.

Die grundlegenden Bedingungen, um Daten für die Erfüllung eigener Geschäftszwecke erheben, speichern, verändern oder übermitteln zu dürfen, finden sich in Absatz 1. Auch ein wesentliches Prinzip des Datenschutzrechts – die *Zweckbindung* – findet sich an dieser Stelle:

„Bei der Erhebung personenbezogener Daten sind die Zwecke, für die die Daten verarbeitet oder genutzt werden sollen, konkret festzulegen.“ (§28 Abs. 1 Satz 2 BDSG).

Von dem festgelegten Zweck darf unter gewissen Umständen allerdings abgewichen werden; diese finden sich in Abs. 2. Die weiteren Absätze enthalten Regelungen zu

- der Verarbeitung und Nutzung personenbezogener Daten für Adresshandel, Werbung, Markt- und Meinungsforschung (Abs. 3 und 4),
- den Anforderungen an die datenschutzrechtliche Einwilligung (Abs. 3a und 3b),
- dem Umgang mit personenbezogenen Daten, die an einen Dritten übermittelt wurden, durch diesen Dritten (Abs. 5),
- dem Umgang mit „besonderen Arten personenbezogener Daten“ (definiert in § 3 Abs. 9 BDSG; beispielsweise handelt es sich um Daten zu Gesundheit oder Sexualität) in Abs. 6 bis 9.

Angeichts der wirtschaftlichen Bedeutung von Auskunftsteilen (wichtigstes Beispiel sind Unternehmen, die die Kreditwürdigkeit von Verbrauchern einschätzen und diese Daten ihren Kunden zur Verfügung stellen) gibt es mit § 28a eine Sonderregelung für die Übermittlung personenbezogener Daten an solche Unternehmen. Häufig stellen Auskunftsteile sogenannte Score-Werte zur Verfügung, die etwas über die Wahrscheinlichkeit aussagen sollen, dass ein Kunde eine Forderung bedienen wird. Die Voraussetzungen für die Verwendung solcher Score-Werte sind in § 28b geregelt.

§ 29 behandelt die Tätigkeit der Auskunftsteile selbst. Weitere spezielle Regelungen für einzelne Branchen finden sich in § 30 und § 30a. Schließlich sei noch § 32 BDSG erwähnt, der in sehr knapper Form Grundsätze für die Verarbeitung von Daten in einem Beschäftigungsverhältnis enthält.

Als nächstes wollen wir eine datenschutzrechtliche Fragestellung lösen. Dazu ziehen wir das BDSG heran.

Beispiel

Darf ein Versandhändler Ihre Anschrift an andere Unternehmen weitergeben, damit diese Ihnen Werbung zuschicken können? Falls ja, wie können Sie das verhindern?

Lösung: Schauen Sie in die Gliederung des BDSG: Zunächst könnte eine Regelung in den „allgemeinen und gemeinsamen Bestimmungen“ enthalten sein. Andererseits kann man vermuten, dass die Antwort, ob Daten für Werbezwecke weitergegeben werden dürfen, sich zwischen nicht-öffentlichen und öffentlichen Stellen unterscheidet. Wir schauen also in den Abschnitt „Datenverarbeitung nicht-öffentlicher Stellen und öffentlich-rechtlicher Wettbewerbsunternehmen“. Dieser Abschnitt hat nur 14 Paragraphen, deren Überschrift wir lesen. Wir werfen einen kurzen Blick auf § 27 und wissen nun sicher, im richtigen Abschnitt gelandet zu sein. Lediglich zwei Überschriften klingen passend: § 28 („Datenerhebung und -speicherung für eigene Geschäftszwecke“) und § 29 („Geschäftsmäßige Datenerhebung und -speicherung zum Zweck der Übermittlung“). § 29 hat eine andere Zielrichtung (nämlich, dass die Erhebung/Speicherung/Veränderung/Nutzung personenbezogener Daten nicht nur ein

Nebeneffekt wie im Versandhandel ist). Das ist aber keineswegs auf den ersten Blick klar – es lohnt sich daher unbedingt, beide Normen zu lesen.

Wir klären die gestellte Frage nun mit § 28.

- § 28 Abs. 1 bezieht sich auf die „Erfüllung eigener Geschäftszwecke“ und passt damit nicht.
- § 28 Abs. 2 passt auch nicht, da die Frage keinen Anhaltspunkt für eine der Möglichkeiten bietet.
- § 28 Abs. 3 Satz 1 passt nicht, da keine Einwilligung vorliegt.
- § 28 Abs. 3 Satz 2 („Listenprivileg“) kommt schließlich zum Tragen.

Der Versandhändler darf Ihre berufliche Anschrift an andere Unternehmen weitergeben, damit diese Ihnen Werbung im Hinblick auf Ihre berufliche Tätigkeit zuschicken können. (Dies ist nicht die einzige Verarbeitung, die in § 28 Abs. 3 Satz 2 erlaubt wird, aber die einzige, die genau zur Frage passt.)

In § 28 Abs. 4 Satz 1 BDSG finden wir schließlich noch die Lösung auf die Frage, wie die Verarbeitung bzw. Nutzung personenbezogener Daten für Werbezwecke verhindert werden kann: „Widerspricht der Betroffene bei der verantwortlichen Stelle der Verarbeitung oder Nutzung seiner Daten für Zwecke der Werbung oder der Markt- oder Meinungsforschung, ist eine Verarbeitung oder Nutzung für diese Zwecke unzulässig.“

11.3.5 Auskunftsanspruch

In § 34 Abs. 1 BDSG ist der *Auskunftsanspruch* normiert:

„Die verantwortliche Stelle hat dem Betroffenen auf Verlangen Auskunft zu erteilen über

1. die zu seiner Person gespeicherten Daten, auch soweit sie sich auf die Herkunft dieser Daten beziehen,
2. den Empfänger oder die Kategorien von Empfängern, an die Daten weitergegeben werden, und
3. den Zweck der Speicherung.“

§ 34 BDSG ist insgesamt noch wesentlich umfangreicher und enthält u. a. Regelungen für das Scoring oder den Grundsatz der Unentgeltlichkeit der Auskunft. Wie die meisten juristischen Grundsätze ist auch dieser nicht ohne Ausnahme: Sie finden sie – ebenso wie die Ausnahmen von der Ausnahme – in Abs. 8.

In den nächsten beiden Abschnitten werden wir uns mit den spezielleren Vorschriften zum Datenschutz im Bereich der Telemedien sowie der Telekommunikation beschäftigen.

11.4 Bereichsspezifischer Datenschutz

Neben dem Bundesdatenschutzgesetz (BDSG) ist der Datenschutz in Deutschland noch in weiteren Gesetzen verankert. In diesem Abschnitt beschäftigen wir uns mit den für uns in diesem Buch relevanten bereichsspezifischen Datenschutz-Normen bei *Telemedien* und bei der *Telekommunikation*. Wir hatten bereits erwähnt, dass diese Regelungen den Regelungen im BDSG vorgehen, d. h. dass das BDSG als Auffanggesetz dient. Die Vielfalt an Datenschutz-Regelungen erklärt sich dadurch, dass das BDSG zu allgemein ist und speziellen Problemen nicht gerecht wird.

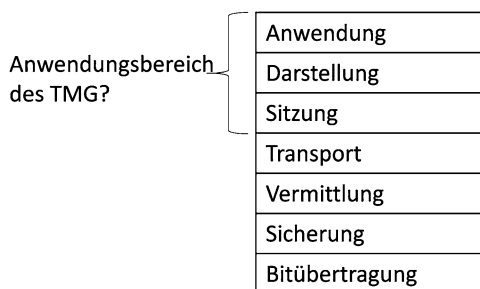
11.4.1 Anwendungsbereich

In der Praxis erweist es sich allerdings als schwierig, die Anwendungsbereiche von Bundesdatenschutzgesetz (BDSG), Telemediengesetz (TMG) und Telekommunikationsgesetz (TKG) sauber abzugrenzen. Die Grundidee ist, dass das TMG nur bei der Erbringung eines Informations-/Kommunikationsdienstes *über* ein Telekommunikationsnetz (z. B. das Internet) angewendet wird. Für das Telekommunikationsnetz selbst greift hingegen das TKG. *Inhaltsbezogenes* wird wiederum im BDSG geregelt. Bei einem Online-Shop greift demnach, sobald es um den Inhalt, also den reinen Kaufvorgang geht, das BDSG und nicht mehr das TMG. Das liegt daran, dass der Vorgang des Online-Kaufs nichts anderes ist als bspw. Einkaufen per Katalog und Telefon – wo das TMG auch nicht greift.

In der juristischen Literatur wird zum Teil die Abgrenzung zwischen TMG und TKG gemäß dem ISO/OSI-Schichtenmodell vertreten. Ab Schicht 5, wie in Abb. 11.1 dargestellt, gilt demnach das TMG.

So allgemein ist diese Abgrenzung allerdings zu sehr vereinfacht. Ein reiner E-Mail-Dienst ist nach herrschender Literaturmeinung reine Telekommunikation, obwohl er oberhalb von Schicht 4 angesiedelt ist. Wir werden bei der Betrachtung von TMG und TKG noch einmal genauer auf den jeweiligen Anwendungsbereich eingehen.

Abb. 11.1 Anwendungsbereich des TMG



11.4.2 Telemediengesetz

Das *Telemediengesetz (TMG)*, das auch als „Gesetz für das Web“ bezeichnet werden könnte, setzt die Europäische Richtlinie über den elektronischen Geschäftsverkehr (RL 2000/31/EG, auch „E-Commerce-Richtlinie“ genannt) um. Das TMG enthält nicht nur Bestimmungen zum Datenschutz, mit denen wir uns hier vordergründig beschäftigen werden, sondern es finden sich auch z. B. Haftungsbestimmungen im TMG. In diesem Zusammenhang dürfte Ihnen vielleicht die „Störerhaftung“ beim Betrieb von WLANs bekannt sein. Sie ist allerdings nur im Zusammenspiel zwischen dem Bürgerlichen Gesetzbuch und den Regelungen des TMG zu verstehen; sollte jemand auf Grundlage der Störerhaftung einen Anspruch gegen Sie geltend machen, ist von einer „Selbstdiagnose“ daher abzuraten.

Anwendungsbereich

Der Anwendungsbereich des Telemediengesetzes ist in § 1 Abs. 1 TMG geregelt. Telemedien sind demnach alle elektronischen Informations- und Kommunikationsdienste außer

- Telekommunikationsdienste, die ganz in der Übertragung von Signalen über Telekommunikationsnetze bestehen,
- telekommunikationsgestützte Dienste,
- Rundfunk.

Telekommunikationsdienste, die ganz in der Übertragung von Signalen über Telekommunikationsnetze bestehen, sind „alle Dienste, die ohne die Aufbereitung und Ansehung von Inhalten Daten übertragen und die Leistung sich somit auf die reine Transportfunktion beschränkt.“² Auch VoIP und E-Mail gehören dazu – aber nicht Webmailer und auch nicht Web-Interfaces zur Konfiguration von VoIP-Diensten.

Telekommunikationsgestützte Dienste (§3 Nr. 25 TKG) sind „Dienste, die keinen räumlich und zeitlich trennbaren Leistungsfluss auslösen, sondern bei denen die Inhaltsleistung noch während der Telekommunikationsverbindung erfüllt wird.“ Gemeint sind sogenannte Mehrwertdienste, die typischerweise über 0900-Nummern angeboten werden (z. B. zum Informationsabruf). Aus dem Wortlaut der Definition wird das nicht klar, hier muss die Gesetzesbegründung herangezogen werden. Bei wortlautgetreuer Auslegung bleibt nämlich kein Bereich mehr übrig, in dem das TMG angewendet werden kann. SCHMITZ (in Spindler/Schuster, Kommentar zu § 1 TMG, Rn. 24) weist zu Recht darauf hin, dass die Ausnahme von „Mehrwertdiensten“ aus dem Anwendungsbereich des TMG unsystematisch ist. Der einzige Unterschied zum Informationsabruf von einer Webseite sind die Verwendung von Leitungsvermittlung und das andere Adressierungsschema –

²Schmitz in Spindler/Schuster, Recht der elektronischen Medien, 1. Auflage 2008, Kommentar zu § 1 TMG, Rn. 16.

beides wird mit zunehmender Verbreitung von VoIP aber zunehmend verwässert (so auch Schmitz mit Bezug auf die Leitungsvermittlung).

Rundfunk ist nach § 2 Abs. 1 des Rundfunk-Staatsvertrags „die für die Allgemeinheit bestimmte Veranstaltung und Verbreitung von Darbietungen aller Art in Wort, in Ton und in Bild unter Benutzung elektromagnetischer Schwingungen ohne Verbindungsleitung oder längs oder mittels eines Leiters.“ Der Begriff der Darbietung wird so ausgelegt, dass er (nur) Inhalte umfasst, die einen Beitrag zur Meinungsbildung leisten können und sollen.³ Dies gilt übrigens auch bei der Übermittlung über das Internet.

Zu Beginn haben wir gesagt, dass das TMG als „Gesetz für das Web“ bezeichnet werden könnte. Juristen denken beim TMG tatsächlich überwiegend an Webseiten. Der Definition des Telemediengesetzes nach fallen bspw. auch Peer-to-Peer-Systeme zum Download von Kinofilmen unter den Telemedienbegriff [3], was allerdings in der juristischen Literatur kaum thematisiert wird; es finden sich auch kaum passende Regelungen, da dem TMG implizit das Client/Server-Modell zugrunde liegt. Auch die Frage, ob es sich bei einzelnen Angebotsseiten in einem Online-Auktionsangebot oder Profildaten in Social Media um eigenständige Telemedien handelt, drängt sich bei Lektüre des TMG auf; erst in jüngerer Vergangenheit wurde die Frage in der Rechtsprechung thematisiert.

Datenschutz im TMG

Die Datenschutzbestimmungen des TMG sind in den §§ 11–15 zu finden. Zunächst schränkt § 11 TMG den Anwendungsbereich ein. Die Bestimmungen des TMG finden keine Anwendung für im Dienst- und Arbeitsverhältnis zu ausschließlich beruflichen oder dienstlichen Zwecken bereitgestellten Telemedien. Des weiteren finden sie keine Anwendung bei Telemedien, die ausschließlich zur Steuerung von Arbeits- oder Geschäftsprozessen eingesetzt werden.

Das TMG enthält gegenüber dem BDSG weitergehende Einschränkungen bezüglich des Datenschutzes. In § 12 Abs. 1 TMG ist etwa ein strengeres Verbot mit Erlaubnisvorbehalt normiert. Eine Erhebung und Verwendung personenbezogener Daten durch den Diensteanbieter ist nur mit Einwilligung, Erlaubnis im TMG oder Erlaubnis in einer Rechtsvorschrift *mit ausdrücklichem Bezug auf Telemedien* erlaubt. Dabei geht es sowohl um die Bereitstellung der Telemedien selbst, als auch um die spätere Verwendung personenbezogener Daten für andere Zwecke. Die allgemeinen Vorschriften des BDSG reichen hier also gerade nicht aus.

In § 13 TMG finden sich Regelungen zum Systemdatenschutz, also zum Datenschutz durch die Gestaltung informationsverarbeitender Systeme. Gegenüber dem allgemeinen Grundsatz der Datenvermeidung und Datensparsamkeit in § 3a BDSG, den wir bereits besprochen haben, gibt es in § 13 TMG eine konkretere Ausgestaltung für Telemedien. Die Pflichten des Diensteanbieters gemäß § 13 TMG lauten folgendermaßen:

³Siehe z. B. Holznagel in Spindler/Schuster, Recht der elektronischen Medien, Kommentar zu § 2 RStV, Rn. 44.

- „(1) Der Diensteanbieter hat den Nutzer zu Beginn des Nutzungsvorgangs über Art, Umfang und Zwecke der Erhebung und Verwendung personenbezogener Daten sowie über die Verarbeitung seiner Daten in Staaten außerhalb des Anwendungsbereichs der Richtlinie 95/46/EG des Europäischen Parlaments und des Rates vom 24. Oktober 1995 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr (ABl. EG Nr. L 281 S. 31) in allgemein verständlicher Form zu unterrichten, sofern eine solche Unterrichtung nicht bereits erfolgt ist. Bei einem automatisierten Verfahren, das eine spätere Identifizierung des Nutzers ermöglicht und eine Erhebung oder Verwendung personenbezogener Daten vorbereitet, ist der Nutzer zu Beginn dieses Verfahrens zu unterrichten. Der Inhalt der Unterrichtung muss für den Nutzer jederzeit abrufbar sein.
- (2) Die Einwilligung kann elektronisch erklärt werden, wenn der Diensteanbieter sicherstellt, dass
1. der Nutzer seine Einwilligung bewusst und eindeutig erteilt hat,
 2. die Einwilligung protokolliert wird,
 3. der Nutzer den Inhalt der Einwilligung jederzeit abrufen kann und
 4. der Nutzer die Einwilligung jederzeit mit Wirkung für die Zukunft widerrufen kann.
- (3) Der Diensteanbieter hat den Nutzer vor Erklärung der Einwilligung auf das Recht nach Absatz 2 Nr. 4 hinzuweisen. Absatz 1 Satz 3 gilt entsprechend.
- (4) Der Diensteanbieter hat durch technische und organisatorische Vorkehrungen sicherzustellen, dass
1. der Nutzer die Nutzung des Dienstes jederzeit beenden kann,
 2. die anfallenden personenbezogenen Daten über den Ablauf des Zugriffs oder der sonstigen Nutzung unmittelbar nach deren Beendigung gelöscht oder in den Fällen des Satzes 2 gesperrt werden,
 3. der Nutzer Telemedien gegen Kenntnisnahme Dritter geschützt in Anspruch nehmen kann,
 4. die personenbezogenen Daten über die Nutzung verschiedener Telemedien durch denselben Nutzer getrennt verwendet werden können,
 5. Daten nach § 15 Abs. 2 nur für Abrechnungszwecke zusammengeführt werden können und
 6. Nutzungsprofile nach § 15 Abs. 3 nicht mit Angaben zur Identifikation des Trägers des Pseudonyms zusammengeführt werden können. An die Stelle der Löschung nach Satz 1 Nr. 2 tritt eine Sperrung, soweit einer Löschung gesetzliche, satzungsmäßige oder vertragliche Aufbewahrungsfristen entgegenstehen.
- (5) Die Weitervermittlung zu einem anderen Diensteanbieter ist dem Nutzer anzuzeigen.
- (6) Der Diensteanbieter hat die Nutzung von Telemedien und ihre Bezahlung anonym oder unter Pseudonym zu ermöglichen, soweit dies technisch möglich und zumutbar ist. Der Nutzer ist über diese Möglichkeit zu informieren.
- (7) Diensteanbieter haben, soweit dies technisch möglich und wirtschaftlich zumutbar ist, im Rahmen ihrer jeweiligen Verantwortlichkeit für geschäftsmäßig angebotene Telemedien durch technische und organisatorische Vorkehrungen sicherzustellen, dass
1. kein unerlaubter Zugriff auf die für ihre Telemedienangebote genutzten technischen Einrichtungen möglich ist und
 2. diese
 - a) gegen Verletzungen des Schutzes personenbezogener Daten und
 - b) gegen Störungen, auch soweit sie durch äußere Angriffe bedingt sind, gesichert sind. Vorkehrungen nach Satz 1 müssen den Stand der Technik berücksichtigen.

Eine Maßnahme nach Satz 1 ist insbesondere die Anwendung eines als sicher anerkannten Verschlüsselungsverfahrens.

- (8) Der Diensteanbieter hat dem Nutzer nach Maßgabe von § 34 des Bundesdatenschutzgesetzes auf Verlangen Auskunft über die zu seiner Person oder zu seinem Pseudonym gespeicherten Daten zu erteilen. Die Auskunft kann auf Verlangen des Nutzers auch elektronisch erteilt werden.“

In der Praxis erfolgt die Unterrichtung nach § 13 Abs. 1 TMG bei Web-Angeboten so, dass eine eigene Seite mit einer *Datenschutzerklärung* angelegt wird und diese von allen „Unterseiten“ des Angebots verlinkt wird. Damit wird auch die jederzeitige Abrufbarkeit sichergestellt. Mit einem „automatisierte[n] Verfahren, das eine spätere Identifizierung des Nutzers ermöglicht und eine Erhebung oder Verwendung personenbezogener Daten vorbereitet“ sind bspw. „Cookies“ gemeint, mit denen wir uns in Abschn. 7.1.1 bereits beschäftigt haben.

- **Trailer** Zusammengefasst ergeben sich als Anforderungen an die Systemgestaltung und die Umsetzung des Datenschutzes durch das TMG also:

- Unterrichtungspflicht
- Anforderungen an Einwilligungen
- Erforderliche technische Datenschutzmaßnahmen
- Transparenzanforderung bei Weitervermittlung
- Anonymitäts-/Pseudonymitätserfordernis
- Anforderungen an IT-Sicherheit
- Modifizierter Auskunftsanspruch

Bestands- und Nutzungsdaten

Im TMG gibt es eine Unterscheidung personenbezogener Daten nach Bestands- und Nutzungsdaten.

Bestandsdaten sind personenbezogene Daten, die für die „Begründung, inhaltliche Ausgestaltung oder Änderung eines Vertragsverhältnisses zwischen dem Diensteanbieter und dem Nutzer über die Nutzung von Telemedien“ (§ 14 Abs. 1 TMG) erforderlich sind. Hierbei handelt es sich typischerweise um statische, auf Nutzer bezogene Daten wie Name, Anschrift oder Kontonummer. Diese Daten dürfen nach § 14 Abs. 1 TMG erhoben und verwendet werden. Ebenso ist in § 14 Abs. 2 TMG eine Auskunft an Strafverfolgungsbehörden und Geheimdienste vorgesehen. Schließlich ist eine Auskunft auch „zum Zweck der Durchsetzung der Rechte am geistigen Eigentum“ vorgesehen.

Nutzungsdaten sind Daten, die erforderlich sind, „um die Inanspruchnahme von Telemedien zu ermöglichen und abzurechnen“ (§ 15 Abs. 1 TMG). Die Erhebung und Verwendung von Nutzungsdaten im erforderlichen Umfang ist nach § 15 Abs. 1 TMG erlaubt. Beispiele für Nutzungsdaten aus dem Gesetz (§ 15 Abs. 1) sind etwa Merkmale zur Identifikation des Nutzers, Angaben über Beginn und Ende sowie Umfangs der jeweiligen Nutzung und Angaben über die vom Nutzer in Anspruch genommenen Telemedien. Ein

typisches Beispiel für Nutzungsdaten sind Daten aus dem Logfile eines Webserver. Es mag gängige Praxis sein, ein solches Logfile pauschal über einige Wochen hinweg aufzubewahren. Dies wird aber so durch § 15 TMG im Allgemeinen nicht erlaubt, sofern die Inhalte des Logfiles personenbezogen sind (inwieweit dies bei vollständig gespeicherten IP-Adressen zutrifft, werden wir noch erörtern; deutlicher wird der Personenbezug noch, wenn Nutzerkennungen mitgespeichert werden).⁴ Allerdings ist die Frage, was noch zur Datenverarbeitung gehört, die erforderlich ist, um die „Inanspruchnahme von Telemedien zu ermöglichen“, einer Auslegung zugänglich. So könnten zur verbesserten Angriffserkennung IP-Adressen durchaus für einen kurzen Zeitraum gespeichert werden, sofern dies für notwendig gehalten wird. Der Europäische Gerichtshof hat am 19.10.2016 der vorher verbreiteten strikteren Auffassung (wonach nur die in § 15 TMG explizit genannten Konstellationen eine Speicherung von personenbezogenen Daten über das Sitzungsende hinaus rechtfertigen) eine Absage erteilt: Diese sei nicht mit der Datenschutzrichtlinie 95/46/EG vereinbar. Das Urteil bedeutet aber gerade nicht, dass das Speichern von Logfiles immer erlaubt wäre. Es kann lediglich nicht durch das TMG pauschal verboten werden.

§ 15 TMG enthält noch eine Reihe besonderer Erlaubnisse und weiterer Regelungen für die Verwendung von Nutzungsdaten. So erlaubt § 15 Abs. 3 TMG die pseudonyme Erstellung von Nutzungsprofilen für Werbung, Marktforschung und die „bedarfsgerechte Gestaltung der Telemedien“. Die Profile dürfen nicht mit Daten über den Träger des Pseudonyms zusammengeführt werden. Der Diensteanbieter muss den Nutzer auf sein Widerspruchsrecht gegen die Profilbildung hinweisen. § 15 Abs. 2, 4, 6, 7 TMG normieren die Abrechnung; § 15 Abs. 5 TMG regelt die Übermittlung aus Gründen der Abrechnung, sowie eine anonymisierte Übermittlung. Das Vorgehen bei Missbrauchsverdacht ist in § 15 Abs. 8 TMG geregelt.

In § 15a TMG findet sich die Pflicht für den Diensteanbieter, Nutzer bei „Daten-Lecks“ zu informieren. Die Voraussetzungen hierfür sind, dass Bestands- oder Nutzungsdaten unrechtmäßig übermittelt wurden oder Dritte anderweitig unrechtmäßig Kenntnis erhalten haben und schwerwiegende Beeinträchtigungen für die Rechte oder schutzwürdigen Interessen des betroffenen Nutzers drohen. § 15a TMG verweist als Folge des Eintretens einer Datenpanne auf § 42a BDSG. Hiernach müssen die Aufsichtsbehörde sowie die Betroffenen informiert werden, und zwar unverzüglich, wenn Sicherungsmaßnahmen abgeschlossen sind und die Strafverfolgung nicht mehr gefährdet ist. Gegebenenfalls ist eine Veröffentlichung, bspw. in einer Tageszeitung, nötig, falls eine individuelle Benachrichtigung einen unverhältnismäßigen Aufwand darstellt.

- **Trailer** Zusammenfassend stellt das TMG das spezielle (Datenschutz-)Recht „für das Web“ dar. Der konkrete Anwendungsbereich ist kompliziert geregelt. Die Abgrenzung gegen andere gesetzliche Regelungen (z.B. TKG)

⁴Zur Anonymisierung von IP-Adressen in Logfiles siehe z. B. LEIBENGER et al. [2].

ist unklar. Es gibt eine de-facto-Beschränkung auf Web-Angebote, andere Dienste werden in der juristischen Fachdiskussion kaum wahrgenommen. Insgesamt gehören die Datenschutzregelungen des TMG aufgrund des Systemdatenschutzes und dem verschärften Verbot mit Erlaubnisvorbehalt (trotz des erwähnten EuGH-Urteils) zu den weltweit restriktivsten. Allerdings gibt es aufgrund der komplizierten Regelungen, der unklaren Abgrenzung und einiger „handwerklichen“ Fehler ein Vollzugsdefizit.

11.4.3 Telekommunikationsgesetz

Das *Telekommunikationsgesetz* (*TKG*) enthält Regelungen zum Fernmeldegeheimnis und spezifische Datenschutzregelungen. Der Datenschutz im engeren Sinne ist in Teil 7, Abschnitt 2 (§§ 91–107) normiert. Die Regelungen sind grundsätzlich ähnlich zum TMG. Wir betrachten hier nur einen Ausschnitt.

Anwendungsbereich der Datenschutzregelungen

Nach § 91 I TKG sind personenbezogene Daten der Teilnehmer und Nutzer von Telekommunikation geschützt. Ein *Teilnehmer* ist dabei jede natürliche oder juristische⁵ Person, die mit einem Anbieter von Telekommunikationsdiensten einen Vertrag über die Erbringung derartiger Dienste geschlossen hat (§ 3 Satz 1 Nr. 20 TKG). Ein *Nutzer* ist jede natürliche Person, die einen Telekommunikationsdienst für private oder geschäftliche Zwecke nutzt, ohne notwendigerweise Teilnehmer zu sein (§ 3 Satz 1 Nr. 14 TKG). Der Schutz besteht „bei der Erhebung und Verwendung dieser Daten durch Unternehmen und Personen, die geschäftsmäßig Telekommunikationsdienste erbringen oder an deren Erbringung mitwirken“ (§ 91 Abs. 1 Satz 1 TKG).

Bestandsdaten

Die Regelung zu *Bestandsdaten* (§ 95 TKG) ist ähnlich dem TMG, aber teilweise etwas konkreter. Erforderliche Daten „für die Begründung, inhaltliche Ausgestaltung, Änderung oder Beendigung eines Vertragsverhältnisses über Telekommunikationsdienste“ dürfen erhoben und verwendet werden. Der Datenaustausch mit anderen Anbietern ist zulässig (Beispiel: Call-by-Call-Verträge). Die Löschung hat zum Ende des Kalenderjahres zu erfolgen, das auf die Vertragsbeendigung folgt. Außerdem ist eine Überprüfung der Bestandsdaten anhand eines amtlichen Ausweises erlaubt. In § 95 Abs. 5 TKG ist zudem das *Kopplungsverbot* normiert:

„Die Erbringung von Telekommunikationsdiensten darf nicht von einer Einwilligung des Teilnehmers in eine Verwendung seiner Daten für andere Zwecke abhängig gemacht werden, wenn dem Teilnehmer ein anderer Zugang zu diesen Telekommunikationsdiensten ohne die

⁵Man beachte, dass hier, anders als im BDSG, auch juristische Personen erfasst sind.

Einwilligung nicht oder in nicht zumutbarer Weise möglich ist. Eine unter solchen Umständen erteilte Einwilligung ist unwirksam.“

Verkehrsdaten

Verkehrsdaten entsprechen ungefähr den Nutzungsdaten aus dem TMG. Es handelt sich um „Daten, die bei der Erbringung eines Telekommunikationsdienstes erhoben, verarbeitet oder genutzt werden“ (§ 3 Nr. 30 TKG). Diese Daten dürfen im Wesentlichen für die Erbringung und Abrechnung der Dienstleistung verwendet werden (§ 96 Abs. 1 und 2 TKG). Die Löschung hat nach Verbindungsende zu erfolgen, soweit die Daten nicht mehr erforderlich sind (§ 96 Abs. 1 TKG). Die Verwendung für Marketing-Zwecke erfordert eine Einwilligung *und* Anonymisierung (§ 96 Abs. 3 TKG). Einwilligung *und* Anonymisierung zu fordern, ist dem Datenschutzrecht eigentlich systemfremd – denn mit der Anonymisierung (falls korrekt durchgeführt) liegen keine personenbezogenen Daten mehr vor. Für Verkehrsdaten, wie sie in § 96 geregelt sind, können auch keine allgemeinen datenschutzrechtlichen Erlaubnistatbestände (z. B. aus dem BDSG) herangezogen werden, denn das speziellere TKG verdrängt die allgemeineren Regelungen.

Vorratsdatenspeicherung

Die „alte“ Vorratsdatenspeicherung hat das Bundesverfassungsgericht 2010 für nichtig erklärt. Davon betroffen waren § 113a, 113b TKG (vollständig) und § 100g StPO (teilweise). Die Kernpunkte aus der Entscheidung des Bundesverfassungsgerichts⁶ waren:

- Die Vorratsdatenspeicherung verletzt das Telekommunikationsgeheimnis (Artikel 10 Grundgesetz (GG)),
- alle bisher gesammelten Daten müssen gelöscht werden,
- die Vorratsdatenspeicherung ist grundsätzlich zulässig,
- die konkrete Ausgestaltung ist unzureichend betreffend
 - Datensicherheit,
 - Anforderungen an unmittelbare Datenverwendung (Abruf nur bei belegter, konkreter Gefahr für „Leib, Leben oder Freiheit einer Person, für den Bestand oder die Sicherheit des Bundes oder eines Landes oder zur Abwehr einer gemeinen Gefahr“),
 - Transparenz und Rechtsschutz.
- die Anforderungen an mittelbare Verwendung der Daten (Identifikation einer Person zu bekannter IP-Adresse) liegen niedriger.

Außerdem hat der Europäische Gerichtshof 2014 die EG-Richtlinie zur Vorratsdatenspeicherung gekippt. Die wesentlichen Kernpunkte aus der Entscheidung:

⁶Eine Zusammenfassung der Entscheidung findet sich unter <http://www.bundesverfassungsgericht.de/pressemitteilungen/bvg10-011.html> (Zugegriffen am 01.08.2016)

- Die Vorratsdatenspeicherung verstößt gegen Grundrechte auf Achtung des Privatlebens und auf Schutz personenbezogener Daten (Artikel 7 und 8 der EU-Grundrechtecharta).
- Die Datenspeicherung ist ein besonders schwerwiegender Eingriff in Grundrechte.
- Aus Daten können genaue Schlüsse auf Gewohnheiten des täglichen Lebens gezogen werden.
- Es gibt keine Information der Bürger über Speicherung und Nutzung der Daten; damit haben die Bürger ein Gefühl, „dass ihr Privatleben Gegenstand einer ständigen Überwachung ist“.
- Die Richtlinie ist zu weit gefasst, sie überschreitet die Grenzen der Verhältnismäßigkeit: Die Speicherung bezieht sich auf alle Personen, auch wenn kein Anhaltspunkt für eine Straftat besteht.

Der Europäische Gerichtshof sieht bei der Richtlinie schon ein Problem bei der Speicherung der Daten (Schritt 1), da die Daten aller Personen erfasst werden. Das Bundesverfassungsgericht hatte vor allem die Umsetzung der Nutzung der Daten (Schritt 2) kritisiert.

Die „neue“ Vorratsdatenspeicherung in Deutschland ist in § 113b TKG geregelt. Sie wurde im Herbst 2015 beschlossen und die neuen Speicherpflichten sind ab dem 01.07.2017 zu erfüllen. Für Anbieter öffentlicher Telefondienste besteht eine Datenspeicherungspflicht für jedes Telefonat, die umfasst:

- Zeitpunkt (inkl. zugrundeliegende Zeitzone), Rufnummern von Anrufer und Angerufenem,
- zusätzlich für Mobilfunkanbieter: Funkzelle und IMEI von Anrufer und Angerufenem,
- zusätzlich für VoIP-Anbieter: IP-Adressen von Anrufer und Angerufenem,
- Anrufversuche (nicht nur erfolgreicher Verbindungsaufbau),
- bei SMS wird auch der Inhalt gespeichert.

Die nun vorgesehene Speicherdauer beträgt für Telefondienste und Internetdienste 10 Wochen. Die Speicherdauer in der „alten“ Vorratsdatenspeicherung lag bei 6 Monaten. Die zugrundeliegende EG-Richtlinie erlaubte Zeiträume zwischen 6 Monaten und 2 Jahren.

Die Datenspeicherungspflicht für Internet Service Provider (ISP) sieht die Speicherung der Anschlusskennung, IP-Adresse und Zeiträume der Internetnutzung vor.

Bei der mobilen Nutzung werden zudem Standortdaten erfasst; das betrifft bei Telefondiensten die Standorte von Anrufer bzw. Angerufenem zu Beginn des Anrufs, bei der Internetnutzung den Standort des Internetnutzers zu Beginn der Internetverbindung. Es müssen die Bezeichnungen der Funkzellen gespeichert sowie die „geografische Lage und die Hauptstrahlrichtungen“ der zugehörigen Funkantennen vorgehalten werden. Es ist im Gesetz und auch der zugehörigen Gesetzesbegründung nicht definiert, was der „Beginn einer Internetverbindung“ ist. Da das Internet per se verbindungslos funktioniert, gibt es auch keinen eindeutigen technischen Anknüpfungspunkt. Das „Radio Resource Control Protocol“, das in UMTS und LTE verwendet wird, sieht verschiedene Aktivitätszustände ja nach Länge der Inaktivität einer Datenverbindung vor; welcher davon als Ende einer

„Internetverbindung“ angesehen wird und wann man folglich vom Beginn einer neuen ausgeht, ist unklar. Da auf aktuellen Smartphones oft mehrere Apps installiert sind, die regelmäßig Daten abrufen, ist es durchaus plausibel, dass aufgrund der Vorratsdatenspeicherung vollständige Bewegungsprofile von deren Nutzern entstehen. Im Gegensatz zu den anderen gespeicherten Daten beträgt die Speicherfrist für Standortdaten aber nur vier Wochen.

Der Zugriff auf die Daten ist in § 113c TKG geregelt. Unter Berufung auf eine gesetzliche Bestimmung⁷ darf eine Strafverfolgungsbehörde⁸ zur Verfolgung von besonders schweren Straftaten auf die Daten zugreifen. Ebenfalls unter Berufung auf eine gesetzliche Bestimmung darf eine Gefahrenabwehrbehörde der Länder⁹ auf die Daten zugreifen, sofern es um die Abwehr einer konkreten Gefahr für Leib, Leben oder Freiheit einer Person oder für den Bestand des Bundes oder eines Landes geht.

Auch die Daten von Geheimnisträgern, etwa Ärzten, Anwälten, Geistlichen, Journalisten etc. werden gespeichert. Diese Daten dürfen von den Behörden allerdings nicht abgerufen werden.

- **Trailer** Insgesamt sind die Regelungen des TKG von hoher praktischer Relevanz. Neben den Datenschutz-Regelungen gibt es weitere, bspw. zum Verbraucherschutz und zur Entgeltregulierung. Die Diskussionsthemen in den letzten Jahren waren in diesem Zusammenhang insbesondere der Schutz vor Werbeanrufen und die Vorratsdatenspeicherung. Das grundlegende Konzept der Datenschutz-Regelungen ist ähnlich dem TMG.

11.5 Datenschutz-Grundverordnung

Mit Geltung der Datenschutz-Grundverordnung ab Mai 2018 ergeben sich zahlreiche Änderungen im Datenschutzrecht. Die Grundprinzipien der Richtlinie 95/46/EG bleiben erhalten – und damit auch die Prinzipien, die im deutschen Datenschutzrecht umgesetzt sind. Im Detail gibt es jedoch viele Neuerungen. Wie diese sich auswirken, lässt sich nicht immer mit letzter Gewissheit sagen: Die Verordnung gibt zum Teil explizite Regelungsaufträge an die nationalen Gesetzgeber, zum Teil werden explizit Regelungsspielräume eröffnet (die genutzt werden können, aber nicht müssen). Zum Zeitpunkt der Drucklegung des vorliegenden Buchs ist allerdings noch kein neues deutsches Datenschutzgesetz verabschiedet, so dass noch unsicher ist, wie die Regelungsspielräume in Deutschland genutzt werden.

Hinzu kommt, dass die Verordnung in vielerlei Hinsicht auslegungsbedürftig ist. Natürlich gilt das im Prinzip für alle Gesetze, doch hat der Verordnungsgeber in besonderem

⁷In dem Gesetz, das der Behörde zugrunde liegt, muss also eine Erlaubnis für die Verwendung der Daten vorliegen.

⁸Dazu zählen insbesondere die Staatsanwaltschaften.

⁹Dazu zählen insbesondere die Polizeien.

Maße mit unbestimmten Rechtsbegriffen gearbeitet. Das ist kein Zufall, sondern wohl dem europäischen Gesetzgebungsprozess geschuldet: Die Notwendigkeit, Kompromisse zu finden, war – gerade bei der Datenschutzgrundverordnung – noch stärker ausgeprägt, als dies typischerweise auf nationaler Ebene der Fall ist. Formulierungen, die keiner der Auffassungen der Verhandlungspartner widersprechen, bieten sich als Kompromiss an und verschieben die eigentliche Entscheidung der Sachfrage in die Rechtsprechung. Da diese Rechtsprechung momentan noch nicht vorliegt, besteht also eine gewisse Unsicherheit bezüglich der zukünftigen Rechtslage.

Ohne damit Anspruch auf Vollständigkeit zu erheben, lassen sich aber jedenfalls folgende wesentliche Neuerungen in der Verordnung feststellen:

Übersicht

- Erweiterung des Anwendungsbereichs: Die Verordnung gilt für verantwortliche Stellen innerhalb der EU – unabhängig vom Ort der Datenverarbeitung. Sie gilt auch für die Verarbeitung personenbezogener Daten sogar durch verantwortliche Stellen außerhalb der EU, sofern es im Zusammenhang mit einem Angebot von Waren oder Dienstleistungen in der EU geschieht oder Verhalten von Personen in der EU beobachtet wird.
- Einführung eines Anspruchs auf Datenübertragbarkeit: Personenbezogene Daten müssen dem Betroffenen unter gewissen Voraussetzungen in einem gängigen, strukturierten und maschinenlesbaren Format zur Verfügung gestellt werden. Das soll die Übertragung an andere Dienstleister erleichtern.
- Einführung einer Rechenschaftspflicht: Verantwortliche Stellen müssen technische und organisatorische Maßnahmen ergreifen, um sicherstellen und *nachweisen* zu können, dass ihre Datenverarbeitung der Verordnung entspricht.
- Gesetzliche Verankerung des „Rechts auf Vergessenwerden“: Der Anspruch, eigene personenbezogene Daten bei der verantwortlichen Stelle löschen zu lassen, wird neu definiert.
- Datenschutz-Folgenabschätzung: Ist durch eine Datenverarbeitung ein besonders hohes Risiko für die Rechte natürlicher Personen zu erwarten, muss *vorab* eine Folgenabschätzung durchgeführt werden. Dabei müssen die Verarbeitungsvorgänge und ihre Zwecke systematisch aufgelistet, die Verhältnismäßigkeit der Verarbeitung überprüft sowie die Risiken bewertet und Gegenmaßnahmen beschrieben werden.
- Einführung spürbarer Sanktionen bei Datenschutzverstößen: Die Geldbußen für einige Verstöße können nun bis zu 20 Millionen Euro oder vier Prozent des weltweiten Jahresumsatzes eines Unternehmens betragen. Es gilt der höhere der beiden Beträge.

11.6 Datenschutzrechtliche Einzelfragen

In diesem Abschnitt beschäftigen wir uns mit einigen datenschutzrechtlichen Einzelproblemen im Detail. Wir lernen dabei, wie wir Schritt für Schritt vorgehen müssen, um rechtliche Fragen zum Datenschutz zu beantworten.

11.6.1 Personenbezug bei IP-Adressen?

Eine der zentralen Fragen im Datenschutzrecht – zumindest, soweit es sich mit kommunikationsbezogenen Sachverhalten befasst – lautet: „Sind IP-Adressen personenbezogene Daten?“.

Zur Beantwortung der Frage sehen wir uns die Definition personenbezogener Daten im Bundesdatenschutzgesetz (BDSG) an. § 3 Abs. 1 BDSG lautet: „Personenbezogene Daten sind Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (Betroffener)“. Unter Einzelangaben versteht man dabei Daten mit Bezug zu einer einzelnen Person (im Gegensatz zu Personengruppen; vgl. dazu Gola/Schomerus, Randnummer 3 zu § 3 BDSG). IP-Adressen weisen oft (näheres dazu weiter unten) einen Bezug zu einer einzelnen Person auf, daher kann von einer Einzelangabe gesprochen werden. Um eine Angabe über „sachliche Verhältnisse“ handelt es sich bei Daten, die Informationen über einen auf den Betroffenen beziehbaren Sachverhalt enthalten (Gola/Schomerus, Rn. 5 zu § 3 BDSG). Eine einzeln gespeicherte IP-Adresse enthält zwar keine solchen Informationen, doch steht die Speicherung in der Praxis fast immer in einem bestimmten Kontext; in einem Webserver-Log sagt sie zum Beispiel aus, dass der Inhaber einer IP-Adresse auf bestimmte Inhalte zugegriffen hat. Zumindest in solchen Fällen kann also der Bezug zu den sachlichen Verhältnissen einer Person bejaht werden.

Es muss sich bei der Person, auf die sich der Sachverhalt bezieht, auch um eine natürliche Person handeln. Auch dies ist bei IP-Adressen nicht immer der Fall, aber doch in sehr vielen Fällen (z. B. wenn eine Person „im Web surft“).

Knackpunkt der Diskussion ist allerdings meistens die Frage, ob die entsprechende Person auch „bestimmt oder bestimmbar“ ist. Hier kann man zunächst danach unterscheiden, *wer* mit der IP-Adresse umgeht, denn nur dann lässt sich die Ausgangsfrage sinnvoll beantworten. So kann der Internet Service Provider (ISP) die IP-Adresse stets mit einem Kunden in Verbindung bringen (ein Streitpunkt besteht hier allenfalls noch, ob dies für den Personenbezug ausreicht, wenn neben dem Vertragspartner des ISP weitere Personen deren Zugang mitnutzen; dies ist aber oft nicht der Fall). Wer nur einen Webserver betreibt, kann die IP-Adresse *alleine* evtl. mit Zusatzinformationen (Login der Person auf geschützter Seite, Suche nach eigenem Namen in einer Suchmaschine) einer Person zuordnen. Schwierig ist dies bei dynamisch zugewiesenen IP-Adressen, einfacher bei statischen (ein einmaliges Zuordnen reicht hier). Cookies können eine erneute Zuordnung

ggf. deutlich vereinfachen. In Zusammenarbeit mit dem ISP wird die Zuordnung in aller Regel unproblematisch möglich sein.

An dieser Stelle entzündet sich die Diskussion, ob von einem relativen oder absoluten Personenbezugsbegriff ausgegangen werden soll. Gola/Schomerus (Rn. 10 zu §3 BDSG) schreiben dazu: „Für die Bestimmbarkeit kommt es auf die Kenntnisse, Mittel und Möglichkeiten der speichernden Stelle an. Sie muss den Bezug mit den ihr normalerweise zur Verfügung stehenden Hilfsmitteln und ohne unverhältnismäßigen Aufwand durchführen können.“ – es wird also von einem relativen Personenbezugsbegriff ausgegangen. Es gibt auch Argumente für die Gegenmeinung (so könnte die Stelle, die die IP-Adresse speichert, auch illegal mit der Stelle zusammenarbeiten, die den Personenbezug herstellen kann; da das Datenschutzrecht gerade vor Missbrauch schützen soll, kann man auch argumentieren, zumindest das Vorhandensein rechtlicher Hürden könne bei der Frage nach dem Personenbezug keine Rolle spielen).

Etwas mehr Klarheit zu dieser Fragestellung brachte das bereits erwähnte Urteil des Europäischen Gerichtshofs (EuGH) vom 19.10.2016. Der Bundesgerichtshof (BGH) hatte dem EuGH 2014 die Frage nach dem Personenbezug dynamischer IP-Adressen vorgelegt (Aktenzeichen: C-582/14). Das Gericht hat den Begriff des Personenbezugs recht weit ausgelegt: Wenn der Betreiber eines Webserver über rechtliche Mittel verfügt, den Personenbezug mit Hilfe Dritter herzustellen, soll dies ausreichen, um einen Personenbezug herzustellen. Dafür genügt es bereits, wenn im Fall eines Angriffs auf den Webserver die Zuordnung der IP-Adresse zu einer Person mit Hilfe von Behörden und dem ISP hergestellt werden kann. Für den Fall, dass diese rechtliche Möglichkeit besteht – wovon der EuGH offenbar ausgeht –, sind auch dynamische IP-Adressen personenbezogene Daten. Die Entscheidung ist noch mit dem relativen Personenbezugsbegriff vereinbar, da nicht *jegliches* Wissen Dritter berücksichtigt werden muss, das die Zuordnung zu einer Person ermöglichen könnte.

Es besteht nach diesem Urteil nun kein Grund mehr, zwischen statischen und dynamischen IP-Adressen differenzieren zu wollen. In der Praxis würde sich das ohnehin schwierig gestalten, da es keinen zuverlässigen Weg gibt, automatisch zwischen beiden Kategorien zu unterscheiden. Da der EuGH – wie oben bereits erwähnt – gleichzeitig entschieden hat, dass eine Abwägungsmöglichkeit bei der Verarbeitung personenbezogener Daten auch im Fall des § 15 TMG bestehen muss, ergeben sich aus dem Urteil sicherlich keine unlösbaren Schwierigkeiten für Webserver-Betreiber.

11.6.2 Einwilligung

Beschäftigen wir uns als nächstes mit der Frage, ob Ihnen ein Online-Shop die Aufgabe einer Bestellung verweigern darf, weil Sie nicht in die Verwendung personenbezogener Daten (z. B. für Werbezwecke) einwilligen. Wir wollen auch klären, ob eine Einwilligung im Rahmen von Allgemeinen Geschäftsbedingungen möglich ist.

Die allgemeinen Anforderungen an eine Einwilligung werden durch § 4a BDSG festgelegt. Zur Frage nach der Einwilligung in AGB findet sich dort folgende Regelung: „Soll die Einwilligung zusammen mit anderen Erklärungen schriftlich erteilt werden, ist sie besonders hervorzuheben.“ Dies *kann* auch im Rahmen von AGB geschehen, wobei dann die besondere Hervorhebung der Einwilligung zu beachten ist. Ein „Verstecken im Kleingedruckten“ ist keinesfalls zulässig und viele in AGB erteilte Einwilligungen dürften auch tatsächlich unwirksam sein.

Zur Frage, ob ein Online-Shop die Aufgabe einer Bestellung wegen nicht gegebener Einwilligung verweigern kann, gibt es eine speziellere Regelung. Da es hier um eine nicht-öffentliche Stelle geht und die Datenerhebung zunächst für eigene Geschäftszwecke des Online-Shops stattfindet, ist § 28 Abs. 3b BDSG einschlägig:

„Die verantwortliche Stelle darf den Abschluss eines Vertrags nicht von einer Einwilligung des Betroffenen nach Absatz 3 Satz 1 abhängig machen, wenn dem Betroffenen ein anderer Zugang zu gleichwertigen vertraglichen Leistungen ohne die Einwilligung nicht oder nicht in zumutbarer Weise möglich ist. Eine unter solchen Umständen erteilte Einwilligung ist unwirksam.“

Dies betrifft insbesondere Anbieter mit einem Monopol auf ihrem Markt. Handelt es sich aber um einen „ganz normalen“ Online-Shop, kann man wohl nicht davon ausgehen, dem Betroffenen sei „ein anderer Zugang zu gleichwertigen vertraglichen Leistungen ohne die Einwilligung nicht oder nicht in zumutbarer Weise möglich“. Daher kann in der Regel ein Online-Shop in der Tat die Bestellannahme von einer Einwilligung abhängig machen.

Wer die Norm übersehen hat, konnte auch anhand von § 4a BDSG zum gleichen Ergebnis kommen, der auch bereits die Freiwilligkeit der Einwilligung regelt. Absatz 1, Satz 1 lautet: „Die Einwilligung ist nur wirksam, wenn sie auf der freien Entscheidung des Betroffenen beruht.“ Bei einem Nicht-Monopolisten als Anbieter kann man davon ausgehen, dass der Betroffene in seiner Entscheidung frei ist, diesen oder einen anderen Online-Shop zu wählen. Anders sieht es aus, wenn der Betroffene keine Auswahl hat – ohne Kenntnis von § 28 Abs. 3b könnte man sogar bei einem Monopol noch in vielen Fällen davon ausgehen, die Entscheidung sei frei: Oft hat der Betroffene auch die Wahl, die Ware oder Dienstleistung gar nicht in Anspruch zu nehmen.

Keine Einwilligung ist erforderlich, soweit die Daten für den Bestellvorgang selbst benötigt werden (§ 28 Abs. 1 Nr. 1 BDSG).

11.6.3 Anwendungsbereich des TMG

In Abschn. 11.4.2 haben wir bereits ausführlich über den Anwendungsbereich des TMG diskutiert. Hier klären wir nun, ob Online-Shops, Webcasts, VoIP-Telefonie, Chatrooms, bzw. die kommerzielle Verbreitung von Informationen über Waren-/Dienstleistungsangebote mit elektronischer Post (z. B. Werbe-Mails) unter das TMG fallen.

Die Beispiele stammen aus der Gesetzesbegründung (Bundestags-Drucksache 16/3078). Erläuterungen finden sich auch im Kommentar „Recht der elektronischen Medien“ von Spindler/Schuster, 2. Auflage 2011.

- Online-Shops sind ein klassisches Beispiel für Telemedien, denn es wird eine über die reine Telekommunikation hinausgehende elektronische Dienstleistung angeboten, die auch nicht unter den Rundfunkbegriff fällt.
- Webcasts: Hierunter wird die ausschließliche Übertragung herkömmlicher Rundfunkprogramme über das Internet verstanden. Da die Rundfunkdefinition auch bei Übertragungen über Internet zutrifft und nicht lediglich bei der Ausstrahlung von Radiowellen, sind Webcasts in diesem Sinne von der Definition der Telemedien ausgenommen (wichtig ist hier, dass die genannten Webcasts zum zeitgleichen Empfang bestimmt sind).
- VoIP-Telefonie besteht an sich nur aus der Übertragung von Signalen über Telekommunikationsnetze und ist damit reiner Telekommunikationsdienst. Es kann aber durchaus Telemedien geben, die mit VoIP-Übertragung im Zusammenhang stehen.
- Chatrooms dienen zwar der Telekommunikation zwischen ihren Teilnehmern, doch wird das Zusammenbringen der Nutzer auf einer Plattform als darüber hinausgehende Dienstleistung angesehen. Chatrooms werden daher als Telemedien eingeordnet.
- Die kommerzielle Verbreitung von Informationen über Waren-/Dienstleistungsangebote mit elektronischer Post wird in § 6 Abs. 2 TMG explizit erwähnt. Reine E-Mail-Übertragung (ohne Webmaildienst o.ä.) fällt zwar unter den Begriff der Telekommunikation – in der Formulierung aus der Aufgabenstellung wird aber explizit auf den Inhalt abgehoben, in dem eine Informationsdienstleistung (eben die Information über Waren- oder Dienstleistungsangebote) enthalten ist, die zudem elektronisch verbreitet wird. Die „kommerzielle Verbreitung von Informationen über Waren-/Dienstleistungsangebote mit elektronischer Post“ zählt daher zu den Telemedien.

11.6.4 Anwendung TMG bei E-Mail

E-Mail-Dienste werden oft als „Telemedien, die überwiegend in der Übertragung von Signalen über Telekommunikationsnetze bestehen“ eingeordnet. Ist das gerechtfertigt und was sind die Konsequenzen dieser Einordnung bezogen auf die Datenschutzregelungen des Telemediengesetzes? Warum hat der Gesetzgeber das wohl so geregelt?

Die Einordnung ist offensichtlich bei der Betrachtung von Webmail-Diensten (bzw. E-Mail-Diensten, die auch Webmail anbieten): Bei diesen wird überwiegend eine Transportdienstleistung erbracht (also ein Übertragen von Signalen über Telekommunikationsnetze), daneben aber auch eine Web-Oberfläche bereit gehalten (was den Anwendungsbereich des TMG eröffnet). Die Gesetzesbegründung (Bundestags-Drucksache 16/3078) sagt aber allgemeiner: „Während die Bereitstellung [...] eines E-Mail-Dienstes

eine besondere Dienstleistung darstellt [...]“. Eine nähere Begründung wird nicht gegeben, aber möglicherweise ist das Vorhalten der E-Mails auf einem IMAP- bzw. POP3-Server gemeint. Für einen normalen E-Mail-Dienst ist das allerdings zweifelhaft, da auch diese Zwischenspeicherung letztlich dem Transport zum Rechner des Empfängers dient.

Konsequenz der Einordnung ist, dass gemäß § 11 Abs. 3 TMG von den Datenschutzregelungen des TMG lediglich § 15 Abs. 8 und § 16 Abs. 2 Nr. 4 anwendbar sind. Diese Regelung vermeidet die gleichzeitige Anwendbarkeit der Datenschutz-Regelungen aus TKG und TMG bzw. die Notwendigkeit, für jeden Teilbereich des Dienstes genau zu trennen, ob TKG oder TMG zur Anwendung kommen. Letztlich wird damit die Rechtsanwendung vereinfacht (wenn erst einmal entschieden ist, wie der Dienst als Ganzes überhaupt einzuordnen ist).

11.6.5 Herausgabe personenbezogener Daten

Gibt § 14 Abs. 2 TMG dem Bundesnachrichtendienst das Recht, von einem Diensteanbieter die Herausgabe personenbezogener Daten seiner Nutzer zu verlangen?

Nein. § 14 Abs. 2 TMG richtet sich an den Diensteanbieter („Auf Anordnung der zuständigen Stellen darf der Diensteanbieter...“). Falls der Bundesnachrichtendienst eine solche Anordnung erteilt, handelt der Diensteanbieter also nicht rechtswidrig, wenn er die angeforderten Bestandsdaten herausgibt. Das TMG regelt aber nicht, unter welchen Umständen die Anordnung erteilt werden darf. Das Recht, die entsprechende Auskunft zu verlangen, ergibt sich vielmehr aus Spezialgesetzen. Die Gesetzesbegründung (Bundestags-Drucksache 16/3078, Seite 16) sagt dazu:

„Die Vorschrift besagt, dass Diensteanbieter aus der Aufgabenerfüllung im Bereich der Strafverfolgung sowie der genannten Behörden erwachsende Auskunftsansprüche nicht aus datenschutzrechtlichen Erwägungen zurückweisen können. Die Anordnung der zuständigen Stellen erfolgt nach Maßgabe der hierfür geltenden Bestimmungen (Strafprozessordnung, Bundes- und Landesverfassungsschutzgesetze, Bundesnachrichtendienstgesetz, Gesetz über den Militärischen Abschirmdienst).“

11.6.6 Auskunft über Datei-Downloads

Das Bundeskriminalamt verlangt von einem Diensteanbieter die Auskunft, welche seiner Nutzer zu welchem Zeitpunkt eine bestimmte Datei heruntergeladen haben. Fällt diese Anfrage unter die Regelung des § 14 Abs. 2 TMG?

Das Verlangen ist aus mehreren Gründen problematisch. Zum einen ist die Anfrage der Fragestellung nach auf Nutzungsdaten und nicht auf Bestandsdaten gerichtet, denn sie sind nicht „für die Begründung, inhaltliche Ausgestaltung oder Änderung eines Vertragsverhältnisses zwischen dem Diensteanbieter und dem Nutzer über die Nutzung von Telemedien erforderlich“. § 14 Abs. 2 TMG betrifft aber nur Bestandsdaten. In

der Rechtsprechung werden Auskünfte über Bestandsdaten, die den Rückgriff auf Nutzungsdaten erfordern, in der Regel aber ebenfalls auf Grundlage von § 14 Abs. 2 TMG gestützt (auch, wenn man diese Auffassung für zweifelhaft halten mag); dies würde dazu führen, dass man die Auskunft über die Identitäten der betreffenden Nutzer für statthaft halten mag. Der hier beschriebene Fall ist aber etwas anders gelagert, denn zumindest der Zeitpunkt des Zugriffs ist kein Bestandsdatum mehr. Daneben stellt sich noch die Frage, ob das Verlangen noch einen „Einzelfall“ betrifft, da die Frage eine Vielzahl von Nutzern betrifft. Da ein Einzelfall nicht zwingend bedeuten muss, dass nur ein einzelner Nutzer betroffen ist, sondern wohl eher routinemäßige Abfragen verhindert werden sollen, kann man im beschriebenen Fall vermutlich davon ausgehen, dass noch ein Einzelfall vorliegt.

11.6.7 Ausweitung der Protokollierung

Sie bieten auf Ihrer Website kostenpflichtige Dienste an und haben den Verdacht, dass einzelne Nutzer Zugangssicherungen überwunden haben, um mehr Inhalte anschauen zu können, als sie bezahlt haben. Dürfen Sie nun einfach Ihre Logfiles, die die Namen der Nutzer enthalten, so lange sichern, bis Sie dem Verdacht erfolgreich nachgegangen sind?

Die Antwort findet sich in § 15 TMG. Absatz 7 regelt, dass „Abrechnungsdaten, die für die Erstellung von Einzelnachweisen über die Inanspruchnahme bestimmter Angebote auf Verlangen des Nutzers verarbeitet werden“, bis zu sechs Monate nach Rechnungsversand gespeichert werden dürfen. Wollen Sie über diesen Zeitraum hinaus speichern oder sind die Daten nicht für die Erstellung von Einzelnachweisen gedacht (wovon Sie in dieser Aufgabe ausgehen durften), kommt lediglich noch Absatz 8 in Frage. Ein bloßer Verdacht reicht hier nicht aus; falls wirklich „tatsächliche Anhaltspunkte vor[liegen], dass seine Dienste von bestimmten Nutzern in der Absicht in Anspruch genommen werden, das Entgelt nicht oder nicht vollständig zu entrichten“, sind diese zu dokumentieren. Nur soweit dies für Zwecke der Rechtsverfolgung erforderlich ist, dürfen die Daten *dieser* Nutzer auch länger gespeichert werden. Die Speicherung kompletter Logfiles, die ja auch andere Nutzer betreffen, ist damit aber nicht abgedeckt.

11.6.8 Rufnummernunterdrückung

Haben Teilnehmer eines Telefonnetzes das Recht, ihre Rufnummer bei ausgehenden Anrufen zu unterdrücken? Welche Ausnahmen gibt es von diesem Grundsatz?

Ja, Teilnehmer haben das Recht, ihre Rufnummer bei ausgehenden Anrufen fallweise oder dauerhaft zu unterdrücken (§ 102 Abs. 1 Satz 1 TKG). Ausgenommen sind Werbeanrufe (§ 102 Abs. 2 TKG) und Dienste für Teilnehmer geschlossener Benutzergruppen (§ 102 Abs. 3 TKG). Bei Anrufen zu bestimmten Notrufnummern (§ 102 Abs. 8 TKG) muss der Diensteanbieter sicherstellen, dass die Rufnummernübermittlung nicht unterdrückt wird.

11.7 Datenschutzrechtliche Betrachtung von Tracking im Web

Im Kap. 7 haben wir uns mit den Techniken des Tracking im Web beschäftigt. Nun wollen wir eine datenschutzrechtliche Bewertung der unterschiedlichen Techniken vornehmen und aufzeigen, wie ein datenschutzkonformer Einsatz möglich ist.

11.7.1 Cookies

Aus datenschutzrechtlicher Sicht gestaltet sich der Einsatz von Cookies wie folgt. § 15 Abs. 3 Telemediengesetz (TMG) erlaubt eine Nutzungsprofilbildung unter Pseudonym. Nutzer haben das Recht, Widerspruch („Opt-out“) gegen diese Profilbildung einzulegen. Daneben gibt es eine europäische „Cookie-Richtlinie“ (E-Privacy-Richtlinie 2009/136/EG), die von einem „Opt-in“ ausgeht; d. h. Nutzer müssten beim Betreten einer Website der Verwendung von Cookies *einwilligen*. Diese Regelung wurde bislang nicht in deutsches Recht umgesetzt (wobei jedoch umstritten ist, ob die deutsche Rechtslage nicht doch schon den Vorgaben der Richtlinie entspricht). In der Praxis könnte sich das Opt-in (auch in Deutschland) trotzdem durchsetzen. Verantwortlich dafür ist ausgerechnet Google, das die Nutzer seiner Dienste (AdSense und DoubleClick) 2015 dazu aufgefordert hat, den in der Cookie-Richtlinie geforderten Cookie-Hinweisen nachzukommen und diese den Nutzern beim Betreten der Website anzuzeigen und eine ausdrückliche Einwilligung zur Verwendung von Cookies einzuholen. Als Grund für diese unerwartete Maßnahme wird vermutet, dass Google mit dieser Art von „Selbstregulierung“ (zumindest Deutschland betreffend) Sanktionen und schärferen gesetzlichen Vorgaben zuvorkommen möchte.

Diansteanbieter sollten jedenfalls in der Datenschutzerklärung über den Einsatz von Cookies informieren. Dabei sind der Zweck, die Speicherdauer und der Verantwortliche für das Cookie zu nennen. Außerdem muss es eine Widerrufslösung für eine einmal erteilte Einwilligung geben.

11.7.2 Google Analytics

In Abschn. 7.1.2 haben wir festgestellt, dass beim Einsatz von Google Analytics auf Websites personenbezogene Daten von Besuchern an Google übermittelt werden. Nach § 12 Absatz 1 Telemediengesetz (TMG) ist für die *Verarbeitung* personenbezogener Daten, worunter auch die Übermittlung fällt, eine *Einwilligung* der Nutzer erforderlich. Ein weiterer, problematischer Punkt ist, dass bei der Verwendung von Google Analytics personenbezogene Daten in die USA übermittelt werden.

Die deutschen Datenschutzaufsichtsbehörden haben mit Google im Jahr 2011 einen Kompromiss im Streit um Google Analytics geschlossen. Demnach ist der Einsatz von Google Analytics unter bestimmten Voraussetzungen – zumindest im nicht-öffentlichen Bereich, d. h. auf Unternehmens-Websites – zulässig. So ist u. A. der zuvor genannte

JavaScript-Code (vom Website-Betreiber) derart anzupassen, dass eine „Anonymisierung“ der IP-Adresse stattfindet, d. h., dass die letzten 8 Bit der IP-Adresse „entfernt“ werden und von Google nicht gespeichert werden. Dies soll verhindern, dass ein Nutzungsprofil einem einzelnen Nutzer zugeordnet werden kann. Durch die Verkürzung der IP-Adresse ist für den Website-Betreiber immer noch ersichtlich, aus welcher Region ein Nutzer kommt. Damit kann § 15 Absatz 3 TMG Anwendung finden, der besagt, dass Nutzungsprofile nur unter Pseudonym erstellt werden dürfen. Nutzern muss demnach aber auch die Möglichkeit des Widerspruchs gegen die Verwendung von Google Analytics gegeben werden. Darauf muss in der Datenschutzerklärung der Website hingewiesen werden. Außerdem muss ein Vertrag für eine *Auftragsdatenverarbeitung (ADV)* nach § 11 Bundesdatenschutzgesetz (BDSG) zwischen dem Website-Betreiber und Google vorliegen, der die rechtliche Grundlage für das „Outsourcing“ der Zugriffsanalyse an Google bildet.

Die Umsetzung der Vorgaben zum rechtssicheren Einsatz von Google Analytics verlaufen allerdings schleppend. So wurde bei einer Prüfung von 12.205 Websites von in Baden-Württemberg ansässigen Unternehmen im Jahr 2014 festgestellt, dass 2.533 Websites Google Analytics nutzen. Bei rund 65 % wurden Mängel bei der Umsetzung der datenschutzrechtlichen Vorgaben ermittelt.¹⁰

11.7.3 Device Fingerprinting

Die Artikel 29-Gruppe, also der Zusammenschluss der nationalen Datenschutzaufsichtsbehörden auf EU-Ebene, hat sich darauf verständigt, dass Device Fingerprinting unter die EU Cookie-Richtlinie fällt, die wir in Abschn. 11.7.1 im Zusammenhang mit Cookies bereits erwähnt haben. Danach müssen Nutzer über die Anwendung von Device Fingerprinting informiert werden und explizit nach deren Zustimmung (Opt-in) gefragt werden. Da beim Device Fingerprinting keine Cookies zum Einsatz kommen, ist die Anwendung dieses Verfahrens für Nutzer auch in der Regel nicht ersichtlich.

11.7.4 Social Plugins

In Abschn. 7.2 haben wir gesehen, dass Social Plugins auf Websites personenbezogene Daten der Besucher an den Betreiber des sozialen Netzwerks übermitteln. Aus datenschutzrechtlicher Sicht ist diese Übermittlung personenbezogener Daten an das soziale Netzwerk problematisch, zumal es sich in den meisten Fällen um einen Anbieter aus den USA handelt. Nach § 13 Absatz 1 Telemediengesetz (TMG) muss der Nutzer zu Beginn

¹⁰<http://www.baden-wuerttemberg.datenschutz.de/wp-content/uploads/2014/02/Datenschutzverst%C3%B6%C3%9F-fe-bei-Internetauftritten-von-baden-w%C3%BCrttembergischen-Firmen-festgestellt.pdf>

der Nutzung der Website über die Übermittlung der personenbezogenen Daten an einen Dritten außerhalb des Anwendungsbereichs der Richtlinie 95/46/EG, wie es etwa bei Facebook der Fall ist, informiert werden.

Als datenschutzfreundliche Lösung zur Einbindung von Social Plugins existiert neben der *Zwei-Klick-Lösung* auch die *Shariff-Lösung*. Bei der Zwei-Klick-Lösung muss der Besucher einer Website die Social Plugins in einem ersten Schritt zunächst aktivieren, sofern er sie verwenden möchte. Aktiviert ein Nutzer die Social Plugins nicht, findet auch keine Datenübertragung an das soziale Netzwerk statt. Bei der Shariff-Lösung werden die Social Plugins nicht über den vom sozialen Netzwerk bereitgestellten iFrame-Code auf der Website eingebunden, sondern als normale HTML-Links. Solange ein Nutzer diese Links nicht betätigt, findet auch keine Übermittlung personenbezogener Daten an das soziale Netzwerk statt.

11.8 Fazit

Das Datenschutzrecht stellt sich in der Praxis aufgrund der sich ständig weiterentwickelnden und neu aufkommenden Technologien als ein kompliziertes Rechtsgebiet dar. Wie wir in diesem Kapitel gesehen haben, ist der Datenschutz in unterschiedlichen Gesetzen geregelt. Neben den in diesem Kapitel betrachteten Gesetzen spielt der Datenschutz außerdem noch im Sozialgesetzbuch und den Datenschutzgesetzen der Länder (für den öffentlichen Bereich) eine Rolle.

Inwiefern die Datenschutz-Grundverordnung tatsächlich die erhoffte Harmonisierung des Datenschutzrechts auf europäischer Ebene bringen wird, vermag zum Zeitpunkt der Entstehung dieses Buchs noch niemand zu beurteilen. Zudem ist zum jetzigen Zeitpunkt noch nicht geklärt, inwiefern das „Privacy Shield“, als Nachfolgeabkommen von „Safe Harbour“ zum Datentransfer in die USA, endgültig datenschutzrechtlich beurteilt werden wird.

11.9 Übungsaufgaben

Aufgabe 1

Ein Unternehmen gibt Kundenkarten mit Fotos aus. Es konvertiert nun die für diesen Zweck gespeicherten hochauflösenden Bitmaps in stark komprimierte JPEGs mit kleinerer Auflösung. Liegt eine Verarbeitung im datenschutzrechtlichen Sinne vor?

Aufgabe 2

Student S hat Schwierigkeiten, einen Kredit zu bekommen. Er befürchtet, eine Auskunftfeei könne falsche Daten über seine Zahlungsmoral gespeichert haben. S verlangt

von der Auskunftfeier daher eine unentgeltliche Auskunft über die zu seiner Person gespeicherten Daten. Hat er ein Recht darauf?

Aufgabe 3

Mit welchen Anliegen können Sie sich an einen Landesbeauftragten für Datenschutz (z. B. in Baden-Württemberg) wenden und welche Aufgaben hat er sonst?

Aufgabe 4

Wie lassen sich Webserver-Logfiles (bspw. bei Apache) derart gestalten, dass IP-Adressen nicht vollständig, sondern „verschleiert“ gespeichert werden? Überlegen Sie sich Ansätze, die eine De-Anonymisierung der IP-Adresse im Angriffsfall erlaubt bzw. recherchieren Sie, ob Apache bereits technische Lösungen dafür bietet.

Aufgabe 5

Nennen Sie jeweils ein Beispiel, wie die Speicherung von Vorratsdaten a) im Bereich der Mobilkommunikation, b) beim Access Provider eines Internetzugangs und c) im Bereich der Kommunikation per E-Mail umgangen werden kann! (Als Umgehung soll hier bereits zählen, wenn der Zweck der Speicherung nicht erreicht wird).

Aufgabe 6

(Unter welchen Umständen) darf Ihr Telefonanbieter Ihre Daten im Telefonbuch veröffentlichen? Falls ja: Welche Daten?

Literatur

1. Peter Gola und Rudolf Schomerus. *Bundesdatenschutzgesetz: Kommentar*, 11. Auflage. C.H. Beck, 2012.
2. Dominik Leibenger, Frederik Möllers, und Ronald Petrlc. Personenbezug verhindern! Verschleierung von IP-Adressen in Webserver-Logfiles. *Datenschutz-Berater: Informationsdienst der Verlagsgruppe Handelsblatt*, (12):269–270, 2014.
3. Christoph Sorge. Datenschutz in P2P-basierten Systemen: Peer-to-Peer-Netze jenseits des Filesharing. *Datenschutz und Datensicherheit*, 31(2):102–106, 2007.
4. Samuel D. Warren und Louis D. Brandeis. The right to privacy. *Harvard Law Review*, 4(5), 1890.
5. Alan Westin. *Privacy and freedom*. Atheneum Publishers, 1967.

Zusammenfassung

Nachdem wir uns in diesem Lehrbuch mit den unterschiedlichsten Themen beschäftigt haben, ist es nun an der Zeit, ein Fazit zu ziehen. Welche spannenden Aufgaben erwarten uns in Zukunft und wie können wir „Privacy by Design“ in der Praxis umsetzen?

In diesem Lehrbuch haben wir uns mit unterschiedlichen Themengebieten beschäftigt. Wir haben gesehen, dass es an zahlreichen Stellen datenschutzrechtliche Probleme gibt. Außerdem haben wir Techniken und Verfahren kennengelernt, die dafür sorgen, dass der Datenschutz gewahrt bleibt. Dieser von uns präsentierte Ansatz des *technischen Datenschutzes* entspricht weitestgehend dem „Privacy by Design“, das in der kommenden EU-Datenschutzgrundverordnung verankert ist. Als Leser sollten Sie nun sowohl durch die Einblicke in das Datenschutzrecht erkennen können, welche datenschutzrechtliche Probleme bei der Entwicklung von Systemen lauern können, als auch über das nötige technische Know-How verfügen, um diesen Problemen von Beginn an zu begegnen und Systeme datenschutzfreundlich zu gestalten. Es warten, nicht nur aus Sicht des Datenschutzes, spannende Aufgaben auf Entwickler. Die weitere Durchdringung von IT in all unsere Lebensbereiche schreitet weiter voran: vom intelligenten Zuhause über das vernetzte Auto bis hin zur personalisierten Medizin. All diese Verfahren werden nur erfolgreich sein, wenn die Sicherheit und der Datenschutz gewährleistet sind. Wir hoffen, dass wir Ihnen in diesem Lehrbuch das spannende Thema Datenschutz mit all seinen Facetten näher bringen konnten und dass Sie das Erlernte sowohl in Ihrem beruflichen, als auch im privaten Alltag anwenden können.

Sachverzeichnis

A

Abstreitbarkeit, 98
Advanced Encryption Standard, 15
 Counter Mode, 99
Angreifer
 globaler, 46
 lokaler, 46
Anonymisierung, 12, 27, 40, 146, 159, 169
Anonymität, 12, 55, 77, 84
Anonymous Credentials, 135
Anwendungsbereich, 162
Arbeitsbeweis, 83
Auftragsdatenverarbeitung, 169
Auskunft, 154, 167
Auskunftei, 150
Auskunftsanspruch, 151
Authentifizierung, 68, 109
Authentizität, 10, 19, 22, 25
Autorisierung, 71

B

Basic Access Control, 111
Bestandsdaten, 156, 158
Bestimmbarkeit, 163
Bewegungsprofil, 160
Bezahlen, anonymes, 77
 nach Chaum, 78
Big Data, 40
Bitcoin (BTC), 82
Blockchain, 83
Brute Force, 19
BTC, *siehe* Bitcoin
Bundesdatenschutzgesetz, 141, 144

C

Certificate Authority, 23, 24
Chip Authentication, 113, 118
Cookie, 156, 169
 Drittanbieter, 91
 Flash, 92
 HTTP (Hypertext Transfer Protocol),
 90
 Opt-out, 92
 Richtlinie, 169

D

Dark Net, 64
Data Protection, 11
Daten, personenbezogene, 145, 163
 Erheben, 148
 Nutzung, 149
 Verarbeiten, 148
Datenübertragbarkeit, 162
Datenpanne, 157
Datenschutz, 4, 11
Datenschutzerklärung, 156
Datenschutzgrundverordnung, 141, 161
Datenschutzrichtlinie, 141
Datensparsamkeit, 148
De-Anonymisierung, 29
Device Fingerprinting, 94, 170
Diensteanbieter, 68
Differential Privacy, 38, 41
Diffie-Hellman-Verfahren, 20
Direct Anonymous Attestation, 130
Do Not Track, 92
Double Spending, 79

E

E-Mail, 98, 166
eID-Server, 126
eIDAS, 124
Einwilligung, 154, 164, 169
ePA, *siehe* Personalausweis,
elektronischer
ePass, *siehe* Reisepass, elektronischer
Erforderlichkeit, 11
Extended Access Control, 113, 118

F

Facebook
Like, 95, 170
Login, 71
Messenger, 107
Folgenabschätzung, 162

G

Geldkarte, 85
Generalisierung, 33
Google Analytics, 93, 169

H

Hashfunktion, kryptographische, 19, 83
Herausgabe, 167
Hintergrundwissen, 38
HIPPA (Health Insurance Portability and
Accountability Act), 40
History Hijacking, 94

I

Identifizierbarkeit, 12, 31
Identifizierung, 109
Identität, 12
Identitätsmanagement, 12, 67
Identitätsprovider, 68
Instant Messaging, 97
Integrität, 10, 19, 22, 25
Internet Protocol (IP)
IP-Adresse, 163
IPsec, 25
Intervenierbarkeit, 11
IT-Sicherheit, 22

K

k-Anonymität, 32
Kopplungsverbot, 158
Kryptographie, 14

L

l-Diversität, 37
Local Shared Object, 92
Logfile, 156

M

Man-in-the-Middle-Angriff, 21, 24
Message Authentication Code (MAC), 19
Metadaten, 46, 47
Mix, 49
Kaskaden, 51

N

Nicht-Verkettbarkeit, 11
Nichtabstreitbarkeit, 10
Nutzungsdaten, 156
Nutzungsprofil, 154, 157

O

OAuth, 71, 74
Oblivious Transfer, 134
Off-the-record Messaging, 98
Onion, 51
Reply, 54
Router, 55
Routing, 55
OpenID, 69, 74
Connect, 74
Opt-in, 169
Opt-out, 169
Outsourcing, 170

P

Passive Authentication, 110
Password Authenticated Connection
Establishment, 117
Peer-to-Peer-Netz, 82
Perfect Forward Secrecy, 21, 98

Personalausweis, elektronischer (ePA),
116
Personenbezug
absoluter, 164
relativer, 164
Platform for Privacy Preferences (P3P),
95
Prepaid-Karte, 86
Pretty Good Privacy (PGP), 98
Privacy, 11, 140
by Design, 4, 109
Shield, 142
Enhancing Technologies, 4, 11
Private Information Retrieval, 133
Private Set Intersection, 134
Privatsphäre, 140
Profilbildung, 169
Protokollierung, 168
Proxy, 48
Pseudonym, 13
Pseudonymisierung, 13
Pseudonymität, 13
Public Key
Infrastructure, 16, 110, 115
Verfahren, 16

Q

Quasi-Identifikator, 30, 94

R

Randomized Response, 39
Re-Identifizierung, 29
Rechenschaftspflicht, 162
Recht auf Vergessenwerden, 162
Reisepass, elektronischer (ePass), 110
Remarketing, 93
Restricted Identification, 120
RSA, 18
Signatur, 18
Verschlüsselung, 16

S

Säulen des Datenschutzes, 4
Safe Harbor, 142
Same Origin Policy, 91

Sanktionen, 162
Schlüsselaustausch, 20
Score-Wert, 150
Secure
Multiparty Computation, 131
Real-Time Transport Protocol, 106
Sockets Layer, 22
Secure/Multipurpose Internet Mail Extensions
(S/MIME), 98
Selbstbestimmungsrecht, informationelles,
143
Selbstregulierung, 169
Shariff, 171
SIGMA, 102
Signal
Messenger, 104
Protokoll, 104
Signatur
blinde, 18
digitale, 17
elektronische, 124
Gruppen, 130
Ring, 131
Single-Sign-On, 68
Sitzungsschlüssel, 21
Smartcard, 85
Social Plugin, 95, 170
Socialist Millionaire's Protocol, 103
Störerhaftung, 153
Standortdaten, 160
Subsidiaritätsprinzip, 147
Systemdatenschutz, 154

T

Telekommunikationsgesetz, 158
Telemediengesetz, 153
Telemedium, 153
Terminal Authentication, 114, 118
Tor, 55
Bridge, 63
Circuit, 55, 56
Hidden Service, 60
Leaky Pipe, 59
Tracking, 89
Pixel, 93
Transparenz, 11
Transport Layer Security, 22

U

Unentdeckbarkeit, 14
Unlinkability, 14
Unterrichtung, 154
Unverkettbarkeit, 14
Unverknüpfbarkeit, 14

V

Verbindlichkeit, 10, 18
Verfügbarkeit, 10
Verkehrsdaten, 159
Verkehrsflussanalyse, 25, 46, 55
Verschlüsselung
 asymmetrische, 14, 16
 hybride, 98
 symmetrische, 14
Vertraulichkeit, 10, 22, 25
Virtual Private Network, 24, 48
Volkszählungsurteil, 141, 143
Vorratsdatenspeicherung, 142, 159

W

WhatsApp, 103
Widerspruch, 169

Y

Yao
 Garbled Circuit, 132
 Millionaires' Problem, 132

Z

Zensurresistenz, 63
Zero-Knowledge Proof, 134
Zertifikat X.509, 23
Zweckbindung, 11, 149
Zwei-Klick-Lösung, 171